Instantons and large $N$
An introduction to non-perturbative methods in QFT

Marcos Mariño
# Contents

1 Introduction

2 Instantons in quantum mechanics
   2.1 QM as a one-dimensional field theory .......................... 6
   2.2 Unstable vacua in quantum mechanics .......................... 10
   2.3 A toy model integral ........................................ 11
   2.4 Path integral around an instanton in QM ......................... 13
   2.5 Calculation of functional determinants ........................... 20
   2.6 Examples .................................................. 25
   2.7 Instantons in the double well .................................. 26
   2.8 Multi-instantons in the double well ............................... 30
   2.9 The dilute instanton approximation ............................... 35
   2.10 Beyond the dilute instanton approximation ....................... 35

3 Unstable vacua in QFT
   3.1 Bounces in scalar QFT ........................................ 40
   3.2 The fate of the false vacuum .................................... 45
   3.3 Instability of the Kaluza–Klein vacuum .......................... 46

4 Large order behavior and Borel summability
   4.1 Perturbation theory at large order .............................. 51
   4.2 The anharmonic oscillator ...................................... 52
   4.3 Large order behaviour and counting of Feynman diagrams ....... 55
   4.4 Asymptotic expansions and the nonperturbative ambiguity ....... 55
   4.5 Borel transform and Borel summability .......................... 57
   4.6 Instantons and large order behavior ............................. 63
   4.7 Large order behavior in quantum mechanics ..................... 64
   4.8 Euler–Heisenberg Lagrangian .................................... 68
   4.9 Resumming perturbative series ................................. 71

5 Nonperturbative aspects of gauge theories
   5.1 Conventions and basics ....................................... 73
   5.2 Topological charge and $\theta$ vacua ............................ 74
   5.3 Chiral symmetry and chiral symmetry breaking ................... 77
   5.4 The anomalous $U(1)$ and the $U(1)$ problem .................... 82
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Instantons and renormalons in gauge theory</td>
<td>84</td>
</tr>
<tr>
<td>6.1</td>
<td>Instantons in Yang–Mills theory</td>
<td>84</td>
</tr>
<tr>
<td>6.2</td>
<td>Instantons and theta vacua</td>
<td>89</td>
</tr>
<tr>
<td>6.3</td>
<td>Renormalons</td>
<td>91</td>
</tr>
<tr>
<td>7</td>
<td>The $\mathbb{P}^{N-1}$ sigma model</td>
<td>97</td>
</tr>
<tr>
<td>7.1</td>
<td>Defining the model</td>
<td>98</td>
</tr>
<tr>
<td>7.2</td>
<td>Instantons</td>
<td>99</td>
</tr>
<tr>
<td>7.3</td>
<td>The effective action at large $N$</td>
<td>101</td>
</tr>
<tr>
<td>7.4</td>
<td>Topological susceptibility at large $N$</td>
<td>105</td>
</tr>
<tr>
<td>8</td>
<td>The $1/N$ expansion in QCD</td>
<td>106</td>
</tr>
<tr>
<td>8.1</td>
<td>Fatgraphs</td>
<td>106</td>
</tr>
<tr>
<td>8.2</td>
<td>Large $N$ rules for correlation functions</td>
<td>111</td>
</tr>
<tr>
<td>8.3</td>
<td>Analyticity in the $1/N$ expansion</td>
<td>115</td>
</tr>
<tr>
<td>8.4</td>
<td>QCD spectroscopy at large $N$: mesons and glueballs</td>
<td>117</td>
</tr>
<tr>
<td>8.5</td>
<td>Baryons at large $N$</td>
<td>118</td>
</tr>
<tr>
<td>9</td>
<td>A solvable toy model: large $N$ matrix quantum mechanics</td>
<td>121</td>
</tr>
<tr>
<td>9.1</td>
<td>Defining the model. Perturbation theory</td>
<td>121</td>
</tr>
<tr>
<td>9.2</td>
<td>Exact ground state energy in the planar approximation</td>
<td>123</td>
</tr>
<tr>
<td>9.3</td>
<td>Excited states, or glueball spectrum</td>
<td>128</td>
</tr>
<tr>
<td>9.4</td>
<td>Some examples</td>
<td>128</td>
</tr>
<tr>
<td>9.5</td>
<td>Adding fermions, or meson spectrum</td>
<td>132</td>
</tr>
<tr>
<td>10</td>
<td>Applications in QCD</td>
<td>138</td>
</tr>
<tr>
<td>10.1</td>
<td>The $U(1)$ problem at large $N$. Witten–Veneziano formula</td>
<td>138</td>
</tr>
<tr>
<td>A</td>
<td>Polology and spectral representation</td>
<td>142</td>
</tr>
<tr>
<td>B</td>
<td>Chiral Lagrangians</td>
<td>144</td>
</tr>
<tr>
<td>C</td>
<td>Effective action for the $\mathbb{P}^{N-1}$ sigma model</td>
<td>149</td>
</tr>
<tr>
<td>D</td>
<td>A simple example of analyticity at large $N$</td>
<td>152</td>
</tr>
</tbody>
</table>
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

$$e^{-A/g}.$$  \hfill (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 for one of the most important quantum-mechanical effects: 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 is a single ground state, corresponding to a symmetric wavefunction, and the energy difference between the ground state and the first excited state is an instanton effect of the form (1.1),

$$E_1(g) - E_0(g) \sim e^{-A/g}.$$  \hfill (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,

$$E_0(g) = \Re E_0(g) + i \Im E_0(g), \quad \Im E_0(g) \sim e^{-A/g}$$ \hfill (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 their dynamics. 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 practitioners 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 [56] for a review). Some aspects of this debate were first pointed out by Witten in his seminal paper [67], and the debate is still going on (see for example [40]).

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 [68], consider the following series:

$$f_0(g) = g - g \log g + g \frac{(\log g)^2}{2} - g \frac{(\log g)^3}{6} + \cdots$$  \hspace{1cm} (1.4)

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

$$\lim_{g \to 0} f_0(g) = 0. \hspace{1cm} (1.5)$$

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

$$\lim_{g \to 0} f_0(g) \neq 0. \hspace{1cm} (1.6)$$

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 [59], 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,

\[
Z(\beta) = \text{tr} e^{-\beta H(\beta)},
\]

as

\[
E = -\lim_{\beta \to \infty} \frac{1}{\beta} \log Z(\beta).
\]

In the path integral formulation,

\[
Z(\beta) = \int D[q(t)] e^{-S(q)},
\]

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

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

and the path integral is over periodic trajectories

\[
q(-\beta/2) = q(\beta/2).
\]

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

\[
S(q) = \int_{-\beta/2}^{\beta/2} dt \left[ \frac{1}{2} (\dot{q}(t))^2 - V(q) \right]
\]

where the potential is

\[
V(q) = -W(q),
\]

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

\[
W(q) = \frac{m}{2} q^2 + W_{\text{int}}(q)
\]

where \( W_{\text{int}}(q) \) is the interaction term. Then, the path integral defining \( Z \) can be computed in standard Feynman perturbation theory by expanding in \( W_{\text{int}}(q) \). To extract the ground state energy we have to take into account the following
• Since we have to consider \( F(\beta) = \log Z(\beta) \), only connected bubble diagrams contribute.

• 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. a factor \( \beta \) that we have to extract in the end. Therefore, we can just perform \( n - 1 \) integrations over \( \mathbb{R} \).

\[
\tau \quad \rightarrow \quad \tau'
\]
\[
\frac{e^{-m|\tau - \tau'|}}{2m}
\]

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

The propagator of this one-dimensional field theory is

\[
\int \frac{dp}{2\pi} \frac{e^{i p \tau}}{p^2 + m^2} = \frac{e^{-m|\tau|}}{2m}.
\]  (2.9)

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

\[
W_{\text{int}}(q) = \frac{g}{4} q^4
\]  (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 to order \( g^3 \) (see Appendix B of [9] for some details). The relevant Feynman diagrams are shown in Fig. 3. For the Feynman
integrals we find (we set \( m = 1 \))

\[
\begin{align*}
1: & \quad \frac{1}{4} \\
2a: & \quad -\frac{1}{16} \int_{-\infty}^{\infty} e^{-2|\tau|} d\tau = -\frac{1}{16} \cdot 1, \\
2b: & \quad -\frac{1}{16} \int_{-\infty}^{\infty} e^{-4|\tau|} d\tau = -\frac{1}{16} \cdot \frac{1}{2}, \\
3a: & \quad \frac{1}{64} \int_{-\infty}^{\infty} e^{-|\tau_1|-|\tau_2|-|\tau_1-\tau_2|} d\tau_1 d\tau_2 = \frac{1}{64} \cdot \frac{3}{2}, \\
3b: & \quad \frac{1}{64} \int_{-\infty}^{\infty} e^{-2|\tau_1|-2|\tau_2|-2|\tau_1-\tau_2|} d\tau_1 d\tau_2 = \frac{1}{64} \cdot \frac{3}{8}, \\
3c: & \quad \frac{1}{64} \int_{-\infty}^{\infty} e^{-|\tau_1-\tau_2|-|\tau_1|-3|\tau_2|} d\tau_1 d\tau_2 = \frac{1}{64} \cdot \frac{5}{8}, \\
3d: & \quad \frac{1}{64} \int_{-\infty}^{\infty} e^{-2|\tau_1-\tau_2|-2|\tau_2|} d\tau_1 d\tau_2 = \frac{1}{64} \cdot 1 \\
\end{align*}
\]  

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

<table>
<thead>
<tr>
<th>diagram</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>3c</th>
<th>3d</th>
</tr>
</thead>
<tbody>
<tr>
<td>symmetry factor</td>
<td>3</td>
<td>36</td>
<td>12</td>
<td>288</td>
<td>576</td>
<td>288</td>
<td>432</td>
</tr>
</tbody>
</table>

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

\[
\frac{1}{n!} \langle (x^4)^n \rangle^{(c)},
\]

where

\[
\langle (x^4)^n \rangle = \frac{\int_{-\infty}^{\infty} dx \, e^{-x^2/2} x^{4n}}{\int_{-\infty}^{\infty} dx \, e^{-x^2/2}}
\]

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

\[
\langle x^{2k} \rangle = (2k - 1)!! = \frac{(2k)!}{2^k k!}
\]

we find

\[
\frac{1}{n!} \langle (x^4)^n \rangle = \frac{(4n - 1)!!}{n!} = \frac{(4n)!}{4^n n!(2n)!}
\]

One finds, for example,

\[
\langle x^4 \rangle^{(c)} = 3,
\]

\[
\frac{1}{2!} \langle (x^4)^2 \rangle^{(c)} = \frac{1}{2} \left( \langle (x^4)^2 \rangle - \langle x^4 \rangle^2 \right) = 48.
\]

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

\[
1 : \frac{1}{4} \cdot 3
\]
\[
2a : - \frac{1}{16} \cdot 1 \cdot 36,
\]
\[
2b : - \frac{1}{16} \cdot 1 \cdot 12,
\]
\[
3a : \frac{1}{64} \cdot \frac{3}{2} \cdot 388
\]
\[
3b : \frac{1}{64} \cdot \frac{3}{8} \cdot 566
\]
\[
3c : \frac{1}{64} \cdot \frac{5}{8} \cdot 288
\]
\[
3d : \frac{1}{64} \cdot 1 \cdot 432
\]
We then find
\[ E = \frac{1}{2} + \frac{3}{4} \left( \frac{g}{4} \right)^{2} - \frac{21}{8} \left( \frac{g}{4} \right)^{2} + \frac{333}{16} \left( \frac{g}{4} \right)^{3} + \mathcal{O}(g^{4}). \]  
\( \text{(2.18)} \)

**Remark 2.1.** In [9], 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.

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
\[ W(x) = \frac{x^{2}}{2} + \frac{g}{4} x^{4}. \]  
\( \text{(2.19)} \)

where
\[ g = -\lambda, \quad \lambda > 0. \]  
\( \text{(2.20)} \)

![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 \( (2/\lambda)^{1/2} \) at \( t = t_{0} \), and comes back to the origin at \( t = +\infty \).]
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

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

(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

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

(2.22)

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 suppressed factor we have

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

(2.23)

and

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

(2.24)

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

### 2.3 A toy model integral

In order to understand how to compute $\text{Im} Z$, it is very instructive to look at a simpler problem [24, 73]. We will then consider the reduction of the anharmonic oscillator to zero dimensions, and we will analyze the simple quartic integral

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

As a complex function of $g$, it can not be analytic at $g = 0$. This is because the integral is divergent for $g < 0$, no matter how small it is. In fact, the formal power series expansion around $g = 0$,

$$I(g) = \sum_{k=0}^{\infty} Z_k g^k,$$

(2.26)

where

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

(2.27)
Figure 5: The complex plane for the saddle-point calculation of (2.25). Here, $C^+$ and $C^-$ 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 $S_1$ and $S_2$.

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

$$Z_k \sim (-4)^k k!$$

(2.28)

It can be shown that $g = 0$ is a branch point, and that $I(g)$ has a branch cut along the negative real axis. There is a discontinuity across this axis that leads to an imaginary part for $I(g)$. The imaginary part can be computed by saddle-point methods. The saddle points occur at $z = 0$ or

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

(2.29)

Therefore we have two nontrivial saddlepoints $S_{1,2}$

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

(2.30)

where $\phi_g$ is the phase of $g$. For $g < 0$, these are on the real axis. The steepest descent trajectories passing through these points are determined by the condition

$$\text{Im} f(z) = \text{Im} f(z_i), \quad f(z) = \frac{z^2}{2} + \frac{g}{4}z^4.$$  

(2.31)

Fo4 $g < 0$ these are hyperbolae

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

(2.32)

passing through the saddlepoints $S_{1,2}$ at $x = \pm|g|^{-\frac{1}{4}}$, $y = 0$, see Fig. 2.3.
In order to define the original integral for \( g < 0 \), we rotate the contour of integration so that
\[
\text{Re} \left( gx^4 \right) > 0.
\] (2.33)

Therefore, we give a phase to \( x \) in such a way that
\[
\text{Arg} \ x = -\frac{1}{4} \text{Arg} \ g.
\] (2.34)

As we approach the negative axis, this can be done in two different ways. For \( g \to -|g| + i0 \), one has
\[
C_+ : \quad \text{Arg} \ x = -\frac{\pi}{4}
\] (2.35)

while for \( g \to -|g| - i0 \), one has
\[
C_- : \quad \text{Arg} \ x = \frac{\pi}{4}
\] (2.36)

Therefore, with this prescription, there will be an analytic continuation of the integral \( I(g) \) to negative \( g \), but with a branch cut along the negative real axis. The discontinuity can be evaluated as
\[
I(g + i0) - I(g - i0) = 2i \text{Im} \ I(g) = \frac{1}{\sqrt{2\pi}} \int_{C_+ - C_-} dz \ e^{-z^2/2 - gz^4/4}.
\] (2.37)

The difference between both contours is in turn given by the sum of the steepest descent trajectories, therefore the imaginary part is given by
\[
\text{Im} \ I(g) \sim e^{\frac{1}{e}}.
\] (2.38)

### 2.4 Path integral around an instanton in QM

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 (2.37), we expect that the quantity
\[
\text{disc} \ Z(-\lambda) = Z(-\lambda + i\epsilon) - Z(-\lambda - i\epsilon) = 2i \text{Im} \ Z(-\lambda)
\] (2.39)
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,
\[
-\ddot{q}_c(t) + V'(q_c) = 0.
\] (2.40)
Figure 6: 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.

