39 Degenerate classical minima and instantons

In this chapter, we study a situation in which instantons play an important role: quantum theories corresponding to classical actions that have non-continuously connected degenerate minima. The simplest examples are provided by one-dimensional quantum systems where the potential has degenerate minima. Classically, the states of minimum energy correspond to a particle sitting at any of the minima of the potential. In the case of symmetric minima, the position of the particle breaks (spontaneously) the symmetry of the system. By contrast, in quantum mechanics (QM) the modulus of the ground-state wave function is expected to be large near all the minima of the potential, as a consequence of *barrier penetration effects*. We illustrate this phenomenon with two typical examples: the double-well potential [393], and the cosine potential, whose periodic structure is closer to field theory examples.

In the context of stochastic dynamics, in Section 39.3, we relate instantons to Arrhenius law. The proof of the existence of instantons relies on an inequality related to supersymmetric structures, and which generalizes to some field theory examples.

In field theory, the problem is more subtle as the study of phase transitions shows. However, the presence of instantons again indicates that the classical minima are connected by quantum tunnelling, and that the symmetry between them is not spontaneously broken [394]. Examples of such a situation are provided, in two dimensions, by the $CP(N-1)$ models and, in four dimensions, by $SU(2)$ gauge theories.

39.1 The quartic double-well potential

We first discuss the Hamiltonian of the quartic double-well potential,

$$H = -\tfrac{1}{2}\left(\mathrm{d}/\mathrm{d}q\right)^2 + V(q\sqrt{g})/g, \quad g > 0\,, \tag{39.1}$$

with (g is a loop expansion parameter),

$$V(q) = \tfrac{1}{2}q^2(1-q)^2. \tag{39.2}$$

The Hamiltonian commutes with the operator P, which acts on wave functions as

$$\mathrm{P}\psi(q) = \psi(g^{-1/2} - q) \;\Rightarrow\; \mathrm{P}^2 = 1, \quad [H, \mathrm{P}] = 0\,.$$

Correspondingly, the potential V has two degenerate minima located at $q = 0$ and at $q = 1/\sqrt{g}$. The symmetry is not essential for the existence of instanton solutions. It is a simplifying feature, which, moreover, is present in several examples of physical interest.

39.1.1 The structure of the ground state

Due to the symmetry of the potential, one can generate a perturbative expansion starting from each of the minima of the potential, and one finds the same expansion to all orders. Therefore, one could conclude that the quantum Hamiltonian has a doubly degenerate ground state, corresponding to two eigenfunctions concentrated, respectively, around each of the classical minima of the potential.

However, due to barrier penetration, the true eigenstates are eigenstates of the reflection operator P, the ground state being an even state.

Quantum Field Theory and Critical Phenomena 5E. Fifth edition. Jean Zinn-Justin, Oxford University Press (2021). © Jean Zinn-Justin. DOI: 10.1093/oso/9780198834625.003.0039

The reflection symmetry cannot be spontaneously broken in QM in the case of regular potentials: correlation functions constructed with a Hamiltonian of this type have, from the point of view of phase transitions, the properties of correlation functions of the one-dimensional Ising model (see Section 14.1).

The partition function. Since one expands in g small first, in the large β limit, the partition function $\operatorname{tr} \mathrm{e}^{-\beta H}$ is dominated by the two lowest eigenvalues E_+ and E_-:

$$\operatorname{tr} \mathrm{e}^{-\beta H} \sim \mathrm{e}^{-\beta H_+} + \mathrm{e}^{-\beta H_-} \sim 2\,\mathrm{e}^{-\beta(E_+ + E_-)/2} \cosh[\beta(E_+ - E_-)/2]$$
$$\text{for } g \to 0\,, \quad \beta \to \infty\,. \tag{39.3}$$

To all orders in g, the partition function only depends on the half sum $\frac{1}{2}(E_+ + E_-)$, and is only sensitive to the non-perturbative difference between the eigenvalues E_+ and E_- at order $(E_+ - E_-)^2$:

$$-\frac{1}{\beta} \ln \operatorname{tr} \mathrm{e}^{-\beta H} = \frac{1}{2}(E_+ + E_-) - \frac{1}{\beta} \ln 2 + O\left[\mathrm{e}^{-\beta}, \beta(E_+ - E_-)^2\right]$$
$$\text{for } g \to 0\,, \quad \beta \to \infty\,. \tag{39.4}$$

By contrast, the difference $(E_+ - E_-)$ dominates the twisted partition function $\operatorname{tr} \mathrm{P}\,\mathrm{e}^{-\beta H}$ (see Section 14.2). Indeed, in the same limits $g \to 0$, and $\beta \to \infty$, one finds

$$\operatorname{tr} \mathrm{P}\,\mathrm{e}^{-\beta H} \sim \mathrm{e}^{-\beta E_+} - \mathrm{e}^{-\beta E_-} \sim -2 \sinh[\beta(E_+ - E_-)/2]\, \mathrm{e}^{-\beta(E_+ + E_-)/2}\,. \tag{39.5}$$
$$\sim -\beta\, \mathrm{e}^{-\beta/2}(E_+ - E_-)\left[1 + O\left(g, \mathrm{e}^{-\beta}\right)\right]. \tag{39.6}$$

Actually, it is convenient to consider the ratio between the quantities (39.3) and (39.5),

$$\operatorname{tr} \mathrm{P}\,\mathrm{e}^{-\beta H} / \operatorname{tr} \mathrm{e}^{-\beta H} \sim -\tfrac{1}{2}\beta(E_+ - E_-)\left[1 + O\left(\mathrm{e}^{-\beta}, (E_+ - E_-)^2\right)\right]. \tag{39.7}$$

The ratio makes it possible to distinguish between a situation in which the ground state is degenerate, and the symmetry spontaneously broken, and a situation in which quantum fluctuations restore the symmetry and lift the degeneracy between the two lowest lying states. Since the ratio vanishes to all orders in perturbation theory, one has to look for non-perturbative effects: they are due here to instantons.

39.1.2 Instanton contributions

The partition function is given by the path integral

$$\mathcal{Z}(\beta) = \int [\mathrm{d}q(t)] \exp\left[-\mathcal{S}(q)\right], \tag{39.8}$$

where

$$\mathcal{S}(q) = \int_{-\beta/2}^{\beta/2} \left[\tfrac{1}{2}\dot{q}^2(t) + V\left(q(t)\sqrt{g}\right)/g\right] \mathrm{d}t\,, \tag{39.9}$$

and the paths satisfy periodic boundary conditions: $q(-\beta/2) = q(\beta/2)$. The path integral representation of the twisted partition function $\operatorname{tr} \mathrm{P}\,\mathrm{e}^{-\beta H}$ only differs by the boundary conditions, which are now $q(-\beta/2) + q(\beta/2) = g^{-1/2}$.

For $g \to 0$, the path integral (39.8) is dominated by the saddle points corresponding to the constant functions $q(t) \equiv 0$ or $q(t) \equiv g^{-1/2}$, and this leads to the usual perturbative expansion.

However, these paths do not contribute to the path integral representation of $\operatorname{tr} \mathrm{P}\, \mathrm{e}^{-\beta H}$ because they do not satisfy the different boundary conditions. This is not too surprising, since the difference $E_+ - E_-$ vanishes to all orders in an expansion in powers of g. Therefore, we have to look for non-constant solutions of the equation of motion, which have a finite action in the infinite β limit. The boundary conditions then impose

$$q(\mp\infty) = 0 \quad \text{and} \quad q(\pm\infty) = g^{-1/2}\,. \tag{39.10}$$

The non-degeneracy of the ground state depends on quantum tunnelling, and the corresponding existence of instanton solutions connecting the two minima of the potential.

