Preprint typeset in JHEP style - PAPER VERSION

Instantons and large N

An introduction to non-perturbative methods in QFT

Marcos Mariño

Département de Physique Théorique et Section de Mathématiques,
Université de Genève, Genève, CH-1211 Switzerland

marcos.marino@unige.ch

Abstract: Lecture notes for a course on non-perturbative methods in QFT.

Contents

---

1. Introduction

A nonperturbative effect in QFT or QM is an effect which can not be seen in perturbation theory. In these notes we will study two types of nonperturbative effects. The first type is due to *instantons*, i.e. to nontrivial solutions to the classical equations of motion. If g is the coupling constant, these effects have the dependence

$$e^{-A/g}. \tag{1.1}$$

Notice that this is *small* if g is small, but on the other hand it is completely invisible in perturbation theory, since it displays an essential singularity at $g = 0$.

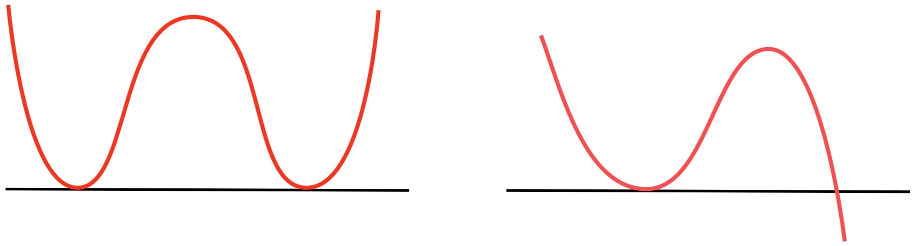


Figure 1: Two quantum-mechanical potentials where instanton effects change qualitatively our understanding of the vacuum structure.

Instanton effects are responsible of one of the most important quantum-mechanical effect: tunneling through a potential barrier. This effect changes qualitatively the structure of the quantum vacuum. In a potential with a perturbative ground state degeneracy, like the one shown on the l.h.s. of Fig. 1, tunneling effects lift the degeneracy. There a single ground state, and the energy difference between the ground state and the first excited state is an instanton effect of the form (1.1),

$$E_1(g) - E_0(g) \sim \mathrm{e}^{-A/g}. \tag{1.2}$$

In a potential with a metastable vacuum, like the one shown in the r.h.s. of Fig. 1, the perturbative vacuum obtained by small quantum fluctuations around this metastable vacuum will eventually decay. This means that the ground state energy has a small imaginary part,

$$E_0(g) = \mathrm{Re}\, E_0(g) + \mathrm{i}\, \mathrm{Im}\, E_0(g), \qquad \mathrm{Im}\, E_0(g) \sim \mathrm{e}^{-A/g} \tag{1.3}$$

which also has the dependence on g typical of an instanton effect.

Some of these instanton effects appear as well in quantum field theories, and they are an important source of information about the dynamics of these theories. However, there are many important strong coupling phenomena in QFT, like confinement and chiral symmetry breaking in QCD, which can not be explained in a satisfactory way in terms of instantons. We should warn the reader that this is a somewhat polemical statement, since for example practicioners of the instanton liquid approach claim that they can explain many aspects of nonperturbative QCD with a semi-phenomenological model based on instanton physics (see [73] for a review). Some aspects of this debate were first pointed out by Witten in his seminal paper [86], and the debate is still going on (see for example [46]).

A different type of nonperturbative method in QFT is based on resumming an infinite subset of diagrams in perturbation theory. This is nonperturbative in the sense that, typically, the effects that one discovers in this way cannot be seen at any finite order of perturbation theory. As an illustration of this, taken from [87], consider the following series:

$$f_0(g) = g - g \log g + g\frac{(\log g)^2}{2} - g\frac{(\log g)^3}{6} + \cdots \tag{1.4}$$

We see that, order by order in perturbation theory, one has the property

$$\lim_{g\to 0} f_0(g) = 0. \tag{1.5}$$

However, each term vanishes more slowly than the one before, and taking into account all the terms in the series one finds $f_0(g) = 1$. Therefore, the property (1.5), which holds at any order in perturbation theory, is not a property of the full resummed series, which satisfies instead

$$\lim_{g \to 0} f_0(g) \neq 0. \tag{1.6}$$

In this sense, the result (1.6) should be also regarded as a nonperturbative effect. Notice that, in this approach, one does not consider a different saddle-point in the path integral, as in instanton physics. Rather, one resums an infinite number of terms in the perturbative series around the conventional vacuum. The most powerful nonperturbative method of this type is probably the $1/N$ expansion of gauge theories [77], where one re-organizes the set of diagrams appearing in perturbation theory according to their dependence on the number N of degrees of freedom.

In these notes we give a pedagogical introduction to these two methods, instantons and large N. We will present general aspects of these methods and we will illustrate them in exactly solvable models.

2. Instantons in quantum mechanics

2.1 QM as a one-dimensional field theory

We first recall that the ground state energy of a quantum mechanical system in a potential $W(q)$ can be extracted from the small temperature behavior of the thermal partition function,

$$Z(\beta) = \mathrm{tr}\, \mathrm{e}^{-\beta H(\beta)}, \tag{2.1}$$

as

$$E = - \lim_{\beta \to \infty} \frac{1}{\beta} \log\, Z(\beta). \tag{2.2}$$

In the path integral formulation,

$$Z(\beta) = \int \mathcal{D}[q(t)] \mathrm{e}^{-S(q)}, \tag{2.3}$$

where $S(q)$ is the action of the Euclidean theory,

$$S(q) = \int_{-\beta/2}^{\beta/2} \mathrm{d}t \left[\frac{1}{2} (\dot{q}(t))^2 + W(q(t)) \right] \tag{2.4}$$

and the path integral is over *periodic* trajectories

$$q(-\beta/2) = q(\beta/2). \tag{2.5}$$

We note that the Euclidean action can be regarded as an action in Lagrangian mechanics,

$$S(q) = \int_{-\beta/2}^{\beta/2} \mathrm{d}t \left[\frac{1}{2} (\dot{q}(t))^2 - V(q) \right] \tag{2.6}$$

where the potential is

$$V(q) = -W(q), \tag{2.7}$$

i.e. it is the inverted potential of the original problem.

It is possible to compute the ground state energy by using Feynman diagrams. We will assume that the potential $W(q)$ is of the form

$$W(q) = \frac{m^2}{2} q^2 + W_{\rm int}(q) \tag{2.8}$$

where $W_{\rm int}(q)$ is the interaction term. Then, the path integral defining Z can be computed in standard Feynman perturbation theory by expanding in $W_{\rm int}(q)$. We will actually work in the limit in which $\beta \to \infty$, since in this limit many features are simpler, like for example the form of the propagator. In this limit, the free energy will be given by β times a β-independent constant, as follows from (2.2). In order to extract the ground state energy we have to take into account the following

1. Since we have to consider $F(\beta) = \log Z(\beta)$, only *connected bubble diagrams* contribute.

2. The standard Feynman rules in position space will lead to n integrations, where n is the number of vertices in the diagram. One of these integrations just gives as an overall factor the "volume," of spacetime i.e. the factor β that we just mentioned. Therefore, in order to extract $E(g)$ we can just perform $n-1$ integrations over $\mathbb{R}$.

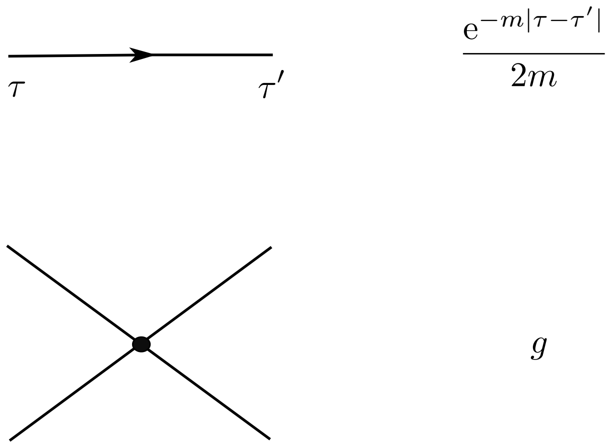


Figure 2: Feynman rules for the quantum mechanical quartic oscillator.

For $\beta \to \infty$ the propagator of this one-dimensional field theory is simply

$$\int \frac{dp}{2\pi} \frac{e^{ip\tau}}{p^2 + m^2} = \frac{e^{-m|\tau|}}{2m}. \tag{2.9}$$

For a theory with a quartic interaction (i.e. the anharmonic, quartic oscillator)

$$W_{\rm int}(q) = \frac{g}{4} q^4 \tag{2.10}$$

the Feynman rules are illustrated in Fig. 2 (an extra factor 4^{-n} has to be introduced at the end, where n is the number of vertices, due to our normalization of the interaction).

One can use these rules to compute the perturbation series of the ground energy of the quartic oscillator. Here we indicate the calculation up tp order g^3 (see Appendix B of [10] for some details). The relevant Feynman diagrams are shown in Fig. 3. For the Feynman integrals we find (we set $m = 1$)

$$
\begin{aligned}
1 :& \quad \frac{1}{4} \\
2a :& \quad -\frac{1}{16}\int_{-\infty}^{\infty} \mathrm{e}^{-2|\tau|}\mathrm{d}\tau = -\frac{1}{16}\cdot 1, \\
2b :& \quad -\frac{1}{16}\int_{-\infty}^{\infty} \mathrm{e}^{-4|\tau|}\mathrm{d}\tau = -\frac{1}{16}\cdot \frac{1}{2}, \\
3a :& \quad \frac{1}{64}\int_{-\infty}^{\infty} \mathrm{e}^{-|\tau_1|-|\tau_2|-|\tau_1-\tau_2|}\mathrm{d}\tau_1\,\mathrm{d}\tau_2 = \frac{1}{64}\cdot\frac{3}{2} \\
3b :& \quad \frac{1}{64}\int_{-\infty}^{\infty} \mathrm{e}^{-2|\tau_1|-2|\tau_2|-2|\tau_1-\tau_2|}\mathrm{d}\tau_1\,\mathrm{d}\tau_2 = \frac{1}{64}\cdot\frac{3}{8} \\
3c :& \quad \frac{1}{64}\int_{-\infty}^{\infty} \mathrm{e}^{-|\tau_1-\tau_2|-|\tau_1|-3|\tau_2|}\mathrm{d}\tau_1\,\mathrm{d}\tau_2 = \frac{1}{64}\cdot\frac{5}{8} \\
3d :& \quad \frac{1}{64}\int_{-\infty}^{\infty} \mathrm{e}^{-2|\tau_1-\tau_2|-2|\tau_2|}\mathrm{d}\tau_1\,\mathrm{d}\tau_2 = \frac{1}{64}\cdot 1
\end{aligned}
\tag{2.11}
$$

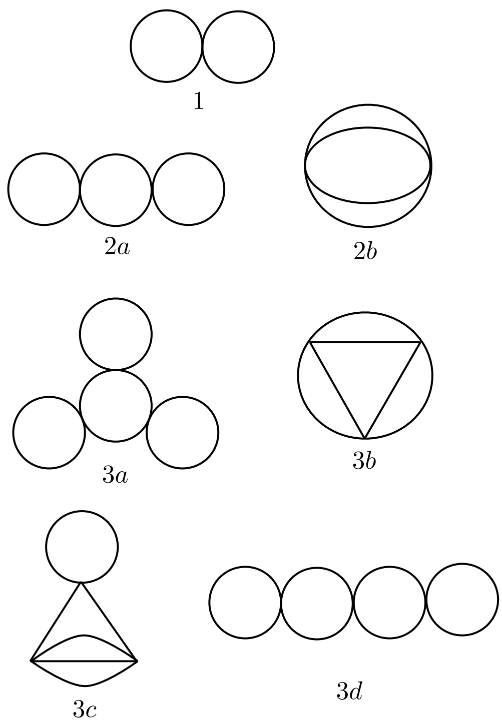


Figure 3: Feynman diagrams contributing to the ground state energy of the quartic oscillator up to order g^3.

The corresponding symmetry factors are given in table 1.

diagram	1	2a	2b	3a	3b	3c	3d
symmetry factor	3	36	12	288	576	288	432

Table 1: Symmetry factors of the Feynman diagrams in Fig. 3.

These numbers can be checked by taking into account that the total symmetry factor for connected diagrams with n quartic vertices is given by

$$\frac{1}{n!}\langle (x^4)^n \rangle^{(c)}, \tag{2.12}$$

where

$$\langle (x^4)^n \rangle = \frac{\int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-x^2/2} x^{4n}}{\int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-x^2/2}} \tag{2.13}$$

is the Gaussian average. By Wick's theorem, this counts all possible pairings among n four-vertices, and we have to take the connected piece. Since

$$\langle x^{2k} \rangle = (2k-1)!! = \frac{(2k)!}{2^k k!} \tag{2.14}$$

we find

$$\frac{1}{n!}\langle (x^4)^n \rangle = \frac{(4n-1)!!}{n!} = \frac{(4n)!}{4^n n!(2n)!} \tag{2.15}$$

One finds, for example,

$$\begin{aligned} \langle x^4 \rangle^{(c)} &= 3, \\ \frac{1}{2!}\langle (x^4)^2 \rangle^{(c)} &= \frac{1}{2}(\langle (x^4)^2 \rangle - \langle x^4 \rangle^2) = 48. \end{aligned} \tag{2.16}$$

We can now compute the first corrections to the ground state energy. Putting together the Feynman integrals with the symmetry factors, we find

$$\begin{aligned} 1 &: \quad \frac{1}{4}\cdot 3 \\ 2a &: \quad -\frac{1}{16}\cdot 1 \cdot 36, \\ 2b &: \quad -\frac{1}{16}\cdot\frac{1}{2}\cdot 12, \\ 3a &: \quad \frac{1}{64}\cdot\frac{3}{2}\cdot 388 \\ 3b &: \quad \frac{1}{64}\cdot\frac{3}{8}\cdot 566 \\ 3c &: \quad \frac{1}{64}\cdot\frac{5}{8}\cdot 288 \\ 3d &: \quad \frac{1}{64}\cdot 1\cdot 432 \end{aligned} \tag{2.17}$$

We then find

$$E = \frac{1}{2} + \frac{3}{4}\left(\frac{g}{4}\right) - \frac{21}{8}\left(\frac{g}{4}\right)^2 + \frac{333}{16}\left(\frac{g}{4}\right)^3 + \mathcal{O}(g^4). \tag{2.18}$$

Remark 2.1. In [10], Bender and Wu give a recursion relation for the coefficients of the perturbative expansion of the ground state energy, starting from the Schrödinger equation.

2.2 Unstable vacua in quantum mechanics

As we explained in the introduction, instanton calculus is relevant for understanding quantum instabilities. We will now calculate the mean lifetime of a particle in the inverted quartic potential by using instanton techniques. Of course, this is a computation which can be also done by using more elementary techniques, like the WKB method. One of the advantages of the path integral/instanton method is that it can be easily generalized to field theory, as we will eventually do.

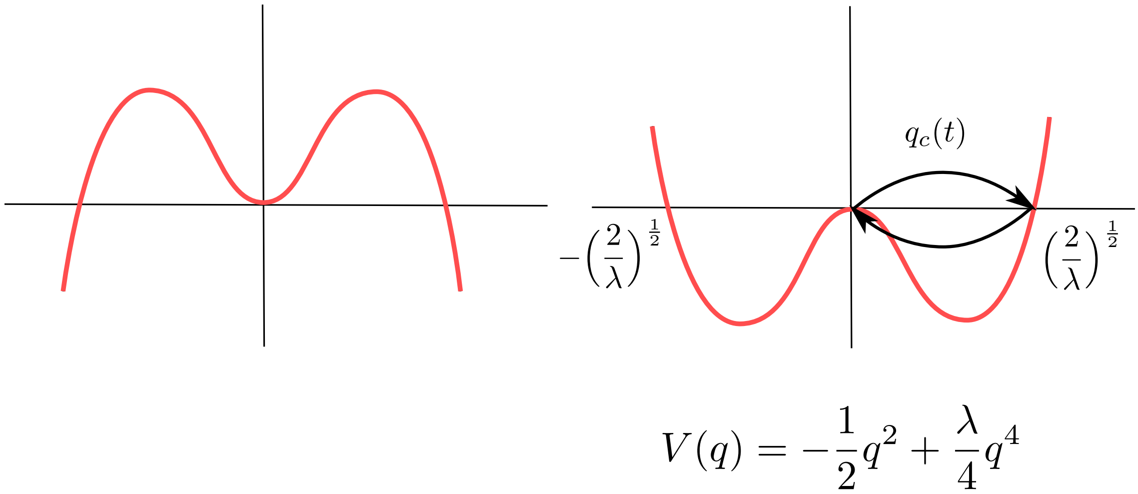


Figure 4: The inverted potential relevant for instanton calculus in the quartic case. The instanton or bounce configuration $q_c(t)$ leaves the origin at $t = -\infty$, reaches the zero $(2/\lambda)^{\frac{1}{2}}$ at $t = t_0$, and comes back to the origin at $t = +\infty$.

Let us now suppose that we have a quantum-mechanical problem with an unstable minimum. A very useful example of such a situation is the inverted *anharmonic oscillator*, with a potential

$$W(x) = \frac{x^2}{2} + \frac{g}{4}x^4. \tag{2.19}$$

where

$$g = -\lambda, \quad \lambda > 0. \tag{2.20}$$

This potential is shown in the left hand side of Fig. 4. The corresponding inverted potential in the Lagrangian interpretation of the Euclidean action is

$$V(q) = -\frac{1}{2}q^2 + \frac{\lambda}{4}q^4 \tag{2.21}$$

and it is shown in the right hand side of Fig. 4.

A particle in the ground state at the bottom of the local unstable minimum will decay by tunneling through the barrier. We want to calculate the mean lifetime of the particle, or equivalently the imaginary part of the ground state energy. This imaginary part is inherited from an imaginary part in the thermal partition function. To see this, we write

$$Z = \mathrm{Re}\, Z + \mathrm{i}\mathrm{Im}\, Z \Rightarrow F(\beta) = -\frac{1}{\beta}\log Z = -\frac{1}{\beta}\log(\mathrm{Re}\, Z) - \frac{\mathrm{i}}{\beta}\frac{\mathrm{Im Z}}{\mathrm{Re}\, Z} + \cdots, \tag{2.22}$$

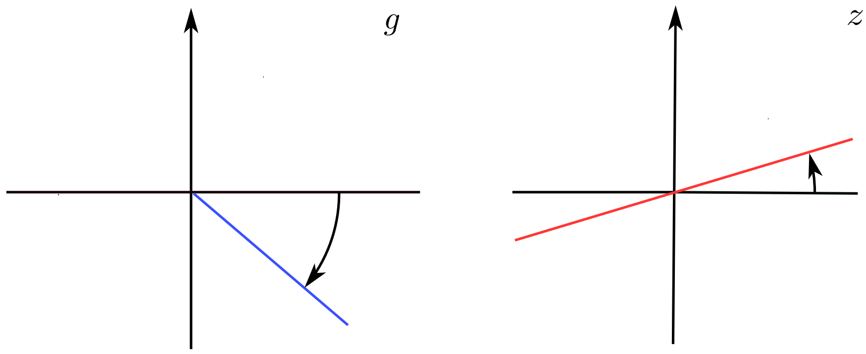


Figure 5: We can analytically continue the integral (2.25) to negative values of g by rotating the integration contour for z. Here we rotate g clockwise, and the integration contour counterclockwise, in such a way that the integral is convergent.

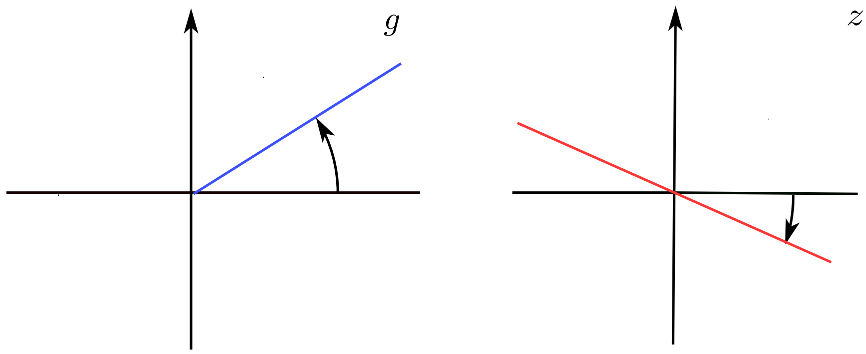


Figure 6: Here we rotate g counterclockwise, and the integration contour clockwise.

where we have taken into account that the imaginary part of Z is exponentially suppressed with respect to the real part (we will verify this in a moment). Therefore, at leading order in the exponentially suppresed factor we have

$$\operatorname{Im} F(\beta) = -\frac{1}{\beta}\frac{\operatorname{Im Z}}{\operatorname{Re} Z}, \tag{2.23}$$

and

$$\operatorname{Im} E(g) = \lim_{\beta\to\infty} \operatorname{Im} F(\beta) = -\lim_{\beta\to\infty}\frac{1}{\beta}\frac{\operatorname{Im Z}}{\operatorname{Re} Z}. \tag{2.24}$$

How do we calculate $\operatorname{Im} Z$ by using path integrals?

2.3 A toy model integral

In order to understand how to compute $\operatorname{Im} Z$, it is very instructive to look at a simpler problem [27, 95]. We will then consider the reduction of the anharmonic oscillator to zero dimensions, and we will analyze the simple quartic integral

$$I(g) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \mathrm{d}z\, \mathrm{e}^{-z^2/2-gz^4/4}. \tag{2.25}$$

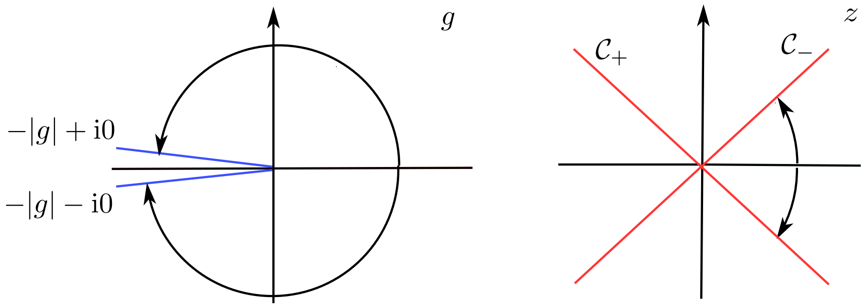


Figure 7: The integration contours $\mathcal{C}_\pm$ correspond to the negative values of $g = -|g| \pm \mathrm{i}0$.

This integral is well defined as long as

$$\mathrm{Re}(g) > 0, \tag{2.26}$$

but we would like to define it for more general, complex values of g, in particular we would like to define it for negative values of g. This can be done by analytic continuation: we *rotate the contour of integration* for the z variable, so that

$$\mathrm{Re}\left(gz^4\right) > 0 \tag{2.27}$$

and the integral is still convergent. Equivalently, we give a phase to z in such a way that

$$\mathrm{Arg}\, z = -\frac{1}{4} \mathrm{Arg}\, g. \tag{2.28}$$

Obviously, this analytic continuation of the integral is no longer real.

In order to define the integral for negative g, we should rotate g towards the negative real axis. But it is clear that this can be done in *two* different ways: clockwise, as in Fig. 5, or counterclockwise, as in Fig. 6. The integration contour for z rotates correspondingly. Since the resulting integration contours are complex conjugate to each other, the two integrals defined in this way are also complex conjugate. For $g \to -|g| + \mathrm{i}0$, one has for z the integration contour

$$\mathcal{C}_+ : \quad \mathrm{Arg}\, z = -\frac{\pi}{4}, \tag{2.29}$$

while for $g \to -|g| - \mathrm{i}0$, one has

$$\mathcal{C}_- : \quad \mathrm{Arg}\, z = \frac{\pi}{4}, \tag{2.30}$$

see Fig. 7. This means that one can indeed obtain an analytic continuation of the integral $I(g)$ to negative g, but the resulting function will have a *branch cut* along the negative real axis. The discontinuity across the cut is given by

$$I(g + \mathrm{i}0) - I(g - \mathrm{i}0) = 2\mathrm{i}\, \mathrm{Im}\, I(g) = \frac{1}{\sqrt{2\pi}} \int_{\mathcal{C}_+ - \mathcal{C}_-} \mathrm{d}z\, \mathrm{e}^{-z^2/2 - gz^4/4}. \tag{2.31}$$

The discontinuity (2.31) can be computed by saddle-point methods. The saddle points of the integral occur at $z = 0$ or

$$z + gz^3 = 0 \Rightarrow z^2 = -\frac{1}{g}. \tag{2.32}$$

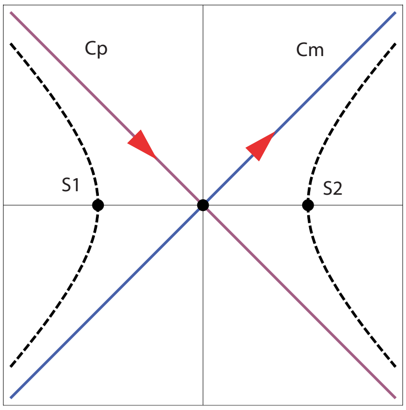


Figure 8: The complex plane for the saddle-point calculation of (2.25). Here, $\mathcal{C}^+$ and $\mathcal{C}^-$ are the rotated contours one needs to consider for $g < 0$. Their sum may be evaluated by the contribution of the saddle–point at the origin. Their difference is evaluated by the contribution of the sub–leading saddle–points, here denoted as $\mathcal{S}_1$ and $\mathcal{S}_2$.

Therefore we have two nontrivial saddlepoints $\mathcal{S}_{1,2}$

$$z_{1,2} = \pm e^{i(\pi/2+\phi_g/2)}|g|^{-\frac{1}{2}} \tag{2.33}$$

where ϕ_g is the phase of g. For $g < 0$, they are on the real axis, see (8). The steepest descent trajectories passing through these points are determined by the condition

$$\mathrm{Im} f(z) = \mathrm{Im}\, f(z_i), \qquad f(z) = \frac{z^2}{2} + \frac{g}{4}z^4. \tag{2.34}$$

For $g < 0$ these are hyperbolae

$$x^2 - y^2 = -\frac{1}{g} \tag{2.35}$$

passing through the saddlepoints $\mathcal{S}_{1,2}$ at $x = \pm|g|^{-\frac{1}{2}}$, $y = 0$, see Fig. 8. From this figure it is also clear that the contour $\mathcal{C}_+ - \mathcal{C}_-$ appearing in (2.31) can be deformed into the *sum* of the steepest descent trajectories passing through $\mathcal{S}_{1,2}$, therefore the imaginary part in (2.31) is given by

$$\mathrm{Im}\, I(g) \sim \exp\left(\frac{1}{4g}\right). \tag{2.36}$$

Since the integral (2.25) is divergent for $g < 0$, the resulting complex function can not be analytic at $g = 0$. One consequence of this lack of analyticity is that the formal power series expansion around $g = 0$,

$$I(g) = \sum_{k=0}^{\infty} a_k g^k, \tag{2.37}$$

where

$$a_k = \frac{(-4)^{-k}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz \frac{z^{4k}}{k!} e^{-z^2/2} = (-4)^{-k}\frac{(4k-1)!!}{k!}, \tag{2.38}$$

has *zero radius of convergence*. Its asymptotic behavior at large k is obtained immediately from Stirling's formula

$$a_k \sim (-4)^k k!. \tag{2.39}$$

This factorial divergence is in fact a generic feature of perturbative series in quantum theory, as we will see.

2.4 Path integral around an instanton in QM

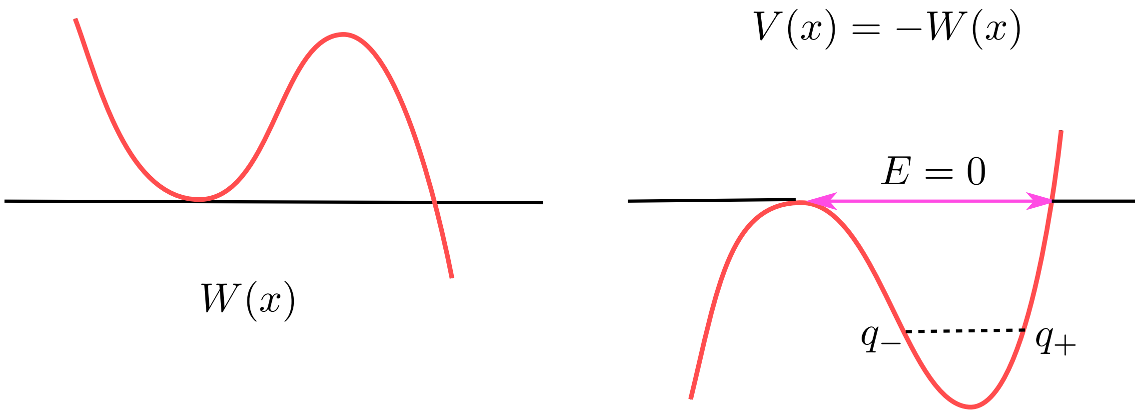


Figure 9: A general unstable potential $W(x)$ and the associated inverted potential $V(x)$. A periodic solution with negative energy moves between the turning points $q_\pm$. The zero energy bounce, relevant to extracting the imaginary part of the ground state energy, is also shown.

The moral of the simple analysis in the previous subsection is that, for negative g, the integral $I(g)$ picks an imaginary part which is given by the contribution of the nontrivial saddlepoints. Let us now come back to our quantum-mechanical problem. By analogy with this example, and in particular from (2.31), we expect that the quantity

$$\text{disc}\, Z(-\lambda) = Z(-\lambda + \mathrm{i}\epsilon) - Z(-\lambda - \mathrm{i}\epsilon) = 2\mathrm{i}\, \mathrm{Im}\, Z(-\lambda) \tag{2.40}$$

is given by the sum of the *nontrivial* saddle-points of the path integral (2.3). These nontrivial saddle points are time-dependent, *periodic* solutions of the EOM for the *inverted* potential,

$$\ddot{q}_c(t) + V'(q_c) = 0. \tag{2.41}$$

Examples of such nontrivial, periodic saddle points are oscillations around the local minima of $V(q)$, as shown in Fig. 9. The period of such an oscillation between the turning points q_- and q_+ is given by

$$\beta = 2 \int_{q_-}^{q_+} \frac{\mathrm{d}q}{\sqrt{2(E - V(q))}}. \tag{2.42}$$

These trajectories satisfy in addition the "energy conservation" constraint

$$\frac{1}{2}\dot{q}^2 + V(q) = E(\beta). \tag{2.43}$$

Notice that the period (2.42) varies between $\beta = \infty$ (corresponding to $E = 0$ in Fig. 9) and a minimum critical value β_c corresponding to the minimum q_0 of the potential. This value can be computed as follows. Near the bottom of the inverted potential one has

$$V(q) = V_0 - \frac{1}{2}\omega^2 (q - q_0)^2 + \cdots \tag{2.44}$$

where

$$\omega^2 = -V''(q_0). \tag{2.45}$$

At this order we can parametrize the energy as

$$E = -V_0 + \frac{1}{2}\omega^2\epsilon^2, \qquad q_\pm = q_0 \pm \epsilon, \tag{2.46}$$

where we just evaluated (2.43) with (2.44) at the turning points $q_0 \pm \epsilon$. We then find,

$$\beta = 2\int_{q_0-\epsilon}^{q_0+\epsilon} \frac{\mathrm{d}q}{\sqrt{\omega^2(\epsilon^2 - (q-q_0)^2)}} = \frac{2}{\omega}\int_{-\epsilon}^{\epsilon} \frac{\mathrm{d}\zeta}{\sqrt{\epsilon^2 - \zeta^2}} = \frac{2\pi}{\omega}. \tag{2.47}$$

As $\epsilon \to 0$ we then find,

$$\beta_c = \frac{2\pi}{\omega}. \tag{2.48}$$

For $\beta < \beta_c$ there are no "instanton" trajectories. In terms of a thermal partition function, this is interpreted as saying that for sufficiently high temperatures the bounce degenerates to a solution $q(t) = q_0$ staying at the top of the barrier. The decay mechanism above the temperature $T_c = 1/\beta_c$ is just thermal excitations over the top of the barrier, see for example [1, 50].

Example 2.2. In the example of the anharmonic oscillator, the EOM reads

$$-\ddot{q}(t) + q(t) - \lambda q^3(t) = 0, \tag{2.49}$$

The inverted potential has minima at

$$q = \pm\lambda^{-\frac{1}{2}}, \tag{2.50}$$

and it has zeroes at

$$q = \pm\left(\frac{2}{\lambda}\right)^{\frac{1}{2}}. \tag{2.51}$$

It is possible to find explicit solutions of (2.49) around the minima (2.50), but they are complicated and involve elliptic functions (see for example [64]). However, in the limit $\beta \to \infty$ they simplify. This limit corresponds to solutions that take infinite time in going from q_- to q_+, and they can only exist if the particle arrives to the turning points with zero energy, therefore q_-, q_+ have to be zeroes of the potential and in addition $E = 0$. One easily finds that the trajectories are given explicitly by

$$q_c(t) = \pm\left(\frac{2}{\lambda}\right)^{\frac{1}{2}} \frac{1}{\cosh(t - t_0)}, \tag{2.52}$$

where, for β finite,

$$-\beta/2 < t_0 < \beta/2 \tag{2.53}$$

is a free parameter. When $\beta \to \infty$, such a trajectory starts at the origin in the infinite past, arrives to the zero (2.51) at $t = t_0$, and returns to the origin in the infinite future, i.e.

$$\begin{aligned} &q_c \to 0, \qquad t \to \pm\infty, \\ &t = t_0, \qquad q_c(t_0) = \pm\left(\frac{2}{\lambda}\right)^{\frac{1}{2}}. \end{aligned} \tag{2.54}$$

An example of (2.52) is shown at Fig. 10.

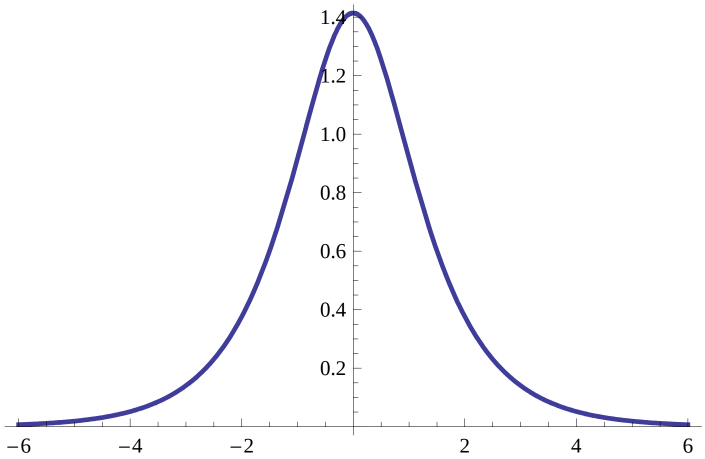


Figure 10: The solution (2.52) with $t_0 = 0$, $\lambda = 1$.

Let us now return to the general case and expand the action around $q_c(t)$. We find, after writing

$$q(t) = q_c(t) + r(t) \tag{2.55}$$

that, at quadratic order in the fluctuations,

$$S(q) \approx S(q_c) + \frac{1}{2}\int \mathrm{d}t_1 \mathrm{d}t_2 r(t_1) M(t_1, t_2) r(t_2) \tag{2.56}$$

where M is the operator defined by

$$M(t_1, t_2) = \frac{\delta^2 S}{\delta q_c(t_1)\delta q_c(t_2)} = \left[-\left(\frac{\mathrm{d}}{\mathrm{d}t_1}\right)^2 - V''(q_c(t_1))\right]\delta(t_1 - t_2). \tag{2.57}$$

In the quadratic (or one-loop) approximation, the path integral around this configuration is then given by

$$\int \mathcal{D}q(t)\,\mathrm{e}^{-S(q)} \approx \mathrm{e}^{-S(q_c)}\int \mathcal{D}r(t)\,\exp\left[-\frac{1}{2}\int \mathrm{d}t_1\mathrm{d}t_2\, r(t_1)M(t_1,t_2)r(t_2)\right]. \tag{2.58}$$

Since we are integrating over periodic configurations, the boundary conditions for $r(t)$ are

$$r(-\beta/2) = r(\beta/2). \tag{2.59}$$

Note that all possible values of the endpoints for $r(t)$ are allowed, since we have to integrate over all possible periodic trajectories and in particular over all possible endpoints. Formally, the Gaussian integration over $r(t)$ gives

$$\int \mathcal{D}r(t)\,\exp\left[-\frac{1}{2}\int \mathrm{d}t_1\mathrm{d}t_2\, r(t_1)M(t_1,t_2)r(t_2)\right] = \left(\det M\right)^{-\frac{1}{2}}. \tag{2.60}$$

The determinant of M is understood here as an (infinite) product over its eigenvalues, and its calculations goes as follows. Let q_n be orthonormal eigenfunctions of M, labeled by $n = 0, 1, \cdots$,

$$\int \mathrm{d}t_2\, M(t_1, t_2) q_n(t_2) = \lambda_n q_n(t_1), \tag{2.61}$$

and satisfying periodic boundary conditions appropriate for (2.58),

$$q_n(-\beta/2) = q_n(\beta/2). \tag{2.62}$$

The eigenvalue problem can be written explicitly as

$$\left[-\frac{\mathrm{d}^2}{\mathrm{d}t^2} - V''(q_c(t))\right] q_n(t) = \lambda_n q_n(t), \quad n \geq 0, \tag{2.63}$$

and orthonormality means that

$$\int_{-\beta/2}^{\beta/2} \mathrm{d}t\, q_n(t) q_m(t) = \delta_{nm}. \tag{2.64}$$

Then,

$$\left(\det M\right)^{-\frac{1}{2}} = \prod_n \lambda_n^{-1/2}. \tag{2.65}$$

In order to makes sense of this formal expression, it is important to notice two properties of M which are crucial to understand the problem.

Firs of all, if we take a further derivative w.r.t. t in (2.41) we find

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\dot{q}_c(t) + V''(q_c(t))\dot{q}_c(t) = 0, \tag{2.66}$$

i.e. $\dot{q}_c(t)$ is a zero mode of M. Since $q_c(t)$ is periodic, $\dot{q}_c(t)$ is periodic as well and the boundary conditions (2.59) are satisfied. It is also a normalizable function, therefore it must be (up to normalization) one of the eigenfunctions $q_n(t)$ of M, say $q_1(t)$ (we will see in a moment that indeed it is the first excited state). The normalized zero mode is

$$q_1(t) = \frac{1}{\|\dot{q}_c\|}\dot{q}_c(t), \tag{2.67}$$

where the norm is given by

$$\|\dot{q}_c\|^2 = \int_{-\beta/2}^{\beta/2} \mathrm{d}t\, (\dot{q}_c(t))^2. \tag{2.68}$$

This norm can be written in many ways. A particularly useful representation is obtained if we use conservation of energy (2.43) and the fact that $E = 0$ for this trajectory,

$$\begin{aligned} \|\dot{q}_c\|^2 &= \int_{-\beta/2}^{\beta/2} \mathrm{d}t\, \frac{1}{2}(\dot{q}_c(t))^2 - \int_{-\beta/2}^{\beta/2} \mathrm{d}t\, V(q_c(t)) \\ &= S(q_c(t)). \end{aligned} \tag{2.69}$$

The origin of this zero mode can be also explained in terms of time translation invariance. When we solve for a nontrivial saddle point we find in general a *family of solutions*, which we can parametrize by an initial time t_0 as in (2.52). A parameter for a family of solutions is called a *modulus* or a *collective coordinate*. Since the Euclidean action is invariant under time translations, the function

$$S(t_0) = S(q_c^{t_0}(t)) \tag{2.70}$$

is *constant.* Here we have explicity indicated the dependence on t_0 in the family of solutions $q_c(t)$. In particular

$$\frac{\mathrm{d}^2 S(t_0)}{\mathrm{d}t_0^2} = 0. \tag{2.71}$$

Notice that

$$\frac{\mathrm{d}S(t_0)}{\mathrm{d}t_0} = \int \mathrm{d}t \frac{\delta S}{\delta q_c(t)} \frac{\delta q_c^{t_0}(t)}{\delta t_0} = 0, \tag{2.72}$$

since $q_c^{t_0}(t)$ is a solution of the classical EOM, and

$$\frac{\mathrm{d}^2 S(t_0)}{\mathrm{d}t_0^2} = \int \mathrm{d}t_1 \mathrm{d}t_2 \frac{\delta^2 S}{\delta q_c(t_1)\delta q_c(t_2)} \frac{\delta q_c^{t_0}(t_1)}{\delta t_0} \frac{\delta q_c^{t_0}(t_2)}{\delta t_0}. \tag{2.73}$$

But the operator appearing here is nothing but $M(t_1, t_2)$, and

$$\frac{\delta q_c^{t_0}(t)}{\delta t_0} = -\dot{q}_c^{t_0}(t), \tag{2.74}$$

therefore we conclude that

$$\int \mathrm{d}t_2\, M(t_1, t_2) \dot{q}_c^{t_0}(t_2) = 0 \tag{2.75}$$

and $\dot{q}_c(t)$ is a zero mode of M.

The second important point is that the operator M has one, and only one, negative mode, therefore (2.58) is imaginary. To see this, we regard

$$-\frac{\mathrm{d}^2}{\mathrm{d}t^2} - V''(q_c(t)) \tag{2.76}$$

as a Schrödinger operator. We have found an eigenfunction $\dot{q}_c(t)$ with zero energy. On the other hand, the spectrum of a Schrödinger operator has the well-known property that the ground state has no nodes, the first excited state has one node, etc. The function $\dot{q}_c(t)$ has one node, since it vanishes at the turning point and then changes sign. Therefore it is the *first* excited state of the above operator, and there must be another eigenfunction with *less* energy (i.e. negative energy) which is the ground state. This is the negative mode of M.

Example 2.3. Let us consider again the quartic oscillator. The operator M is given in this case by

$$M = -\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 1 - \frac{6}{\cosh^2(t - t_0)}. \tag{2.77}$$

Using translation invariance we can just set $t_0 = 0$ to study the spectrum. It is easy to see that

$$M\psi(t) = -3\psi(t), \qquad \psi(t) = \frac{1}{\cosh^2(t)}, \tag{2.78}$$

which is the single negative mode of this operator.

We now address the issue raised by the existence of a zero mode. Naively this leads to an infinite answer for the functional determinant. Indeed, if we expand the fluctuations as

$$r(t) = \sum_{n \geq 0} c_n q_n(t) \tag{2.79}$$

we find

$$(\det M)^{-1/2} = \int \prod_n \frac{\mathrm{d}c_n}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}\sum_{n\geq 0}\lambda_n c_n^2} = \Big(\int \frac{\mathrm{d}c_1}{\sqrt{2\pi}}\Big)(\det{}' M)^{-1/2} \tag{2.80}$$

where

$$\det{}' M = \prod_{n\neq 1} \lambda_n \tag{2.81}$$

is the determinant of the operator M once the zero mode has been removed. Therefore, the infinite answer comes from the integration over c_1. But c_1 stands really for the collective coordinate t_0. To see this, notice that an arbitrary, periodic function of t can be expanded in two equivalent ways, either as in (2.79) or as

$$q_c^{t_0}(t) + \sum_{n\neq 1} c_n q_n(t) \tag{2.82}$$

where t_0 now varies and parametrizes a direction in the space of path configurations. If we change c_1 in (2.79) we obtain

$$q_1(t)\delta c_1 = \frac{1}{\|\dot{q}_c\|}\dot{q}_c^{t_0}(t)\delta c_1 \tag{2.83}$$

while varying t_0 in (2.82) gives

$$-\dot{q}_c^{t_0}(t)\delta t_0. \tag{2.84}$$

Both variations are proportional, therefore (2.82) parametrizes the same fluctuations as (2.79). The Jacobian of the change of variables from c_1 to t_0 can be easily computed by comparing both variations,

$$J = \left|\frac{\delta c_1}{\delta t_0}\right| = \|\dot{q}_c\| = S_c^{1/2} \tag{2.85}$$

Therefore, the integration over c_1 gives

$$\frac{1}{\sqrt{2\pi}}\int \mathrm{d}c_1 = \frac{J}{\sqrt{2\pi}}\int_{-\beta/2}^{\beta/2} \mathrm{d}t_0 = \frac{\beta S_c^{1/2}}{\sqrt{2\pi}}, \tag{2.86}$$

where we have used that the "moduli space" for t_0 is $[-\beta/2, \beta/2]$. Of course, this is infinite when $\beta \to \infty$, and this is the source of the divergence in the determinant of M. But since we have to divide by β in (2.24), we will obtain in the end a finite result.

To summarize: instantons come in families parametrized by collective coordinates or "moduli." This leads to zero modes in the quadratic operators that are obtained by looking at fluctuations around a fixed solution. The integration over these zero modes has to be translated into an integration over collective coordinates, which then expresses the possible infinities appearing in this integration as divergences due to the volume of these zero modes.

We now put everything together, and obtain

$$2\mathrm{i}\,\mathrm{Im}\, Z \approx \mathcal{N}\mathrm{e}^{-S_c}\frac{\beta S_c^{1/2}}{\sqrt{2\pi}}(\det{}' M)^{-1/2}. \tag{2.87}$$

As a consistency check, notice that M has one and only one negative eigenvalue, therefore $\det' M$ is negative and the r.h.s. is then pure imaginary, as it should. In this equation, $\mathcal{N}$ is an overall normalization of the measure in the path integral, which is independent of the potential. In order to fix it, t is convenient to use the (unperturbed) harmonic oscillator with $\omega = 1$ as a reference point. Its thermal partition function will be denoted by $Z_{\mathrm{G}}(\beta)$. Notice that, for large β, we have of course

$$Z_{\mathrm{G}}(\beta) \approx \mathrm{e}^{-\beta/2} \tag{2.88}$$

On the other hand, a path integral evaluation of this partition function gives

$$Z_{\mathrm{G}}(\beta) = \mathcal{N}(\det M_0)^{-1/2}, \tag{2.89}$$

where

$$M_0 = \left[-\left(\frac{\mathrm{d}}{\mathrm{d}t_1}\right)^2 + 1 \right] \delta(t_1 - t_2). \tag{2.90}$$

We then find,

$$\operatorname{Im} Z(\beta) \approx \frac{1}{2\mathrm{i}} Z_{\mathrm{G}}(\beta) \left[\frac{\det' M}{\det M_0}\right]^{-\frac{1}{2}} \frac{\beta S_c^{1/2}}{\sqrt{2\pi}} \mathrm{e}^{-S_c}, \tag{2.91}$$

which is valid at small λ (in fact, it is the one-loop approximation to the full result). We now assume that the inverted potential has the form

$$V(q) = \frac{1}{2} q^2 + \mathcal{O}(\lambda), \tag{2.92}$$

where λ is a coupling constant. Then at leading order in the coupling constant λ our problem is in fact a harmonic oscillator, and

$$\operatorname{Re} Z \approx Z_{\mathrm{G}}(\beta) \tag{2.93}$$

Therefore, at leading order in λ,

$$\operatorname{Im} E(\lambda) \approx \frac{\|\dot{q}_c\|}{2\sqrt{2\pi}} \left[-\frac{\det' M}{\det M_0}\right]^{-\frac{1}{2}} \mathrm{e}^{-S_c}. \tag{2.94}$$

This gives the imaginary part of the free energy for potentials of the form (2.92), and from (2.24) we can deduce the imaginary part of the ground state energy. Before going on, we have to compute the remaining ingredient of this equation, namely the functional determinants.

2.5 Calculation of functional determinants I: solvable models

Let us consider the most possible direct approach to the computation of the determinant of M, by focusing on the quartic case (2.77). The operator one finds in this case belongs to a general family of operators called *Pöschl–Teller potentials*. This family has the structure

$$M_{\ell,m} = -\frac{\mathrm{d}^2}{\mathrm{d}t^2} + m^2 - \frac{\ell(\ell+1)}{\cosh^2(t)}. \tag{2.95}$$

Remarkably, this family of potentials has an exactly solvable spectrum. This is due to a factorization property first studied by Schrödinger, and which can be substantially clarified in the light of supersymmetric quantum mechanics, as we will do later. Useful references for the following calculation include [54, 22]. Let us introduce the operators

$$A_\ell = \frac{\mathrm{d}}{\mathrm{d}t} + \ell \tanh t, \qquad A_\ell^\dagger = -\frac{\mathrm{d}}{\mathrm{d}t} + \ell \tanh t. \tag{2.96}$$

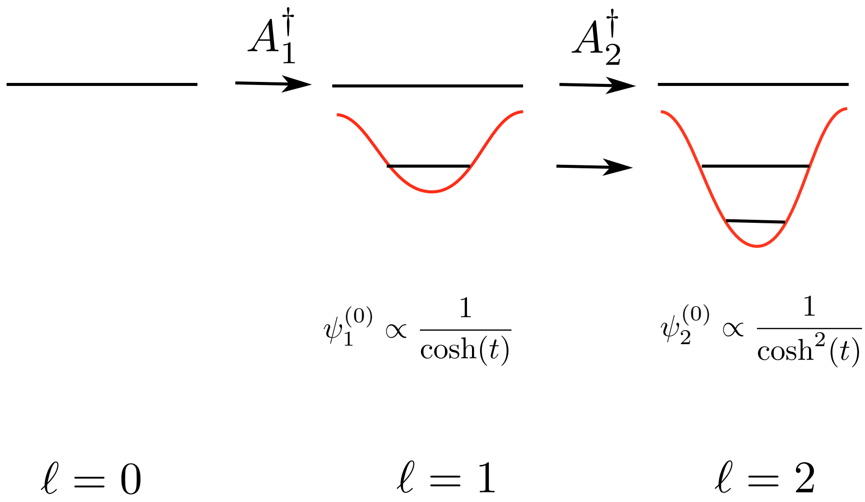


Figure 11: The recursive solution of the spectrum of the Pöschl–Teller potential.

It is immediate to compute that

$$A_\ell^\dagger A_\ell = M_{\ell,m} + \ell^2 - m^2, \qquad A_\ell A_\ell^\dagger = M_{\ell-1,m} + \ell^2 - m^2. \tag{2.97}$$

Notice that for $\ell = 0$ we recover the free particle. Also, we can obtain the ground state for the full family of potentials just by solving

$$A_\ell \psi_0^{(\ell)}(t) = 0. \tag{2.98}$$

This is a first order ODE with solution

$$\psi_0^{(\ell)}(t) \propto \frac{1}{\cosh^\ell(t)}. \tag{2.99}$$

The ground state energy for the operator $M_{\ell,m}$ is simply

$$E_{\ell,m}^{(0)} = m^2 - \ell^2. \tag{2.100}$$

The properties above also make possible to calculate the excited states. To do this, notice that if $\psi^{(\ell-1)}(t)$ is an eigenfunction of $M_{\ell-1,m}$ with eigenvalue $\mu_{\ell-1}$, then

$$\psi^{(\ell)}(t) = A_\ell^\dagger \psi^{(\ell-1)}(t) \tag{2.101}$$

is an eigenfunction of $M_{\ell,m}$ with the same eigenvalue. Indeed,

$$\begin{aligned} M_{\ell,m}\psi^{(\ell)}(t) &= \left(A_\ell^\dagger A_\ell + m^2 - \ell^2\right) A_\ell^\dagger \psi^{(\ell-1)}(t) \\ &= A_\ell^\dagger \left(M_{\ell-1,m} - \ell^2 + m^2\right)\psi^{(\ell-1)}(t) + \left(m^2 - \ell^2\right)\psi^{(\ell)}(t) = \mu_{\ell-1}\psi^{(\ell)}(t) \end{aligned} \tag{2.102}$$

We can then construct the spectrum of $M_{\ell,m}$ by starting with the free particle $\ell = 0$ and applying the operators $A_\ell^\dagger$. For $\ell = 0$, the eigenfunctions are just plane waves (scattering states)

$$\mathrm{e}^{\mathrm{i}kt} \tag{2.103}$$

with energies

$$E_{\ell,m}(k) = k^2 + m^2. \tag{2.104}$$

Applying $A_1^\dagger$ we obtain the scattering states of the $\ell = 1$ potential

$$\psi_1^{(k)}(t) = \frac{A_1^\dagger}{\sqrt{1+k^2}} \frac{\mathrm{e}^{\mathrm{i}kt}}{\sqrt{2\pi}} \tag{2.105}$$

appropriately normalized. On top of that, we have the ground state (2.99) with $\ell = 1$,

$$\psi_1^{(0)} \propto \frac{1}{\cosh(t)}. \tag{2.106}$$

To go to $\ell = 2$, we apply $A_2^\dagger$ to these states, and we obtain the scattering states

$$\psi_2^{(k)}(t) = \frac{A_2^\dagger}{\sqrt{2^2+k^2}} \frac{A_1^\dagger}{\sqrt{1+k^2}} \frac{\mathrm{e}^{\mathrm{i}kt}}{\sqrt{2\pi}}, \tag{2.107}$$

a bound state

$$\psi_2^{(1)}(t) \propto A_2^\dagger \frac{1}{\cosh(t)} \tag{2.108}$$

and the new ground state

$$\psi_2^{(0)}(t) \propto \frac{1}{\cosh^2(t)}. \tag{2.109}$$

Proceeding in this way we obtain the full spectrum of the ℓ-th potential. It consists of scattering states

$$\psi_\ell^{(k)}(t) = \frac{A_\ell^\dagger}{\sqrt{\ell^2+k^2}} \cdots \frac{A_1^\dagger}{\sqrt{1+k^2}} \frac{\mathrm{e}^{\mathrm{i}kt}}{\sqrt{2\pi}}, \tag{2.110}$$

with energy

$$E_{\ell,m} = k^2 + m^2, \tag{2.111}$$

and ℓ bound states

$$\psi_\ell^{(j-1)}(t) \propto A_\ell^\dagger \cdots A_{\ell-j+1}^\dagger \frac{1}{\cosh^{\ell-j+1}(t)}, \qquad j = 1, \cdots, \ell, \tag{2.112}$$

with energy

$$E_{\ell,m}^{(j)} = m^2 - (\ell - j + 1)^2. \tag{2.113}$$

The scattering states are normalized as

$$\int_{-\infty}^{\infty} \psi_\ell^{(k)}(t) \left(\psi_\ell^{(k')}(t)\right)^* \mathrm{d}t = \delta(k - k'). \tag{2.114}$$

In principle, since the spectrum of the operator $M_{\ell,m}$ is known, the determinant is also known. The only subtlety is that we have to regularize the determinant in two ways. The first one is the overall normalization constant, which involves dividing by the determinant of a reference operator as in (2.91). In this case the natural choice is the free particle operator $M_{0,m}$. The other regularization is due to the fact that there is a part of the spectrum which is continuum. If an operator $\mathcal{M}$ has a discrete spectrum $\{\lambda_n\}$ and a continuum spectrum $\lambda(k)$, the determinant should be understood as

$$\log\,\det\mathcal{M} = \sum_n \log(\lambda_n) + \int \mathrm{d}k\,\rho(k)\log\left(\lambda(k)\right), \tag{2.115}$$

where $\rho(k)$ is the density of states for the continuum part. This is easily computed by putting the system in a box. A scattering state in the Pöschl–Teller potential will experience phase shifts. Indeed, as $t \to \pm\infty$, we have

$$A_k^\dagger \sim -\frac{\mathrm{d}}{\mathrm{d}t} \pm k \tag{2.116}$$

and the asymptotic form of the scattering states will be

$$\psi_\ell^{(k)}(t) \to \prod_{j=1}^{\ell}(-\mathrm{i}k \pm j)\,\mathrm{e}^{\mathrm{i}kt} \tag{2.117}$$

Therefore the phase shifts, defined by

$$\psi_\ell^{(k)}(t) \to \exp\left[\mathrm{i}\left(kt \pm \frac{\delta(k)}{2}\right)\right] \tag{2.118}$$

are given by

$$\frac{\delta(k)}{2} = -\sum_{j=1}^{\ell}\tan^{-1}\left(\frac{k}{j}\right) + \frac{\pi}{2} \tag{2.119}$$

The quantization condition once we put these scattering states in a box of length β is just

$$\mathrm{i}k\beta + \mathrm{i}\delta(k) = 2\pi\mathrm{i}n \tag{2.120}$$

and the density of states is

$$\frac{\mathrm{d}n}{\mathrm{d}k} = \rho_{\text{free}}(k) + \rho(k), \tag{2.121}$$

where

$$\rho_{\text{free}}(k) = \frac{\beta}{2\pi}, \qquad \rho(k) = \frac{1}{2\pi}\delta'(k). \tag{2.122}$$

In our case

$$\rho(k) = -\frac{1}{\pi}\sum_{j=1}^{\ell}\frac{j}{k^2+j^2}. \tag{2.123}$$

Subtracting the determinant of the free operator $\det M_{0,m}$ means just keeping $\rho(k)$ in the density. We then find the following result

$$\begin{aligned}\log\det{}' M_{\ell,m} - \log\det M_{0,m} &= \sum_{1\le j\le\ell,\, j\neq m}\log(m^2-j^2) + \int_{-\infty}^{\infty}\mathrm{d}k\,\rho(k)\log(k^2+m^2)\\ &= \sum_{1\le j\le\ell,\, j\neq m}\log(m^2-j^2) - \frac{1}{\pi}\sum_{j=1}^{\ell} j\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{k^2+j^2}\log(k^2+m^2).\end{aligned} \tag{2.124}$$

The last integral can be calculated by residues to be

$$\frac{2\pi}{j}\log(j+m), \tag{2.125}$$

and the prime in the first sum means that we remove the zero mode which appears at $m = j$ when $m \leq \ell$. The end result is

$$\frac{\det' M_{\ell,m}}{\det M_{0,m}} = \frac{\prod_{1\leq j\leq \ell,\, j\neq m}(m^2-j^2)}{\prod_{1\leq j\leq \ell}(m+j)^2} \tag{2.126}$$

2.6 Calculation of functional determinants II: Gelfand–Yaglom method

The determinant of the operator appearing in (2.94) can be computed with a remarkable result known as the Gelfand–Yaglom theorem. This result is very powerful since it does not involve a precise knowledge of the spectrum. We will follow the approach in the Appendix of [50] (see for example [38] for more information).