Examples of such nontrivial, periodic saddle points are oscillations around the local minima of $V(q)$, as shown in Fig. 6. The period of such an oscillation between the turning points $q_{-}$ and $q_{+}$ is given by

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

These trajectories satisfy in addition the “energy conservation” constraint

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

Notice that the period (2.41) varies between $\beta = \infty$ (corresponding to $E = 0$ in Fig. 6) and a minimum critical value $\beta_c$ 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

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

where

$$
\omega^2 = -V''(q_0).
$$

At this order we can parametrize the energy as

$$
E = -V_0 + \frac{1}{2} \omega^2 \epsilon^2, \quad q_{\pm} = q_0 \pm \epsilon,
$$
where we just evaluated (2.42) with (2.43) at the turning points \( q_0 \pm \epsilon \). We then find,

\[
\beta = 2 \int_{q_0-\epsilon}^{q_0+\epsilon} \frac{dq}{\sqrt{\omega^2 (\epsilon^2 - (q - q_0)^2)}} = \frac{2}{\omega} \int_{-\epsilon}^{\epsilon} \frac{d\zeta}{\sqrt{\epsilon^2 - \zeta^2}} = \frac{2\pi}{\omega}.
\] (2.46)

As \( \epsilon \to 0 \) we then find,

\[
\beta_c = \frac{2\pi}{\omega}.
\] (2.47)

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.

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

\[
-\ddot{q}(t) + q(t) - \lambda q^3(t) = 0,
\] (2.48)

The inverted potential has minima at

\[
q = \pm \lambda^{-\frac{1}{2}},
\] (2.49)

and it has zeroes at

\[
q = \pm \left(\frac{2}{\lambda}\right)^{\frac{1}{2}}.
\] (2.50)

It is possible to find explicit solutions of (2.48) around the minima (2.49), but they are complicated and involve elliptic functions (see for example [52]). 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

\[
q_c(t) = \pm \left(\frac{2}{\lambda}\right)^{\frac{1}{2}} \frac{1}{\cosh(t-t_0)}.
\] (2.51)

where

\[
-\infty < t_0 < \infty
\] (2.52)

is a free parameter. Such a trajectory starts at the origin in the infinite past, arrives to the zero (2.50) \( \pm (2/\lambda)^{1/2} \) at \( t = t_0 \), and returns to the origin in the infinite future, i.e.

\[
q_c \to 0, \quad t \to \pm \infty,
\]

\[
t = t_0, \quad q_c(t_0) = \pm \left(\frac{2}{\lambda}\right)^{\frac{1}{2}}.
\] (2.53)

An example of (2.51) is shown at Fig. 7.
Let us now return to the general case and expand the action around $q_c(t)$. We find, after writing
\[ q(t) = q_c(t) + r(t) \]
that
\[ S(q) = S(q_c) + \int dt_1 dt_2 r(t_1) M(t_1, t_2) r(t_2) \]
where $M$ is the operator defined by
\[ M(t_1, t_2) = \frac{\delta^2 S}{\delta q_c(t_1) \delta q_c(t_2)} = \left[ -\left( \frac{d}{dt_1} \right)^2 + V''(q_c(t_1)) \right] \delta(t_1 - t_2). \]
In the quadratic (or one-loop) approximation, the path integral around this configuration is then given by
\[ \int \mathcal{D}q(t) e^{-S(q)} \approx e^{-S(q_c)} \int \mathcal{D}r(t) \exp \left[ -\frac{1}{2} \int dt_1 dt_2 r(t_1) M(t_1, t_2) r(t_2) \right]. \]
Since we are integrating over periodic configurations, the boundary conditions for $r(t)$ are
\[ r(-\beta/2) = r(\beta/2). \]
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.
The calculation of the determinant goes formally as follows. Let $q_n$ be orthonormal eigenfunctions of $M$, labeled by $n = 0, 1, \cdots$,

$$
\int dt_2 M(t_1, t_2)q_n(t_2) = \lambda_n q_n(t_1), \quad (2.59)
$$

and satisfying periodic boundary conditions appropriate for (2.57),

$$
q_n(-\beta/2) = q_n(\beta/2). \quad (2.60)
$$

The eigenvalue problem can be written explicitly as

$$
\left[ -\frac{d^2}{dt^2} + V''(q_c(t)) \right] q_n(t) = \lambda_n q_n(t), \quad n \geq 0, \quad (2.61)
$$

and orthonormality means that

$$
\int_{-\beta/2}^{\beta/2} dt q_n(t)q_m(t) = \delta_{nm}. \quad (2.62)
$$

A general configuration of $r(t)$ can then be written as

$$
r(t) = \sum_{n \geq 0} c_n q_n(t), \quad (2.63)
$$

and

$$
\int \mathcal{D}r(t) \exp \int dt_1 dt_2 r(t_1)M(t_1, t_2)r(t_2) = \mathcal{N} \int \prod_n \frac{dc_n}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \sum_n \lambda_n c_n^2 \right] = \mathcal{N} \prod_n \lambda_n^{-1/2}, \quad (2.64)
$$

where $\mathcal{N}$ is an overall, universal normalization factor in the path integral (in the sense that it does not depend on the potential we are considering). One usually writes

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

However, in order to makes sense of this expression, it is important to notice two properties of $M$ which are crucial to understand the problem:

- If we take a further derivative w.r.t. $t$ in (2.40) we find

$$
-\frac{d^2}{dt^2} \dot{q}_c(t) + V''(q_c(t))\ddot{q}_c(t) = 0, \quad (2.66)
$$
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 (2.58) are satisfied. It is also a normalizable function, therefore it must be (up to normalization) one of the functions \( q_n(t) \) appearing in (2.63), say \( q_0(t) \). The normalized zero mode is

\[
q_0(t) = \frac{1}{\|\dot{q}_c\|} \dot{q}_c(t),
\]

(2.67)

where the norm is given by

\[
\|\dot{q}_c\|^2 = \int_{-\beta/2}^{\beta/2} dt \ (\dot{q}_c(t))^2 = \int_{-\beta/2}^{\beta/2} dt \ [2(E - V(q))]^{1/2}.
\]

(2.68)

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 (2.51). A parameter for a family of solutions is called a modulus or a collective coordinate. If in a solution we change the modulus we still have a stationary solution, and the action does not change, i.e.

\[
S(q^{t_0}_c(t)) = S(q^{t_0+\delta t_0}_c(t)) \Rightarrow \frac{\delta^2 S}{\delta t_0^2} = 0,
\]

(2.69)

where we have explicitly indicated the dependence on \( t_0 \) in the family of solutions \( q_c(t) \). Equivalently,

\[
\int dt_1 M(t_1, t_2) \frac{\delta q^{t_0}_c(t)}{\delta t_0} = 0.
\]

(2.70)

But

\[
\frac{\delta q^{t_0}_c(t)}{\delta t_0} = -\dot{q}^{t_0}_c(t),
\]

(2.71)

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

- \( M \) has one, and only one, negative mode, therefore (2.57) is imaginary. To see this, we regard

\[
-\frac{d^2}{dt^2} + V''(q_c(t))
\]

(2.72)

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 \( M \).
Since \( \lambda_0 = 0 \), the contribution of \( c_0 \) to the path integral is simply
\[
\frac{1}{\sqrt{2\pi}} \int dc_0. \tag{2.73}
\]
On the other hand, we know that \( c_0 \), the parameter corresponding to the zero mode, stands really for the collective coordinate \( t_0 \), and in order to perform the above integral we have to change variables from \( c_0 \) to \( t_0 \). This is easily done by using the trick of Coleman ([23], p. 276). We expand the general configuration as
\[
q(t) = \sum_{n=0}^{\infty} c_n q_n(t) = q_c^0(t) + \sum_{n=1}^{\infty} c_n q_n(t) \tag{2.74}
\]
and we compare the variations of \( q(t) \) due to a change of \( t_0 \), and to a change of \( c_0 \):
\[
\delta q = -\dot{q}_c^0(t) \delta t_0 = q_0(t) \delta c_0 \tag{2.75}
\]
It follows that the Jacobian in absolute value is
\[
J = \left| \frac{\delta c_0}{\delta t_0} \right| = ||\dot{q}_c||. \tag{2.76}
\]
Therefore, the integration over \( c_0 \) gives
\[
\frac{1}{\sqrt{2\pi}} \int dc_0 = \frac{J}{\sqrt{2\pi}} \int_{-\beta/2}^{\beta/2} dt_0 = ||\dot{q}_c|| \sqrt{\frac{2\pi}{\beta}}, \tag{2.77}
\]
where we have used that the “moduli space” for \( t_0 \) is \([-\beta/2, \beta/2]\).

We now put everything together. In order to take into account the overall normalization \( N \) of the measure in the path integral, it is convenient to use the (unperturbed) harmonic oscillator with \( \omega = 1 \) as a reference point. Its thermal partition function will be denoted by \( Z_G \). A path integral evaluation of this partition function gives immediately
\[
Z_G = N (\det M_0)^{-1/2}, \tag{2.78}
\]
where
\[
M_0 = \left[ -\left( \frac{d}{dt_1} \right)^2 + 1 \right] \delta(t_1 - t_2). \tag{2.79}
\]
We then find,
\[
\text{Im} \ Z(\beta) = \frac{1}{2i} Z_G(\beta) \frac{\det' M}{\det M_0} \left( \frac{\beta}{\sqrt{2\pi}} \right)^{1/2} ||\dot{q}_c|| e^{-S_c}, \tag{2.80}
\]
where
\[
\det' M = \prod_{n \geq 1} \lambda_n \tag{2.81}
\]
is the determinant of the operator \( M \) once the zero mode has been removed. Notice that \( \det' M \) is negative, therefore the above quantity is real, as it should.

We now assume that the inverted potential has the form
\[
V(q) = \frac{1}{2} q^2 + \mathcal{O}(\lambda), \tag{2.82}
\]
where \( \lambda \) is a coupling constant. Then at leading order in the coupling constant \( \lambda \) our problem is in fact a harmonic oscillator, and
\[
\text{Re} Z \approx Z_G \tag{2.83}
\]
Therefore, at leading order in \( \lambda \),
\[
\text{Im} F(\beta) = \frac{1}{2} \sqrt{\frac{\left\| \dot{q}_c \right\|^2}{2\pi}} \left[ \det M_0 \left( -\det' M \right)^{-1} \right]^{\frac{1}{2}} e^{-S_c}. \tag{2.84}
\]
This gives the imaginary part of the free energy for potentials of the form (2.82), 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

We will now explain some general results for computing functional determinants which are very much used in instanton calculations. We will use a slightly indirect approach, in the sense that we will not compute the original determinant in (2.84) with the zero modes remove (this can be easily done by adapting the methods below, see for example the Appendix A of [42] or [52] for a more detailed analysis). We will follow the strategy of [73] and actually compute a determinant with no zero modes.

The first result we need on functional determinants is the so-called Gelfand–Yaglom theorem (see [32] for statement and examples). Let us consider the eigenvalue problem for two Schrödinger operators in the interval \( [-\beta/2, \beta/2] \), and with Dirichlet boundary conditions \( \psi(-\beta/2) = \psi(\beta/2) = 0 \),
\[
\begin{align*}
\left[ -\frac{d^2}{dt^2} + W_1(t) \right] \psi(t) &= \lambda \psi(t), \\
\left[ -\frac{d^2}{dt^2} + W_2(t) \right] \psi(t) &= \lambda \psi(t). \tag{2.85}
\end{align*}
\]
Then,
\[
\frac{\det\left(-\frac{d^2}{dt^2} + W_1(t)\right)}{\det\left(-\frac{d^2}{dt^2} + W_2(t)\right)} = \frac{\phi_1(\beta/2)}{\phi_2(\beta/2)},
\]
where \(\phi_i(t), i = 1, 2\), are solutions of the zero mode problem
\[
\left[-\frac{d^2}{dt^2} + W_1(t)\right]\phi_i(t) = 0
\]
with the boundary condition
\[
\phi_i(-\beta/2) = 0, \quad \phi_i'(\beta/2) = 1.
\]
A short proof of this result can be found in [23], and an explicit proof is easily found by using the so-called “shifting method” (see for example [26, 73]).