In the infinite β limit, in terms of $u(t) = \sqrt{g}q_c(t)$, the Euclidean equation of motion yields

$$-\ddot{u}(t) + V'\big(u(t)\big) = 0 \;\Rightarrow\; \tfrac{1}{2}\dot{u}^2(t) = V\big(u(t)\big),$$

in which the boundary conditions (39.10) have been taken into account. The equation has two one-parameter family of solutions with finite classical action, which we call instanton and anti-instanton when it is necessary to distinguish between them. There are given by

$$q_c^{\pm}(t) = u^{\pm}(t)/\sqrt{g}, \quad \text{with} \quad u^{\pm}(t) = \frac{1}{1+\mathrm{e}^{\mp(t-t_0)}}\,, \tag{39.11}$$

and, therefore,

$$\mathcal{S}(q_c) = \frac{1}{g}\int \mathrm{d}t\,\left[\tfrac{1}{2}\dot{u}^2(t) + V\big(u(t)\big)\right] = \frac{1}{6g}\,. \tag{39.12}$$

Large β expansion. The methods of Section 37.6.1 can be adapted to the present problem. For β large but finite, from equations analogous to equations (37.84) and (37.85), one infers the expansions of the classical energy and action:

$$E(\beta) = -2\,\mathrm{e}^{-\beta} + O\left(\mathrm{e}^{-2\beta}\right), \tag{39.13}$$

$$g\mathcal{S}(q_c) = \tfrac{1}{6} - 2\,\mathrm{e}^{-\beta} + O(\mathrm{e}^{-2\beta}). \tag{39.14}$$

The determinant resulting from the integration around the saddle point can also be evaluated by the method explained in Section 37.5. The only noticeable modification stems from the property that $\dot{q}_c(t)$ has no zero: it corresponds to the ground state of the differential operator $\partial^2\mathcal{S}/\delta q(t)\delta q(t')|_{q=q_c}$, which, therefore, is a positive operator. The final result is real, as expected. We can use the expression (37.83) to obtain it, except that no $1/2i$ factor appears here, and one has to multiply by a factor 2, since the two solutions q_c^+ and q_c^- give identical contributions:

$$\operatorname{tr} \mathrm{P}\, \mathrm{e}^{-\beta H} \sim \frac{2}{\sqrt{\pi g}}\beta\, \mathrm{e}^{-\beta/2}\, \mathrm{e}^{-1/6g}\big(1+O(g)\big), \quad \text{for} \quad g \to 0\,, \quad \beta \to \infty\,. \tag{39.15}$$

From equation (39.7), one then infers the asymptotic behaviour of $E_+ - E_-$ for g small,

$$E_+ - E_- \underset{g\to 0}{=} -\frac{2}{\sqrt{\pi g}}\,\mathrm{e}^{-1/6g}\,(1+O(g))\,. \tag{39.16}$$

The difference is exponentially small in $1/g$, a result that, as expected, for $g \to 0$, vanishes to all orders in g.

39.2 The periodic cosine potential

We now consider the slightly more complicated problem of the Hamiltonian

$$H = -\tfrac{1}{2}\left(\mathrm{d}/\mathrm{d}q\right)^2 + V(q\sqrt{g})/g\,, \tag{39.17}$$

with the periodic potential,

$$V(q) = 1 - \cos q\,. \tag{39.18}$$

Since the potential is periodic, the action has an infinite number of degenerate classical minima. Starting from any of these minima one obtains, to all orders in powers of g, the same perturbative spectrum and, therefore, the quantum Hamiltonian seems to have an infinite number of degenerate ground states. Actually, we know that the spectrum of the Hamiltonian H is continuous and has, at least for g small enough, a band structure: this property, for g small, again is due to barrier penetration.

39.2.1 The structure of the ground state

We introduce the unitary translation operator T, which acting on a wave function $\psi(q)$, translates it by one period,

$$T\psi(q) = \psi(q + 2\pi/\sqrt{g}) \;\Rightarrow\; [T, H] = 0\,. \tag{39.19}$$

Since T commutes with the Hamiltonian, both operators can be diagonalized simultaneously. The eigenvalues of T are pure phases. Each eigenfunction $\psi_N(\varphi, g, q)$ of H, which denote by $|N, \varphi, g\rangle$, is then characterized by a phase $\mathrm{e}^{i\varphi}$, eigenvalue of T:

$$H\,|N, \varphi, g\rangle = \mathcal{E}_N(\varphi, g)\,|N, \varphi, g\rangle\,, \quad T\,|N, \varphi, g\rangle = \mathrm{e}^{i\varphi}\,|N, \varphi, g\rangle\,. \tag{39.20}$$

For $g = 0$, H has the spectrum of the harmonic oscillator, $\mathcal{E}_N(\varphi, 0) = (N + \tfrac{1}{2})$. Moreover, to all orders in an expansion in powers of g, $\mathcal{E}_N(\varphi, g)$ is independent of φ: $\mathcal{E}_N(\varphi, g) \equiv \mathcal{E}_N(g)$. However, beyond the perturbative expansion, due to barrier penetration, for $g \neq 0$, $\mathcal{E}_N(\varphi, g)$ depends on φ and, for g small enough, to each value of N is associated a band.

Globally, the spectrum of H is periodic in φ. Moreover, because, for g small, the bands for different values of N do not overlap, in a band the energy eigenvalue itself is a periodic function of φ, which can be expanded in a Fourier series:

$$\mathcal{E}_N(\varphi, g) = \sum_{-\infty}^{+\infty} \mathcal{E}_N(l, g)\,\mathrm{e}^{il\varphi}, \quad \mathcal{E}_N(l, g) = \mathcal{E}_N(-l, g). \tag{39.21}$$

All coefficients $\mathcal{E}_N(l, g)$, except $\mathcal{E}_N(0, g)$, vanish to all orders in an expansion in g.

We now consider the partition function, which is here $\mathrm{tr}'\,\mathrm{e}^{-\beta H}$. The notation tr' has the following meaning: since the diagonal matrix elements of $\mathrm{e}^{-\beta H}$ in configuration space are periodic functions, we only integrate over one period.

For g small, the large β limit selects the lowest band, and we obtain (see equation (A39.8)):

$$\mathrm{tr}'\,\mathrm{e}^{-\beta H} \underset{\beta\to\infty}{\sim} \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\varphi\; \mathrm{e}^{-\beta\mathcal{E}_0(\varphi, g)}\,. \tag{39.22}$$

Like in the case of the double-well potential, we note that it is difficult to determine the dependence on φ of the energy levels by calculating the partition function. By contrast, we can consider (see equation (A39.8)),

$$\operatorname{tr}' T\,\mathrm{e}^{-\beta H} \underset{\beta\to\infty}{\sim} \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\varphi\ \mathrm{e}^{i\varphi}\,\mathrm{e}^{-\beta\mathcal{E}_0(\varphi,g)}\,. \tag{39.23}$$

To simplify the notation, we focus on the lowest band $N=0$, and define

$$\mathcal{E}_0(\varphi,g)\equiv E(\varphi,g),\quad \mathcal{E}_0(l,g)\equiv E_l(g).$$

For g small, $E(\varphi,g)-E_0(g)$ vanishes faster than any power of g. Therefore, we can expand equation (39.23), for $g\to 0$ and then $\beta\to\infty$, as

$$\operatorname{tr}'\ T\ \mathrm{e}^{-\beta H}\sim \mathrm{e}^{-\beta E_0(g)}\int\frac{\mathrm{d}\varphi}{2\pi}\,\mathrm{e}^{i\varphi}\left[1-\beta\big(E(\varphi,g)-E_0(g)\big)+\cdots\right]. \tag{39.24}$$

The integration over φ selects $E_1(g)$. Therefore,

$$\operatorname{tr}' T\,\mathrm{e}^{-\beta H}\sim -\beta\,\mathrm{e}^{-\beta E_0(g)}\,E_1(g),\quad g\to 0\,,\quad \beta\to\infty\,. \tag{39.25}$$

This equation can be more conveniently rewritten as

$$\operatorname{tr}' T\,\mathrm{e}^{-\beta H}\,/\operatorname{tr}'\,\mathrm{e}^{-\beta H}\sim -\beta E_1(g). \tag{39.26}$$

As explained previously, if E_1 does not vanish this implies that the translation symmetry is not spontaneously broken.