Let us consider the eigenvalue problem for a Schrödinger operator in the interval $[-\beta/2, \beta/2]$, and with periodic boundary conditions $\psi(-\beta/2) = \psi(\beta/2)$,

$$\left[-\frac{\mathrm{d}^2}{\mathrm{d}t^2} + W(t)\right]\psi(t) = \lambda\psi(t). \tag{2.127}$$

Let us denote by $\psi_\lambda^{1,2}(t)$ the solutions with initial conditions

$$\begin{aligned} \psi_\lambda^1(-\beta/2) &= 1, \qquad & \dot{\psi}_\lambda^1(-\beta/2) &= 0,\\ \psi_\lambda^2(-\beta/2) &= 0, \qquad & \dot{\psi}_\lambda^1(-\beta/2) &= 1. \end{aligned} \tag{2.128}$$

Consider now the matrix,

$$M_\lambda(t) = \begin{pmatrix} \psi_\lambda^1(t) & \psi_\lambda^2(t) \\ \dot{\psi}_\lambda^1(t) & \dot{\psi}_\lambda^2(t) \end{pmatrix} \tag{2.129}$$

Any solution of the eigenvalue problem can be written as

$$\begin{pmatrix} \psi_\lambda(t) \\ \dot{\psi}_\lambda(t) \end{pmatrix} = M_\lambda(t) \begin{pmatrix} \psi_\lambda(-\beta/2) \\ \dot{\psi}_\lambda(-\beta/2) \end{pmatrix}. \tag{2.130}$$

If there is a periodic solution of the eigenvalue problem, then

$$(M_\lambda(\beta/2) - \mathbf{1}) \begin{pmatrix} \psi_\lambda(-\beta/2) \\ \dot{\psi}_\lambda(-\beta/2) \end{pmatrix} = 0. \tag{2.131}$$

Therefore,

$$\det\left(M_\lambda(\beta/2) - \mathbf{1}\right) = \det(M_\lambda(\beta/2)) - \mathrm{Tr}\left(M_\lambda(\beta/2)\right) - 1 = 0. \tag{2.132}$$

Since the Wronskian is constant in time,

$$\det(M_\lambda(\beta/2)) = 1 \tag{2.133}$$

and we end up with the condition

$$\mathrm{Tr}\left(M_\lambda(\beta/2) - \mathbf{1}\right) = 0 \tag{2.134}$$

for λ to be a solution of the eigenvalue problem. If we regard this quantity as a meromorphic function of λ, we conclude that

$$\det\left(-\partial_t^2 + W(t) - \lambda\right) = \mathrm{Tr}\left(M_\lambda(\beta/2) - \mathbf{1}\right) \tag{2.135}$$

since both sides have the same set of poles and zeros and the same behavior at infinity. Notice that the choice of normalization is incorporated in the appropriate way, since for $W(t) = 0$ and $\lambda = -\omega^2$ we have

$$M_{-\omega^2}(t) = \begin{pmatrix} \cosh(\omega t) & \omega^{-1}\sinh(\omega t) \\ \omega\sinh(\omega t) & \cosh(\omega t) \end{pmatrix} \tag{2.136}$$

therefore

$$\mathrm{Tr}\left(M_{-\omega^2}(\beta/2) - \mathbf{1}\right) = \left(2\sinh\frac{\hbar\omega}{2}\right)^2 \tag{2.137}$$

The partition function of the harmonic oscillator is then given by

$$Z(\beta) = \left[\det\left(-\partial_t^2 + \omega^2\right)\right]^{-1/2} = \frac{1}{2\sinh\frac{\hbar\omega}{2}} \tag{2.138}$$

which is the standard result.

The calculation of determinants after removing the zero modes is also relatively straightforward. If we want to calculate

$$\det{}'\left(-\partial_t^2 + V''(q_c(t))\right) \tag{2.139}$$

we just write

$$\begin{aligned} \det{}'\left(-\partial_t^2 + V''(q_c(t))\right) &= -\frac{\partial}{\partial\lambda}\det\left(-\partial_t^2 + V''(q_c(t)) - \lambda\right)\Big|_{\lambda=0} \\ &= -\frac{\partial}{\partial\lambda}\mathrm{Tr}\left(M_\lambda(\beta/2) - \mathbf{1}\right)\Big|_{\lambda=0}. \end{aligned} \tag{2.140}$$

We then have to compute $\psi_\lambda^{1,2}(t)$ to first order in λ. It is easy to find two solutions for $\lambda = 0$. To do this we have to find a solution to the zero mode problem

$$\left[-\frac{\mathrm{d}^2}{\mathrm{d}t^2} + V''(q_c(t))\right]\chi(t) = 0 \tag{2.141}$$

with the right boundary conditions. The general solution of (2.141) is given by

$$\chi(t) = A\dot{q}_c(t) + B\dot{q}_c(t)\int_{-\beta/2}^{t}\frac{\mathrm{d}\tau}{(\dot{q}_c(t))^2}. \tag{2.142}$$

That $\dot{q}_c(t)$ is a zero mode follows from (2.66), and an easy calculation shows that

$$\dot{q}_c(t)f(t) \tag{2.143}$$

solves (2.141) if $f(t)$ solves

$$\dot{q}_c(t)\ddot{f}(t) + 2\ddot{q}_c(t)\dot{f}(t) = 0, \tag{2.144}$$

which indeed is the case for

$$f(t) = \int_{-\beta/2}^{t} \frac{\mathrm{d}\tau}{(\dot{q}_c(t))^2}. \tag{2.145}$$

We now look for the particular solutions which are needed in the Gelfand–Yaglom theorem. By translation in time, we can construct a bounce satisfying

$$\ddot{q}_c(-\beta/2) = 0 \tag{2.146}$$

so that the solutions (2.128) for $\lambda = 0$ are given by

$$\psi_0^1(t) = \frac{\dot{q}_c(t)}{\dot{q}_c(-\beta/2)}, \qquad \psi_0^2(t) = \dot{q}_c(-\beta/2)\dot{q}_c(t) \int_{-\beta/2}^{t} \frac{\mathrm{d}t'}{(\dot{q}_c(t')^2}. \tag{2.147}$$

The solutions $\psi_\lambda^{1,2}(t)$ for $\lambda \neq 0$ can now be computed as a power series in λ, using for example degenerate perturbation theory. The result is

$$\psi_\lambda^{1,2}(t) = \psi_0^{1,2}(t) \left(1 + \lambda \int_{-\beta/2}^{t} \mathrm{d}t' \left[\psi_0^2(t)\psi_0^1(t') - \psi_0^1(t)\psi_0^2(t') \right] \right) + \mathcal{O}(\lambda^2). \tag{2.148}$$

We then obtain

$$\mathrm{Tr}\left(M_\lambda(\beta/2) - 1\right) = \lambda \int_{-\beta/2}^{\beta/2} \mathrm{d}t \int_{-\beta/2}^{\beta/2} \mathrm{d}t' \left(\frac{\dot{q}_c(t)}{\dot{q}_c(t')} \right)^2 + \mathcal{O}(\lambda^2). \tag{2.149}$$

We conclude that

$$\det{}' \left(-\partial_t^2 + V''(q_c(t)) \right) = -\int_{-\beta/2}^{\beta/2} \mathrm{d}t\, (\dot{q}_c(t))^2 \int_{-\beta/2}^{\beta/2} \frac{\mathrm{d}t'}{(\dot{q}_c(t'))^2} \tag{2.150}$$

Notice that this quantity is manifestly negative. Notice that the first factor is just the classical action (2.68), while the second factor can be expressed as follows. Taking derivatives in (2.42) w.r.t. β, we find

$$1 = -\frac{\partial E}{\partial \beta} \int_{x'}^{x} \frac{\mathrm{d}q}{\left[2(E - V(q))\right]^{\frac{3}{2}}}. \tag{2.151}$$

Because of energy conservation we have

$$\dot{q}_c(t) = \left[2(E - V(q_c))\right]^{\frac{1}{2}} \Rightarrow \mathrm{d}q_c(t) = \left[2(E - V(q_c))\right]^{\frac{1}{2}} \mathrm{d}t, \tag{2.152}$$

therefore

$$-\frac{\partial E}{\partial \beta} = \left[\int_{x'}^{x} \frac{\mathrm{d}q}{\left[2(E - V(q))\right]^{\frac{3}{2}}} \right]^{-1} = \left[\int_{-\beta/2}^{\beta/2} \frac{\mathrm{d}t}{2(E - V(q))} \right]^{-1} \tag{2.153}$$

and

$$\int_{-\beta/2}^{\beta/2} \frac{\mathrm{d}t}{(\dot{q}_c(t))^2} = \left(-\frac{\partial E}{\partial \beta} \right)^{-1}. \tag{2.154}$$

We finally conclude that

$$\frac{\det' M}{\det M_0} = -\frac{S_c}{4 \sinh^2 \frac{\beta}{2}} \left(\frac{\partial E}{\partial \beta} \right)^{-1}. \tag{2.155}$$

This result is essentially derived (with a slightly different method) in [32]. See also [16].

The above expression for the determinant has been obtained for finite β, but to extract the ground state energy we are interested in the limit $\beta \to \infty$. This can be done as follows (for simplicity we normalize our potential in such a way that the unstable minimum is at $q = 0$). Let us first consider the classical action. At large β, the only periodic trajectories which survive have $E = 0$ and go from the unstable minimum $q_- = 0$ to a zero of the potential q_+. Their action becomes

$$S_c = \int_{-\beta/2}^{\beta/2} \mathrm{d}t \Big(\frac{1}{2}\dot{q}_c(t)^2 - V(q_c(t))\Big) = 2\int_{q_0}^{q_1} \mathrm{d}q\,(2W)^{\frac{1}{2}}, \tag{2.156}$$

where we have used conservation of energy (2.43). Let us now consider the derivative of E w.r.t. β. In the large β limit, we write (2.42) as

$$\beta = 2\int_{q_-}^{q_+} \mathrm{d}x \left[\frac{1}{\left[2(E-V(x))\right]^{\frac{1}{2}}} - \frac{1}{\left[2(E+x^2)\right]^{\frac{1}{2}}} + \frac{1}{\left[2(E+x^2)\right]^{\frac{1}{2}}}\right]. \tag{2.157}$$

The last term gives

$$2\log\Big(x + \sqrt{x^2+2E}\Big)\Big|_0^{q_+} = \log q_+^2 - \log(E/2) + \mathcal{O}(E), \qquad E \to 0. \tag{2.158}$$

The first two terms, again up to corrections of order $\mathcal{O}(E)$, give

$$2\int_0^{q_+} \mathrm{d}x \left(\frac{1}{\sqrt{2W(x)}} - \frac{1}{x}\right). \tag{2.159}$$

It follows that

$$E(\beta) \approx -2q_+^2 \exp\left[2\int_0^{q_+} \mathrm{d}x \left(\frac{1}{\sqrt{2W(x)}} - \frac{1}{x}\right)\right] \mathrm{e}^{-\beta}, \tag{2.160}$$

therefore

$$\frac{\partial E}{\partial \beta} \approx 2q_+^2 \exp\left[2\int_0^{q_+} \mathrm{d}x \left(\frac{1}{\sqrt{2W(x)}} - \frac{1}{x}\right)\right] \mathrm{e}^{-\beta}. \tag{2.161}$$

We then obtain the following expression for the (normalized) functional determinant relevant to the trajectories with zero energy

$$\frac{\det' M}{\det M_0} = -\frac{S_c}{2q_+^2} \exp\left[-2\int_0^{q_+} \mathrm{d}x \left(\frac{1}{\sqrt{2W(x)}} - \frac{1}{x}\right)\right]. \tag{2.162}$$

Plugging this expression in (2.94) we finally obtain a general formula for the width of an unstable level in quantum mechanics (at one-loop):

$$\mathrm{Im}\, E_0 \approx \frac{1}{2\sqrt{\pi}} q_+ \exp\left[\int_0^{q_+} \mathrm{d}x \left(\frac{1}{\sqrt{2W(x)}} - \frac{1}{x}\right)\right] \mathrm{e}^{-S(q_c)}. \tag{2.163}$$

It is an interesting exercise to evaluate (2.162) for concrete potentials and check that indeed it agrees with (2.126).

Example 2.4. *Anharmonic oscillator.* The action of the bounce is given by

$$S[q_c(t)] = 2\int_0^{\sqrt{2/\lambda}} x\sqrt{1-\frac{\lambda}{2}x^2}\,\mathrm{d}x = -\frac{2\left(2-\lambda x^2\right)^{3/2}}{3\sqrt{2}\lambda}\Bigg|_0^{\sqrt{2/\lambda}} = \frac{4}{3\lambda}. \tag{2.164}$$

The exponent in (2.162) is

$$\begin{aligned}\int_0^{q_+}\mathrm{d}x\left(\frac{1}{\sqrt{2W(x)}}-\frac{1}{x}\right) &= \int_0^{\sqrt{2/\lambda}}\mathrm{d}x\,\frac{\sqrt{2}-\sqrt{2-\lambda x^2}}{x\sqrt{2-\lambda x^2}} \\ &= -\log\left(\sqrt{2}\sqrt{2-\lambda x^2}+2\right)\Bigg|_0^{\sqrt{2/\lambda}} = \log 2.\end{aligned} \tag{2.165}$$

The determinant is then given by

$$\frac{\det' M}{\det M_0} = -\frac{1}{12} \tag{2.166}$$

and the imaginary part of the ground state energy is

$$\mathrm{Im}\,E_0 \approx \frac{2}{2\sqrt{\pi}}\cdot\sqrt{\frac{2}{\lambda}}\cdot 2\cdot \mathrm{e}^{-\frac{4}{3\lambda}} = \frac{4}{\sqrt{2\pi\lambda}}\mathrm{e}^{-\frac{4}{3\lambda}}. \tag{2.167}$$

Example 2.5. *Cubic oscillator.* Let us now study the cubic potential. The potential is given by

$$V(x) = \frac{1}{2}x^2 - gx^3. \tag{2.168}$$

The turning points are $q_- = 0$ and

$$q_+ = \frac{1}{2g}. \tag{2.169}$$

The instanton solution is

$$q_c(t) = \frac{1}{2g\cosh^2\left(\frac{t}{2}\right)} \tag{2.170}$$

and the operator M reads

$$M = -\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 1 - \frac{3}{\cosh^2\left(\frac{t}{2}\right)}. \tag{2.171}$$

The action of the instanton is

$$S_c = 2\int_0^{1/(2g)}(x^2-2gx^3)^{\frac{1}{2}}\mathrm{d}x = \frac{2}{15g^2}. \tag{2.172}$$

The nontrivial integral involved in the one-loop fluctuation is

$$\int_0^{1/(2g)}\mathrm{d}x\,\frac{x-\sqrt{x^2-2gx^3}}{x\sqrt{x^2-2gx^3}} = \log 4, \tag{2.173}$$

and we find

$$\frac{\det' M}{\det M_0} = -\frac{1}{60} \tag{2.174}$$

and

$$\mathrm{Im}\, E_0(g) \approx \frac{1}{2\pi^{\frac{1}{2}}} \cdot \frac{1}{2g} \cdot 4 \cdot \mathrm{e}^{-2/(15g^2)} = \frac{1}{\sqrt{\pi g^2}} \mathrm{e}^{-2/(15g^2)}. \tag{2.175}$$

This agrees with the results of [4]. The result (2.174) can be also obtained with the Pöschl–Teller potential. After rescaling $t \to 2t$ we find that

$$M = \frac{1}{4} M_{3,2}. \tag{2.176}$$

The behavior of the determinant of an operator under rescaling is not completely trivial (see for example [71], section 3.6). After normalizing we find that

$$\frac{\det' M}{\det M_0} = \left(\frac{1}{4}\right)^{N'_{3,2} - N_{0,2}} \frac{\det' M_{3,2}}{\det M_{0,2}} \tag{2.177}$$

where $N'_{3,2} - N_{0,2}$ is the number of non-zero modes of $M_{3,2}$ minus the number of modes of $M_{0,2}$. This can be computed in general by mimicking the procedure in (2.124). Notice that $M_{\ell,m}$ has $j-1$ discrete non-zero modes for $m \leq \ell$, plus a continuum. To calculate the difference between the zero modes in the continuum for $M_{\ell,m}$ and $M_{0,m}$ we can use again the spectral density. We find,

$$\begin{aligned} N'_{\ell,m} - N_{0,m} &= j - 1 + \int_{-\infty}^{\infty} \mathrm{d}k\, \rho(k) = j - 1 - \frac{1}{\pi} \sum_{j=1}^{\ell} j \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{k^2 + j^2} \\ &= j - 1 - j = -1 \end{aligned} \tag{2.178}$$

Therefore, we conclude that the determinant (2.162) for the cubic oscillator and (2.126) for $\ell = 3$, $m = 2$ should be related by

$$\frac{\det' M}{\det M_0} = 4 \frac{\det' M_{3,2}}{\det M_{0,2}}, \tag{2.179}$$

which is indeed the case.

The result (2.135) can be generalized to other boundary conditions. Let us consider for example the eigenvalue problem (2.127) but now with Dirichlet boundary condition

$$\psi(-\beta/2) = \psi(\beta/2) = 0. \tag{2.180}$$

Then, it is possible to show that (see for example [26])

$$\det\left(-\frac{\mathrm{d}^2}{\mathrm{d}t^2} + W(t)\right) = \phi(\beta/2) \tag{2.181}$$

where $\phi(t)$ is a solution of the zero mode problem

$$\left[-\frac{\mathrm{d}^2}{\mathrm{d}t^2} + W(t)\right] \phi(t) = 0 \tag{2.182}$$

with the boundary condition

$$\phi(-\beta/2) = 0, \qquad \phi'(-\beta/2) = 1. \tag{2.183}$$

An explicit proof is easily found by using the so-called "shifting method" [32, 95].

As an application of this second version of the Gelfand–Yaglom theorem, we can calculate the one-loop approximation to the propagator

$$\langle x'|\mathrm{e}^{-\beta H}|x\rangle = \mathcal{N}\int \mathcal{D}q(t)\mathrm{e}^{-S(q)} \tag{2.184}$$

where the Euclidean action is given by (2.6). The integration is now over trajectories with

$$q(-\beta/2) = x, \qquad q(\beta/2) = x'. \tag{2.185}$$

As before, the calculation reduces, in the Gaussian approximation, to

$$\langle x'|\mathrm{e}^{-\beta H}|x\rangle \approx \mathrm{e}^{-S(q_c)}\int \mathcal{D}r(t)\,\exp\left[-\frac{1}{2}\int \mathrm{d}t_1\mathrm{d}t_2\, r(t_1)M(t_1,t_2)r(t_2)\right], \tag{2.186}$$

where M is the operator (2.57), $q_c(t)$ is a classical solution which satisfies the boundary conditions (2.185), and $r(t)$ satisfies now Dirichlet boundary conditions

$$r(-\beta/2) = r(\beta/2) = 0. \tag{2.187}$$

There are now no zero modes for the operator M, and

$$\int \mathcal{D}r(t)\,\exp\left[-\frac{1}{2}\int \mathrm{d}t_1\mathrm{d}t_2\, r(t_1)M(t_1,t_2)r(t_2)\right] = \Big(\det M\Big)^{-1/2}. \tag{2.188}$$

We can compute the determinant by using the Gelfand–Yaglom theorem. The function $\phi(t)$ satisfying the boundary conditions (2.183) is nothing but the function $\psi_0^2(t)$ considered before, i.e. it is given by

$$\phi(t) = \dot{q}_c(-\beta/2)\dot{q}_c(t)\int_{-\beta/2}^{t}\frac{\mathrm{d}t'}{\left(q_c(t')\right)^2} \tag{2.189}$$

Up to an overall constant $\mathcal{C}$, and at one loop, we have

$$\langle x'|\mathrm{e}^{-\beta H}|x\rangle \approx \mathcal{C}\mathrm{e}^{-S(q_c)}(\phi(\beta/2))^{-\frac{1}{2}}, \tag{2.190}$$

or, explicitly,

$$\langle x'|\mathrm{e}^{-\beta H}|x\rangle \approx \mathcal{C}\mathrm{e}^{-S(q_c)}\left[\dot{q}_c(-\beta/2)\dot{q}_c(\beta/2)\int_{-\beta/2}^{\beta/2}\frac{\mathrm{d}\tau}{(\dot{q}_c(t))^2}\right]^{-1/2}. \tag{2.191}$$

To determine $\mathcal{C}$ we just notice that for the free particle

$$\langle x'|\mathrm{e}^{-\beta H}|x\rangle = (2\pi\beta)^{-1/2}\mathrm{e}^{-(x-x')^2/2\beta}. \tag{2.192}$$

The classical trajectory for a free particle is a straight line with appropriate boundary conditions,

$$q_c(t) = \frac{x+x'}{2} + \frac{x'-x}{\beta}t, \tag{2.193}$$

therefore $\dot{q}_c(t)$ is a constant and

$$\left[\dot{q}_c(-\beta/2)\dot{q}_c(\beta/2)\int_{-\beta/2}^{\beta/2}\frac{\mathrm{d}\tau}{(\dot{q}_c(t))^2}\right]^{-1/2}=\frac{1}{\sqrt{\beta}}. \tag{2.194}$$

We conclude, by comparing the two calculations, that

$$\mathcal{C}=\frac{1}{\sqrt{2\pi}} \tag{2.195}$$

and we find our final expression for the one-loop propagator of a particle in an arbitrary potential:

$$\langle x'|\mathrm{e}^{-\beta H}|x\rangle \approx \mathrm{e}^{-S(q_c)}\left[2\pi\dot{q}_c(-\beta/2)\dot{q}_c(\beta/2)\int_{-\beta/2}^{\beta/2}\frac{\mathrm{d}t}{(\dot{q}_c(t))^2}\right]^{-1/2}. \tag{2.196}$$

This formula can be re-written in many ways. For example, using that

$$\frac{\partial^2 S}{\partial x'\partial x}=\frac{1}{\dot{q}_c(-\beta/2)\dot{q}_c(\beta/2)}\frac{\partial E}{\partial \beta}, \tag{2.197}$$

one can write

$$\langle x'|\mathrm{e}^{-\beta H}|x\rangle \approx \mathrm{e}^{-S(q_c)}\left(-\frac{1}{2\pi}\frac{\partial^2 S}{\partial x'\partial x}\right)^{-1/2}. \tag{2.198}$$

This can be also derived with the WKB method [95] and it is known as Van Vleck's formula.

2.7 Instantons in the double well

The double-well illustrates one of the most important applications of instantons: their ability to lift perturbation theory degeneracies. Indeed, the double-well potential has, in perturbation theory, two different ground states located at the two degenerate minima. This implies, in particular, that in perturbation theory parity symmetry is spontaneously broken. This cannot be the case. We know from elementary quantum mechanics that the spectrum of the Schrödinger operator in this bound-state problem must be discrete, and that the true vacuum is described by a symmetric wavefunction. This wavefunction corresponds, in the limit of vanishing coupling, to the symmetric combination of the two perturbative vacua. The energy split between the symmetric and antisymmetric combination is however invisible in perturbation theory and goes like $\exp(-1/g)$ –a typical instanton effect. A systematic exposition of instanton effects in the double-well potential can be found in [95, 96].

Consider the double well potential with Hamiltonian

$$H=-\frac{1}{2}\left(\frac{\mathrm{d}}{\mathrm{d}q}\right)^2+W(g,q)\,,\qquad W(g,q)=\frac{\lambda^2}{2}\left(q^2-\frac{\mu^2}{\lambda}\right)^2. \tag{2.199}$$

In perturbation theory one finds two degenerate ground states, located around the minima

$$q=\pm\frac{\mu}{\sqrt{\lambda}}, \tag{2.200}$$

and with energy given by

$$\frac{E_0(g)}{\omega}=\frac{1}{2}-g-\frac{9}{2}g^2-\frac{89}{2}g^3-\frac{5013}{8}g^4-\cdots\,, \tag{2.201}$$

where

$$\omega = 2\mu\sqrt{\lambda} \tag{2.202}$$

is the frequency of oscillations around the minima, and

$$g = \frac{\sqrt{\lambda}}{8\mu^3}. \tag{2.203}$$

The Hamiltonian is invariant under the parity symmetry

$$q \leftrightarrow -q \tag{2.204}$$

and thus commutes with the corresponding parity operator P, whose action on wave functions is

$$P\,\psi(q) = \psi(-q). \tag{2.205}$$

The eigenfunctions of H satisfy

$$H\,\psi_{\epsilon,N}(q) = E_{\epsilon,N}(g)\psi_{\epsilon,N}(q)\,, \quad P\,\psi_{\epsilon,N}(q) = \epsilon\,\psi_{\epsilon,N}(q), \tag{2.206}$$

where $\epsilon = \pm 1$ is the parity and the quantum number N can be uniquely assigned to a given state by the requirement that, as $g \to 0$,

$$\frac{E_{\epsilon,N}(g)}{\omega} = N + 1/2 + \mathcal{O}(g), \tag{2.207}$$

i.e. it corresponds to the N-th energy level of the unperturbed harmonic oscillator.

For the double-well potential, one can separate eigenvalues corresponding to symmetric and antisymmetric eigenfunctions by considering, in addition to the standard partition function, the "twisted" partition function

$$Z_{\rm a}(\beta) = {\rm Tr}\left(P\,{\rm e}^{-\beta H}\right) \tag{2.208}$$

where P is the parity operator (2.205). For large β and small coupling constant one has

$$Z_{\rm a}(\beta) \approx {\rm e}^{-\beta E_{+,0}} - {\rm e}^{-\beta E_{-,0}} \approx -\beta {\rm e}^{-\beta\omega/2}\left(E_{+,0} - E_{0,-}\right) \tag{2.209}$$

$Z_{\rm a}(\beta)$ can be written in terms of a path integral with "twisted" boundary conditions,

$$Z_{\rm a}(\beta) = \int_{q(\beta/2)=P(q(-\beta/2))} \mathcal{D}q(t)\,\exp\left[-S\big(q(t)\big)\right], \tag{2.210}$$

In the case of the double well potential we are studying, the boundary condition reads

$$q(-\beta/2) = -q(\beta/2). \tag{2.211}$$

In the infinite β limit, the leading contributions to the path integral come from paths which are solutions of the Euclidean equations of motion and have zero energy. In the case of $Z_{\rm a}(\beta)$, constant solutions of the equation of motion do not satisfy the boundary conditions. Therefore we have to sum over paths which connect the two minima of the potential (2.200), like in Fig. 12. These correspond to nontrivial instanton configurations. In the example of the double-well potential (2.199), such solutions are

$$q_{\pm}^{t_0}(t) = \pm\frac{\mu}{\sqrt{\lambda}}\tanh\left(\mu\sqrt{\lambda}(t - t_0)\right) \tag{2.212}$$

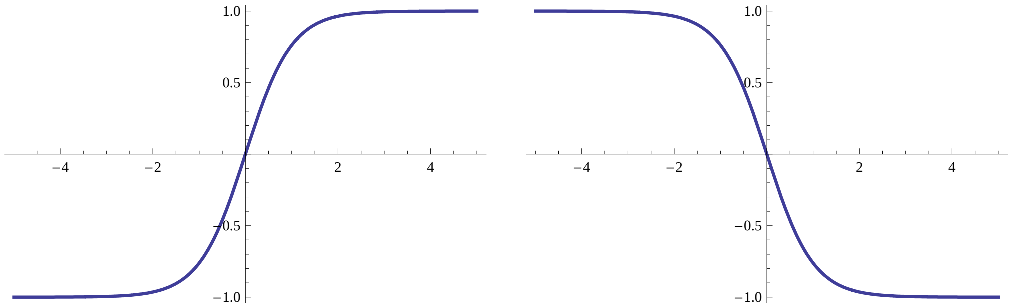


Figure 12: Left: an instanton configuration with center at $t_0 = 0$. Right: an anti-instanton configuration with center at $t_0 = 0$.

The solutions $q_{\pm}^{t_0}$ respectively, are called *(anti)instantons* of center t_0. They are represented in Fig. 12 . Since both solutions depend on an integration constant t_0, there are two one-parameter families of degenerate saddle points.

The operator M in (2.57) is now given by

$$M = -\partial_t^2 + \lambda\mu^2 \left(4 - \frac{6}{\cosh^2\left(\mu\sqrt{\lambda}(t - t_0)\right)} \right) \tag{2.213}$$

which is proportional to the Pöschl–Teller potential $M_{2,2}$:

$$M = \mu^2 \lambda M_{2,2} \tag{2.214}$$

after rescaling $t \to \mu\sqrt{\lambda}t$. Notice that this operator has a zero mode but does *not* have a negative mode. This reflects the fact that the quantum-mechanical state we are now studying is stable. To evaluate the contribution of these configurations to the path integral, we can repeat the arguments that led to (2.91). We obtain

$$Z_{\rm a}(\beta) \approx 2 Z_{\rm G}(\beta) \left[\frac{\det' M}{\det M_0} \right]^{-\frac{1}{2}} \frac{\beta S_c^{1/2}}{\sqrt{2\pi}} {\rm e}^{-S_c}, \tag{2.215}$$

where the extra factor of 2 is due to the fact that the solutions $q_{\pm}^{t_0}(t)$ give the same contribution. From the relation (2.209) we deduce

$$E_{+,0}(g) - E_{-,0} = -2 \frac{S_c^{1/2}}{\sqrt{2\pi}} {\rm e}^{-S_c} \left[\frac{\det' M}{\det M_0} \right]^{-\frac{1}{2}}. \tag{2.216}$$

Let us now compute the quantities involved in this expression. First of all, we have, as in (2.156),

$$S_c = \frac{4}{3} \frac{\mu^3}{\lambda^{1/2}}. \tag{2.217}$$

We see that, at leading order in the effective coupling constant g and for $\beta \to \infty$, (2.216) is proportional to ${\rm e}^{-1/(6g)}$ and therefore it is nonperturbative. For the quotient of determinants we can use the general result for Pöschl–Teller potentials,

$$\frac{\det' M}{\det M_0} = \frac{1}{\mu^2 \lambda} \frac{\det' M_{2,2}}{\det M_{0,2}} = \frac{1}{48 \mu^2 \lambda}. \tag{2.218}$$

One finally obtains the *nonperturbative splitting* between the symmetric and the antisymmetric wavefunctions as

$$\frac{E_{+,0}(g) - E_{-,0}}{\omega} = -\frac{2}{\sqrt{\pi g}} e^{-1/6g} \big(1 + \mathcal{O}(g)\big). \tag{2.219}$$

at leading order in g and $e^{-1/(6g)}$. It follows that the true ground state corresponds to the symmetric wavefunction. The energies of these states can be written as

$$E_{\epsilon,0}(g) = E_0^{(0)}(g) + E_{\epsilon,0}^{(1)}(g), \tag{2.220}$$

where $\epsilon = \pm$ and

$$\frac{E_0^{(0)}(g)}{\omega} = \frac{1}{2} + \mathcal{O}(g), \qquad \frac{E_{\epsilon,0}^{(1)}(g)}{\omega} = -\frac{\epsilon}{\sqrt{\pi g}} e^{-1/6g} \big(1 + \mathcal{O}(g)\big) \tag{2.221}$$

and correspond respectively to the perturbative and the one-instanton contribution.

2.8 Multi-instantons in the double well

In fact, the energies (2.220) have *multi-instanton corrections.* We now provide a brief discussion of these. A more detailed treatment of this beautiful subject can be found in the encyclopedic account by Zinn–Justin and Jentschura [96], which we will follow closely. In the following we will set $\omega = 1$ to simplify our notations, so that the only coupling will be (2.203).

It is easy to see that the existence of a one-instanton correction $E_0^{(1)}(g)$ to the perturbative ground state energy $E_0^{(0)}(g)$ implies the existence of n-instanton contributions to the partition function, since

$$Z_\epsilon(\beta) = \frac{1}{2}(Z(\beta) + \epsilon Z_{\rm a}(\beta)) \sim e^{-\beta(E_0^{(0)} + E_{\epsilon,0}^{(1)})} \sim e^{-\beta/2} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\epsilon \beta}{\sqrt{\pi g}} \right)^n e^{-n/6g}. \tag{2.222}$$

The n-instanton contribution is proportional to β^n. As we will see, the form (2.222) for the n-th instanton contribution is precisely what one finds in the *dilute instanton approximation*, in which one neglects instanton interactions.

The n-th instanton configurations captured in (2.222) do not correspond, in general, to solutions of the classical equation of motion, but rather to configurations of largely separated instantons, connected in a way which we shall discuss, which become solutions of the equation of motion only asymptotically, in the limit of infinite separation. These configurations depend on n times more collective coordinates than the one-instanton configuration. We will call them *quasi-instantons.* Notice that there are no instanton solutions which start from $q = 0$ at $t = -\infty$ and return to it at $t = +\infty$. But there are quasiinstanton solutions that have this property, see Fig. 13.

Finally, notice that the sum in (2.222) can be written as

$$Z_\epsilon(\beta) \sim e^{-\beta/2} \sum_{k=0}^{\infty} \frac{1}{(2k)!} \left(\frac{\beta}{\sqrt{\pi g}} \right)^{2k} e^{-2k/6g} + \epsilon e^{-\beta/2} \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} \left(\frac{\beta}{\sqrt{\pi g}} \right)^{2k+1} e^{-(2k+1)/6g}. \tag{2.223}$$

We see that n *even* contributes to $Z(\beta)$, while n odd contributes to $Z_{\rm a}(\beta)$. This is because a configuration with $n = 2k$ even can be regarded as a chain of k instanton-antiinstanton pairs, which satisfy the boundary condition

$$q(-\beta/2) = q(\beta/2). \tag{2.224}$$

For example, in Fig. 13 we show a configuration contributing to $n = 2$ and with $q(\beta/2) = q(-\beta/2) = 0$. There is a similar contribution coming from a configuration with $q(\beta/2) = q(-\beta/2) = 1$. Similarly, $n = 2k+1$ can be regarded as a chain of k instanton-antiinstanton pairs, followed by an instanton or an antiinstanton, therefore satisfying the boundary condition

$$q(-\beta/2) = -q(\beta/2). \tag{2.225}$$

Since we now know that multiinstanton configurations are expected, let us calculate their effects at leading order for $g \to 0$. We first construct a *two-instanton configuration*. The relevant configurations are instanton-anti-instanton pairs. These configurations depend on one additional time parameter, the separation between instantons, and they decompose in the limit of infinite separation into two instantons.

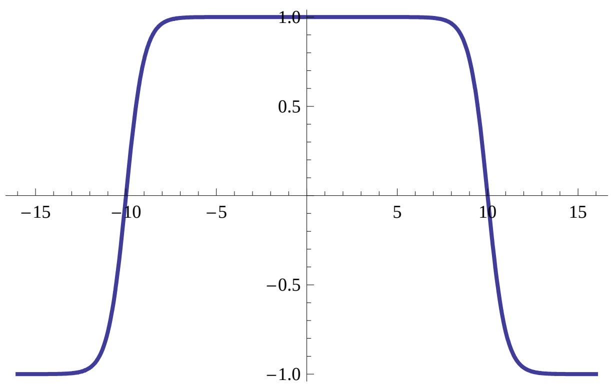


Figure 13: A two-instanton configuration of the form (2.226), for $\theta = 20$.

We consider a configuration $q_{\rm c}^{\theta}(t)$ that is the sum of instantons separated by a distance θ, up to an additive constant adjusted in such a way as to satisfy the boundary conditions (Fig. 13):

$$q_{\rm c}^{\theta}(t) = q_{+}^{-\theta/2}(t) + q_{-}^{\theta/2}(t) - 1 = q_{-}^{\theta/2}(t) - q_{-}^{-\theta/2}(t). \tag{2.226}$$

This path has the following properties:

- It is continuous and differentiable.
- It represents, roughly speaking, an instanton centered at $-\theta/2$, joined to an anti-instanton centered at $\theta/2$.
- When θ is large it differs, near each instanton, from the instanton solution only by exponentially small terms of order $e^{-\theta}$.

We now calculate the action evaluated on this path, as a function of θ. It is convenient to introduce some additional notation:

$$\begin{aligned} u(t) &= q_-^{\theta/2}(t), \\ v(t) &= u(t+\theta), \end{aligned} \tag{2.227}$$

therefore it follows from (2.226) that $q_c^\theta = u - v$. The action corresponding to the path (2.226) can be written as

$$\begin{aligned} S(q_c^\theta) &= \int \mathrm{d}t \left(\frac{1}{2} \dot{q}_c^2 + V(q_c) \right) \\ &= 2 \times \frac{1}{6} + \int \mathrm{d}t \left[-\dot{u}\,\dot{v} + V(u-v) - V(u) - V(v) \right] . \end{aligned} \tag{2.228}$$

Since q_c is even as a function of t, the integral is twice the integral for $t > 0$, where v is at least of order $\mathrm{e}^{-\theta/2}$ for large θ. After an integration by parts of the term $\dot{v}\dot{u}$, one finds

$$S(q_c^\theta) = \frac{1}{3} + 2 \left\{ v(0)\,\dot{u}(0) + \int_0^{+\infty} \mathrm{d}t \left[v\,\ddot{u} + V(u-v) - V(u) - V(v) \right] \right\} . \tag{2.229}$$

One then expands the integrand in powers of v. Since the leading correction to S is of order $\mathrm{e}^{-\theta}$, one needs the expansion only up to order v^2. The term linear in v vanishes as a consequence of the u-equation of motion. One obtains

$$S(q_c^\theta) - \frac{1}{3} \sim 2\,v(0)\,\dot{u}(0) + 2 \left\{ \int_0^{+\infty} \mathrm{d}t \left[\frac{1}{2} v^2\, V''(u) - \frac{1}{2} V''(0)\, v^2 \right] \right\} . \tag{2.230}$$

The function v decreases exponentially away from the origin so the main contributions to the integral come from the neighbourhood of $t = 0$, where $u = 1 + \mathcal{O}(\mathrm{e}^{-\theta/2})$ and thus $V''(u) \sim V''(1) = V''(0)$. Therefore, at leading order the two terms in the integral cancel. At leading order,

$$v(0)\,\dot{u}(0) \sim -\mathrm{e}^{-\theta} \tag{2.231}$$

and thus

$$S(q_c^\theta) = \frac{1}{3} - 2\mathrm{e}^{-\theta} + \mathcal{O}\left(\mathrm{e}^{-2\theta}\right) . \tag{2.232}$$

As we will see, in order to compute the n-instanton contribution but at leading order in g, the corrections of higher order in $\mathrm{e}^{-\theta}$ are not needed. The reason is that, for g small (and negative), the action favours instanton configurations in which the instantons are far apart and θ is large. In analogy with the partition function of a classical gas (instantons being identified with particles), one calls the quantity $-2\mathrm{e}^{-\theta}$ the interaction potential between instantons.

Actually, it is simple to extend the result to β large but finite. To do this, we have to notice that β is a periodic variable, so it lives in a circle. The two instantons separated by a distance θ are also separated by a distance $\beta - \theta$, see Fig. 14. Symmetry between θ and $\beta - \theta$ then implies

$$S(q_c^\theta) = \frac{1}{3} - 2\mathrm{e}^{-\theta} - 2\mathrm{e}^{-(\beta-\theta)} + \cdots \tag{2.233}$$

We now consider an *n-instanton configuration*, i.e. a succession of n instantons (more precisely, alternatively instantons and anti-instantons) separated by times θ_i with

$$\textstyle\sum_{i=1}^n \theta_i = \beta. \tag{2.234}$$

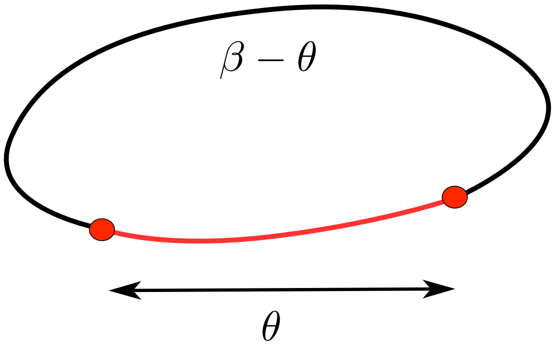


Figure 14: Two instantons separated by θ in one segment of the circle, and by $\beta-\theta$ in the other segment.

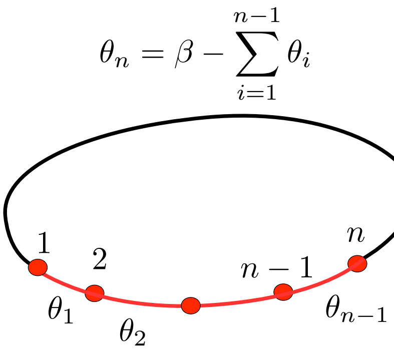


Figure 15: n instantons separated by the distances θ_i along the circle of length β.

We can represent them as in Fig. 15. As noted above, for n even, n-instanton configurations contribute to $\mathrm{Tr}\,\mathrm{e}^{-\beta H}$, while for n odd they contribute to $\mathrm{Tr}\left(P\mathrm{e}^{-\beta H}\right)$.

At leading order, we need only consider "interactions" between nearest neighbour instantons. Other interactions are negligible because they are of higher order in $\mathrm{e}^{-\theta}$. This is an essential simplifying feature of quantum mechanics compared to quantum field theory. The classical action $S_{\mathrm{c}}(\theta_i)$ can then be directly inferred from expression (2.233):

$$S_{\mathrm{c}}(\theta_i)=\frac{n}{6}-2\sum_{i=1}^{n}\mathrm{e}^{-\theta_i}+\mathcal{O}\left(\mathrm{e}^{-(\theta_i+\theta_j)}\right). \tag{2.235}$$

We have calculated the n-instanton action. We now evaluate, at leading order, the contribution to the path integral of the neighbourhood of the n-instanton configuration. We expand the action up to second order in the deviation from the classical path. Although the path is not a solution of the equation of motion, it has been chosen in such a way that the linear terms in the expansion can be neglected at large θ. The Gaussian integration involves then the determinant of the operator M in (2.57). It can be seen that, at leading order in the separation between instantons, the spectrum of M is just the spectrum arising in the one-instanton problem but n-times degenerate, with corrections which are exponentially small in the separation. Therefore, the determinant of M is just the determinant arising the in the $n = 1$ case to the power n. Since we have n collective time variables we also have the Jacobian of the one-instanton case to the power n.

Therefore, the n-instanton contribution to the partition function (2.222) is given by

$$Z_\epsilon^{(n)}(\beta) = \mathrm{e}^{-\beta/2}\, \frac{\beta}{n} \left(\epsilon \frac{\mathrm{e}^{-1/6g}}{\sqrt{\pi g}} \right)^n \int_{\theta_i \geq 0} \delta\left(\sum \theta_i - \beta \right) \prod_i \mathrm{d}\theta_i \exp\left[\frac{2}{g} \sum_{i=1}^n \mathrm{e}^{-\theta_i} \right]. \tag{2.236}$$

The overall factor β comes from the integration over a global time translation, and the factor $1/n$ arises because the configuration is invariant under a cyclic permutation of the θ_i. Finally, the normalization factor $\mathrm{e}^{-\beta/2}$ corresponds to the partition function of the harmonic oscillator. Odd-n instanton effects contribute positively to $Z_+^{(n)}(\beta)$, and negatively to $Z_-^{(n)}(\beta)$. The expression (2.236) is the final expression for the contribution of an n-instanton configuration.

2.9 The dilute instanton approximation

We will now evaluate (2.236) in which instanton interactions are neglected. This is called the *dilute instanton approximation*. Formally, to suppress the interactions, we should take the limit

$$g \to 0^-, \tag{2.237}$$

since in this case

$$\exp\left[\frac{2}{g} \sum_{i=1}^n \mathrm{e}^{-\theta_i} \right] \to 0. \tag{2.238}$$

In fact, as we will see in a moment, the multiinstanton computation is only well-defined for $g < 0$, and the dilute instanton approximation corresponds to g negative and small.

When the interaction term is suppressed, the integration over the θ_i's is straightforward, since

$$\int_{\theta_i \geq 0} \delta\left(\sum \theta_i - \beta \right) \prod_i \mathrm{d}\theta_i = \frac{\beta^{n-1}}{(n-1)!}, \tag{2.239}$$

and

$$Z_\epsilon^{(n)}(\beta, g) = \mathrm{e}^{-\beta/2} \frac{\beta}{n} \left(\epsilon \frac{\mathrm{e}^{-1/6g}}{\sqrt{\pi g}} \right)^n \frac{\beta^{n-1}}{(n-1)!} = \frac{\mathrm{e}^{-\beta/2}}{n!} \left(\epsilon\beta \frac{\mathrm{e}^{-1/6g}}{\sqrt{\pi g}} \right)^n. \tag{2.240}$$

The sum of the leading order n-instanton contributions

$$Z_\epsilon(\beta, g) = \mathrm{e}^{-\beta/2} + \sum_{n=1}^{\infty} Z_\epsilon^{(n)}(\beta, g) \tag{2.241}$$

can now be calculated:

$$Z_\epsilon(\beta, g) \approx \mathrm{e}^{-\beta/2} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\epsilon\beta \frac{\mathrm{e}^{-1/6g}}{\sqrt{\pi g}} \right)^n = \mathrm{e}^{-\beta E_{\epsilon,0}(g)} \tag{2.242}$$

with

$$E_{\epsilon,0}(g) = \frac{1}{2} + \mathcal{O}(g) - \frac{\epsilon}{\sqrt{\pi g}} \mathrm{e}^{-1/6g} \left(1 + \mathcal{O}(g) \right). \tag{2.243}$$

We recognize the perturbative and one-instanton contribution, at leading order, to $E_{\epsilon,0}(g)$, the ground state and the first excited state energies. This is what we could have expected based on (2.222).

2.10 Beyond the dilute instanton approximation

To go beyond the dilute instanton approximation, which only gives the one-instanton contribution to the energy levels (free energy), it is necessary to take into account the interaction between instantons and resum the series. Unfortunately, and as we pointed out before, the interaction between instantons is *attractive* for g positive. In particular, in the limit $g \to 0^+$, the dominant contributions to the integral come from configurations in which the instantons are close and θ_i are small. In this situation, our approximation scheme assuming that the instantons are well separated is not consistent. In fact, when instantons are close, the concept of instanton is no longer meaningful, since the corresponding configurations cannot be distinguished from fluctuations around the constant or the one-instanton solution.

In order to solve this problem, we proceed in two steps: first, we calculate the instanton contribution for g small and *negative*. For negative g the interaction between instantons is *repulsive* and the approximation in terms of well separated instantons becomes meaningful. In a second step, we perform an analytic continuation to g positive of all quantities consistently. It turns out that there are many ways of performing this continuation, but as we will see later these choices are correlated with the choices of Borel resummation of the perturbative series.

Let us now introduce the "fugacity" $\lambda(g)$ of the instanton gas, which is half the one-instanton contribution at leading order,

$$\lambda(g) = \frac{\epsilon}{\sqrt{\pi g}}\, \mathrm{e}^{-1/6g}. \tag{2.244}$$

Notice that, for $g < 0$, λ is imaginary. We also introduce the parameter

$$\mu = -\frac{2}{g}. \tag{2.245}$$

To calculate the integral (2.236), we factorize the integral over the θ_i, by introducing a complex contour integral representation for the δ-function,

$$\delta\left(\sum_{i=1}^{n} \theta_i - \beta\right) = \frac{1}{2\pi \mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}s \exp\left[-s\left(\beta - \sum_{i=1}^{n} \theta_i\right)\right]. \tag{2.246}$$

In terms of the function

$$\mathcal{I}(s,\mu) = \int_0^{+\infty} \exp\left(s\,\theta - \mu\,\mathrm{e}^{-\theta}\right) \mathrm{d}\theta\,, \tag{2.247}$$

$Z_\epsilon^{(n)}(\beta)$ can be rewritten as

$$Z_\epsilon^{(n)}(\beta) \sim \frac{\beta\,\mathrm{e}^{-\beta/2}\,\lambda^n}{2\pi \mathrm{i} n} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}s\,\mathrm{e}^{-\beta\, s}\, \left[\mathcal{I}(s,\mu)\right]^n. \tag{2.248}$$

In order to compute $Z_\epsilon^{(n)}(\beta)$, we will compute its Laplace transform

$$G_\epsilon^{(n)}(E) = \int_0^{\infty} \mathrm{d}\beta\, \mathrm{e}^{\beta E} Z_\epsilon^{(n)}(\beta) \tag{2.249}$$

This gives the n-instanton contribution to the trace of the resolvent

$$G(E) = \mathrm{Tr}\frac{1}{H-E} = \int_0^\infty \mathrm{d}\beta\, \mathrm{e}^{\beta E} Z(\beta)\,, \tag{2.250}$$

The poles of $G(E)$ then yield the spectrum of the Hamiltonian H, i.e. the energy levels. We have,

$$\begin{aligned}
G_\epsilon^{(n)}(E) &= \int_0^\infty \mathrm{d}\beta\, \mathrm{e}^{\beta E} Z_\epsilon^{(n)}(\beta) \\
&= \int_0^\infty \mathrm{d}\beta\, \frac{\beta\, \mathrm{e}^{\beta(E-1/2)}\, \lambda^n}{2\pi \mathrm{i} n} \int\limits_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}s\, \mathrm{e}^{-\beta s}\, [\mathcal{I}(s,\mu)]^n \\
&= \frac{\partial}{\partial E} \int\limits_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}s\, \frac{\lambda^n}{2\pi \mathrm{i} n}\, [\mathcal{I}(s,\mu)]^n \int_0^\infty \mathrm{d}\beta\, \mathrm{e}^{\beta(E-s-1/2)} \\
&= \frac{\partial}{\partial E} \int\limits_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}s\, \frac{\lambda^n}{2\pi \mathrm{i} n}\, [\mathcal{I}(s,\mu)]^n\, \frac{1}{s+1/2-E} \\
&= \frac{\partial}{\partial E} \frac{\lambda^n}{n} \left[\mathcal{I}(E-\frac{1}{2},\mu)\right]^n .
\end{aligned} \tag{2.251}$$

In the last line we have deformed the integration contour to pick the pole at $s = E - 1/2$. We can now sum over all $n \geq 1$ to obtain

$$\sum_{n=1}^\infty G_\epsilon^{(n)}(E) = -\frac{\partial}{\partial E} \ln \phi_\epsilon(E) \tag{2.252}$$

where

$$\phi_\epsilon(E) = \log\Big(1 - \lambda \mathcal{I}(E-\frac{1}{2},\mu)\Big). \tag{2.253}$$

The zero-instanton contribution has not yet been included at all, hence to obtain the trace of the resolvent summed over all sectors $\mathcal{G}_\epsilon(E,g)$ we add the trace of the resolvent of the harmonic oscillator $G_0(E)$

$$\begin{aligned}
\mathcal{G}_\epsilon(E,g) &= G_0(E) - \frac{\partial}{\partial E} \ln \phi_\epsilon(E) \\
&= \frac{\partial}{\partial E} \ln \Gamma\left(\frac{1}{2} - E\right) - \frac{\partial}{\partial E} \ln \phi_\epsilon(E) \\
&= -\frac{\partial}{\partial E} \ln \frac{\phi_\epsilon(E)}{\Gamma\left(\frac{1}{2}-E\right)} \\
&= -\frac{\partial}{\partial E} \ln \Delta_\epsilon(E),
\end{aligned} \tag{2.254}$$

where

$$\Delta_\epsilon(E) = \frac{1}{\Gamma(\frac{1}{2}-E)} - \lambda \frac{\mathcal{I}(E-\frac{1}{2},\mu)}{\Gamma(\frac{1}{2}-E)}. \tag{2.255}$$

Let us now evaluate the integral (2.247) in the limit $\mu \to +\infty$, and thus $g \to 0^-$. We change variables, setting $\mu\, \mathrm{e}^{-\theta} = t$, and the integral becomes

$$\mathcal{I}(s,\mu) = \mu^s \int_0^{\mu} \mathrm{d}t\, t^{-1-s}\, \mathrm{e}^{-t} = \mu^s \int_0^{+\infty} \mathrm{d}t\, t^{-1-s}\, \mathrm{e}^{-t} + \mathcal{O}\left(\mathrm{e}^{-\mu}/\sqrt{\mu}\right). \tag{2.256}$$

We thus obtain

$$\mathcal{I}(s,\mu) \approx \mu^s\, \Gamma(-s), \tag{2.257}$$

since for $\mu \to +\infty$ the difference is exponentially small. Therefore, our ansatz is that the estimate (2.257) gives the correct leading behaviour of the true function. Using this ansatz, we find

$$\Delta_\epsilon(E) = \frac{1}{\Gamma(\frac{1}{2}-E)} - \lambda \mu^{E-\frac{1}{2}} = \frac{1}{\Gamma(\frac{1}{2}-E)} + \epsilon \mathrm{i} \left(-\frac{2}{g}\right)^E \frac{\mathrm{e}^{-1/6g}}{\sqrt{2\pi}} \tag{2.258}$$

The energies are located at the poles of $\mathcal{G}_\epsilon(E,g)$, but since

$$\mathcal{G}_\epsilon(E,g) = -\frac{\Delta'_\epsilon(E)}{\Delta_\epsilon(E)}. \tag{2.259}$$

the poles occur at the zeroes of $\Delta_\epsilon(E)$. These zeros can be obtained as a power series in λ. In order to do that, it is convenient to rewrite the equation

$$\Delta_\epsilon(E) = 0 \tag{2.260}$$

as

$$\frac{\sin \pi(E-1/2)}{\pi} = -\frac{\lambda \mu^{E-1/2}}{\Gamma(E+1/2)}. \tag{2.261}$$

Notice that, for $\lambda = 0$, the zeroes indeed take place at $1/2 + N$, therefore

$$E^{(0)}_{\epsilon,N} = \frac{1}{2} + N + \mathcal{O}(\lambda). \tag{2.262}$$

Using now (2.261) we find a series of the form,

$$E_{\epsilon,N}(g) = \sum_{n=0}^{\infty} E^{(n)}_{\epsilon,N}(g) \lambda^n, \tag{2.263}$$

where $E^{(n)}_N(g)$ is the perturbative series around the n–th instanton solution. To find the leading order in g of $E^{(n)}_{\epsilon,N}(g)$, we write

$$E = \frac{1}{2} + N + x, \qquad x = \sum_{n=1}^{\infty} E^{(n)}_{\epsilon,N}(g)\lambda^n, \tag{2.264}$$

where x solves the implicit equation

$$\frac{\sin \pi x}{\pi} + \frac{\hat{\lambda}\, \mathrm{e}^{\xi x}}{\Gamma(1+N+x)} = 0, \tag{2.265}$$

and

$$\hat{\lambda} = \left(\frac{2}{g}\right)^N \lambda, \qquad \xi = \log \mu = \log\left(-\frac{2}{g}\right). \tag{2.266}$$

This equation can be solved for x as a power series in λ after expanding in x. One finds,

$$x+\frac{\hat{\lambda}}{N!}+\hat{\lambda}x\frac{\xi-\psi(1+N)}{N!}+\cdots=0 \tag{2.267}$$

Therefore, at leading order one has

$$x=-\frac{\hat{\lambda}}{N!}+\frac{\xi-\psi(1+N)}{(N!)^2}\hat{\lambda}^2+\mathcal{O}(\hat{\lambda}^3), \tag{2.268}$$

which means that the one and the two-instanton contributions at one-loop are given by,

$$\begin{aligned} E^{(1)}_{\epsilon,N}(g)&=-\epsilon\frac{1}{N!}\left(\frac{2}{g}\right)^{N+1/2}\frac{\mathrm{e}^{-1/6g}}{\sqrt{2\pi}}\big(1+\mathcal{O}(g)\big),\\ E^{(2)}_{\epsilon,N}(g)&=\frac{1}{(N!)^2}\left(\frac{2}{g}\right)^{2N+1}\frac{\mathrm{e}^{-1/3g}}{2\pi}\Big\{\ln(-2/g)-\psi(N+1)+\mathcal{O}\left(g\ln g\right)\Big\}. \end{aligned} \tag{2.269}$$

Notice that a single equation, (2.261), gives *all* the multi-instanton contributions to *all* energy eigenvalues $E_{\epsilon,N}(g)$ of the double-well potential at leading order in g. It is obvious from the form of (2.261) that the n-th instanton contribution has at leading order the form

$$E^{(n)}_{\epsilon,N}(g)=\left(\frac{2}{g}\right)^{n(N+1/2)}\left(-\epsilon\frac{\mathrm{e}^{-1/6g}}{\sqrt{2\pi}}\right)^{n}\Big\{P^N_n\left(\ln(-g/2)\right)+\mathcal{O}\left(g\left(\ln g\right)^{n-1}\right)\Big\}, \tag{2.270}$$

in which $P^N_n(\xi)$ is a polynomial of degree $n-1$. The first three polynomials are

$$\begin{aligned} P^N_1(\xi)&=1,\\ P^N_2(\xi)&=\xi+\psi(1+N),\\ P^0_3(\xi)&=\frac{3}{2}\left(\xi+\psi(1+N)\right)^2-\frac{1}{2}\psi'(1+N). \end{aligned} \tag{2.271}$$

3. Unstable vacua in QFT

Classic references for this topic are Coleman's papers [25, 20], reviewed in [26].

3.1 Bounces in scalar QFT

We now consider a self-interacting scalar field theory in $d=4$ with an Euclidean action of the form

$$S(\phi)=\int \mathrm{d}^dx\left(\frac{1}{2}(\partial_\mu\phi)^2+U(\phi)\right) \tag{3.1}$$

where the potential $U(\phi)$ has two non-degenerate minima: a *false* vacuum at ϕ_+ which is quantum–mechanically unstable, and a true vacuum at ϕ_-. An example of this situation, which we will analyze in some detail, is given by

$$U(\phi)=\frac{1}{2}\phi^2-\frac{1}{2}\phi^3+\frac{\alpha}{8}\phi^4 \tag{3.2}$$

where

$$0<\alpha<1. \tag{3.3}$$

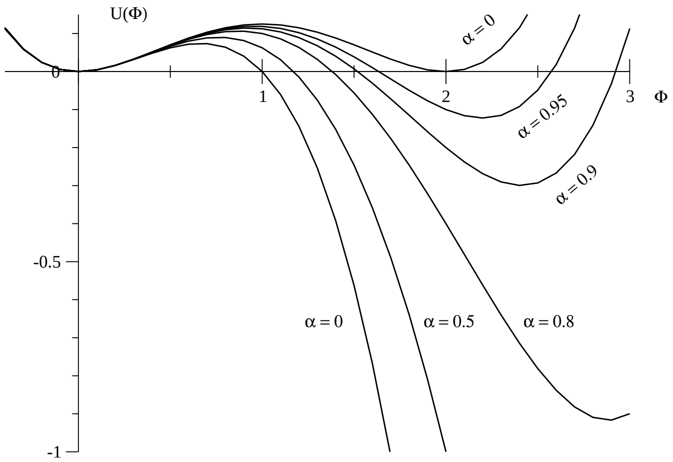


Figure 16: The potential (3.2) for various values of α (from [5]).

This potential is represented in Fig. 16 for different values of α.

This potential has a relative, "false" minimum at $\phi_+ = 0$, a true minimum at

$$\phi_- = \frac{3}{2\alpha} + \frac{\sqrt{9-8\alpha}}{2\alpha}, \tag{3.4}$$

and a local maximum at

$$\phi = \frac{3}{2\alpha} - \frac{\sqrt{9-8\alpha}}{2\alpha}. \tag{3.5}$$

We also have that

$$U(\phi_-) - U(\phi_+) = \frac{4\left(-2\alpha + 2\sqrt{9-8\alpha} + 9\right)\alpha - 9\left(\sqrt{9-8\alpha}+3\right)}{16\alpha^3}, \tag{3.6}$$

and for $\alpha = 1$ the two minima are degenerate. The limit in which $\alpha \to 1$ is called the *thin wall limit*, for reasons that will become clear in the following.

As in the quantum-mechanical case, we want to compute the imaginary part of the ground state energy, in order to derive the decay rate. The steps we will follow are just a carbon copy of what we did in quantum mechanics.

First, we have to look at the solutions of the Euclidean equation of motion. This is simply

$$\left(-\nabla^2 - \frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\right)\phi + U'(\phi) = 0, \tag{3.7}$$

where ∇ is the gradient in three spatial dimensions, and τ is Euclidean time. We also have to impose the relevant boundary conditions. As in the bounce problem in quantum mechanics, we want to start from the false vacuum in the infinite past, and come back to it in the infinite future. Therefore,

$$\phi(\vec{x}, \tau) \to \phi_+, \qquad \tau \to \pm\infty. \tag{3.8}$$

In order to have a finite action for the bounce, we also need it to go to the vacuum value at *spatial* infinity (this is a condition which should be familiar from soliton physics). Hence we have

$$\phi(\vec{x}, \tau) \to \phi_+, \qquad |\vec{x}| \to \infty. \tag{3.9}$$

We can interpret this solution in terms of the formation of a "bubble" in the middle of the false vacuum: asymptotically in Euclidean space, the field configuration is in the false vacuum. But the "core" of the bubble is in a different state.