We will now use this result to calculate
\[
\langle x'| e^{-\beta H} | x \rangle = \mathcal{N} \int \mathcal{D}q(t) e^{-S(q)}
\]
where the Euclidean action is given by (2.6), and for arbitrary potential \(V(q)\) and finite \(\beta\). Notice however that now the integration is not over periodic trajectories, but over trajectories with
\[
q(-\beta/2) = x, \quad q(\beta/2) = x'.
\]
As before, the calculation reduces, in the Gaussian approximation, to
\[
\langle x'| e^{-\beta H} | x \rangle \approx e^{-S(q_c)} \int \mathcal{D}r(t) \exp \left[ -\frac{1}{2} \int dt_1dt_2 r(t_1)M(t_1,t_2)r(t_2) \right],
\]
where \(M\) is the operator (2.56), \(q_c(t)\) is a classical solution which satisfies the boundary conditions (2.90), and \(r(t)\) satisfies now Dirichlet boundary conditions
\[
r(-\beta/2) = r(\beta/2) = 0.
\]
The advantage of these boundary conditions is that there are now no zero modes for the operator \(M\), and
\[
\int \mathcal{D}r(t) \exp \left[ -\frac{1}{2} \int dt_1dt_2 r(t_1)M(t_1,t_2)r(t_2) \right] = \left(\det M\right)^{-1/2}.
\]
Another advantage is that we can compute the determinant we can use the Gelfand–Yaglom theorem. To do this we have to find a solution to the zero mode problem
\[
\left[-\frac{d^2}{dt^2} + V''(q_c(t))\right]\chi(t) = 0
\]
with the right boundary conditions. The general solution of (2.94) is given by

\[ \chi(t) = A \dot{q}_c(t) + B \ddot{q}_c(t) \int_{-\beta/2}^{t} \frac{d\tau}{(\dot{q}_c(t))^2}. \]  

(2.95)

That \( q_c(t) \) is a zero mode follows from (2.66), and an easy calculation shows that

\[ \dot{q}_c(t)f(t) \]

solves (2.94) if \( f(t) \) solves

\[ \dot{q}_c(t) \ddot{f}(t) + 2\ddot{q}_c(t) \dot{f}(t) = 0, \]

(2.96)

which indeed is the case for

\[ f(t) = \int_{-\beta/2}^{t} \frac{d\tau}{(\dot{q}_c(t))^2}. \]  

(2.97)

We now look for the particular solution which is needed in the Gelfand–Yaglom theorem. Imposing first \( \chi(-\beta/2) = 0 \) leads to \( A = 0 \), and imposing \( \chi'(-\beta/2) = 1 \) leads to

\[ \phi(t) = \dot{q}_c(-\beta/2)\dot{q}_c(\beta/2) \int_{-\beta/2}^{\beta/2} \frac{d\tau}{(\dot{q}_c(t))^2}. \]

(2.98)

It then follows that, up to an overall constant \( C \), and at one loop

\[ \langle x' | e^{-\beta H} | x \rangle \approx C e^{-S(q_c)}(\phi(\beta/2))^{-1/2}, \]

(2.100)

or, explicitly,

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

(2.101)

To determine \( C \) we just notice that for the free particle

\[ \langle x' | e^{-\beta H} | x \rangle = (2\pi\beta)^{-1/2} e^{-(x-x')^2/2\beta}. \]

(2.102)

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

\[ q_c(t) = \frac{x + x'}{2} + \frac{x' - x}{\beta} t, \]

(2.103)

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

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

(2.104)
We conclude, by comparing the two calculations, that

\[ C = \frac{1}{\sqrt{2\pi}} \]  

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

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

This formula can be re-expressed in many ways. We can for example write the equation for the duration of the trajectory as

\[ \beta = \int_{x'}^x \frac{dq}{[2(E - V(q))]^{1/2}}. \]  

Taking derivatives w.r.t. \( \beta \) we find

\[ 1 = -\frac{\partial E}{\partial \beta} \int_{x'}^x \frac{dq}{[2(E - V(q))]^{1/2}}. \]  

Because of energy conservation we have

\[ \dot{q}_c(t) = \left[ 2(E - V(q_c)) \right]^{1/2} \Rightarrow dq_c(t) = \left[ 2(E - V(q_c)) \right]^{1/2} dt, \]  

therefore

\[ -\frac{\partial E}{\partial \beta} = \left[ \int_{x'}^x \frac{dq}{[2(E - V(q))]^{1/2}} \right]^{-1} = \left[ \int_{-\beta/2}^{\beta/2} \frac{dt}{2(E - V(q))} \right]^{-1}. \]

and

\[ \int_{-\beta/2}^{\beta/2} \frac{dt}{(\dot{q}_c(t))^2} = \left( -\frac{\partial E}{\partial \beta} \right)^{-1}. \]

**Remark 2.3.** Using that

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

one can write

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

This can be also derived with the WKB method [73] and it is known as Van Vleck’s formula.
We want to calculate now
\[ \text{Tr } e^{-\beta H} = \int dx \langle x | e^{-\beta H} | x \rangle, \] (2.114)
which we evaluate at one loop around a periodic trajectory with turning points \( q_- \) and \( q_+ \), as we did in the previous section. We therefore impose periodic boundary conditions \( q_-(\beta/2) = q_+(\beta/2) = x \), where \( x \) can be any point of this trajectory, and we find
\[ \dot{q}_-(\beta/2)\dot{q}_+(\beta/2) = \dot{q}_c(x)^2 = 2(E - V(x)). \] (2.115)
Since we are expanding around a nontrivial saddle-point with an instability, we are actually computing
\[ 2i \text{Im } Z = \int dx \langle x | e^{-\beta H} | x \rangle. \] (2.116)
The action \( S_c \) and the factor (2.111) depend only on \( \beta \), and not on the point \( x \), therefore we find
\[ \text{Im } Z = \frac{1}{2i} \left( 2 \int_{q_-}^{q_+} \frac{dx}{[2(E - V(x))]^{1/2}} \right) \left( -\frac{1}{2\pi} \frac{\partial E}{\partial \beta} \right)^{1/2} e^{-S(q_c)} = -\frac{1}{2} \frac{\beta}{\sqrt{2\pi}} \left( \frac{\partial E}{\partial \beta} \right)^{1/2} e^{-S(q_c)}, \] (2.117)
where the first integral has been computed by using (2.41). We remind that this result is valid for all \( \beta > \beta_c \), where \( \beta_c \) is given in (2.47), since for \( \beta < \beta_c \) there are no periodic trajectories like the ones we have considered.

If we now compare (2.117) with (2.80) we find that
\[ \frac{\text{det' } M}{\text{det } M_0} = -\frac{\|\dot{q}_c\|^2}{4 \sinh^2 \frac{\beta}{2}} \left( \frac{\partial E}{\partial \beta} \right)^{-1}. \] (2.118)
A direct derivation of this result can be found in Appendix A of [42].

We can now analyze the large \( \beta \) limit of our final expression in order to extract the ground state energy. For simplicity we normalize our potential in such a way that the unstable minimum is at \( q = 0 \).

- At large \( \beta \), 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
\[ S_c = \int dt \left( \frac{1}{2} \dot{q}_c(t)^2 - V(q_c(t)) \right) = \int dt \left( \dot{q}_c(t) \right)^2 = 2 \int_{q_0}^{q_1} dq (2W)^{1/2}, \] (2.119)
where we have used conservation of energy (2.42).
To compute $\partial E/\partial \beta$ in the large $\beta$ limit, we write (2.107) as

$$
\beta = 2 \int_{q^{-}}^{q^{+}} \mathrm{d}x \left[ \frac{1}{2(E - V(x))^{2}} - \frac{1}{2(E + x^{2})^{2}} + \frac{1}{2(E + x^{2})^{2}} \right]. \quad (2.120)
$$

The last term gives

$$
2 \log \left( x + \sqrt{x^{2} + 2E} \right) \bigg|_{0}^{q^{+}} = \log q_{+}^{2} - \log(E/2) + O(E), \quad E \to 0. \quad (2.121)
$$

The first two terms, again up to corrections of order $O(E)$, gives

$$
2 \int_{0}^{q^{+}} \mathrm{d}x \left( \frac{1}{\sqrt{2W(x)}} - \frac{1}{x} \right). \quad (2.122)
$$

It follows that

$$
E(\beta) \approx -2q_{+}^{2} \exp \left[ 2 \int_{0}^{q^{+}} \mathrm{d}x \left( \frac{1}{\sqrt{2W(x)}} - \frac{1}{x} \right) \right] e^{-\beta}, \quad (2.123)
$$

therefore

$$
\frac{\partial E}{\partial \beta} \approx 2q_{+}^{2} \exp \left[ 2 \int_{0}^{q^{+}} \mathrm{d}x \left( \frac{1}{\sqrt{2W(x)}} - \frac{1}{x} \right) \right] e^{-\beta}. \quad (2.124)
$$

Finally, at leading order in $\beta$ and the coupling constants, we have that

$$
\text{Re} Z \sim e^{-\beta/2}. \quad (2.125)
$$

After putting everything together we finally obtain a general formula for the width of an unstable level in quantum mechanics,

$$
\text{Im} E_{0} \approx \frac{1}{2\sqrt{\pi}} q_{+} \exp \left[ \int_{0}^{q^{+}} \mathrm{d}x \left( \frac{1}{\sqrt{2W(x)}} - \frac{1}{x} \right) \right] e^{-S(q_{c})}. \quad (2.126)
$$

### 2.6 Examples

We now compute (2.126) in various examples.

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

$$
S[q_{c}(t)] = 2 \int_{0}^{\sqrt{2/\lambda}} x \sqrt{1 - \frac{\lambda}{2} x^{2}} \mathrm{d}x = -\frac{2 (2 - \lambda x^{2})^{3/2}}{3 \sqrt{2 \lambda}} \bigg|_{0}^{\sqrt{2/\lambda}} = \frac{4}{3 \lambda}. \quad (2.127)
$$
To compute the prefactor we have to compute

\[
\int_0^{q_c} dx \left( \frac{1}{\sqrt{2W(x)}} - \frac{1}{x} \right) = \int_0^{\sqrt{2/\lambda}} dx \frac{\sqrt{2 - \sqrt{2 - \lambda x^2}}}{x \sqrt{2 - \lambda x^2}}
\]

\[
= - \log \left( \sqrt{2\sqrt{2 - \lambda x^2} + 2} \right) \bigg|_0^{\sqrt{2/\lambda}} = \log 2.
\]

Therefore, putting everything together, we find

\[
\text{Im} E_0 \approx \frac{2}{2\sqrt{\pi}} \cdot \sqrt{\frac{2}{\lambda}} \cdot 2 \cdot e^{-\frac{4}{\pi \lambda}} = \frac{4}{\sqrt{2\pi \lambda}} e^{-\frac{4}{\pi \lambda}}.
\]

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

\[
V(x) = \frac{1}{2} x^2 - g x^3.
\]

The turning points are \( x_- = 0 \) and

\[
x_+ = \frac{1}{2g}.
\]

The action of the instanton is

\[
S_c = 2 \int_0^{1/(2g)} (x^2 - 2gx^3)^{\frac{1}{2}} dx = \frac{2}{15g^2}.
\]

The nontrivial integral involved in the one-loop fluctuation is

\[
\int_0^{1/(2g)} dx \frac{x - \sqrt{x^2 - 2gx^3}}{x \sqrt{x^2 - 2gx^3}} = \log 4,
\]

and we find

\[
\text{Im} E_0(g) \approx \frac{1}{2\sqrt{\pi}} \cdot \frac{1}{2g} \cdot 4 \cdot e^{-2/(15g^2)} = \frac{1}{\sqrt{\pi g^2}} e^{-2/(15g^2)}.
\]

This agrees with the results of [3].

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. In this section we follow the exposition and notations in [73, 74].

Consider the double well potential with Hamiltonian

\[
H = -\frac{1}{2} \left( \frac{d}{dq} \right)^2 + W(g,q), \quad W(g,q) = \frac{1}{2} q^2 (1 - \sqrt{g} q)^2. 
\]  

(2.135)

In perturbation theory one finds two degenerate ground states, located around the minima, and with energy given by

\[
E_0(g) = \frac{1}{2} - g - \frac{9}{2} g^2 - \frac{89}{2} g^3 - \frac{5013}{8} g^4 - \ldots 
\]

(2.136)

For all further discussions, we scale the coordinate \( q \) in (2.135) as

\[
q \rightarrow \frac{1}{\sqrt{g}} q. 
\]

(2.137)

The Hamiltonian corresponding to the double-well potential can therefore be written as

\[
H = -\frac{g}{2} \left( \frac{d}{dq} \right)^2 + \frac{1}{g} W(q), 
\]

(2.138a)

\[
W(q) = \frac{1}{2} q^2 (1 - q)^2. 
\]

(2.138b)

In this representation, the minima of the potential are at

\[
q = 0, \quad q = 1 
\]

(2.139)

and \( g \) plays the role of \( \hbar \). The Hamiltonian is symmetric in the exchange \( q \leftrightarrow 1 - q \) and thus commutes with the corresponding parity operator \( P \), whose action on wave functions is

\[
P \psi(q) = \psi(1 - q) \Rightarrow [H, P] = 0. 
\]

(2.140)

The eigenfunctions of \( H \) satisfy

\[
H \psi_{\epsilon,N}(q) = E_{\epsilon,N}(g) \psi_{\epsilon,N}(q), \quad P \psi_{\epsilon,N}(q) = \epsilon \psi_{\epsilon,N}(q), 
\]

(2.141)

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 \( g \to 0 \), \( E_{\epsilon,N}(g) = N + 1/2 + \mathcal{O}(g) \), 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

$$Z_a(\beta) = \text{Tr } P e^{-\beta H}$$

(2.142)

where $P$ is the parity operator (2.140). For large $\beta$ one has

$$Z_a(\beta) \sim e^{-\beta E_{+,0} - \beta E_{-,0}} \sim -2 \sinh \left[ \beta(E_{+,0} - E_{0,-})/2 \right] e^{-\beta(E_{+,0} + E_{0,-})}$$

(2.143)

where we have used that $E_{+,0} - E_{0,-}$ vanishes in perturbation theory and that, at leading order in $g$, $E_{+,0} + E_{0,-} = 1$. $Z_a(\beta)$ can be written in terms of a path integral with “twisted” boundary conditions,

$$Z_a(\beta) \equiv \text{Tr } (Pe^{-\beta H}) \propto \int_{q(\beta/2) = P(q(-\beta/2))} \mathcal{D}q(t) \exp \left[ -\frac{1}{g} S(q(t)) \right],$$

(2.144)

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

$$q(-\beta/2) + q(\beta/2) = 1.$$  

(2.145)

In the infinite $\beta$ limit, the leading contributions to the path integral come from paths which are solutions of the Euclidean equations of motion and have a finite action. In the case of $Z_a(\beta)$, 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 0 and 1, like in Fig. 8. These correspond to nontrivial instanton configurations. In the example of the double-well potential (2.138), such solutions are

$$q^{t_0}_\pm(t) = \frac{1}{1 + e^{t-t_0}}.$$  

(2.146)

The solutions $q^{t_0}_\pm$, which go from $q = 0$ to $q = 1$, and from $q = 1$ to $q = 0$, respectively, are called (anti)instantons of center $t_0$. They are represented in Fig. 8 and Fig. 9. Since both solutions depend on an integration constant $t_0$, there are two one-parameter families of degenerate saddle points.