Remark. To evaluate the other Fourier series coefficients $E_2,E_3,\ldots$, for g small, the most convenient method is to consider $\operatorname{tr}' T^k\,\mathrm{e}^{-\beta H}$ for $k=2,3,\ldots$ (see Chapter 42).

39.2.2 The instanton contributions

The path integral representations of the partition function $\operatorname{tr}'\mathrm{e}^{-\beta H}$ and of $\operatorname{tr}' T\,\mathrm{e}^{-\beta H}$ again only differ by the boundary conditions. The operator T has the effect of translating the argument q in the matrix element $\langle q'|\operatorname{tr}'\,\mathrm{e}^{-\beta H}|q\rangle$ before taking the trace:

$$\operatorname{tr}' T\,\mathrm{e}^{-\beta H}=\int_{q(\beta/2)=q(-\beta/2)+2\pi/\sqrt{g}}[\mathrm{d}q(t)]\exp\left[-\mathcal{S}(q)\right], \tag{39.27}$$

$$\mathcal{S}(q)=\int_{-\beta/2}^{\beta/2}\left[\tfrac12\dot q^2(t)+V\big(q(t)\sqrt{g}\big)/g\right]\mathrm{d}t\,. \tag{39.28}$$

We recall that $q(-\beta/2)$ only varies over one period of the potential. For β large and g small, due to the boundary conditions, the path integral is dominated by instanton configurations which connect two consecutive minima of the potential. The explicit solution of the equation of motion is

$$q_c(t)=\frac{4}{\sqrt{g}}\tan^{-1}\mathrm{e}^{(t-t_0)}, \tag{39.29}$$

and the corresponding classical action, in the infinite β limit, is

$$\mathcal{S}(q_c)=8/g\,. \tag{39.30}$$

For all potentials for which the minima can be exchanged by a reflection, the analogue of expression (37.92) is

$$E(\beta) \sim -\,\mathrm{e}^{-\beta}\,\frac{x_0^2}{2}\exp\left[2\int_0^{x_0/2}\left(\frac{1}{\sqrt{2V(x)}}-\frac{1}{x}\right)\mathrm{d}x\right], \tag{39.31}$$

in which x_0 is the location of the other minimum. Applying equation (39.31) to the analogue of equation (37.83), one obtains

$$-\beta\,\mathrm{e}^{-\beta/2}\,E_1(g) \underset{g\to 0}{\sim} \frac{4\beta\,\mathrm{e}^{-\beta/2}}{\sqrt{\pi g}}\,\mathrm{e}^{-8/g}, \tag{39.32}$$

or

$$E_1(g) \underset{g\to 0}{\sim} -\frac{4}{\sqrt{\pi g}}\,\mathrm{e}^{-8/g}\,. \tag{39.33}$$

Without evaluating E_n for $n \geq 2$ explicitly, one verifies that the corresponding boundary conditions for $\operatorname{tr} T^n\,\mathrm{e}^{-\beta H}$, which are

$$q(\beta/2) = q(-\beta/2) + n\frac{2\pi}{\sqrt{g}}\,,$$

then select an instanton solution which for β large has an action $8n/g$. Therefore, E_1 gives the dominant non-perturbative contribution for g small and

$$E(\varphi, g) = E_0(g) - \frac{8}{\sqrt{\pi g}}\,\mathrm{e}^{-8/g}\,[1+O(g)]\cos\varphi + O\left(\mathrm{e}^{-16/g}\right). \tag{39.34}$$

Discussion. The two examples have illustrated that, as anticipated, in a theory in which, at the classical level, a discrete symmetry is spontaneously broken, because the classical potential has degenerate minima, the existence of instantons implies that quantum fluctuations restore the symmetry. However, we have also shown that, by contrast, spontaneous symmetry breaking of discrete symmetries is possible in higher dimensions. Analogous conclusions have been reached on the lattice in Sections 14.1 and 14.2.

Note that, in contrast to discrete symmetries, where quantum fluctuations lead to exponentially small effects in $1/\hbar$, or the equivalent coupling constant, in the case of continuous symmetries, the effects of quantum fluctuations show up already at first order in perturbation theory as a consequence of the Goldstone phenomenon (see Section 19.3).

While in theories in which the dynamical variables live in flat Euclidean space, instantons are associated with a degeneracy of the classical minimum of the potential, this is no longer necessarily the case when the space has curvature or is topologically non-trivial. An example is provided by the cosine potential with compactified space, the coordinate q representing a point on a circle of radius $2\pi/\sqrt{g}$. The Hamiltonian then corresponds to a $O(2)$ rotator in a potential (Section 3.4), or a one-dimensional classical spin chain in a magnetic field. The classical minimum is no longer degenerate, because all minima correspond to one point on the circle. The quantum ground state is equally unique, since the Hilbert space consists in strictly periodic eigenfunctions ($\varphi = 0$). The same instanton solutions still exist, which now start from and return to the same classical minimum, winding around the circle. They are stable because the circle is topologically non-trivial. They generate the same exponentially small corrections, which we have described previously.

Further insight into the problem can be gained by generalizing the Hamiltonian to the $O(N)$ symmetric rotator of Section 3.5 in a potential $1-q_1$. The classical solutions are the same, but the degeneracy and the stability properties are different. For $N>2$, the solutions which wind around the sphere have $(N-2)$ directions of instability. Their contributions have to be discussed in the context of the large-order behaviour of perturbation theory (see Chapter 40).

39.3 Instantons and stochastic dynamics

We now describe the role of instantons in the context of stochastic dynamics. We consider the problem of evaluating the decay probability of a metastable state by thermal fluctuations. At first, one could think that this topic should be discussed in Chapter 38, simultaneously with the problem of decay by quantum fluctuations. We show here that, technically, the problem has a more direct relation with degenerate classical minima. At the end of the section, we also briefly examine the role of instantons when the equilibrium distribution has degenerate minima.

39.3.1 Random walk

We consider the Langevin equation (34.34) (see Sections 34.4–34.6),

$$\dot{\mathbf{q}}(t)=-\tfrac{1}{2}\Omega\nabla E\bigl(\mathbf{q}(t)\bigr)+\boldsymbol{\nu}(t), \tag{39.35}$$

with the Gaussian white noise distribution (34.3) defined by (Ω is a positive constant)

$$\langle\nu_i(t)\rangle_\nu=0\,,\quad \langle\nu_i(t)\nu_j(t')\rangle_\nu=\Omega\,\delta_{ij}\,\delta(t-t'). \tag{39.36}$$

To the Langevin equation is associated the *Hermitian* Hamiltonian (34.39),

$$H=\tfrac{1}{2}\Omega\left[-\nabla^2+\tfrac{1}{4}\left(\nabla E(\mathbf{q})\right)^2-\tfrac{1}{2}\nabla^2 E(\mathbf{q})\right]. \tag{39.37}$$

Observables can be calculated with a path integral with the functional measure $\mathrm{e}^{-\mathcal{S}(\mathbf{q})}[\mathrm{d}q]$, where the corresponding dynamic action is given by (equation (34.61), with $\mathcal{S}/\Omega\mapsto\mathcal{S}$),

$$\mathcal{S}(\mathbf{q})=\tfrac{1}{2}\Omega^{-1}\int\left\{\dot{\mathbf{q}}^2(t)+\tfrac{1}{4}\Omega^2\left[\nabla E\bigl(\mathbf{q}(t)\bigr)\right]^2-\tfrac{1}{2}\Omega^2\nabla^2 E\bigl(\mathbf{q}(t)\bigr)\right\}\mathrm{d}t\,. \tag{39.38}$$

The classical limit here is replaced by the small Ω and thus low temperature limit. At leading order in a semi-classical analysis, the term $\Omega^2\nabla^2 E$ can be omitted. Therefore, the classical minima of the action (39.38) correspond to all points where $\nabla E(\mathbf{q})$ vanishes, thus all critical points (extrema or saddle points) of the function $E(\mathbf{q})$. If more than one critical point can be found, the classical equations may have instanton solutions.