The solution to the EOM must have a negative mode

$$\det \frac{\delta^2 S}{\delta \phi_*^2} < 0 \tag{3.10}$$

reflecting the instability. Otherwise, the solution does not contribute to the probability of decay (we have to extract an imaginary part to the energy, which gives the decay rate). Since the EOM is invariant under full $O(4)$ rotations of Euclidean space, it is reasonable to look for solutions which are $O(4)$ symmetric, i.e.

$$\phi(r), \quad r = \sqrt{\vec{x}^2 + \tau^2}. \tag{3.11}$$

We also expect that the most symmetric solution is the one with least action, and indeed this turns out to be the case (see [26], chapter 7, section 6.2, and references therein). The equation of motion reduces to

$$\frac{\mathrm{d}^2\phi_c}{\mathrm{d}r^2} + \frac{3}{r}\frac{\mathrm{d}\phi_c}{\mathrm{d}r} = U'(\phi). \tag{3.12}$$

The boundary condition translates into

$$\lim_{r\to\infty} \phi_c = \phi_-. \tag{3.13}$$

Regularity at the origin demands that

$$\left.\frac{\mathrm{d}\phi_c}{\mathrm{d}r}\right|_{r=0} = 0. \tag{3.14}$$

Analytic solutions to (3.12) are not available for nontrivial potentials, but one can show that solutions indeed exist (see [26] for an overshoot/undershoot argument). The equation is an ODE and can be solved numerically with high precision. Results for the potential (3.2) are shown in Fig. 17 for different values of α (this figure is taken from the paper [5]).

One interesting feature of these solutions is that, as $\alpha \to 1$, the solution becomes closer and closer to a step function. Therefore, in the thin wall approximation, the bounce starts at $\phi(0)$ very near the true vacuum ϕ_+ and stays there for a long time $r \sim R$. Then, it moves quickly through the valley between the two minima and stays there. This explains the name thin wall approximation: the bounce looks here like a big bubble of true vacuum of radius R, centered at the origin, separated by a thin wall from the false vacuum that extends to infinity.

We now give a particularly useful form for the action evaluated at the bounce. We will compute

$$S(\phi_c, \lambda) = S(\phi_c(\lambda x)) = \int \mathrm{d}^d x \left(\frac{1}{2}(\partial_\mu \phi_c(\lambda x))^2 + U(\phi_c(\lambda x))\right). \tag{3.15}$$

If we change variables

$$x \to \lambda x \tag{3.16}$$

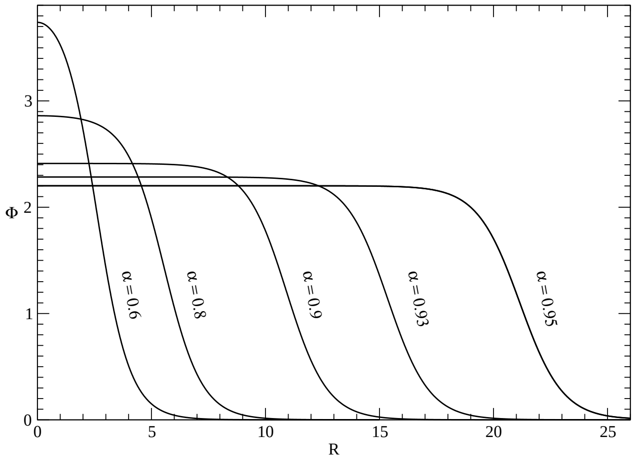


Figure 17: Solutions to (3.12) for the potential (3.2), for various values of α (from [5]).

we find

$$S(\phi_c, \lambda) = \lambda^{2-d} \int \mathrm{d}^d x \frac{1}{2} (\partial_\mu \phi_c(x))^2 + \lambda^{-d} \int \mathrm{d}^d x \, U(\phi_c(x)). \tag{3.17}$$

Since $\phi_c(x)$ satisfies the EOM, the action is stationary under variations of λ:

$$\frac{\mathrm{d}S(\phi_c, \lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=1} = (2-d) \int \mathrm{d}^d x \frac{1}{2} (\partial_\mu \phi_c(x))^2 - d \int \mathrm{d}^d x \, U(\phi_c(x)) = 0. \tag{3.18}$$

Therefore,

$$\int \mathrm{d}^d x \, U(\phi_c(x)) = \frac{2-d}{d} \int \mathrm{d}^d x \frac{1}{2} (\partial_\mu \phi_c(x))^2 \tag{3.19}$$

and

$$S_c = S(\phi_c) = \frac{1}{d} \int \mathrm{d}^d x (\partial_\mu \phi_c(x))^2. \tag{3.20}$$

Notice that there are now d zero modes, corresponding to translation invariance of the bounce in d dimensions. The corresponding functions are

$$\phi_\mu = \partial_\mu \phi_c, \tag{3.21}$$

with norm

$$\int \mathrm{d}^d x \phi_\mu \phi_\nu = \frac{1}{d} \delta_{\mu\nu} \int \mathrm{d}^d x (\partial_\mu \phi_c(x))^2 = \delta_{\mu\nu} S_c, \tag{3.22}$$

The second equality is due to $O(d)$ invariance of the solution. The normalized zero modes are

$$\phi_\mu^{(0)} = \frac{1}{S_c^{\frac{1}{2}}} \partial_\mu \phi_c. \tag{3.23}$$

We have then

$$\delta\phi = \partial_\mu \phi \, \delta x^\mu = \phi_\mu^{(0)} \delta c_\mu^{(0)}. \tag{3.24}$$

In analogy with QM, the zero modes contribute to the integral

$$\frac{1}{(2\pi)^{d/2}} \int \prod_{\mu=1}^{d} \mathrm{d}c_\mu^{(0)} = \frac{S_c^{d/2}}{(2\pi)^{d/2}} \int \prod_{\mu=1}^{d} \mathrm{d}x_\mu = \frac{S_c^{d/2} V \beta}{(2\pi)^{d/2}}, \tag{3.25}$$

where V is the volume of $(d-1)$-dimensional space and β is the total time. We can now proceed with analogy with the derivation in QM. At leading order in coupling constant, our problem is a quadratic theory characterized by the operator

$$-\frac{\mathrm{d}^2}{\mathrm{d}\tau^2} - \nabla^2 + U''(\phi_-), \tag{3.26}$$

This plays the role of M_0 in the QM case. We can then write

$$\operatorname{Im} E/V = \frac{1}{2}\frac{S_c^{d/2}\beta}{(2\pi)^{d/2}}\left|\frac{\det'\big(-\mathrm{d}^2/\mathrm{d}\tau^2 - \nabla^2 + U''(\phi_c)\big)}{\det\big(-\mathrm{d}^2/\mathrm{d}\tau^2 - \nabla^2 + U''(\phi_-)\big)}\right|^{-\frac{1}{2}} \mathrm{e}^{-S_c} \tag{3.27}$$

at leading order (i.e. at one loop). This is the final formula for the decay rate in a scalar theory.

The only subtlety here which does not appear in QM is the issue of renormalization. In a scalar theory there will be divergences which have to be removed by adding counterterms. We then have the renormalized action

$$S = S_R + \sum_{n=1}^{\infty} \hbar^n S^{(n)} \tag{3.28}$$

where $S^{(n)}$ includes the counterterms related to a calculation at n loops (we have included $\hbar$ factors explicitly). We then perform the calculations above with the *renormalized* action, and then we incorporate the effects of loops. In the full theory, it might happen that

$$S(\phi_-) \neq 0, \tag{3.29}$$

therefore we change

$$\mathrm{e}^{-S_c} \to \mathrm{e}^{-(S_c - S(\phi_-))}. \tag{3.30}$$

The bounce ϕ_c is now computed for S_R. If we compute it for the full action, it will have corrections as

$$\phi_c \to \phi_c + \hbar\phi^{(1)} + \cdots \tag{3.31}$$

where $\phi^{(1)}$ is induced by the first order correction to the action, and as we will see immediately, at one-loop is not necessary to compute it. We then have

$$\begin{aligned} S(\phi) &= S_R(\phi_c + \hbar\phi^{(1)} + \cdots) + \hbar S^{(1)}(\phi_c + \hbar\phi^{(1)} + \cdots) + \cdots \\ &= S_R(\phi_c) + \hbar S^{(1)}(\phi_c) + \cdots \end{aligned} \tag{3.32}$$

since

$$\frac{\delta S_R}{\delta\phi}(\phi_c) = 0 \tag{3.33}$$

by construction. We then find

$$\begin{aligned} \operatorname{Im} E/V &= \frac{1}{2}\frac{S_c^{d/2}\beta}{(2\pi)^{d/2}}\left|\frac{\det'\big(-\mathrm{d}^2/\mathrm{d}\tau^2 - \nabla^2 + U''(\phi_c)\big)}{\det\big(-\mathrm{d}^2/\mathrm{d}\tau^2 - \nabla^2 + U''(\phi_-)\big)}\right|^{-\frac{1}{2}} \mathrm{e}^{-S_c + S(\phi_-)} \\ &\approx \frac{1}{2}\frac{S_R^{d/2}(\phi_c)V\beta}{(2\pi)^{d/2}}\left|\frac{\det'\big(-\mathrm{d}^2/\mathrm{d}\tau^2 - \nabla^2 + U''(\phi_c)\big)}{\det\big(-\mathrm{d}^2/\mathrm{d}\tau^2 - \nabla^2 + U''(\phi_-)\big)}\right|^{-\frac{1}{2}} \mathrm{e}^{-S_R(\phi_c) - \hbar S^{(1)}(\phi_c) + S^{(1)}(\phi_-)} \end{aligned} \tag{3.34}$$

where we have used that $S_R(\phi_-) = 0$. This is our final, UV finite expression, since the divergences of the one-loop determinants are taken care of by the one-loop counterterms of the effective action. In physical terms, what we have calculated is the probability per unit time for the formation of a tiny bubble of true vacuum in a given unit volume of space. At leading order we assume that bubbles do not interact and this probability is simply proportional to the volume.

As in the QM example, the only nontrivial piece in the expression for the decay rate (3.34) is the functional determinant, which can be calculated by generalizing the QM results. For the potential (3.2) very detailed results are presented in [5, 39].

3.2 The fate of the false vacuum

What happens after the quantum bubble has materialized? This is very similar to what happens to a particle which has crossed a potential barrier. Such a particle materializes at the point where the potential energy is zero, which is the point $q_c(t_0)$ of the trajectory (see for example Fig. 9). It has zero kinetic energy at that point. Starting from those conditions it propagates in the potential, and we can describe this process with classical mechanics.

Something similar happens with the bubble. After materializing past the barrier at the time $t = 0$ it will evolve with initial conditions

$$\begin{aligned} \phi(t=0,\vec{x}) &= \phi_c(\vec{x},\tau=0), \\ \partial_t\phi(t=0,\vec{x}) &= 0. \end{aligned} \tag{3.35}$$

The last condition is the analogue of $\dot{q}_t = 0$. The evolution will be governed by the wave equation

$$(\nabla^2 - \partial_t^2)\phi = U'(\phi). \tag{3.36}$$

Interestingly, we can solve this equation easily. Take the $O(4)$ invariant bounce $\phi_c(r)$ and define

$$\phi(t,\vec{x}) = \phi_c(r = (\vec{x}^2 - t^2)^{\frac{1}{2}}). \tag{3.37}$$

This solves the equation above with the same initial conditions (3.35). The first condition is obvious. Since

$$\partial_t\phi = \frac{\mathrm{d}\phi_c}{\mathrm{d}r}\partial_t r = -\frac{t}{r}\frac{\mathrm{d}\phi_c}{\mathrm{d}r} \tag{3.38}$$

vanishes at $t = 0$, the second condition is also satisfied.

What is then the evolution of the bubble? Let us assume for simplicity that the bounce is of the form depicted in (17) for $\alpha \approx 1$. Then, at $t = 0$ we have a bubble of true vacuum at the origin, of radius R. The boundary of the bubble simply expands at the speed of light, following the hyperboloid

$$\vec{x}^2 = t^2 + R^2. \tag{3.39}$$

Notice that this is a Lorentz–invariant evolution, i.e. it has $O(3,1)$ symmetry inherited from the $O(4)$ invariance of the bounce.

3.3 Instability of the Kaluza–Klein vacuum

The same techniques we have used to discuss unstable vacua in scalar field theories can be used to analyze other theories. A particularly striking application of these ideas is the semiclassical instability of the Kaluza–Klein vacuum, discovered by Witten in [90].

In this case, the field is the Riemannian metric of a five-dimensional manifold. The classical theory of such a field is of course general relativity. In the Kaluza–Klein approach one assumes that the ground state (the vacuum) is a manifold of the form

$$X_5 = M_4 \times \mathbb{S}^1, \tag{3.40}$$

where M_4 is Minkowski space and $\mathbb{S}^1$ is a circle of radius R. A complete analysis of this problem is only possible in a full quantum theory of gravity in five dimensions, which at this stage can be only obtained by some appropriate compactification of string theory. However, one can use semiclassical considerations, in the spirit of Euclidean quantum gravity, to decide about the stability of the Kaluza–Klein vacuum.

Indeed, it is clear from the analysis in previous sections that semiclassical stability can be determined with purely classical data. To look for an instability one has to look for a bounce solution to the classical euclidean field equations, i.e. a solution which asymptotically approaches the vacuum we want to analyze, and such that it has one negative mode –and these are questions that in principle can be addressed without having a complete quantum treatment of the model.

We start then with the metric for the standard KK vacuum, continued to Euclidean space. This is a *constant* metric:

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \mathrm{d}\tau^2 + \mathrm{d}\phi^2. \tag{3.41}$$

The first four terms correspond to the Euclidean metric in $\mathbb{R}^4$, while the last term is the angle for $\mathbb{S}^1$ (therefore, it is a periodic variable). Using polar coordinates for $\mathbb{R}^4$ we have

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\Omega^2 + \mathrm{d}\phi^2, \tag{3.42}$$

where

$$r = \sqrt{x^2 + y^2 + z^2 + \tau^2}. \tag{3.43}$$

Is there a bounce? It turns out that

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1-\alpha/r^2} + r^2\mathrm{d}\Omega^2 + \Big(1 - \frac{\alpha}{r^2}\Big)\mathrm{d}\phi^2 \tag{3.44}$$

is asymptotically constant and solves Einstein's equations in 5d. This is in fact the Euclidean section of the 5d Schwarzschild solution. There is a singularity at $r = \alpha$, and to analyze it we follow the same logic as in the Euclidean continuation of the 4d Schwarzschild solution [45], i.e. we want to interpret the singularity at $r = \alpha$ as an apparent singularity due to the fact that the the coordinate ϕ is *periodic*. Indeed, the flat metric in polar coordinates

$$\mathrm{d}s^2 = \mathrm{d}\rho^2 + \rho^2\mathrm{d}\phi^2 \tag{3.45}$$

has an apparent singularity at $\rho = 0$, but this is just due to the choice of coordinates. We want to find a new coordinate such that the part involving r, ϕ in the metric (3.44) looks like (3.45) near $x = 0$, so that the singularity at $r = \alpha$ can be removed by making ϕ periodic with an appropriate period. Let us set

$$\rho = c\Big(1 - \frac{\alpha}{r^2}\Big)^{\beta}, \tag{3.46}$$

where c, β are constants to be determined by our requirements. We have

$$\begin{aligned} r^2 &= \frac{\alpha}{1-\left(\frac{\rho}{c}\right)^{\frac{1}{\beta}}}, \\ \mathrm{d}r &= \frac{r^3}{2c\alpha\beta}\left(1-\frac{\alpha}{r^2}\right)^{1-\beta}, \end{aligned} \tag{3.47}$$

and we deduce

$$\mathrm{d}r^2 = \frac{\alpha}{4c^2\beta^2}\left(\frac{\rho}{c}\right)^{\frac{2(1-\beta)}{\beta}} \frac{\mathrm{d}\rho^2}{\left(1-\left(\frac{\rho}{c}\right)^{\frac{1}{\beta}}\right)^3}, \tag{3.48}$$

as well as

$$\frac{\mathrm{d}r^2}{1-\alpha/r^2} = \frac{\alpha}{4c^2\beta^2}\left(\frac{\rho}{c}\right)^{\frac{1-2\beta}{\beta}} \frac{\mathrm{d}\rho^2}{\left(1-\left(\frac{\rho}{c}\right)^{\frac{1}{\beta}}\right)^3}. \tag{3.49}$$

We want this to look like $\mathrm{d}\rho^2$ near $\rho = 0$, therefore

$$\beta = \frac{1}{2}, \qquad c = \alpha^{\frac{1}{2}}. \tag{3.50}$$

We can now analyze the periodic part,

$$\left(1-\frac{\alpha}{r^2}\right)\mathrm{d}\phi^2 = \left(\frac{\rho}{c}\right)^{\frac{1}{\beta}}\mathrm{d}\phi^2, \tag{3.51}$$

which for $\beta = 1/2$ indeed gives

$$\frac{\rho^2}{c^2}\mathrm{d}\phi^2 = \rho^2\mathrm{d}\left(\frac{\phi}{c}\right)^{\frac{1}{2}}. \tag{3.52}$$

We conclude that ϕ is periodic with period

$$2\pi c = 2\pi\sqrt{\alpha}. \tag{3.53}$$

Since in the original Kaluza–Klein metric ϕ has period $2\pi R$, where R is the radius of the fifth compact dimension, we find

$$\alpha = R^2. \tag{3.54}$$

Therefore, the metric reads

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1-(R/r)^2} + r^2\mathrm{d}\Omega_3^2 + \left(1-\left(\frac{R}{r}\right)^2\right)\mathrm{d}\phi^2 \tag{3.55}$$

Notice that the radial coordinate x starts at $x = 0$, but this means by looking at (3.46) that

$$r \geq R. \tag{3.56}$$

In order to see what is the instability associated to this solution, we recall the analysis of the bounce in the scalar field theory. After continuation to Minkowski space, the bounce solution represents a bubble of true vacuum. Therefore, in analogy with scalar field theory, we want to rotate the metric (3.55) to Minkowski signature. If we did this in the flat case, we would require

$$\mathrm{d}r^2 + r^2\mathrm{d}\Omega_3^2 \to \mathrm{d}x^2 + x^2\mathrm{d}\Omega_2^2 - \mathrm{d}t^2, \tag{3.57}$$

where

$$x^2 = x_1^2 + x_2^2 + x_3^2 \tag{3.58}$$

is the radius of the space-like $\mathbb{R}^3$ coordinates in Minkowski space. To do that, we pick a polar angle θ and we write

$$\mathrm{d}\Omega_3^2 = \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\Omega_2^2. \tag{3.59}$$

We then define a new angle ψ by

$$\theta = \frac{\pi}{2} + \mathrm{i}\psi \tag{3.60}$$

and we introduce the coordinates

$$x = r\cosh\psi, \qquad t = r\sinh\psi, \tag{3.61}$$

so that

$$\begin{aligned} r &= \sqrt{x^2 - t^2}, \qquad &\mathrm{d}r = \frac{1}{r}(x\mathrm{d}x - t\mathrm{d}t), \\ \psi &= \tanh^{-1}\frac{t}{x}, \qquad &\mathrm{d}\psi = \frac{1}{r^2}(x\mathrm{d}t - t\mathrm{d}x). \end{aligned} \tag{3.62}$$

The line element of the three-sphere becomes

$$\mathrm{d}\Omega_3^2 = -\mathrm{d}\psi^2 + \cosh^2\psi \mathrm{d}\Omega_2^2, \tag{3.63}$$

and the 4d Euclidean space metric becomes,

$$\begin{aligned} \mathrm{d}r^2 + r^2\mathrm{d}\Omega_3^2 &= \mathrm{d}r^2 + r^2\left(-\mathrm{d}\psi^2 + \cosh^2\psi \mathrm{d}\Omega_2^2\right) \\ &= \mathrm{d}x^2 - \mathrm{d}t^2 + x^2\mathrm{d}\Omega_2^2, \end{aligned} \tag{3.64}$$

as we wanted. Notice however that in this parametrization $x^2 - t^2 = r^2 > 0$ is always positive. Therefore, to be precise, the above continuation (3.64) describes rather the *exterior of the light cone* in Minkowski space.

Now, for the bounce solution, the same continuation and change of variables can be performed, and we obtain the metric

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1-(R/r)^2} + r^2\left(-\mathrm{d}\psi^2 + \cosh^2\psi \mathrm{d}\Omega_2^2\right) + \left(1 - \left(\frac{R}{r}\right)^2\right)\mathrm{d}\phi^2. \tag{3.65}$$

Here, $r^2 = x^2 - t^2$, and on top of that it starts at $r = R$. Therefore, the omitted part of space here is the full hyperboloid interior bounded by

$$x^2 - t^2 = R^2, \tag{3.66}$$

see Fig. 18. One could think that this is an ugly space with a boundary. But the presence of the fifth dimension gives in the end a space which is non-singular and geodesically complete, since the circle has now radius

$$R(1 - R^2/r^2)^{\frac{1}{2}} \tag{3.67}$$

which is zero when $r \to R$. Restricted to the x axis, this smooths out what would be the complement of an interval into two *discs*, see Fig. 19. We can now interpret this solution. The instability is generated by the nucleation of a *hole* of radius R (therefore very small) in three-dimensional space with $O(3)$ distance x. From the point of view of a 4d Minkowski observer, this is a *hole of nothing* which forms at $t = 0$ and then expands at the speed of light according to

$$x^2 = R^2 + t^2. \tag{3.68}$$

Therefore, the Kaluza–Klein vacuum decays into literally *nothing*.

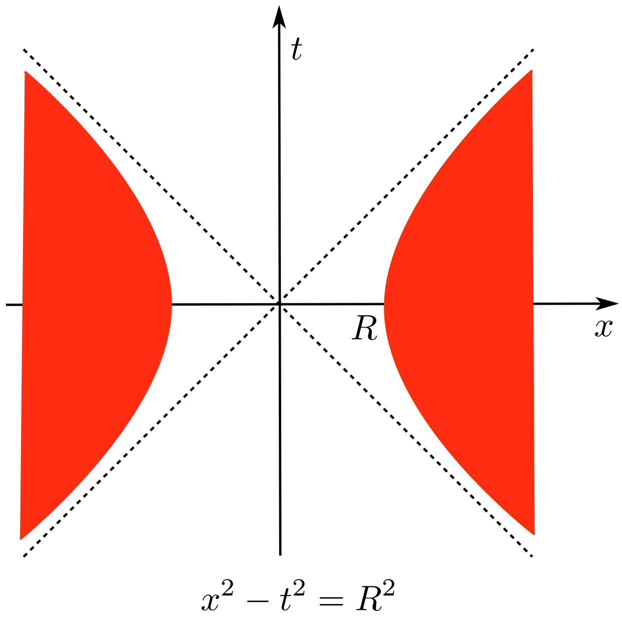


Figure 18: Restricted to the $x - t$ plane, the space described by the metric (3.65) is the exterior of the hyperboloid $x^2 - t^2 = R^2$ (in red in the figure).

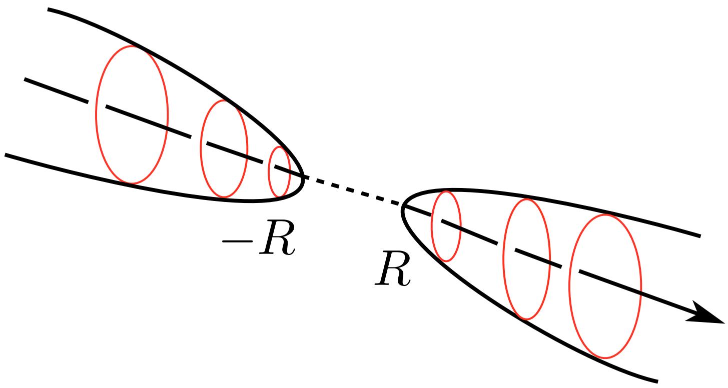


Figure 19: The fifth dimension in (3.65) is a circle fibered over the four-dimensional space, with radius (3.67). As we approach the boundary $r^2 = R^2$, the circle shrinks to zero size and the total space is smooth. When restricted to the x axis, as in the figure, the fifth dimension smooths out the complement of the interval into two discs.

4. Large order behavior and Borel summability

4.1 Perturbation theory at large order

Let us consider a quantum system in which one computes a quantity Z as a perturbation series in a parameter g,

$$E(g) = \sum_k a_k g^k. \tag{4.1}$$

A typical example of this is the anharmonic oscillator with Hamiltonian

$$H = \frac{p^2}{2} + \frac{x^2}{2} + \frac{g}{4}x^4. \tag{4.2}$$

where $E(g)$ is the energy of the ground state computed in stationary perturbation theory. There are various questions that we can ask about this kind of series:

1. *Large order behavior*. What is the radius of convergence of (4.1)? We will see that very often we have *zero radius of convergence*.
2. *Summability*. In case the above series has zero radius of convergence, is there a way to make sense of the perturbative series? We will see that in some situations this can be done (for example, if the series is Borel summable).

Dyson [40] has provided a general argument why series like the one for the anharmonic oscillator have zero radius of convergence. If this radius was finite, the series for $E(g)$ would describe the physics of the problem also for a small $g < 0$. But for negative coupling, the physics is completely different: we have an unstable particle which will eventually decay. Therefore, we should *not* expect a nonzero radius of convergence[1]

Dyson's argument indicates that there is a deep connection between the imaginary part of Z that gives the *tunneling amplitude* in the unstable potential with $g < 0$, and the large order behavior of perturbation theory. In general, there will be a connection between the *instantons* of the theory (which compute tunneling amplitudes) and the large order behavior. However, in renormalizable field theories there are sources of divergence (the *renormalons*) which dominate over the instantons.

4.2 The toy model integral, revisited

As a first approach to the problem of large orders in perturbation theory, we will revisit the toy model integral. We will see, by a simple contour deformation analysis, that the behavior of its power series expansion around $g = 0$ at large k, given in (2.39), is intimately related to its imaginary part (2.36), which is non-perturbative in g.

Let us first list some of the properties of $I(x)$ as a function on the complex x-plane:

1. It is analytic with a branch cut along $(-\infty, 0)$.
2. At the origin it behaves like
$$\lim_{g\to 0} gI(g) = 0. \tag{4.3}$$
This is because the series (2.37) is asymptotic
3. At infinity it goes like
$$I(g) \sim g^{-1/4}, \qquad |g| \to \infty. \tag{4.4}$$

The last property follows from the following scaling argument. Make the following change of variables:

$$z = g^{-1/4}u, \tag{4.5}$$

[1]One should be careful with this general argument, since there are counterarguments to it. See [76], p. 4.

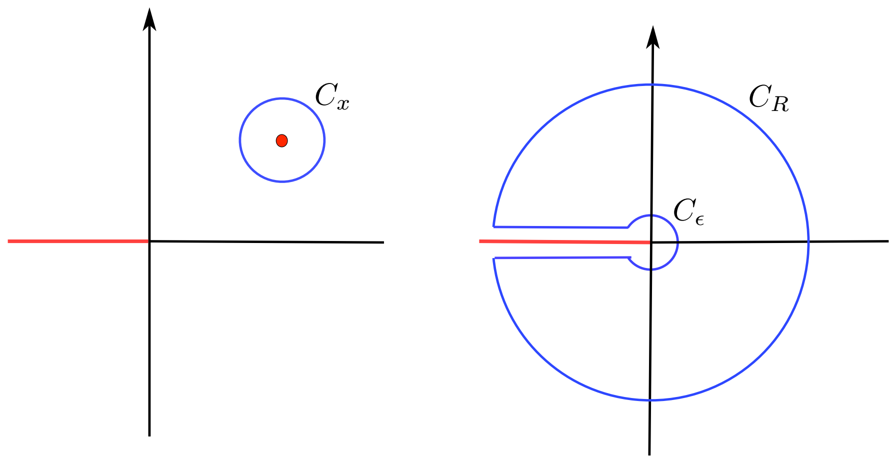


Figure 20: Contour deformation from the contour C_x around $x \in \mathbb{C}$. The red line represents the branch cut at $(-\infty, 0)$.

so that the integral reads

$$I(g) = \frac{g^{-\frac{1}{4}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}u \, \exp\left(-\frac{u^4}{4} - g^{-\frac{1}{2}}\frac{u^2}{2}\right). \tag{4.6}$$

As $g \to \infty$ the Gaussian part becomes unimportant, and we find

$$I(g) \approx \frac{g^{-\frac{1}{4}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}u \, \exp\left(-\frac{u^4}{4}\right) = \frac{\Gamma\left(\frac{1}{4}\right)}{\sqrt{4\pi}} g^{-\frac{1}{4}}, \qquad g \to \infty \tag{4.7}$$

Let C_x be a contour around a point $x \in \mathbb{C}$ away from the branch cut, as in the left hand side of Fig. 20. Cauchy's theorem gives

$$I(z) = \frac{1}{2\pi \mathrm{i}} \oint_{C_x} \mathrm{d}x \frac{I(x)}{x - z}. \tag{4.8}$$

We can now deform the contour to encircle the branch cut in the negative real axis, as in the right hand side of Fig. 20. The contributions from the contours at infinity C_R and around the origin C_ϵ vanish, thanks to (4.4) and (4.3), respectively. The only remaining contribution comes from the lines which are parallel to the branch cut,

$$I(z) = \frac{1}{2\pi \mathrm{i}} \int_{-\infty}^{0} \mathrm{d}x \frac{D(x)}{x - z}, \tag{4.9}$$

where $D(x)$ is the discontinuity across the negative, real axis,

$$D(x) = \lim_{\epsilon \to 0} (I(x + \mathrm{i}\epsilon) - I(x - \mathrm{i}\epsilon)) = 2\mathrm{i} \, \mathrm{Im} \, I(x + \mathrm{i}0_+), \tag{4.10}$$

i.e. we have

$$I(z) = \frac{1}{\pi} \int_{-\infty}^{0} \mathrm{d}x \frac{\mathrm{Im} \, I(x)}{x - z}. \tag{4.11}$$

We the find the following integral representation for the coefficients a_k in (2.37)

$$a_k = \frac{1}{\pi}\int_{-\infty}^{0} \mathrm{d}x\, \frac{\mathrm{Im}\, I(x)}{x^{k+1}} = \frac{(-1)^{k+1}}{\pi}\int_0^{\infty} \mathrm{d}x \frac{\mathrm{Im}\, I(-x)}{x^{k+1}}, \quad k \geq 1. \tag{4.12}$$

Therefore, if we know $\mathrm{Im}\, I(g)$, we can plug it in here to obtain the asymptotics of a_k. Moreover, at large k the above integral will be controlled by the behavior of $\mathrm{Im}\, I(g)$ at *small, negative* g. This is precisely the quantity which is governed by the *non-trivial* saddle points in (2.33)!

This result is just an example of a *dispersion relation*, which makes possible to relate the behavior of a quantity at different regimes of its control parameter. In this case, we have been able to relate a phenomenon at strong coupling (the large g behavior) with a phenomenon at weak coupling (a saddle point calculation).

Let us assume that the discontinuity across the cut

$$\mathrm{disc}\, I(-g) = \lim_{\epsilon\to 0}(I(-g+\mathrm{i}\epsilon) - I(-g-\mathrm{i}\epsilon)) = 2\mathrm{i}\,\mathrm{Im}\, I(-g) \tag{4.13}$$

is of the form

$$\mathrm{disc}\, I(-g) = \mathrm{i} g^{-b} \mathrm{e}^{-A/g} \sum_{n=0}^{\infty} c_n g^n, \qquad g > 0 \tag{4.14}$$

This leads to the following behavior for a_k:

$$\begin{aligned} a_k &= \frac{(-1)^{k+1}}{2\pi} \sum_{n=0}^{\infty} c_n \int_0^{\infty} \frac{\mathrm{d}s}{s^{k+1}} s^{-b+n} \mathrm{e}^{-A/s} \\ &= \frac{(-1)^{k+1}}{2\pi} \sum_{n=0}^{\infty} c_n A^{-k-b+n} \int_0^{\infty} \mathrm{d}x\, x^{k+b+1-n} \mathrm{e}^{-x} = \frac{(-1)^{k+1}}{2\pi} \sum_{n=0}^{\infty} c_n A^{-k-b+n} \Gamma(k+b-n), \end{aligned} \tag{4.15}$$

which can be also written as (see [27])

$$a_k \sim \frac{(-1)^{k+1} A^{-b-k}}{2\pi} \Gamma(k+b) \left[c_0 + \frac{c_1 A}{k+b-1} + \frac{c_2 A^2}{(k+b-2)(k+b-1)} + \cdots \right]. \tag{4.16}$$

Note that the *leading term* is the factorial $k!$. The subleading piece is captured by A^{-k}. Since we know from (2.36) that $A = 1/4$, we reproduce precisely the asymptotic behavior (2.39).

We then conclude that the large order behavior of the perturbative series around the trivial saddle point encodes the information about the non-trivial saddle points.

4.3 The anharmonic oscillator

The above considerations can be repeated almost *verbatim* for the quartic anharmonic oscillator. The energy of the ground state can be computed as a power series in g in perturbation theory. One possibility to do that is, as in section 2.1, to use Feynman diagrams. The resulting series is of the form:

$$E(g) = \sum_{k=0}^{\infty} a_k g^k, \quad a_0 = \frac{1}{2} \tag{4.17}$$

Since this system is unstable for $g < 0$, we expect this series to have zero radius of convergence, by Dyson's argument. This connection can be made very precise, as first done in the pioneering paper of Bender and Wu [11]. The argument follows the one made in the last subsection. First, we introduce the function

$$f(z) = \frac{1}{z}(E(z) - a_0) = \sum_{k=0}^{\infty} f_k z^k, \quad f_k = a_{k+1}. \tag{4.18}$$

As a function on the complex z-plane it has the following properties:

1. As in the case of the quartic integral (2.25), it is analytic in the complex plane with a cut along $(-\infty, 0)$.
2. At the origin it behaves like
$$\lim_{z\to 0} z f(z) = 0. \tag{4.19}$$
This is because the series (4.17) is asymptotic.
3. At infinity it goes like
$$|f(z)| \sim |z|^{-2/3}. \tag{4.20}$$

The first and the second property can be proved rigorously (see for example [76] for a review and references). The last property follows from a simple scaling argument (see for example [42], p. 171). At large g, we have that

$$H \sim \frac{p^2}{2} + g\frac{x^4}{4}. \tag{4.21}$$

If we rescale $x \to g^{-1/6}x$, we have

$$H \to g^{\frac{1}{3}}\Big(\frac{p^2}{2} + \frac{x^4}{4}\Big), \tag{4.22}$$

therefore the energy will be

$$E(g) \sim C g^{\frac{1}{3}}, \qquad g \to \infty, \tag{4.23}$$

where C is the energy of the ground state of the Hamiltonian $p^2/2 + x^4/4$ in (4.22).

We can now apply to $f(x)$ the same argument we applied above to $I(x)$, and consider the contour deformation of Fig. 20. We find again,

$$f(z) = \frac{1}{\pi}\int_{-\infty}^{0} \mathrm{d}x \frac{\mathrm{Im}\, f(x)}{x - z}. \tag{4.24}$$

In terms of the original quantity, the ground state energy, we have

$$E(g) = a_0 + \frac{g}{\pi}\int_{-\infty}^{0} \mathrm{d}g' \frac{\mathrm{Im}\, E(g')}{g'(g' - g)}. \tag{4.25}$$

Since

$$\frac{g}{g'}\frac{\mathrm{Im}\, E(g')}{g' - g} = \sum_{k\geq 0} g^{k+1} \frac{\mathrm{Im}\, E(g')}{(g')^{k+2}}. \tag{4.26}$$

we find again the integral representation

$$a_k = \frac{(-1)^{k+1}}{\pi} \int_0^\infty \mathrm{d}z \frac{\mathrm{Im}\, E(-z)}{z^{k+1}}, \quad k \geq 1. \tag{4.27}$$

Equivalently, we can write

$$a_k = \frac{(-1)^{k+1}}{2\pi \mathrm{i}} \int_0^\infty \mathrm{d}z \frac{\mathrm{disc}\, E(-z)}{z^{k+1}}. \tag{4.28}$$

This now relates the coefficients of the power series in stationary perturbation theory to the tunneling amplitude which we computed with instanton calculus.

Assuming for $\mathrm{disc}\, E(-z)$ the same structure than in (4.14), which is indeed what we found in the instanton calculation, we find that the coefficients a_k in (4.17) have the large k asymptotics given in (4.16). This asymptotics involves all the data of the one-instanton amplitude, like the classical action of the instanton A and the one-loop contribution c_0.

Remark 4.1. What one computes in the path integral, instanton calculation is precisely $\mathrm{disc}\, E(-z)$. In other words, the instanton contribution is 2i the imaginary part of $E(z)$.

We can now use the formula (2.167) for $\mathrm{Im}\, E$ in the quartic oscillator and read from it the quantities appearing in (4.14) at one-loop,

$$b = \frac{1}{2}, \quad c_0 = 2\sqrt{\frac{2}{\pi}}, \quad A = 4/3. \tag{4.29}$$

We then find

$$a_k \sim (-1)^{k+1} \frac{\sqrt{6}}{\pi^{3/2}} \left(\frac{3}{4}\right)^k \Gamma\left(k + \frac{1}{2}\right), \tag{4.30}$$

which is the famous result of [11]. We will see later on a more general structure for the asymptotic behaviour and its relation to Borel transforms.

The *main conclusion* of this analysis is that, indeed, the perturbative series for the anharmonic oscillator has zero radius of convergence, as expected from Dyson's argument: for negative coupling the theory becomes unstable, so analyticity at $g = 0$ is impossible. Moreover, an analysis of this instability, in terms of instanton configurations, makes possible to give a precise and quantitative characterization of the asymptotics of perturbation theory.

In this example, and also in the toy quartic integral, it is possible to show that the factorial growth of a_k is due to the *factorial growth of the number of Feynman diagrams* (see [7] for an excellent survey). Recall from section 2.1 that a_n can be computed as a sum over connected quartic graphs. The total number of disconnected graphs is simply given by the quartic integral (2.25), which asymptotically as $n \to \infty$ behaves like (see (2.39))

$$4^n n!, \tag{4.31}$$

i.e. there is a *factorial growth* in the number of disconnected diagrams. One could think that there might be a substantial reduction in this number when we consider *connected* diagrams, but a detailed analysis [8] shows that this is not the case: at large n, the quotient of the number of connected and disconnected diagrams differs from 1 only in $\mathcal{O}(1/n)$ corrections. We conclude that there are $\sim n!$ diagrams that contribute to a_n. This leads to the factorial behavior in (4.30). In fact, one can derive the leading and subleading behavior in (4.30) by a detailed statistical analysis of Feynman diagrams [12].

4.4 Asymptotic expansions and Borel resummation

We have seen that the perturbative series for the ground state energy of the quartic oscillator is divergent, and this turns out to be a *generic* feature of perturbative series in quantum theories. The type of series that we find are *asymptotic* in the sense of Poincaré, i.e. if the series is given by

$$S(w) = \sum_{n=0}^{\infty} a_n w^n \tag{4.32}$$

then one has that

$$\lim_{w \to 0} w^{-N} \Big(S(w) - \sum_{n=0}^{N} a_n w^n \Big) = 0 \tag{4.33}$$

for all $N > 0$. Analytic functions might have asymptotic expansions. For example, the Stirling series for the Gamma function

$$\left(\frac{z}{2\pi}\right)^{1/2} \left(\frac{z}{e}\right)^{-z} \Gamma(z) = 1 + \frac{1}{12z} + \frac{1}{288z^2} + \cdots \tag{4.34}$$

is an asymptotic series for $|z| \to \infty$. We will consider series of the form (4.32) where, as in the quantum quartic oscillator, the coefficients a_n grow factorially at large n,

$$a_n \sim A^{-n} n! \tag{4.35}$$

The partial sums of the series (4.32) are defined, as usual, by

$$S_N(w) = \sum_{n=1}^{N} a_n w^n \tag{4.36}$$

We say that the series (4.32) is a *strong asymptotic series* for $f(w)$ if for all N there exists a bound,

$$\big| f(w) - S_N(w) \big| \le C_{N+1} |w|^{N+1} \tag{4.37}$$

with

$$C_N = c A^{-N} N!. \tag{4.38}$$

Notice that different functions may have the same asymptotic expansion, since

$$f(w) + C \mathrm{e}^{-A/w} \tag{4.39}$$

has the same expansion around $w = 0$ than $f(w)$, for any C, A. This can be also seen as follows. Let us suppose that we have a function $f(w)$ with a strong asymptotic expansion like (4.32) at $w = 0$. The partial sums (4.36) will first approach the true value $f(w)$, and then, for N sufficiently big, they will diverge. Imagine that you want to obtain the partial sum which gives the best possible estimate of $f(w)$. Then, one has to find the N that truncates the asymptotic expansion in an optimal way, and such a procedure is called *optimal truncation*. In order to do that, we should find the N that minimizes the bound in (4.37)

$$C_N |w|^N = c N! \left(\frac{|w|}{A} \right)^N. \tag{4.40}$$

By using the Stirling approximation, we rewrite this as

$$c\exp\Big\{N\big(\log N - 1 - \log X\big)\Big\}, \tag{4.41}$$

where

$$X = \frac{A}{|w|}. \tag{4.42}$$

The above function has a saddle at large N given by

$$N_* = X, \tag{4.43}$$

and for this value of N the bound on the asymptotics is of the form

$$\epsilon(w) = C_{N_*}|w|^{N_*} \sim \mathrm{e}^{-A/|w|}. \tag{4.44}$$

Therefore, the maximal "resolution" we can expect when we reconstruct a function $f(w)$ from its an asymptotic expansion is of order $\epsilon(w)$. This ambiguity present in an asymptotic series is sometimes called (in the context of quantum theory) the *nonperturbative ambiguity*.

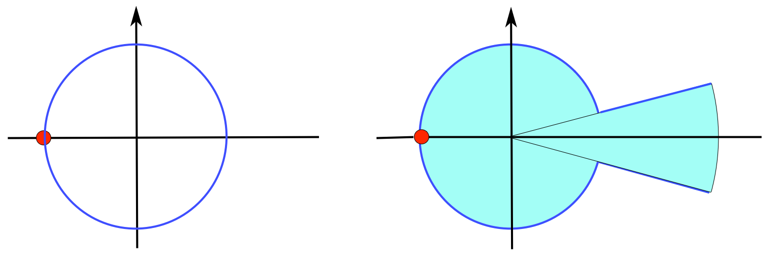


Figure 21: The Borel transform is analytic in a neighbourhood of $z = 0$, of radius $\rho = A$. Typically we encounter a singularity on the circle $|z| = A$, but we can analytically continue to a wider region. If this region includes a neighbourhood of the positive real axis, and the resulting function decays sufficiently fast at infinity, we say that the series is Borel summable.

In fact, we can do better than optimal truncation and take into account the information in *all* the terms of the series. The way to do that is *Borel resummation*, which is the standard tool to deal with divergent series. Let us consider a series (4.32), where the coefficients a_n behave like (4.35) when n is large (such series are called *Gevrey-1* in mathematics). The *Borel transform* of S, $B_S(z)$, is defined as the series

$$B_S(z) = \sum_{n=0}^{\infty} \frac{a_n}{n!} z^n. \tag{4.45}$$

Notice that, due (4.35), the series $B_S(z)$ has a finite radius of convergence $\rho = A$ and it defines an analytic function in the circle $|z| < A$. Typically, there is a singularity at the boundary of this region $|z| = A$, like a pole or a branch cut, but very often the resulting function can be analytically continued to a wider region of the complex plane.

Example 4.2. Consider

$$S(w) = \sum_{n=0}^{\infty} (-1)^n n! w^n. \tag{4.46}$$

In this case, the Borel transform is

$$B_S(z) = \sum_{n=0}^{\infty} (-1)^n z^n, \tag{4.47}$$

which is a series with radius of convergence $\rho = 1$. However, it is an elementary fact that this series can be analytically continued to a meromorphic function with a single pole at $z = -1$, namely

$$B_S(z) = \frac{1}{1+z} \tag{4.48}$$

Example 4.3. Consider now the series

$$S(w) = \sum_{k=0}^{\infty} \frac{\Gamma(k+b)}{\Gamma(b)} A^{-k} s^k. \tag{4.49}$$

The Borel transform is given by

$$B_S(z) = \sum_{k=0}^{\infty} \frac{\Gamma(k+b)}{k!\Gamma(b)} A^{-k} z^k = (1 - z/A)^{-b}, \tag{4.50}$$

which has a singularity at $z = A$ as well as a branch cut starting at that point. We can extend this to $b = 0$, and we obtain in this way a logarithmic branch cut for $B_S(s)$.

Suppose now that the Borel transform $B_S(z)$ has an analytic continuation to the whole sector $|\arg(z)| < \epsilon$ (i.e. to a neighbourhood of $(0, \infty)$), and that the integral

$$f(w) = \int_0^{\infty} \mathrm{d}t \, \mathrm{e}^{-t} B_S(tw) = w^{-1} \int_0^{\infty} \mathrm{d}t \, \mathrm{e}^{-t/w} B_S(t), \tag{4.51}$$

is absolutely convergent for small w. In this case, we say that $S(w)$ is *Borel summable*. Notice that, by construction, $f(w)$ has an asymptotic expansion around $w = 0$ which coincides with the original series $S(w)$, since

$$f(w) = w^{-1} \sum_{n \geq 0} \frac{a_n}{n!} \int_0^{\infty} \mathrm{d}t \, \mathrm{e}^{-t/w} t^n = \sum_{n \geq 0} a_n w^n = S(w) \tag{4.52}$$

However, if this series is Borel summable, $f(w)$ is an analytic function at $w = 0$, and we have then reconstructed the exact nonperturbative result from the asymptotic series.

Example 4.4. In the Example (4.2), the Borel transform extends to an analytic function on $\mathbb{C} \backslash \{-1\}$, and the integral (4.51) is

$$f(w) = \int_0^{\infty} \mathrm{d}t \, \mathrm{e}^{-t} \frac{1}{1+wt}, \tag{4.53}$$

which exists and is well defined for all $w \leq 0$. Therefore, we can resum the original divergent series (4.46) for all $w \leq 0$.

Therefore, if a divergent series is Borel summable, we can in principle use the method of Borel transforms to obtain its true value, at least for some values of the parameters. However, in practice one only knows a few coefficients of the original series, and this makes very difficult the procedure of analytic continuation to a neighbourhood of the positive axis. We need a practical method to find accurate approximations to the resulting function. A useful method, first proposed in [47], is to use *Padé approximants*. Given a series

$$S(z) = \sum_{k=0}^{\infty} a_k z^k \tag{4.54}$$

the Padé approximant $[l/m]$ is given by a rational function

$$[l/m]_S(z) = \frac{p_0 + p_1 z + \cdots + p_l z^l}{q_0 + q_1 z + \cdots + q_m z^m}, \tag{4.55}$$

where q_0 is fixed to 1, and one requires that

$$f(z) - [l/m]_S(z) = \mathcal{O}(z^{l+m+1}). \tag{4.56}$$

This fixes the coefficients involved in (4.55).

Given a series $S(z)$ we can construct the Padé approximant of its Borel transform

$$\mathcal{P}_n^S(z) = \Big[[n/2]/[(n+1)/2]\Big]_{B_S} \tag{4.57}$$

which requires knowledge of its first $n+1$ coefficients. This is a rational function with various poles on the complex plane. If the Borel transform has for example a branch cut, the Padé approximant will mimick this by a series of poles along the cut. The first pole of the approximant will be close to the branch point of the Borel transform, and increasingly so as n grows. A good approximation to the Borel resummed series will then be an integral of the form (4.51) where one integrates instead $\mathcal{P}_n^S(z)$,

$$f_n(w) = w^{-1} \int_0^{\infty} \mathrm{d}t \, \mathrm{e}^{-t/w} \mathcal{P}_n^S(t). \tag{4.58}$$

Example 4.5. *The quartic integral.* The simplest example of this procedure is the quartic integral (2.25). The coefficients in the power series expansion (2.37) are factorially divergent, as shown in (2.39). Therefore, we can define the Borel transform of the original series. If we compare with the Borel transform in example 4.3, we see that in this example $A = -1/4$, and there should be a singularity in the Borel transform at

$$w = -\frac{1}{4} = -0.25. \tag{4.59}$$

If we compute the Padé approximants (4.57) we will find that for $n < 50$, all of their poles are on the real negative axis. The rightmost pole occurs, for $n = 10, 20, 30, 40$ at

$$-0.26185, \quad -0.253241, \quad -0.25149, \quad -0.250863, \tag{4.60}$$

which approaches the position of the true singularity. For $n = 40$ and $g = 0.4$, the integral of the Padé approximant is

$$f_{40}(0.4) = 0.85760858538... \tag{4.61}$$

This value can be compared to the numerical evaluation of the integral

$$Z(0.4) = 0.85760858529... \tag{4.62}$$

As we can see, the Borel resummation gives a value in remarkable agreement with the exact one.

The above procedure to reconstruct an analytic function $f(w)$ starting from the series $S(w)$ only holds when $S(w)$ is Borel summable, since in this case there is no ambiguity in the reconstruction. Suppose now that $B_S(z)$ has singularities on the *positive* real axis, and that it can be extended in a neigbourhood of the positive real axis as a meromorphic or multivalued function which decreases sufficiently fast at infinity. An example would be the function (4.50) when A is real and positive, corresponding to a series which is factorially divergent and *non-alternating*. Then, the integral (4.51) is ill-defined, but we can define it by deforming the contour in (4.51) appropriately in order to avoid the singularities and/or branch cut in the positive real axis. For example, we could choose contours $\mathcal{C}_\pm$ that avoid the singularities by encircling them from above or from below, respectively, as in Fig. 22. The functions obtained in this way

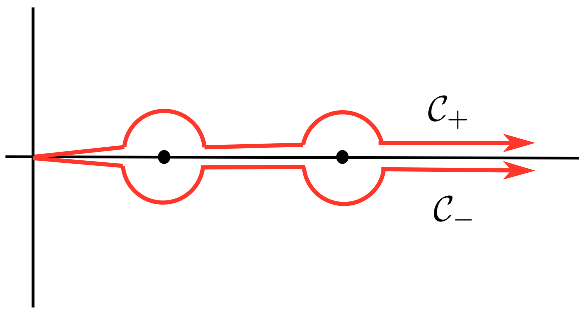


Figure 22: The paths $\mathcal{C}_\pm$ avoiding the singularities of the Borel transform from above (respectively, below).

$$f_\pm(w) = w^{-1} \int_{\mathcal{C}_\pm} \mathrm{d}t\, \mathrm{e}^{-t/w} B_S(t) \tag{4.63}$$

are called *lateral Borel transforms*. They pick an imaginary part due to the contour deformation, and their difference, which is purely imaginary is encoded in the discontinuity function

$$\epsilon(w) = \frac{1}{2\pi \mathrm{i}}(f_+(w) - f_-(w)) = \frac{1}{\pi} \mathrm{Im}\, f(w), \tag{4.64}$$

and is nonperturbative in w. Different choices of contour in the Laplace transform (4.51) lead to different functions $f(w)$, and in this case the nonperturbative ambiguity can be reformulated as the ambiguity in choosing a contour which avoids the singularities on the positive real axis. For example, the Borel transform $B_S(s)$ in (4.50), with a branch cut starting at A and with exponent $-b$, leads to the following form for $\epsilon(w)$

$$\epsilon(z) \sim z^{-b} \mathrm{e}^{-A/z}. \tag{4.65}$$

This can be seen by direct computation

$$
\begin{aligned}
\epsilon(z) &= \frac{1}{\pi} z^{-1} \int_A^\infty \mathrm{d}t\, \mathrm{e}^{-t/z} \mathrm{Im}\, B_S(t) = \frac{1}{\pi} z^{-1} \int_A^\infty \mathrm{d}t\, \mathrm{e}^{-t/z} \mathrm{Im}(1 - t/A)^{-b} \\
&= \frac{1}{\pi} z^{-1} \mathrm{e}^{-A/z} A^b \int_0^\infty \mathrm{d}u\, \mathrm{e}^{-u/z} u^{-b} \mathrm{Im}\, \mathrm{e}^{-\pi \mathrm{i} b} \\
&= \frac{1}{\pi} \mathrm{e}^{-A/z} z^{-b} A^b \Gamma(1-b) \sin \pi(1-b) \\
&= \frac{A^b}{\Gamma(b)} \mathrm{e}^{-A/z} z^{-b},
\end{aligned}
\tag{4.66}
$$

where in the second line we introduced the variable $u = t - A$, and in the last line we used the identity

$$
\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}. \tag{4.67}
$$

Example 4.6. Let us consider the series

$$
S(w) = \sum_{n=0}^{\infty} n! w^n. \tag{4.68}
$$

which is (4.46) but without the sign $(-1)^n$ (or we consider negative values of w). Then, the Borel transform is

$$
f(w) = \frac{1}{w} \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-t/w} \frac{1}{1-t}. \tag{4.69}
$$

This integral is ill-defined, and we have to give a prescription to avoid the pole at $t = 1$. The lateral Borel summations differ in this case by

$$
f_+(w) - f_-(w) = \frac{1}{w} \oint_C \mathrm{d}t\, \mathrm{e}^{-t/w} \frac{1}{1-t} = \frac{2\pi \mathrm{i}}{w} \mathrm{e}^{-1/w}, \tag{4.70}
$$

where C is a circle surrounding the pole. This is in fact a particular case of the computation we did in (4.66).

Whe conclude that, when there are singular points of $B_S(z)$ along the positive real axis, it is not possible to reconstruct the function $f(w)$ via Borel resummation just from its asymptotic expansion. Typically, one has to provide additional nonperturbative information to fix the ambiguity, or equivalently one has to choose a contour which avoids the singularities.

One can also use the method of Padé approximants to calculate the lateral resummations (4.63) of non-Borel summable functions. In this case there will be poles of $\mathcal{P}_n^S(t)$ along the real axis, but they are avoided by the contours $\mathcal{C}_\pm$. When there are also zeros away from the real axis, more complicated contours have been proposed which are useful in numerical computations [55].

4.5 Borel transforms and large order behavior

In many series appearing in quantum field theory and quantum mechanics, even if the coefficients are not as simple as in (4.49), their large k the asymptotic behavior is

$$
a_k \sim A^{-k} \Gamma(k+b), \quad k \gg 1, \tag{4.71}
$$

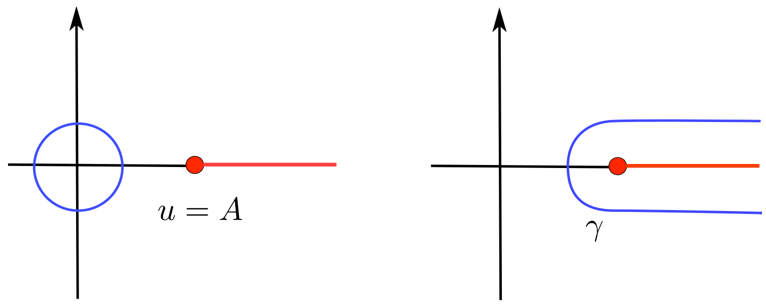


Figure 23: Contour deformation in the derivation of (4.75).

where A is a constant. This is for example the case for the quartic oscillator, see (4.30). As we have seen in the example 4.3, the Borel transform for such a series will be of the form

$$B_S(z) = C\big(1 - z/A\big)^{-b} + \cdots \tag{4.72}$$

with corrections due to the subleading terms in the asymptotics of a_k. The above expression shows that the Borel transform encodes the data in (4.71) determining the large order behavior of the coefficients a_k, independently of wether the series is Borel summable or not. Specifically, we have that

1. the factor A, which in the QM example discussed above is the action of the instanton, is given by the location of the singularity of the Borel transform which is closest to the origin.

2. b is the exponent characterizing the branch cut/pole structure of the Borel transform at $z = A$.

We conclude that *the large order behavior of a divergent series is controlled by the singularities of its Borel transform*, in particular by the singularity which is closest to the origin. Notice that we can obtain information about non-perturbative effects (like the instanton action) by looking at the Borel transform of the perturbative series.

Let us now present a general formula giving a quantitative expression for the relation between large order behavior and Borel transforms, assuming for concreteness that $A > 0$. If

$$B_S(z) = \sum_{k \geq 0} b_k z^k \tag{4.73}$$

then

$$a_k = k! b_k = \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-t} t^k b_k = \frac{1}{2\pi \mathrm{i}} \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-t} t^k \oint_0 \frac{\mathrm{d}u}{u^{k+1}} B_S(u). \tag{4.74}$$

Suppose then that $B_S(u)$ has a singularity at A along the positive real axis, with a cut going to infinity. By deforming the contour around the origin to the contour γ encircling the cut and the singularity, as shown in Fig. 23, we get

$$a_k = \frac{1}{2\pi \mathrm{i}} \int_0^\infty \mathrm{d}t \oint_\gamma \mathrm{d}s\, \mathrm{e}^{-t} \frac{t^k}{s^{k+1}} B_S(s) = \frac{1}{2\pi \mathrm{i}} \int_0^\infty \frac{\mathrm{d}z}{z^{k+1}} \left[z^{-1} \oint_\gamma \mathrm{d}u\, \mathrm{e}^{-u/z} B_S(u) \right] \tag{4.75}$$

where we have introduced the variable z as $t = u/z$. We have assumed that the behavior at infinity is such that the there is no contribution to take into account. This has to be proved in a case-by-case basis. We now write the contour γ as the difference of two contours,

$$\gamma = C_{0,+} - C_{0,-}, \tag{4.76}$$

where $C_{0,\pm}$ are lines starting from the origin and going above (respectively, below) the real axis. Therefore,

$$\begin{aligned} z^{-1} \oint_\gamma \mathrm{d}u\, \mathrm{e}^{-u/z} B_S(u) &= z^{-1} \left[\oint_{C_{0,+}} - \oint_{C_{0,-}} \right] \mathrm{d}u\, \mathrm{e}^{-u/z} B_S(u) \\ &= f_+(z) - f_-(z), \end{aligned} \tag{4.77}$$

i.e. the difference between the functions f above and below the cut. Finally we can write

$$a_k = \frac{1}{2\pi \mathrm{i}} \int_0^\infty \frac{\mathrm{d}z}{z^{k+1}} \big(f_+(z) - f_-(z) \big), \tag{4.78}$$

i.e.

$$a_k = \int_0^\infty \frac{\mathrm{d}z}{z^{k+1}} \epsilon(z). \tag{4.79}$$

This generalizes (4.27). As a simple test of this formula, we can see that a discontinuity of the form (4.65) gives the expected behavior for a_k:

$$a_k = \int_0^\infty \frac{\mathrm{d}z}{z^{k+1}} z^{-b} \mathrm{e}^{-A/z} \sim A^{-k-b} \int_0^\infty \mathrm{d}x\, x^{k+b-1} \mathrm{e}^{-x} \sim A^{-k-b} \Gamma(k+b). \tag{4.80}$$

4.6 Instantons and large order behavior in quantum theory

We now consider the typical perturbation series which appear in quantum mechanics and quantum field theory. As we have seen in the example of the quartic oscillator, these series diverge factorially, so their Borel transforms are analytic in a neighborhood of the origin. What are the possible sources of the singularities in the Borel transform? In the case of the quartic oscillator, the discontinuity in the energy is given by an instanton calculation, and the singularity in the Borel plane A is nothing but the action of the instanton. This is expected to be a general feature of quantum theories: *if a QFT admits an instanton configuration ϕ_* with finite action $S(\phi_*)$, the Borel transform of any correlation function will be singular at $S(\phi_*)$ and the corresponding perturbative series has a zero radius of convergence.*

There is a heuristic argument for this due to 't Hooft [79]. We write a correlation function

$$W(\alpha) = \int \mathrm{d}\phi\, \mathrm{e}^{-\frac{1}{\alpha} S(\phi)} \phi(x_1) \cdots \phi(x_n) \tag{4.81}$$

as

$$\begin{aligned} W(\alpha) &= \alpha \int_0^\infty \mathrm{d}t \int D\phi\, \delta(\alpha t - S(\phi)) \mathrm{e}^{-\frac{1}{\alpha} S(\phi)} \phi(x_1) \cdots \phi(x_n) \\ &= \alpha \int_0^\infty \mathrm{d}t\, F(\alpha t) \mathrm{e}^{-t}, \end{aligned} \tag{4.82}$$

where we used that

$$\alpha \int_0^\infty \mathrm{d}t\, \delta(\alpha t - S(\phi)) = 1, \tag{4.83}$$

and we wrote

$$F(z) = \int D\phi\, \delta(z - S(\phi))\phi(x_1)\cdots\phi(x_N). \tag{4.84}$$

By comparing (4.82) to (4.51) we see that $F(z)$ is essentially the Borel transform of $W(\alpha)$. If the theory admits a finite action instanton with $z_* = S(\phi_*)$, then the function $F(z)$ will be singular at z_*.