To evaluate the contribution of these configurations to the path integral, we can use (2.117) but without the factor $1/(2i)$ (the operator (2.112) is now positive definite), therefore

$$\text{Tr } P e^{-\beta H} = 2 \frac{\beta}{\sqrt{2\pi}} \left( -\frac{\partial E}{\partial \beta} \right)^{1/2} e^{-S(q_c)}.$$  

(2.147)
The extra factor of 2 is due to the fact that the solutions $q^{l_0}_\pm(t)$ give the same contribution. We can easily compute

$$E(\beta) = -\frac{2}{g} e^{-\beta} + \mathcal{O}(e^{-2\beta}), \quad S(q^{l_0}_\pm) = \frac{1}{6g} + \mathcal{O}(e^{-\beta})$$  \hspace{1cm} (2.148)$$

We see that, at leading order in $g$ and for $\beta \to \infty$, this contribution is proportional to $e^{-1/(6g)}$ and therefore it is nonperturbative. One finally obtains

$$Z_a(\beta) \sim \frac{2}{\sqrt{\pi g}} \beta e^{-\beta/2} e^{-1/6g}.$$  \hspace{1cm} (2.149)$$

Therefore, from (2.143) we find the nonperturbative splitting between the symmetric and the antisymmetric wavefunctions as

$$E_{+,0}(g) - E_{-,0} = -\frac{2}{\sqrt{\pi g}} e^{-1/6g} (1 + \mathcal{O}(g)).$$  \hspace{1cm} (2.150)$$

29
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

\[ E_{\epsilon,0}(g) = E_0^{(0)}(g) + E_{\epsilon,0}^{(1)}(g), \]

(2.151)

where $\epsilon = \pm$ and

\[ E_0^{(0)}(g) = \frac{1}{2} + \mathcal{O}(g), \quad E_{\epsilon,0}^{(1)}(g) = -\frac{\epsilon}{\sqrt{\pi g}} e^{-1/6g} (1 + \mathcal{O}(g)) \]

(2.152)

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

### 2.8 Multi-instantons in the double well

In fact, the energies (2.151) 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 [74].

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

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

(2.153)

The $n$-instanton contribution is proportional to $\beta^n$. As we will see, the form (2.153) 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 (2.153) 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. 10.

Finally, notice that the sum in (2.153) can be written as

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

(2.154)
We see that \( n \) even contributes to \( Z(\beta) \), while \( n \) odd contributes to \( Z_a(\beta) \). 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

\[
q(-\beta/2) = q(\beta/2).
\]  

(2.155)

For example, in Fig. 10 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

\[
q(-\beta/2) + q(\beta/2) = 1.
\]  

(2.156)

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 \textit{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 10: A two-instanton configuration of the form (2.157), for \( \theta = 20 \).](image)

We consider a configuration \( q_c^\theta(t) \) that is the sum of instantons separated by a distance \( \theta \), up to an additive constant adjusted in such a way as to satisfy the boundary conditions (Fig. 10):

\[
q_c^\theta(t) = q_+^{-\theta/2}(t) + q_-^{\theta/2}(t) - 1 = q_+^{\theta/2}(t) - q_-^{-\theta/2}(t).
\]  

(2.157)

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 $\theta$ 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 $\theta$. It is convenient to introduce some additional notation:

\[ u(t) = q_{-\theta/2}(t), \]
\[ v(t) = u(t + \theta), \]  

therefore it follows from (2.157) that $q_c^\theta = u - v$. The action corresponding to the path (2.157) can be written as

\[
S(q_c^\theta) = \int dt \left( \frac{1}{2} q_c^2 + V(q_c) \right) = 2 \times \frac{1}{6} + \int dt \left[ -\dot{u}\dot{v} + V(u - v) - V(u) - V(v) \right].
\]  

(2.159)

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 $e^{-\theta/2}$ for large $\theta$. After an integration by parts of the term $\dot{v}\dot{u}$, one finds

\[
S(q_c^\theta) = \frac{1}{3} + 2 \left\{ v(0) \dot{u}(0) + \int_0^{+\infty} dt \left[ v\ddot{u} + V(u - v) - V(u) - V(v) \right] \right\}.
\]  

(2.160)

One then expands the integrand in powers of $v$. Since the leading correction to $S$ is of order $e^{-\theta}$, 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

\[
S(q_c^\theta) - \frac{1}{3} \sim 2v(0)\dot{u}(0) + 2 \left\{ \int_0^{+\infty} dt \left[ \frac{1}{2} v^2 V''(u) - \frac{1}{2} V''(0) v^2 \right] \right\}.
\]  

(2.161)

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

\[
v(0)\dot{u}(0) \sim -e^{-\theta}
\]  

(2.162)

and thus

\[
S(q_c^\theta) = \frac{1}{3} - 2e^{-\theta} + O(e^{-2\theta}).
\]  

(2.163)

32
As we will see, in order to compute the $n$-instanton contribution but at leading order in $g$, the corrections of higher order in $e^{-\theta}$ 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 $\theta$ is large. In analogy with the partition function of a classical gas (instantons being identified with particles), one calls the quantity $-2e^{-\theta}$ the interaction potential between instantons.

Actually, it is simple to extend the result to $\beta$ large but finite. To do this, we have to notice that $\beta$ is a periodic variable, so it lives in a circle. The two instantons separated by a distance $\theta$ are also separated by a distance $\beta - \theta$, see Fig. 11. Symmetry between $\theta$ and $\beta - \theta$ then implies

$$S(q_\theta^c) = \frac{1}{3} - 2e^{-\theta} - 2e^{-(\beta-\theta)} + \cdots$$

(2.164)

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

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

(2.165)

We can represent them as in Fig. 12. As noted above, for $n$ even, $n$-instanton configurations contribute to $\text{Tr}e^{-\beta H}$, while for $n$ odd they contribute to $\text{Tr}(P e^{-\beta H})$.

At leading order, we need only consider “interactions” between nearest neighbour instantons. Other interactions are negligible because they are of higher order in $e^{-\theta}$. This is an essential simplifying feature of quantum mechanics compared to quantum field theory. The classical action $S_c(\theta_i)$ can then be directly inferred from expression (2.164):

$$S_c(\theta_i) = \frac{n}{6} - 2\sum_{i=1}^n e^{-\theta_i} + O(e^{-(\theta_i+\theta_j)}).$$

(2.166)
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 $\theta$. The Gaussian integration involves then the determinant of the second derivative of the action at the classical path

$$M(t',t) = -\left(\frac{d}{dt}\right)^2 + V''(q_c(t)) \delta(t-t').$$

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 (2.153) is given by

$$Z_n^+(\beta) = e^{-\beta^2/2} \frac{\beta}{n} \left(\frac{e^{-1/\sqrt{n}g}}{\sqrt{n}g}\right)^n \int_{\theta_i \geq 0} \delta\left(\sum \theta_i - \beta\right) \prod_i d\theta_i \exp\left[\frac{2}{g} \sum_{i=1}^n e^{-\theta_i}\right]. \quad (2.168)$$

The overall factor $\beta$ 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 $\theta_i$. Finally, the normalization factor $e^{-\beta^2/2}$ 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 (2.168) is the final expression for the contribution of an $n$-instanton configuration.
2.9 The dilute instanton approximation

We will now evaluate (2.168) in which instanton interactions are neglected. This is called the dilute instanton approximation. Formally, to suppress the interactions, we should take the limit

$$g \rightarrow 0^-,$$

(2.169)
since in this case

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

(2.170)

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 $\theta_i$'s is straightforward, since

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

(2.171)

and

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

(2.172)

The sum of the leading order $n$-instanton contributions

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

(2.173)
can now be calculated:

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

(2.174)

with

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

(2.175)

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 (2.153).

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 $\theta_i$ 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,

$$\lambda(g) = \frac{e}{\sqrt{\pi g}} e^{-1/6g}.$$  \hfill (2.176)

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

$$\mu = \frac{-2}{g}.$$  \hfill (2.177)

To calculate the integral (2.168), we factorize the integral over the $\theta_i$, by introducing a complex contour integral representation for the $\delta$-function,

$$\delta \left( \sum_{i=1}^{n} \theta_i - \beta \right) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} ds \exp \left[ -s \left( \beta - \sum_{i=1}^{n} \theta_i \right) \right].$$  \hfill (2.178)

In terms of the function

$$\mathcal{I}(s, \mu) = \int_{0}^{+\infty} \exp (s \theta - \mu e^{-\theta}) \, d\theta,$$  \hfill (2.179)

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

$$Z^{(n)}_{\epsilon}(\beta) \sim \frac{\beta e^{-\beta/2}}{2\pi i n} \int_{-\infty}^{\infty} ds \, e^{-\beta s} \left[ \mathcal{I}(s, \mu) \right]^n.$$  \hfill (2.180)
In order to compute $Z^{(n)}_{\epsilon}(\beta)$, we will compute its Laplace transform

$$G^{(n)}_{\epsilon}(E) = \int_{0}^{\infty} d\beta e^{\beta E} Z^{(n)}_{\epsilon}(\beta) \quad (2.181)$$

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

$$G(E) = \text{Tr} \frac{1}{H - E} = \int_{0}^{\infty} d\beta e^{\beta E} Z(\beta), \quad (2.182)$$

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

$$G^{(n)}_{\epsilon}(E) = \int_{0}^{\infty} d\beta e^{\beta E} Z^{(n)}_{\epsilon}(\beta) = \int_{0}^{\infty} d\beta \frac{\beta e^{\beta E - 1/2}}{2\pi i} \int_{-i\infty}^{i\infty} ds e^{-\beta s} [I(s, \mu)]^{n}$$

$$= \frac{\partial}{\partial E} \int_{-i\infty}^{i\infty} ds \frac{\lambda^{n}}{2\pi i} [I(s, \mu)]^{n} \int_{0}^{\infty} d\beta e^{\beta(E-s-1/2)}$$

$$= \frac{\partial}{\partial E} \int_{-i\infty}^{i\infty} ds \frac{\lambda^{n}}{2\pi i} [I(s, \mu)]^{n} \frac{1}{s + 1/2 - E}$$

$$= \frac{\partial}{\partial E} \frac{\lambda^{n}}{n} [I(E - \frac{1}{2}, \mu)]^{n}. \quad (2.183)$$

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

$$\sum_{n=1}^{\infty} G^{(n)}_{\epsilon}(E) = -\frac{\partial}{\partial E} \ln \phi_{\epsilon}(E) \quad (2.184)$$

where

$$\phi_{\epsilon}(E) = \log \left( 1 - \lambda I(E - \frac{1}{2}, \mu) \right). \quad (2.185)$$

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

$$G_\epsilon(E, g) = G_0(E) - \frac{\partial}{\partial E} \ln \phi_\epsilon(E)$$

$$= \frac{\partial}{\partial E} \ln \Gamma \left( \frac{1}{2} - E \right) - \frac{\partial}{\partial E} \ln \phi_\epsilon(E)$$

$$= - \frac{\partial}{\partial E} \ln \frac{\phi_\epsilon(E)}{\Gamma \left( \frac{1}{2} - E \right)}$$

$$= - \frac{\partial}{\partial E} \ln \Delta_\epsilon(E), \quad (2.186)$$

where

$$\Delta_\epsilon(E) = \frac{1}{\Gamma \left( \frac{1}{2} - E \right)} - \frac{\lambda}{\Gamma \left( \frac{1}{2} - E \right)} \frac{\mathcal{I}(E - \frac{1}{2}, \mu)}{\Gamma \left( \frac{1}{2} - E \right)} \quad (2.187)$$

Let us now evaluate the integral (2.179) in the limit $\mu \to +\infty$, and thus $g \to 0^-$. We change variables, setting $\mu e^{-\theta} = t$, and the integral becomes

$$\mathcal{I}(s, \mu) = \mu^s \int_0^\mu dt \ t^{-1-s} e^{-t} = \mu^s \int_0^{+\infty} dt \ t^{-1-s} e^{-t} + O \left( e^{-\mu/\sqrt{\mu}} \right). \quad (2.188)$$

We thus obtain

$$\mathcal{I}(s, \mu) \approx \mu^s \Gamma(-s), \quad (2.189)$$

since for $\mu \to +\infty$ the difference is exponentially small. Therefore, our ansatz is that the estimate (2.189) gives the correct leading behaviour of the true function. Using this ansatz, we find

$$\Delta_\epsilon(E) = \frac{1}{\Gamma \left( \frac{1}{2} - E \right)} - \lambda \mu e^{-\frac{1}{2}} = \frac{1}{\Gamma \left( \frac{1}{2} - E \right)} + \frac{\lambda}{\Gamma \left( \frac{1}{2} - E \right)} \frac{\mathcal{I}(E - \frac{1}{2}, \mu)}{\Gamma \left( \frac{1}{2} - E \right)} \quad (2.190)$$

The energies are located at the poles of $G_\epsilon(E, g)$, but since

$$G_\epsilon(E, g) = -\frac{\Delta'_\epsilon(E)}{\Delta_\epsilon(E)} \quad (2.191)$$

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

$$\Delta_\epsilon(E) = 0 \quad (2.192)$$

as

$$\frac{\sin \pi (E - 1/2)}{\pi} = \frac{\lambda \mu^{E-1/2}}{\Gamma(E + 1/2)}. \quad (2.193)$$

38
Notice that, for \( \lambda = 0 \), the zeroes indeed take place at \( 1/2 + N \), therefore
\[
E_{c,N}^{(0)} = \frac{1}{2} + N + \mathcal{O}(\lambda).
\]
(2.194)