Examples. We consider an analytic function $E(\mathbf{q})$ which has only a relative minimum at $\mathbf{q}=0$:

$$E(\mathbf{q})=\tfrac{1}{2}\omega^2\mathbf{q}^2+O\left(|\mathbf{q}|^3\right).$$

Then the function $E(\mathbf{q})$ must also have elsewhere a saddle point or a relative maximum. Physically, we then know that if we put a particle at time 0 at the relative minimum $\mathbf{q}=0$, then after some time the particle will escape from the well, as a result of thermal fluctuations, as described by the Langevin equation. The problem is to evaluate, in the small Ω limit, the escape probability per unit time, or the average escape time τ.

A class of examples corresponds to functions such that the distribution $\mathrm{e}^{-E(\mathbf{q})}$ is not normalizable, like in one dimension

$$E(q) = q^2 - 2q^3/3 \,. \tag{39.39}$$

We then know that τ is the inverse of the smallest eigenvalue of the Hamiltonian (39.37) (see Section 34.3.1), and this eigenvalue is strictly positive.

However, in all examples to all orders in a perturbative expansion in powers of Ω starting from the saddle point $\mathbf{q} = 0$, the function $\mathrm{e}^{-E(\mathbf{q})/2}$ is the formal ground state eigenvector associated with the eigenvalue 0. It follows that the calculation of the eigenvalue is not perturbative. We now show that the instantons connecting the critical points of the function $E(\mathbf{q})$ provide a solution to the problem [395].

Instantons. An instanton connects the minimum at $\mathbf{q} = 0$ to another critical point $\mathbf{q}_0$ where ∇E vanishes. For an instanton solution $\mathbf{q}_c$ the following inequality holds,

$$\int_{-\infty}^{+\infty} \mathrm{d}t \left[\dot{\mathbf{q}}_c(t) \pm \tfrac{1}{2}\Omega\nabla E\big(\mathbf{q}_c(t)\big)\right]^2 \geq 0 \tag{39.40}$$

and, therefore,

$$\mathcal{S}(\mathbf{q}_c) \geq |Q(\mathbf{q_c})| \,, \tag{39.41}$$

with

$$Q(\mathbf{q_c}) = \tfrac{1}{2}\int_{-\infty}^{+\infty} \mathrm{d}t\, \dot{\mathbf{q}}_c(t)\cdot\nabla E\big(\mathbf{q}_c(t)\big) = \tfrac{1}{2}\big(E(\mathbf{q}_0) - E(0)\big). \tag{39.42}$$

We conclude that the action satisfies [396]

$$\mathcal{S}(\mathbf{q}_c) \geq \tfrac{1}{2}\big|E(\mathbf{q}_0) - E(0)\big|. \tag{39.43}$$

The equality corresponds to a local minimum of the action and $\mathbf{q}_c$ then is a solution of a first-order differential equation,

$$\dot{\mathbf{q}}(t) = \pm\tfrac{1}{2}\Omega\nabla E\big(\mathbf{q}(t)\big). \tag{39.44}$$

However, this is not the end of the story. Indeed, in Section 34.5, we have shown that the degeneracy between the minima and maxima of the function $E(\mathbf{q})$ is lifted by the first quantum correction. Therefore, the two minima of the action are not really degenerate, and no instanton can connect them. What really happens is that we have to only consider closed trajectories passing through the origin. If we consider a finite time interval β, we can find such trajectories. In the infinite β limit, they decompose into a succession of instantons and anti-instantons. The limit of the classical action is an even multiple of the instanton action. The leading contribution thus is (Arrhenius law)

$$\mathcal{S}(\mathbf{q}_c) = E(\mathbf{q}_0) - E(0). \tag{39.45}$$

We conclude, quite generally, that if the function $E(\mathbf{q})$ has a relative minimum where $E = E_{\min}$, separated from a lower minimum (possibly $E = -\infty$) by a local maximum $E = E_{\max}$, then the eigenvalue corresponding to an eigenfunction concentrated around the first minimum is of the order $\mathrm{e}^{-\Delta E}$, in which ΔE is the variation of the function E:

$$\Delta E = E_{\max} - E_{\min} \,.$$

The time $\tau = O(\mathrm{e}^{\Delta E})$ characterizes the exponential decay of the probability of finding $\mathbf{q}(t)$ near the origin when the initial conditions at $t = 0$ are $\mathbf{q}(t = 0) = 0$.

Finally, in order to complete the calculation of the eigenvalue, it is necessary to use multi-instanton techniques of the kind explained in Chapter 42.

Note that the inequality (39.40), which we have used, corresponds to a structure that is typical for *supersymmetric* theories (*cf.*, the action (34.74)).

Degenerate minima. A new problem arises when the function $E(\mathbf{q})$ has a degenerate minimum. Let us assume that the corresponding distribution is normalizable. Then the ground state eigenvalue vanishes. The interesting question is how to calculate the difference between the two first eigenvalues, difference which vanishes to all orders in perturbation theory. This is the problem we have solved in Section 39.1 for the double-well potential. However, here the set-up is slightly different because, if $E(\mathbf{q})$ is regular, as we always assume, the two minima are necessarily separated by a maximum or a saddle point and, therefore, $(\nabla E)^2$ has at least three minima. A one-dimensional example is

$$E(q) = q^2(1-q)^2 \tag{39.46}$$

and, therefore,

$$E'^2(q) = 4(1-2q)^2 q^2 (1-q)^2 . \tag{39.47}$$

This time we look for instanton solutions connecting $q = 0$ to $q = 1$. However, in the infinite time limit, only instantons which go from 0 to 1/2 or 1/2 to 1 survive. From the analysis of the previous problem, we guess that the relevant configurations correspond to gluing together two instantons. Therefore, the difference between the two leading eigenvalues, which is also the second eigenvalue ϵ_1, is again of the form

$$\ln \epsilon_1 = -\ln \tau \sim -\left(E_{\max} - E_{\min}\right) . \tag{39.48}$$

in which $E_{\min}$ and $E_{\max}$ are respectively the values of the function $E(q)$ at the degenerate minima and at the maximum which connects them.

39.3.2 Quantum field theory (QFT)

We consider a dynamics governed by a purely dissipative Langevin equation (equation (35.35)), which formally converges towards an equilibrium distribution corresponding to the field integral of a d-dimensional Euclidean QFT. Moreover, the field theory, with the Euclidean action

$$\mathcal{A}(\phi) = \int \mathrm{d}^d x \left[\tfrac{1}{2} \big(\nabla\phi(x)\big)^2 + \tfrac{1}{2} m^2 \phi^2(x) + \tfrac{1}{6} g \phi^3(x) \right] , \tag{39.49}$$

has a metastable minimum at $\phi(x) = 0$. We know that a quantum state concentrated around the minimum will decay due to quantum fluctuations, and we have calculated the rate by instanton methods.

We now want to evaluate the decay probability due to thermal fluctuations. The relevant dynamic action, which is now $(d+1)$ dimensional, at leading order, reads:

$$\mathcal{S}(\phi) = \tfrac{1}{2} \int \mathrm{d}^d x \, \mathrm{d}t \left\{ \frac{\dot{\phi}^2(t,x)}{\Omega} + \tfrac{1}{4}\Omega \left[-\nabla_x^2 \phi(t,x) + m^2 \phi(t,x) + \tfrac{1}{2} g \phi^2(t,x) \right]^2 \right\} . \tag{39.50}$$

Formally, the discussion follows the same line as in the case of a finite number of degrees of freedom. The problem is to identify the minimum and the maximum of the action.