In general, the information provided by instantons about large order behavior seems to encode the growth of the terms in the perturbative series due to the growth in the number of diagrams contributing to each order (as in the case of the quartic oscillator). There are other, very different sources of factorial divergence in perturbation theory, encoded in the so-called renormalons, which we will study later.

We can now discuss various possible behaviors that can arise in quantum theory (and in particular in Quantum Mechanics) concerning the behavior of perturbation theory.

4.6.1 Stable vacua

If we expand around a stable vacuum (like the absolute minimum of a potential in QM), there are no *positive action* instanton solutions. The perturbative series is in principle Borel summable, and the poles of the Borel transform are not on the positive real axis. This is what happens in the case of the quartic, anharmonic oscillator with positive coupling $g > 0$.

4.6.2 Unstable vacua

If we consider the perturbation series around an *unstable* minimum there is always an instanton with real, positive action mediating the decay of the particle. This is what happens for example in the quartic oscillator with negative coupling constant, or in the cubic oscillator. The perturbative series is *not* expected to be Borel summable, since the Borel transform will have singularities in the positive real axis. This is a case in which lateral Borel transforms have an interesting physical meaning. If we consider the lateral Borel resummations of the ground state energy, we will pick a small imaginary part

$$E_{0,\pm}(g) = \mathrm{Re}\, E_0(g) \pm \mathrm{i}\, \mathrm{Im}\, E_0(g), \qquad \mathrm{Im}\, E_0(g) \sim \mathrm{e}^{-A/g}. \tag{4.85}$$

The imaginary part of the ground state energy is nothing but half the width of the level,

$$\mathrm{Im}\, E_0(g) = -\frac{\Gamma}{2} \tag{4.86}$$

and represents the probability of decay of the particle in a metastable vacuum. Therefore, the fact that the Borel transform of the perturbative series is a complex quantity is precisely what is needed in order to capture the physics of the problem.

Example 4.7. *The cubic oscillator.* Perturbation theory gives a series for the ground state energy $E_0(g)$ of the form

$$E_0(g) = \sum_{n=0}^{\infty} a_n g^{2n}. \tag{4.87}$$

The one-instanton contribution to the imaginary part of the ground state energy was computed in (2.175). Since $b = 1/2$, the leading asymptotics is

$$a_k \sim \frac{1}{2\pi}\Gamma(k + 1/2) A^{-k-1/2} c_0, \tag{4.88}$$

where the various quantities A, c_0 can be read from (2.175). The large order behavior is

$$a_n \sim -\frac{(60)^{n+1/2}}{(2\pi)^{\frac{3}{2}} 2^{3n}} \Gamma(n+1/2) \tag{4.89}$$

as computed in for example [4]. Notice that the series is nonalternating, therefore $E_0(g)$ is not Borel summable. As we explained above, this is just reflecting the fact that the energy levels are *unstable*. The "true" energies are complex, and their imaginary parts give the width of the energy levels. A calculation of the widths doing lateral Borel resummation can be found in [4].

4.6.3 Complex instantons

In general, we will have *complex* instanton solutions with complex actions. This leads to perturbative series which are Borel summable and with an oscillatory character. The large order behavior is given by the instanton with the smallest action in absolute value. The phase of the action determines the oscillation period of the series.

Example 4.8. Consider a particle situated at the origin of the potential

$$V(x) = \frac{1}{2}x^2 - \gamma x^3 + \frac{1}{2}x^4. \tag{4.90}$$

There are two different situations here:

1. For $|\gamma| > 1$, the origin is not an absolute minimum, which is in fact at

$$x_0 = \frac{3\gamma + \sqrt{-8 + 9\gamma^2}}{4}. \tag{4.91}$$

2. For $|\gamma| < 1$, the origin is the absolute minimum.

In the first case, the vacuum is quantum–mechanically unstable, and there is an instanton given by a trajectory from $x = 0$ to the turning point

$$x_+ = \gamma - \sqrt{\gamma^2 - 1}. \tag{4.92}$$

The action of this instanton can be written as

$$A = \int_0^{x_+} \mathrm{d}x \, (2V(x))^{\frac{1}{2}} = -\frac{2}{3} + \gamma^2 - \frac{1}{2}\gamma(\gamma^2 - 1) \log \frac{\gamma + 1}{\gamma - 1}, \tag{4.93}$$

while the prefactor reads,

$$C = -\frac{1}{\pi^{\frac{3}{2}}}(\gamma^2 - 1)^{-1/2} A^{-1/2}. \tag{4.94}$$

In the second case, we have to *analytically continue* the results of the first case and in particular consider the instanton above, which is now complex. In fact, there are *two complex conjugate instantons* described by a particle which goes from $x = 0$ to

$$x = g \pm \mathrm{i}\sqrt{1 - g^2}. \tag{4.95}$$

We have then to *add* the contributions of both instantons,

$$\begin{aligned} E_k &= -\frac{1}{\pi^{\frac{3}{2}}}\Gamma(k+1/2)\left[A^{-k-1/2}\mathrm{i}(1-\gamma^2)^{-1/2} - \overline{A}^{-k-1/2}\mathrm{i}(1-\gamma^2)^{-1/2}\right] \\ &= \frac{2}{\pi^{\frac{3}{2}}}\Gamma(k+1/2)(1-\gamma^2)^{-1/2}\mathrm{Im}\, A^{-k-1/2}. \end{aligned} \tag{4.96}$$

For $\gamma = 0$ (the quartic potential) we find,

$$\mathrm{Im}\, A^{-k-1/2} = (-1)^{1+k}\left(\frac{3}{2}\right)^{k+1/2}, \tag{4.97}$$

and the final result for the asymptotics is

$$E_k = \frac{(-1)^{k+1}\sqrt{6}}{\pi^{\frac{3}{2}}}\left(\frac{3}{2}\right)^k, \tag{4.98}$$

which agrees with the previous result after taking into account the relative normalization of g, which adds a factor 2^k.

4.6.4 Cancellation of nonperturbative ambiguities

In some cases, the perturbative series is not Borel summable and on the other hand we know that the true result must be real. This is for example what happens in the double-well potential. The energy of the ground state

$$E^{(0)}(g) = \sum_{k\geq 0} a_k g^k \tag{4.99}$$

is given by the series (2.201). The coefficients in this series grow factorially, and the series is not Borel summable (all coefficients except the leading term have negative sign). In this case, lateral Borel summations cannot lead to the true answer, since they involve an imaginary part:

$$E^{(0)}_{\pm}(g) = \mathrm{Re}E^{(0)}(g) \pm \mathrm{i}\,\mathrm{Im}E^{(0)}(g). \tag{4.100}$$

This imaginary part is exponentially small, and it is or order

$$\mathrm{Im}E^{(0)}(g) \sim \mathrm{e}^{-1/(3g)}. \tag{4.101}$$

It turns out that it is still possible to extract the exact ground state energy from Borel resummations of perturbative series, but we have to consider the *full series* of instanton corrections:

$$E(g) = E^{(0)}(g) + E^{(1)}(g) + E^{(2)}(g) + \cdots \tag{4.102}$$

where $E^{(k)}(g)$ denotes the perturbative expansion around the k-instanton, and it is itself given by a perturbative series which is also non-Borel summable. We can then consider lateral Borel resummations of all the series involved in the multi-instanton expansion,

$$E_{\pm}(g) = E^{(0)}_{\pm}(g) + E^{(1)}_{\pm}(g) + E^{(2)}_{\pm}(g) + \cdots \tag{4.103}$$

Alternatively, we can define the lateral Borel summations by a procedure of analytic continuation. We first consider the series $E^{(k)}(g)$ for *negative* g. In this case, the series is

Borel summable, and this is in turn related to the fact that the instanton corrections are only well-defined for negative g as well, as we saw in section 2. The resummed perturbative series $E_\pm^{(k)}(g)$ is then given analytic continuation of this Borel sum from g negative to $g = |g| \pm i0$, see Fig. 24. Notice that, with this definition, the two-instanton configuration computed in (2.269) picks an imaginary part since

$$\log\left(-\frac{2}{g}\right) \to \log\left(\frac{2}{|g|}\right) \pm \pi \mathrm{i}. \tag{4.104}$$

therefore

$$\mathrm{Im}\, E_\pm^{(2)} \sim \pm\left(\frac{2}{g}\right)\frac{\mathrm{e}^{-1/3g}}{2\pi}\mathrm{Im}\,\log\left(-\frac{2}{g}\right) = \pm\frac{1}{g}\mathrm{e}^{-1/3g}. \tag{4.105}$$

We now impose the physical requirement that $E_\pm(g)$must be *independent* of the resum-

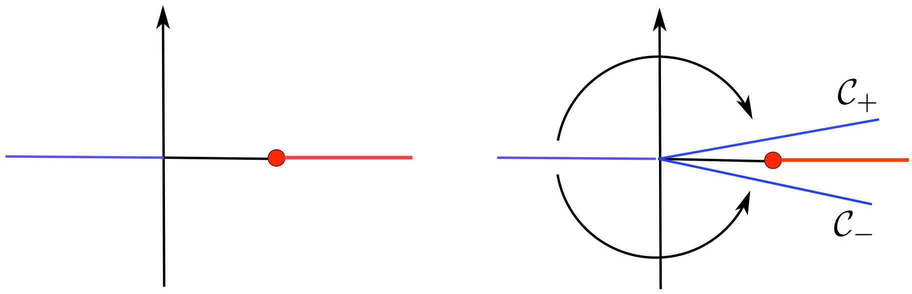


Figure 24: Lateral resummations can be obtained as *two* different analytic continuations of a Borel sum along the negative real axis (where there are no singularities) to the positive real axis.

mation prescription and *real* (since it is the energy of a bound state!). These means that imaginary parts must *cancel* in the total sum (4.103). At leading order in the instanton expansion, this implies in particular that the imaginary part of the perturbative sums $E_\pm^{(0)}$ is equal but opposite in sign to the imaginary part of the two–instanton contribution $\mathrm{Im}\, E_\pm^{(2)}$:

$$\mathrm{Im}\, E_\pm^{(0)} = -\mathrm{Im}\, E_\pm^{(2)} \Rightarrow \mathrm{Im}\, E^{(0)} \sim -\frac{1}{g}\mathrm{e}^{-1/3g}. \tag{4.106}$$

The one-instanton contribution has an imaginary part, but it is proportional to $\mathrm{e}^{-1/2g}$ and cancels against the third-instanton contribution, so we don't have to consider it at this order. The cancellation (4.106) determines the large order behavior of the perturbative series by using (4.79) and taking into account that

$$\epsilon(z) \sim -\frac{1}{z}\mathrm{e}^{-1/3z}. \tag{4.107}$$

One then finds,

$$a_k \sim -\frac{1}{\pi}3^{k+1}\Gamma(k+1) = -\frac{1}{\pi}3^{k+1}k! \tag{4.108}$$

which can be tested against the explicit results for the perturbative series [96] providing in this way a confirmation of the cancellation mechanism (4.106).

The cancellation between perturbative and nonperturbative contributions appearing in the double-well has been argued to be relevant in more general situations in quantum theory. These situations involve non-Borel summable series which however should lead to well-defined, non-perturbative real quantities, and include realistic examples in quantum field theory and in particular in QCD; see [33, 51] for examples involving renormalons and [63] for examples in unitary matrix models and string theory.

5. Nonperturbative aspects of gauge theories

5.1 Conventions and basics

We follow the gauge theory conventions in [26]. The generators of the Lie algebra T^a are taken to be *anti–Hermitian*, and satisfy the commutation relations

$$[T^a, T^b] = f^{abc} T^c. \tag{5.1}$$

For $SU(2)$, for example, we take

$$T^a = -\frac{\mathrm{i}}{2}\sigma^a, \tag{5.2}$$

and the structure constants are

$$f^{abc} = \epsilon^{abc}. \tag{5.3}$$

The Cartan inner product is defined by

$$(T^a, T^b) = \delta^{ab}. \tag{5.4}$$

and it can be shown that

$$(T^a, T^b) = -2\mathrm{Tr}(T^a T^b). \tag{5.5}$$

The Euclidean action for pure Yang–Mills is

$$S_E = \frac{1}{4g^2} \int \mathrm{d}^4 x \, (F_{\mu\nu}, F^{\mu\nu}). \tag{5.6}$$

The Lagrangian of QCD will be written as

$$\mathcal{L} = \frac{1}{g^2} \left[\frac{1}{4}(F_{\mu\nu}, F^{\mu\nu}) + \sum_{f=1}^{N_f} \bar{q}_f \, (i\not{D} - m_f) \, q_f \right] \tag{5.7}$$

where the covariant derivative is defined by

$$D_\mu = \partial_\mu + \mathrm{i} A_\mu \tag{5.8}$$

Very often it is more convenient to use rescaled fields, in such a way that the coupling constant appears only in the vertices of the theory. These fields are defined by

$$A_\mu = g\hat{A}_\mu, \qquad q = g\hat{q}. \tag{5.9}$$

At the quantum level, theories of the Yang-Mills type are renormalizable (g is dimensionless), and they exhibit a *running coupling constant* and *asymptotic freedom.* Let us denote by

$$\alpha_s(\mu) = \frac{g^2(\mu)}{4\pi} \tag{5.10}$$

the renormalized coupling constant in the $\overline{\mathrm{MS}}$ scheme, at the subtraction point μ. The β-function is written as

$$\beta(\alpha_s) = \mu^2 \frac{\partial \alpha_s}{\partial \mu^2} = \beta_0 \alpha_s^2 + \beta_1 \alpha_s^3 + \dots . \tag{5.11}$$

The β-function is scheme-dependent, but the first two coefficients are scheme-independent in the class of massless subtraction schemes. The one-loop coefficient is given by

$$\beta_0 = \beta_{0g} + \beta_{0f} = -\frac{1}{4\pi} \left(\frac{11 N_c}{3} - \frac{2 N_f}{3} \right), \tag{5.12}$$

where N_c is the number of colors and N_f the number of massless quark flavours. β_{0g} and β_{0f} denote respectively the gluon and fermion contribution to the one-loop β-function. If the number of flavours is small enough as compared to the number of colors, the first coefficient of the beta function is negative and the theory is asymptotically free. It follows from the running of the coupling constant that the quantity

$$\Lambda^2 = \mu^2 \mathrm{e}^{1/(\beta_0 \alpha_s(\mu))} \tag{5.13}$$

is in fact independent of μ, at leading order, and therefore defines a RG-invariant scale. This is the so-called *dynamically generated scale* of QCD. The fact that a theory with a dimensionless coupling constant g generates a dimensionful scale is called *dimensional transmutation*.

5.2 Topological charge and θ vacua

Good references for this subsection are [92, 82].

In Yang–Mills theory, besides the standard YM action, there is another term that can be added to the action. This term is called the *topological charge* for reasons that will become clear later on, and it is given by

$$Q = \int \mathrm{d}^4 x \, q(x), \tag{5.14}$$

where

$$q(x) = \frac{1}{32\pi^2} (F, \tilde{F}) = \frac{1}{64\pi^2} \epsilon_{\mu\nu\rho\sigma} (F^{\mu\nu}, F^{\rho\sigma}). \tag{5.15}$$

This term is allowed by gauge invariance and renormalizability, so it is natural to add it to the action and to take as the Euclidean YM Lagrangian

$$\mathcal{L}_\theta = \frac{1}{4g^2} (F^{\mu\nu}, F_{\mu\nu}) - \mathrm{i}\theta q(x), \tag{5.16}$$

where θ is a new parameter in the QCD Lagrangian. We will see below that (5.14) is *quantized* for any classical, continuous field configuration with a *finite* action.

The different observables of QCD should be sensitive to the θ parameter. One such quantity is the ground state energy density $E(\theta)$, computed at large, finite volume V as

$$\exp(-V E(\theta)) = \int [\mathcal{D}A] \mathrm{e}^{-\int \mathrm{d}^4 x \, \mathcal{L}_\theta}. \tag{5.17}$$

The function $E(\theta)$ has two properties. First of all, the path integral with the insertion of $e^{i\theta Q}$, $\theta \neq 0$ should be smaller than the path integral without the insertion, at $\theta = 0$. This is because when $\theta \neq 0$ we are integrating an oscillating function with a positive measure. We conclude that

$$E(0) \leq E(\theta), \qquad \theta \neq 0, \tag{5.18}$$

and the ground state energy has an absolute minimum at $\theta = 0$ [80]. Second, as we will show in a moment, smooth field configurations with a finite action have quantized values of Q. Thus we expect $E(\theta)$ to be *periodic*, with period 2π:

$$E(\theta + 2\pi) = E(\theta). \tag{5.19}$$

Notice that, in the limit of infinite volume, smooth configurations of finite action give just a zero-measure set in the path integral, and we could think that the value of $E(\theta)$ is dominated by field configurations in which Q is not an integer. However, using a fully non-perturbative definition in the lattice, as the one proposed in [59], Q takes integer values for any discretized lattice configuration, and we have periodicity in θ in the continuum, large volume limit.

The function $E(\theta)$ can be expanded around $\theta = 0$ as

$$E(\theta) - E(0) = \frac{1}{2}\chi_t^V \theta^2 s(\theta), \quad s(\theta) = 1 + \sum_{n=1}^{\infty} b_{2n}\theta^{2n}. \tag{5.20}$$

Since $q(x)$ is odd under parity reversal, only even powers of $q(x)$ have nonzero vacuum expectation values (since the vacuum is invariant under parity), and only even powers of θ appear in the expansion of $E(\theta)$. The coefficient χ_t^V is an important quantity and measures the leading dependence of $E(\theta)$ on the θ angle around $\theta = 0$. It is called the *topological susceptibility* and it can be written as

$$\chi_t^V = \left(\frac{d^2E}{d\theta^2}\right)_{\theta=0} = \frac{\langle Q^2\rangle}{V} = \int_V d^4x \langle q(x)q(0)\rangle. \tag{5.21}$$

The last equality follows from

$$\langle Q^2\rangle = \int_V d^4x \int_V d^4y \, \langle 0|q(x)q(y)|0\rangle = \int_V d^4x \int_V d^4y \, \langle 0|q(x-y)q(0)|0\rangle = V\chi_t^V, \tag{5.22}$$

where translation invariance of the vacuum has been used. Of course, since $\theta = 0$ is a minimum of $E(\theta)$, we have $\chi_t^V \geq 0$. The infinite-volume limit of the quantity χ_t^V will be simply denoted by

$$\chi_t = \lim_{V\to\infty} \chi_t^V \tag{5.23}$$

Although we have said that observables in YM theory should be sensitive to the θ parameter, this dependence is very subtle. The reason is that (5.15) is a total divergence,

$$q(x) = \partial_\mu K^\mu, \tag{5.24}$$

where

$$K_\mu = \frac{1}{16\pi^2}\epsilon_{\mu\nu\rho\sigma}(A_\nu, \partial_\rho A_\sigma + \frac{2}{3}A_\rho A_\sigma). \tag{5.25}$$

The three–form appearing here is the so-called *Chern–Simons term*. This means, in particular, that

$$\tilde{q}(p) = \int \mathrm{d}^4 x\, \mathrm{e}^{-\mathrm{i}px} q(x) \tag{5.26}$$

vanishes at zero momentum, since it is of the form $p^\mu \tilde{K}_\mu(p)$. But the topological susceptibility is given by

$$\chi_t = \lim_{k\to 0} U(k), \tag{5.27}$$

where

$$U(k) = \int \mathrm{d}^4 x\, \mathrm{e}^{\mathrm{i}kx} \langle q(x) q(0)\rangle. \tag{5.28}$$

We can write

$$U(k) = \int \frac{\mathrm{d}^4 p'}{(2\pi)^4} \langle \tilde{q}(-k)\tilde{q}(p')\rangle \tag{5.29}$$

Since $\tilde{q}(0) = 0$, this quantity vanishes order by order in perturbation theory. However, as noticed by Witten in [87], this does not mean that it vanishes *tout court*. It might happen that after adding an infinite number of diagrams (or a subset of them), and then taking the limit $k \to 0$, one obtains a nonzero result. Indeed, this is the kind of situation we briefly illustrated in the example (1.4). We will see below that, in the $1/N$ expansion, after adding up an infinite number of diagrams (the so-called *planar* diagrams), one finds a nonzero value for the topological susceptibility.

Using Stokes theorem, we can now write the topological charge as

$$Q = \int \mathrm{d}\Sigma_\mu K^\mu. \tag{5.30}$$

Let us take as surface of integration two spatial planes at $t = \pm\infty$, so that

$$Q = \int \mathrm{d}^3\vec{x}\, K^0(t\to\infty, \vec{x}) - \int \mathrm{d}^3\vec{x}\, K^0(t\to -\infty,\vec{x}) \equiv K_+ - K_-. \tag{5.31}$$

These operators are Hermitian, and related to each other by time reversal, so their spectra coincide. Let $|n_\pm\rangle$ denote their eigenstates,

$$K_\pm |n_\pm\rangle = n|n_\pm\rangle. \tag{5.32}$$

We can now expand the physical vacuum as

$$|\theta\rangle = \sum_n c_n(\theta)|n_+\rangle = \sum_n c_n(\theta)|n_-\rangle. \tag{5.33}$$

This follows from time reversal invariance of the vacuum: if we apply the time reversal operator, the vacuum is unchanged and the first sum becomes the second one. Notice that $|\theta\rangle$ is just the vacuum for the Yang–Mills field theory which includes a theta term. We also have the following identity,

$$\begin{aligned} \mathrm{i}\frac{\partial}{\partial\theta}\langle\theta|\mathcal{O}|\theta\rangle &= \mathrm{i}\frac{\partial}{\partial\theta}\langle 0|\mathcal{O}\mathrm{e}^{-\int \mathrm{d}^4x \mathcal{L}_\theta}|0\rangle \\ &= \int \mathrm{d}^4 x \langle 0|q(x)\mathcal{O}\mathrm{e}^{-\int \mathrm{d}^4x\mathcal{L}_\theta}|0\rangle \\ &= \int \mathrm{d}^4x \langle\theta|q(x)\mathcal{O}|\theta\rangle, \end{aligned} \tag{5.34}$$

so the operator $\mathrm{i}\partial_\theta$ is equivalent to the insertion of Q. But because of (5.31) we find

$$\mathrm{i}\frac{\partial}{\partial\theta}\langle\theta|\mathcal{O}|\theta\rangle = \langle\theta|K_+\mathcal{O}|\theta\rangle - \langle\theta|\mathcal{O}K_-|\theta\rangle. \tag{5.35}$$

Here we have used a time-ordering prescription which says that K_+ should be inserted to the left and K_- to the right. If we now plug in the expansion (5.33), we find,

$$\mathrm{i}\frac{\partial}{\partial\theta}\sum_{n,k} c_n^*(\theta)c_k(\theta) = \sum_{n,k}(n-k)c_n^*(\theta)c_k(\theta), \tag{5.36}$$

which leads to

$$c_n = C\mathrm{e}^{\mathrm{i}n\theta}, \tag{5.37}$$

where C is an overall constant. In terms of the eigenstates of $K_\pm$, we find that

$$|\theta\rangle = \sum_n \mathrm{e}^{\mathrm{i}n\theta}|n\rangle, \tag{5.38}$$

and we set the overall constant C to 1 for simplicity.

So far we don't have more information about the structure of the spectrum of $K_\pm$. It might happen that all of the n are identical, so that the structure above collapses to something trivial. But as we will see, the existence of YM instantons implies that all $n \in \mathbb{Z}$ exist.

5.3 Instantons in Yang–Mills theory

We will now look for *instantons* in Yang–Mills theory. These are, by definition, field configurations which solve the equations of motion and have *finite action.* These configurations are important in a semi-classical analysis, since they might lead to a starting point for a perturbation expansion.

The condition of finite action gives constraints on the large distance behavior of the fields. In order to see how they must behave as $r \to \infty$, we notice that schematically the Euclidean action can be written as

$$S_E \sim \int \mathrm{d}r\, r^3 F^2 \tag{5.39}$$

If we want this to be finite, the integrand has to go at least like $1/r^2$. For example, we could have

$$F \sim \frac{1}{r^3} \tag{5.40}$$

as $r \to \infty$. This leads to the following behavior for $A(r)$,

$$A(r) \sim \frac{1}{r^2}, \qquad r \to \infty. \tag{5.41}$$

However, A is only well defined up to a gauge transformation, so we can have the more general behavior

$$A_\mu \to g\partial_\mu g^{-1} + \mathcal{O}(r^{-2}), \qquad r \to \infty. \tag{5.42}$$

This means that the gauge potential is pure gauge at infinity. Since the limiting behavior has to be well-defined as $r \to \infty$, we can define the function g on the boundary at infinity

$\mathbb{S}^3 \subset \mathbb{R}^4$. This is for example achieved if g depends only on the angular variables of $\mathbb{R}^4$. Therefore, any solution like the above defines a map from $\mathbb{S}^3$ to the gauge group, i.e.

$$g : \mathbb{S}^3 \to G. \tag{5.43}$$

Under gauge transformation, g will change. Therefore, what is a gauge-invariant concept is the *homotopy* type of mappings from $\mathbb{S}^3$ to G. As in the theory of solitons, these homotopy types are classified by

$$\pi_3(G). \tag{5.44}$$

Example 5.1. A toy example are instantons in Euclidean two–dimensional space with $U(1)$ gauge group. Here, the homotopy group is $\pi_1(\mathbb{S}^1) = \mathbb{Z}$. Homotopy classes are classified by an integer n. A map in the class characterized by n is the covering

$$g^{(n)}(\theta) = \mathrm{e}^{\mathrm{i}n\theta}. \tag{5.45}$$

Example 5.2. Let us consider the case $G = SU(2)$. Any element of $SU(2)$ can be written as

$$g = a + \mathrm{i}\mathbf{b} \cdot \boldsymbol{\sigma}, \quad a^2 + \mathbf{b}^2 = 1, \tag{5.46}$$

hence $SU(2)$ is homeomorphic to $\mathbb{S}^3$. We have then to consider maps of the form

$$g : \mathbb{S}^3 \to \mathbb{S}^3. \tag{5.47}$$

The relevant homotopy group is

$$\pi_3(\mathbb{S}^3) = \mathbb{Z}. \tag{5.48}$$

This can be computed by using Hurewicz isomorphism theorem (see for example [15]), which holds in this case due to the fact that

$$\pi_1(\mathbb{S}^3) = 0. \tag{5.49}$$

This theorem relates homotopy groups to homology groups, which are typically much easier to calculate, and in this case it says that

$$\pi_2(\mathbb{S}^3) = H_2(\mathbb{S}^3), \qquad \pi_3(\mathbb{S}^3) = H_3(\mathbb{S}^3) = \mathbb{Z}. \tag{5.50}$$

It follows from (5.48) that the homotopy classes relevant to the gauge group $SU(2)$ are labelled by an integer n, which is called the *winding number*. An explicit expression for a map

$$g : \mathbb{S}^3 \to SU(2) \tag{5.51}$$

with winding number n is given by

$$g^{(n)}(x) = \left(\frac{x_4 + \mathrm{i}\mathbf{x} \cdot \boldsymbol{\sigma}}{r} \right)^n. \tag{5.52}$$

For $n = 0$, this is the trivial map, while for $n = 1$ it is the identity. Notice that this map can be expressed solely in terms of angular variables.

We than have learned that, at least when the gauge group is $SU(2)$, *every field configuration of finite action is characterized by its winding number n.* It can be shown that the winding number of a gauge field is the value of the topological charge (5.14). We sketch here some steps of the argument, referring to [70] for further details. We start from the expression (5.30), integrated over the boundary at infinity, which is a three-sphere $\mathbb{S}^3$. For a gauge field satisfying (5.42), the field strength $F_{\mu\nu}$ vanishes at infinity. Therefore, on $\mathbb{S}^3$,

$$\epsilon_{\mu\nu\alpha\beta}\partial^\alpha A^\beta = -\epsilon_{\mu\nu\alpha\beta}A^\alpha A^\beta, \tag{5.53}$$

and one finds,

$$Q = -\frac{1}{48\pi^2}\int \mathrm{d}\Sigma^\mu \epsilon_{\mu\nu\alpha\beta}(A^\nu, A^\alpha A^\beta). \tag{5.54}$$

By using the boundary behavior of the gauge potentials, one can also write this quantity as

$$Q = \frac{1}{48\pi^2}\int \mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}\theta_3\, \epsilon^{ijk}(g^{-1}\partial_i g, g^{-1}\partial_j g g^{-1}\partial_k g). \tag{5.55}$$

This quantity is a homotopy invariant and gives the winding number associated to the homotopy class of g.

Example 5.3. Let us use the integral expression (5.54) to verify that $g^{(1)}$, as given in (5.52), indeed has $n = 1$. The inverse map is

$$g^{-1} = \frac{x_4 - \mathrm{i}\vec{x}\cdot\vec{\sigma}}{r}. \tag{5.56}$$

One finds,

$$Q = -\frac{1}{24\pi^2}\int \mathrm{d}\Sigma_\mu\Big(-\frac{12x^\mu}{|x|^4}\Big). \tag{5.57}$$

Using now

$$\mathrm{d}\Sigma^\mu = x^\mu|x|^2\mathrm{d}\Omega_3, \tag{5.58}$$

we obtain

$$Q = \frac{1}{2\pi^2}\int \mathrm{d}\Omega_3 = 1. \tag{5.59}$$

So far, we have seen that, if there are field configurations of finite action, they will be classified by an integer winding number. We now have to construct explicitly configurations with finite action which solve the equations of motion, and therefore lead to different vacua of the Yang–Mills theory. We will see that, in each of the topological sectors, there is a configuration which minimizes the action, and therefore solves the equation of motion. In other words, we will see that there is *an infinite set of classical vacua enumerated by an integer n.*

Start with the identity

$$\int \mathrm{d}^4x(F \pm \tilde{F})^2 \geq 0. \tag{5.60}$$

From here we find

$$\frac{1}{4g^2}\int \mathrm{d}^4x\,(F,F) \geq \mp\frac{1}{4g^2}\int \mathrm{d}^4x\,(F,\tilde{F}), \tag{5.61}$$

or equivalently

$$S \pm \frac{8\pi^2 n}{g^2} \geq 0. \tag{5.62}$$

We conclude that

$$S \geq \frac{8\pi^2|n|}{g^2}. \tag{5.63}$$

To saturate the inequality, notice that S is always positive. Therefore if $n > 0$ is positive we have

$$F = \tilde{F}, \qquad S = \frac{8\pi^2 n}{g^2}, \tag{5.64}$$

i.e. the gauge field is *self-dual* (SD) and we have a gauge theory instanton. If $n < 0$ is negative we have

$$F = -\tilde{F}, \qquad S = -\frac{8\pi^2 n}{g^2} \tag{5.65}$$

i.e. the gauge field is *anti-self-dual* (ASD) and we have a gauge theory anti-instanton. If any of these conditions holds, the corresponding gauge field minimizes the action for a fixed topological class given by n, and in particular solves the EOM. Notice that, in contrast to the standard EOM of Yang–Mills theory, these are *first order* equations. This can be related to BPS conditions in supersymmetry, as we will see.

It is possible to write down explicitly the asymptotic expression of the gauge field for the instanton configuration with gauge group $SU(2)$ and $n = 1$ (the one-instanton solution). To do this, we simply set

$$A_\mu = -(\partial_\mu g) g^{-1}, \tag{5.66}$$

where $g = g^{(1)}$, and $g^{(n)}$ is the the map (5.52). Since

$$\begin{aligned} \partial_4 g &= -\frac{x_\mu}{r^2} g + \frac{1}{r}, \\ \partial_i g &= -\frac{x_\mu}{r^2} g + \frac{\mathrm{i}\sigma_i}{r}, \quad i = 1,2,3. \end{aligned} \tag{5.67}$$

One then finds

$$\begin{aligned} A_4 &= \mathrm{i}\frac{\vec{x}\cdot\vec{\sigma}}{r^2}, \\ A_i &= -\frac{\mathrm{i}}{r^2}\Big(x_4\sigma_i + \epsilon_{ijk} x_j \sigma_k\Big), \end{aligned} \tag{5.68}$$

where we used that

$$\sigma_i \vec{x}\cdot\vec{\sigma} = x_i + \mathrm{i}\epsilon_{ijk} x_j \sigma_k. \tag{5.69}$$

If we write

$$A_\mu = -\frac{\mathrm{i}}{2}\sigma_a A^a_\mu, \tag{5.70}$$

and we introduce the 't Hooft matrices $\eta^a_{\mu\nu}$ by

$$\eta^a_{ij} = \epsilon_{aij}, \quad \eta^a_{i4} = \delta_{ai}, \quad \eta^a_{4i} = -\delta_{ai}, \tag{5.71}$$

where $i, j = 1, 2, 3$, we find that

$$A^a_\mu = 2\eta^a_{\mu\nu}\frac{x^\nu}{r^2}. \tag{5.72}$$

This asymptotic form suggests the following ansatz for the *exact* form

$$A^a_\mu = 2\eta^a_{\mu\nu}\frac{x^\nu}{r^2} f(r^2), \tag{5.73}$$

where

$$f(r^2) \to 1, \qquad r \to \infty. \tag{5.74}$$

Also, regularity at the origin requires that

$$f(r^2) \sim r^2, \qquad r \to 0. \tag{5.75}$$

We can now plug the ansatz (5.73) in the gauge theory action and compute

$$S \propto \int_0^\infty \mathrm{d}r \left[\frac{r}{2} \left(f'\right)^2 + \frac{2}{r} f^2 (1-f)^2 \right] \tag{5.76}$$

The second order EOM for f gives

$$-\frac{\mathrm{d}}{\mathrm{d}r}\left(r \frac{\mathrm{d}f}{\mathrm{d}r} \right) + \frac{4}{r} f(1-f)(1-2f) = 0. \tag{5.77}$$

There are three constant solutions: $f = 0$ is the trivial gauge connection, $f = 1$ is a pure gauge transformation with winding number 1, and $f = 1/2$ is a "half gauge transformation" also called *meron*. On top of that we have a space-dependent solution

$$f(r) = \frac{r^2}{r^2 + \rho^2}. \tag{5.78}$$

This gives the one-instanton solution of $SU(2)$ Yang–Mills theory. Notice that the resulting configuration interpolates between the trivial vacuum $f = 0$ at the origin and the homotopically non-trivial gauge transformation with $n = 1$ as $r \to \infty$, and at large r it is indeed of the form (5.42).

In (5.78) ρ is an integration constant which can be regarded as the *size* of the instanton. There is an interesting contrast between the instantons of Yang–Mills theory and the instantons or bounces of the scalar theory studied in chapter 3. The size of a "bubble" in the scalar theory was fixed by the parameters of the potential, while the size of an instanton in Yang–Mills theory is a free parameter. It is yet another example of a collective coordinate, and as it is always the case, its existence is due to a symmetry of the theory. In this case, the symmetry is the *scale invariance* of the classical Yang–Mills action. In fact, in writing (5.78) we have already fixed some integration constants: the above solution is centered at the origin, but one can write a more general solution down,

$$A^a_\mu = 2\eta^a_{\mu\nu} \frac{(x - x_0)^\nu}{(x - x_0)^2 + \rho^2}, \tag{5.79}$$

where x_0 is the position of the center of the instanton. This gives four extra collective coordinates due to translation invariance.

We have found (5.78) by solving the original EOM of the Yang–Mills action, which are second order, but one can solve instead the first order equation (5.64). By plugging in the ansatz (5.73) we find the following first order equation for f,

$$f(1-f) - r^2 \frac{\mathrm{d}f}{\mathrm{d}r^2} = 0, \tag{5.80}$$

which leads again to the constant solutions $f = 0, 1$ and to the one-instanton solution (5.78). Notice however that the meron solution $f = 1/2$ does not solve the first order equation, but indeed it leads to an infinite action.

5.4 Instantons and theta vacua

Like in the Quantum Mechanics examples, instantons in Yang–Mills theory can be interpreted as tunneling configurations between different vacua [21], i.e. the instanton field in a sector with winding number n can be interpreted as a field which goes from one give vacuum at infinite past in the Euclidean theory $\tau = -\infty$, to another vacuum at infinite future, $\tau = +\infty$.

To see this, let us choose a gauge in which $A_0 = 0$. The winding number (5.30) is given by an integral over an $\mathbb{S}^3$ at infinity. Let us deform this boundary into a cylinder parallel to the $x^0 = \tau$ axis. In the axial gauge $A_0 = 0$, the curved surface of the cylinder does not make any contribution, and we can write

$$n = n_+ - n_-, \tag{5.81}$$

where

$$n_\pm = -\frac{1}{48\pi^2}\int \mathrm{d}^3x\, \epsilon_{ijk}(A_i, A_jA_k)\Big|_{\tau=\pm\infty}. \tag{5.82}$$

The field configurations at $\tau \to \pm\infty$ correspond to different vacua whose homotopy numbers differ by n, the charge of the instanton. We know that n is an integer, and one can choose the gauge in such a way that $n_- = 0$. Therefore, we find an explicit semiclassical realization of all the vacua $|n_\pm\rangle$ that we introduced in (5.33). They are labeled by integers. In particular, the transition amplitude between two vacua is given by

$$\langle n|\mathrm{e}^{-HT}|m\rangle = \int \mathcal{D}A_{n-m}\, \exp\Big[-\int \mathrm{d}^4x\, \mathcal{L}(A)\Big], \tag{5.83}$$

where the measure $\mathcal{D}A_{n-m}$ means that we integrate over all gauge fields with fixed winding number $n-m$, and it is understood that we are considering the limit $T \to \infty$. We will denote by

$$Z_\nu = \int \mathcal{D}A_\nu\, \exp\Big[-\int \mathrm{d}^4x\, \mathcal{L}(A)\Big] \tag{5.84}$$

the partition function in the sector with winding number ν. We can then write

$$\langle\theta'|\mathrm{e}^{-HT}|\theta\rangle = \sum_{n,m} \mathrm{e}^{\mathrm{i}n\theta - \mathrm{i}m\theta'} Z_{n-m} = \sum_{n,\nu} \mathrm{e}^{\mathrm{i}m(\theta-\theta')+\mathrm{i}\nu\theta} Z_\nu = \delta(\theta-\theta')\sum_\nu \mathrm{e}^{\mathrm{i}\nu\theta} Z_\nu \tag{5.85}$$

where in the second line we changed variables to $\nu = n = m$. We now define the total, θ-dependent partition function

$$Z(\theta) \equiv \int \mathcal{D}A\, \exp\Big[-\int \mathrm{d}^4x\, \mathcal{L}_\theta(A)\Big] \tag{5.86}$$

where we have introduced the Lagrangian with a θ term (5.16) and we integrate now over *all* possible gauge fields (belonging to all possible homotopy classes). We can then write

$$\langle\theta'|\mathrm{e}^{-HT}|\theta\rangle = \delta(\theta-\theta')Z(\theta) \tag{5.87}$$

This confirms that, indeed, the theta vacuum is the vacuum which is obtained by quantizing the theory with the Lagrangian (5.16).

In terms of these partition functions, we have that

$$VE(\theta) = -\log\left\{\sum_\nu e^{i\nu\theta}\int \mathcal{D}A_\nu \exp\Big[-\int d^4x\,\mathcal{L}(A)\Big]\right\} = -\log Z(\theta) \tag{5.88}$$

and also

$$\chi_t^V = \frac{1}{V}\sum_\nu \nu^2 P_\nu, \qquad P_\nu = \frac{Z_\nu}{Z(0)}. \tag{5.89}$$

Notice that P_ν can be interpreted as the probability of finding a gauge field with charge k.

The leading contributions in these sums over instantons come from the one-instanton as well as the one-anti-instanton, since both have

$$S_c = \frac{8\pi^2}{g^2} \tag{5.90}$$

but opposite $\nu = \pm 1$. Therefore,

$$VE(\theta) = -\log Z_0 - \log\Big\{1 + e^{i\theta - \frac{8\pi^2}{g^2}}K_1 + e^{-i\theta - \frac{8\pi^2}{g^2}}K_{-1} + \cdots\Big\} \tag{5.91}$$

where $K_{\pm 1} = KV$ are given, at leading order in g, by the one-loop fluctuation around the instanton/anti-instanton solutions, and we have factored out the volume V which is obtained by integrating over the zero mode x_0 due to translation invariance. We have also taken into account that the one-loop fluctuations are the same around the instanton and the anti-instanton. We then obtain

$$E(\theta) - E(0) \approx 2\,(1-\cos\theta)\,K\,e^{-\frac{8\pi^2}{g^2}}. \tag{5.92}$$

This approximation gives, for the topological susceptibility,

$$\chi_t \sim K e^{-\frac{8\pi^2}{g^2}}. \tag{5.93}$$

In the calculation of the path integral as a sum over instantons, the topological susceptibility is indeed fully nonperturbative, and it is invisible in perturbation theory.

We must now compute K, and to do that we must be careful with the collective coordinates. There are eight in total in the case of the instanton. Four of them correspond to the location of the instanton, and as usual integrating over them gives the total volume of space-time V, which we already factored out in (5.92). Another collective coordinate is the size of the instanton ρ, and finally there are three extra parameters coming from gauge rotations. In total, we have 8 parameters that lead to a factor

$$S_c^4 = \Big(\frac{8\pi^2}{g^2}\Big)^4. \tag{5.94}$$

The integral over gauge transformations leads to a constant factor. The integral over ρ must be of the form

$$\int_0^\infty \frac{d\rho}{\rho^5} f(\rho\mu) \tag{5.95}$$

just based on dimensional reasons: recall that we are computing an energy density, therefore it has units of length^{-4}, while ρ has dimensions of length. $f(\rho\mu)$ is a function which depends on μ, which is needed to renormalize a quantum gauge theory. The form of f can be determined by noticing that the final answer must involve RG invariant quantities. This means, on one hand, that in the above computation we must use the running coupling constant $g^2(\mu)$

$$e^{-\frac{8\pi^2}{g^2(\mu)}} = e^{-\frac{2\pi}{\alpha_s(\mu)}}. \tag{5.96}$$

In view of (5.13), this must combine with

$$\mu^{-4\pi\beta_0} \tag{5.97}$$

in order to produce a RG-invariant integrand, and this fixes the form of $f(\rho\mu)$ at leading order as

$$f(\rho\mu) = (\rho\mu)^{-4\pi\beta_0}. \tag{5.98}$$

We can now write

$$e^{-\frac{2\pi}{\alpha_s(\mu)}}\mu^{-4\pi\beta_0} = e^{-\frac{2\pi}{\alpha_s(1/\rho)}}\rho^{4\pi\beta_0} = \Lambda^{-4\pi\beta_0} \tag{5.99}$$

because of RG invariance, and the integral becomes

$$e^{-\frac{2\pi}{\alpha_s(\mu)}}\int_0^\infty \frac{d\rho}{\rho^5}(\rho\mu)^{-4\pi\beta_0} = \int_0^\infty \frac{d\rho}{\rho^5}e^{-\frac{2\pi}{\alpha_s(1/\rho)}} \tag{5.100}$$

which is the RG-invariant way of writing the integral over instanton sizes. At small ρ we can use asymptotic freedom and the one-loop beta function to write the integral as

$$\int_0^\infty \frac{d\rho}{\rho^5}(\rho\Lambda)^{11N_c/3} \tag{5.101}$$

for pure Yang–Mills theory. This integral is convergent in the UV $\rho \to 0$, for all $N_c \geq 2$, but it diverges in the IR $\rho \to \infty$. This is the famous IR embarrassment for instanton calculus due to instantons of large size.

Of course, what is really going on is that in the regime $\rho \to \infty$ the integral (5.101) is not really the right answer. As the instanton size becomes large, the running coupling constant $\alpha_s(1/\rho)$ in (5.100) enters the strong coupling regime and we are unable to perform reliable instanton calculations. The only way to do instanton calculus in a gauge theory is to have an IR cutoff in the instanton size which avoids the problems of strong coupling. This is the case, for example, if we do the instanton calculation in a space-time with finite, small volume V (for example, a four-sphere $\mathbb{S}^4$). In this case, since the size of the instanton cannot be bigger than the characteristic scale of spacetime $V^{1/4}$, the integral over the instanton size ρ has a natural cutoff. One should then be able to calculate P_1 by using instanton calculus. Since the natural scale in the problem is V we expect by dimensional transmutation

$$P_1 \sim \exp\left\{-\frac{8\pi^2}{\alpha_s(\Lambda V^{1/4})}\right\} \sim (V\Lambda^4)^{\frac{11N_c}{12}} \tag{5.102}$$

This is indeed what is found in the calculation of [58].

Another way of having a natural cutoff is to have a Higgs-like eld with a large VEV which sets the scale (as in supersymmetric gauge theories), or to consider the theory at finite temperature. In both cases one can use instanton calculus profitably (see for example [50] for a instanton calculation of the topological susceptibility at finite temperature). Otherwise, instanton calculations in QCD are doubtful. Indeed, the θ dependence in (5.92), which seems to be a universal feature of instanton-based approaches to the topological susceptibility, is currently disfavoured by lattice calculations [46].

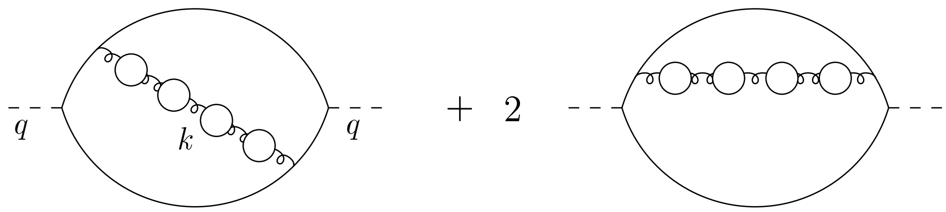


Figure 25: The simplest set of 'bubble' diagrams for the Adler function consists of all diagrams with any number of fermion loops inserted into a single gluon line.

5.5 Renormalons

As we saw in Chapter 4, instantons dominate the large order behavior in quantum mechanics. There are other simple quantum models where this is still the case, in the sense that the large order of perturbation theory is determined by the factorial growth of the number of diagrams. For example, the large order behavior of super-renormalizable quantum field theories is supposed to be dominated by instantons. In *renormalizable* quantum field theories, however, the large order behavior of perturbation theory seems to be dominated by another type of divergences called *renormalon* divergences, or renormalons for sort. Renormalon divergences also lead to factorial behavior $n!$ at order n in perturbation theory. However, this is not due to the proliferation of diagrams, but to integration over momenta in some special Feynman diagrams.

We will analyze here, following [13, 3], a classical example of renormalons in QCD. Consider the correlation functions of two vector currents $j_\mu = \bar{q}\gamma_\mu q$ of massless quarks

$$(-\mathrm{i})\int \mathrm{d}^4x\, \mathrm{e}^{-\mathrm{i}qx}\, \langle 0|T\,(j_\mu(x)j_\nu(0))|0\rangle = \left(q_\mu q_\nu - q^2 g_{\mu\nu}\right)\, \Pi(Q^2) \tag{5.103}$$

with $Q^2 = -q^2$. We now compute the contribution of the fermion bubble diagrams shown in Fig. 25 to the Adler function

$$D(Q^2) = 4\pi^2\, \frac{\mathrm{d}\Pi(Q^2)}{\mathrm{d}Q^2}. \tag{5.104}$$

The renormalized fermion loop is given by

$$-\beta_{0f}\alpha_s \left[\ln\left(-\frac{k^2}{\mu^2}\right) + C\right] \tag{5.105}$$

where α_s is the running coupling constant (5.10) and C is a scheme-dependent constant (in the $\overline{\mathrm{MS}}$ scheme $C = -5/3$). Let us consider the type of diagrams contributing to (5.105) shown in Fig. 25. To calculate their contribtution, we integrate over the loop momentum of the 'large' fermion loop and the angles of the gluon momentum k. Defining $\hat{k}^2 = -k^2/Q^2$, we obtain

$$D = \sum_{n=0}^{\infty} \alpha_s \int_0^\infty \frac{d\hat{k}^2}{\hat{k}^2}\, F(\hat{k}^2) \left[\beta_{0f}\alpha_s \ln\left(\hat{k}^2 \frac{Q^2 e^{-5/3}}{\mu^2}\right)\right]^n . \tag{5.106}$$

where we have plugged in the leading logarithmic term of the fermion loop in each blob, and $F(\hat{k}^2)$ is a function of $\hat{k}^2$ which can be computed explicitly, see [13]. For us it will be enough to know that F can be expanded around $\hat{k}^2 = 0$ in power series of $\hat{k}^2$, and near $\hat{k}^2 = \infty$ in inverse powers $(\hat{k}^2)^{-1}$.

Let us proceed with the computation of (5.106). We have a geometric series in the logs that can be summed up, and we obtain

$$D = \int_0^\infty \frac{d\hat{k}^2}{\hat{k}^2} F(\hat{k}^2) \frac{\alpha_s}{1 - \beta_{0f}\alpha_s \log\left(\hat{k}^2 \frac{Q^2 e^{-5/3}}{\mu^2}\right)}. \tag{5.107}$$

It remains to perform the integral over $\hat{k}^2$. We then split the integral in two regions

$$(0, \mu^2/(Q^2 e^{-5/3}), \qquad (\mu^2/(Q^2 e^{-5/3}), \infty) \tag{5.108}$$

around $\hat{k}^2 = 0, \infty$, respectively. For the first integral in the IR region we find terms with the generic form

$$\int_0^{\mu^2/(Q^2 e^{-5/3})} \mathrm{d}\hat{k}^2 (\hat{k}^2)^{h-1} \frac{\alpha_s}{1 - \beta_{0f}\alpha_s \ln\left(\hat{k}^2 \frac{Q^2 e^{-5/3}}{\mu^2}\right)}. \tag{5.109}$$

Let us introduce the variable

$$t = \log \frac{\mu^2/(Q^2 e^{-5/3})}{\hat{k}^2}. \tag{5.110}$$

The integral reads,

$$\left(\frac{\mu^2}{Q^2 e^{-5/3}}\right)^h \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-th} \frac{\alpha_s}{1 + \beta_{0f}\alpha_s t}. \tag{5.111}$$

A further normalization $t \to th$ leads to

$$\frac{1}{h}\left(\frac{\mu^2}{Q^2 e^{-5/3}}\right)^h \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-t} \frac{\alpha_s}{1 + \frac{\beta_{0f}}{h}\alpha_s t}. \tag{5.112}$$

This has precisely the structure of the Borel transform (4.51), with poles at

$$t = -\frac{h}{\beta_{0f}}. \tag{5.113}$$

An explicit computation shows that $h \geq 2$ (indeed, F goes near $\hat{k}^2 = 0$ as $(\hat{k}^2)^2$), therefore there are poles at $-2/\beta_{0f}$, $-3/\beta_{0f}$, $\cdots$. These are called *IR renormalons*. We can now repeat the procedure with the UV integral,

$$\int_{\mu^2/(Q^2 e^{-5/3})}^\infty \mathrm{d}\hat{k}^2 (\hat{k}^2)^{-1-r} \frac{\alpha_s}{1 - \beta_{0f}\alpha_s \ln\left(\hat{k}^2 \frac{Q^2 e^{-5/3}}{\mu^2}\right)}. \tag{5.114}$$

Here we introduce the variable

$$t = \log \frac{\hat{k}^2}{\mu^2/(Q^2 e^{-5/3})}, \tag{5.115}$$

and we find

$$\left(\frac{Q^2 e^{-5/3}}{\mu^2}\right)^r \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-rt} \frac{\alpha_s}{1 + \beta_{0f}\alpha_s t}. \tag{5.116}$$

A further normalization $t \to tr$ leads to

$$\frac{1}{r}\left(\frac{Q^2 e^{-5/3}}{\mu^2}\right)^r \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-rt} \frac{\alpha_s}{1+\frac{\beta_{0f}}{r}\alpha_s t}. \tag{5.117}$$

Now the poles are at

$$t = \frac{r}{\beta_{0f}}. \tag{5.118}$$

These are called *UV renormalons*. An explicit computation shows that $r \geq 1$, hence there are poles at $1/\beta_{0f}$, $2/\beta_{0f}$, $\cdots$.

One can argue that, after including other effects, the coefficient β_{0f} becomes β_0, the full coefficient of the beta function at one loop [13]. Since $\beta_0 < 0$ in asymptotically free theories, we see that IR renormalons lead to poles on the positive real axis. The resulting perturbative expansion is therefore not Borel summable, and the ambiguities due to these poles are of the form

$$\left(\frac{\mu^2}{Q^2}\right)^h \mathrm{e}^{h/(\beta_0 \alpha_s(\mu))} = \left(\frac{\Lambda^2}{Q^2}\right)^h, \tag{5.119}$$

where Λ is the dynamically generated scale of QCD. These are power corrections due to nonperturbative effects.

The fact that IR renormalons are located at (5.113) but with β_0 instead of β_{0f} can be argued by using RG arguments, as we will now show following [51] (the original argument is due to Parisi [68]). A generic correlation function in QCD can be written as

$$F(\alpha_s) = F_{\mathrm{p}}(\alpha_s) + F_{\mathrm{np}}(\alpha_s) \tag{5.120}$$

where

$$F_{\mathrm{p}}(\alpha_s) = \sum_{n=0}^{\infty} f_n \alpha_s^{n+1} \tag{5.121}$$

is the perturbative contribution, and $F_{\mathrm{np}}(\alpha_s)$ is a nonperturbative contribution. If $F(\alpha_s)$ is given by Adler function that we discussed above we can calculate it through an operator product expansion (OPE), and $F_{\mathrm{np}}(\alpha_s)$ can be calculated as a *condensate* [74]. For example, the contribution of the gluon condensate would be

$$F_{\mathrm{np}}(\alpha_s) = \frac{1}{Q^4}\langle 0|F^2|0\rangle_\mu C(Q/\mu, \alpha_s), \tag{5.122}$$

where we have indicated explicitly the dependence on the renormalization scale μ. Since both $F(\alpha_s)$ and $F_{\mathrm{p}}(\alpha_s)$ are separately RG-invariant, the same must happen to $F_{\mathrm{np}}(\alpha_s)$. We then have the equation

$$\left(\mu^2\frac{\partial}{\partial\mu^2} + \beta(\alpha_s)\frac{\partial}{\partial\alpha_s} + \gamma_n(\alpha_s)\right) F_{\mathrm{np}}(\alpha_s) = 0 \tag{5.123}$$

where

$$\gamma_n(\alpha_s) = \gamma_1\alpha_s + \cdots \tag{5.124}$$

is the anomalous dimension of the operator. Using this equation, one can determine $F_{\mathrm{np}}(\alpha_s)$ up to an overall constant,

$$F_{\mathrm{np}}(\alpha_s) = C\left(\frac{\mu^2}{Q^2}\right)^{d/2} \alpha_s^\delta \exp\left[\frac{d}{2\beta_0\alpha_s}\right]\left(1+\mathcal{O}(\alpha_s)\right). \tag{5.125}$$

where d is the dimension of the condensate. This can be easily checked

$$\mu^2 \frac{\partial}{\partial \mu^2} F_{\rm np}(\alpha_s) = \frac{d}{2} F_{\rm np}(\alpha_s), \tag{5.126}$$

$$\begin{aligned} \beta(\alpha_s)\frac{\partial}{\partial \alpha_s} F_{\rm np}(\alpha_s) &= \Big(\beta_0 \alpha_s^2 + \beta_1 \alpha_s^3 + \cdots\Big)\Big[\frac{\delta}{\alpha_s} - \frac{d}{\beta_2 \alpha_s^2} + \cdots\Big] F_{\rm np}(\alpha_s) \\ &= \Big(-\frac{d}{2} + \delta \beta_0 \alpha_s - \frac{d\beta_1}{2\beta_0 \alpha_s} + \cdots\Big) F_{\rm np}(\alpha_s), \end{aligned} \tag{5.127}$$

so consistency requires

$$\delta = \frac{d}{2}\frac{\beta_1}{\beta_0^2} - \frac{\gamma_1}{\beta_0}. \tag{5.128}$$

Now, the perturbative part $F_{\rm p}(\alpha_s)$ is typically not Borel summable. If we define the Borel transform

$$B_F(z) = \sum_{n=0}^{\infty} \frac{f_n}{n!} z^n, \tag{5.129}$$

we obtain a representation

$$F_{\rm p}(\alpha_s) = \int_0^\infty {\rm d}z \, {\rm e}^{-z/\alpha_s} B_F(z) \tag{5.130}$$

(the fact that there is no $1/\alpha_s$ in front of this integral is due to the fact that our original perturbative series has an extra factor of α_s). Assuming now a divergence of the type (4.49), one obtains an *ambiguity* in the Borel representation leading to an imaginary part of the form

$${\rm Im}\, F_{\rm p}(\alpha_s) \sim \alpha_s^{1-b} {\rm e}^{-A/\alpha_s}. \tag{5.131}$$

As in the calculation of the ground state energy for the double-well, the full correlation function must be real, and this requires that this imaginary part *cancels* against an imaginary part coming from the nonperturbative condensate above. In other words, the condensate must be also ambiguous in a correlated way. This means that the coefficient C in (5.125) should have an imaginary part which *cancels* agains the imaginary part of the Borel resummation (5.131). A necessary condition for this cancellation to take place is that the location of the first pole in the Borel plane A, which appears in the exponent of (5.131), equals the corresponding exponent in (5.125). We then obtain

$$A = -\frac{d}{2\beta_0}. \tag{5.132}$$

This gives the location of the IR renormalon corresponding to the condensate of dimension d. Also, comparing the exponent of the leading term in g we find

$$1 - b = \delta. \tag{5.133}$$

We then see that, by using RG arguments, assuming the validity of the OPE for the vevs of currents, and requiring consistency of the underlying field theory, we can relate perturbative and nonperturbative effects in a nontrivial way. In particular we find that the location of the IR renormalon involves the one-loop coefficient of the *full* beta function, and the poles found at (5.132) should be identified with those found in (5.113) after setting $h = d/2$.

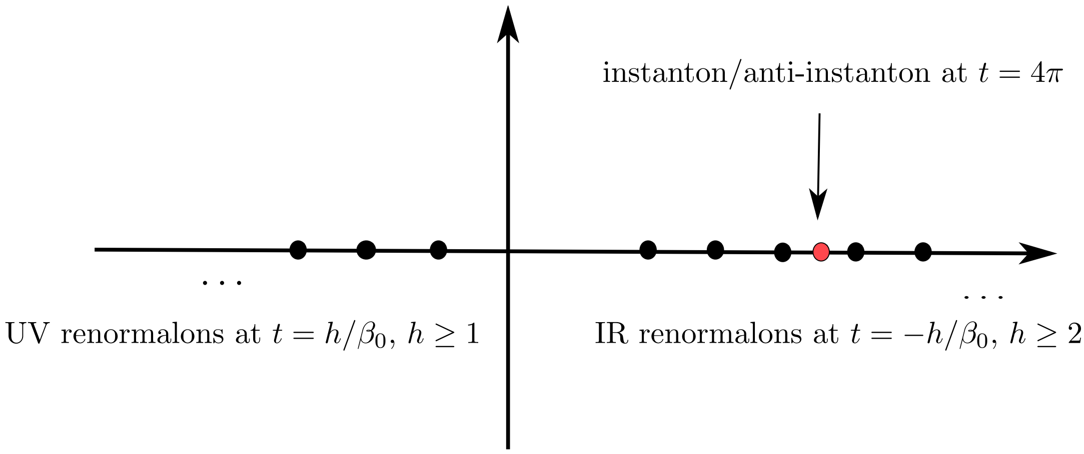


Figure 26: The conjectural structure of the Borel plane for the current-current correlation function in QCD.