Using now (2.193) we find a series of the form,
\[
E_{c,N}(g) = \sum_{n=0}^{\infty} E_{c,N}^{(n)}(g) \lambda^n,
\]
(2.195)
where \( E_{N}^{(n)}(g) \) is the perturbative series around the \( n \)–th instanton solution. To find the leading order in \( g \) of \( E_{c,N}(g) \), we write
\[
E = \frac{1}{2} + N + x, \quad x = \sum_{n=1}^{\infty} E_{c,N}^{(n)}(g) \lambda^n,
\]
(2.196)
where \( x \) solves the implicit equation
\[
\frac{\sin \pi x}{\pi} + \frac{\hat{\lambda} e^{\xi x}}{\Gamma(1 + N + x)} = 0,
\]
(2.197)
and
\[
\hat{\lambda} = \left( \frac{2}{g} \right)^N \lambda, \quad \xi = \log \mu = \log \left( -\frac{2}{g} \right).
\]
(2.198)
This equation can be solved for \( x \) as a power series in \( \lambda \) after expanding in \( x \). One finds,
\[
x + \frac{\hat{\lambda}}{N!} + \hat{\lambda} x \frac{\xi - \psi(1 + N)}{N!} + \cdots = 0
\]
(2.199)
Therefore, at leading order one has
\[
x = -\frac{\hat{\lambda}}{N!} + \frac{\xi - \psi(1 + N)}{(N!)^2} \hat{\lambda}^2 + \mathcal{O}(\hat{\lambda}^3),
\]
(2.200)
which means that the one and the two-instanton contributions at one-loop are given by,
\[
E_{c,N}^{(1)}(g) = -\frac{1}{N!} \left( \frac{2}{g} \right)^{N+1/2} \frac{e^{-1/6g}}{\sqrt{2\pi}} (1 + \mathcal{O}(g)),
\]
\[
E_{c,N}^{(2)}(g) = \frac{1}{(N!)^2} \left( \frac{2}{g} \right)^{2N+1} \frac{e^{-1/3g}}{2\pi} \left\{ \ln(-2/g) - \psi(N + 1) + \mathcal{O}(g \ln g) \right\}.
\]
(2.201)
Notice that a single equation, (2.193), gives all the multi-instanton contributions to all energy eigenvalues \( E_{c,N}(g) \) of the double-well potential at leading order in \( g \). It is obvious
from the form of (2.193) that the \( n \)-th instanton contribution has at leading order the form

\[
E_{\epsilon,N}^{(n)}(g) = \left( \frac{2}{g} \right)^{n(N+1/2)} \left( -\epsilon \frac{e^{-1/6g}}{\sqrt{2\pi}} \right)^n \left\{ P_n^N(\ln(-g/2)) + \mathcal{O}(g(\ln g)^{n-1}) \right\},
\]

(2.202)
in which \( P_n^N(\xi) \) is a polynomial of degree \( n - 1 \). The first three polynomials are

\[
P_1^N(\xi) = 1, \\
P_2^N(\xi) = \xi + \psi(1 + N), \\
P_3^N(\xi) = \frac{3}{2} (\xi + \psi(1 + N))^2 - \frac{1}{2} \psi'(1 + N).
\]

(2.203)

3 Unstable vacua in QFT

Classic references for this topic are Coleman’s papers [22, 19], reviewed in [23].

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

\[
S(\phi) = \int \! d^d x \left( \frac{1}{2} (\partial_\mu \phi)^2 + U(\phi) \right)
\]

(3.1)

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

\[
U(\phi) = \frac{1}{2} \phi^2 - \frac{1}{2} \phi^3 + \frac{\alpha}{8} \phi^4
\]

(3.2)

where

\[
0 < \alpha < 1.
\]

(3.3)

This potential is represented in Fig. 13 for different values of \( \alpha \).

This potential has a relative, “false” minimum at \( \phi_- = 0 \), a true minimum at

\[
\phi_+ = \frac{3}{2\alpha} + \frac{\sqrt{9 - 8\alpha}}{2\alpha},
\]

(3.4)

and a local maximum at

\[
\phi = \frac{3}{2\alpha} - \frac{\sqrt{9 - 8\alpha}}{2\alpha}.
\]

(3.5)
We also have that
\[ U(\phi_+) - U(\phi_-) = \frac{4 (-2\alpha + 2\sqrt{9 - 8\alpha} + 9) \alpha - 9 (\sqrt{9 - 8\alpha} + 3)}{16\alpha^3}, \] (3.6)
and for \( \alpha = 1 \) the two minima are degenerate. This is called the thin wall limit.

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
\[ \left( -\nabla^2 - \frac{d^2}{d\tau^2} \right) \phi + U'(\phi) = 0, \] (3.7)
where \( \nabla \) is the gradient in three spatial dimensions, and \( \tau \) 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,
\[ \phi(\vec{x}, \tau) \to \phi_+, \quad \tau \to \pm \infty. \] (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
\[ \phi(\vec{x}, \tau) \to \phi_+, \quad |\vec{x}| \to \infty. \] (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

$$\det \frac{\delta^2 S}{\delta \phi^2} < 0$$  \hspace{1cm} (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.

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

We also expect that the most symmetric solution is the one with least action, and indeed this turns out to be the case (see [23], chapter 7, section 6.2, and references therein). The equation of motion reduces to

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

The boundary condition translates into

$$\lim_{r \to \infty} \phi_c = \phi_-. \quad \text{(3.13)}$$

Regularity at the origin demands that

$$\frac{d \phi_c}{dr} \bigg|_{r=0} = 0. \quad \text{(3.14)}$$

Analytic solutions to (3.12) are not available for nontrivial potentials, but one can show that solutions indeed exist (see [23] 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. 14 for different values of $\alpha$ (this figure is taken from the paper [4]).

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

If we change variables
\[ x \rightarrow \lambda x \] (3.16)
we find
\[ S(\phi_c, \lambda) = \lambda^{2-d} \int d^d x \frac{1}{2} (\partial_\mu \phi_c(x))^2 + \lambda^{-d} \int d^d x U(\phi_c(x)). \] (3.17)

Since \( \phi_c(x) \) satisfies the EOM, the action is stationary under variations of \( \lambda \):
\[ \frac{dS(\phi_c, \lambda)}{d \lambda} \bigg|_{\lambda=1} = (2 - d) \int d^d x \frac{1}{2} (\partial_\mu \phi_c(x))^2 - d \int d^d x U(\phi_c(x)) = 0. \] (3.18)

Therefore,
\[ \int d^d x U(\phi_c(x)) = \frac{2 - d}{d} \int d^d x \frac{1}{2} (\partial_\mu \phi_c(x))^2 \] (3.19)
and
\[ S_c = S(\phi_c) = \frac{1}{d} \int d^d x (\partial_\mu \phi_c(x))^2. \] (3.20)

Notice that there are now \( d \) zero modes, corresponding to translation invariance of the bounce in \( d \) dimensions. The corresponding functions are
\[ \phi_\mu = \partial_\mu \phi_c. \] (3.21)
with norm
\[ \int d^d x \phi_\mu \phi_\nu = \frac{1}{d} \delta_{\mu\nu} \int d^d x (\partial_\mu \phi_c(x))^2 = \delta_{\mu\nu} S_c, \]  \tag{3.22}
and the normalized zero modes are
\[ \phi^{(0)}_\mu = \frac{1}{S_c^2} \partial_\mu \phi_c. \]  \tag{3.23}
We have then
\[ \delta \phi = \partial_\mu \phi \delta x^\mu = \phi^{(0)}_\mu \delta c^{(0)}_\mu. \]  \tag{3.24}
The second equality is due to \( O(d) \) invariance of the solution. In analogy with QM, the zero modes contribute to the integral
\[ \frac{1}{(2\pi)^{d/2}} \int \prod_{\mu=1}^d dc^{(0)}_\mu = S_{c}^{d/2} \int \prod_{\mu=1}^d dx_\mu = \frac{S_{c}^{d/2} V \beta}{(2\pi)^{d/2}}, \]  \tag{3.25}
where \( V \) is the volume of \((d-1)\)-dimensional space and \( \beta \) 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
\[ -\frac{d^2}{d\tau^2} - \nabla^2 + U''(\phi_-), \]  \tag{3.26}
This plays the role of \( M_0 \) in the QM case. We can then write
\[ \text{Im} \frac{E}{V} = \frac{1}{2(2\pi)^{d/2}} \frac{S_{c}^{d/2} V \beta}{|\det(-d^2/d\tau^2 - \nabla^2 + U''(\phi_-))|^{\frac{1}{2}}} e^{-S_c} \]  \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

\[ e^{-S_c} \rightarrow e^{-(S_c-S(\phi_-))}. \] (3.30)

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

\[ \phi_c \rightarrow \phi_c + \hbar \phi^{(1)} + \cdots \] (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

\[
S(\phi) = S_R(\phi_c + \hbar \phi^{(1)} + \cdots) + \hbar S^{(1)}(\phi_c + \hbar \phi^{(1)} + \cdots) + \cdots
\] (3.32)

since

\[
\frac{\delta S_R}{\delta \phi}(\phi_c) = 0
\] (3.33)

by construction. We then find

\[
\text{Im } E/V = \frac{1}{2} \frac{S_R^{d/2} V \beta}{(2\pi)^{d/2}} \left| \det \left( \frac{-d^2}{d\tau^2} - \nabla^2 + U''(\phi_c) \right) \right|^{-\frac{1}{2}} e^{-S_c + S(\phi_-)}
\]

\[
\approx \frac{1}{2} \frac{S_R^{d/2}(\phi_c) V \beta}{(2\pi)^{d/2}} \left| \frac{\det \left( \frac{-d^2}{d\tau^2} - \nabla^2 + U''(\phi_c) \right) \left| \det \left( \frac{-d^2}{d\tau^2} - \nabla^2 + U''(\phi_-) \right) \right|^{-\frac{1}{2}} e^{-S_R(\phi_c) - \hbar S^{(1)}(\phi_c) + S^{(1)}(\phi_-)}
\] (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 [4, 33].

### 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. 6). 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

$$\phi(t = 0, \vec{x}) = \phi_c(\vec{x}, \tau = 0),$$
$$\partial_t \phi(t = 0, \vec{x}) = 0.$$  \hspace{1cm} (3.35)

The evolution will be governed by the wave equation

$$(\nabla^2 - \partial_t^2)\phi = U'(\phi).$$  \hspace{1cm} (3.36)

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

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

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

$$\partial_t \phi = \frac{d\phi_c}{dr} \partial_t r = -t \frac{d\phi_c}{dr}$$  \hspace{1cm} (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 (14) 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

$$\vec{x}^2 = t^2 + R^2.$$  \hspace{1cm} (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 [70].

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

$$X_5 = M_4 \times S^1,$$  \hspace{1cm} (3.40)
where $M_4$ is Minkowski space and $S^1$ 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:

$$ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2 + d\phi^2. \quad (3.41)$$

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

$$ds^2 = dr^2 + r^2 d\Omega^2 + d\phi^2, \quad (3.42)$$

where

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

Is there a bounce? It turns out that

$$ds^2 = \frac{dr^2}{1 - \alpha/r^2} + r^2 d\Omega^2 + \left(1 - \frac{\alpha}{r^2}\right)d\phi^2 \quad (3.44)$$

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 [39], i.e. we want to interpret the singularity at $r = \alpha$ as an apparent singularity due to the fact that the the coordinate $\phi$ is periodic. Indeed, the flat metric in polar coordinates

$$ds^2 = dx^2 + x^2 d\phi^2 \quad (3.45)$$

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

$$\rho = c \left(1 - \frac{\alpha}{r^2}\right)^\beta, \quad (3.46)$$

47
where $c, \beta$ are constants to be determined by our requirements. We have

\begin{equation}
    r^2 = \frac{\alpha}{1 - \left(\frac{\rho}{c}\right)^{\frac{1}{\beta}}},
\end{equation}

\begin{equation}
    dr = \frac{r^3}{2c\alpha\beta} \left(1 - \frac{\alpha}{r^2}\right)^{1-\beta},
\end{equation}

and we deduce

\begin{equation}
    dr^2 = \frac{\alpha}{4c^2\beta^2} \left(\frac{\rho}{c}\right)^{\frac{2(1-\beta)}{\beta}} \frac{d\rho^2}{\left(1 - \left(\frac{\rho}{c}\right)^{\frac{1}{\beta}}\right)^3},
\end{equation}

as well as

\begin{equation}
    \frac{dr^2}{1 - \alpha/r^2} = \frac{\alpha}{4c^2\beta^2} \left(\frac{\rho}{c}\right)^{\frac{1-2\beta}{\beta}} \frac{d\rho^2}{\left(1 - \left(\frac{\rho}{c}\right)^{\frac{1}{\beta}}\right)^3}.
\end{equation}

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

\begin{equation}
    \beta = \frac{1}{2}, \quad c = \alpha^{\frac{1}{2}}.
\end{equation}

We can now analyze the periodic part,

\begin{equation}
    \left(1 - \frac{\alpha}{r^2}\right) d\phi^2 = \left(\frac{\rho}{c}\right)^{\frac{1}{\beta}} d\phi^2,
\end{equation}

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

\begin{equation}
    \frac{\rho^2}{c^2} d\phi^2 = \rho^2 d\left(\frac{\phi}{c}\right)^{\frac{1}{2}}.
\end{equation}

We conclude that $\phi$ is periodic with period

\begin{equation}
    2\pi c = 2\pi \sqrt{\alpha}.
\end{equation}

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

\begin{equation}
    \alpha = R^2.
\end{equation}

Therefore, the metric reads

\begin{equation}
    ds^2 = \frac{dr^2}{1 - (R/r)^2} + r^2 d\Omega_3^2 + \left(1 - \left(\frac{R}{r}\right)^2\right) d\phi^2
\end{equation}

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

\begin{equation}
    r \geq R.
\end{equation}
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. This continuation has now two steps. First of all, we must rotate the resulting metric (3.55) to Minkowski signature. To do that, we recall that
\[ d\Omega_3^2 = d\theta^2 + \sin^2 \theta d\Omega_2^2 \] (3.57)
and we define a new angle \( \psi \) by
\[ \theta = \frac{\pi}{2} + i\psi. \] (3.58)
The line element of the three-sphere becomes
\[ d\Omega_3^2 = -d\psi^2 + \cosh^2 \psi d\Omega_2^2, \] (3.59)
and 4d Euclidean space becomes,
\[ dr^2 + r^2 d\Omega_3^2 \rightarrow dr^2 + r^2(-d\psi^2 + \cosh^2 \psi d\Omega_2^2). \] (3.60)
This is 4d Minkowski. Indeed, if we write
\[ x = r \cosh \psi, \quad t = r \sinh \psi, \] (3.61)
we find
\[ dr^2 - r^2 d\psi^2 + r^2 \cosh^2 \psi d\Omega_2^2 = dx^2 - dt^2 + x^2 d\Omega_2^2, \] (3.62)
so that \( x \) is here the standard \( O(3) \)-invariant radial distance. To compute this, we use that
\[ r = \sqrt{x^2 - t^2}, \quad dr = \frac{1}{r}(xdx - tdt), \]
\[ \psi = \tanh^{-1} t \frac{1}{x}, \quad d\psi = \frac{1}{r^2}(xdt - tdx). \] (3.63)
Notice however that in this parametrization \( x^2 - t^2 = r^2 > 0 \) is always positive. Therefore, to be precise, the above continuation (3.60) 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
\[ ds^2 = \frac{dr^2}{1 - (R/r)^2} + r^2(-d\psi^2 + \cosh^2 \psi d\Omega_2^2) + \left(1 - \left(\frac{R}{r}\right)^2\right)d\phi^2. \] (3.64)
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
\[ x^2 - t^2 = R^2, \] (3.65)
Figure 15: Restricted to the $x-t$ plane, the space described by the metric (3.64) is the exterior of the hyperboloid $x^2 - t^2 = R^2$ (in red in the figure).

see Fig. 15. 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

$$R(1 - R^2/r^2)^{1/2}$$

and so it becomes zero as $r \to R$. This smooths out what would be the complement of an interval into two discs.