The minimum is easy to find: $\phi \equiv 0$. The maximum requires some more thought. It does not correspond to a constant field configuration: $\phi(x) = -2m^2/g$. Indeed, it is sufficient that some part of the field starts passing the barrier. Instead, the relevant maximum of the action corresponds to a static instanton configuration. The arguments of Section 39.2.2, then lead to the estimate $\ln \tau \sim \exp\left(\mathcal{S}_{\text{inst.}} - \mathcal{S}(\phi = 0)\right)$, where $\mathcal{S}_{\text{inst.}}$ is the instanton action.

39.4 Instantons in stable boson field theories: General remarks

We now briefly study the existence of instantons in stable theories, connecting, for example, degenerate classical minima. The most interesting examples correspond, unfortunately, to scale invariant classical theories. The evaluation of the instanton contributions at leading order, which formally follows the method described in Chapter 38, leads to difficulties due both to ultraviolet (UV) and infrared (IR) divergences. Some of them are examined in Chapter 38. Since, for the two examples we consider in Sections 39.5 and 39.6, they have not been satisfactorily solved, we restrict ourselves here to semi-classical considerations.

We begin with a few general remarks about the possible existence of instantons in stable field theories.

Scalar field theories. First, we assume that the action for a multicomponent scalar boson field ϕ^i has the form

$$\mathcal{S}(\phi) = \int \left[K\big(\boldsymbol{\phi}(x)\big) + V\big(\boldsymbol{\phi}(x)\big)\right] \mathrm{d}^d x\,, \tag{39.51}$$

with

$$K\big(\boldsymbol{\phi}(x)\big) = \tfrac{1}{2}\sum_{i,j} g_{ij}\big(\boldsymbol{\phi}(x)\big)\nabla\phi^i(x)\nabla\phi^j(x),$$

in which $g_{ij}(\boldsymbol{\phi})$ a positive matrix (positive definite almost everywhere) and

$$\min_{\{\boldsymbol{\phi}\}} V(\boldsymbol{\phi}) = 0\,. \tag{39.52}$$

Equation (A38.3) immediately generalizes to

$$(2-d)\int K\big(\boldsymbol{\phi}(x)\big)\mathrm{d}^d x = d\int V\big(\boldsymbol{\phi}(x)\big)\mathrm{d}^d x\,.$$

This equation has no solution for $d > 2$. For $d = 2$, it has solutions only if

$$V\big(\boldsymbol{\phi}_c(x)\big) = 0\,. \tag{39.53}$$

The condition (39.52) then implies that $\boldsymbol{\phi}_c(x)$ is for all x a minimum of the potential,

$$\frac{\partial V(\boldsymbol{\phi}_c)}{\partial \phi} = 0\,,$$

and, therefore, $\boldsymbol{\phi}_c(x)$ is a solution of the field equations

$$\frac{\delta}{\partial \phi^i(y)}\int K\big(\boldsymbol{\phi}(x)\big)\mathrm{d}^2 x = 0\,.$$

These two equations are in general incompatible, except if $V(\boldsymbol{\phi})$ vanishes identically. In the latter case, the action (39.51) corresponds to a two-dimensional model on a Riemannian manifold. A particular class of such models based on homogeneous spaces has been discussed in Chapters 19 and 29. Among them, the $CP(N-1)$ models are known to admit instanton solutions, which are described in Section 39.5.

Gauge theories. If, in addition to scalar fields, the theory contains gauge fields A^a_μ (see Chapter 22), the gauge invariant action has the form (D_μ is the covariant derivative)

$$\mathcal{S}(\boldsymbol{\phi},\mathbf{A})=\mathcal{S}(\mathbf{A})+\Sigma(\boldsymbol{\phi},\mathbf{A})+\int \mathrm{d}^d x\, V\bigl(\boldsymbol{\phi}(x)\bigr), \tag{39.54}$$

with $V(\phi)\geq 0$, and

$$\mathcal{S}(\mathbf{A})=\sum_{a,\mu,\nu}\frac{1}{4g}\int \mathrm{d}^d x\, F^a_{\mu\nu}(x)F^a_{\mu\nu}(x),\ \text{with } g>0\,,\quad \Sigma(\boldsymbol{\phi},\mathbf{A})=\tfrac{1}{2}\int \mathrm{d}^d x\sum_{\mu,i}\bigl(\mathrm{D}_\mu\phi_i(x)\bigr)^2.$$

Assuming the existence of a finite action solution $\{\boldsymbol{\phi}^c,\ \mathbf{A}^c_\mu\}$ (in which $\mathbf{A}^c_\mu$ is not a pure gauge), one calculates the action for $\lambda\mathbf{A}^c_\mu(\lambda x)$ and $\boldsymbol{\phi}^c(\lambda x)$. After the change of variables $\lambda x\mapsto x$, one obtains

$$\mathcal{S}(\boldsymbol{\phi}^c,\mathbf{A}^c;\lambda)=\lambda^{4-d}\mathcal{S}(\mathbf{A}^c)+\lambda^{2-d}\Sigma(\boldsymbol{\phi}^c,\mathbf{A}^c)+\lambda^{-d}\int V\bigl(\boldsymbol{\phi}^c(x)\bigr)\mathrm{d}^d x\,. \tag{39.55}$$

Stationarity at $\lambda=1$ implies

$$(4-d)\mathcal{S}(\mathbf{A}^c)+(2-d)\Sigma(\boldsymbol{\phi}^c,\mathbf{A}^c)-d\int V\bigl(\boldsymbol{\phi}^c(x)\bigr)\mathrm{d}^d x=0\,. \tag{39.56}$$

We see that no solution can exist for $d>4$, since a sum of negative terms cannot vanish.

For $d=4$, we find two conditions:

$$V\bigl(\boldsymbol{\phi}^c(x)\bigr)=0\,, \tag{39.57a}$$
$$\mathrm{D}_\mu\boldsymbol{\phi}^c(x)=0\,. \tag{39.57b}$$

Applying these conditions to the field equations, we conclude that $\mathbf{A}^c_\mu$ is the solution of the pure gauge field equations. As we show in Section 39.6, instantons can indeed be found in pure gauge theories. Equation (39.57*b*), which now is an equation for $\boldsymbol{\phi}^c$, then leads to the integrability conditions:

$$[\mathrm{D}_\mu,\mathrm{D}_\nu]=F_{\mu\nu}\Rightarrow\sum_{a,j}\bigl(F^a_{\mu\nu}(x)\bigr)^c t^a_{ij}\phi^c_j(x)=0\,, \tag{39.58}$$

in which the matrices $\mathbf{t}^a$ are the generators of the Lie algebra. The conditions (39.58) together with the equation (39.57*a*) show that, in general, the system has only the trivial solution $\boldsymbol{\phi}^c(x)=0$.

39.5 Instantons in $CP(N-1)$ models

The preceding considerations can be illustrated by the two-dimensional $CP(N-1)$ models. We mainly describe the nature of the instanton solutions and refer the reader to the literature for a detailed discussion [397].

We consider a set of N complex fields φ_α, subject to the condition

$$\bar{\boldsymbol{\varphi}}(x) \cdot \boldsymbol{\varphi}(x) = 1\,. \tag{39.59}$$

Moreover, two vectors $\boldsymbol{\varphi}$ and $\boldsymbol{\varphi}'$ are equivalent if they are related by the $U(1)$ gauge transformation,

$$\boldsymbol{\varphi}'(x) = \mathrm{e}^{i\Lambda(x)}\,\boldsymbol{\varphi}(x)\,, \tag{39.60}$$

where $\Lambda(x)$ is an arbitrary real function. These conditions characterize the manifold $CP(N-1)$ (for $(N-1)$-dimensional Complex Projective), which is isomorphic to the complex Grassmannian manifold $U(N)/U(1)/U(N-1)$, one of the symmetric spaces exhibited in Section A29.4.3. One form of the unique classical action is

$$\mathcal{S}(\boldsymbol{\varphi}, A) = \frac{1}{g}\sum_\mu \int \mathrm{d}^2x\, \overline{\mathrm{D}_\mu\boldsymbol{\varphi}}(x)\cdot \mathrm{D}_\mu\boldsymbol{\varphi}(x), \quad g>0\,, \tag{39.61}$$

in which D_μ is the covariant derivative:

$$\mathrm{D}_\mu = \partial_\mu + iA_\mu\,. \tag{39.62}$$

The gauge field A_μ implements the invariance of the action under the $U(1)$ gauge transformations (39.60). Since the action contains no kinetic term for the gauge field A_μ, A_μ is an auxiliary field that can be integrated out. The integral is Gaussian, and the result is obtained by replacing in the action A_μ by the solution of the A_μ-field equation, $\delta\mathcal{S}/\delta A_\mu(x)=0$. Using equation (39.59), one finds

$$A_\mu(x) = i\bar{\boldsymbol{\varphi}}(x)\cdot\partial_\mu\boldsymbol{\varphi}(x). \tag{39.63}$$

After this substitution $\bar{\boldsymbol{\varphi}}(x)\cdot\partial_\mu\boldsymbol{\varphi}(x)$ plays the role of a composite gauge field.