To check the consistency of this picture, we point out that (5.119) are precisely the nonperturbative effects that one finds in OPEs due to condensates. The fact that the first condensate contributing to the Adler function is the gluon condensate, with $d = 4$, is also consistent with the fact that $h = 2$ is the first contribution appearing in perturbation theory in (5.112). Conversely, one can use renormalons to obtain hints about nonperturbative effects in the computation of correlation functions, see [13].

What about the role of instantons in QCD and their effect on the large order behavior of perturbation theory? It has been argued [18] that, just as the large order behavior in the double well is due to an instanton-anti-instanton pair, in the same way the instanton-induced large order behavior of QCD is due to the same configuration, with total topological charge zero but with action equal to twice the action of a Yang–Mills instanton ($n = 2$ in (5.63)):

$$S = \frac{16\pi^2}{g^2} = \frac{4\pi}{\alpha_s}. \tag{5.134}$$

This would lead to a singularity in the Borel plane at

$$z_{\text{inst}} = 4\pi. \tag{5.135}$$

A configuration of n instanton-anti-instanton pairs would lead to singularities at $4\pi n$, $n \geq 2$. Notice that, in general, renormalons are *more important* than instantons in determining the large order behavior. For example, for the Adler function, the renormalon singularity which is closest to the origin is located at

$$|z_{\text{ren}}| \leq \frac{12\pi}{11 N_c} < 4\pi \tag{5.136}$$

and corresponds to the UV renormalon. We depict in Fig. 26 the conjectural structure of the Borel plane for the current-current correlation function in QCD.

6. Instantons, fermions and supersymmetry

6.1 Instantons in supersymmetric quantum mechanics

We will now study a very interesting variant of the quantum mechanical models that we have been looking at: supersymmetric quantum mechanics. This model was famously first considered in [89], and studied in detail in [72], which we will mainly follow.

6.1.1 General aspects

On top of the usual bosonic operators $\hat{q}, \hat{p}$, we introduce *Grassmann* variables $\hat{\psi}_{1,2}$ which obey anticommutation relations,

$$\{\hat{\psi}_\alpha, \hat{\psi}_\beta\} = \delta_{\alpha\beta}. \tag{6.1}$$

It is more useful to consider the creation and annihilation operators

$$\hat{\psi}_\pm = \frac{1}{\sqrt{2}}\left(\hat{\psi}_1 \pm \mathrm{i}\hat{\psi}_2\right), \tag{6.2}$$

which satisfy

$$\{\hat{\psi}_+, \hat{\psi}_-\} = 1, \qquad \hat{\psi}_\pm^2 = 0. \tag{6.3}$$

This algebra can be represented by the matrices

$$\hat{\psi}_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \hat{\psi}_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{6.4}$$

Wave-functions are then represented by vector-valued objects,

$$\Psi(x) = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \end{pmatrix}. \tag{6.5}$$

We have the following representation for the operators:

$$\hat{\psi}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \hat{\psi}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \tag{6.6}$$

and the commutator

$$[\hat{\psi}_1, \hat{\psi}_2] = \frac{\mathrm{i}}{2}\sigma_3. \tag{6.7}$$

The Hamiltonian of the system is taken to be

$$\hat{H} = \frac{1}{2}\hat{p}^2 + V(\hat{q}) - \frac{\mathrm{i}}{2}Y(\hat{q})[\hat{\psi}_1, \hat{\psi}_2] \tag{6.8}$$

which on the space of wavefunctions becomes

$$\hat{H} = -\frac{1}{2}\frac{\partial^2}{\partial q^2} + V(q) + \frac{1}{2}Y(q)\sigma_3. \tag{6.9}$$

Since σ_3 commutes with the Hamiltonian, we can diagonalize it simultaneously. Therefore, we can study the spectrum by considering wavefunctions of the form

$$\begin{pmatrix} \phi_1(q) \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \phi_2(q) \end{pmatrix}. \tag{6.10}$$

When the functions $V(q)$, $Y(q)$ appearing in (6.9) satisfy

$$V(q) = \frac{1}{2}W(q)^2, \qquad Y(q) = W'(q), \tag{6.11}$$

where $W(q)$ is called the *superpotential*, the above quantum-mechanical system is supersymmetric. There are two equivalent ways to see this. In the Hamiltonian picture, we simply note that there are two conserved fermionic charges,

$$\begin{aligned} \hat{Q}_+ &= (p - \mathrm{i}W)\,\hat{\psi}_+, \\ \hat{Q}_- &= (p + \mathrm{i}W)\,\hat{\psi}_- \end{aligned} \tag{6.12}$$

which satisfy

$$H = \frac{1}{2}\{\hat{Q}_+, \hat{Q}_-\}. \tag{6.13}$$

In matrix notation, they can be written as

$$\hat{Q}_+ = \begin{pmatrix} 0 & -\mathrm{i}\,(\partial_q + W(q)) \\ 0 & 0 \end{pmatrix}, \qquad \hat{Q}_- = \begin{pmatrix} 0 & 0 \\ -\mathrm{i}\,(\partial_q - W(q)) & 0 \end{pmatrix}. \tag{6.14}$$

In the Lagrangian picture, we just have to show that with the above choice there are two fermionic symmetries in the Lagrangian. This can be done in detail by using standard superspace techniques (see Appendix A of [28] for a detailed derivation). The outcome is that the Lagrangian (in components, and in Minkowski space)

$$L = \frac{1}{2}\dot{q}^2 + \frac{\mathrm{i}}{2}(\psi_-\dot{\psi}_+ - \dot{\psi}_-\psi_+) + \frac{1}{2}D^2 + Df'(q) + \frac{[\psi_-, \psi_+]}{2}f''(q) \tag{6.15}$$

is invariant under

$$\begin{aligned} \mathrm{i}\delta q &= \epsilon_-\psi_- - \psi_+\epsilon_+, \\ \delta\psi_\pm &= \mp\mathrm{i}\epsilon_\mp D + \epsilon_\mp\dot{q}, \\ \delta D &= \epsilon\dot{\psi}_+ + \dot{\psi}_-\epsilon_-. \end{aligned} \tag{6.16}$$

Since D is an auxiliary field, we can integrate it out to obtain $D = -f'(q)$, and upon setting

$$f'(q) = W(q) \tag{6.17}$$

we recover the Lagrangian above.

Notice that in the theory with Hamiltonian

$$H = \frac{1}{2}p^2 + \frac{1}{2}W(q)^2 + \frac{1}{2}W'(q)\sigma_3 \tag{6.18}$$

the fermionic sectors with σ_3 eigenvalues ± 1 have different potentials,

$$V_\pm(q) = \frac{1}{2}W(q)^2 \pm \frac{1}{2}W'(q). \tag{6.19}$$

6.1.2 Supersymmetry breaking

One important question in this type of theories is: is SUSY spontaneously broken or not? In order to answer this question, let us first make some general remarks.

First, in SUSY theories, the energy of any state is positive or zero. Moreover, a state can have zero energy if and only if it is annihilated by all supercharges. Indeed, if

$$H|0\rangle = 0, \tag{6.20}$$

then

$$0 = \langle 0|H|0\rangle = \langle 0|\left(\hat{Q}_+\hat{Q}_- + \hat{Q}_+\hat{Q}_-\right)|0\rangle = \|\hat{Q}_-|0\rangle\|^2 + \|\hat{Q}_+|0\rangle\|^2 \tag{6.21}$$

where we used that $\hat{Q}_\pm$ are Hermitian conjugate to each other. Since the last term is a sum of positive definite quantities, we must have

$$\hat{Q}_+|0\rangle = \hat{Q}_-|0\rangle = 0, \tag{6.22}$$

which is the condition to have a supersymmetric ground state.

Conversely, any state which is not annihilated by both supercharges must have a non-zero energy. If SUSY is spontaneously broken, then at least one of the SUSY charges does not annihilate the ground state, and this means that there are actually *two* states with the same energy (but different fermion number). Indeed, let $|\psi\rangle$ be a non-supersymmetric ground state:

$$\hat{H}|\psi\rangle = E_0|\psi\rangle. \tag{6.23}$$

Then, *if*

$$|\psi_+\rangle = \hat{Q}_+|\psi\rangle \neq 0 \quad \text{or} \quad |\psi_-\rangle = \hat{Q}_-|\psi\rangle \neq 0 \tag{6.24}$$

we have

$$\hat{H}|\psi_\pm\rangle = H\hat{Q}_\pm|\psi\rangle = \hat{Q}_\pm H|\psi\rangle = E_0|\psi_\pm\rangle. \tag{6.25}$$

and the ground state is doubly degenerate.

Witten [89] has pointed out a very simple criterium to know whether SUSY is broken for a given superpotential. A SUSY ground state must satisfy the equations

$$\left(\frac{\partial}{\partial q} - W(q)\right)\phi_1(q) = 0, \qquad \left(\frac{\partial}{\partial q} + W(q)\right)\phi_2(q) = 0, \tag{6.26}$$

with the immediate solution

$$\phi_1(q) = \phi_1(q_0)\exp\left(\int_{q_0}^{q} W(q')\mathrm{d}q'\right), \qquad \phi_2(q) = \phi_2(q_0)\exp\left(-\int_{q_0}^{q} W(q')\mathrm{d}q'\right). \tag{6.27}$$

There are two cases here (assuming $W(x)$ to be a polynomial)

1. If the highest power of $W(q)$ is even, $\int W(q')\mathrm{d}q'$ will have its highest power of odd degree, and none of the above functions is normalizable. In this case SUSY is broken, and we expect a degenerate ground state of nonzero energy.

2. If the highest power of $W(q)$ is odd, $\int W(q')\mathrm{d}q'$ will have its highest power of even degree, and we will have one normalizable ground state with $\phi_1 = 0$ or $\phi_2 = 0$ (depending on the sign of the highest power). In this case supersymmetry is unbroken.

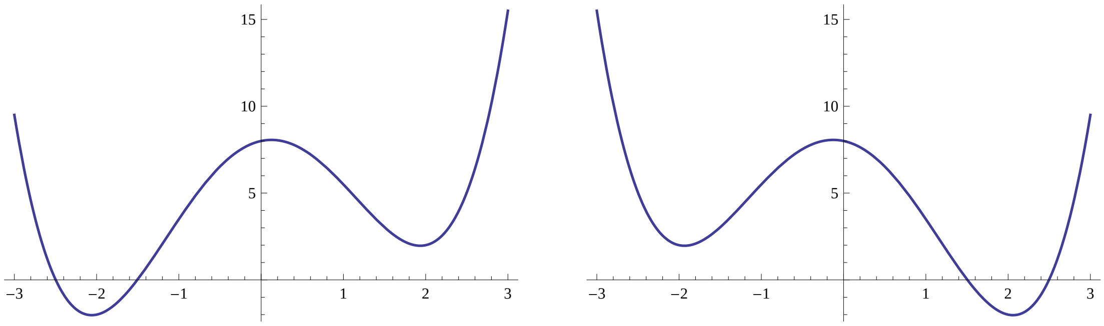


Figure 28: The potentials $V_\pm(q)$ (left and right, respectively) for the Hamiltonian (6.32), represented here for $\lambda = 1$, $\mu = 2$.

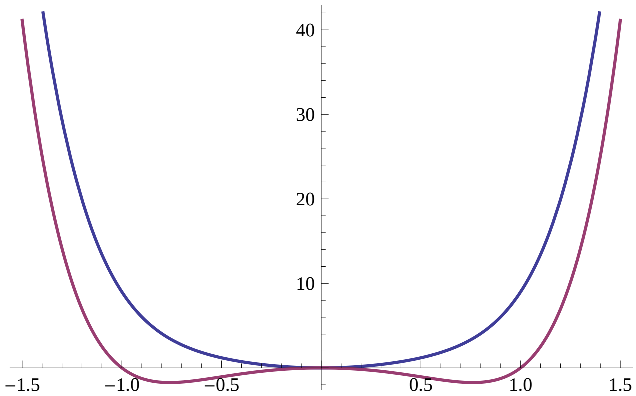


Figure 27: The potentials $V_\pm(q)$ (top and bottom line, respectively) for the Hamiltonian (6.29), represented here for $g = 3$.

As an example of the second situation, let us take

$$W(q) = gq^3, \tag{6.28}$$

so that

$$H = \frac{1}{2}p^2 + \frac{g^2q^6}{2} + \frac{3gq^2}{2}\sigma_3. \tag{6.29}$$

The ground state is in this case

$$\Phi = \begin{pmatrix} 0 \\ \phi_2(q) \end{pmatrix}, \qquad \phi_2(q) = Ce^{-gq^4/4} \tag{6.30}$$

which is normalizable. Notice that the potentials $V_\pm(q)$ are quite different in this case (see Fig. 27), and the ground state is an eigenstate for $V_-(q)$.

An example of the first situation is the quadratic superpotential

$$W(q) = \lambda q^2 - \mu^2. \tag{6.31}$$

In this case, the Hamiltonian is

$$H = \frac{1}{2}p^2 + \frac{\lambda^2}{2}\left(q^2 - \frac{\mu^2}{\lambda}\right)^2 + \lambda q\sigma_3 \tag{6.32}$$

The potentials $V_\pm(q)$ are shown in Fig. 28. Notice that the effect of the fermions is to lift the degeneracy between the two wells of the potential $W(x)^2/2$. There are now *two* ground states with $\sigma_3 = \pm 1$ and nonzero energy, localized in different wells of the potential. Clearly, it should be possible to transform one state into the other by just changing the fermion number (i.e. going from $\sigma_3 = +1$ to $\sigma_3 = -1$), but in order to do this one has in addition to tunnel from one of the vacua to the other. This is only possible by means of an instanton process. We then expect the transition amplitude

$$\langle q_+, \uparrow, T = -\infty | \hat{\psi}_+(t) | q_-, \downarrow, T = +\infty \rangle, \tag{6.33}$$

where $q_\pm$ are the minima of $V_\pm(q)$, to be nonzero and given by an instanton amplitude. In the path integral formalism, this transition amplitude is given by

$$\int \mathcal{D}q\,\mathcal{D}\psi_+\,\mathcal{D}\psi_-\;\psi_+(t)\mathrm{e}^{-S} \tag{6.34}$$

As we will see, for this integral not to be zero, the field ψ_+ must have a zero mode in the background of an instanton, which can be used to "soak up" the insertion of the operator $\hat{\psi}_+$ in the path integral.

6.1.3 Instantons and fermionic zero modes

Let us now consider the Euclidean version of the Lagrangian (6.15). After Wick rotating $t \to \mathrm{i}t$, we find

$$L_E = \frac{1}{2}\dot{q}^2 + \frac{1}{2}W^2(q) - \frac{1}{2}\psi_-\dot{\psi}_+ - \frac{1}{2}\psi_+\dot{\psi}_- + \psi_-\psi_+W'(q). \tag{6.35}$$

The equations of motion for the Euclidean Lagrangian are

$$\begin{aligned} \ddot{q} - W'(q)W(q) - \psi_-\psi_+W''(q) &= 0, \\ \dot{\psi}_\pm \mp W'(q)\psi_\pm &= 0. \end{aligned} \tag{6.36}$$

Clearly, a solution of this EOM is simply given by

$$\psi_+^c = 0, \qquad \dot{q}_c(t) = \pm W(q_c(t)), \tag{6.37}$$

therefore there are two types of instantons depending on the choice of sign in this equation.

Remark 6.1. Already in this system we can see two general properties of SUSY models which make the analysis of instantons much easier. First,we have been able to integrate the EOM once and obtain an instanton equation for $x(t)$ which is only *first order*. Second, this first order equation is equivalent to the conditions

$$\delta\psi_\pm = 0, \tag{6.38}$$

Indeed, as we can see from (6.16), in the Euclidean theory we have

$$\delta\psi_\pm = -\mathrm{i}\epsilon_\mp\left(\dot{q} \pm W(q)\right). \tag{6.39}$$

The condition (6.38) is very important in supersymmetric systems, and defines *supersymmetric configurations*, i.e. configurations which preserve supersymmetry.

The EOM for $q_c(t)$ can be integrated immediately:

$$t = \pm\int_{q_c(0)}^{q_c(t)} \frac{\mathrm{d}q'}{W(q')} \tag{6.40}$$

Example 6.2. For the superpotential (6.31) the solutions to (6.37) are

$$q_c(t) = \mp\frac{\mu}{\sqrt{\lambda}}\tanh\left(\mu\sqrt{\lambda}(t-t_0)\right), \tag{6.41}$$

corresponding to the $\pm$ sign in (6.37). which are precisely the (anti)instantons of the double well (2.212). They are depicted in Fig. 12.

As in the non-supersymmetric case, we have a bosonic zero mode corresponding to $\dot{q}_c(t)$, i.e.

$$q_0(t) \propto W(q_c(t)). \tag{6.42}$$

There is also a normalizable *fermionic* zero mode:

$$\psi_\pm(t) \propto W(q_c(t)) \qquad \text{if} \quad \dot{q}_c(t) = \pm W(q_c(t)), \tag{6.43}$$

i.e. there are fermionic zero modes for ψ_+ of for ψ_- (but not for both) depending on the choice of instanton or anti-instanton background in (6.37). In the above equation, ψ_0 is a constant Grassmann variables. Notice also that these zero modes are the supersymmetry transformations of the bosonic zero modes, i.e. we can write

$$\zeta_0^\pm \psi_\pm(t) \propto \delta_\pm \left(c_0 W(q_c(t))\right). \tag{6.44}$$

where c_0, $\zeta_0^\pm$ are commuting and Grassmann constants, respectively, related by supersymmetry transformations. Finally, if we denote by $\nu = \pm 1$ the "instanton number", which is $+1$ (-1) for instantons (anti-instantons, respectively), and if we denote by $N_\pm$ the number of zero modes of $\psi_\pm$, we have the equality

$$N_+ - N_- = -\nu. \tag{6.45}$$

This is our first example of an "index theorem," relating the number of zero modes of fermions to the topological class of the instanton background.

Let us now look at the operator describing fluctuations around this solution. For $q(t)$, this operator is given by

$$M_B = -\partial_t^2 + W''(q_c(t))W(q_c(t)) + (W'(q_c(t)))^2. \tag{6.46}$$

One important point is that M_B can be factorized as

$$M_B = -\left(\partial_t + W'(q_c(t))\right)\left(\partial_t - W'(q_c(t))\right) = M_F^\dagger M_F \tag{6.47}$$

for the first choice of sign in (6.37). Here we have denoted

$$M_F = \partial_t - W'(q_c(t)), \qquad M_F^\dagger = -\partial_t - W'(q_c(t)). \tag{6.48}$$

Example 6.3. In the case of the quadratic superpotential (6.31) the bosonic operator is given by (2.213), which is proportional to the Pöschl–Teller potential $M_{2,2}$. In this case, the factorization (6.47) is nothing but the factorization of the Hamiltonian in (2.97), since

$$M_F = \partial_t + 2\mu\sqrt{\lambda}\tanh\left(\mu\sqrt{\lambda}(t - t_0)\right) \tag{6.49}$$

which is the operator A_2 up to a rescaling of t. Indeed, the solvability of the Pöschl–Teller potential is a consequence of a hidden supersymmetry, see for example [29].

We now introduce the eigenvalue problem for the operators M_F, $M_F^\dagger$ as follows:

$$\begin{aligned} M_F \psi_+^n &= \omega_F^n \psi_-^n, \\ M_F^\dagger \psi_-^n &= \omega_F^n \psi_+^n, \end{aligned} \tag{6.50}$$

since they are related by Hermitian conjugation. By acting with $M_F^\dagger$ on the first equation we find that ψ_+^n satisfies

$$M_B \psi_+^n = (\omega_F^n)^2 \, \psi_+^n \tag{6.51}$$

i.e. it is an eigenvector of the bosonic operator M_B. Therefore, for any solution of the eigenvalue equation

$$M_B q_n = (\omega_B^n)^2 q_n \tag{6.52}$$

we could obtain a solution of the fermionic equations by setting

$$\psi_+^n(t) = q_n(t), \qquad \omega_F^n = \omega_B^n, \qquad n \geq 0. \tag{6.53}$$

This pairing of the bosonic and fermionic modes is typical of supersymmetric systems, as first pointed out in [31]. The eigenfunctions ψ_-^n are then given by

$$\psi_-^n(t) = \frac{1}{\omega_F^n} M_F q_n(t), \qquad n \geq 1, \tag{6.54}$$

since the zero mode q_0 is annihilated by M_F. Therefore, there is no zero mode for ψ_-^n. In fact, there cannot be a zero mode of both M_F and $M_F^\dagger$, since the solution to the zero mode equations are, as in (6.27),

$$\psi_\pm^0(t) \propto \exp\left(\pm \int^t W'(q_c(t'))\, \mathrm{d}t' \right), \tag{6.55}$$

and they cannot be both normalizable.

Remark 6.4. For the quadratic superpotential (6.31) we can write the modes $\psi_\pm^n$ very explicitly by using the results for the Pöschl–Teller potential with $\ell = 2$ (for simplicity we set $\mu = \lambda = 1$). For $\psi_+^n(t) = q_n(t)$ we have two bound states, given in (2.109), (2.108), and a continuum of scattering states (2.107) with the explicit expression

$$\psi_+^{(k)}(t) = -\left(k^2 - 2 + 3\mathrm{sech}^2(t) + 3\mathrm{i}k \tanh(t)\right) \frac{\mathrm{e}^{\mathrm{i}kt}}{\sqrt{2\pi(k^2+1)(k^2+4)}}. \tag{6.56}$$

For the states $\psi_-^n(t)$ we have a bound state

$$\psi_-^1(t) \propto \frac{1}{\cosh(t)} \tag{6.57}$$

and a continuum given by (2.105), or explicitly,

$$\psi_-^{(k)}(t) = (-\mathrm{i}k + \tanh(t)) \frac{\mathrm{e}^{\mathrm{i}kt}}{\sqrt{2\pi(k^2+1)}}. \tag{6.58}$$

Using that $A_2 A_2^\dagger = M_{1,2}$ it is easy to verify these statements, where the eigenvalue for the scattering states is given by

$$\omega_F^{(k)} = \sqrt{k^2+4}. \tag{6.59}$$

We are now ready to evaluate the path integral (6.34) at one-loop. The path connecting the vacua q_- to q_+ is precisely the anti-instanton, corresponding to the $+$ sign in (6.37) and to the $-$ sign in (2.212). This leads to a fermionic zero mode for ψ_+. The expansion of $\psi_+(t)$ is given by

$$\psi_+(t) = \zeta_0^+ \frac{\dot{q}_c^{t_0}(t)}{\|\dot{q}_c\|} + \sum_{n\geq 1} \zeta_n^+ q_n(t), \tag{6.60}$$

while

$$\psi_-(t) = \sum_{n\geq 1} \zeta_n^- \psi_-^n(t), \tag{6.61}$$

The kinetic term for the fermions in $-S_E$ can be written as

$$\frac{1}{2} \left(\psi_+ \ \psi_-\right) \mathcal{M}_F \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix} \tag{6.62}$$

where

$$\mathcal{M}_F = \begin{pmatrix} 0 & -M_F^\dagger \\ M_F & 0 \end{pmatrix}. \tag{6.63}$$

When acting on the basis (6.50) of orthonormal modes, the operator $\mathcal{M}_F$ becomes diagonal

$$\mathcal{M}_F \rightarrow \begin{pmatrix} -\omega_F^n & 0 \\ 0 & \omega_F^n \end{pmatrix} \tag{6.64}$$

and (6.62) is given by

$$-\sum_{n\geq 1} \omega_F^n \zeta_n^+ \zeta_n^- \tag{6.65}$$

We have the following measure for zero modes

$$\int \frac{\mathrm{d}c_0}{\sqrt{2\pi}} \int \mathrm{d}\zeta_0^+ = \frac{S_c^{1/2}}{\sqrt{2\pi}} \int \mathrm{d}t_0 \int \mathrm{d}\zeta_0^+. \tag{6.66}$$

The integration over ζ_0^+ picks up the first term in (6.60). Notice that, if there was no insertion of the fermionic field $\psi_+(t)$, the path integral would *vanish* due to the presence of the fermionic zero mode. The Gaussian integration over the fermionic non-zero modes is simply

$$\int \prod_{n\geq 1} \mathrm{d}\zeta_n^+ \mathrm{d}\zeta_n^- \exp\left[-\sum_{n\geq 1} \omega_F^n \zeta_n^+ \zeta_n^-\right] = \prod_{n\geq 1} \omega_F^n. \tag{6.67}$$

Putting everything together we obtain

$$\frac{\mathrm{e}^{-S_c}}{\sqrt{2\pi}} \left[\frac{\det' \mathcal{M}_F}{\det' M_B}\right]^{1/2} \int_{-\infty}^{\infty} \mathrm{d}t_0 \, \dot{q}_c^{t_0}(t) \tag{6.68}$$

where

$$\left[\frac{\det' \mathcal{M}_F}{\det' M_B}\right]^{1/2} = \prod_{n\geq 1} \frac{\omega_F^n}{\omega_B^n}. \tag{6.69}$$

Due to (6.53), one could think that the above quotient of determinants is just one. Indeed, this would be the case if the spectrum was discrete. But in the presence of a continuous spectrum this "cancellation between fermionic and bosonic degrees of freedom" does not take place, contrary to what was argued in [30]. This was first noticed in the context of corrections to the mass of supersymmetric kinks in [56]. An explicit computation [72] shows that indeed this quotient has in general the value

$$\frac{\det' \mathcal{M}_F}{\det' M_B} = 2W'(q_+). \tag{6.70}$$

Putting everything together, we find

$$\int \mathcal{D}q\, \mathcal{D}\psi_+\, \mathcal{D}\psi_-\, \psi_+(t) \mathrm{e}^{-S} = \sqrt{\frac{W'(q_+)}{\pi}} \mathrm{e}^{-S_c} (q_+ - q_-). \tag{6.71}$$

One way to see that there is no cancellation between the fermionic and the bosonic degrees of freedom is to look carefully at the densities of states of the fields $\psi_\pm$, as in the soliton calculation of [56]. Since ψ_+ satisfies the same equation than the bosonic modes, the continuous eigenvectors have the same phase shifts that we found in (2.118):

$$\psi_k^+(t) \sim \exp\left[\mathrm{i} \left(kt \pm \frac{\delta(k)}{2} \right) \right], \qquad t \to \pm\infty. \tag{6.72}$$

On the other hand, the asymptotic behavior of ψ_- can be deduced from the first equation in (6.50):

$$\psi_k^-(t) \sim \left(k \mp 2\mathrm{i}\mu\sqrt{\lambda} \right) \exp\left[\mathrm{i} \left(kt \pm \frac{\delta(k)}{2} \right) \right], \qquad t \to \pm\infty. \tag{6.73}$$

This means that there is an extra phase shift $\theta(k)/2$ given by

$$\tan\left(\frac{\theta(k)}{2} \right) = -\frac{2\mu\sqrt{\lambda}}{k} \tag{6.74}$$

and the densities of states are then given by

$$\rho_+(k) = \frac{1}{2\pi}\delta'(k), \qquad \rho_-(k) = \frac{1}{2\pi}\left(\delta'(k) + \theta'(k) \right). \tag{6.75}$$

We then see that the eigenstates of the Dirac operator $\mathcal{M}_F$ with different chiralities have different densities. This phenomenon is called the *spectral asymmetry* of the Dirac operator.

We can now proceed with the calculation of the determinant. The log of the square root of (6.81) is then

$$\int_{-\infty}^{\infty} \mathrm{d}k \, \log\left(\sqrt{k^2 + 4\lambda\mu^2} \right) \rho_t(k) \tag{6.76}$$

where

$$\rho_t(k) = \rho_+(k) + \rho_-(k) - 2\rho(k) = \frac{1}{2\pi}\theta'(k) = \frac{1}{\pi}\frac{\sqrt{\lambda}\mu}{k^2 + 4\lambda\mu^2}, \tag{6.77}$$

and we finally obtain

$$\int_{-\infty}^{\infty} \mathrm{d}k \, \log\left(\sqrt{k^2 + 4\lambda\mu^2} \right) \rho_t(k) = \frac{1}{2}\log\left(4\sqrt{\lambda}\mu \right) \tag{6.78}$$

in agreement with (6.86).

A different way to compute the above determinant is to compute instead its absolute square

$$\det'(\mathcal{M}_F\mathcal{M}_F^\dagger) = \det' \begin{pmatrix} M_F^\dagger M_F & 0 \\ 0 & M_F M_F^\dagger \end{pmatrix} = \det'\left(M_F^\dagger M_F\right)\det\left(M_F M_F^\dagger\right) \tag{6.79}$$

where the operator

$$M_F M_F^\dagger = -\partial_t^2 + \left(W'(q_c(t))\right)^2 - W''(q_c(t))W(q_c(t)) \tag{6.80}$$

has no zero modes. Therefore, up to an overall phase,

$$\frac{\det'\mathcal{M}_F}{\det' M_B} = \left(\frac{\det\left(-\partial_t^2 + (W'(q_c(t)))^2 - W''(q_c(t))W(q_c(t))\right)}{\det'\left(-\partial_t^2 + (W'(q_c(t)))^2 + W''(q_c(t))W(q_c(t))\right)}\right)^{\frac{1}{2}}. \tag{6.81}$$

This can be computed by using the Gelfand–Yaglom theorem, or, in the case of the quadratic superpotential, by using the results on the spectrum of the Pöschl–Teller potential.

Example 6.5. For (6.31), we find

$$-\partial_t^2 + \left(W'(q_c(t))\right)^2 - W''(q_c(t))W(q_c(t)) = -\partial_t^2 + \mu^2\lambda\left(4 - \frac{2}{\cosh^2\left(\mu\sqrt{\lambda}(t-t_0)\right)}\right) \tag{6.82}$$

The resulting operator is, up to a rescaling of t, the Pöschl–Teller potential $M_{1,2}$:

$$-\partial_t^2 + \mu^2\lambda\left(4 - \frac{2}{\cosh^2\left(\mu\sqrt{\lambda}(t-t_0)\right)}\right) = \mu^2\lambda M_{1,2} \tag{6.83}$$

Notice that in this case $M_{1,2}$ has no zero mode, and one has, by using (2.126), that

$$\frac{\det M_{1,2}}{\det M_{0,2}} = \frac{1}{3}. \tag{6.84}$$

Therefore,

$$\frac{\det M_{1,2}}{\det' M_{2,2}} = 16 \tag{6.85}$$

and we conclude that

$$\frac{\det'\mathcal{M}_F}{\det' M_B} = 4\mu\sqrt{\lambda} \tag{6.86}$$

in agreement with (6.70).

6.2 Fermions and anomalies in Yang–Mills theory

Let us consider the QCD Lagrangian with N_f flavors,

$$g^2\mathcal{L} = \mathrm{i}\sum_{f=1}^{N_f} \bar{\psi}_f D\psi_f - \sum_{f=1}^{N_f} m_f \bar{\psi}_f \psi_f + \cdots \tag{6.87}$$

We can write this in terms of lef-handed and right-handed components

$$\psi_{L,f} = \frac{1-\gamma_5}{2}\psi_f, \qquad \psi_{R,f} = \frac{1+\gamma_5}{2}\psi_f \tag{6.88}$$

as follows

$$g^2\mathcal{L} = \mathrm{i}\sum_{f=1}^{N_f}\left(\bar{\psi}_{L,f}D\psi_{L,f} + \bar{\psi}_{R,f}D\psi_{R,f}\right) - \sum_{f=1}^{N_f} m_f\left(\bar{\psi}_{R,f}\psi_{L,f} + \bar{\psi}_{L,f}\psi_{R,f}\right) + \cdots \tag{6.89}$$

In the world of massless quarks this Lagrangian has two classical $U(1)$ symmetries. The first one is a vectorial $U_V(1)$

$$\psi_f \to \mathrm{e}^{\mathrm{i}\theta}\psi_f, \qquad f = 1, \cdots, N_f. \tag{6.90}$$

The associated current

$$Q^\mu = \sum_{f=1}^{N_f} \bar{\psi}_f \gamma^\mu \psi_f \tag{6.91}$$

is conserved quantum-mechanically. This leads to a conserved quantum number which is just the number of quarks minus the number of antiquarks.

The second classical symmetry is the axial $U_A(1)$

$$\psi_f \to \mathrm{e}^{\mathrm{i}\theta\gamma_5}\psi_f, \qquad f = 1, \cdots, N_f. \tag{6.92}$$

with current

$$J^\mu = \frac{1}{g^2}\sum_{f=1}^{N_f} \bar{\psi}_f \gamma^\mu \gamma_5 \psi_f, \tag{6.93}$$

where we used again the un-hatted quark fields. If this current was conserved quantum-mechanically, there would be an extra conserved quantum number. If it was spontaneously broken it would lead to a Goldstone boson. As we will study in more detail in chapter 10, none of these possibilities are realized in Nature: the $U_A(1)$ symmetry is *anomalous*. The divergence of J^μ can be computed to be

$$\partial_\mu J^\mu = 2N_f q(x) + \frac{2\mathrm{i}}{g^2}\sum_{f=1}^{N_f} m_f \bar{\psi}_f \gamma_5 \psi_f. \tag{6.94}$$

The first term in (6.94) is the anomaly, and leads to a non-vanishing of the divergence even for masless quarks. A particular elegant derivation of this relation, originally due to Fujikawa [41], can be obtained in the path integral formulation by studying the change in the path integral measure. This goes as follows. First, we define the measure on the space of Dirac spinors (we will eliminate the flavour subindex in what follows). To do this we consider the eigenfunctions of the Dirac operator:

$$\mathrm{i}D\psi_n = \lambda_n \psi_n \tag{6.95}$$

and we define the inner product of Dirac fields in the usual way, with an appropriate normalization of the eigenfunctions:

$$\langle \psi_i | \psi_j \rangle = \int \mathrm{d}^{2n}x\, \psi_i{}^{\dagger}(x)\psi_j(x) = \delta_{ij} \tag{6.96}$$

We expand the fields in modes of the Dirac operator as:

$$\psi = \sum_n a_n \psi_n, \qquad \overline{\psi} = \sum_n \bar{b}_n \psi_n^{\dagger}. \tag{6.97}$$

As the change of variables is unitary, we can define the path integral measure as:

$$\mathcal{D}\psi\mathcal{D}\bar{\psi} = \prod_n \mathrm{d}a_n \prod_n \mathrm{d}\bar{b}_n \tag{6.98}$$

Notice that a_n, $\bar{b}_n$ are Grassmann variables. Under an infinitesimal chiral transformation the fields transform as:

$$\begin{aligned} \psi(x) &\to \psi(x) + \mathrm{i}\alpha(x)\gamma_5\psi(x), \\ \overline{\psi}(x) &\to \overline{\psi}(x) + \overline{\psi}(x)\mathrm{i}\gamma_5\alpha(x) \end{aligned} \tag{6.99}$$

The action changes under this transformation as:

$$\frac{1}{g^2}\int \mathrm{d}^{2n}x\, \overline{\psi}\mathrm{i}D\psi \to \frac{1}{g^2}\int \mathrm{d}^{2n}x\, \overline{\psi}\mathrm{i}D\psi - \mathrm{i}\int \mathrm{d}^{2n}x\, \alpha(x)\partial_\mu J^\mu(x) \tag{6.100}$$

where J^μ is the current (6.93). However, in order to obtain the *quantum* conservation law, we must take into account the change in the measure. Define the chiral rotated fields as

$$\begin{aligned} \psi'(x) &= \psi(x) + \mathrm{i}\alpha(x)\gamma_5\psi(x) = \sum_n a'_n \psi_n(x), \\ \overline{\psi}'(x) &= \overline{\psi}(x) + \overline{\psi}(x)\mathrm{i}\gamma_5\alpha(x) = \sum_n \bar{b}'_n \psi_n^{\dagger}. \end{aligned} \tag{6.101}$$

The coefficients a'_n, $\bar{b}'_n$ are computed using the inner product defined above:

$$\begin{aligned} a'_n =& \langle \psi_n | \psi' \rangle = \langle \psi_i | 1 + \mathrm{i}\alpha(x)\gamma_5 | \psi \rangle \\ =& \sum_m \langle \psi_n | 1 + \mathrm{i}\alpha(x)\gamma_5 | \psi_m \rangle a_m = \sum_m C_{nm} a_m \end{aligned} \tag{6.102}$$

where

$$C_{nm} = \langle \psi_n | 1 + \mathrm{i}\alpha(x)\gamma_5 | \psi_m \rangle = \delta_{nm} + \mathrm{i}\langle \psi_n | \alpha(x)\gamma_5 | \psi_m \rangle. \tag{6.103}$$

The measure for the chiral-rotated fields is $\prod_n \mathrm{d}a'_n$, and it can be obtained from (6.103). Since we are dealing with Grassmann quantities the change of variables involves *inverse* of the Jacobian:

$$
\begin{aligned}
\prod_n \mathrm{d}a'_n =&[\det C_{nm}]^{-1}\prod_m \mathrm{d}a_m = \mathrm{e}^{-\mathrm{Tr}\log C_{mn}}\prod_m \mathrm{d}a_m \\
=&\exp\left[-\mathrm{Tr}\log(\delta_{nm}+\mathrm{i}\langle\psi_n|\alpha(x)\gamma_5|\psi_m\rangle)\right]\prod_m \mathrm{d}a_m \\
\approx&\exp\left[-\mathrm{Tr}\,(\mathrm{i}\langle\psi_n|\alpha(x)\gamma_5|\psi_m\rangle)\right]\prod_m \mathrm{d}a_m \\
=&\exp\left[-\mathrm{i}\sum_n\langle\psi_n|\alpha(x)\gamma_5|\psi_n\rangle\right]\prod_m \mathrm{d}a_m
\end{aligned}
\tag{6.104}
$$

A similar analysis can be done for the $\bar{b}$ variables, which give the same contribution to the change in the measure. Performing the inner product in (6.104) we obtain, for the change in the measure,

$$
\prod_n \mathrm{d}a_n\prod_n \mathrm{d}\bar{b}_n \to \prod_n \mathrm{d}a'_n\prod_n \mathrm{d}\bar{b}'_n \exp\Big(-2\mathrm{i}\int \mathrm{d}^{2n}x\,\alpha(x)\sum_n\psi_n^\dagger(x)\gamma_5\psi_n(x)\Big) \tag{6.105}
$$

The effective action changes under an infinitesimal chiral transformation to:

$$
\int\prod_n \mathrm{d}a_n\prod_n \mathrm{d}\bar{b}_n \exp\left[-\frac{1}{g^2}\int \mathrm{d}^{2n}x\,\overline{\psi}\mathrm{i}D\psi+\mathrm{i}\int \mathrm{d}^{2n}x\,\alpha(x)\partial_\mu J^\mu(x)-2\mathrm{i}\int \mathrm{d}^{2n}x\,\alpha(x)A(x)\right], \tag{6.106}
$$

and we have defined the anomaly density $A(x)$ as

$$
A(x)=\sum_n\psi_n^\dagger(x)\gamma_5\psi_n(x). \tag{6.107}
$$

Therefore the anomalous conservation law for the chiral current is

$$
\partial_\mu J^\mu(x)=2A(x). \tag{6.108}
$$

The anomaly density can be evaluated to give [41]

$$
A(x)=N_f q(x), \tag{6.109}
$$

where $q(x)$ is the topological density (5.15). In this way we recover (6.94) in the massless case.

One of the most remarkable consequences of the anomaly is that it leads to the existence of zero modes for the spinors. One way to see this is simply to integrate the density $A(x)$. Let us suppose that $|\psi_n\rangle$ is an eigenfunction of D with a non-zero eigenvalue λ_n. The state

$$
|\psi_n\rangle^\chi=\gamma_5|\psi_n\rangle \tag{6.110}
$$

verifies:

$$
\mathrm{i}D|\psi_n\rangle^\chi=\mathrm{i}D\gamma_5|\psi_n\rangle=-\gamma_5 iD|\psi_n\rangle=-\lambda_n|\psi_n\rangle \tag{6.111}
$$

Since $\mathrm{i}D$ is hermitian, the eigenfunctions for different eigenvalues are orthogonal:

$$
\langle\psi_n|\psi_n\rangle^\chi=\langle\psi_n|\gamma_5|\psi_n\rangle=0 \tag{6.112}
$$

The conclusion is that, in the sum appearing in $A(x)$, only zero modes give non-zero contribution. These zero modes, $|0,i\rangle_\pm$, can be classified according to their chirality:

$$\gamma_5|0,i\rangle_\pm = \pm|0,i\rangle_\pm \tag{6.113}$$

The integral of $A(x)$ can then be written as:

$$\int \mathrm{d}^{2n}x\, A(x) = \sum_{n=1}^{\nu_+}\langle 0,i|\gamma_5|0,i\rangle_+ + \sum_{n=1}^{\nu_-}\langle 0,i|\gamma_5|0,i\rangle_- = N_f\,(\nu_+ - \nu_-)\,, \tag{6.114}$$

which counts the difference between zero modes of positive chirality and negative chirality for one single flavor. For each flavor we have

$$\nu_+ - \nu_- = Q \tag{6.115}$$

where Q is the topological charge. We conclude that in the background of an instanton there are zero modes for the fermions.

As a consequence of the presence of fermionic zero modes, the vacuum-vacuum amplitude in the background of an instanton is always zero. Consider now the correlator:

$$G(x_i,y_i) = \int D\psi D\overline{\psi}(\psi(x_1)\cdots\psi(x_N)\overline{\psi}(y_1)\cdots\overline{\psi}(y_N))\mathrm{e}^{-S} \tag{6.116}$$

We will compute this quantity by an expansion of the Dirac fields taking into account zero-modes:

$$\begin{aligned}\psi &= \sum_{\lambda_n\neq 0} a_n\psi_n + \sum_\alpha C_\alpha\psi_\alpha\\ \overline{\psi} &= \sum_{\lambda_n\neq 0}\bar{b}_n\psi_n^\dagger + \sum_\alpha \bar{C}_\alpha\psi_\alpha^\dagger\end{aligned} \tag{6.117}$$

where ψ_α are the zero-modes of the Dirac operator. The measure reads now:

$$\mathcal{D}\psi\,\mathcal{D}\overline{\psi} = \prod_n \mathrm{d}a_n\mathrm{d}\bar{b}_n\prod_\alpha \mathrm{d}C_\alpha\mathrm{d}\bar{C}_\alpha. \tag{6.118}$$

Introducing this expansion in the Green function (6.116) we obtain:

$$\begin{aligned}G(x_i,y_i) = \int &\prod_n \mathrm{d}a_n\mathrm{d}\bar{b}_n\prod_\alpha \mathrm{d}C_\alpha\mathrm{d}\bar{C}_\alpha\,\mathrm{e}^{-\sum_n \bar{b}_n a_n\lambda_n}\\ &\cdot\prod_{i=1}^{N}\left(\sum_{\lambda_n\neq 0} a_n\psi_n(x_i) + \sum_\alpha C_\alpha\psi_\alpha(x_i)\right)\left(\sum_{\lambda_n\neq 0}\bar{b}_n\psi_n^\dagger(y_i) + \sum_\alpha \bar{C}_\alpha\psi_\alpha^\dagger(y_i)\right)\end{aligned} \tag{6.119}$$

In order to obtain a non-zero result for this integral, we must saturate the zero modes which appear in the measure with an appropriate term from the product. It is clear that the only way to do that is to have as many fields in the correlator as zero modes: $N = \nu_+ + \nu_-$. Once this is guaranteed, the Green function reads:

$$G(x_i,y_i) = (\det{}'\,\mathrm{i}D)\det\left(\psi_\alpha(x_\beta)\right)\det\left(\psi_\alpha^\dagger(y_\beta)\right), \tag{6.120}$$

where the prime in the determinant denotes deletion of the zero eigenvalues. Notice that this selection rule needs more information than the one provided by the index of the operator. Whenever we have a non-zero index we know that there must be zero-modes. However, from the index we don't know the total number of zero modes, but only the difference $\nu_+ - \nu_-$. Notice that in a Green function we must saturate both ν_+ and ν_-. There are many interesting cases in which one is able to obtain, from what are called "vanishing theorems", the dimension of each kernel separately, and therefore one knows what to insert in the path integral to get sensible results.

7. Sigma models at large N

7.1 The $O(N)$ non-linear sigma model

The $O(N)$ non-linear sigma model in two dimensions is one of the best toy models in QFT. It shares two important properties with non-abelian Yang–Mills theory: it is asymptotically free and it has a mass gap. However, many of its properties can be obtained exactly, like for example its S-matrix [93]. It can be also solved at large N, and indeed its large N solution has been crucial in the determination of the exact S-matrix. Good references on the $1/N$ expansion of the $O(N)$ sigma model are [67, 91].

The $O(N)$ sigma model is a theory of N fields σ^a, $a = 1, \cdots, N$, defined on the unit sphere:

$$\sigma^a \sigma^a = 1. \tag{7.1}$$

There is an $O(N)$ global symmetry, since the σ^a transform in the vector representation of $O(N)$. The action is given by

$$S = \frac{1}{2g^2} \int \mathrm{d}^2 x \, \partial_\mu \sigma^a \partial^\mu \sigma^a = \frac{N}{2t} \int \mathrm{d}^2 x \, \partial_\mu \sigma^a \partial^\mu \sigma^a \tag{7.2}$$

where

$$t = N g^2 \tag{7.3}$$

is the so-called *'t Hooft parameter*. We will take the limit in which N is large, g^2 is small so that t is fixed. The theory described by the classical action (7.2) is a complicated non-linear theory of $N-1$ independent fields, as it can easily seen by solving the constraint (7.1). Perturbatively, we have a theory of $N-1$ massless bosons in two dimensions. These can be regarded as the Goldstone bosons of a Higgs theory with $SO(N)$ symmetry. However, there are no Goldstone bosons in two dimensions (this is a famous theorem of Coleman–Mermin–Wagner), so we don't expect the perturbative picture to be correct. Indeed, the non-perturbative spectrum consists of N *massive* particles in a vector representation of $O(N)$. Their mass m gives the mass gap of the theory. Although these particles are invisible in perturbation theory, they can be seen in a large N solution of the model.

The large N analysis proceeds as follows. First, we renormalize the fields in order to have a canonically normalized kinetic term,

$$\sigma^a \to \sqrt{\frac{t}{N}} \sigma^a. \tag{7.4}$$

We now impose the constraint (7.1) through an extra field α, so we get the action

$$S = \frac{1}{2} \int \mathrm{d}^d x \left\{ \partial_\mu \sigma^a \partial^\mu \sigma^a - \mathrm{i}\alpha \left(\sigma^a \sigma^a - \frac{N}{t} \right) \right\}. \tag{7.5}$$

We want to calculate the generating functional of correlation functions,

$$Z[J] = \int \mathcal{D}\sigma\mathcal{D}\alpha \exp\left\{-S + \int \mathrm{d}^2x\, J^a(x)\sigma^a(x)\right\} \tag{7.6}$$

We can now integrate σ^a and produce an effective action for α

$$Z[J] = \int \mathcal{D}\alpha \exp\left\{-S_{\text{eff}} + \int \mathrm{d}^2x\mathrm{d}^2y\, J^a(x)\left(-\partial^2 - \mathrm{i}\alpha\right)^{-1}(x,y)J^a(y)\right\} \tag{7.7}$$

where

$$S_{\text{eff}}(\alpha) = \frac{N}{2}\mathrm{Tr}\,\log\,\left(-\partial^2 - \mathrm{i}\alpha(x)\right) + \frac{\mathrm{i}N}{2t}\int \mathrm{d}^2x\,\alpha(x). \tag{7.8}$$

and

$$\left(-\partial^2 - \mathrm{i}\alpha\right)^{-1}(x,y) \tag{7.9}$$

denotes the Green function of the operator of $-\partial^2 - \mathrm{i}\alpha(x)$. Notice that in this effective action N plays the role of $1/\hbar$. For large N it makes sense to evaluate the path integral by looking at stationary points of the form

$$\alpha = \text{constant}. \tag{7.10}$$

This is of course what one expects from Lorentz invariance. The EOM for α is obtained from

$$\frac{\delta}{\delta\alpha}\left[\frac{\mathrm{i}N}{g}\int \mathrm{d}^2x\alpha(x) + N\mathrm{Tr}\log\Big(-\partial^2 - \mathrm{i}\alpha\Big)\right] = 0 \tag{7.11}$$

or

$$\frac{1}{t} - \mathrm{Tr}\frac{1}{-\partial_\mu^2 - \mathrm{i}\alpha} = 0 \tag{7.12}$$

Evaluating the trace in momentum space we find

$$\frac{1}{t} - \int \frac{\mathrm{d}^2k}{(2\pi)^2}\frac{1}{k^2 - \mathrm{i}\alpha} = 0. \tag{7.13}$$

This is a divergent integral and we introduce a cutoff Λ for $|k|$. Going to polar variables we have to evaluate

$$\begin{aligned}\int \frac{\mathrm{d}^2k}{(2\pi)^2}\frac{1}{k^2 - \mathrm{i}\alpha} &= \int \frac{\mathrm{d}k}{2\pi}\frac{k}{k^2 - \mathrm{i}\alpha} = \frac{1}{4\pi}\log\Big(k^2 - \mathrm{i}\alpha\Big)\Big|_0^\Lambda \\ &= \frac{1}{4\pi}\log\Big(\frac{\mathrm{i}\Lambda^2}{\alpha} + 1\Big) \approx \frac{1}{4\pi}\log\frac{\Lambda^2}{m^2}, \qquad \Lambda \gg 1.\end{aligned} \tag{7.14}$$

where we have assumed that the solution to the equation (7.13) is of the form

$$\alpha = \mathrm{i}m^2, \qquad m^2 > 0. \tag{7.15}$$

(We will see in a moment that this is indeed the case). We then find the equation

$$\frac{1}{t} - \frac{1}{4\pi}\log\frac{\Lambda^2}{m^2} = 0. \tag{7.16}$$

Since we have an explicit cutoff, we have to renormalize the coupling constant. Let μ be a renormalization scale, and let us introduce the running coupling constant

$$\frac{1}{t(\mu)} = \frac{1}{t} + \frac{1}{4\pi} \log \frac{\mu^2}{\Lambda^2}. \tag{7.17}$$

Then, the above equation reads

$$\frac{1}{t(\mu)} - \frac{1}{4\pi} \log \frac{\mu^2}{m^2} = 0, \tag{7.18}$$

which is solved by

$$m^2 = \mu^2 \mathrm{e}^{-4\pi/t(\mu)}, \tag{7.19}$$

confirming our ansatz (7.15).

This result is very important. First of all, by comparing it to (5.13), we see that it corresponds to the phenomenon of dimensional transmutation in an asymptotically free theory. m^2 is a dynamically generated dimensionful parameter which is the analogue in this model of Λ^2_{QCD}. Indeed, if we define $\alpha_s(\mu)$ and the β function as we did in QCD, we find that

$$\beta_0 = -N \tag{7.20}$$

at large N. Second, if we now plug in the expectation value for α in (7.69), we see that m^2 is indeed *a mass* for the σ^a fields. This is the first dynamical effect that can be obtained at large N.

Once we have found this constant field configuration which gives the main nonperturbative properties of the theory, we expand around this vacuum. For simplicity, we will denote the fluctuation of the α field around its vev (7.15) by α as well:

$$\alpha \to \mathrm{i}m^2 + \alpha. \tag{7.21}$$

It turns out that the natural normalization for the fluctuation α is

$$\alpha \to \alpha/\sqrt{N} \tag{7.22}$$

as we will see. The effective action is now given by

$$\frac{N}{2} \mathrm{Tr} \log \left(-\partial^2 + m^2 - \mathrm{i}\frac{\alpha(x)}{\sqrt{N}} \right), \tag{7.23}$$

which can be written as

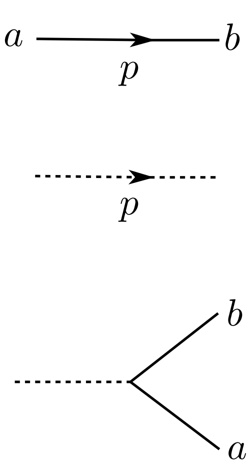
$$\frac{N}{2} \log(-\partial^2 + m^2) + \frac{N}{2} \log\Big[1 + \Delta\Big(-\mathrm{i}\alpha(x)/\sqrt{N} \Big) \Big] \tag{7.24}$$

where

$$\Delta = (-\partial^2 + m^2)^{-1}. \tag{7.25}$$

In these equations, α is regarded as a two-point operator $\alpha(x,y) = \alpha(x)\delta(x-y)$, and

$$(\Delta\alpha)(x,y) = \int \mathrm{d}^2 z\, \Delta(x,z)\alpha(z,y) = \Delta(x,y)\alpha(y). \tag{7.26}$$



$$G_{ab}(p) = \frac{\delta_{ab}}{p^2 + m^2}$$

$$D^{\alpha}(p) = -\frac{2}{f(p)}$$

$$-\frac{\mathrm{i}}{\sqrt{N}}\delta_{ab}$$

Figure 29: The Feynman rules for the $O(N)$ sigma model in the $1/N$ expansion. The dashed lines correspond to the α particle.

We now expand in inverse powers of N. Notice that, by definition of saddle point, there is no linear term in α, therefore at leading order in $1/N$ we have

$$\begin{aligned} -\frac{1}{4}\mathrm{Tr}\,(\Delta\alpha)^2 &= -\frac{1}{4}\int \mathrm{d}^2x\,\mathrm{d}^2y\,(\Delta\alpha)(x,y)(\Delta\alpha)(y,x) \\ &= -\frac{1}{4}\int \mathrm{d}^2x\,\Delta(x,y)\alpha(y)\int \mathrm{d}^2y\,\Delta(y,x)\alpha(x). \end{aligned} \tag{7.27}$$

where

$$\Delta(x,y) = \int \frac{\mathrm{d}^2p}{(2\pi)^2}\frac{\mathrm{e}^{\mathrm{i}p(x-y)}}{p^2+m^2}. \tag{7.28}$$

In terms of Fourier-transformed fields, we can write this as

$$-\frac{}{4}\int \frac{\mathrm{d}^2p}{(2\pi)^2}\tilde{\alpha}(p)\widetilde{\Gamma}^{\mathrm{s}}(p)\tilde{\alpha}(-p), \tag{7.29}$$

where

$$\widetilde{\Gamma}^{s}(p) = \int \frac{\mathrm{d}^2q}{(2\pi)^2}\frac{1}{(q^2+m^2)((p+q)^2+m^2)}. \tag{7.30}$$

The computation of (7.30) is a standard but somewhat long exercise in Feynman calculus. The details can be found in Appendix D. The result is the following:

$$\widetilde{\Gamma}^{\alpha}(p) = f(p) \equiv \frac{1}{2\pi\sqrt{p^2(p^2+4m^2)}}\log\frac{\sqrt{p^2+4m^2}+\sqrt{p^2}}{\sqrt{p^2+4m^2}-\sqrt{p^2}}, \tag{7.31}$$

The calculation of correlation functions in the original theory can now be written as a $1/N$ expansion by using the following rules: we have N massive particles with Green function

$$G_{ab}(p) = \frac{\delta_{ab}}{p^2+m^2} \tag{7.32}$$

as well as an α particle with propagator

$$D^{\alpha}(p) = -\frac{2}{f(p)}. \tag{7.33}$$

These particles interact through a trivalent vertex with

$$-\frac{\mathrm{i}\delta_{ab}}{\sqrt{N}}. \tag{7.34}$$

Therefore, interactions at suppressed at large N, and $1/\sqrt{N}$ becomes the effective coupling constant of the theory. The Feynman rules for this theory are summarized in Fig. 29.

7.2 The $\mathbb{P}^{N-1}$ sigma model

Our second example of the $1/N$ expansion, and of the type of nonperturbative methods associated to resumming an infinite number of diagrams, will be the another two-dimensional toy model: the $\mathbb{P}^{N-1}$ sigma model [30, 86]. In particular, we will be able to obtain a purely nonperturbative result: a nonzero value for a two-dimensional analogue of the topological susceptibility.

7.2.1 The model and its instantons

The basic field of the $\mathbb{P}^{N-1}$ sigma model is an N-component complex vector of norm 1, defined on a two-dimensional spacetime:

$$z_1(x), \cdots, z_N(x), \qquad \sum_{i=1}^{N} |z_i|^2 = 1. \tag{7.35}$$

There is also a $U(1)$ gauge symmetry

$$z_i \to \mathrm{e}^{\mathrm{i}\alpha(x)} z_i \tag{7.36}$$

We can cook up a gauge field out of the z_i, since the composite field

$$A_\mu = \frac{\mathrm{i}}{2}\Big(\bar{z}_i \partial_\mu z_i - (\partial_\mu z_i)\bar{z}_i\Big), \tag{7.37}$$

which is real, transforms as

$$A_\mu \to A_\mu - \partial_\mu \alpha(x). \tag{7.38}$$

Let us check this:

$$\partial_\mu z_i \to \mathrm{e}^{\mathrm{i}\alpha(x)}\Big(\mathrm{i}\partial_\mu \alpha + \partial_\mu z_i\Big) \tag{7.39}$$

Therefore

$$\bar{z}_i \partial_\mu z_i \to \bar{z}_i \partial_\mu z_i + \mathrm{i}\partial_\mu \alpha \bar{z}_i z_i = \bar{z}_i \partial_\mu z_i + \mathrm{i}\partial_\mu \alpha. \tag{7.40}$$

The dynamics of this field is described by the gauge invariant action

$$S = \frac{1}{g^2}\int \mathrm{d}^2 x\, \overline{D_\mu z} D^\mu z, \qquad D_\mu = \partial_\mu + \mathrm{i}A_\mu. \tag{7.41}$$

This action defines the $\mathbb{P}^{N-1}$ sigma model. Notice that the gauge field is in fact an auxiliary field, since it does not have a kinetic term. If we expand the Lagrangian of (7.41)

$$\mathcal{L} = \overline{D_\mu z} D^\mu z \tag{7.42}$$

we find

$$\mathcal{L} = \partial^\mu \bar{z}_i \partial_\mu z_i - \mathrm{i} A_\mu \bar{z}_i \partial^\mu z_i + \mathrm{i} A_\mu \partial^\mu \bar{z}_i z_i + A_\mu A^\mu \bar{z}_i z_i \tag{7.43}$$

which is

$$\mathcal{L} = \partial^\mu \bar{z}_i \partial_\mu z_i + A_\mu^2 - A_\mu \mathrm{i}\Big(\bar{z}_i \partial^\mu z_i - (\partial^\mu \bar{z}_i) z_i\Big). \tag{7.44}$$

The classical EOM for A_μ gives precisely the definition (7.37). Notice that

$$z_i \bar{z}_i = 1 \Rightarrow (\partial^\mu \bar{z}_i) z_i + \bar{z}_i \partial^\mu z_i = 0 \tag{7.45}$$

therefore we can write

$$A^\mu = \mathrm{i} \bar{z}_i \partial^\mu z_i = -\mathrm{i}(\partial^\mu \bar{z}_i) z_i, \tag{7.46}$$

and

$$\mathcal{L} = \partial^\mu \bar{z}_i \partial_\mu z_i - A_\mu^2 = \partial^\mu \bar{z}_i \partial_\mu z_i - (\bar{z}_i \partial^\mu z_i)(z_j \partial_\mu \bar{z}_j) \tag{7.47}$$

or equivalently

$$\mathcal{L} = \partial^\mu \bar{z}_i \partial_\mu z_i + (\bar{z}_i \partial^\mu z_i)(\bar{z}_j \partial_\mu z_j). \tag{7.48}$$

It is interesting to point out that the $\mathbb{P}^{N-1}$ model has instanton solutions, similar in many respects to the instantons of Yang–Mills theory. Classical aspects of instantons in the $\mathbb{P}^{N-1}$ model are discussed in the original paper [30] as well as in section 4.5 of [70]. These instantons are topologically nontrivial configurations with finite action. Notice that finite action means here that

$$D_\mu z_i = 0, \quad \text{at } |x| \to \infty, \quad i = 1, \cdots, n, \tag{7.49}$$

therefore, at infinity, z_i is covariantly constant, i.e. it must be a constant vector up to a phase. We write

$$z_i = n_i \mathrm{e}^{\mathrm{i}\sigma(x)}, \quad |x| \to \infty, \qquad n_i \bar{n}^i = 1. \tag{7.50}$$

This can be seen in detail by spelling out the condition (7.49). It means that

$$-\mathrm{i} A_\mu = \frac{\partial_\mu z_i}{z_i} = \frac{\partial |z_i|}{|z_i|} + \mathrm{i} \partial_\mu \phi_i, \tag{7.51}$$

where ϕ_i is the phase of z_i. Since $\mathrm{i} A_\mu$ is pure imaginary and indepedent of the index i, we deduce that, at infinity,

$$\partial_\mu |z_i| = 0, \qquad \phi_i = \sigma(\theta), \quad i = 1, \cdots, N, \tag{7.52}$$

which is precisely (7.50).