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

$$x^2 = R^2 + t^2.$$  \hspace{1cm} (3.67)

Therefore, the Kaluza–Klein vacuum decays into literally nothing.
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 \( g \),

\[
Z = \sum_k Z_k g^k.
\]

(4.1)

A typical example of this is the anharmonic oscillator with Hamiltonian

\[
H = \frac{p^2}{2} + \frac{x^2}{2} + \frac{g}{4}x^4.
\]

(4.2)

where \( Z = E(g) \) 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:

- **Large order behavior.** What is the radius of convergence of (4.1)? We will see that very often we have zero radius of convergence.

- **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 [34] 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.

\(^1\)One should be careful with this general argument, since there are counterarguments to it. See [58], p. 4.
4.2 The anharmonic oscillator

As a first approach to the problem of large orders in perturbation theory, we study the case of the anharmonic oscillator. We consider a Hamiltonian of the form

\[ H = \frac{p^2}{2} + \frac{x^2}{4} + \frac{g}{4}x^4 \]  

(4.3)

where we follow the normalizations of [10]. The energy of the ground state can be computed as a power series in \( g \) in perturbation theory:

\[ E(g) = \sum_{k=0}^{\infty} a_k g^k, \quad a_0 = \frac{\hbar \omega}{2} \]  

(4.4)

Since the above system is unstable for \( g < 0 \) (the Hamiltonian is unbounded from below) we expect this series to have zero radius of convergence (by Dyson’s argument). The behavior of \( E_k \) at large \( k \) can be obtained in a more precise way as follows.

First, we establish the relation between the analytic structure and the large order behavior, following the original paper of Bender and Wu [10]. Consider the function

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

(4.5)

The function \( f(z) \), as a function in the complex \( z \)-plane, has the following properties:

- As in the case of the quartic integral studied in the Appendix, it is analytic in the complex plane with a cut along \((-\infty, 0)\).
- At the origin it behaves like
  \[ \lim_{z \to 0} zf(z) = 0. \]  

(4.6)

- At infinity it goes like
  \[ |f(z)| \sim |z|^{-2/3}. \]  

(4.7)

The last property follows from the following (see for example [36], p. 171). At large \( g \), we have that

\[ H \sim \frac{p^2}{2} + g\frac{x^4}{4}. \]  

(4.8)

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

\[ H \to g^{\frac{1}{6}}\left(\frac{p^2}{2} + \frac{x^4}{4}\right), \]  

(4.9)
therefore the energy will be
\[ E(g) \sim Cg^\frac{4}{3}, \quad g \to \infty, \] (4.10)
where \( C \) is the energy of the ground state of the Hamiltonian \( p^2/2 + x^4/4 \) in (4.9).

If we now use the Cauchy representation
\[ f(z) = \frac{1}{2\pi i} \oint_{C_z} dx \frac{f(x)}{z-x}, \] (4.11)
we can deform the contour to encircle the branch cut in the negative real axis. This gives,
\[ f(z) = \frac{1}{2\pi i} \int_{-\infty}^{0} \frac{D(x)}{x-z}, \] (4.12)
where \( D(x) \) is the discontinuity across the negative, real axis,
\[ D(x) = \lim_{\epsilon \to 0} (f(x+i\epsilon) - f(x-i\epsilon)) = 2i \text{Im} f(x+i0^+), \] (4.13)
i.e. we have
\[ f(z) = \frac{1}{\pi} \int_{-\infty}^{0} \frac{\text{Im} f(x)}{x-z}. \] (4.14)
In terms of the original quantity, we have
\[ E(g) = a_0 + \frac{g}{\pi} \int_{-\infty}^{0} \frac{d g'}{g'} \frac{\text{Im} E(g')}{g'(g'-g)}. \] (4.15)

From the above representation we find,
\[ f_k = \frac{1}{2\pi i} \int_{-\infty}^{0} dz \frac{D(z)}{z^{k+1}}, \] (4.16)
and
\[ a_k = \frac{1}{2\pi i} \int_{-\infty}^{0} dz \frac{D(z)}{z^k} = \frac{(-1)^k}{2\pi i} \int_{0}^{\infty} dz \frac{D(-z)}{z^k}, \quad k \geq 1. \] (4.17)
Therefore, if we know \( D(-z) \), we can plug it in here to obtain the asymptotics of \( a_k \).
Moreover, at large \( k \) this will be controlled by the behavior of \( D(z) \) at small, negative \( z \).
This is precisely the tunneling amplitude at weak coupling, which can be computed by instanton calculus!

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 (an instanton calculation).
Let us assume that
\[
\text{disc } E(-z) = \lim_{\varepsilon \to 0} (E(-z + i\varepsilon) - E(-z - i\varepsilon)) = 2i \text{Im } E(-z + i0_+)
\] (4.18)
is of the form
\[
\text{disc } E(-z) = i|z|^{-b} e^{-A/|z|} \sum_{n=0}^{\infty} c_n |z|^n,
\] (4.19)
where \(c_n\) is given by the \(n\)-loop fluctuation around an instanton solution. Therefore, \(D(-z)\) is of the form
\[
D(-z) = i|z|^{-b-1} e^{-A/|z|} \sum_{n=0}^{\infty} c_n |z|^n,
\] (4.20)
This leads to the following behavior for \(a_k\):
\[
a_k = \frac{(-1)^k}{2\pi} \sum_{n=0}^{\infty} c_n \int_0^{\infty} \frac{ds}{s^{k+b+n}} s^{-b-1+n} e^{-A/s}
\]
\[
= \frac{(-1)^k}{2\pi} \sum_{n=0}^{\infty} c_n A^{-k-b+n} \int_0^{\infty} dx x^{k+b+1-n} e^{-x} = \frac{(-1)^{k+1}}{2\pi} \sum_{n=0}^{\infty} c_n A^{-k-b+n} \Gamma(k+b-n)
\] (4.21)
This can be also written as (see [24])
\[
a_k \sim \frac{(-1)^k A^{-b-k}}{2\pi} \Gamma(k+b) \left[ c_0 + \frac{c_1 A}{k+b-1} + \frac{c_2 A^2}{(k+b-2)(k+b-1)} + \cdots \right].
\] (4.22)
Note that the leading term is the factorial \(k!\). The subleading piece is captured by \(A^{-k}\).

**Remark 4.1.** What one computes in an instanton calculation is precisely \(D(-z)\). In other words, the instanton contribution is \(2i\) the imaginary part of \(f(z)\).

We can now use the formula (2.129) for \(\text{Im } E\) in the quartic oscillator and read,
\[
b = \frac{1}{2}, \quad c_0 = 2 \sqrt{\frac{2}{\pi}}, \quad A = 4/3
\] (4.23)
and we find
\[
a_k \sim (-1)^{k+1} \frac{\sqrt{6}}{\pi^{3/2}} \left( \frac{3}{4} \right)^k \Gamma \left( k + \frac{1}{2} \right),
\] (4.24)
which is the famous result of [10]. 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. We will see how to make sense of this.
4.3 Large order behaviour and counting of Feynman diagrams

What is the source of the divergence of perturbation theory; reflected in the factorial growth of $a_k$? In this quantum-mechanical example, this turns out to be due to the factorial growth of the number of Feynman diagrams (see [6] for an excellent survey). Recall from section 2.1 that $a_n$ can be computed from a sum over connected quartic graphs. The total number of disconnected graphs is simply given by the quartic integral (4.59), which asymptotically as $n \to \infty$ behaves like (see (2.28)

$$ (16)^n n! , $$

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 [7] shows that this is not the case: at large $n$, the number of connected and disconnected diagrams differs from 1 only in $O(1/n)$ corrections. We conclude that there are $(16)^n n!$ diagrams that contribute to $a_n$. This leads to the factorial behavior in (4.24). In fact, one can derive the leading and subleading behavior in (4.24) by a statistical analysis of Feynman diagrams [11].

4.4 Asymptotic expansions and the nonperturbative ambiguity

Given the fact that the perturbative series is divergent, we have to make sense of it. The first try is to deduce that the perturbation expansion is the asymptotic expansion of an analytic function. This is a good try, but it still has some problems: there is an ambiguity in going from the asymptotic expansion to the function. Let us see this.

A series

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

is called asymptotic (in the Poincaré sense) if

$$ \lim_{w \to 0} w^{-N} \left( S(w) - \sum_{n=0}^{N} a_n w^n \right) = 0. $$

Let us assume that at large $n$

$$ a_n \sim (\beta n)!, $$

and let $f(w)$ be a function analytic in a region

$$ D = \{|w| < R, \arg(z) < \alpha \pi\}. $$
We say that the series (4.26) is a strong asymptotic series for \( f(w) \) if for all \( N \) there exists a bound,
\[
|f(w) - \sum_{n=0}^{N} a_n w^n| \leq C_{N+1}|w|^{N+1}
\]
with
\[
C_N = cA^{-N}(\beta N)!.
\]
Notice that (4.26) has in fact zero radius of convergence.

The asymptotic expansion above does not define uniquely the function \( f(w) \). This can be seen as follows: imagine that you want to find the best estimate of \( f(w) \). Then, one has to find the \( N \) that truncates the asymptotic expansion in an optimal way. In order to do that, we find the minimum of
\[
C_N|w|^N = M(\beta N)! \left( \frac{|w|}{A} \right)^N.
\]
By using the Stirling approximation, we rewrite this as
\[
M \exp N \{ \beta \log(\beta N) - \beta - \log X \},
\]
where
\[
X = \frac{A}{|w|}
\]
The above function has a saddle at large \( N \) given by
\[
N^* = \frac{1}{\beta} X^{\frac{1}{\beta}},
\]
and for this value of \( N \) the bound on the asymptotics is of the form
\[
\epsilon(w) = C_{N^*}|w|^{N^*} \sim \exp \left( -\frac{A}{|w|} \right)^{\frac{1}{\beta}}.
\]
It follows that we can add to the asymptotic expansion in a given region a function which is analytic and smaller than \( \epsilon(w) \). This ambiguity present in an asymptotic series is sometimes called (in the context of QFT and string theory) the nonperturbative ambiguity.

There are some cases in which this ambiguity is not present. For example, if \( \alpha = 1/2 \) in the region \( D \) in (4.29), a theorem due to Carleman shows that \( f(w) \) is uniquely determined by its asymptotic expansion \( S(w) \).
4.5 Borel transform and Borel summability

Given a divergent series, there are two questions we may ask about it:

- How do we assign a numerical value to the series?
- How is the series of its sum related to the exact answer or function \( f(w) \)?

In order to answer the second question, we need a nonperturbative definition of the theory which allows an independent definition of \( f(w) \), and this is not always available. In any case, an interesting procedure to address the first point is Borel resummation. This goes as follows. If we represent

\[
n! = \int_0^\infty dt \, e^{-t} t^n, \quad (4.37)
\]

we find

\[
S(w) = \sum_{n=0}^\infty \frac{a_n}{n!} \int_0^\infty dt \, e^{-t} (wt)^n \quad (4.38)
\]

Let us define the Borel transform of \( S \), \( B_S(z) \), as the series

\[
B_S(z) = \sum_{n=0}^\infty \frac{a_n}{n!} z^n. \quad (4.39)
\]

It follows that

\[
f(w) = \int_0^\infty dt \, e^{-t} B_S(tw), \quad (4.40)
\]

or equivalently

\[
f(w) = w^{-1} \int_0^\infty dt \, e^{-t/w} B_S(t) \quad (4.41)
\]

defines an analytic continuation of \( S(w) \). Notice that the inverse of the Borel transform, (4.41), is essentially the familiar Laplace transform.

If \( S(w) \) has a finite radius of convergence, then \( B_S(z) \) is analytic on the full complex plane. But even if \( S(w) \) has zero radius of convergence, the Borel transform can still lead to a nice \( f(w) \).

**Definition 4.2.** We say that the series \( S(w) \) is Borel summable if

- the series \( B_S(z) \) has a finite radius of convergence and it has an analytic continuation to the whole sector \( |\arg(z)| < \epsilon \) (i.e. in a neighbourhood of \((0, \infty)\)), and
- The integral (4.40) is absolutely convergent for small \( w \).
Example 4.3. Consider
\[ S(w) = \sum_{n=0}^{\infty} (-1)^n n! w^n. \] (4.42)

In this case, the Borel transform is
\[ B_S(z) = \sum_{n=0}^{\infty} (-1)^n z^n = \frac{1}{1 + z}, \] (4.43)
and we have
\[ f(w) = \int_0^{\infty} dt e^{-t} \frac{1}{1 + wt} \] (4.44)
which exists and is well defined as long as \( w > 0 \).