Instantons. A proof of the existence of locally stable non-trivial minima of the action follows from the inequality (note the analogy with equation (39.40))

$$\sum_\mu \int \mathrm{d}^2x \left| \mathrm{D}_\mu\boldsymbol{\varphi}(x) \mp i\sum_\nu \epsilon_{\mu\nu}\mathrm{D}_\nu\boldsymbol{\varphi}(x)\right|^2 \ge 0\,, \tag{39.64}$$

($\epsilon_{\mu\nu}=-\epsilon_{\nu\mu}$ and $\epsilon_{12}=1$). Expanding the expression, one obtains

$$\mathcal{S}(\boldsymbol{\varphi}) \ge |Q(\boldsymbol{\varphi})|/g\,, \tag{39.65}$$

with

$$Q(\boldsymbol{\varphi}) = -i\sum_{\mu,\nu}\epsilon_{\mu\nu}\int \mathrm{d}^2x\,\mathrm{D}_\mu\boldsymbol{\varphi}(x)\cdot\overline{\mathrm{D}_\nu\boldsymbol{\varphi}}(x) = i\sum_{\mu,\nu}\int \mathrm{d}^2x\,\epsilon_{\mu\nu}\mathrm{D}_\nu\mathrm{D}_\mu\boldsymbol{\varphi}(x)\cdot\bar{\boldsymbol{\varphi}}(x), \tag{39.66}$$

after an integration by parts. Then, in the representation (39.61)

$$i\sum_{\mu,\nu}\epsilon_{\mu\nu}\mathrm{D}_\nu\mathrm{D}_\mu = \tfrac{1}{2}i\sum_{\mu,\nu}\epsilon_{\mu\nu}[\mathrm{D}_\nu,\mathrm{D}_\mu] = -\tfrac{1}{2}\sum_{\mu,\nu}\epsilon_{\mu\nu}F_{\mu\nu}\,, \tag{39.67}$$

where

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x)$$

is the curvature.

Therefore, using equation (39.59), one finds

$$Q(\boldsymbol{\varphi}) = -\frac{1}{2}\sum_{\mu,\nu}\int \mathrm{d}^2x\, \epsilon_{\mu\nu}F_{\mu\nu}(x). \tag{39.68}$$

The integrand is proportional to the two-dimensional Abelian chiral anomaly (see the general expression (23.92), or equation ($A30.3$)), and thus is a total divergence:

$$\tfrac{1}{2}\sum_{\mu,\nu}\epsilon_{\mu\nu}F_{\mu\nu}(x) = \sum_{\mu,\nu}\partial_\mu\epsilon_{\mu\nu}A_\nu(x).$$

Substituting this form into equation (39.68), and integrating in a large disk of radius R, one infers

$$Q(\boldsymbol{\varphi}) = -\lim_{R\to\infty}\oint_{|x|=R}\mathrm{d}\mathbf{x}\cdot\mathbf{A}(x).$$

$Q(\boldsymbol{\varphi})$ thus only depends on the behaviour of the classical solution for $|x|$ large and is a *topological charge*. Finiteness of the action demands that at large distances $\mathrm{D}_\mu\boldsymbol{\varphi}$ vanishes. Equation (39.67) then implies that $F_{\mu\nu}$ vanishes, and thus A_μ is a pure gauge (and $\boldsymbol{\varphi}$ a gauge transform of a constant vector),

$$A_\mu(x) = \partial_\mu\Lambda(x) \ \Rightarrow\ Q(\boldsymbol{\varphi}) = -\lim_{R\to\infty}\oint_{|x|=R}\mathrm{d}\mathbf{x}\cdot\nabla\Lambda(x)\,. \tag{39.69}$$

The topological charge measures the variation of the angle $\Lambda(x)$ on a large circle, which necessarily is a multiple of 2π because $\boldsymbol{\varphi}$ is regular. One is thus led to the consideration of the homotopy classes of mappings from $U(1)$, that is, S_1 (the circle) to S_1, which are characterized by an integer n, the *winding number*, and

$$Q(\boldsymbol{\varphi}) = 2\pi n \ \Rightarrow \mathcal{S}(\boldsymbol{\varphi}) \ge 2\pi|n|/g\,. \tag{39.70}$$

The equality $\mathcal{S}(\boldsymbol{\varphi}) = 2\pi|n|/g$ corresponds to a local minimum, and implies that the classical solutions satisfy the first-order partial differential (self-duality) equations

$$\mathrm{D}_\mu\boldsymbol{\varphi}(x) = \pm i\sum_\nu \epsilon_{\mu\nu}\mathrm{D}_\nu\boldsymbol{\varphi}(x). \tag{39.71}$$

It can be shown that, in the variable $z = x_1 + ix_2$, the solutions of the equations (39.71) are proportional to holomorphic or anti-holomorphic (depending on the sign) vectors (this reflects the conformal invariance of the classical field theory). Using then equation (39.59) and a gauge transformation (39.60), one can cast the holomorphic solution into the form

$$\varphi_\alpha(z) = P_\alpha(z)/\left[P(z)\cdot\bar{P}(z)\right]^{1/2}\,, \tag{39.72}$$

where the $P_\alpha(z)$ are polynomials in z without common roots. The anti-holomorphic solution corresponds to interchanging $\boldsymbol{\varphi}$ and $\bar{\boldsymbol{\varphi}}$.

The semi-classical vacuum. In contrast to the method used in Sections 39.1 and 39.2, the existence of instantons has been discussed here without reference to the structure of the classical vacuum. To find an interpretation of instantons in gauge theories, it is convenient to express the results in the temporal gauge.

39.5.1 The semi-classical vacuum: Temporal gauge

In the temporal gauge, classical minima of the potential correspond to fields $\boldsymbol{\varphi}(x_1)$, where x_1 is only the space variable, gauge transforms of a constant vector:

$$\boldsymbol{\varphi}(x_1) = \mathrm{e}^{i\Lambda(x_1)}\,\mathbf{v}\,, \quad \bar{\mathbf{v}}\cdot\mathbf{v} = 1\,.$$

If the vacuum state is invariant under space reflection, then $\boldsymbol{\varphi}(+\infty) = \boldsymbol{\varphi}(-\infty)$, and thus

$$\Lambda(+\infty) - \Lambda(-\infty) = 2n\pi \quad n \in \mathbb{Z}\,.$$

The integer n is a topological number that classifies the degenerate classical minima, and the semi-classical vacuum thus has a periodic structure. This analysis is consistent with Gauss's law (Section 22.3), which only implies that states are invariant under infinitesimal gauge transformations and, therefore, under gauge transformations of the class $n = 0$, which are continuously connected to the identity.