The topological charge classifying instantons is given by

$$Q = \frac{1}{2\pi} \int \mathrm{d}^2 x \, \epsilon_{\mu\nu} \partial_\mu A_\nu. \tag{7.53}$$

Since

$$\epsilon_{\mu\nu} \partial_\mu A_\nu = \mathrm{i} \epsilon_{\mu\nu} \partial_\mu \bar{z}_i \partial_\nu z_i \tag{7.54}$$

this can be rewritten as

$$Q = \frac{1}{2\pi \mathrm{i}} \int \mathrm{d}^2 x \, \epsilon_{\mu\nu} \partial_\nu (\bar{z}_i \partial_\mu z_i). \tag{7.55}$$

In order to be able to talk about instantons, we have to show that the topological charge (7.53) is quantized. We follow the discussion in [86]. Using Stokes theorem, we can write (7.55) as an integral at the boundary, i.e. at infinity

$$Q = \frac{1}{2\pi \mathrm{i}} \oint \mathrm{d}x^\mu \, \bar{z}_i \partial_\mu z_i. \tag{7.56}$$

Plugging in here the boundary behavior (7.50), we obtain

$$Q = \frac{1}{2\pi} \oint \mathrm{d}x^\mu \frac{\partial \sigma}{\partial x^\mu} = \frac{1}{2\pi} \Delta\sigma, \tag{7.57}$$

where $\Delta\sigma$ is just the change of σ as we go around a circle at infinity. Since a phase is defined up to an integer multiple of 2π, it is clear that $\Delta\sigma$ is quantized.

Another important property of instantons is that they minimize the action in their topological sector. To see that this also holds in this model, let us write the topological density as

$$q(x) = \frac{1}{2\pi} \epsilon_{\mu\nu} \partial_\mu A_\nu = \frac{\mathrm{i}}{2\pi} \epsilon_{\mu\nu} \overline{D}_\mu z \cdot D_\nu z. \tag{7.58}$$

To see this, notice that the last term equals

$$\frac{\mathrm{i}}{2\pi} \epsilon_{\mu\nu} (\partial_\mu \bar{z}_i - \mathrm{i} A_\mu \bar{z}_i)(\partial_\nu z + \mathrm{i} A_\nu z_i) \tag{7.59}$$

and due to antisymmetry of $\epsilon_{\mu\nu}$ we only have to check that

$$-\mathrm{i}\epsilon_{\mu\nu} (A_\mu \bar{z}_i \partial_\nu z_i - A_\nu z_i \partial_\mu \bar{z}_i) \tag{7.60}$$

vanishes. Using (7.45) we can write it as

$$-\mathrm{i}\epsilon_{\mu\nu} (A_\mu \bar{z}_i \partial_\nu z_i + A_\nu \bar{z}_i \partial_\mu z_i) = 0, \tag{7.61}$$

therefore the topological charge can be written as

$$Q = \frac{\mathrm{i}}{2\pi} \int \mathrm{d}^2 x \, \epsilon_{\mu\nu} \overline{D}_\mu z \cdot D_\nu z. \tag{7.62}$$

From the obvious inequality

$$\left| D_\mu z \mp \mathrm{i}\epsilon_{\mu\nu} \overline{D_\nu z} \right|^2 \geq 0 \tag{7.63}$$

we find

$$\overline{D_\mu z} \cdot D_\mu z + \epsilon_{\mu\rho}\epsilon_{\mu\sigma} \overline{D_\rho z} \cdot D_\sigma z \mp 2\mathrm{i}\epsilon_{\mu\nu} \overline{D}_\mu z \cdot D_\nu z \geq 0, \tag{7.64}$$

and since $\epsilon_{\mu\rho}\epsilon_{\mu\sigma} = \delta_{\rho\sigma}$ we get at the end of the day

$$\overline{D_\mu z} \cdot D_\mu z \geq \mathrm{i}\epsilon_{\mu\nu} \overline{D_\mu z} \cdot D_\nu z, \tag{7.65}$$

after integration one finds,

$$\frac{1}{g^2} \int \mathrm{d}^2 x \, \overline{D_\mu z} \cdot D_\mu z \geq \frac{\mathrm{i}}{g^2} \int \mathrm{d}^2 x \, \epsilon_{\mu\nu} \overline{D_\mu z} \cdot D_\nu z, \tag{7.66}$$

i.e.

$$S \geq \frac{2\pi}{g^2} |Q|. \tag{7.67}$$

This is the typical BPS bound. Equality holds only if the bound is saturated, and from here we derive the equation describing instanton configurations in this model:

$$D_\mu z \mp \mathrm{i}\epsilon_{\mu\nu}\overline{D}_\nu z = 0. \tag{7.68}$$

The $\pm$ signs give instanton and anti-instanton solutions respectively. These are the analogues of the (anti) self-duality conditions for instantons in QCD.

7.2.2 The effective action at large N

The (Euclidean) action of the $\mathbb{P}^{N-1}$ model is given by

$$S = \int \mathrm{d}^2x \left[\frac{1}{g^2}\overline{D_\mu z}\cdot D_\mu z - \frac{\mathrm{i}\lambda}{g^2}(z_i\bar{z}_i - 1) + \frac{\mathrm{i}\theta}{2\pi}\epsilon_{\mu\nu}\partial_\mu A_\nu\right] \tag{7.69}$$

where we have introduced a Lagrange multiplier λ to impose the constraint (7.35), as well as the analogue of a theta term. We define the 't Hooft parameter as in (7.3), and we take again the limit in which N is large, g^2 is small so that t is fixed. In (7.69) we treat A_μ and λ as auxiliary fields. When we integrate them out we obtain the action for the fields z_i together with the constraint (7.35). But since the action is quadratic in z_i, of the form

$$\int \mathrm{d}^2x\bar{z}_i\Delta z_i, \tag{7.70}$$

where

$$\Delta = -\frac{N}{t}D_\mu D^\mu - \frac{N\mathrm{i}\lambda}{t}, \tag{7.71}$$

we can integrate out the N bosonic, complex variables z_i. Each of them gives a factor

$$\frac{1}{\det\Delta} \tag{7.72}$$

and since we have N of them, we obtain, after writing the determinant as the exponential of a trace of a log,

$$\exp\left[-N\mathrm{Tr}\log\Big(-(\partial_\mu + \mathrm{i}A_\mu)^2 - \mathrm{i}\lambda\Big)\right]. \tag{7.73}$$

This leads to the effective action

$$S_{\text{eff}} = N\mathrm{Tr}\log\Big(-(\partial_\mu + \mathrm{i}A_\mu)^2 - \mathrm{i}\lambda\Big) + \frac{\mathrm{i}N\lambda}{t} - \frac{\mathrm{i}\theta}{2\pi}\epsilon_{\mu\nu}\partial_\mu A_\nu \tag{7.74}$$

which depends on the fields A_μ and λ. We will often Fourier-transform the fields as

$$\tilde{\lambda}(p) = \int \mathrm{d}^2x\, \mathrm{e}^{-\mathrm{i}px}\lambda(x). \tag{7.75}$$

Notice that in this effective action N plays the role of $1/\hbar$. For large N it makes sense to evaluate the path integral by looking at stationary points of the form

$$A_\mu = 0, \qquad \lambda = \text{constant}. \tag{7.76}$$

This is of course what one expects from Lorentz invariance. The EOM for λ is obtained from

$$\frac{\delta}{\delta\lambda}\left[\frac{\mathrm{i}N}{t}\int \mathrm{d}^2x\lambda + N\mathrm{Tr}\log\Big(-(\partial_\mu + \mathrm{i}A_\mu)^2 - \mathrm{i}\lambda\Big)\right] = 0 \tag{7.77}$$

and it is identical to the one we obtained for α in the $O(N)$ sigma model. The solution is also of the form

$$\lambda = \mathrm{i}m^2, \qquad m^2 > 0, \tag{7.78}$$

where m^2 is the dynamically generated scale of the theory. There is a running coupling constant $t(\mu)$ which satisfies the same RG equation (7.19) of the $O(N)$ sigma model. In particular, we find that the $\mathbb{P}^{N-1}$ model is also asymptotically free, and the expectation value for λ in (7.69) gives again a mass for the z_i fields.

For simplicity, we will denote the fluctuation of the λ field around its vev (7.78) by λ as well. It turns out that the natural normalizations for A_μ and the fluctuation λ are

$$A_\mu \to \frac{1}{\sqrt{N}} A_\mu, \qquad \lambda \to \lambda/\sqrt{N}. \tag{7.79}$$

We write

$$N \mathrm{Tr} \log\Big(-(\partial_\mu + \mathrm{i}A_\mu/\sqrt{N})^2 + m^2 - \mathrm{i}\frac{\lambda}{\sqrt{N}}\Big) \tag{7.80}$$

as

$$N \mathrm{Tr} \log(-\partial^2 + m^2) + N \mathrm{Tr} \log\Big[1 + \Delta\Big(A^2/N - \mathrm{i}\lambda/\sqrt{N} - \mathrm{i}\{A, \partial\}/\sqrt{N}\Big)\Big] \tag{7.81}$$

where Δ is given in (7.25). If we expand in inverse powers of N, we find, schematically, and at leading order,

$$\Delta A^2 + \frac{1}{2}\Delta^2 (\partial A + 2A\partial)^2 + \frac{1}{2}(\Delta\lambda)^2 . \tag{7.82}$$

The last term has been computed for the $O(N)$ sigma model, and it leads to

$$\frac{1}{2}\int \frac{\mathrm{d}^2 p}{(2\pi)^2} \tilde{\lambda}(p) \widetilde{\Gamma}^{\mathrm{s}}(p) \tilde{\lambda}(-p), \tag{7.83}$$

where $\widetilde{\Gamma}^{\mathrm{s}}(p)$ is given in (7.30). For the quadratic term in the A_μ fields, we find,

$$\frac{1}{2}\int \frac{\mathrm{d}^2 p}{(2\pi)^2} \tilde{A}^\mu(p) \widetilde{\Gamma}^A_{\mu\nu}(p) \tilde{A}^\nu(-p), \tag{7.84}$$

where

$$\widetilde{\Gamma}^A_{\mu\nu}(p) = 2\delta_{\mu\nu}\int \frac{\mathrm{d}^2 q}{(2\pi)^2}\frac{1}{(q^2+m^2)} - \int \frac{\mathrm{d}^2 q}{(2\pi)^2}\frac{(p_\mu + 2q_\mu)(p_\nu + 2q_\nu)}{(q^2+m^2)((p+q)^2+m^2)}. \tag{7.85}$$

The computation of (7.85) can be also found in Appendix D, and gives

$$\widetilde{\Gamma}^A_{\mu\nu}(p) = \Big(\delta_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\Big)\Big\{(p^2 + 4m^2) f(p) - \frac{1}{\pi}\Big\}, \tag{7.86}$$

where $f(p)$ is given in (7.31). Since

$$f(p) = \frac{1}{4\pi m^2} - \frac{p^2}{24\pi m^4} + \mathcal{O}(p^4) \tag{7.87}$$

near $p^2 = 0$, the quadratic term in $\tilde{A}$ is of the form

$$(\delta_{\mu\nu} p^2 - p_\mu p_\nu)(c + \mathcal{O}(p^2)) \tag{7.88}$$

where

$$c = \frac{1}{12\pi m^2}. \tag{7.89}$$

This structure is a consequence of gauge invariance, and leads to the standard gauge field kinetic energy

$$(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 \tag{7.90}$$

written in momentum space. In other words, *the quantum corrections have generated a kinetic energy for A_μ*. This is the second dynamical effect that can be seen at large N, and it can be seen that is obtained in the original z variables after resumming an infinite number of conventional diagrams –those which dominate at large N.

The excitations associated to the fields z_i and $\bar{z}_i$ can be regarded as quarks and antiquarks of the model. These particles will interact through the gauge field A_μ. But a $U(1)$ gauge field in two dimensions is actually confining, since Coulomb's law in two dimensions leads to a linear potential. Therefore, an extra consequence of the emergence of a dynamical gauge field in this model is confinement of charges, which can only appear as singlets or triplets.

7.2.3 Topological susceptibility at large N

Another truly nonperturbative effect that can be seen at large N is a nonzero value for the topological susceptibility. Remember from the discussion in the context of YM theory that χ_t is given by the limit (5.27). We will then compute

$$U(p) = \int \mathrm{d}^2x\, \mathrm{e}^{\mathrm{i}px} \langle q(x) q(0)\rangle = \int \frac{\mathrm{d}^2 p'}{(2\pi)^2} \langle \tilde{q}(-p)\tilde{q}(p')\rangle \tag{7.91}$$

where $q(x)$ is the topological density defined in (7.58). This quantity has now a factor $1/N$ which comes from the normalization of A_μ. The Fourier transform of $q(p)$ is given by

$$\tilde{q}(p) = -\frac{\mathrm{i}}{2\pi\sqrt{N}} \epsilon_{\mu\nu} p_\mu \tilde{A}_\nu. \tag{7.92}$$

Therefore,

$$\langle \tilde{q}(-p)\tilde{q}(p')\rangle = \frac{1}{4\pi^2 N} \epsilon_{\mu\nu}\epsilon_{\rho\sigma} p_\mu p'_\rho \langle \tilde{A}_\nu(-p)\tilde{A}_\sigma(p')\rangle. \tag{7.93}$$

To calculate the two-point function of the gauge field we first choose the Lorentz gauge

$$\partial_\mu A_\mu = 0. \tag{7.94}$$

In this gauge the two-point function can be immediately deduced from (7.86), and one finds

$$\langle \tilde{A}_\nu(p)\tilde{A}_\sigma(-p')\rangle = (2\pi)^2 \delta(p-p') \Big(\delta_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\Big) D_A(p), \tag{7.95}$$

where

$$D_A(p) = \Big\{ (p^2 + 4m^2) f(p) - \frac{1}{\pi} \Big\}^{-1}. \tag{7.96}$$

The $(2\pi)^2$ factor in (7.95) comes from the kinetic term (7.84) in momentum space. Since

$$\epsilon_{\mu\nu}\epsilon_{\rho\sigma} p_\mu p_\rho \Big(\delta_{\nu\sigma} - \frac{p_\nu p_\sigma}{p^2}\Big) = (\delta_{\mu\rho}\delta_{\nu\sigma} - \delta_{\mu\sigma}\delta_{\nu\rho}) p_\mu p_\rho \Big(\delta_{\nu\sigma} - \frac{p_\nu p_\sigma}{p^2}\Big) = p^2 \tag{7.97}$$

we find

$$\langle \tilde{q}(-p)\tilde{q}(p')\rangle = \frac{p^2}{4\pi^2 N}(2\pi)^2 D_A(p)\delta(p-p'). \tag{7.98}$$

Therefore,

$$\int \frac{\mathrm{d}^2 p'}{(2\pi)^2}\langle \tilde{q}(p)\tilde{q}(p')\rangle = \frac{p^2}{N}D_A(p) = \frac{3m^2}{\pi N} + \mathcal{O}(p^2), \tag{7.99}$$

and the topological susceptibility reads, at leading order in the $1/N$ expansion,

$$\chi_t^{\text{large } N} = \frac{3m^2}{\pi N}. \tag{7.100}$$

This is a rather remarkable result, since this quantity vanishes order by order in perturbation theory, as we saw in the context of Yang–Mills theory in (5.29). The reasons that we have not obtained a vanishing result is because we have resummed an *infinite* number of diagrams (those dominating at large N) *before* taking the $p \to 0$ limit. That this can be the case was already mentioned in the introduction, following the argument by Witten in [87].

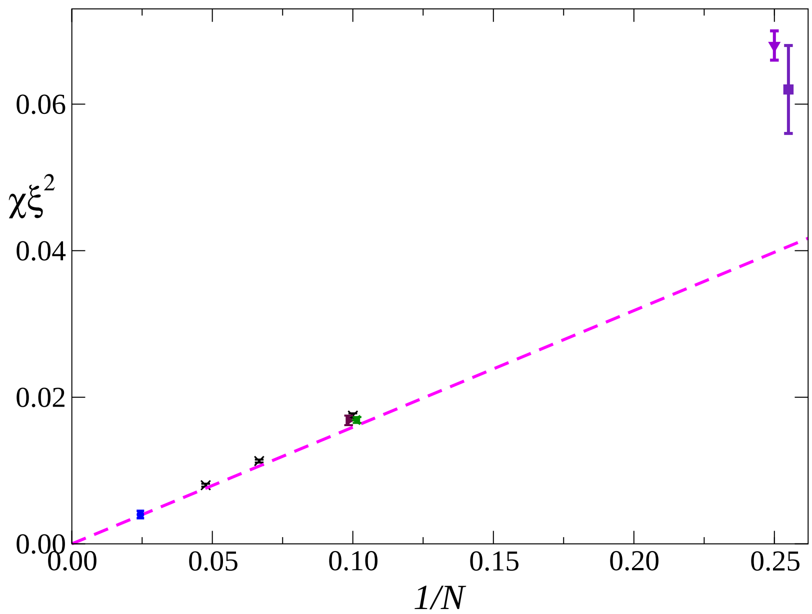


Figure 30: The data points give the results for $\chi_t \xi^2$ in the $\mathbb{CP}^{N-1}$ model, calculated in the lattice for various values of N, and represented as a function of $1/N$. The quantity $\xi^2 = (6m^2)^{-1}$ sets the length scale. The diagonal line, with slope $(2\pi)^{-1}$, is the the large N result (7.100). This figure courtesy of [82].

The result (7.100) has been tested in lattice calculations of the topological susceptibility, see the figure Fig. 30 extracted from [82]. As we can see, the calculations agree very well with the large N result for $N \geq 10$. Other recent lattice calculations of this quantity can be found in [57].

We can now summarize the nonperturbative effects obtained for this model at large N, i.e. by resumming an infinite number of conventional Feynman diagrams.

1. A mass is generated for the quarks and antiquarks z_i, $\bar{z}_i$. This mass is invisible in perturbation theory in the coupling constant (and in the 't Hooft parameter).
2. The field A_μ, which started its life as an auxiliary variable, becomes a dynamic gauge field which leads to quark confinement.
3. The topological susceptibility is nonzero, and of order $\mathcal{O}(1/N)$.

Not all of these features are shared by other field theories at large N, but the appearance of a nontrivial topological susceptibility of order $\mathcal{O}(1/N)$ will also appear in QCD.

8. The $1/N$ expansion in QCD

The $1/N$ expansion in QCD was introduced by 't Hooft in [77]. Classic reviews of this topic are [26, 88]. A more modern reference is [61].

8.1 Fatgraphs

We will write down the QCD Lagrangian (5.7) as

$$\mathcal{L} = \frac{N}{t}\left[\frac{1}{4}(F_{\mu\nu}, F^{\mu\nu}) + \sum_{f=1}^{N_f} \bar{q}_f \left(i\not{D} - m_f\right) q_f\right] \tag{8.1}$$

where we have introduced the *'t Hooft parameter* as

$$t = g^2 N. \tag{8.2}$$

The large N limit is defined as

$$N \to \infty, \quad g^2 \to 0, \qquad t \text{ fixed}. \tag{8.3}$$

In this way the theory is still nontrivial. A first indication of this is the one-loop β function of QCD, (5.11)–(5.12), which can be written as

$$\mu \frac{\mathrm{d}g}{\mathrm{d}\mu} = -\left(\frac{11}{3}N - \frac{2}{3}N_f\right)\frac{g^3}{16\pi^2}, \tag{8.4}$$

and becomes, after multiplying by $N^{\frac{1}{2}}$,

$$\mu \frac{\mathrm{d}t}{\mathrm{d}\mu} = -\left(\frac{11}{3}N^{\frac{3}{2}} - \frac{2}{3}N^{\frac{1}{2}}N_f\right)\frac{t^3/N^{\frac{3}{2}}}{16\pi^2} = -\left(\frac{11}{3} - \frac{2}{3}\frac{N_f}{N}\right)\frac{t^3}{16\pi^2}, \tag{8.5}$$

so it is well-defined in the large N limit. We also see that the effect of the quark loops (which gives the contribution proportional to N_f) is suppressed, and this will be explained diagramatically in what follows. We will also see that all interesting quantities in QCD have an expansion in powers of $1/N$, and the large N limit (8.3) keeps the leading term (which, for reasons that will become clear in a moment, is called the *planar part*). We will be also interested in the $1/N$ corrections to this limit.

We note for future use that the rescaling (5.9) reads, in terms of the 't Hooft parameter,

$$A_\mu = \frac{t}{\sqrt{N}}\hat{A}_\mu, \qquad q = \frac{t}{\sqrt{N}}\hat{q}. \tag{8.6}$$

The key idea in the $1/N$ expansion is that in $SU(N)$ gauge theories there is, in addition to the coupling constants appearing in the model (like for example g), a hidden variable, namely N, the rank of the gauge group. The N dependence in the perturbative expansion comes from the *group factors* associated to Feynman diagrams, and in general a single Feynman diagram gives rise to a polynomial in N involving different powers of N. Therefore, the standard Feynman diagrams, which are good in order to keep track of powers of the coupling constants, are not good in order to keep track of powers of N. If we want to keep track of the N dependence we have to "split" each diagram into different pieces which correspond to a definite power of N. To do that, one writes the Feynman diagrams of the theory as "fatgraphs" or double line graphs, as first indicated by 't Hooft [79]. Let us see how this works.

Figure 31: The quark propagator.

The quark propagator is

$$\langle \psi^i (x) \bar{\psi}^j (y) \rangle = \frac{t}{N} \delta^{ij} S(x-y), \qquad i,j = 1, \cdots, N. \tag{8.7}$$

This is represented diagrammatically by a single line, and the color at the beginning of the line is the same as at the end of the line, because of the δ^{ij} in eq. (8.7), see Fig. 31. The gluon propagator is

$$\langle A^a_\mu (x) A^b_\nu (y) \rangle = \frac{t}{N} \delta^{ab} D_{\mu\nu} (x-y), \tag{8.8}$$

where a and b are indices in the adjoint representation. Instead of treating a gluon as a field with a single adjoint index, it is preferable to treat it as an $N \times N$ matrix with two indices in the N and $\overline{N}$ representations, i.e.

$$(A_\mu)^i{}_j = A^a_\mu (T_a)^i{}_j \tag{8.9}$$

Here, $(T_a)^i{}_j$ is a basis of the Lie algebra which satisfies the normalization condition

$$\mathrm{Tr}\,(T_a T_b) = \delta_{ab}, \qquad a,b = 1, \cdots, N^2. \tag{8.10}$$

They also satisfy,

$$\sum_a (T_a)^i{}_j (T_a)^k{}_l = \delta^i_l \, \delta^k_j \tag{8.11}$$

for $U(N)$, and

$$\sum_a (T_a)^i{}_j (T_a)^k{}_l = \delta^i_l \, \delta^k_j - \frac{1}{N} \delta^i_j \, \delta^k_l \tag{8.12}$$

for $SU(N)$. Therefore, we can rewrite the $U(N)$ gluon propagator as

$$\langle A^i_{\mu j} (x) A^k_{\nu l} (y) \rangle = \frac{t}{N} D_{\mu\nu} (x-y) \; \delta^i_l \, \delta^k_j. \tag{8.13}$$

The group structure of this propagator can be represented by a *double line*, as in Fig. 32.

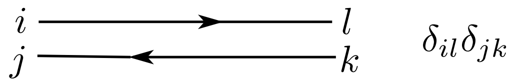


Figure 32: The gluon propagator in the double line notation.

We can also write the interaction vertices in the double line notation. The three-gluon vertex involves the structure constants f_{abc} of the Lie algebra, which are defined by

$$[T_a, T_b] = f_{abc} T_c. \tag{8.14}$$

By multiplying by T_d and taking a trace, we find the relation

$$f_{abc} = \mathrm{Tr}\,(T_a T_b T_c) - \mathrm{Tr}\,(T_b T_a T_c). \tag{8.15}$$

The trace of three generators of the Lie algebra can be interpreted as a cubic vertex. Indeed, it comes from

$$\mathrm{Tr}(A_\mu A_\nu A_\rho) = A^a_\mu A^b_\nu A^c_\rho \, \mathrm{Tr}(T_a T_b T_c) \tag{8.16}$$

but in the double line notation it leads to the index structure

$$\sum_{i,j,k} (A_\mu)^i{}_j \, (A_\nu)^j{}_k \, (A_\rho)^k{}_i \tag{8.17}$$

which can be depicted as in Fig. 33. Since we have a commutator, we get an additional term

$$-\sum_{i,j,k} (A_\nu)^i{}_j \, (A_\mu)^j{}_k \, (A_\rho)^k{}_i, \tag{8.18}$$

which can be also represented as double-line vertex. Notice however that it is twisted in comparison to the previous one, see Fig. 34.

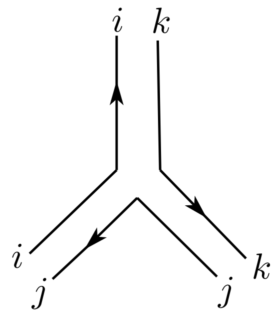


Figure 33: The cubic vertex (8.17) in the double line notation.

Our final rule concerns the quark-gluon vertex describing the interaction between a quark bilinear and a gluon. The group structure of this vertex

$$\psi^i(x) \, (A_\mu)^j{}_i \, (x) \overline{\psi}^j(x), \tag{8.19}$$

therefore it can be represented, in the double line notation, as shown in Fig. 35.

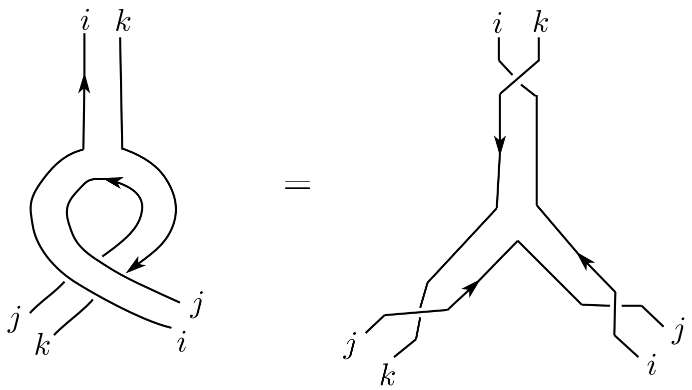


Figure 34: The twisted vertex (8.18) in the double line notation.

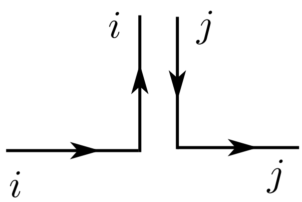


Figure 35: The interaction vertex between a quark (vertical line) and a gluon (horizontal double line), in the double line notation.

In general, fatgraphs (which have no external lines) are characterized topologically by the number of propagators or edges E, the number of vertices V, and the number of closed loops h. By (8.13), each propagator gives a factor of g, while each interaction vertex gives a power of g. Finally, each closed loop involves a sum over a color index and gives a factor of h. Therefore, we have a total factor

$$N^h g^{2(E-V)}, \tag{8.20}$$

but in terms of the ‘t Hooft parameter this is

$$N^{V-E+h} t^{V-E}. \tag{8.21}$$

We can now regard each fatgraph as a Riemann surface which will be *closed* in the absence of quarks loops. To see this, we think about each closed loop as the perimeter of a polygon. A double-line is then interpreted as an instruction to glue polygons: we identify one edge of a polygon with one edge of another polygon if they both lie on the same double line. Finally, each closed quark (single-line) loop is interpreted as a boundary for the surface. With this interpretation, we can use Euler’s relation to write

$$h + V - E = \chi = 2 - 2g - b \tag{8.22}$$

where g is the genus of the Riemann surface and b the number of boundaries. Therefore the factor of N in (8.20) is

$$N^{2-2g-b} = N^{\chi} \tag{8.23}$$

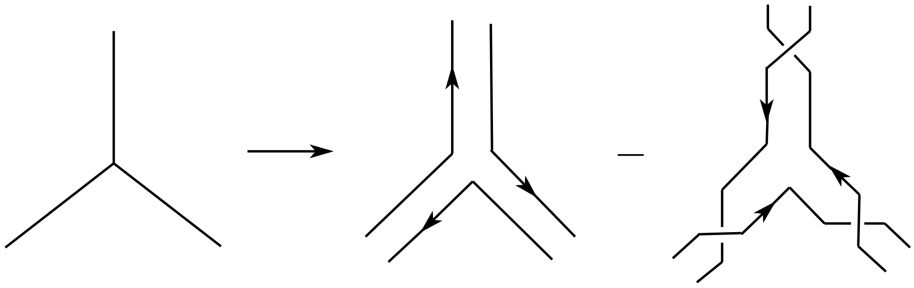


Figure 36: The standard cubic vertex of QCD becomes a sum of two fatgraphs.

The fatgraphs with $g = 0$ are called *planar*, while the ones with $g > 0$ are called *nonplanar*. It is easy to see that each conventional Feynman diagram gives rise to many different fatgraphs with different genera.

We can now formalize the procedure to compute the group factor of any diagram in QCD. Notice that, from the point of view of the group theory structure, the quartic vertex of Yang–Mills can be reduced to two cubic vertices joined by an extra edge, therefore any diagram will be written in the end in terms of trivalent diagrams. Given a trivalent diagram G, with V vertices, we use (8.15) to get a sum over the 2^V possible "resolutions" of the vertices. This is represented graphically in Fig. 36. Each of these diagrams will be a fatgraph Σ, which we will weight by N to the power $h(\Sigma)$, the number of closed loops. We then have

$$r(G) = \sum_{\Sigma} N^{h(\Sigma)}. \tag{8.24}$$

The contribution of such a diagram to the large N expansion will include in addition a factor $g^{2(E-V)}$ (again, from the power counting point of view we can treat a quartic vertex as a two cubic vertices joined by an extra edge, since $E - V$ remains invariant).

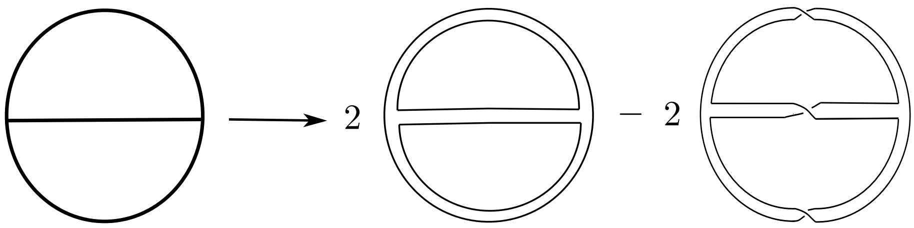


Figure 37: The two fatgraphs associated to the theta diagram. The first one has $g = 0$, while the second one has $g = 1$.

Example 8.1. Let us consider the simplest two-loop graph made out of cubic vertices, the so-called theta diagram. After "resolving" the vertices according to (8.15), we find two different graphs, as shown in Fig. 37: an "untwisted" graph with multiplicity 2, and a "twisted" graph, also with multiplicity 2. Therefore, its group factor

$$2N^3 - 2N. \tag{8.25}$$

The weight of the first fatgraph (with $g = 0$) in the $1/N$ expansion is

$$N^3 g^2 = N^2 t^2, \tag{8.26}$$

while the second one of $g = 1$ has the weight

$$Ng^2 = t. \tag{8.27}$$

Example 8.2. In Fig. 38 we show a diagram with a quark line and one of the fatgraphs that it generates. This graph has three closed cycles, four vertices and six propagators, and so its weight is

$$N^3 g^4 = Nt^2 \tag{8.28}$$

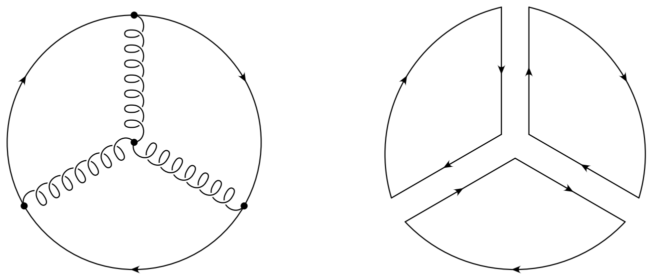


Figure 38: A typical QCD diagram with one closed quark line (left) and one of the fatgraphs it generates (right). This figure courtesy of [61].

8.2 Large N rules for correlation functions

We can now use the diagrammatic representation in terms of fatgraphs to analyze the large N counting rules of correlation functions (of course, all of our conclusions will be also valid for any quantum theory with a $U(N)$ symmetry with fields in the adjoint and the fundamental).

We have seen that, when we reorganize the perturbative expansion in terms of fatgraphs, the Feynman diagrams become two-dimensional surfaces labelled by two topological quantities: the genus g and the number of boundaries, with a weight (8.23). The largest values of χ are 2 in the case of closed surfaces, corresponding to $g = 0$, and $\chi = 1$ for the surfaces with boundaries, corresponding to $g = 0$ and $b = 1$. It follows immediately that

1. The leading connected vacuum-to-vacuum graphs are of order N^2. They are planar graphs made out of gluons.
2. The leading connected vacuum-to-vacuum graphs with quark lines are of order N. They are planar graphs with only one quark loop forming the boundary of the graph.

In particular, we deduce that the free energy of a pure $U(N)$ gauge theory (which is given by the sum over all connected, vacuum-to-vacuum digrams) is given by a sum of the form

$$F(N,t) = \sum_{g=0}^{\infty} F_g(t) N^{2-2g}, \tag{8.29}$$

where

$$F_g(t) = \sum_{h \geq 0} a_{g,h} t^{2g-2+h} \tag{8.30}$$

is a sum over all fatgraphs with a fixed topology. In the large N limit (8.3), only the planar diagrams $g = 0$ survive.

We can now study correlation functions. Let G_i be a gauge-invariant operator made out of gluons only. Examples of such operators are

$$\mathrm{Tr}\, F_{\mu\nu}F^{\mu\nu}, \qquad \mathrm{Tr}_R U_\gamma, \tag{8.31}$$

where

$$U_\gamma = \mathrm{P}\,\exp \oint_\gamma A \tag{8.32}$$

is a Wilson line operator around the closed loop γ. We add to the action

$$S \to S + N\sum_i J_i G_i \tag{8.33}$$

where J_i are sources. Due to the overall factor of N, the counting rules for the new Lagrangian are the same as before. On the other hand, we know that the sum of connected vacuum-to-vacuum graphs with these sources is a generating functional of *connected* correlation functions. We then conclude,

$$\langle G_1 \cdots G_r \rangle^{(c)} = \frac{1}{N^r}\frac{\partial^r \Gamma(J)}{\partial J_1 \cdots \partial J_r}\Big|_{J=0}. \tag{8.34}$$

Since the leading contribution to this generating functional is again of order N^2, we conclude that

$$\langle G_1 \cdots G_r \rangle^{(c)} \sim N^{2-r} \tag{8.35}$$

at leading order in N. If we consider the full $1/N$ expansion of this correlation function, we will obtain an expansion of the form

$$W^{(r)}(N,t) = \langle G_1 \cdots G_r \rangle^{(c)} = \sum_{g=0}^{\infty} W_g^{(r)}(t) N^{2-2g-r} \tag{8.36}$$

where

$$W_g^{(r)}(t) = \sum_{n \geq 0} W_{n,g}^{(r)} t^n \tag{8.37}$$

is the sum over all fatgraphs contributing to the correlation function and with a fixed topology.

Similarly, we can consider gauge-invariant operators M_i involving quark bilinears, like

$$\bar\psi\psi, \quad \bar\psi(y)\mathrm{P}\,\mathrm{e}^{\int_x^y A}\psi(x), \tag{8.38}$$

ans so on. We now perturb the action as

$$S \to S + N\sum_i J_i M_i \tag{8.39}$$

where b_i are sources, and

$$\langle M_1 \cdots M_r \rangle^{(c)} = \frac{1}{N^r}\frac{\partial^r \Gamma(b)}{\partial J_1 \cdots \partial J_r}\Big|_{J=0}. \tag{8.40}$$

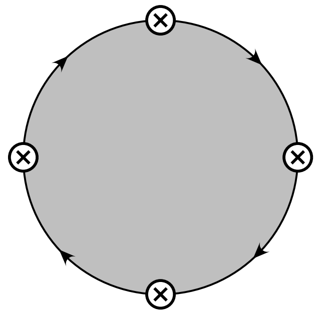


Figure 39: A graph with $g = 0$, $b = 1$ where quark bilinear operators are inserted at the quark loop.

The leading contribution to this generating functional is of order N, and it involves a quark loop at the boundary where we insert the bilinears, see Fig. 39. We conclude that

$$\langle M_1 \cdots M_r \rangle^{(c)} \sim N^{1-r}. \tag{8.41}$$

We now use these rules to derive counting rules for meson and glueball scattering amplitudes. Gluon operators G_i create glueball states, while quark bilinears B_i create meson states

$$G_i|0\rangle \sim |G_i\rangle, \qquad M_i|0\rangle \sim |M_i\rangle. \tag{8.42}$$

To look for appropriately normalized states, we notice that

$$\langle G_1|G_2\rangle \sim \langle G_1 G_2\rangle^{(c)} \sim \mathcal{O}(N^0), \tag{8.43}$$

therefore G_i creates glueball states with unit amplitude. Similarly,

$$\langle M_1|M_2\rangle \sim \langle M_1 M_2\rangle^{(c)} \sim \mathcal{O}(1/N), \tag{8.44}$$

therefore the appropriately normalized meson state is

$$\sqrt{N} M_i|0\rangle \tag{8.45}$$

We can now see that meson and glueball interactions are suppressed by factors of N. An r-glueball vertex is suppressed by N^{2-r}, and each additional glueball adds a $1/N$ suppression. Similarly, a normalized r meson vertex will be suppressed as

$$\langle \sqrt{N} M_1 \cdots \sqrt{N} M_r \rangle^{(c)} \sim N^{1-r/2} \tag{8.46}$$

and each additional meson adds a $1/\sqrt{N}$ suppression. Finally, mixed glueball-meson correlators will be suppressed as

$$\langle G_1 \cdots G_s \sqrt{N} M_1 \cdots \sqrt{N} M_r \rangle^{(c)} \sim N^{1-s-r/2} \tag{8.47}$$

In other words, if we think about $1/N$ as a coupling constant, we have reorganized QCD into a theory of weakly interacting glueballs and mesons. Finally, notice that we can obtain counting rules for the original fields of the Lagrangian by using the rescaling (8.6).

Example 8.3. Consider for example

$$\langle 0|\mathrm{Tr}(FF)|M\rangle, \qquad \langle 0|\mathrm{Tr}(FF)|G\rangle \tag{8.48}$$

Using the rules above we find

$$\langle 0|\mathrm{Tr}(FF)|M\rangle \sim \frac{1}{\sqrt{N}}, \qquad \langle 0|\mathrm{Tr}(FF)|G\rangle \sim \mathcal{O}(1). \tag{8.49}$$

In terms of rescaled fields, we have $\mathrm{Tr}(\hat{F}\hat{F}) \sim \sqrt{N}\mathrm{Tr}(FF)$, therefore

$$\langle 0|\mathrm{Tr}(\hat{F}\hat{F})|M\rangle \sim \sqrt{N}, \qquad \langle 0|\mathrm{Tr}(\hat{F}\hat{F})|G\rangle \sim N. \tag{8.50}$$

We will use these results later one, when analyzing the $U(1)$ problem from the viewpoint of the $1/N$ expansion.

Example 8.4. *Large N scaling of F_π.* The pion decay constant is defined by (10.19)-(10.20). This has the structure

$$\langle 0|M_1|M_2\rangle \sim 1/\sqrt{N}. \tag{8.51}$$

Since $\hat{q} \sim \sqrt{N}q$, it follows that $\hat{A}_{ud} \sim NA_{ud}$, and we finally obtain

$$F_\pi \sim \sqrt{N}. \tag{8.52}$$

8.3 QCD spectroscopy at large N: mesons and glueballs

We can now extract lessons from the above behavior for the spectrum of QCD. We will first analyze mesons and glueballs. The results are the following:

- At large N, both mesons and gluons are free, stable and non-interacting. Their masses have a smooth large N limit, and their number is infinite.
- Meson decay amplitudes are of order $1/\sqrt{N}$, and the large N limit is described by the tree diagrams of an effective local Lagrangian involving meson fields.
- To lowest order in $1/N$, glueball states are decoupled from mesons. The mixing between glueballs and mesons is of order $1/\sqrt{N}$, while the mixing between glueballs is of order $1/N$.

We now sketch an argument to establish the first property, following [88], where more details can be found. Let us consider the two-point function of a current J made of quark bilinears (and that can therefore create a meson, like in (10.19). As for any other two-point function, its spectral representation expresses it as a sum over poles, plus a more complicated part coming from multiparticle states. The first important result is that, at large N, only the sum over poles contributes, in other words

$$\langle J(k)J(-k)\rangle = \sum_n \frac{a_n^2}{k^2 - m_n^2}. \tag{8.53}$$

Here the sum is over one-particle meson states $|n\rangle$ with masses m_n, and

$$a_n = \langle 0|J|n\rangle \tag{8.54}$$

up to a kinematic factor. This can be established by noticing that the Feynman diagrams that contribute to this correlator at large N are diagrams with one single quark loop at the boundary. Therefore, when we cut this diagram as in Fig. 47 to detect intermediate states, we find exactly one $q\bar{q}$ pair. If we assume that confinement holds, this state must be a single meson.

From (8.53) we can also deduce that the spectrum of mesons contains an infinite number of states whose masses are well-defined at large N. This is because the r.h.s. of (8.53) is well-defined at large N. For example, if we normalize the currents as in (8.46), the r.h.s. is independent of N, and the meson masses m_n^2 also have a smooth limit which is independent of N. The number of states must be infinite, since at large k^2 we know from asymptotic freedom that the two-point function is logarithmic in k^2. The logarithmic behavior can only be obtained at large k^2 from the r.h.s. if the number of terms in the sum is infinite, otherwise we would find a k^{-2} behavior.

8.4 Baryons at large N

Baryons are color singlet hadrons made up of quarks. The $SU(N)$ invariant ϵ-symbol has N indices, so a baryon is an N-quark state,

$$\epsilon_{i_1 \cdots i_N} q^{i_1} \cdots q^{i_N}.$$

A baryon can be thought of as containing N quarks, one of each color, since all the indices on the ϵ-symbol must be different for it to be non-zero. Quarks obey Fermi statistics, and the ϵ-symbol is antisymmetric in color, so the baryon must be completely symmetric in the other quantum numbers such as spin and flavor.

The number of quarks in a baryon grows with N, so one might think that large N baryons have little to do with baryons for $N = 3$. However, one can compute baryonic properties in a systematic semiclassical expansion in $1/N$. The results are in good agreement with the experimental data, and provide information on the spin-flavor structure of baryons. We refer to [61] and references therein for an update on more recent results, and here we will content ourselves with some basic results from [88].

The N-counting rules for baryon graphs can be derived using previous results for meson graphs. Draw the incoming baryon as N-quarks with colors arranged in order, $1 \cdots N$. The colors of the outgoing quark lines are then a permutation of $1 \cdots N$. It is convenient to derive the N-counting rules for connected graphs. For this purpose, the incoming and outgoing quark lines are to be treated as ending on independent vertices, so that the connected piece of Fig. 40(a) is Fig. 40(b).

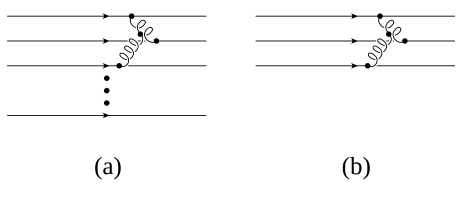


Figure 40: A baryon interaction and its corresponding connected component.

A connected piece that contains n quark lines will be referred to as an n-body interaction. The colors on the outgoing quarks in an n-body interaction are a permutation of the colors on the incoming quarks, and the colors are distinct. Each outgoing line can be identified with an incoming line of the same color in a unique way. One can now relate connected graphs for baryons interactions with planar diagrams with a single quark loop. The leading in N diagrams for the n-body interaction are given by taking a planar diagram with a single quark loop, cutting the loop in n places, and setting the color on each cut line to equal the color of one of the incoming (or outgoing) quarks. For example, the interaction in Fig. 40(b) is given by cutting Fig. 38 once at each of the three fermion lines. Planar meson diagrams contain a single closed quark loop as the outer edge of the diagram. Baryon n-body graphs obtained from cutting the quark loop require that one twist the quark lines to orient them with their arrows pointing in the same direction, and do not necessarily look planar when drawn on a sheet of paper. For example, Fig. 41 is a "planar" diagram for a two-body interaction. Baryon graphs in the double-line notation can have color index lines crossing each other due to the fermion line twists.

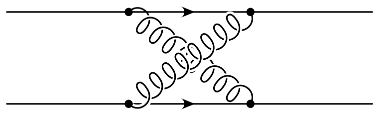


Figure 41: An example of a "planar" two-body baryon graph.

The relationship between meson and baryon graphs immediately gives us the N-counting rules for an n-body interaction in baryons: an n-body interaction is of order N^{1-n}, since planar quark diagrams are of order N, and n index sums over quark colors have been eliminated by cutting n fermion lines. Baryons contain N quarks, so n-body interactions are equally important for any n. n-body interactions are of order N^{1-n}, but there are $\mathcal{O}(N^n)$ ways of choosing n-quarks from a N-quark baryon. Thus the net effect of n-body interactions is of order N.

The result of this discussion suggests to use a Hartree–Fock strategy, since for large N we have a problem involving many particles with weak interactions. Interactions of quarks in a baryon can be described by a non-relativistic Hamiltonian if the quarks are very heavy. The Hamiltonian has the form

$$H = Nm + \sum_i \frac{p_i^2}{2m} + \frac{1}{N} \sum_{i \neq j} V\left(x_i - x_j\right) + \frac{1}{N^2} \sum_{i \neq j \neq k} V\left(x_i - x_j, x_i - x_k\right) + \ldots . \tag{8.55}$$

Each term contributes $\mathcal{O}(N)$ to the total energy. The interaction terms in the Hamiltonian eq. (8.55) are the sum of many small contributions, so fluctuations are small, and each quark can be considered to move in an average background potential. Consequently, the Hartree approximation is exact in the large N limit. The ground state wavefunction can be written as

$$\psi_0\left(x_1, \ldots, x_N\right) = \prod_{i=1}^{N} \phi_0\left(x_i\right), \tag{8.56}$$

where x_i are the positions of the quarks. The spatial wavefunction $\phi_0(x)$ is N-independent, so the baryon size is fixed in the $N \to \infty$ limit. The first excited state wavefunction is

$$\psi_1\left(x_1, \ldots, x_N\right) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \phi_i\left(x_k\right) \prod_{i=1, i \neq k}^{N} \phi_0\left(x_i\right). \tag{8.57}$$

Further details about this approach can be found in [26, 88].

8.5 Analyticity in the $1/N$ expansion

Standard perturbation theory (even in the absence of renormalons) is divergent due to the factorial growth of the number of diagrams. In the $1/N$ expansion, however, the computation of the genus g contribution (like in (8.29) involves a sum over fatgraphs with a fixed topology. It turns out that the number of such graphs does not grow factorially, but only *exponentially.* Therefore, barring problems associated to renormalons, the genus g amplitudes are in principle analytic functions in the 't Hooft parameter t with a finite radius of convergence around $t = 0$. This has been verified in various models where renormalons are absent, like matrix models and matrix quantum mechanics [17], $N = 4$ super Yang–Mills theory [14] and Chern–Simons theory [43].

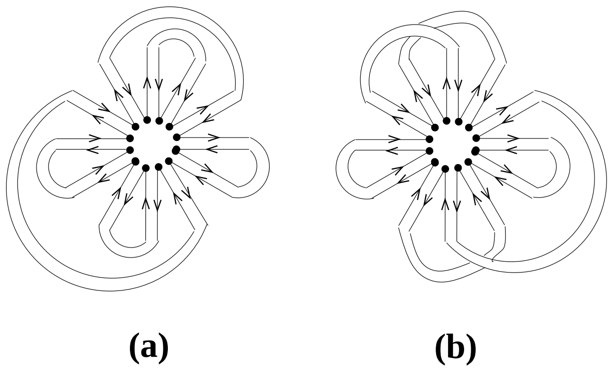


Figure 42: An example of planar (or "petal") diagram (a) and a non-planar one (b) obtained from a vertex with $2k$ edges (in this example, $k = 6$) by contracting them.

One simple example which shows the factorial *versus* exponential behavior is the fatgraph version of the calculation (2.14). This counts the number of possible contractions in a vertex with $2k$ legs, and grows factorially. This factorial growth gets inherited in the large order behavior of quantum-mechanical models. If we now consider the vertex to be a fatgraph, and we consider the possible contractions, we will of course get planar and nonplanar diagrams, see Fig. 42. The total number of contractions remains the same, and given by (2.14), but if one considers contractions that lead to *planar* diagrams, the number is much smaller. One way to derive this number [35] is to notice that the planar diagrams have a "petal" structure, in which the petals are either juxtaposed or included into one-another (with no edges-crossings). The counting of these petal diagrams is a standard problem in combinatorics which might be solved by using a recursion relation. Let us imagine that we want to obtain a petal diagram with $2k$ edges. We first fix one edge (say at position 1), and then we sum over the positions of the edges which can be contracted with the first one. These edges are at positions $2j$, where $j = 1, 2, ..., k$ (other positions will lead to crossing edges, which are forbidden due to the planarity condition). The petal obtained with this contractions has two halves, with $2(j - 1)$ edges in one of them and $2(k - j)$ edges in the other one, and therefore might lead to $c_{j-1} \cdot c_{k-j}$. Summing over all the possible positions gives the recursion relation

$$c_k = \sum_{j=1}^{k} c_{j-1} c_{k-j} \qquad c_0 = 1 \tag{8.58}$$

which is solved by the Catalan numbers

$$c_k = \frac{(2k)!}{(k+1)!k!}. \tag{8.59}$$

This number, in contrast to (2.14), grows only exponentially

$$c_k \sim 4^k. \tag{8.60}$$

This is an indication that, if we sum over fatgraphs with the same topology, we might obtain amplitudes with a finite radius of convergence.

Let us now give a more precise argument for the analyticity of physical amplitudes. For simplicity, we will consider the free energy of the theory (8.29), since in this case one just looks at connected bubble diagrams Γ. We will denote by $\mathcal{A}_n^{(c)}$ the set of independent, connected Feynman diagrams with

$$n = E - V = L - 1 \tag{8.61}$$

where E is the number of edges (propagators), V is the number of vertices, and L is the number of loops. We then have,

$$F(\mathbf{g}, g_s) = \sum_{n=1}^{\infty} \sum_{\Gamma \in \mathcal{A}_n^{(c)}} c_\Gamma W_{\mathbf{g}}(\Gamma) g_s^n, \tag{8.62}$$

where c_Γ is the Feynman integral associated to the diagram Γ, and $W_{\mathbf{g}}(\Gamma)$ is the group factor. Let us now see how this looks in the $1/N$ expansion. Due to the thickening rules, each diagram Γ gives a formal linear combination of fatgraphs $\Gamma_{g,h}$, which can be classified topologically by their genus g and the number of closed loops h:

$$\Gamma \rightarrow \sum_{g,h} p_{g,h}(\Gamma) \Gamma_{g,h} \tag{8.63}$$

and we have

$$W_{\mathrm{u}(N)}(\Gamma) = \sum_{g,h} p_{g,h}(\Gamma) N^h. \tag{8.64}$$

One then finds the following expression for the free energy:

$$F(\mathrm{u}(N), g_s) = \sum_{g=0}^{\infty} \sum_{\Gamma_{g,h}} c_\Gamma p_{g,h}(\Gamma) N^h g_s^{E-V} \tag{8.65}$$

where $E(\Gamma), V(\Gamma)$ are the number of edges and vertices in Γ (these topological data do not depend on the fattening of the graph). If we now use Euler's relation (8.22), we find the series (8.29), (8.30), where $a_{g,h}$ is given by the following finite sum

$$a_{g,h} = \sum_{\Gamma_{g,h}} c_\Gamma p_{g,h}(\Gamma). \tag{8.66}$$

It is easy to see that

$$p_{g,h}(\Gamma) \leq C_p^{2g-2+h}. \tag{8.67}$$

For example, in a theory with a pure cubic interaction, each vertex gives two resolutions, and the maximum number of terms is 2^V. Since in a theory with cubic interactions we have

$$3V = 2E \tag{8.68}$$

we deduce

$$p_{g,h}(\Gamma) \leq 2^V = 4^{E-V} = 4^{2g-2+h}. \tag{8.69}$$

In QCD there are also quartic vertices, which from the point of view of this counting can be regarded as two cubic vertices joined by an edge, and lead to a similar estimate. The next step is to analyze the Feynman integrals, c_Γ. If the theory has renormalons, they can grow factorially with the number of vertices. But in a theory without renormalons they grow only exponentially in the number of vertices, and we can write

$$|c_\Gamma| \sim C_F^{2g-2+h}. \tag{8.70}$$

This has been shown to be the case for a large class of diagrams in Quantum Mechanics [7], and it has been proved to be the case in Chern–Simons theory, by using the formulation in terms of the LMO invariant [44]. We then have,

$$a_{g,h} \sim (C_p C_F)^{2g-2+h} \sum_{\Gamma_{g,h}} 1. \tag{8.71}$$

In the last equation, we sum over all diagrams with the appropriate weight. Although we have just set it equal to 1, depending on the way we normalize the interaction we have a specific counting. For example, if we normalize all vertices of degree p with a factor $1/p!$, the weight of a diagram $\Gamma_{g,h}$ is given by

$$\frac{1}{|\mathrm{Aut}(\Gamma_{g,h})|} \tag{8.72}$$

i.e. the inverse of the order of the automorphism group. The counting of fatgraphs (weighted by their automorphism group, as above) has been developed very much both in combinatorics and in mathematical physics. The main result we have in this respect is that

$$\sum_{\Gamma_{g,h}} 1 \sim C_D^V C_G^g (2g)!, \tag{8.73}$$

see for example [43]. We conclude that

$$a_{g,h} \sim (2g)! C_1^g C_2^h. \tag{8.74}$$

Therefore, for *fixed genus*, we have that $F_g(t)$ is analytic at $t = 0$ with a finite radius of convergence ρ common to all g. This is the analiticity result we wanted to establish. Generically $\rho < \infty$, and there is typically a singularity t_c in the t-plane somewhere in the circle of radius ρ,

$$|t_c| = \rho. \tag{8.75}$$

Notice as well that *for fixed* t the sequence $F_g(t)$ will diverge like $(2g)!$. This will be important in the next subsection, when we discuss the rôle of instantons in the $1/N$ expansion.

8.6 Large N instantons

What happens to instantons in large N theories? Let us consider a theory with a coupling constant g_s (for a gauge theory, one has $g_s = g^2$ in our previous conventions) and 't Hooft parameter

$$t = g_s N. \tag{8.76}$$

Let us consider an instanton solution whose action (including the coupling) is given by

$$S_{\text{inst}} = \frac{c_0}{g_s}, \qquad c_0 \sim \mathcal{O}(1) \tag{8.77}$$

i.e. we assume that the action of the instanton is of order one at large N. This is generally the case. In Yang–Mills theory, this is due to the fact we can build an instanton by using just an $SU(2)$ subgroup of $U(N)$. In the example of Matrix Quantum Mechanics of the next section, the theory turns out to be equivalent to a theory of N free fermions. An instanton configuration can then be obtained by tunneling a single fermion out of N. Of course, there are instanton configurations whose action is of the same order than N, but these are "giant instantons" which will not be considered here. We can now do perturbation theory around the instanton configuration. The one-loop fluctuations give a term with the generic form (at large N)

$$\left(\frac{c}{g_s}\right)^{c_1 N} \tag{8.78}$$

where $c_1 N$ is the number of zero modes, or collective coordinates of the instanton, at large N. This factor comes from the canonical normalization of the modes in the path integral, since we can always normalize the fields in such a way that the action has an overall power of $1/g_s$. Putting both things together and, expressing everything in terms of g_s and the 't Hooft parameter, we find

$$\left(\frac{c}{g_s}\right)^{c_1 N} \mathrm{e}^{-c_0/g_s} \approx \exp\left(-\frac{A(t)}{g_s}\right) \tag{8.79}$$

where

$$A(t) = c_0 - c_1 t \log\left(\frac{c}{t}\right) + \mathcal{O}(t), \tag{8.80}$$

is called the large N instanton action. It is given by a series in t which incorporates, on top of the classical action and the one-loop fluctuations which we have written down explicitly, all vacuum, connected bubble planar diagrams (at all loops) in the background of the classical instanton action. To see how these appear, let us focus for simplicity on the interaction given by the cubic vertex Fig. 33. Let us consider fluctuations around the instanton solution $\overline{A}$

$$A = \overline{A} + A', \tag{8.81}$$

The action for the fluctuations will include a vertex of the form

$$\sum_{i,j,k} (\overline{A}_\mu)^i{}_j \, (A_\nu)^j{}_k \, (A_\rho)^k{}_i \tag{8.82}$$

and involving the instanton background. We can represent this vertex in the double-line notation as in Fig. 43, where the red line ending on the blob corresponds to the instanton background. It gives a factor of g_s, but only the interior line gives a factor of N after

Figure 43: The instanton vertex (8.82) (left) and a planar diagram contributing to the large N instanton action $A(t)$ a term of order t^3 (right).

tracing over. A simple example of a diagram contributing to the instanton action is the one depicted on the r.h.s. of Fig. 43. The inner closed lines gives a factor of N^3, and the diagram is proportional to

$$\mathrm{Tr}\big(\overline{A}^3\big)N^3 g_s^2 = \frac{1}{g_s} t^3 \mathrm{Tr}\big(\overline{A}^3\big), \tag{8.83}$$

since there are nine edges $E = 9$ and seven vertices $V = 7$, so the power of g_s^{E-V} is two. This diagram gives a correction of order t^3 to $A(t)$. We see that the calculation of large N instanton actions in realistic theories is of course difficult, since we have to sum up an infinite number of planar diagrams (in the same way that calculating the planar free energy involves adding up an infinite number of diagrams at all loops).