Example 4.4. Consider now the series
\[ S(w) = \sum_{k=0}^{\infty} \frac{\Gamma(k + b)}{\Gamma(b)} A^{-k} s^k. \] (4.45)

The Borel transform is given by
\[ B_S(z) = \sum_{k=0}^{\infty} \frac{\Gamma(k + b)}{k! \Gamma(b)} A^{-k} z^k = (1 - z/A)^{-b}, \] (4.46)
which has a singularity at \( z = A \) as well as a branch cut starting at that point. In many cases in quantum field theory and quantum mechanics, even if the coefficients of a series are not as simple as in (4.45), they have at large \( k \) the asymptotic behavior is
\[ a_k \sim A^{-k} \Gamma(k + b), \quad k \gg 1, \] (4.47)
where \( A \) is a constant. If we use that
\[ \frac{\Gamma(k + b)}{\Gamma(k + 1)} \sim k^{b-1} \] (4.48)
which can be derived from Stirling’s formula, we can also write
\[ a_k \sim k! A^{-k} k^{b-1} \] (4.49)
The Borel transform for such a series will be of the form
\[ B_S(z) = C (1 - z/A)^{-b} + \cdots \] (4.50)
and there will be corrections due to the subleading corrections to the asymptotics of $a_k$. When $b = 0$ we have a logarithmic branch cut for $B_S(s)$. For example, it follows from (2.28) that the formal power series expansion of the quartic integral (2.25) is of the form (4.49) with

$$A = -\frac{1}{4},$$

and its Borel transform is

$$B(z) = -\frac{1}{\pi \sqrt{2}} \log (1 + 4z),$$

with a logarithmic branch point at $s = A$.

The above example shows, conversely, that the Borel transform encodes the data in (4.49) determining the large order behavior of the coefficients $a_k$: the factor $A$ is given by the location of the singularity of the Borel transform which is closest to the origin, and $b$ is the characteristic exponent of its branch cut. We conclude that the large order behavior of a divergent series is controlled by the singularities of its Borel transform.

Let us now present a general formula giving a quantitative expression for this relation. If

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

then

$$a_k = k! b_k = \int_0^\infty dt e^{-t} t^k b_k = \frac{1}{2\pi i} \int_0^\infty dt e^{-t} t^k \int_0^{\infty} \frac{du}{u^{k+1}} B_S(u).$$

Suppose then that $B_S(u)$ has a singularity at $A$ with a cut going to infinity (for simplicity, we will assume that $A$ is on the real axis). By deforming the contour around the origin to the contour $\gamma$ encircling the cut and the singularity, as shown in Fig. 16, we get

$$a_k = \frac{1}{2\pi i} \int_0^\infty dt \oint_\gamma ds e^{-t} \frac{t^k}{s^{k+1}} B_S(s) = \frac{1}{2\pi i} \int_0^\infty \frac{dz}{z^{k+1}} \left[ z^{-1} \oint_\gamma du e^{-u/z} B_S(u) \right]$$

Figure 16: Contour deformation in the derivation of (4.55).
where we have introduced the variable \( z = t/u \) We can write the contour \( \gamma \) as the difference of two contours,

\[
\gamma = C_{0,+} - C_{0,-},
\]

where \( C_{0,\pm} \) are lines starting from the origin and going above (respectively, below) the real axis. Therefore,

\[
z^{-1} \oint_{\gamma} du \, e^{-u/z} B_S(u) = z^{-1} \left[ \oint_{C_{0,+}} - \oint_{C_{0,-}} \right] du \, e^{-u/z} B_S(u) = f_+(z) - f_-(z),
\]

i.e. the difference between the functions \( f_+ \) above and below the cut. Finally we can write

\[
a_k = \frac{1}{2\pi i} \int_0^\infty \frac{dz}{z^{k+1}} (f_+(z) - f_-(z)).
\]

We will denote the discontinuity in \( f \) by

\[
\epsilon(s) = \frac{1}{2\pi i} (f_+(s) - f_-(s)) = \frac{1}{\pi} \text{Im} \, f(s),
\]

so that

\[
a_k = \int_0^\infty \frac{dz}{z^{k+1}} \epsilon(z).
\]

This generalizes (4.16).

Notice that \( f_\pm \) can be regarded as two different functions with the same asymptotics. Their difference should be exponentially suppressed, as we noticed above when discussing the nonperturbative ambiguity. Indeed, a Borel transform \( B_S(s) \) with a branch cut starting at \( A \) and with exponent \(-b\) leads to the following form for \( \epsilon(w) \)

\[
\epsilon(z) \sim z^{-b} e^{-A/z}.
\]

This can be seen by direct computation

\[
\epsilon(z) = \frac{1}{\pi} z^{-1} \int_A^\infty dt \, e^{-t/z} \text{Im} \, B_S(t) = \frac{1}{\pi} z^{-1} \int_A^\infty dt \, e^{-t/z} \text{Im} \, (1 - t/A)^{-b}
\]

\[
= \frac{1}{\pi} z^{-1} e^{-A/z} A^b \int_0^\infty du \, e^{-u/z} u^{-b} \text{Im} \, e^{-\pi ib}
\]

\[
= \frac{1}{\pi} e^{-A/z} \frac{A^b}{\Gamma(1 - b)} \sin \pi (1 - b)
\]

\[
= \frac{A^b}{\Gamma(b)} \frac{e^{-A/z}}{z^{-b}},
\]

60
where in the second line we introduced the variable $u = t - A$, and in the last line we used the identity
\[
\Gamma(z)\Gamma(1 - z) = \frac{\pi}{\sin \pi z}.
\] (4.63)

We can also check that (4.61) gives the expected behavior for $a_k$:
\[
a_k = \int_0^\infty \frac{dz}{z^{k+1}} z^{-b} e^{-A/z} \sim A^{-k-b} \int_0^\infty dx x^{k+b-1} e^{-x} \sim A^{-k-b} \Gamma(k + b). \tag{4.64}
\]

Example 4.5. Let us obtain $a_k$ in the case (4.62). We obtain,
\[
a_k \sim A^{-k-b} \Gamma(k + b) \frac{1}{\pi} \frac{\pi A^b}{\Gamma(b)} = \frac{\Gamma(k + b)}{\Gamma(b)} A^{-k}, \tag{4.65}
\]
so we recover the original expression for the coefficients of the series.

Example 4.6. A closely related example is the series
\[
S(w) = \sum_{k=0}^\infty \frac{\Gamma(k + \gamma)}{\Gamma(k)} k! w^k \tag{4.66}
\]
which is analyzed in [46]. It turns out that this is the asymptotic expansion of the function
\[
f(w) = \Gamma(\gamma) w \frac{d}{dw} {}_2F_0(1, \gamma; w), \tag{4.67}
\]
This function has a branch cut on the positive real axis, and its discontinuity is given by
\[
2\text{Im } f(w) = 2\pi (1 - w^\gamma) w^{-\gamma-1} e^{-1/w}. \tag{4.68}
\]
We can check that this discontinuity gives back the form of the $a_k$ by using (4.22). Since the cut is in the positive real axis, we don’t have to change sign in that derivation. We have
\[
b = \gamma + 1, \quad A = 1, \quad c_0 = 2\pi, \quad c_1 = -2\pi\gamma. \tag{4.69}
\]
Therefore,
\[
a_k = \frac{1}{2\pi} \Gamma(k + \gamma + 1) \left[ 2\pi - \frac{2\pi\gamma}{k + \gamma} \right] = \Gamma(k + \gamma + 1) \frac{k}{k + \gamma} \frac{\Gamma(k + \gamma)}{\Gamma(k)} k!, \tag{4.70}
\]
which is exactly the original expression for the coefficients. Notice that the Borel transform of $S(w)$ is
\[
B_S(w) = \sum_{k=0}^\infty \frac{\Gamma(k + \gamma)}{\Gamma(k)} w^k = w \frac{d}{dw} \sum_{k=0}^\infty \frac{\Gamma(k + \gamma)}{\Gamma(k + 1)} w^{k+1}
\]
\[
= \Gamma(\gamma) w \frac{d}{dw} (1 - w)^{-1-\gamma} = \Gamma(1 + \gamma) \frac{z}{(1 - z)^{\gamma+1}}. \tag{4.71}
\]
If we are interested in understanding the large order behavior of the original series, it is irrelevant whether $A$ is positive or negative, the only difference being that for $A$ negative the series is alternating. However, if we want to use (4.41) to reconstruct the analytic function $f(w)$, a singularity of $B_S$ on the positive real axis gives an obstruction to performing the integral. This is typically the case when the starting series is factorially divergent and nonalternating in sign. It is usually said that in such a case the series $S(w)$ is not Borel summable. More precisely, in order to define $f(w)$ we have to choose a contour in (4.41) which avoids the singularities in the real axis. For example, we could choose contours $C_\pm$ that avoid the singularities by encircling them from above or from below, respectively, as in Fig. 17. The functions

$$f_\pm(z) = w^{-1} \int_{C_\pm} dt \, e^{-t/w} B_S(t)$$

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 (4.59) and is nonperturbative. Different choices of contour in the Laplace transform (4.41) 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.

**Example 4.7.** Let us consider the series

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

![Figure 17: The paths $C_\pm$ avoiding the singularities of the Borel transform from above (respectively, below).](image-url)
which is (4.42) but without the sign \((-1)^n\) (or we consider negative values of \(w\)). Then, the Borel transform is

\[
f(w) = \frac{1}{w} \int_0^\infty dt \frac{e^{-t/w}}{1 - t}.
\]  

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

\[
f_+(w) - f_-(w) = \frac{1}{w} \oint_C dt \frac{e^{-t/w}}{1 - t} = \frac{2\pi i}{w} e^{-1/w},
\]  

where \(C\) is a circle surrounding the pole. This is in fact a particular case of the computation we did in (4.62).

When there are poles of \(B_S(z)\) along the 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.

### 4.6 Instantons and large order behavior

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? As we noticed above, 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 \(\phi_\ast\) with finite action \(S(\phi_\ast)\), the Borel transform of any correlation function will be singular at \(S(\phi_\ast)\) and the corresponding perturbative series has a zero radius of convergence.

There is a heuristic argument for this due to ‘t Hooft [61]. We write a correlation function

\[
W(\alpha) = \int d\phi e^{-\frac{\alpha}{\pi}S(\phi)} \phi(x_1) \cdots \phi(x_n)
\]  

as

\[
W(\alpha) = \alpha \int_0^\infty dt \int D\phi \delta(\alpha t - S(\phi)) e^{-\frac{\alpha}{\pi}S(\phi)} \phi(x_1) \cdots \phi(x_n)
\]

\[
= \alpha \int_0^\infty dt F(\alpha t) e^{-t},
\]

where we used that

\[
\alpha \int_0^\infty dt \delta(\alpha t - S(\phi)) = 1,
\]
and we wrote

\[ F(z) = \int D\phi \delta(z - S(\phi))\phi(x_1)\cdots\phi(x_N). \]  \hspace{1cm} (4.79)

By comparing (4.77) to (4.40) 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, we will have complex instanton solutions with non-positive action \( S(\phi_*) \).

The leading large order behavior is given by the solution with the smallest action in absolute value. The phase of the action determines the oscillatory character of the series.

As we saw in the example of the quartic oscillator, the divergence of the perturbative series is due to the growth in the number of diagrams. 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. There are other, very different sources of factorial divergence in perturbation theory, encoded in the so-called renormalons, which we will study later.

### 4.7 Large order behavior in quantum mechanics

Taking into account the general discussion above, we can have two different behaviors in QM:

- If we expand around an absolute minimum of a potential in QM, there are no positive instanton solutions. We expect that the poles of the Borel transform are not on the real axis and the perturbative series for the ground state energy is in principle Borel summable. This is what happens in the case of the quartic, anharmonic oscillator.

- 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, and for real values of the coupling the perturbation series is not Borel summable. In some situations, however, one can still extract the physical quantities from the perturbative series. This leads in general to a complex perturbation series whose imaginary part is exponentially suppressed and represents the decay amplitude. This is what happens for example for the Stark effect or for the cubic anharmonic oscillator.

We now discuss various examples.

**Example 4.8. The cubic oscillator.** Perturbation theory gives a series for the ground state energy \( E_0(g) \) of the form

\[ E_0(g) = \sum_{n=0}^{\infty} a_ng^{2n}. \]  \hspace{1cm} (4.80)
The one-instanton contribution to the imaginary part of the ground state energy was computed in (2.134). Since \( b = 1/2 \), the leading asymptotics is
\[
a_k \sim \frac{1}{2\pi} \Gamma(k + 1/2) A^{-k-1/2} c_0, \tag{4.81}
\]
where the various quantities \( A, c_0 \) can be read from (2.134). The large order behavior is
\[
a_n \sim -\frac{(60)^{n+1/2}}{(2\pi)^{3/2} 2^{3n}} \Gamma(n + 1/2) \tag{4.82}
\]
as computed in for example [3].

Notice that the series is nonalternating, therefore \( E_0(g) \) is not Borel summable in the sense explained above. This is just reflecting the fact that the energy levels are unstable. The “true” energies are complex, and their imaginary parts have a very clear, physical meaning: they give the width of the energy levels. These complex energies can be computed by first calculating the Borel resummation of the series \( E_0(g) \) for complex values of the coupling constant, and then by analytically continuing the result to real values of \( g \). In fact, if \( g \) is pure imaginary, the series (4.80) is real and Borel summable: the energy \( E_0(g) \) can be computed by Borel resummation and it is real when \( g \) is imaginary (see for example, [8]).

In practice, when \( g \) is real, one computes
\[
E_0(g) = e^{i\theta} \int_0^\infty e^{-e^{it} B_S(e^{it}g^2)} dt, \tag{4.83}
\]
where \( \theta \) is chosen in such a way that \( B_S(re^{it}g^2) \) is not singular for \( t > 0 \). One can choose for example \( \theta = \pi/4 \) as in [3]. Of course, the choice of \( \theta \) can be regarded as a choice of integration contour in the complex \( t \) plane, and the prescription (4.83) is closely related to lateral Borel resummations.

**Example 4.9. More general potentials.** Consider a particle situated at the origin of the potential
\[
V(x) = \frac{1}{2} x^2 - \gamma x^3 + \frac{1}{2} x^4. \tag{4.84}
\]
There are two different situations here:

- For \( |\gamma| > 1 \), the origin is not an absolute minimum, which is in fact at
  \[
x_0 = \frac{3\gamma + \sqrt{-8 + 9\gamma^2}}{4}. \tag{4.85}
  \]
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

$$x_+ = \gamma - \sqrt{\gamma^2 - 1}. \quad (4.86)$$

The action of this instanton can be written as

$$A = \int_0^{x_+} (2V(x))^{1/2} = -\frac{2}{3} + \gamma^2 - \frac{1}{2}\gamma(\gamma^2 - 1) \log \frac{\gamma + 1}{\gamma - 1}, \quad (4.87)$$

while the prefactor reads,

$$C = -\frac{1}{\pi^{3/2}}(\gamma^2 - 1)^{-1/2}A^{-1/2}. \quad (4.88)$$

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

$$x = g \pm i\sqrt{1 - g^2}. \quad (4.89)$$

We have then to add the contributions of both instantons,

$$E_k = -\frac{1}{\pi^{3/2}}\Gamma(k + 1/2) \left[ A^{-k-1/2}i(1 - \gamma^2)^{-1/2} - A^{-k-1/2}i(1 - \gamma^2)^{-1/2} \right], \quad (4.90)$$

For $\gamma = 0$ (the quartic potential) we find,

$$\text{Im} A^{-k-1/2} = (-1)^{1+k} \left( 3 \sqrt{2} \right)^{k+1/2}, \quad (4.91)$$

and the final result for the asymptotics is

$$E_k = \frac{(-1)^{k+1} \sqrt{6}}{\pi^{3/2}} \left( 3 \sqrt{2} \right)^k, \quad (4.92)$$

which agrees with the previous result after taking into account the relative normalization of $g$, which adds a factor $2^k$.