We now consider a large rectangle with an extension R in the space direction and T in the Euclidean time direction, and by a smooth gauge transformation continue the instanton solution to the temporal gauge. Then, the variation of the pure gauge comes entirely from the sides with $x_2 = 0$ and $x_2 = T$. One finds for $R \to \infty$,

$$\Lambda(+\infty, 0) - \Lambda(-\infty, 0) - [\Lambda(+\infty, T) - \Lambda(-\infty, T)] = 2n\pi\,.$$

Therefore, instantons interpolate between different classical minima. One can project onto a proper quantum eigenstate, the θ-vacuum, corresponding to an angle θ, by adding a topological term to the classical action,

$$\mathcal{S}(\boldsymbol{\varphi}) \mapsto \mathcal{S}(\boldsymbol{\varphi}) + i\frac{\theta}{4\pi}\sum_{\mu,\nu}\int \mathrm{d}^2x\,\epsilon_{\mu\nu}F_{\mu\nu}(x),$$

in analogy with expressions (23.40), or as in the example of the cosine potential (42.61, 42.60).

The $CP(1)$ model. The $CP(1)$ model is locally isomorphic to the $O(3)$ non-linear σ-model, with the identification

$$\boldsymbol{\phi}(x) = \sum_{\alpha,\beta}\bar{\varphi}_\alpha(x)\boldsymbol{\sigma}_{\alpha\beta}\varphi_\beta(x), \tag{39.73}$$

where σ_i are the three Pauli matrices. In the $O(3)$ σ-model, the $CP(1)$ minimal instanton solution becomes the stereographic mapping of the sphere S_2 onto the plane [398]:

$$P_1(z) = z\,,\ P_2(z) = 1 \ \Rightarrow\ \phi_1 = \frac{z+\bar{z}}{1+\bar{z}z}\,, \quad \phi_2 = i\frac{z-\bar{z}}{1+\bar{z}z}\,, \quad \phi_3 = \frac{1-\bar{z}z}{1+\bar{z}z}\,.$$

39.6 Instantons in the $SU(2)$ gauge theory

In four dimensions, non-Abelian gauge theories provide an example of instantons related to the vacuum structure of quantum chromodynamics (QCD, see Chapter 22) [399]. According to the analysis of Section 39.4, we can consider only pure gauge theories. Actually, it is sufficient to consider the gauge group $SU(2)$, since a general theorem states that for a Lie group containing $SU(2)$ as a subgroup the instantons are those of the $SU(2)$ subgroup.

In $SO(3)$ notation, the gauge field $\mathbf{A}_\mu$ is a vector and the gauge action reads

$$\mathcal{S}(\mathbf{A}) = \frac{1}{4g} \sum_{\mu,\nu} \int \left[\mathbf{F}_{\mu\nu}(x)\right]^2 \mathrm{d}^4 x\,, \tag{39.74}$$

with

$$\mathbf{F}_{\mu\nu}(x) = \partial_\mu \mathbf{A}_\nu(x) - \partial_\nu \mathbf{A}_\mu(x) + \mathbf{A}_\mu(x) \times \mathbf{A}_\nu(x). \tag{39.75}$$

Generalizing the arguments used for the $CP(N-1)$ model, one derives the existence, and some properties, of instantons in this theory.

The dual of the tensor $\mathbf{F}_{\mu\nu}$ is defined by

$$\tilde{\mathbf{F}}_{\mu\nu}(x) = \tfrac{1}{2} \sum_{\rho,\sigma} \epsilon_{\mu\nu\rho\sigma} \mathbf{F}_{\rho\sigma}(x). \tag{39.76}$$

Then the inequality [396]

$$\sum_{\mu,\nu} \int \mathrm{d}^4 x \left[\mathbf{F}_{\mu\nu}(x) \pm \tilde{\mathbf{F}}_{\mu\nu}(x)\right]^2 \geq 0\,, \tag{39.77}$$

implies,

$$\mathcal{S}(\mathbf{A}_\mu) \geq |Q(\mathbf{A}_\mu)|/4g\,, \tag{39.78}$$

where $Q(\mathbf{A})$ is an expression one also meets in Section 23.6.3 (equation (23.95), here written in $SO(3)$ notation) in the calculation of the axial anomaly

$$Q(\mathbf{A}) = \int \mathrm{d}^4 x \sum_{\mu,\nu} \mathbf{F}_{\mu\nu}(x) \cdot \tilde{\mathbf{F}}_{\mu\nu}(x). \tag{39.79}$$

There it is shown that the quantity $\sum_{\mu,\nu} \mathbf{F}_{\mu\nu} \cdot \tilde{\mathbf{F}}_{\mu\nu}$ is a pure divergence (equation (23.96)). Indeed, one verifies that

$$\sum_{\mu,\nu} \mathbf{F}_{\mu\nu}(x) \cdot \tilde{\mathbf{F}}_{\mu\nu}(x) = \nabla \cdot \mathbf{V}(x), \tag{39.80}$$

with

$$V_\mu(x) = 2 \sum_{\nu,\rho,\sigma} \epsilon_{\mu\nu\rho\sigma} \left[\mathbf{A}_\nu(x) \cdot \partial_\rho \mathbf{A}_\sigma(x) + \tfrac{1}{3} \mathbf{A}_\nu(x) \cdot \left(\mathbf{A}_\rho(x) \times \mathbf{A}_\sigma(x)\right)\right]. \tag{39.81}$$

The integral thus only depends on the behaviour of the gauge field at large distances, and its values are quantized (equation (23.102)). Here again, as in the example of the $CP(N-1)$ model, the bound involves a topological charge, $Q(\mathbf{A}_\mu)$.

The finiteness of the action implies that the classical solution must asymptotically become a pure gauge, that is, with our conventions,

$$-\tfrac{1}{2} i \mathbf{A}_\mu(x) \cdot \boldsymbol{\sigma} = \mathbf{g}(x) \partial_\mu \mathbf{g}^{-1}(x) + O\left(|x|^{-2}\right) \quad |x| \to \infty\,, \tag{39.82}$$

in which $\boldsymbol{\sigma}$ are Pauli matrices and $\mathbf{g}(x)$ is an element of $SU(2)$.

Since $SU(2)$ is topologically equivalent to S_3, we are now led to consider the homotopy classes of mappings from S_3 to S_3, which are also classified by an integer, the *winding number*. The one to one mapping corresponds to an element of the form

$$\mathbf{g}(x) = \frac{x_4 + i\mathbf{x}\cdot\boldsymbol{\sigma}}{r}, \quad r = (x_4^2 + \mathbf{x}^2)^{1/2}, \tag{39.83}$$

and, thus,

$$A^i_m(x) \underset{r\to\infty}{\sim} 2\left(x_4\delta_{im} + \sum_k \epsilon_{imk}x_k\right)\frac{1}{r^2}, \quad \text{for } m \le 3, \quad A^i_4(x) = -2\frac{x_i}{r^2}. \tag{39.84}$$

It follows that

$$\int \mathrm{d}^4x \sum_{\mu,\nu} \mathbf{F}_{\mu\nu}(x)\cdot\tilde{\mathbf{F}}_{\mu\nu}(x) = \int \mathrm{d}\Omega\,\hat{\mathbf{n}}\cdot\mathbf{V}(x) = 32\pi^2, \tag{39.85}$$

where $\mathrm{d}\Omega$ is the invariant measure on the sphere, and $\hat{\mathbf{n}}$ the unit vector normal to surface of the sphere.