An alternative, more general way to think about large N instantons is in terms of large N effective actions. The idea of the master field suggests that ordinary theories at large N can be reformulated in terms of a "large N effective action" with coupling constant (or $\hbar$ constant) equal to $1/N$. In this theory, correlation functions at large N are obtained simply by solving the classical equations of motion of the effective action in the presence of sources. A large N instanton is simply an instanton solution of this large N effective theory, i.e. a saddle point with finite action. This is in general different from the usual instanton configurations, which are saddle point of the *classical* action. In some cases, large N instantons can be thought of as deformations of the classical instantons, where the deformation parameter is the 't Hooft parameter: as it is manifest in (8.80), when $t \to 0$ we recover the gauge theory instanton. Explicit examples of large N instantons were obtained in the $\mathbb{CP}^N$ model in [**?**, 65], as deformations of classical instantons. A particularly beautiful example is the large N instanton of two-dimensional Yang–Mills theory obtained in [48].

In the same way that the standard factorial growth of the perturbative expansion is related to "standard" instantons, the growth of the $1/N$ expansion should be related to large N instantons. From the estimates in the previous subsection we deduce that quantities like the genus g free energy must diverge doubly-factorially. In fact we have the asymptotics,

$$F_g(t) \sim (2g)!(A(t))^{-2g}, \qquad g \gg 1. \tag{8.84}$$

where $A(t)$ is the action of a large N instanton. This type of growth has been found in many simple models with a nontrivial $1/N$ expansion.

Notice that, in the calculation of observables, large N instantons are weighted by

$$e^{-A(t)/g_s} \tag{8.85}$$

where $A(t)$, the large N instanton action, is in general a non-trivial function of the 't Hooft parameter. If $\mathrm{Re}(A(t)/g_s) > 0$, large N instantons are suppressed exponentially at large N (or small g_s), as in standard instanton physics. This might lead to think that "instantons are suppressed at large N," but as Neuberger pointed out in [66], this is not necessarily the case. It might happen for example that $A(t)$ vanishes at a particular value of t, and in this case the contribution of instantons become as important as the perturbative contributions. The value of the 't Hooft parameter for which $A(t)$ vanishes signals very often a *large N phase transition*, or a critical point, in the theory. The critical value of the 't Hooft parameter is also, in many cases, the first singularity t_c in the t-plane which we found in (8.75).

Remark 8.5. There has been some lot of confusion in the literature concerning the role of instantons at large N. It was noted in [86] that instanton methods and large N methods seem to be incompatible, in the sense that they give different qualitative predictions for the N-dependence of some quantities. For example, an instanton calculation of the topological susceptibility in Yang–Mills theory gives an N-dependence of the form e^{-N}, while large N methods predict a dependence of the form $1/N$. The solution to these apparent paradoxes is that, as we have noticed before, instanton calculus is not well-defined in the absence of an infrared (IR) cutoff, so it shouldn't come as a surprise that naive predictions based on instantons are not consistent with other procedures like the $1/N$ expansion. However, in theories with an IR cutoff (like finite temperature or finite volume), both instanton calculus and large N methods make sense, and the results are perfectly compatible (see [2] for an early discussion of this point).

Remark 8.6. Notice that, effectively, the diagrams contributing to the large N instanton action are similar to the diagrams involving an external boundary, coming from a Wilson loop for example. This means that, in the dual formulation in terms of Riemann surfaces, the large N instanton is associated to a hole in the worldsheet.

9. A solvable toy model: large N matrix quantum mechanics

9.1 Defining the model. Perturbation theory

We will consider a quantum-mechanical model where the degrees of freedom are the entries of a Hermitian $N \times N$ matrix M. This model is described by the Euclidean Lagrangian

$$L_M = \mathrm{Tr}\Big[\frac{1}{2}\dot{M}^2 + V(M)\Big], \tag{9.1}$$

where $V(M)$ is a polynomial in M. Notice that this problem has a symmetry

$$M \to UMU^\dagger \tag{9.2}$$

where U is a constant unitary matrix. This model is sometimes called *matrix quantum mechanics* (MQM). It can be regarded as a one-dimensional field theory for a quantum field $M(t)$ taking values in the adjoint representation of $U(N)$. As any other field theory, it can be studied in perturbation theory. We will assume that the potential $V(M)$ is of the form

$$V(M) = \frac{1}{2}M^2 + V_{\text{int}}(M) \tag{9.3}$$

where $V_{\text{int}}(M)$ is the interaction term. The Feynman rules are the same as in the case of quantum mechanics, with the only difference that we will now have "group factors" due to the fact that M is matrix valued.

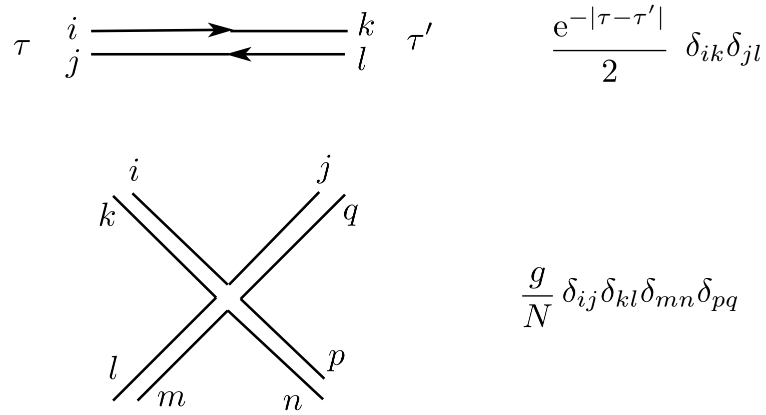


Figure 44: Feynman rules for matrix quantum mechanics.

The propagator of MQM is

$$\frac{e^{-|\tau|}}{2}\,\delta_{ik}\delta_{jl}. \tag{9.4}$$

For a theory with a quartic interaction

$$V_{\text{int}}(M) = \frac{g}{N}M^4 \tag{9.5}$$

the Feynman rules are illustrated in Fig. 44. The factor of N in (9.5) is introduced in order to have a standard large N limit, as we will see in more detail later.

One can use these rules to compute the perturbation series of the ground state energy of MQM, which is obtained by considering connected bubble diagrams. Here we indicate the calculation up tp order g^3. The relevant Feynman diagrams are shown in Fig. 3. As in any field theory for fields in the adjoint representation, each Feynman diagram leads to a group factor which depends on N, i.e. each conventional Feynman diagram gives various fatgraphs that can be classified according to their topology. A fatgraph with V vertices and h boundaries will have a factor

$$g^V N^{h-V} = g^V N^{2-2g}, \tag{9.6}$$

since the number of edges is twice the number of vertices, $E = 2V$ (this is a quartic interaction!) and

$$h + E - V = h - V. \tag{9.7}$$

Planar diagrams, as usual, are proportional to N^2. The symmetry factors for the first few planar diagrams are given in table 2 (see [17]). These numbers can be checked by taking

diagram	1	2a	2b	3a	3b	3c	3d
symmetry factor	2	16	2	256/3	32/3	64	128

Table 2: Symmetry factors of the planar quartic diagrams shown in Fig. 3.

into account that the total symmetry factor for connected diagrams with n quartic vertices is given by the Gaussian average

$$\frac{1}{n!}\langle(\mathrm{Tr}M^4)^n\rangle^{(c)}. \tag{9.8}$$

where M is a Hermitian $N \times N$ matrix. For example,

$$\begin{aligned}\langle \mathrm{Tr}M^4\rangle &= 2N^3 + N,\\ \frac{1}{2!}\langle(\mathrm{Tr}M^4)^2\rangle^{(c)} &= \frac{1}{2}(\langle(\mathrm{Tr}M^4)^2\rangle - \langle\mathrm{Tr}M^4\rangle^2) = 18N^4 + 30N^2,\end{aligned} \tag{9.9}$$

where $18 = 16 + 2$ corresponds to planar diagrams, in agreement with table 2.

We can now compute the first corrections to the *planar* ground state energy. For the Feynman integrals we find the same values we found in (2.11) for conventional quantum mechanics. Putting together the Feynman integrals with the symmetry factors, we obtain

$$\begin{aligned} 1:&\quad \frac{1}{4}\cdot 2\\ 2a:&\quad -\frac{1}{16}\cdot 1\cdot 16,\\ 2b:&\quad -\frac{1}{16}\cdot\frac{1}{2}\cdot 2,\\ 3a:&\quad \frac{1}{64}\cdot\frac{3}{2}\cdot\frac{256}{3}\\ 3b:&\quad \frac{1}{64}\cdot\frac{3}{8}\cdot\frac{32}{3}\\ 3c:&\quad \frac{1}{64}\cdot\frac{5}{8}\cdot 64\\ 3d:&\quad \frac{1}{64}\cdot 1\cdot 128 \end{aligned} \tag{9.10}$$

We then find

$$E_0(N) = N^2\mathcal{E}_0(g) + \mathcal{E}_1(g) + \cdots \tag{9.11}$$

where

$$\mathcal{E}_0(g) = \frac{1}{2} + \frac{1}{2}g - \frac{17}{16}g^2 + \frac{75}{16}g^3 + \cdots \tag{9.12}$$

9.2 Exact ground state energy in the planar approximation

Remarkably, the planar ground state energy in MQM can be obtained *exactly* by using a free fermion formulation. This exact result *resums* in closed form all the planar diagrams of MQM contributing to the ground state energy. This was noted in the classic paper [17], which we now explain.

After quantization of the system we obtain a Hamiltonian operator

$$H = \mathrm{Tr}\Big[-\frac{1}{2}\frac{\partial^2}{\partial M^2} + V(M)\Big], \tag{9.13}$$

where

$$\mathrm{Tr}\frac{\partial^2}{\partial M^2} = \sum_{ab} \frac{\partial^2}{\partial M_{ab} M_{ba}} \tag{9.14}$$

In order to study the spectrum of this Hamiltonian, it is useful to change variables

$$M = U \Lambda U^\dagger \tag{9.15}$$

where

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_N) \tag{9.16}$$

is a diagonal matrix. It can be shown that

$$\mathrm{Tr}\frac{\partial^2}{\partial M^2} = \frac{1}{\Delta(\lambda)} \sum_{a=1}^{N} \Big(\frac{\partial}{\partial \lambda_a}\Big)^2 \Delta(\lambda) + \sum_{a<b} \frac{\mathcal{F}_{ab}}{(\lambda_a - \lambda_b)^2}, \tag{9.17}$$

where

$$\Delta(\lambda) = \prod_{a<b} (\lambda_a - \lambda_b) \tag{9.18}$$

is the Vandermonde determinant, and $\mathcal{F}_{ab}$ are differential operators w.r.t. the angular coordinates in U (see [62] for a statement of this result). Notice that the first term in (9.17) can be written as

$$\sum_{a=1}^{N} \Big(\frac{\partial}{\partial \lambda_a}\Big)^2 + \frac{2}{\Delta} \sum_{a=1}^{N} \frac{\partial \Delta}{\partial \lambda_a} \frac{\partial}{\partial \lambda_a} + \frac{1}{\Delta} \sum_{a=1}^{N} \frac{\partial^2 \Delta}{\partial \lambda_a^2} \tag{9.19}$$

Since

$$\log \Delta = \sum_{a<b} \log(\lambda_a - \lambda_b) \tag{9.20}$$

we find

$$\frac{1}{\Delta}\frac{\partial \Delta}{\partial \lambda_a} = \frac{\partial \log \Delta}{\partial \lambda_a} = \sum_{b \neq a} \frac{1}{\lambda_a - \lambda_b} \tag{9.21}$$

We also have

$$\sum_{a=1}^{N} \frac{\partial^2 \Delta}{\partial \lambda_a^2} = 0. \tag{9.22}$$

Therefore, we find, acting on singlet states (i.e., states that are invariant under the full $U(N)$ group)

$$-\frac{1}{2}\mathrm{Tr}\frac{\partial^2}{\partial M^2} = -\frac{1}{2} \sum_{a=1}^{N} \frac{\partial^2}{\partial \lambda_a^2} + \sum_{b \neq a} \frac{1}{\lambda_b - \lambda_a} \frac{\partial}{\partial \lambda_a}. \tag{9.23}$$

After reduction to eigenvalues, the $U(N)$ group still acts through the Weyl group, i.e. by permuting eigenvalues. Therefore, singlet states will be represented by a *symmetric* function of N eigenvalues,

$$\Psi(\lambda_i), \tag{9.24}$$

which in particular does not depend on the angular coordinates of U. If we are now interested in computing the spectrum of the Hamiltonian for singlet states, we can reformulate the problem as a the problem of N fermions in the potential $V(x)$. To see this, we introduce a completely *antisymmetric* wavefunction

$$\Phi(\lambda) = \Delta(\lambda)\Psi(\lambda) \tag{9.25}$$

The equation

$$H\Psi = E\Psi \tag{9.26}$$

can be written as

$$\left(\sum_{i=1}^{N} h(\lambda_i)\right)\Phi(\lambda_j) = E\Phi(\lambda_j) \tag{9.27}$$

where $h(\lambda)$ is the Hamiltonian

$$h(\lambda) = -\frac{\hbar^2}{2}\frac{\partial^2}{\partial\lambda^2} + V_N(\lambda). \tag{9.28}$$

We have explicitly introduced the Planck constant and relabel the potential as V_N. Since the fermions are not interacting, we can just solve the Schröndiger equation for a single particle of unit mass,

$$h(\lambda)\phi_n(\lambda) = E_n\phi_n(\lambda) \tag{9.29}$$

In particular, the ground state of the system (in the singlet sector) will be obtained by putting the N fermions in the first N energy levels of the potential, and its energy will be

$$E(N) = \sum_{n=1}^{N} E_n \tag{9.30}$$

We want to compute the ground state energy in the limit in which N is very large. We assume that $V(\lambda)$ has the "right" factors of N, more precisely

$$V_N(\lambda) = NV(\lambda/\sqrt{N}). \tag{9.31}$$

In this case, one can see that λ is of order $N^{\frac{1}{2}}$, so we can redefine

$$\lambda \to N^{\frac{1}{2}}\lambda, \tag{9.32}$$

and λ is now of order 1. In this way we find a Hamiltonian where the potential only depends on λ

$$\frac{1}{N}h(\lambda) = -\frac{\hbar^2}{2N^2}\frac{\partial^2}{\partial\lambda^2} + V(\lambda). \tag{9.33}$$

and the problem to be solved is

$$\left\{-\frac{\hbar^2}{2N^2}\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} + V(\lambda)\right\}\phi_n(\lambda) = e_n\phi_n(\lambda). \tag{9.34}$$

where we denoted

$$e_n = \frac{1}{N}E_n \tag{9.35}$$

For example,

$$V_N(\lambda) = \frac{1}{2}\lambda^2 + \frac{g}{N}\lambda^4, \tag{9.36}$$

has the right scaling properties. This can be interpreted as saying that

$$g = N g_s \tag{9.37}$$

is the 't Hooft parameter of the model, which is kept fixed as $N \to \infty$.

Notice that, since the quantum effects are controlled by $\hbar/N$, large N is equivalent to $\hbar$ small and in the large N limit we can use the *semiclassical approximation*. The total energy of the ground state is

$$E_0(N) = \sum_{k=1}^{N} E_k = N \sum_{k=1}^{N} e_k = N^2 \mathcal{E}_0 + \cdots , \tag{9.38}$$

where $\mathcal{E}_0$ is independent of N.

In order to solve this problem, we notice that, since the effective Planck constant in this problem is $\hbar/N$, when N is large we can use the WKB approximation. In particular, we can use the Bohr–Sommerfeld formula to find the energy spectrum at leading order in $\hbar/N$. We will write this semiclassical quantization condition as

$$N J(e_n) = n - \frac{1}{2}, \qquad n \geq 1, \tag{9.39}$$

where

$$J(e) = \frac{1}{\pi\hbar} \int_{\lambda_1(e)}^{\lambda_2(e)} \mathrm{d}\lambda \sqrt{2(e - V(\lambda))} \tag{9.40}$$

and $\lambda_{1,2}(e)$ are the turning points of the potential. If we denote

$$\xi = \frac{n - \frac{1}{2}}{N}, \tag{9.41}$$

we see that (9.39) defines implicitly a function $e(\xi)$. The total ground state energy can then be written as

$$E_0(N) = N \sum_{n=1}^{N} e(\xi). \tag{9.42}$$

At large N, the spectrum becomes denser and denser, and the variable ξ becomes a continuous variable

$$\xi \in [0, 1]. \tag{9.43}$$

At large N, the sum in (9.42) becomes an integral through the rule

$$\sum_{n=1}^{N} \to N \int_0^1 \mathrm{d}\xi \tag{9.44}$$

and we find

$$E_0(N) \to N^2 \int_0^1 \mathrm{d}\xi e(\xi), \tag{9.45}$$

in other words,

$$\mathcal{E}_0 = \int_0^1 \mathrm{d}\xi e(\xi). \tag{9.46}$$

To evaluate this integral, we change variables from ξ to e. We define the Fermi energy by the condition

$$J(e_F) = 1 \tag{9.47}$$

therefore $\xi = 1$ corresponds to $e = e_F$, while $\xi = 0$ corresponds to $e = \min V(\lambda)$. We then find,

$$\mathcal{E}_0 = \int_0^1 \mathrm{d}\xi e(\xi) = \int_{\min V(\lambda)}^{e_F} \mathrm{d}e\, e J'(e), \tag{9.48}$$

where

$$J'(e) = \frac{1}{\pi\hbar}\int_{\lambda_1(e)}^{\lambda_2(e)} \frac{\mathrm{d}\lambda}{\sqrt{2(e - V(\lambda))}}. \tag{9.49}$$

An easy calculation gives,

$$\begin{aligned} \mathcal{E}_0 &= \frac{1}{\pi\hbar}\int_{\min V(\lambda)}^{e_F} \mathrm{d}e \int_{\lambda_1(e)}^{\lambda_2(e)} \mathrm{d}\lambda\, \frac{e}{\sqrt{2(e-V(\lambda))}} \\ &= \frac{1}{\pi\hbar}\int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \mathrm{d}\lambda \int_{V(\lambda)}^{e_F} \mathrm{d}e\, \frac{e}{\sqrt{2(e-V(\lambda))}} \\ &= \frac{1}{3\pi\hbar}\int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \mathrm{d}\lambda (2V(\lambda) + e_F)\sqrt{2(e_F - V(\lambda))} \end{aligned} \tag{9.50}$$

and the final expression we obtain is

$$\mathcal{E}_0 = e_F - \frac{1}{3\pi\hbar}\int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \mathrm{d}\lambda \Big[2(e_F - V(\lambda))\Big]^{\frac{3}{2}}. \tag{9.51}$$

The previous development suggests to introduce a *density of eigenvalues* $\rho(\lambda)$. Using (9.47) we find that

$$\frac{1}{\pi\hbar}\int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \mathrm{d}\lambda\, \sqrt{2(e_F - V(\lambda))} = 1, \tag{9.52}$$

therefore

$$\rho(\lambda) = \frac{1}{\pi\hbar}\sqrt{2(e_F - V(\lambda))} \tag{9.53}$$

is a normalized distribution of eigenvalues which can be regarded as the master field of matrix quantum mechanics.

9.3 Excited states, or glueball spectrum

We can now compute the analogue of the glueball spectrum in MQM. This discussion is based on [62].

The first excited state can be obtained by exciting the last fermion in the Fermi sea, i.e.

$$E_1(N) = E(N) + E_{N+1} - E_N, \tag{9.54}$$

therefore

$$E_1(N) - E(N) = E_{N+1} - E_N = N(e_{N+1} - e_N) = N\Big(e(\xi + 1/N) - e(\xi)\Big)_{\xi=1} \tag{9.55}$$

which at leading order in $1/N$ is given by

$$E_1(N) - E(N) = \left.\frac{\mathrm{d}e(\xi)}{\mathrm{d}\xi}\right|_{\xi=1} = \omega, \tag{9.56}$$

where

$$\omega = \frac{1}{J'(e_F)}. \tag{9.57}$$

Notice that ω is just the frequency of a classical particle with the Fermi energy. A general excited singlet state is obtained by exciting r fermions from the Fermi sea. It is characterized by the integers

$$0 \leq h_1 < h_2 < \cdots < h_r, \qquad 1 \leq p_1 < p_2 \leq \cdots < p_r, \tag{9.58}$$

and its energy is

$$E_{h,p}(N) - E_0(N) \sim \omega \sum_{i=1}^{r} (h_i + p_i). \tag{9.59}$$

9.4 Some examples

Example 9.1. *Harmonic oscillator.* A simple case occurs for

$$V_N(\lambda) = \frac{1}{2}\omega^2\lambda^2. \tag{9.60}$$

The exact answer for the ground state energy is

$$E_0(N) = N \sum_{n=1}^{N} \frac{\hbar}{N}\omega\Big(n - \frac{1}{2}\Big) = \frac{\hbar\omega}{2}N^2, \tag{9.61}$$

therefore

$$\mathcal{E}_0 = \frac{\hbar\omega}{2}. \tag{9.62}$$

Let us now compute this with the formulae above. First of all, we have that

$$J(\theta) = \frac{\theta}{\hbar\omega} \Rightarrow e(\phi) = \hbar\omega\phi, \tag{9.63}$$

therefore

$$e_F = e(1) = \hbar\omega. \tag{9.64}$$

One also finds,

$$\mathcal{E}_0 = e_F - \frac{1}{2}\frac{e_F^2}{\hbar^2\omega^2} = \frac{1}{2}\hbar\omega, \tag{9.65}$$

which agrees with the direct computation above.

Example 9.2. *The quartic potential.* This is the potential originally considered in [17]. The potential is given by

$$V(\lambda) = \frac{1}{2}\lambda^2 + g\lambda^4. \tag{9.66}$$

We first compute the Fermi energy, which is defined by (9.47). The integral involved here can be computed in terms of elliptic functions. We first write

$$2e - \lambda^2 - 2g\lambda^4 = 2g(a^2 - \lambda^2)(b^2 + \lambda^2), \tag{9.67}$$

where

$$a^2 = \frac{\sqrt{16eg+1} - 1}{4g}, \qquad b^2 = \frac{\sqrt{16eg+1} + 1}{4g}. \tag{9.68}$$

We introduce the elliptic modulus

$$k^2 = \frac{a^2}{a^2 + b^2}. \tag{9.69}$$

Then, we have that

$$J(e) = \frac{2}{3\pi\hbar}(2g)^{\frac{1}{2}}(a^2+b^2)^{\frac{1}{2}}\Big[b^2 K(k) + (a^2 - b^2)E(k)\Big]. \tag{9.70}$$

The implicit function $e_F(g)$ is easy to compute in perturbation theory in g, by using the series expansion of the elliptic functions. We find (we set $\hbar = 1$ in the following)

$$e_F(g) = 1 + \frac{3g}{2} - \frac{17g^2}{4} + \frac{375g^3}{16} + \mathcal{O}(g^4). \tag{9.71}$$

The planar free energy is given by

$$\mathcal{E}_0(g) = e_F(g) - \frac{1}{3\pi}\mathcal{I}(g, e_F(g)). \tag{9.72}$$

which involves the integral

$$\begin{aligned}\mathcal{I}(g,e) = \int_{-a}^{a} \mathrm{d}t \Big[(a^2 - t^2)(b^2 + t^2)\Big]^{\frac{3}{2}} = \\ \frac{2}{35}\sqrt{a^2+b^2}\Big\{2(a^2-b^2)(a^4+6a^2b^2+b^4)E(k) + b^2(2b^4+9a^2b^2-a^4)K(k)\Big\}.\end{aligned} \tag{9.73}$$

This can be also be computed in perturbation theory in g, and it gives

$$\mathcal{E}_0 = \frac{1}{2} + \frac{g}{2} - \frac{17g^2}{16} + \frac{75g^3}{16} + \mathcal{O}(g^4) \tag{9.74}$$

in perfect agreement with the calculation in planar perturbation theory (9.12).

One important remark on this result is that $\mathcal{E}_0(g)$ is an analytic function of g at $g = 0$, with a finite radius of convergence. This follows from the explicit expression for $\mathcal{E}_0$ in terms of elliptic functions (in order to calculate $e_F(g)$ we have to invert an analytic function, and this preserves analyticity). The radius of convergence of the expansion (9.74) can be calculated by locating the position of the nearest singularity in the g_c plane. This singularity occurs when the modulus (9.69) becomes $-\infty$, i.e. when

$$e_F(g_c) = -\frac{1}{16g_c}. \tag{9.75}$$

It can be easily checked that this happens when

$$g_c = -\frac{\sqrt{2}}{6\pi}. \tag{9.76}$$

This has a nice interpretation in terms of the fermion picture. Since g_c is negative, we have an inverted quartic potential. The critical value of g corresponds to the moment in which the Fermi sea reaches the maximum of the potential, see Fig. 45.

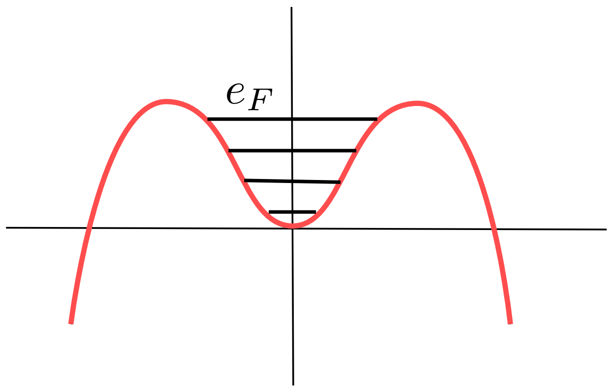


Figure 45: The Fermi level e_F in the quartic potential with negative coupling $g < 0$. The nearest singularity corresponds to the critical value in which e_F reaches the maximum of the potential.

Remark 9.3. The function $e_F(g)$ defined by $J(e_F(g)) = 1$, where $J(e)$ is given by (9.70), can be easily obtained numerically, and we plot it in the left hand side of Fig. 46. Using this result, we evaluate the planar free energy, which we plot it on the right hand side of the same figure (red line). We also show there (in green) the energy of the ground state as computed by the WKB/Bohr–Sommerfeld condition. Notice that, as $g \to 0$, both tend to $1/2$, which is the energy of the ground state of the harmonic oscillator. In principle, if we consider $N = 1$, we obtain the planar approximation for the ground state energy of a particle in the quartic potential. Surprisingly, by looking at the exact values computed numerically, which are above both curves, one observes that the planar approximation is slightly better than the WKB/Bohr–Sommerfeld approximation!

9.5 Large N instantons in matrix quantum mechanics

We can now try to calculate large N instanton effects in matrix quantum mechanics. For $g = -\kappa < 0$ we have a metastable vacuum at the origin and we should expect some instanton configuration mediating vacuum decay. In principle, one should write down an instanton solution with "small" action and calculate the path integral around this solution. This solution can be found by tunneling one single diagonal eigenvalue of the matrix M, which has an action of order $\mathcal{O}(1)$, i.e. we consider the matrix instanton,

$$M_c(t) = \text{diag}\,(0, \cdots, 0, q_c(t), 0, \cdots, 0)\,, \tag{9.77}$$

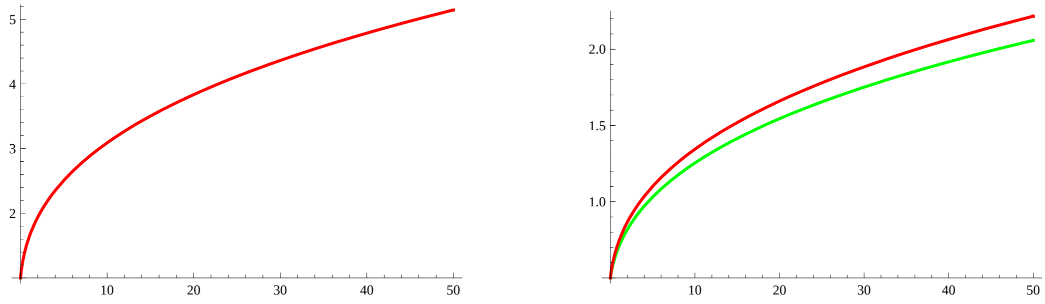


Figure 46: On the left: the Fermi energy e_F as a function of g. For $g = 0$, one has $e_F = 1$. On the right: the planar free energy $\mathcal{E}_0$ (red) and the ground state energy e_0 in the WKB approximation (green), as a function of g.

where $q_c(t)$ is the bounce (2.52) with $\lambda = 4g_s$. In principle, one could expand the path integral of matrix quantum mechanics around this configuration and compute quantum planar fluctuations to determine the large N instanton action (this calculation was originally proposed in [66]).

However, the fermion picture, which gives us a compact way of computing the planar ground state energy, should also give us an efficient way to compute the large N instanton action in a single strike. In this picture, the ground state is given by a filled Fermi level. As in any Fermi system, tunneling effects will first affect fermions which are near the Fermi surface. The instanton action of such a fermion is just given by the standard WKB action,

$$\frac{A(\kappa)}{g_s} = 2N\,(2\kappa)^{1/2}\int_a^b \sqrt{(\lambda^2 - a^2)(b^2 - \lambda^2)}. \tag{9.78}$$

where a, b are the turning points associated to the Fermi energy e_F, and they are non-trivial functions of κ (which, remember, plays the role of the 't Hooft parameter in this problem).The factor of 2 is due to the symmetry of the problem, and the factor of N is due to the fact that the effective Planck constant in this problem is in fact $1/N$, as we remarked in (9.34).

The integral in (9.78) can be explicitly computed by using elliptic functions, and the final result is

$$\frac{A(\kappa)}{\kappa} = \frac{2}{3}(2\kappa)^{\frac{1}{2}} b\Big[(a^2 + b^2)E(k) - 2a^2 K(k)\Big] \tag{9.79}$$

where the elliptic modulus is now given by

$$k^2 = \frac{b^2 - a^2}{b^2}. \tag{9.80}$$

The above function has the following expansion around $\kappa = 0$,

$$A(\kappa) = \frac{1}{3} - \kappa \log\left(\frac{4e}{\kappa}\right) + \frac{17\kappa^2}{4} + \frac{125\kappa^3}{8} + \cdots \tag{9.81}$$

This is precisely the expected structure for an instanton action in (8.80): the leading term is the action for an instanton (2.164) in the $N = 1$ quantum mechanical problem. They differ in a factor of 4 since the coupling λ used there is related to g_s defined in (9.37) by

$\lambda = 4g_s$. The log term is a one-loop factor in disguise, and the rest of the series is a sum of loop corrections in the background of the "classical" instanton. An interesting property of $A(\kappa)$ is that it *vanishes* at the critical value

$$\kappa_c = \frac{\sqrt{2}}{6\pi}, \tag{9.82}$$

which is indeed the first singularity in the complex g plane and corresponds to the critical behavior in which the Fermi sea reaches the local maximum.

9.6 Adding fermions, or meson spectrum

MQM as we have described it is a toy model to study Yang–Mills theory, since there is a single field, M, in the adjoint representation of $U(N)$. If we want to study other aspects of large N QCD, like for example meson spectra, we must introduce the toy analogue of quark fields, in the fundamental and the antifundamental representation of the group. This enlarged version of MQM was studied by Affleck in [2].

To start with, we have to add fermions to the model. For this, we add to (9.1) the fermionic piece

$$L_F = q^\dagger \Big[\frac{\mathrm{d}}{\mathrm{d}\tau} + m\sigma_3 + \frac{g_F}{\sqrt{N}} M\sigma_1\Big] q. \tag{9.83}$$

Here, $q^i_{f,\alpha}$ are two-component Fermi fields, with $\alpha = 1,2$ and i, f color and flavor indices, respectively. The standard vacuum is defined by

$$q^i_{f,1}|0\rangle = q^{i\dagger}_{f,2}|0\rangle = 0, \tag{9.84}$$

and the canonical commutation relations are

$$\{q^{i\dagger}_{f,\alpha}, q^j_{f',\beta}\} = \delta^{ij}\delta_{ff'}\delta_{\alpha\beta}. \tag{9.85}$$

The fermion propagator is given by

$$\langle 0|\mathrm{T}(q^i_{f,\alpha}(\tau) q^{j\dagger}_{f',\beta}(\tau))|0\rangle = \delta^{ij}\delta_{ff'} \int_{-\infty}^{\infty} \frac{\mathrm{d}p}{2\pi} \mathrm{e}^{\mathrm{i}p(\tau-\tau')} \Big(\frac{1}{\mathrm{i}p + m\sigma_3}\Big)_{\alpha\beta} \tag{9.86}$$

This is computed as follows

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{\mathrm{d}p}{2\pi} \frac{\mathrm{e}^{\mathrm{i}p(\tau-\tau')}}{\mathrm{i}p + m\sigma_3} &= \int_{-\infty}^{\infty} \frac{\mathrm{d}p}{2\pi} \mathrm{e}^{\mathrm{i}p(\tau-\tau')} \frac{-\mathrm{i}p + m\sigma_3}{p^2+m^2} \\
&= m\sigma_3 \int_{-\infty}^{\infty} \frac{\mathrm{d}p}{2\pi} \frac{\mathrm{e}^{\mathrm{i}p(\tau-\tau')}}{p^2+m^2} - \mathrm{i} \int_{-\infty}^{\infty} \frac{\mathrm{d}p}{2\pi} \frac{\mathrm{e}^{\mathrm{i}p(\tau-\tau')}p}{\mathrm{i}p + m\sigma_3} \\
&= m\sigma_3 \frac{1}{2m} \mathrm{e}^{-m|\tau-\tau'|} - \mathrm{i}\frac{2\pi\mathrm{i}}{2\pi} \epsilon(\tau-\tau') \mathrm{e}^{-m|\tau-\tau'|} \\
&= \frac{1}{2}\mathrm{e}^{-m|\tau-\tau'|} \Big[\sigma_3 + \epsilon(\tau-\tau')\Big].
\end{aligned} \tag{9.87}$$

Therefore,

$$\langle 0|\mathrm{T}(q^i_{f,\alpha}(\tau) q^{j,\dagger}_{f',\beta}(\tau))|0\rangle = \frac{1}{2}\mathrm{e}^{-m|\tau-\tau'|} \Big[\sigma_3 + \epsilon(\tau-\tau')\Big]_{\alpha\beta}. \tag{9.88}$$

In order to study the spectrum of the problem, we diagonalize M as in (9.15) and change variables to

$$q^i \to \tilde{q}^i = U^i{}_j q^j \tag{9.89}$$

so that the coupling between M and q, $q^\dagger$ in (9.83) is diagonal. It is easy to compute that

$$\frac{\partial \tilde{q}^a}{\partial M_{ij}} = \frac{\partial U_{ab}}{\partial M_{ij}} q^b = \sum_{b\neq a} \frac{U_{ai} U^\dagger_{jb}}{\lambda_a - \lambda_b} \tilde{q}^b. \tag{9.90}$$

We can write this as

$$\frac{\partial \tilde{q}^a_{f,\alpha}}{\partial M_{ij}} = [\mathcal{O}_{ij}, \tilde{q}^a_{f,\alpha}], \tag{9.91}$$

where

$$\mathcal{O}_{ij} = \sum_{a\neq b,\, \alpha, g} \frac{\tilde{q}^{a\dagger}_{g,\alpha} U_{ai} U_{jb} \tilde{q}^b_{g,\beta}}{\lambda_a - \lambda_b} \tag{9.92}$$

The Hamiltonian of the problem includes now a fermionic part H_F, which has two terms. The first one comes from the explicit fermionic piece of the Lagrangian and it is a bilinear. The second one comes from the kinetic piece, and it appears due to the fact that states made out of $\tilde{q}$ depend on U, therefore on M. It is derived in detail in [2, 60], and leads to a quartic term in fermion fields. When acting on singlets it reads (we set $\tilde{q} = q$ from now on)

$$H_F = \sum_{i=1}^{N} q^{i\dagger} \Big(m\sigma_3 + \frac{g_F}{\sqrt{N}} \lambda_i \Big) q^i + \frac{1}{2} \sum_{i\neq j} \frac{q^{i\dagger}_{f,\alpha} q^i_{g,\beta} q^j_{f,\alpha} q^{j\dagger}_{g,\beta}}{(\lambda_i - \lambda_j)^2}. \tag{9.93}$$

To find the vacuum, we define a new set of Fermi operators a^i_f, b^i_f which annihilate by definition the true ground state:

$$a^i_f |\theta\rangle = b^i_f |\theta\rangle = 0. \tag{9.94}$$

The new operators are obtained from the old ones by a rotation of the form

$$\begin{pmatrix} q^i_{f,1} \\ q^i_{f,2} \end{pmatrix} = \exp\Big\{ \frac{\mathrm{i}}{2} \sigma_2 \theta_i \Big\} \begin{pmatrix} a^i_f \\ b^{i\dagger}_f \end{pmatrix}, \tag{9.95}$$

and this leads to a Hamiltonian [2, 60]

$$\begin{aligned}
H_F = {} & \frac{1}{2} \sum_{j\neq i} \frac{\cos^2\left[\frac{1}{2}(\theta_i - \theta_j)\right]}{(\lambda_i - \lambda_j)^2} (a^{i\dagger}_f a^i_f + b^{i\dagger}_f b^i_f + a^{i\dagger}_f a^{j\dagger}_g a^i_g a^j_f + 2 a^{i\dagger}_f b^{i\dagger}_g a^j_f b^j_g + b^{i\dagger}_f b^{j\dagger}_g b^i_g b^j_f) \\
& + \frac{1}{2} \sum_{j\neq i} \frac{\sin^2\left[\frac{1}{2}(\theta_i - \theta_j)\right]}{(\lambda_i - \lambda_j)^2} (1 - a^{i\dagger}_f a^i_f - b^{i\dagger}_f b^i_f - a^{i\dagger}_f a^{j\dagger}_g b^{i\dagger}_g b^{j\dagger}_f - 2 a^{i\dagger}_f b^{j\dagger}_f a^i_g b^j_g - a^i_f a^j_g b^i_g b^j_f) \\
& + \frac{1}{2} \sum_{j\neq i} \frac{\sin(\theta_i - \theta_j)}{(\lambda_i - \lambda_j)^2} (a^{i\dagger}_f b^{i\dagger}_f - a^i_f b^i_f - a^{i\dagger}_f a^i_g a^{j\dagger}_f b^{j\dagger}_g - a^{i\dagger}_f b^{i\dagger}_g a^{j\dagger}_g a^j_f - a^{i\dagger}_f b^{i\dagger}_g b^{j\dagger}_f b^j_g + b^{i\dagger}_f b^i_g b^j_f a^j_g) \\
& + m \sum_i \cos\theta_i (a^{i\dagger}_f a^i_f + b^{i,\dagger}_f b^i_f - 1) + m \sum_i \sin\theta_i (a^{i\dagger}_f b^{i,\dagger}_f - a^i_f b^i_f) \\
& - g_F \sum_i \frac{\lambda_i}{\sqrt{N}} \sin\theta_i (a^{i\dagger}_f a^i_f + b^{i,\dagger}_f b^i_f - 1) + g_F \sum_i \frac{\lambda_i}{\sqrt{N}} \cos\theta_i (a^{i\dagger}_f b^{i\dagger}_f - a^i_f b^i_f).
\end{aligned} \tag{9.96}$$

This is a complicated Hamiltonian, since it involves quartic operators. However, it can be shown that the quartic operators can be treated as perturbations and give subleading corrections in $1/N$ [2]. For simplicity we will consider the case in which there is one single flavour, as in [2]. The vacuum is simply determined by requiring that the quadratic part of H_F contains no fermion-number-changing operators, so that this part is proportional to the occupation number. This condition was obtained in [6] in the context of the 't Hooft model for QCD_2. The theta angles are then fixed by the condition

$$\frac{1}{2}\sum_{j\neq i}\frac{\sin(\theta_i-\theta_j)}{(\lambda_i-\lambda_j)^2}+m\sin\theta_i+\frac{g_F}{\sqrt{N}}\lambda_i\cos\theta_i=0. \tag{9.97}$$

We now assume that at large N the angles θ_i become functions of the eigenvalue λ, whose distribution $\rho(\lambda)$ is given by (9.53). In other words, we assume that the dynamics of the eigenvalues is given, at large N, by the planar limit of matrix quantum mechanics *without fermions*. This is a consequence of the fact that at large N mesons and glueballs do not mix. We then have,

$$\theta_i\to\theta(\lambda) \tag{9.98}$$

and

$$\sum_i h(\lambda_i)\to N\int \mathrm{d}\lambda\,\rho(\lambda)h(\lambda). \tag{9.99}$$

At large N the equation for the angle (9.97) becomes an integral equation

$$\frac{1}{2}\mathrm{P}\int \mathrm{d}\lambda'\,\rho(\lambda')\frac{\sin\left[\theta(\lambda)-\theta(\lambda')\right]}{(\lambda-\lambda')^2}+m\sin\theta(\lambda)+g_F\lambda\cos\theta(\lambda)=0 \tag{9.100}$$

where we rescaled the eigenvalues λ_i in the way (9.32) appropriate for the large N limit.

Once this equation is solved, we can easily compute the subleading correction (of order N) to the ground state energy as

$$E_F=\langle\theta|H_F|\theta\rangle. \tag{9.101}$$

This is trivial to compute from the normal-ordered Hamiltonian (9.96), since only the constant terms contribute. In the large N limit we obtain

$$\begin{aligned}\frac{E_F}{N}&=\frac{1}{2}\int \mathrm{d}\lambda\,\mathrm{d}\lambda'\,\rho(\lambda)\rho(\lambda')\frac{\sin^2\left[\frac{1}{2}(\theta(\lambda)-\theta(\lambda'))\right]}{(\lambda-\lambda')^2}\\&\quad+\int \mathrm{d}\lambda\,\rho(\lambda)\Big[-m\cos\theta(\lambda)+g_F\lambda\sin\theta(\lambda)\Big].\end{aligned} \tag{9.102}$$

This gives the sum over all planar diagrams with one quark loop at the boundary, as expected from the large N counting rules.

Notice that, for $g_F=0$, the integral equation (9.100) is solved by the trivial solution

$$\theta(\lambda)=0 \tag{9.103}$$

therefore

$$E_F=-mN, \tag{9.104}$$

i.e. we obtain the energy of N free particles of mass m. It is possible to consider the corrections in g_F to this result by studying (9.100), see the Appendix of [2].

Let us now study mesons. Their wavefunctions have the structure

$$|\phi_{\mathrm{M}}\rangle = \sum_i f_i a^{i\dagger} b^{i\dagger} |\theta\rangle, \tag{9.105}$$

since they correspond to $\bar{q}q$ states. At leading order in $1/N$ we find the equation

$$(H_F - E_F)|\phi_{\mathrm{M}}\rangle = E_M |\phi_{\mathrm{M}}\rangle \tag{9.106}$$

Again, the dynamics for the mesons takes place in the background of the master field for the pure "glue" theory.

To analyze (9.106) we consider the fermion-number conserving terms in the Hamiltonian H_F. After subtracting E_F, the quadratic part, proportional to $a^{i\dagger}a^i + b^{i\dagger}b^i$, can be written as

$$\sum_i \left[\frac{1}{2}\sum_{j\neq i} \frac{\cos(\theta_i - \theta_j)}{(\lambda_i - \lambda_j)^2} + m\cos\theta_i - g_F \frac{\lambda_i}{\sqrt{N}}\sin\theta_i\right](a_f^{i\dagger}a_f^i + b_f^{i\dagger}b_f^i) \tag{9.107}$$

where we have combined the terms appearing in the first and second lines of (9.96) as

$$\cos^2\left[\frac{1}{2}(\theta_i - \theta_j)\right] - \sin^2\left[\frac{1}{2}(\theta_i - \theta_j)\right] = \cos(\theta_i - \theta_j). \tag{9.108}$$

The operator (9.107) acting on (9.105) leads to

$$\sum_i \left[\sum_{j\neq i} \frac{\cos^2\left[\frac{1}{2}(\theta_i - \theta_j)\right]}{(\lambda_i - \lambda_j)^2} + 2m\cos\theta_i - 2g_F \frac{\lambda_i}{\sqrt{N}}\sin\theta_i\right] f_i a^{i\dagger} b^{i\dagger}|\theta\rangle \tag{9.109}$$

while the quartic term

$$\sum_i \left[\sum_{j\neq i} \frac{\cos^2\left[\frac{1}{2}(\theta_i - \theta_j)\right]}{(\lambda_i - \lambda_j)^2}\right] a_f^{i\dagger} b_f^{j\dagger} a_f^i b_g^j \tag{9.110}$$

gives

$$-\sum_i \left[\sum_{j\neq i} \frac{\cos^2\left[\frac{1}{2}(\theta_i - \theta_j)\right]}{(\lambda_i - \lambda_j)^2} f_j\right] a^{i\dagger} b^{i\dagger}|\theta\rangle, \tag{9.111}$$

where the $-$ sign arises from anticommutation. The Schrödinger equation for the mesons becomes

$$\left[\sum_{j\neq i} \frac{\cos(\theta_i - \theta_j)}{(\lambda_i - \lambda_j)^2} + 2m\cos\theta_i - 2g_F \frac{\lambda_i}{\sqrt{N}}\sin\theta_i\right] f_i - \sum_{j\neq i} \frac{\cos^2\left[\frac{1}{2}(\theta_i - \theta_j)\right]}{(\lambda_i - \lambda_j)^2} f_j = E_M f_i. \tag{9.112}$$

In the large N limit we have

$$f_i \to f(\lambda) \tag{9.113}$$

and we obtain the integral equation

$$\begin{aligned} &\mathrm{P}\int \frac{\mathrm{d}\lambda'\rho(\lambda')}{(\lambda - \lambda')^2}\left[\cos(\theta(\lambda) - \theta(\lambda'))(f(\lambda) - f(\lambda')) - \sin^2\frac{\theta(\lambda) - \theta(\lambda')}{2} f(\lambda')\right] \\ &+ \left[2m\cos\theta(\lambda) - 2g_F\lambda\sin\theta(\lambda)\right] f(\lambda) = E_{\mathrm{M}} f(\lambda). \end{aligned} \tag{9.114}$$

This determines the meson spectrum at large N. Notice that the meson mass spectrum is smooth at large N, as expected from general large N arguments. One can also show that there is an infinite number of meson states. One way to see this is to solve the Schrödinger equation at large energies. To do this, we introduce

$$\tilde{f}(\lambda) = \rho(\lambda) f(\lambda). \tag{9.115}$$

The integral equation (9.114) becomes

$$-\rho(\lambda)\mathrm{P}\int \frac{\mathrm{d}\lambda' \tilde{f}(\lambda')}{(\lambda-\lambda')^2}\Big[1-\sin^2(\theta(\lambda)-\theta(\lambda'))/2\Big] + \tilde{f}(\lambda)\tilde{q}(\lambda) = E\tilde{f}(\lambda), \tag{9.116}$$

where

$$\tilde{q}(\lambda) = \int \frac{\mathrm{d}\lambda' \rho(\lambda')}{(\lambda-\lambda')^2}\cos\Big[\theta(\lambda)-\theta(\lambda')\Big] + 2m\cos\theta(\lambda) - 2g_F\lambda\sin\theta(\lambda). \tag{9.117}$$

Let us assume that

$$\tilde{f}(\lambda) \sim \mathrm{e}^{\mathrm{i}Eg(\lambda)+\cdots} \tag{9.118}$$

for large E. The integral

$$I(\lambda) = \mathrm{P}\int \frac{\mathrm{d}\lambda' \mathrm{e}^{\mathrm{i}Eg(\lambda')}}{(\lambda-\lambda')^2} \tag{9.119}$$

can be evaluated by the saddle-point method, and the largest contribution comes from $\lambda \sim \lambda'$. The computation gives

$$I(\lambda) \sim -\mathrm{e}^{\mathrm{i}Eg(\lambda)}\pi E|g'(\lambda)| + \mathcal{O}(1/E). \tag{9.120}$$

The term $\sin^2(\theta(\lambda)-\theta(\lambda'))/2$ is subleading in this expansion, and (9.116) becomes

$$\pi E\rho(\lambda)|g'(\lambda)| + \tilde{q}(\lambda) = E + \mathcal{O}(1/E). \tag{9.121}$$

This is solved by

$$g(\lambda) = \pm\frac{1}{\pi}\int_{\lambda_1(e_F)}^{\lambda} \frac{\mathrm{d}\lambda'}{\rho_0(\lambda')}\left[1-\frac{\tilde{q}(\lambda')}{E}\right] + \text{constant}. \tag{9.122}$$

We then find the real solution

$$f(\lambda) \sim \frac{1}{\rho(\lambda)}\sin\left\{\frac{E}{\pi}\int_{\lambda_1(e_F)}^{\lambda} \frac{\mathrm{d}\lambda'}{\rho_0(\lambda')}\left[1-\frac{\tilde{q}(\lambda')}{E}\right] + \phi\right\}. \tag{9.123}$$

Since $\rho(\lambda_2(e_F)) = 0$, the sine function must vanish at the endpoint of the distribution, and this gives

$$E_n = \left[\int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \frac{\mathrm{d}\lambda}{\rho(\lambda)}\right]^{-1}\left[\pi^2 n + \int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \mathrm{d}\lambda\frac{\tilde{q}(\lambda)}{\rho(\lambda)} - \pi\phi\right] + \mathcal{O}(1/n) \tag{9.124}$$

where n is a large number. This gives the meson spectrum at large n and shows that it is asymptotically linear, i.e. at large n the spectrum fits into "Regge trajectories." The number of mesons is infinite, as expected based on general large N arguments. Notice that the slope of the meson spectrum is indeed

$$\pi^2\Big[\int_{\lambda_1(e_F)}^{\lambda_2(e_F)}\frac{\mathrm{d}\lambda}{\rho(\lambda)}\Big]^{-1}=\omega \tag{9.125}$$

where ω was defined in (9.57), and it coincides in this case with the slope of the glueball spectrum.

One can also study baryons at large N in this model, following the ideas in [88], see [60].

Remark 9.4. The model we have just analyzed, matrix quantum mechanics with fermions, is very similar to the two-dimensional version of QCD first analyzed by 't Hooft in the large N expansion [78]. This theory is defined by the Hamiltonian density

$$\mathcal{H}=\frac{1}{2}\mathrm{Tr}E^2+\overline{\Psi}(\mathrm{i}\gamma_1\partial_1+m)\Psi \tag{9.126}$$

and the constraint

$$\partial_1 E=J(x)=-g\Psi^\dagger\Psi. \tag{9.127}$$

One can integrate out E to obtain a quartic Hamiltonian

$$H=-\frac{1}{4}\int\mathrm{d}x\,\mathrm{d}y\,\mathrm{Tr}\big(J(x)|x-y|J(y)\big)+\int\mathrm{d}x\,\overline{\Psi}(\mathrm{i}\gamma_1\partial_1+m)\Psi. \tag{9.128}$$

This fermionic Hamiltonian can be analyzed with the same techniques we have used, see [2] and specially [6] for a detailed study.

10. Applications in QCD

10.1 Chiral symmetry and chiral symmetry breaking

Excellent references for this section are [92], Chapter 7, and [36]. In this subsection we will use the hatted fields defined in (5.9), but for notational simplicity we will remove the hats.

Let us consider the QCD Lagrangian with N_f flavors,

$$\mathcal{L}=\mathrm{i}\sum_{f=1}^{N_f}\bar{q}_f D q_f-\sum_{f=1}^{N_f}m_f\bar{q}_f q_f+\cdots \tag{10.1}$$

We can write this in terms of lef-handed and right-handed components

$$q_{L,f}=\frac{1-\gamma_5}{2}q_f,\qquad q_{R,f}=\frac{1+\gamma_5}{2}q_f \tag{10.2}$$

as follows

$$\mathcal{L}=\mathrm{i}\sum_{f=1}^{N_f}\Big(\bar{q}_{L,f}Dq_{L,f}+\bar{q}_{R,f}Dq_{R,f}\Big)-\sum_{f=1}^{N_f}m_f\Big(\bar{q}_{R,f}q_{L,f}+\bar{q}_{L,f}q_{R,f}\Big)+\cdots \tag{10.3}$$

Let us consider the global symmetry group

$$SU_L(N_f)\times SU_R(N_f) \tag{10.4}$$

acting as

$$q_{L,f} \to \sum_{f'=1}^{N_f} L_{ff'} q_{L,f'}, \qquad q_{R,f} \to \sum_{f'=1}^{N_f} R_{ff'} q_{R,f'}. \tag{10.5}$$

If we denote by T^a the generators of $SU(N_f)$, we can write the rotation matrices as

$$L = \mathrm{e}^{-\mathrm{i}\theta_L^a T^a}, \qquad R = \mathrm{e}^{-\mathrm{i}\theta_L^a T^a}. \tag{10.6}$$

It is also very useful to parametrize this symmetry group in terms of a vectorial and an axial part, i.e.

$$SU_V(N_f) \times SU_A(N_f) \tag{10.7}$$

with an action on the Dirac spinors

$$q \to \mathrm{e}^{-\mathrm{i}\theta_V^a T^a} q, \qquad q \to \mathrm{e}^{-\mathrm{i}\theta_A^a T^a \gamma_5} q \tag{10.8}$$

and angles

$$\theta_V = \frac{1}{2}(\theta_L + \theta_R), \qquad \theta_A = \frac{1}{2}(\theta_L - \theta_R). \tag{10.9}$$

The vectorial part corresponds to the diagonal subgroup of (10.4), while the axial group is the anti-diagonal part. $SU_V(N_f)$ is often called the *isospin* symmetry of QCD. The only the term in the QCD Lagrangian which is not invariant under this symmetry is the term for the quark masses. Therefore, in a world of massless quarks, the above group is a symmetry of QCD. The corresponding conserved currents are

$$J_L^{a\mu} = \sum_{f,f'} \bar{q}_{L,f} T^a_{ff'} \gamma^\mu q_{L,f'}, \qquad J_R^{a\mu} = \sum_{f,f'} \bar{q}_{R,f} T^a_{ff'} \gamma^\mu q_{R,f'}, \tag{10.10}$$

where T^a, $a = 1, \cdots, N_f^2 - 1$ are generators of $SU(N_f)$. Equivalently, we can consider axial and vector currents

$$V_a^\mu = \sum_{f,f'} V^\mu_{ff'} T^a_{ff'}, \qquad A_a^\mu = \sum_{f,f'} A^\mu_{ff'} T^a_{ff'}. \tag{10.11}$$

with

$$V^\mu_{ff'} = \bar{q}_f \gamma^\mu q_{f'}, \qquad A^\mu_{ff'} = \bar{q}_f \gamma^\mu \gamma_5 q_{f'}. \tag{10.12}$$

Notice that the vectorial current corresponds to the diagonal of (10.7). One fundamental, nonperturbative aspect of Nature is that the chiral symmetry (10.7) is *spontaneously broken*. Of course, since this symmetry is only approximate, one has to be careful about this statement, but in any case in a world with massless quarks this seems to be the case. The symmetry breaking pattern is that (10.7) is broken down to the vectorial part,

$$SU_V(N_f) \times SU_A N_f) \to SU_V(N_f). \tag{10.13}$$

In other words, the charges

$$Q_a^5(t) = \int \mathrm{d}^3\vec{x}\, A_a^0(t, \vec{x}) \tag{10.14}$$

do not leave the vacuum invariant. This is called *chiral symmetry breaking* (χSB, in short). χSB is manifested in the fact that the *quark condensate*

$$\langle 0|\bar{q}_f q_f|0\rangle \neq 0 \tag{10.15}$$

in the vacuum (notice that, due to isospin symmetry, this vev is the same for any flavour).

We now recall Goldstone's theorem, which says that, for each generator that fails to annihilate the vacuum, there is a massless boson with the quantum numbers of this generator. In other words, there must be $N_f^2 - 1$ pseudoscalar Goldstone bosons as a consequence of this symmetry breaking. These are the *pions*. Of course, it only makes sense to talk about pions if one only considers light quarks, since we know that chiral symmetry is *explicitly* broken by quark masses. Taking the quarks u, d, s as light, we have eight pions. These are the three π, the four K, and the η_8.

Of course, pions are not massless in the real world, but we would expect their masses to go to zero as the masses of the quarks go to zero. It is possible to use various field-theoretic arguments to find a quantitative expression for this fact. Let us consider the current

$$A^\mu_{ud}(x) = \bar{u}\gamma^\mu\gamma_5 d(x). \tag{10.16}$$

This current is not conserved in the real world where u, d are massive, and its divergence is given by

$$\partial_\mu A^\mu_{ud}(x) = \mathrm{i}(m_u + m_d)\bar{u}\gamma_5 d \tag{10.17}$$

This current has the same quantum numbers as π^+, so we can use it as a composite pion field operator. In other words, if

$$|\pi(p)\rangle \tag{10.18}$$

is the state of a pion with momentum p, we must have

$$\langle 0|A^\mu_{ud}(x)|\pi(p)\rangle = \mathrm{i}p^\mu C_\pi \mathrm{e}^{-\mathrm{i}p\cdot x}. \tag{10.19}$$

where C_π is a constant. This constant is typically parametrized as

$$C_\pi = \frac{\sqrt{2}F_\pi}{(2\pi)^{3/2}\sqrt{2E_p}}, \tag{10.20}$$

where F_π is called the *pion decay constant*. It can be determined experimentally from the weak decay $\pi^+ \to \mu^+\nu$, and one finds

$$F_\pi \sim 93\,\mathrm{MeV}. \tag{10.21}$$

If we introduce the normalized pion field

$$\langle 0|\phi_\pi(x)|\pi(p)\rangle = \frac{1}{(2\pi)^{3/2}\sqrt{2E}}\mathrm{e}^{-\mathrm{i}p\cdot x} \tag{10.22}$$

we can write

$$\partial_\mu A^\mu_{ud}(x) = \sqrt{2}F_\pi m_\pi^2 \phi_\pi(x). \tag{10.23}$$

Using now (10.23) and (10.17) we can obtain a formula for m_π^2 in terms of F_π and m_u, m_d. The basic idea for the formula is the following. If we sandwich (10.23) with the vacuum and a pion state, we find

$$\langle \pi(q)|\partial_\mu A^\mu(0)|0\rangle = \frac{\sqrt{2}F_\pi m_\pi^2}{(2\pi)^{3/2}\sqrt{2E}}, \tag{10.24}$$

where we used (10.22). On the other hand, this equals

$$\langle \pi(q)|\mathrm{i}(m_u + m_d)\bar{u}\gamma_5 d|0\rangle. \tag{10.25}$$

The strategy is to evaluate this correlator to relate m_π^2 to m_u, m_d. To calculate (10.25), we need the soft pion theorem. We follow the short treatment in [36], section IV-5. Chapter 2 of [26] contains a more detailed treatment. We consider the matrix element for the process

$$\alpha \to \beta + \pi(q) \tag{10.26}$$

where $\pi(q)$ is a pion state. The LSZ reduction formula states that

$$\langle \pi(q)\beta|\mathcal{O}(0)|\alpha\rangle = \frac{\mathrm{i}}{(2\pi)^{3/2}\sqrt{2E}} \int \mathrm{d}^4x\, \mathrm{e}^{\mathrm{i}q\cdot x}(m_\pi^2 - q^2)\langle\beta|T\phi_\pi(x)\mathcal{O}(0)|\alpha\rangle. \tag{10.27}$$

We can use (10.23) again to write (10.27) as

$$\frac{\mathrm{i}}{(2\pi)^{3/2}\sqrt{2E}} \frac{m_\pi^2 - q^2}{\sqrt{2}F_\pi^2 m_\pi^2} \int \mathrm{d}^4x\, \mathrm{e}^{\mathrm{i}q\cdot x}\langle\beta|T\partial_\mu A^\mu(x)\mathcal{O}(0)|\alpha\rangle. \tag{10.28}$$

We remind that

$$T\mathcal{O}_1(x)\mathcal{O}_2(0) = \theta(x^0)\mathcal{O}_1(x)\mathcal{O}_2(0) + \theta(-x^0)\mathcal{O}_2(0)\mathcal{O}_2(x) \tag{10.29}$$

therefore

$$\partial^\mu T A_\mu(x)\mathcal{O}(0) = T(\partial^\mu A_\mu(x)\mathcal{O}(0)) + \delta(x^0)[A_0(x), \mathcal{O}(0)] \tag{10.30}$$

which holds as an operator equality. We then find, after integrating by parts,

$$\begin{aligned}\langle \pi(q)\beta|\mathcal{O}(0)|\alpha\rangle &= \frac{\mathrm{i}}{(2\pi)^{3/2}\sqrt{2E}} \frac{m_\pi^2 - q^2}{\sqrt{2}F_\pi^2 m_\pi^2} \\ &\quad\cdot \int \mathrm{d}^4x\, \mathrm{e}^{\mathrm{i}q\cdot x}\Big\{-\delta(x^0)\langle\beta|[A_0(x), \mathcal{O}(0)]|\alpha\rangle - \mathrm{i}q^\mu\langle\beta|TA_\mu(x)\mathcal{O}(0)|\alpha\rangle\Big\}.\end{aligned} \tag{10.31}$$

If we now take the limit as $q \to 0$ of this equation, we find

$$\begin{aligned}\lim_{q\to 0}\langle \pi(q)\beta|\mathcal{O}(0)|\alpha\rangle &= -\frac{\mathrm{i}}{(2\pi)^{3/2}\sqrt{2E}} \frac{1}{\sqrt{2}F_\pi} \int \mathrm{d}^4x\, \delta(x^0)\langle\beta|[A_0(x), \mathcal{O}(0)]|\alpha\rangle \\ &\quad + \lim_{q\to 0} q^\mu R_\mu\end{aligned} \tag{10.32}$$

where

$$R_\mu = -\frac{\mathrm{i}}{(2\pi)^{3/2}\sqrt{2E}} \frac{1}{\sqrt{2}F_\pi} \int \mathrm{d}^4x\, \mathrm{e}^{\mathrm{i}q\cdot x}\langle\beta|TA_\mu(x)\mathcal{O}(0)|\alpha\rangle. \tag{10.33}$$

In our case,

$$\mathcal{O} = \mathrm{i}(m_u + m_d)\bar{u}\gamma_5 d. \tag{10.34}$$

To compute the commutator in (10.32) we can use the general current commutation relations, which are easily derived from the equal time commutation relations of the quark fields (see, for example, [92]),

$$\delta(x^0 - y^0)[A^0_{ff'}(x), q_{f''}(y)] = -\delta(x - y)\delta_{ff''}\gamma_5 q'_f(x). \tag{10.35}$$

Using this, we find,

$$\begin{aligned}
&-\mathrm{i}\delta(x^0)[A^0(x), \partial_\mu A^\mu(0)] \\
&= (m_u + m_d)\Big\{\delta(x^0)[A^0(x), \bar{u}(0)]\gamma_5 d(0) + \bar{u}(0)\gamma_5\delta(x^0)[A^0(x), d(0)]\Big\} \\
&= -(m_u + m_d)\delta(x)\Big\{\bar{u}(0)u(0) + \bar{d}(0)d(0)\Big\}.
\end{aligned} \tag{10.36}$$

We finally obtain,

$$\lim_{q\to 0}\langle\pi(q)|\mathrm{i}(m_u + m_d)\bar{u}\gamma_5 d|0\rangle = -\frac{1}{(2\pi)^{3/2}\sqrt{2E_q}}\frac{m_u + m_d}{\sqrt{2}F_\pi}\langle 0|\Big\{\bar{u}(0)u(0) + \bar{d}(0)d(0)\Big\}|0\rangle. \tag{10.37}$$

By chiral symmetry, and to leading order in the quark masses, we can set

$$\langle 0|\bar{u}(0)u(0)|0\rangle = \langle 0|\bar{d}(0)d(0)|0\rangle \equiv \langle 0|\bar{q}\,q|0\rangle \tag{10.38}$$

and taking into account (10.24) we finally obtain the formula relating the mass of the pion to the masses of the quarks and the χSB order parameter $\langle 0|\bar{q}\,q|0\rangle$:

$$m_\pi^2 = -\frac{m_u + m_d}{F_\pi^2}\langle 0|\bar{q}\,q|0\rangle \tag{10.39}$$

An alternative, elegant derivation of this relation using chiral Lagrangians can be found in Appendix C.