Comparing this result with equation (23.102), we note that we have indeed found the minimal action solution. In general, we then expect

$$\int \mathrm{d}^4x\,\mathbf{F}_{\mu\nu}(x)\cdot\tilde{\mathbf{F}}_{\mu\nu}(x) = 32\pi^2 n\,, \tag{39.86}$$

and therefore,

$$\mathcal{S}(\mathbf{A}_\mu) \ge 8\pi^2|n|/g\,. \tag{39.87}$$

The equality, which corresponds to a local minimum of the action, is obtained for fields satisfying the self-duality equations

$$\mathbf{F}_{\mu\nu}(x) = \pm\tilde{\mathbf{F}}_{\mu\nu}(x), \tag{39.88}$$

which are first-order partial differential equations. The one-instanton solution, which depends on an arbitrary scale parameter λ, is ($r = |\mathbf{x}|$)

$$A^i_m(x) = \frac{2}{r^2+\lambda^2}\left(x_4\delta_{im} + \sum_k \epsilon_{imk}x_k\right), \quad A^i_4(x) = -\frac{2x_i}{r^2+\lambda^2}. \tag{39.89}$$

The semi-classical vacuum. In analogy with the analysis of the $CP(N-1)$ model, we quantize in the temporal gauge $\mathbf{A}_4 = 0$. The classical minima of the potential correspond to gauge field components $\mathbf{A}_i$, $i = 1,2,3$, which are pure gauge functions of the three space variables x_i:

$$-\tfrac{1}{2}i\mathbf{A}_m(x)\cdot\boldsymbol{\sigma} = \mathbf{g}(x_i)\partial_m\mathbf{g}^{-1}(x_i)\,. \tag{39.90}$$

The structure of the classical minima is related to the homotopy classes of mappings of the group elements $\mathbf{g}$ into compactified $\mathbb{R}^3$ (because $\mathbf{g}(x)$ goes to a constant for $|x| \to \infty$), that is, again of S_3 into S_3, and thus the semi-classical vacuum has a periodic structure. One verifies that the gauge equivalent in the temporal gauge of the instanton solution (39.89) connects minima with different winding numbers.

Therefore, as in the example of the $CP(N-1)$ model, to project onto a θ vacuum, one can add the topological term to the classical action of gauge theories,

$$\mathcal{S}_\theta(\mathbf{A}) = \mathcal{S}(\mathbf{A}) + i\frac{\theta}{32\pi^2}\int d^4x \sum_{\mu,\nu} \mathbf{F}_{\mu\nu}(x)\cdot\tilde{\mathbf{F}}_{\mu\nu}(x), \tag{39.91}$$

and then integrate over all fields $\mathbf{A}_\mu$ without restriction. At least in the semi-classical approximation, the gauge theory thus depends on one additional parameter, the angle θ. For non-vanishing values of θ, the additional term violates CP (charge conjugation parity) conservation, and is at the origin of the *strong CP violation* problem: experimental bounds fix the value with of θ with an unnatural precision of the order of 10^{-9}.

39.6.1 Fermions in an instanton background

In QCD (see Section 23.4.1), gauge fields are coupled to quarks ($\mathbf{Q}$, $\bar{\mathbf{Q}}$) with an action of the form (using here a $SU(3)$ notation)

$$\mathcal{S}(\mathbf{A}_\mu, \bar{\mathbf{Q}}, \mathbf{Q}) = -\int \mathrm{d}^4x \left(\frac{1}{4g^2}\sum_{mu,\nu} \operatorname{tr}\mathbf{F}^2_{\mu\nu}(x) + \sum_{f=1}^{N_f} \bar{\mathbf{Q}}_f(x)\,(\not{\mathbf{D}} + m_f)\,\mathbf{Q}_f(x)\right),$$

where N_f is the number of quark flavours.

Then, first if the θ term in (39.91) contributes and one fermion field is massless, according to the analysis of Section 23.6.4, the Dirac operator has at least one vanishing eigenvalue, and the determinant resulting from the fermion integration vanishes. Then, the instantons do not contribute to the field integral, and the strong CP violation problem is solved. However, such a hypothesis seems to be inconsistent with experimental estimates of quark masses.

Second, as we have already discussed in Section 23.7.2, if the instantons contribute, they solve the $U(1)$ problem, that is, the absence of a Goldstone boson associated with the almost spontaneous breaking of the axial $U(1)$ current.

The Gaussian integration. In $CP(N-1)$ models, and in non-Abelian gauge theories, the classical theory is scale invariant. Therefore, instanton solutions depend on a scale parameter, which is a collective coordinate over which one has to integrate. This leads to difficult problems, as the analysis of the massless $\phi^4_{d=4}$ field theory reveals (see Chapter 38). Both theories are asymptotically free, and the problems come from the IR region, that is, from instantons of large size for which the semi-classical approximation is no longer legitimate, because the interaction increases with distance (see Chapters 24 and 25).

The role of instantons thus is not fully understood, a complete calculation being possible only with an IR cut-off, provided, for example, by a finite volume. One piece of information presently available concerns the $O(3)$ non-linear σ-model, whose instantons are derived from those of the $CP(1)$-model. It has been rather indirectly argued, by mapping the σ-model onto a one-dimensional quantum spin chain, that instantons are only relevant for $\theta = \pi$, but then they drastically alter the physical picture.

A39 Trace formula for periodic potentials

We consider a Hamiltonian H corresponding to a real periodic analytic potential $V(x)$ with period τ:

$$V(x+\tau) = V(x). \tag{A39.1}$$

Then H commutes with the unitary translation operator T, which on wave functions acts like

$$T\psi(x) = \psi(x+\tau) \;\Rightarrow\; T^\dagger\psi(x) = \psi(x-\tau).$$

Both operators T and H can be diagonalized simultaneously (see Section 39.2). At φ fixed, the spectrum of H is discrete. We define

$$H\psi_n(\varphi,x) = E_n(\varphi)\psi_n(\varphi,x), \quad T\psi_n(\varphi,x) = \mathrm{e}^{i\varphi}\,\psi_n(\varphi,x), \tag{A39.2}$$

with $\|\psi_n(\varphi,x)\| = 1$.

In an interval of size $N\tau$ with periodic boundary conditions, φ is quantized:

$$\mathrm{e}^{iN\varphi} = 1 \Rightarrow \varphi = \varphi_p \equiv \frac{2\pi p}{N}, \quad 0 \le p < N\,. \tag{A39.3}$$

We express the matrix elements of $T^k\,\mathrm{e}^{-\beta H}$ $(k \in \mathbb{Z})$ in terms of eigenfunctions as

$$\langle x'|\,T^k\,\mathrm{e}^{-\beta H}\,|x\rangle = \sum_{p,n} \psi_n^*(\varphi_p,x')\,\mathrm{e}^{-\beta E_n(\varphi_p)+ik\varphi_p}\,\psi_n(\varphi_p,x). \tag{A39.4}$$

This implies for the diagonal elements,

$$\langle x|\,T^k\,\mathrm{e}^{-\beta H}\,|x\rangle = \sum_{p,n} |\psi_n(\varphi_p,x)|^2\,\mathrm{e}^{-\beta E_n(\varphi_p)+ik\varphi_p}\,. \tag{A39.5}$$

Because in a translation of a period (equation (A39.2)), the eigenfunctions are multiplied by a phase, equation (A39.5) shows that the left-hand side is a periodic function of x with period τ. Therefore,

$$\int_0^\tau \langle x|\,T^k\,\mathrm{e}^{-\beta H}\,|x\rangle\,\mathrm{d}x = \sum_{n,p} \mathrm{e}^{ik\varphi_p-\beta E_n(\varphi_p)}\,\frac{1}{N}\int_0^{N\tau} |\psi_n(\varphi_p,x)|^2\,\mathrm{d}x\,. \tag{A39.6}$$

Since the eigenfunctions $\psi_{\varphi_p,n}(x)$ are orthonormal over $N\tau$,

$$\int_0^\tau \langle x|\,T^k\,\mathrm{e}^{-\beta H}\,|x\rangle\,\mathrm{d}x = \frac{1}{N}\sum_{n,p} \mathrm{e}^{ik\varphi_p-\beta E_n(\varphi_p)}\,. \tag{A39.7}$$

Taking the large N limit, one obtains the expression

$$\int_0^\tau \langle x|\,T^k\,\mathrm{e}^{-\beta H}\,|x\rangle\,\mathrm{d}x = \frac{1}{2\pi}\sum_n \int_0^{2\pi} \mathrm{e}^{ik\varphi-\beta E_n(\varphi)}\,\mathrm{d}\varphi\,. \tag{A39.8}$$