10.2 The $U(1)$ problem

The axial current introduced in (6.93)

$$J^\mu = \frac{1}{g^2}\sum_{f=1}^{N_f}\bar{q}_f\gamma^\mu\gamma_5 q_f, \tag{10.40}$$

is classically conserved. As we mentioned in chapter 7, If this current was conserved quantum-mechanically, there would be an extra conserved quantum number. If it was spontaneously broken, there would be a ninth Goldstone boson, in addition to the other mesons. In Appendix C we show that a chiral Lagrangian for a ninth Goldstone boson η' predicts that its mass will be equal to the pion mass. This derivation assumes that $F_{\eta'} = F_\pi$, but even relaxing this assumption one finds that the ninth Goldstone boson would have a squared mass no larger than $\sqrt{3}m_\pi^2$ [83]. The $U(1)$ *problem* (reviewed in for example [24]) comes from the fact that none of these two things happen: there is no extra conserved quantum number, and the lightest flavour-singlet, pseudoscalar is the η', with a mass of almost 1 GeV. This is far too heavy to be the ninth Goldstone boson.

As we explained in (6.94), the resolution of this problem is that the axial current is anomalous. We can in principle apply to this current and the corresponding η' particle the same type of current algebra arguments that we used before to analyze pions. In other words, if

$$|\eta'(p)\rangle \tag{10.41}$$

is the state of the η' with momentum p, we must have

$$\langle 0|J^\mu(x)|\eta'(p)\rangle = \mathrm{i}p^\mu\frac{F_{\eta'}}{(2\pi)^{3/2}\sqrt{2E_p}}\mathrm{e}^{-\mathrm{i}p\cdot x}. \tag{10.42}$$

It follows that

$$\langle 0|\partial_\mu J^\mu(0)|\eta'(p)\rangle = \frac{F_{\eta'} m_{\eta'}^2}{(2\pi)^{3/2}\sqrt{2E}} \tag{10.43}$$

is equal to

$$\langle 0|\Big\{2N_f q(x) + \frac{2\mathrm{i}}{g^2}\sum_{f=1}^{N_f} m_f \bar{q}_f \gamma_5 q_f\Big\}|\eta'(p)\rangle \tag{10.44}$$

If there was no anomalous term here, one could deduce that, as for the true Goldstone bosons,

$$m_{\eta'}^2 = \mathcal{O}(m_f). \tag{10.45}$$

The anomalous term, although being a total derivative, has nontrivial effects, and in principle gives a mass to the η'. We will be able to quantify this effect in the context of the $1/N$ expansion of QCD.

An important consequence of the anomalous $U(1)$ is that, in a world with massless quarks, the theta dependence in the QCD path integral disappears. One quick way to see this is by doing a change of variables in the path integral

$$q_f' = q_f + \mathrm{i}\alpha\gamma_5 q_f, \qquad \bar{q}_f' = \bar{q}_f + \bar{q}_f \mathrm{i}\gamma_5\alpha \tag{10.46}$$

which is just a chiral rotation with arbitrary angle α. The fermion measure in the path integral changes as [41]

$$\mathcal{D}q\,\mathcal{D}\bar{q} = \mathcal{D}q'\,\mathcal{D}\bar{q}' \exp\Big\{2N_f\alpha\int \mathrm{d}^4x\, q(x)\Big\}. \tag{10.47}$$

Precisely because the quarks are massless there is no other change in the path integral. Therefore, after this change of variables we have that

$$Z(\theta) = Z(\theta + 2N_f\alpha), \tag{10.48}$$

where α is an arbitrary angle. In other words, the partition function (and any other observable) will be independent of θ.

10.3 The $U(1)$ problem at large N. Witten–Veneziano formula

As we explained above, the $U(1)_A$ flavor symmetry is broken by anomalies, so the η' is not a Goldstone boson. To understand this more quantitatively, we should understand how the anomaly gives a mass to the η'. This was solved by Witten [87] (and subsequently by Veneziano [81]) by using large N techniques, who obtained a remarkable expression for the mass of the η' at leading order in $1/N$ known as the Witten–Veneziano formula. This formula has been spectacularly confirmed by lattice gauge theory calculations [34].

A first observation one can do is that the anomalous contribution to the divergence of the $U(1)_A$ current *vanishes* in the large N limit (8.3), since one has from (6.94) that

$$\partial_\mu J^\mu = \frac{2N_f}{N}\frac{t}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}(\hat{F}^{\mu\nu}, \hat{F}^{\rho\sigma}) \tag{10.49}$$

when expressed in terms of normalized fields (8.6). Therefore, at large N the η' is a true Goldstone boson, and we can regard $1/N$ as a symmetry breaking parameter.

To derive the Witten–Veneziano formula, let us come back to (10.24)-(10.25). To compute $m^2_{\eta'}$ we have to compute

$$\langle \eta'(q)|q(x)|0\rangle. \tag{10.50}$$

To have a handle on this, we study the two point function (5.28), evaluated in the theory with quarks. This function can be organized as

$$U(k) = \sum_{L\geq 0} U_L(k), \tag{10.51}$$

where L denotes the number of quark loops and $U_L(k)$ is the contribution to $U(k)$ of diagrams with L quark loops. We know from the general rules of the $1/N$ expansion that (see for example (8.43))

$$U_0(k) \sim \mathcal{O}(N^0), \qquad U_1(k) \sim \mathcal{O}(N^{-1}). \tag{10.52}$$

In fact, we can get more precise information about these functions. At leading order in $1/N$, the only singularities of two-point functions of gauge-invariant operators are meson and glueball poles,

$$U(k) = \sum_{G_i} \frac{a_n^2}{k^2 - m_i^2} + \sum_{M_i} \frac{c_i^2}{N(k^2 - m_i^2)} \tag{10.53}$$

where the first sum is over glueball states, and the second sum is over meson states. In this equation,

$$a_i = (2\pi)^{3/2}\sqrt{2E}\langle 0|q(0)|G_i\rangle, \qquad \frac{c_i}{\sqrt{N}} = (2\pi)^{3/2}\sqrt{2E}\langle 0|q(0)|M_i\rangle, \tag{10.54}$$

see for example (A.10). We have already extracted the leading N dependence as it follows from (8.49), so that c_i, a_i are of order one. We have that, at leading order in the $1/N$ expansion,

$$U_0(k) \sim \sum_{G_i} \frac{a_n^2}{k^2 - m_i^2}, \qquad U_1(k) \sim \sum_{M_i} \frac{c_i^2}{N(k^2 - m_i^2)}. \tag{10.55}$$

We know that, in a world of massless quarks, there is no θ dependence, and the topological susceptibility vanishes. But

$$\chi_t = U(0). \tag{10.56}$$

Therefore, the contributions from quark loops to $U(k)$ must *cancel* the contributions from gluons at $k = 0$. This seems difficult to achieve from the standpoint of the large N expansion, since $U_0(k)$ is of order $\mathcal{O}(N^0)$, and $U_1(k)$ is of order $1/N$. As pointed out by Witten, this cancellation can happen at $k = 0$ *if there is a pseudoscalar, flavor single meson (so that it contributes to c_i) whose mass squared is of order* $1/N$. Let us call this meson the η'. If this is the case, the term

$$\frac{c^2_{\eta'}}{N(k^2 - m^2_{\eta'})} \tag{10.57}$$

in the sum over meson resonances becomes at $k = 0$

$$-\frac{c_{\eta'}^2}{N m_{\eta'}^2} \sim \mathcal{O}(N^0) \tag{10.58}$$

and can kill the glueball contribution. Notice that this contribution is precisely

$$U_0(0) \tag{10.59}$$

at leading order in N. We deduce

$$\frac{c_{\eta'}^2}{N m_{\eta'}^2} = U_0(0). \tag{10.60}$$

We can now put everything together to deduce a formula for $m_{\eta'}^2$. We have from (10.43) that

$$\frac{F_{\eta'} m_{\eta'}^2}{(2\pi)^{3/2}\sqrt{2E}} = \langle \eta'(p)|\partial_\mu J^\mu(0)|0\rangle = 2N_f \langle \eta'(p)|q(0)|0\rangle = \frac{1}{(2\pi)^{3/2}\sqrt{2E}} \frac{2N_f c_{\eta'}}{\sqrt{N}}, \tag{10.61}$$

in other words,

$$c_{\eta'} = \frac{\sqrt{N}}{2N_f} F_{\eta'} m_{\eta'}^2. \tag{10.62}$$

Plugging this into (10.60) we find

$$m_{\eta'}^2 = \frac{4N_f^2}{F_{\eta'}^2} U_0(0). \tag{10.63}$$

After an appropriate normalization, $F_{\eta'}$ equals F_π at leading order in the $1/N$ expansion,

$$F_{\eta'} = \sqrt{2N_f} F_\pi. \tag{10.64}$$

This follows from the full chiral symmetry $U_L(N_f) \times U_R(N_f)$ at large N. We then obtain the Witten–Veneziano formula in the form

$$m_{\eta'}^2 = \frac{2N_f}{F_\pi^2} \chi_t^{\rm YM} \tag{10.65}$$

where $\chi_t^{\rm YM}$ is the topological susceptibility *in pure gluodynamics*. In principle, $\chi_t^{\rm YM}$ vanishes order by order in perturbation theory, since $(F, \tilde{F})$ is a total divergence, so its matrix elements vanish at zero momentum, as we explained in section 7.2. But it might happen that the sum of *all* planar diagrams does not vanish at $k = 0$. This is indeed what happens in the $\mathbb{P}^N$ sigma model, as we showed before following [86, 30]. The consistent picture of the η' developed by Witten in [87] requires that this is also the case in QCD.

The formula (10.65) for the mass of the η' in the world with $N_f = 3$ actually assumes that there is no mixing with the other mesons. A more refined analysis can be done by taking into account the detailed structure of the chiral Lagrangian [81]. The mass matrix $\mathcal{M}_{\eta-\eta'}$ for the η, η' in the approximation $m_u = m_d$ is written down in (B.38). The diagonalization of this matrix leads to the masses m_η^2, $m_{\eta'}^2$. Since the trace is a unitary invariant, we find

$$\mathrm{Tr}\,\mathcal{M}_{\eta-\eta'} = m_\eta^2 + m_{\eta'}^2 = 2m_K^2 + \frac{6}{F_\pi^2}\chi_t^{\rm YM} \tag{10.66}$$

where we set $N_f = 3$. This leads to a surprising relation between the topological susceptibility of *pure* Yang–Mills theory (i.e. in the theory *without* quarks) and the meson masses,

$$\chi_t^{\rm YM} = \frac{F_\pi^2}{6}\Big(m_{\eta'}^2 + m_\eta^2 - 2m_{\rm K}^2\Big). \tag{10.67}$$

Interestingly, the Witten–Veneziano solution of the $U(1)$ problem the large N limit does not involve instantons, as originally proposed by 't Hooft. According to the argument put forward by Witten, the topological susceptibility of pure Yang–Mills theory is indeed nonzero in the full nonperturbative theory, but this nonzero value is not due to instantons: it shows up already in the large N expansion and it is due to an infinite sum of planar diagrams.

Fortunately, recen lattice calculations have been able to determine $\chi_t^{\rm YM}$ for $N = 3$. One finds [34]

$$\chi_t^{\rm YM} = (191 \pm 5\,\mathrm{MeV})^4 \tag{10.68}$$

On the other hand, by plugging the experimental values of the pion masses in (10.67) we get

$$\frac{F_\pi^2}{6}\Big(m_{\eta'}^2 + m_\eta^2 - 2m_{\rm K}^2\Big) \approx (180\,\mathrm{MeV})^4. \tag{10.69}$$

This is a quite remarkable qualitative agreement, since after all the Witten–Veneziano formula is only supposed to be valid at leading order in the $1/N$ expansion and in a world with massless quarks. The explicit computation of the topological susceptibility also suggests that it is not captured by instanton configurations [46].

A. Polology and spectral representation

Let us consider a general correlation function in momentum space

$$G(q_1, \cdots, q_n) = \int \mathrm{d}^4x_1 \cdots \mathrm{d}^4x_n\, \mathrm{e}^{-\mathrm{i}q_1\cdot x_1} \cdots \mathrm{e}^{-\mathrm{i}q_n\cdot x_n} \langle A_1(x_1)\cdots A_n(x_n)\rangle. \tag{A.1}$$

The analytic structure of this function in momentum space is quite complicated. Following [84], let us consider this as a function of q^2, where

$$q = q_1 + \cdots + q_r = -q_{r-1} - \cdots - q_n \tag{A.2}$$

and $1 \le r \le n-1$. A general nonperturbative result in QFT says that G has a pole at

$$q^2 = -m^2 \tag{A.3}$$

where m^2 is the mass of any one-particle state that has nonvanishing matrix elements with the states

$$A_1^\dagger \cdots A_r^\dagger|0\rangle, \qquad A_{r+1}\cdots A_n|0\rangle. \tag{A.4}$$

The pole has the structure

$$-\frac{2\mathrm{i}\sqrt{\vec{q}^2+m^2}}{q^2+m^2-\mathrm{i}\epsilon}(2\pi)^7\delta^4(q_1+\cdots+q_n)\sum_\sigma M_{0|\vec{q},\sigma}(q_2,\cdots,q_r)M_{\vec{q},\sigma|0}(q_{r+2},\cdots,q_n) \tag{A.5}$$

where the Ms are defined by

$$\begin{aligned}
&(2\pi)^4\delta^4(q_1+\cdots+q_r-p)M_{0|\vec{q},\sigma}(q_2,\cdots,q_r)\\
&=\int \mathrm{d}^4x_1\cdots\mathrm{d}^4x_r\,\mathrm{e}^{-\mathrm{i}q_1\cdot x_1}\cdots\mathrm{e}^{-\mathrm{i}q_n\cdot x_n}\langle 0|A_1(x_1)\cdots A_r(x_r)|\vec{p},\sigma\rangle,\\
&(2\pi)^4\delta^4(q_{r+1}+\cdots+q_n-p)M_{\vec{p},\sigma|0}(q_{r+1},\cdots,q_n)\\
&=\int \mathrm{d}^4x_{r+1}\cdots\mathrm{d}^4x_n\,\mathrm{e}^{-\mathrm{i}q_{r+1}\cdot x_{r+1}}\cdots\mathrm{e}^{-\mathrm{i}q_n\cdot x_n}\langle\vec{p},\sigma|A_{r+1}(x_{r+1})\cdots A_n(x_n)|0\rangle.
\end{aligned} \tag{A.6}$$

Therefore, we can read the residue at the pole from (A.5). The factors

$$(2\pi)^{3/2}\left[2\sqrt{\vec{k}^2+m^2}\right]^{1/2} \tag{A.7}$$

in (A.5) just serve to remove kinematic factors associated with the mass m external line in $M_{\vec{p},\sigma|0}$ and $M_{0|\vec{p},\sigma}$. However, on top of the poles associated to one-particle states, G will have branch cuts associated to multi-particle states in the spectrum.

The above result for the structure of $G(q_1,\cdots,q_r)$ is what we would expect from a Feynman diagram with a single internal line for a particle of mass m connecting the first r and the last $n-r$ external lines. However, the particle of mass m is not necessarily an elementary field appearing in the Lagrangian. Rather, if we consider the Feynman diagrams that contribute to $G(q_1,\cdots,q_r)$, we will find diagrams like the one shown in Fig. 47, with two internal lines associated to elementary particles which interact through some other particle. The pole would be in that case be due to a bound state made of the two elementary particles.

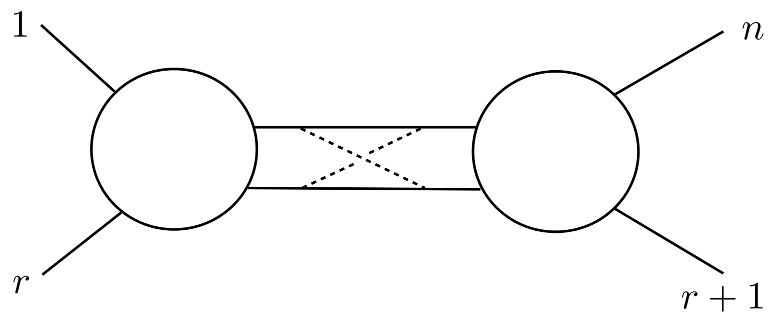


Figure 47: A Feynman diagram contributing to $G(q_1,\cdots,q_n)$. The pole is due to a bound state of two elementary particles, represented as an intermediate channel of two elementary particles interacting through the dashed lines.

As a particular case of the above, let us consider a complex scalar operator $\Phi(x)$, and the two-point function

$$\langle\Phi(x)\Phi^\dagger(y)\rangle. \tag{A.8}$$

Let $|\vec{p}, \sigma\rangle$ be a one-particle state of mass m which has a non-vanishing matrix element with

$$\langle 0|\Phi(0). \tag{A.9}$$

Therefore, we have that

$$\langle 0|\Phi(0)|\vec{p}, \sigma\rangle = (2\pi)^{-3/2}\Big(2\sqrt{\vec{p}^2 + m^2}\Big)^{-1/2} N_\sigma \tag{A.10}$$

where N is a constant (this follows from Lorentz invariance, and it features in the mode expansion of quantized scalar fields). Translation invariance implies that

$$\langle 0|\Phi(x)|\vec{p}, \sigma\rangle = \mathrm{e}^{\mathrm{i}p\cdot x}\langle 0|\Phi(0)|\vec{p}, \sigma\rangle \tag{A.11}$$

therefore in this case

$$M_{0|\vec{p},\sigma} = (2\pi)^{-3/2}\Big(2\sqrt{\vec{p}^2 + m^2}\Big)^{-1/2} N_\sigma, \qquad M_{\vec{p},\sigma|0} = (2\pi)^{-3/2}\Big(2\sqrt{\vec{p}^2 + m^2}\Big)^{-1/2} N_\sigma^*, \tag{A.12}$$

and the above result implies that the momentum space function

$$-\mathrm{i}\Delta(q) = \langle \Phi(q)\Phi^\dagger(-q)\rangle = \int \mathrm{d}^4x \, \exp^{-\mathrm{i}q\cdot(x-y)}\langle 0|\Phi(x)\Phi^\dagger(y)|0\rangle \tag{A.13}$$

has a pole at

$$q^2 = -m^2 \tag{A.14}$$

with residue

$$Z = |N_\sigma|^2 \tag{A.15}$$

This result can be rephrased in yet another way by using the Käller-Lehman representation of the propagator,

$$\Delta'(p) = \int_0^\infty \rho(\mu^2)\frac{\mathrm{d}\mu^2}{p^2 + \mu^2 - \mathrm{i}\epsilon}. \tag{A.16}$$

The existence of a pole in the propagator at $q^2 = -m^2$ and with residue (A.15) means that, near $\mu^2 = m^2$ we have

$$\rho(\mu^2) = Z\delta(\mu^2 - m^2) + \sigma(\mu^2), \tag{A.17}$$

where $\sigma(mu^2)$ is the contribution of multi-particle states.

B. Chiral Lagrangians

The chiral Lagrangian is a particular example of the effective theory of Goldstone bosons that one can obtain in theories with spontaneously broken global symmetries, see [85] for a detailed treatment. We will here collect some basic facts which are useful.

Chiral $SU(N_f)_V \times SU(N_f)_A$ symmetry acts on the quark fields as

$$q \to \exp\Big[\mathrm{i}\sum_a \theta_a^V T^a + \theta_a^A T^a \gamma_5\Big] q. \tag{B.1}$$

In order to write a Lagrangian for Goldstone bosons, we rewrite the quark fields as

$$q = \exp\Big(-\mathrm{i}\gamma_5 \sum_a \xi_a T^a\Big)\tilde{q}, \tag{B.2}$$

where ξ_a are the Goldstone bosons associated to the broken axial symmetry. The new quark fields $\tilde{q}$ transform only under the unbroken, vectorial symmetry, i.e.

$$\tilde{q}' = \exp\Big(\mathrm{i}\sum_a \theta_a T^a\Big)\tilde{q}, \tag{B.3}$$

and this imposes on the Goldstone bosons the transformation rule

$$\exp\Big[\mathrm{i}\sum_a \theta_a^V T^a + \theta_a^A T^a\gamma_5\Big]\exp\Big(-\mathrm{i}\gamma_5\sum_a \xi_a T^a\Big) = \exp\Big(-\mathrm{i}\gamma_5\sum_a \xi'_a T^a\Big)\exp\Big(\mathrm{i}\sum_a \theta_a T^a\Big). \tag{B.4}$$

In terms of left and right moving angles, we find

$$\begin{aligned}\exp\Big(\mathrm{i}\sum_a \theta_a^L T^a\Big)\exp\Big(-\mathrm{i}\sum_a \xi_a T^a\Big) &= \exp\Big(-\mathrm{i}\sum_a \xi'_a T^a\Big)\exp\Big(\mathrm{i}\sum_a \theta_a T^a\Big),\\ \exp\Big(\mathrm{i}\sum_a \theta_a^R T^a\Big)\exp\Big(\mathrm{i}\sum_a \xi_a T^a\Big) &= \exp\Big(\mathrm{i}\sum_a \xi'_a T^a\Big)\exp\Big(\mathrm{i}\sum_a \theta_a T^a\Big),\end{aligned} \tag{B.5}$$

where

$$\theta_a^L = \theta_a^V + \theta_a^A, \qquad \theta_a^R = \theta_a^V - \theta_a^A. \tag{B.6}$$

This means that

$$U = \exp\Big(-2\mathrm{i}\sum_a \xi_a T^a\Big) \tag{B.7}$$

transforms as

$$U' = \exp\Big(\mathrm{i}\sum_a \theta_a^L T^a\Big)U\exp\Big(-\mathrm{i}\sum_a \theta_a^R T^a\Big). \tag{B.8}$$

i.e. it belongs to the representation $(\overline{N}_f, N_f)$ of $SU(N_f)_R \times SU(N_f)_L$. We will write the chiral Lagrangian for the Goldstone bosons in terms of U. It has the form

$$\mathcal{L} = \frac{F^2}{4}\mathrm{Tr}\Big[\partial_\mu U \partial^\mu U\Big] \tag{B.9}$$

at leading order in derivatives.

One thing we want to do with this Lagrangian is to calculate matrix elements of currents of the underlying, microscopic Lagrangian. To do this, we add sources to the microscopic Lagrangian:

$$\begin{aligned}\mathcal{L}_{\mathrm{QCD}}(\ell, r, s, p) = \mathrm{i}\sum_{f=1}^{N_f}\Big(\bar{q}_{L,f}Dq_{L,f} + \bar{q}_{R,f}Dq_{R,f}\Big) - \bar{q}_L(s+\mathrm{i}p)q_R - \bar{q}_R(s+\mathrm{i}p)^\dagger q_L\\ - \bar{q}\gamma_\mu\frac{1+\gamma_5}{2}\ell^\mu q - \bar{q}\gamma_\mu\frac{1-\gamma_5}{2}\ell^\mu q,\end{aligned} \tag{B.10}$$

where ℓ_μ, r_μ, s, p are $N_f \times N_f$ matrices. The standard QCD Lagrangian is obtained by setting

$$\ell_\mu, = r_\mu = p = 0, \qquad s = m \tag{B.11}$$

where m is the quark mass matrix. It is clear that an insertion of the bilinear $\bar{q}q$ is obtained by taking a derivative of the free energy w.r.t. the source s^0, where s^0 is the component of s multiplying the identity matrix. In terms of the effective Lagrangian, we have

$$\langle \bar{q}q \rangle = -\frac{\delta \mathcal{L}_{\rm eff}}{\delta s^0} \tag{B.12}$$

To see how the sources appear in the effective Lagrangian, we *gauge* the chiral symmetry and we promote the sources to gauge fields [44]. Under the transformations

$$q_L \to L(x) q_L, \quad q_R \to R(x) q_R \tag{B.13}$$

we have, as we just have seen,

$$U \to L(x) U R^\dagger(x). \tag{B.14}$$

Ths sources ℓ_μ, r_μ behave as gauge potentials for L, R, respectively, and they transform as

$$\begin{aligned} \ell_\mu &\to L(x) \ell_\mu L^\dagger(x) + \mathrm{i}(\partial_\mu L)(x) L^\dagger(x), \\ r_\mu &\to R(x) r_\mu R^\dagger(x) + \mathrm{i}(\partial_\mu R)(x) R^\dagger(x). \end{aligned} \tag{B.15}$$

The Lagrangian (B.10) is now gauge invariant under the gauged chiral symmetry. The goal is to construct a gauge-invariant low-energy Lagrangian. The gauge covariant derivative acting on U is

$$D_\mu U = \partial_\mu U + \mathrm{i}\ell_\mu U - \mathrm{i} U r_\mu, \tag{B.16}$$

and transforms covariantly

$$D_\mu U \to L(x) D_\mu U R^\dagger(x). \tag{B.17}$$

It is easy to construct an effective Lagrangian which is invariant under the gauged symmetry. At leading order in the derivative expansion it is just

$$\mathcal{L} = \frac{F_\pi^2}{4} \mathrm{Tr}(D_\mu U D^\mu U^\dagger) + \frac{F_\pi^2}{4} \mathrm{Tr}(\chi U^\dagger + U \chi^\dagger), \tag{B.18}$$

where

$$\chi = 2B_0(s + \mathrm{i}p) \tag{B.19}$$

and B_0 is a constant. If we now evaluate this for (B.11) we can read off the masses of the pions. First we write

$$2 \sum_a \xi_a T^a = \frac{\sqrt{2}}{F_\pi} \mathcal{B}, \tag{B.20}$$

where

$$\mathcal{B} = \begin{pmatrix} \frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta_8 & \pi^+ & K^+ \\ \pi^- & -\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta_8 & K^0 \\ K^- & \overline{K}^0 & -\sqrt{\frac{2}{3}}\eta_8 \end{pmatrix} \tag{B.21}$$

With this representation, the kinetic term of the Lagrangian is canonically normalized

$$\begin{aligned} &\frac{F_\pi^2}{4} \mathrm{Tr}(D_\mu U D^\mu U^\dagger) \\ &= \frac{1}{2}\partial_\mu \pi^0 \partial^\mu \pi^0 + \partial_\mu \pi^+ \partial_\mu \pi^- + \partial_\mu K^+ \partial_\mu K^- + \partial_\mu K^0 \partial_\mu \overline{K}^0 + \frac{1}{2}\partial_\mu \eta_8 \partial^\mu \eta_8. \end{aligned} \tag{B.22}$$

The quark mass matrix reads

$$\mathcal{M} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{pmatrix} \tag{B.23}$$

and

$$U + U^\dagger = 2 - \frac{2}{F_\pi^2}\mathcal{B}^2 + \cdots . \tag{B.24}$$

The mass term in the effective Lagrangian is

$$\begin{aligned} -B_0 \mathrm{Tr}(\mathcal{B}^2\mathcal{M}) = -B_0 \Big\{ & m_u\Big(\big(\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta_8\big)^2 + \pi^+\pi^- + K^+K^-\Big) \\ & + m_d\Big(\big(-\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta_8\big)^2 + \pi^+\pi^- + K^0\overline{K}^0\Big) \\ & + m_s\Big(K^-K^+ + K^0\overline{K}^0 + \frac{2}{3}\eta_8^2\Big)\Big\} \end{aligned} \tag{B.25}$$

and from here we can read the masses of the Goldstone bosons:

$$\begin{aligned} m_\pi^2 &= B_0(m_u + m_d), \\ m_{K^\pm}^2 &= B_0(m_u + m_s), \\ m_{K^0}^2 &= B_0(m_d + m_s), \\ m_{\eta_8}^2 &= B_0\frac{m_u + m_d + 4m_s}{3}. \end{aligned} \tag{B.26}$$

There is also a mixing term between the η_8 and the π^0,

$$m_{\pi\eta}^2 = B_0\frac{m_u - m_d}{\sqrt{3}}. \tag{B.27}$$

One particular prediction of chiral symmetry, following from (B.26), is that

$$m_{\eta_8}^2 = \frac{1}{3}(2m_{K^\pm}^2 + 2m_{K^0}^2 - m_\pi^2) \tag{B.28}$$

which is called the *Gell-Mann–Okubo mass formula*. Taking as data the masses of the kaons and the pions, it predicts

$$m_{\eta_8}^2 = 566 \text{ MeV} \tag{B.29}$$

which is not far from the experimental value 549 MeV.

We can also relate B_0 to the quark condensate by using the relation (B.12). We have

$$-\frac{\partial \mathcal{L}}{\partial s_{ff'}^0} = -\frac{\partial}{\partial s_{ff'}^0}\frac{F_\pi^2}{4}\mathrm{Tr}(\chi U^\dagger + U\chi^\dagger) = -\frac{F_\pi^2 B_0}{2}(U_{ff'}^\dagger + U_{f'f}) \tag{B.30}$$

Evaluating this in the vacuum $U = \mathbf{1}$, we obtain

$$\langle 0|\bar{q}_f q_f'|0\rangle = -F_\pi^2 B_0 \delta_{ff'}. \tag{B.31}$$

Expressing B_0 in terms of the quark condensate, we recover from (B.26) the relation (10.39).

We can now formulate the $U(1)$ problem in the language of chiral Lagrangians. Let us assume that the axial $U(1)$ is spontaneously broken, so that we enlarge U with an extra Goldstone boson

$$\mathcal{B} + \frac{1}{\sqrt{3}}\eta_0 \mathbf{1}. \tag{B.32}$$

Under the axial $U(1)$, ζ transforms as required,

$$\zeta \to e^{i\theta^L} \eta_0 e^{-i\theta^R} = e^{2i\theta^A}\eta_0. \tag{B.33}$$

The new mass term in the effective Lagrangian is

$$\begin{aligned} - B_0\Big\{ m_u \Big(\big(\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta_8 + \frac{1}{\sqrt{3}}\eta_0\big)^2 + \pi^+\pi^- + K^+K^- \Big) \\ + m_d \Big(\big(-\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta_8 + \frac{1}{\sqrt{3}}\eta_0\big)^2 + \pi^+\pi^- + K^0\overline{K}^0 \Big) \\ + m_s \Big(K^-K^+ + K^0\overline{K}^0 + \big(\frac{1}{\sqrt{3}}\eta_0 - \sqrt{\frac{2}{3}}\eta_8\big)^2 \Big) \Big\} \end{aligned} \tag{B.34}$$

This leads to a mixing matrix for the neutral mesons π^0, η_8, η_0 (listed in this order) given by

$$B_0 \begin{pmatrix} m_u + m_d & \frac{1}{\sqrt{3}}(m_u - m_d) & \sqrt{\frac{2}{3}}(m_u - m_d) \\ \frac{1}{\sqrt{3}}(m_u - m_d) & \frac{1}{3}(m_u + m_d + 4m_s) & \frac{\sqrt{2}}{3}(m_u + m_d - 2m_s) \\ \sqrt{\frac{2}{3}}(m_u - m_d) & \frac{\sqrt{2}}{3}(m_u + m_d - 2m_s) & \frac{2}{3}(m_u + m_d + m_s) \end{pmatrix} \tag{B.35}$$

Let us continue the analysis assuming for simplicity that $m_u = m_d$. This eliminates all mixings but η_8–η_0, which leads to a matrix

$$\begin{pmatrix} \frac{4}{3}m_K^2 - \frac{1}{3}m_\pi^2 & -\frac{2\sqrt{2}}{3}(m_K^2 - m_\pi^2) \\ -\frac{2\sqrt{2}}{3}(m_K^2 - m_\pi^2) & \frac{2}{3}m_K^2 + \frac{1}{3}m_\pi^2 \end{pmatrix} \tag{B.36}$$

The eigenvalues of this matrix are

$$m_\pi^2, \qquad 2m_K^2 - m_\pi^2. \tag{B.37}$$

Therefore, if the $U(1)$ anomaly was spontaneously broken, there will be an extra isoscalar state degenerate in mass with the pion. Even including more general values for the parameters, it can be shown that the extra Goldstone boson must have a mass squared of less than $\sqrt{3}m_\pi^2$ [83]. Using however the Witten–Veneziano formula, we see that the above matrix gets an extra contribution due to the anomaly [81],

$$\mathcal{M}_{\eta-\eta'} = \begin{pmatrix} \frac{4}{3}m_K^2 - \frac{1}{3}m_\pi^2 & -\frac{2\sqrt{2}}{3}(m_K^2 - m_\pi^2) \\ -\frac{2\sqrt{2}}{3}(m_K^2 - m_\pi^2) & \frac{2}{3}m_K^2 + \frac{1}{3}m_\pi^2 + \frac{\epsilon}{N_c} \end{pmatrix} \tag{B.38}$$

where

$$\frac{\epsilon}{N_c} = \frac{2N_f}{F_\pi^2}\chi_t^{\mathrm{YM}}. \tag{B.39}$$

C. Effective action for large N sigma models

Here we compute the large N, effective propagators (7.30) and (7.85).

We start by calculating (7.30). Using the standard trick of introducing Feynman parameters (see for example [69], p. 189), we find

$$\begin{aligned}\widetilde{\Gamma}^{\lambda}(p) &= \int \frac{\mathrm{d}^2 q}{(2\pi)^2} \int_0^1 \mathrm{d}x \, \frac{1}{\Big[x(m^2+q^2)+(1-x)((p+q)^2+m^2)\Big]^2} \\ &= \int \frac{\mathrm{d}^2 q}{(2\pi)^2} \int_0^1 \mathrm{d}x \, \frac{1}{\Big[m^2+q^2+xp^2+2xp\cdot q\Big]^2} \\ &= \int \frac{\mathrm{d}^2 \ell}{(2\pi)^2} \int_0^1 \mathrm{d}x \, \frac{1}{\Big[m^2+\ell^2+x(1-x)p^2\Big]^2}\end{aligned} \tag{C.1}$$

where we have introduced

$$\ell = q + xp \Rightarrow q^2 + 2xp\cdot q = \ell^2 - x^2 p^2. \tag{C.2}$$

We end up with

$$\widetilde{\Gamma}^{\lambda}(p) = \int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} \frac{1}{(\ell^2+\Delta)^2}, \quad \Delta = m^2 + x(1-x)p^2. \tag{C.3}$$

We now recall the standard formula in dimensional regularization (see for example [69], p. 250),

$$\int \frac{\mathrm{d}^d\ell}{(2\pi)^d} \frac{1}{(\ell^2+\Delta)^2} = \frac{1}{(4\pi)^{d/2}} \Gamma(2-d/2)\Delta^{d/2-2}. \tag{C.4}$$

In our case $d=2$ and the integral is convergent, and we simply find

$$\widetilde{\Gamma}^{\lambda}(p) = \int_0^1 \mathrm{d}x \frac{1}{4\pi\Delta} = \frac{1}{4\pi} \int_0^1 \frac{\mathrm{d}x}{m^2+x(1-x)p^2}. \tag{C.5}$$

This integral is elementary. We write the denominator as

$$-p^2(x-a)(x-b), \tag{C.6}$$

where

$$\begin{aligned} a &= \frac{1}{2} - \frac{1}{2}\sqrt{1+\frac{4m^2}{p^2}}, \\ b &= \frac{1}{2} + \frac{1}{2}\sqrt{1+\frac{4m^2}{p^2}}. \end{aligned} \tag{C.7}$$

The integral reads now

$$\begin{aligned} \frac{1}{4\pi}\int_0^1 \frac{\mathrm{d}x}{m^2+x(1-x)p^2} &= -\frac{1}{4\pi p^2}\frac{1}{a-b}\Big[\log(x-a)-\log(b-x)\Big]_0^1 \\ &= \frac{1}{4\pi p^2}\frac{1}{b-a}\log\frac{b(1-a)}{a(1-b)} = \frac{1}{2\pi p^2}\frac{1}{b-a}\log\Big(\frac{b}{a}\Big) \\ &= \frac{1}{2\pi\sqrt{p^2(p^2+4m^2)}}\log\frac{\sqrt{p^2+4m^2}+\sqrt{p^2}}{\sqrt{p^2+4m^2}-\sqrt{p^2}}, \end{aligned} \tag{C.8}$$

where we used $1-a=b$, $1-b=a$. Therefore, we find

$$\widetilde{\Gamma}^\lambda(p) = f(p) \equiv \frac{1}{2\pi\sqrt{p^2(p^2+4m^2)}} \log \frac{\sqrt{p^2+4m^2}+\sqrt{p^2}}{\sqrt{p^2+4m^2}-\sqrt{p^2}}. \tag{C.9}$$

We now compute $\widetilde{\Gamma}^A_{\mu\nu}(p)$. Both integrals appearing in (7.85) are divergent, but their divergences cancel. This is easily seen in dimensional regularization. We first massage the last piece in the second integral, after doing the change of variables (C.2). We find,

$$\begin{aligned}
&\int \frac{\mathrm{d}^2 q}{(2\pi)^2} \frac{(p_\mu+2q_\mu)(p_\nu+2q_\nu)}{(q^2+m^2)((p+q)^2+m^2)} = \int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} \frac{p_\mu p_\nu(1-4x+4x^2)+4\ell_\mu\ell_\nu}{\left[m^2+\ell^2+x(1-x)p^2\right]^2} \\
&= p_\mu p_\nu \int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} \frac{1-4x+4x^2}{\left[m^2+\ell^2+x(1-x)p^2\right]^2} \\
&+ \frac{4\delta_{\mu\nu}}{d} \int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} \frac{\ell^2}{\left[m^2+\ell^2+x(1-x)p^2\right]^2},
\end{aligned} \tag{C.10}$$

where in the first line we have set to zero the integrals over linear terms in ℓ_μ, and in the last line we have set

$$\ell_\mu\ell_\nu \to \frac{\delta_{\mu\nu}}{d}\ell^2 \tag{C.11}$$

since this is the only contribution to the integral (see again [69], p.). In total, we find that (7.85) has two contributions. One has the tensorial structure of $\delta_{\mu\nu}$, with coefficient

$$2\int \frac{\mathrm{d}^2 q}{(2\pi)^2}\frac{1}{(q^2+m^2)} - \frac{4}{d}\int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} \frac{\ell^2}{\left[m^2+\ell^2+x(1-x)p^2\right]^2}, \tag{C.12}$$

while the other one has the tensorial structure of $p_\mu p_\nu$, and coefficient

$$-\int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^2\ell}{(2\pi)^2} \frac{1-4x+4x^2}{\left[m^2+\ell^2+x(1-x)p^2\right]^2}. \tag{C.13}$$

Let us compute (C.12). Using dimensional regularization and the integral

$$\int \frac{\mathrm{d}^d\ell}{(2\pi)^d}\frac{\ell^2}{(\ell^2+\Delta)^2} = \frac{1}{(4\pi)^{d/2}}\frac{d}{2}\Gamma(1-d/2)\Delta^{d/2-1}. \tag{C.14}$$

we obtain

$$2\int_0^1 \mathrm{d}x \left[\frac{1}{(4\pi)^{d/2}}\frac{\Gamma(1-d/2)}{\Delta_1^{1-d/2}} - \frac{2}{d}\frac{1}{(4\pi)^{d/2}}\frac{d}{2}\frac{\Gamma(1-d/2)}{\Delta_1^{1-d/2}}\right], \tag{C.15}$$

where

$$\Delta_1 = m^2, \qquad \Delta_2 = m^2 + x(1-x)p^2. \tag{C.16}$$

This can be written as

$$\frac{2\Gamma(1-d/2)}{(4\pi)^{d/2}}\int_0^1 \mathrm{d}x \left[\Delta_1^{d/2-1} - \Delta_2^{d/2-1}\right]. \tag{C.17}$$

We now expand around

$$d = 2 - \epsilon \Rightarrow d/2 - 1 = -\epsilon/2. \tag{C.18}$$

Since

$$\Delta_1^{d/2-1} - \Delta_2^{d/2-1} = \mathrm{e}^{-\epsilon \log \Delta_1/2} - \mathrm{e}^{-\epsilon \log \Delta_2/2} = \frac{\epsilon}{2} \log \frac{\Delta_2}{\Delta_1} + \cdots , \tag{C.19}$$

and

$$\Gamma(\epsilon/2) = \frac{2}{\epsilon} - \gamma + \mathcal{O}(\epsilon), \tag{C.20}$$

the total result is finite as $\epsilon \to 0$ and given by

$$2 \int_0^1 \mathrm{d}x \, \log\Big[1 + x(1-x)\frac{p^2}{m^2}\Big]. \tag{C.21}$$

References

[1] I. Affleck, "Testing The Instanton Method," Phys. Lett. B **92**, 149 (1980).

[2] I. Affleck, "Quantum Statistical Metastability," Phys. Rev. Lett. **46**, 388 (1981).

[3] I. Affleck, "Mesons In The Large N Collective Field Method," Nucl. Phys. B **185**, 346 (1981).

[4] G. Altarelli, "Introduction to renormalons," in Erice Subnuclear 1995, p. 221-248.

[5] G. Álvarez, "Coupling-constant behavior of the resonances of the cubic anharmonic oscillator," Phys. Rev. A **37** (1988) 4079-4083.

[6] J. Baacke and G. Lavrelashvili, "One-loop corrections to the metastable vacuum decay," Phys. Rev. D **69**, 025009 (2004) [arXiv:hep-th/0307202].

[7] I. Bars and M. B. Green, "Poincare And Gauge Invariant Two-Dimensional QCD," Phys. Rev. D **17**, 537 (1978).

[8] C. M. Bender, "Perturbation Theory In Large Order," Adv. Math. **30**, 250 (1978).

[9] C. M. Bender and W. E. Caswell, "Asymptotic Graph Counting Techniques In ψ^{2N} Field Theory," J. Math. Phys. **19**, 2579 (1978).

[10] C. M. Bender and G. V. Dunne, "Large-order Perturbation Theory for a Non-Hermitian PT-symmetric Hamiltonian," J. Math. Phys. **40**, 4616 (1999) [arXiv:quant-ph/9812039].

[11] C. M. Bender and T. T. Wu, "Anharmonic oscillator," Phys. Rev. **184**, 1231 (1969).

[12] C. M. Bender and T. T. Wu, "Anharmonic Oscillator. 2: A Study Of Perturbation Theory In Large Order," Phys. Rev. D **7**, 1620 (1973).

[13] C. M. Bender and T. T. Wu, "Statistical Analysis Of Feynman Diagrams," Phys. Rev. Lett. **37**, 117 (1976).

[14] M. Beneke, "Renormalons," Phys. Rept. **317**, 1 (1999) [arXiv:hep-ph/9807443].

[15] M. K. Benna, S. Benvenuti, I. R. Klebanov and A. Scardicchio, "A test of the AdS/CFT correspondence using high-spin operators," Phys. Rev. Lett. **98**, 131603 (2007) [arXiv:hep-th/0611135].

[16] R. Bott and L. Tu, *Differential forms in algebraic topology*, Springer-Verlag.

[17] E. Brézin, J. C. Le Guillou and J. Zinn-Justin, "Perturbation Theory At Large Order. 2. Role Of The Vacuum Instability," Phys. Rev. D **15**, 1558 (1977).

[18] E. Brézin, C. Itzykson, G. Parisi and J. B. Zuber, "Planar Diagrams," Commun. Math. Phys. **59**, 35 (1978).

[19] E. B. Bogomolny and V. A. Fateev, "Large order calculations in gauge theories," Phys. Lett. B **71**, 93 (1977).

[20] E. Caliceti, M. Meyer-Hermann, P. Ribeca, A. Surzhykov and U. D. Jentschura, "From Useful Algorithms for Slowly Convergent Series to Physical Predictions Based on Divergent Perturbative Expansions," Phys. Rep. **446** (2007) 1 [arXiv:0707.1596 [physics.comp-ph]].

[21] C. G. Callan and S. R. Coleman, "The Fate Of The False Vacuum. 2. First Quantum Corrections," Phys. Rev. D **16**, 1762 (1977).

[22] C. G. Callan, R. F. Dashen and D. J. Gross, "The structure of the gauge theory vacuum," Phys. Lett. B **63**, 334 (1976).

[23] J. Casahorrán, "Nonperturbative contributions in quantum-mechanical models: The instantonic approach," Commun. Math. Sci. **1**, 245 (2003) [arXiv:hep-th/0012038].

[24] S. Chadha and P. Olesen, "On Borel Singularities In Quantum Field Theory," Phys. Lett. B **72**, 87 (1977).

[25] G. A. Christos, "Chiral Symmetry And The U(1) Problem," Phys. Rept. **116**, 251 (1984).

[26] S. R. Coleman, "The Fate Of The False Vacuum. 1. Semiclassical Theory," Phys. Rev. D **15**, 2929 (1977) [Erratum-ibid. D **16**, 1248 (1977)].

[27] S. Coleman, *Aspects of symmetry*, Cambridge University Press.

[28] J. C. Collins and D. E. Soper, "Large Order Expansion In Perturbation Theory," Annals Phys. **112**, 209 (1978).

[29] F. Cooper and B. Freedman, "Aspects Of Supersymmetric Quantum Mechanics," Annals Phys. **146**, 262 (1983).

[30] F. Cooper, A. Khare and U. Sukhatme, "Supersymmetry and quantum mechanics," Phys. Rept. **251**, 267 (1995) [arXiv:hep-th/9405029].

[31] A. D'Adda, M. Luscher and P. Di Vecchia, "A 1/N Expandable Series Of Nonlinear Sigma Models With Instantons," Nucl. Phys. B **146**, 63 (1978).

[32] A. D'Adda and P. Di Vecchia, "Supersymmetry and instantons," Phys. Lett. B **73**, 162 (1978).

[33] R. F. Dashen, B. Hasslacher and A. Neveu, "Nonperturbative Methods And Extended Hadron Models In Field Theory. 1. Semiclassical Functional Methods," Phys. Rev. D **10**, 4114 (1974).

[34] F. David, "On The Ambiguity Of Composite Operators, Ir Renormalons And The Status Of The Operator Product Expansion," Nucl. Phys. B **234**, 237 (1984).

[35] L. Del Debbio, L. Giusti and C. Pica, "Topological susceptibility in the SU(3) gauge theory," Phys. Rev. Lett. **94**, 032003 (2005) [arXiv:hep-th/0407052].

[36] P. Di Francesco, "2D quantum gravity, matrix models and graph combinatorics," in *Applications of random matrices in physics*, arXiv:math-ph/0406013.

[37] J.F. Donoghue, E. Golwich and B.R. Holstein, *Dynamics of the Standard Model*, Cambridge University Press.

[38] G. V. Dunne, "Perturbative - nonperturbative connection in quantum mechanics and field theory," arXiv:hep-th/0207046.

[39] G. V. Dunne, "Functional Determinants in Quantum Field Theory," arXiv:0711.1178 [hep-th].

[40] G. V. Dunne and H. Min, "Beyond the thin-wall approximation: Precise numerical computation of prefactors in false vacuum decay," Phys. Rev. D **72**, 125004 (2005) [arXiv:hep-th/0511156].

[41] F. J. Dyson, "Divergence Of Perturbation Theory In Quantum Electrodynamics," Phys. Rev. **85** (1952) 631.

[42] K. Fujikawa, "Path Integral For Gauge Theories With Fermions," Phys. Rev. D **21**, 2848 (1980) [Erratum-ibid. D **22**, 1499 (1980)].

[43] A. Galindo and P. Pascual, *Mecánica Cuántica*, vol. 2, Eudema (Madrid), 1989.

[44] S. Garoufalidis, T.T.Q. Lê and M. Mariño, "Analyticity of the free energy of a closed 3-manifold," arXiv:0809.2572 [math.GT].

[45] J. Gasser and H. Leutwyler, "Chiral Perturbation Theory To One Loop," Annals Phys. **158**, 142 (1984).

[46] G. W. Gibbons and S. W. Hawking, "Action Integrals And Partition Functions In Quantum Gravity," Phys. Rev. D **15**, 2752 (1977).

[47] L. Giusti, S. Petrarca and B. Taglienti, "Theta dependence of the vacuum energy in the SU(3) gauge theory from the lattice," Phys. Rev. D **76**, 094510 (2007) [arXiv:0705.2352 [hep-th]].

[48] S. Graffi, V. Grecchi and B. Simon, "Borel Summability: Application To The Anharmonic Oscillator," Phys. Lett. B **32**, 631 (1970).

[49] D. J. Gross and A. Matytsin, "Instanton induced large N phase transitions in two-dimensional and four-dimensional QCD," Nucl. Phys. B **429**, 50 (1994) [arXiv:hep-th/9404004].

[50] D. J. Gross, M. J. Perry and L. G. Yaffe, "Instability Of Flat Space At Finite Temperature," Phys. Rev. D **25**, 330 (1982).

[51] D. J. Gross, R. D. Pisarski and L. G. Yaffe, "QCD And Instantons At Finite Temperature," Rev. Mod. Phys. **53**, 43 (1981).

[52] G. Grunberg, "Perturbation theory and condensates," Phys. Lett. B **325**, 441 (1994).

[53] S. Hikami and E. Brézin, "Large Order Behavior Of The 1/N Expansion In Zero-Dimensions And One-Dimensions," J. Phys. A **12**, 759 (1979).

[54] C. Itzykson and J.B. Zuber, *Quantum field theory*, McGraw Hill, 1980.

[55] M.A. Jafarizadeh and H. Fakhri, "Calculation of the determinant of shape invariant operators," Phys. Lett. A **230** (1997) 157.

[56] U. D. Jentschura, "The resummation of nonalternating divergent perturbative expansions," Phys. Rev. D **62**, 076001 (2000) [arXiv:hep-ph/0001135].

[57] R. K. Kaul and R. Rajaraman, "Soliton Energies In Supersymmetric Theories," Phys. Lett. B **131**, 357 (1983).

[58] P. Keith-Hynes and H. B. Thacker, "Fractionally charged Wilson loops as a probe of θ-dependence in $\mathbb{CP}^{N-1}$ sigma models: Instantons vs. large N," Phys. Rev. D **78**, 025009 (2008) [arXiv:0804.1534 [hep-lat]].

[59] M. Luscher, "A Semiclassical Formula For The Topological Susceptibility In A Finite Space-Time Volume," Nucl. Phys. B **205**, 483 (1982).

[60] M. Luscher, "Topological effects in QCD and the problem of short-distance singularities," Phys. Lett. B **593**, 296 (2004) [arXiv:hep-th/0404034].

[61] J. D. Lykken, "Baryons In A Large N Matrix Model," Phys. Rev. D **25**, 1653 (1982).

[62] A. V. Manohar, "Large N QCD," arXiv:hep-ph/9802419.

[63] G. Marchesini and E. Onofri, "Planar Limit For SU(N) Symmetric Quantum Dynamical Systems," J. Math. Phys. **21**, 1103 (1980).

[64] M. Mariño, "Nonperturbative effects and nonperturbative definitions in matrix models and topological strings," JHEP **0812**, 114 (2008) [arXiv:0805.3033 [hep-th]].

[65] A. J. McKane and M. B. Tarlie, "Regularisation of functional determinants using boundary perturbations," J. Phys. A **28**, 6931 (1995) [arXiv:cond-mat/9509126].

[66] G. Münster, "The $1/N$ Expansion And Instantons In $\mathbb{CP}^{N-1}$ Models On A Sphere," Phys. Lett. B **118**, 380 (1982).

[67] H. Neuberger, "Instantons As A Bridgehead At N = Infinity," Phys. Lett. B **94**, 199 (1980). "Nonperturbative Contributions In Models With A Nonanalytic Behavior At Infinite N," Nucl. Phys. B **179**, 253 (1981).

[68] V. A. Novikov, M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, "Two-Dimensional Sigma Models: Modeling Nonperturbative Effects Of Quantum Chromodynamics," Phys. Rept. **116**, 103 (1984) [Sov. J. Part. Nucl. **17**, 204.1986 FECAA,17,472 (1986 FECAA,17,472-545.1986)].

[69] G. Parisi, "Singularities Of The Borel Transform In Renormalizable Theories," Phys. Lett. B **76**, 65 (1978).

[70] M. Peskin, *An introduction to quantum field theory*, Addison–Weasley.

[71] R. Rajaraman, *Solitons and instantons*, North Holland, Amsterdam, 1982.

[72] P. Ramond, *Field theory. A modern primer*, second edition, Westview Press.

[73] P. Salomonson and J. W. van Holten, "Fermionic Coordinates And Supersymmetry In Quantum Mechanics," Nucl. Phys. B **196**, 509 (1982).

[74] T. Schafer and E. V. Shuryak, "Instantons in QCD," Rev. Mod. Phys. **70**, 323 (1998) [arXiv:hep-ph/9610451].

[75] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, "QCD And Resonance Physics. Sum Rules," Nucl. Phys. B **147**, 385 (1979).

[76] W. Siegel, "Fields," hep-th/9912205.

[77] B. Simon, "Large orders and summability of eigenvalue perturbation theory: a mathematical overview," Int. J. Quant. Chem. **21** (1982) 3-25.

[78] G. 't Hooft, "A planar diagram theory for strong interactions," Nucl. Phys. B **72**, 461 (1974).

[79] G. 't Hooft, "A Two-Dimensional Model For Mesons," Nucl. Phys. B **75**, 461 (1974).

[80] G. 't Hooft, "Can We Make Sense Out Of 'Quantum Chromodynamics'?," in G. 't Hooft, *Under the spell of the gauge principle*, World Scientific, p. 547-575.

[81] C. Vafa and E. Witten, "Parity Conservation In QCD," Phys. Rev. Lett. **53**, 535 (1984).

[82] G. Veneziano, "$U(1)$ Without Instantons," Nucl. Phys. B **159**, 213 (1979).

[83] E. Vicari and H. Panagopoulos, "Theta dependence of SU(N) gauge theories in the presence of a topological term," arXiv:0803.1593 [hep-th].

[84] S. Weinberg, "The $U(1)$ Problem," Phys. Rev. D **11**, 3583 (1975).

[85] S. Weinberg, *The quantum theory of fields. Volume I: Foundations*. Cambridge University Press, 1995.

[86] S. Weinberg, *The quantum theory of fields. Volume II: Modern applications*. Cambridge University Press, 1996.

[87] E. Witten, "Instantons, The Quark Model, And The 1/N Expansion," Nucl. Phys. B **149**, 285 (1979).

[88] E. Witten, "Current Algebra Theorems For The U(1) Goldstone Boson," Nucl. Phys. B **156**, 269 (1979).

[89] E. Witten, "Baryons In The 1/N Expansion," Nucl. Phys. B **160**, 57 (1979).

[90] E. Witten, "Dynamical Breaking Of Supersymmetry," Nucl. Phys. B **188**, 513 (1981).

[91] E. Witten, "Instability Of The Kaluza-Klein Vacuum," Nucl. Phys. B **195**, 481 (1982).

[92] E. Witten, "Dynamical aspects of QFT," in *Quantum fields and strings: A course for mathematicians*, vol. 2, American Mathematical Society, Providence, 1999.

[93] F. J. Ynduráin, *The theory of quark and gluon interactions*, Springer–Verlag.

[94] A. B. Zamolodchikov and A. B. Zamolodchikov, "Relativistic Factorized S Matrix In Two-Dimensions Having O(N) Isotopic Symmetry," Nucl. Phys. B **133**, 525 (1978) [JETP Lett. **26**, 457 (1977)].

[95] J. Zinn-Justin, "Multi - Instanton Contributions In Quantum Mechanics. 2," Nucl. Phys. B **218**, 333 (1983).

[96] J. Zinn–Justin, *Quantum field theory and critical phenomena*, Oxford University Press.

[97] J. Zinn-Justin and U. D. Jentschura, "Multi-instantons and exact results I: Conjectures, WKB expansions, and instanton interactions," Annals Phys. **313**, 197 (2004) [arXiv:quant-ph/0501136]; "Multi-Instantons And Exact Results Ii: Specific Cases, Higher-Order Effects, And Numerical Calculations," Annals Phys. **313**, 269 (2004) [arXiv:quant-ph/0501137].