**Example 4.10. The double-well.** The perturbative energy of the ground state

$$E^{(0)}(g) \equiv E_{+,0}(g) \quad (4.93)$$
is given by the series (2.136). The coefficients in this series grow factorially, and the series is not Borel summable (all coefficients except the leading term have negative sign). The nonperturbative ambiguities in the perturbative expansion are of order

\[ e^{-1/3g} \]  

(4.94)

which correspond to the two–instanton amplitude \( E^{(2)}(g) \). It is possible to show (see for example [74]) that the perturbation series is Borel summable for \( g < 0 \). This is related to the fact that the instanton corrections are only well-defined for negative \( g \) as well, as we saw in section 2.

We then define the resummed perturbative series \( E^{(0)}_\pm(g) \) as the analytic continuation of this Borel sum from \( g \) negative to \( g = |g| \pm i0 \). Equivalently, we can obtain \( E^{(0)}_\pm(g) \) by using any of the two lateral Borel resummations along the contours \( C_\pm \) in Fig. 17. The difference between these two prescriptions is purely imaginary, i.e.

\[ \text{Re} E^{(0)}_+(g) = \text{Re} E^{(0)}_-(g), \quad \text{Im} E^{(0)}_+(g) = -\text{Im} E^{(0)}_-(g). \]  

(4.95)

It turns out that the perturbative expansion around each multi-instanton configuration is also non-Borel summable for \( g > 0 \), and the resummed series \( E^{(k)}_\pm \) are defined by the same procedure. Notice that, with this definition, the two-instanton configuration computed in (2.201) picks an imaginary part since

\[ \log(-2g) \rightarrow \log(\frac{2}{|g|}) \pm \pi i. \]  

(4.96)

The total energy of the ground state is given by the sum over all instanton contributions,

\[ E_\pm(g) = E^{(0)}_\pm(g) + E^{(1)}_\pm(g) + E^{(2)}_\pm(g) + \cdots \]  

(4.97)

However, the physical energy must be independent of the resummation prescription and real (since it is the energy of a bound state!). This two requirements can be achieved simultaneously, at leading order in the instanton expansion, if the imaginary part of the perturbative sums \( E^{(0)}_\pm \) is equal but opposite in sign to the imaginary part of the two–instanton contribution \( \text{Im} E^{(2)}_\pm \):

\[ \text{Im} E^{(0)}_\pm = -\text{Im} E^{(2)}_\pm \sim -\left(\frac{2}{g}\right) e^{-1/3g} \frac{1}{2\pi} \text{Im} \log\left(-\frac{2}{g}\right) = -\frac{1}{g} e^{-1/3g}. \]  

(4.98)

The one-instanton contribution has an imaginary part, but it is proportional to \( e^{-1/2g} \) and cancels against the third-instanton contribution, so we don’t have to consider it at this
order. The cancellation (4.98) determines the large order behavior of the perturbative series by using (4.60) and taking into account that

$$\epsilon(z) \sim -\frac{1}{z}e^{-1/3z}. \quad (4.99)$$

One then finds,

$$a_k \sim -\frac{1}{\pi}3^{k+1}\Gamma(k+1) = -\frac{1}{\pi}3^{k+1}k! \quad (4.100)$$

which can be tested against the explicit results for the perturbative series, see [74], providing in this way a confirmation of the cancellation mechanism (4.98). The cancellation between perturbative and nonperturbative contributions appearing in the double-well has been argued to be relevant in more general situations, see for example [27, 43] for examples involving renormalons and [51] for examples in unitary matrix models and string theory.

### 4.8 Euler–Heisenberg Lagrangian

To illustrate the considerations above, we analyze a situation where the exact sum of the perturbative series is known in closed form in integral form: the Euler–Heisenberg effective action for a scalar particle in a constant electromagnetic field. The analysis of this problem from the point of view of Borel resummation was first done in [20], and is reviewed in the excellent survey [31].

The quantum effective action for the EM field after integrating out exactly the matter–field interactions is the Euler–Heisenberg Lagrangian

$$S(A) = \frac{1}{8\pi^2} \int_0^\infty \frac{ds}{s^3} e^{-im^2s} \left\{ (es)^2|G| \cot\left[ es(\sqrt{F^2 + G^2 + F})^{1/2} \right] \right\}$$

$$\times \coth\left[ es(\sqrt{F^2 + G^2 - F})^{1/2} \right] - 1 + \frac{2}{3}(es)^2F \right\}, \quad (4.101)$$

where

$$F = \frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}(\vec{B}^2 - \vec{E}^2), \quad G = \frac{1}{4}F_{\mu\nu}G^{\mu\nu} = -\vec{E} \cdot \vec{B}. \quad (4.102)$$

We will now consider the cases of a pure electric or a pure magnetic field, i.e. $G = 0$. Since

$$\coth\left[ es(\sqrt{F^2 + G^2 - F})^{1/2} \right] = \frac{\sqrt{2F}}{esG} + \cdots, \quad G \to 0, \quad (4.103)$$

we find,

$$S = \frac{1}{8\pi^2} \int_0^\infty \frac{ds}{s^3} e^{-im^2s} \left\{ es\sqrt{2F} \cot\left[ es\sqrt{2F} \right] - 1 + \frac{2}{3}(es)^2F \right\}, \quad (4.104)$$

68
The series expansion of \( x \cot x \) is

\[
x \cot x = 1 - \sum_{k=1}^{\infty} \frac{4^k |B_{2k}|}{(2k)!} x^{2k},
\]

therefore we are left with

\[
-\frac{1}{8\pi^2} \int_0^\infty ds \ e^{-ims} \left\{ \sum_{k=2}^{\infty} \frac{8^k |B_{2k}|}{(2k)!} (e^2 F)^k s^{2k-3} \right\},
\]

This can be expanded in powers of \( F \) to obtain a perturbative action,

\[
S = \frac{m^4}{8\pi^2} \sum_{k=2}^{\infty} \frac{|B_{2k}|}{(2k)!} \Gamma(2k-2) \left( \frac{8e^2 F}{m^4} \right)^k.
\]

For a purely electric field, this is precisely the result in eq. (8) of [20].

In order to understand the asymptotics of the coefficients in this expansion, we notice that

\[
|B_{2k}|(2k)! = \frac{2}{(2\pi)^{2k}} \zeta(2k).
\]

The zeta function behaves as

\[
\zeta(2k) \to 1, \quad k \to \infty.
\]

Therefore, the coefficients of this series behave asymptotically as

\[
(2\pi)^{-2k}(2k)!
\]

The sign depends crucially on the type of field we have. Since

\[
2F = \begin{cases} 
-\vec{E}^2, & \text{purely electric} \\
\vec{B}^2, & \text{purely magnetic},
\end{cases}
\]

we see that the sign of \( F \) depends on having an electric field or a magnetic field. If there is a magnetic field, \( -F \) is negative and the series is alternating. For an electric field \( -F \) is negative and all the terms are positive. In the first case, the series is Borel summable, and the process of Borel resummation of the series gives back the Euler–Heisenberg Lagrangian, as it can be easily checked [31]. In the second case, the series is not Borel summable. This just reflects the fact that, in a constant electric field, the vacuum is unstable and it is possible to produce pairs from the vacuum. This can be seen by noticing that the full, nonperturbative expression has an imaginary part that can be
computed as follows. Let us assume that in (4.104) we have an electric field, and let us denote

\[ E = \sqrt{E^2}. \]  

(4.112)

Using now that

\[ \sqrt{2F} = iE, \]  

(4.113)

and that

\[ \cot ix = -i \coth x \]  

(4.114)

we find the integral

\[ S = \frac{1}{8\pi^2} \int_0^\infty ds s^3 e^{-im^2s} \left\{ eEs \coth(eEs) - 1 - \frac{1}{3}(eEs)^2 \right\}, \]  

(4.115)

with an imaginary part

\[ \text{Im} S = -\frac{1}{8\pi^2} \int_0^\infty ds s^3 \sin(m^2s) \left\{ eEs \coth(eEs) - 1 - \frac{1}{3}(eEs)^2 \right\}. \]  

(4.116)

This can be rewritten as

\[ \text{Im} S = -\frac{i}{16\pi^2} \int_{-\infty}^\infty ds s^3 e^{-im^2s} \left\{ eEs \coth(eEs) - 1 - \frac{1}{3}(eEs)^2 \right\}. \]  

(4.117)

We now evaluate this by residues, by deforming the contour along the negative imaginary axis, and picking all the poles of the coth (see [45], p. 195) located at

\[ s_n = \frac{-in\pi}{eE}. \]  

(4.118)

We then find,

\[ \text{Im} S = -\frac{1}{8\pi} \sum_{n=1}^\infty \text{Res}_{s=s_n} \frac{eE \coth(eEs)}{s^2} \]  

(4.119)

which equals

\[ \text{Im} S = \frac{1}{8\pi} \left( \frac{eE}{\pi} \right)^2 \sum_{n=1}^\infty \frac{1}{n^2} \exp \left( -\frac{n\pi m^2}{eE} \right). \]  

(4.120)

The situation we find in this example is very similar to what we found in the case of the cubic anharmonic oscillator. When the coupling is purely imaginary (which corresponds here to a magnetic field), the perturbation series is Borel summable to the exact result. When the coupling is real (an electric field), the series is not Borel summable. However, we can regard the case of the electric field as an analytic continuation from the situation
with a general complex coupling. This is precisely what we have done in (4.115), and we verified that this analytic continuation leads to an imaginary part for the amplitude which corresponds to the decay problem.

Notice that the “instanton action” appearing in (4.120) is

\[ A = \frac{\pi m^2}{eE}. \]  

(4.121)

This should give a large order behavior of the form

\[ \left( \frac{\pi m^2}{eE} \right)^{-k} k! \]  

(4.122)

which is in perfect agreement with the asymptotic behaviour of (4.107), which is given by

\[ (2\pi)^{-2g} \left( -\frac{8e^2 F m^4}{g} \right)^g (2g)!, \]  

(4.123)

since only even powers of \( k \) contribute in the series (4.107).

### 4.9 Resumming perturbative series

We now discuss practical resummation methods for divergent series, focusing on Padé-Borel resummation. An excellent survey can be found in the recent review [18].

Given a Gevrey-1 series, its Borel transform has typically a finite radius of convergence. But in order to compute the integral (4.40), we also need an analytic continuation of the Borel transform along the integration contour, which is well beyond its radius of convergence. In realistic situations, one only knows a few coefficients of the original series, and a practical method is needed in order to find accurate approximations to the Borel transformed series.

A useful method, first proposed in [41], is to use Padé approximants. Given a series

\[ S(z) = \sum_{k=0}^{\infty} a_k z^k \]  

(4.124)

the Padé approximant \([l/m]\) is given by a rational function

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

(4.125)

where \( q_0 \) is fixed to 1, and one requires that

\[ f(z) - [l/m]S(z) = O(z^{l+m+1}). \]  

(4.126)
This fixes the coefficients involved in (4.125).

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

$$P_n^S(z) = \left[ \frac{[n/2]/((n + 1)/2)} {B_s} \right]_{B_s}$$

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 mimic 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 (4.41) where one integrates instead $P_n^S(z)$,

$$f_n(w) = w^{-1} \int_0^\infty dt \frac{e^{-t/w}}{P_n^S(t)}.$$  

We have also to be careful with the choice of contour. When there are poles along the real axis, there will be various choices of contour avoiding these poles. The obvious choices are the contours slightly above or slightly below the real axis considered in (17). When there are also zeros away from the real axis, more complicated contours have been proposed which are also useful [46].

**Example 4.11. The quartic integral.** The simplest example of this procedure is the power series expansion of the quartic integral, which is divergent and has zero radius of convergence. The Borel resummation leads to a power series with radius of convergence given by the location of the first singularity, namely

$$t = -\frac{1}{16} = -0.0625.$$  

If we compute the Padé approximants (4.127) 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

$$-0.0654626, \quad -0.0633103, \quad -0.0628726, \quad -0.0627157,$$

which approaches the position of the true singularity. For $n = 40$ and $g = 0.1$, the integral of the Padé approximant is

$$f_{40}(0.1) = 0.8576085853832414...$$

to be compared to the numerical evaluation of the integral

$$Z(0.1) = 0.8576085852902494...$$
5 Nonperturbative aspects of gauge theories

5.1 Conventions and basics

We follow the gauge theory conventions in [23]. The generators of the Lie algebra $T^a$ are taken to be anti–Hermitian, and satisfy the commutation relations

$$[T^a, T^b] = C^{abc} T^c.$$  \hfill (5.1)

For $SU(2)$, for example, we take

$$T^a = -\frac{i}{2} \sigma^a,$$  \hfill (5.2)

and the structure constants are

$$C^{abc} = \epsilon^{abc}.$$  \hfill (5.3)

The Cartan inner product is defined by

$$(T^a, T^b) = \delta^{ab}.$$  \hfill (5.4)

and it can be shown that

$$(T^a, T^b) = -2 \text{Tr}(T^a T^b).$$  \hfill (5.5)

The Euclidean action for pure Yang–Mills is

$$S_E = \frac{1}{4 g^2} \int d^4 x \left( F_{\mu\nu}^a F^{a\mu\nu} \right).$$  \hfill (5.6)

The Lagrangian of QCD will be written as

$$\mathcal{L} = \frac{1}{g^2} \left[ \frac{1}{4} (F_{\mu\nu}^a F^{a\mu\nu}) + \sum_{f=1}^{N_f} \bar{q}_f (i \not{D} - m_f) q_f \right]$$  \hfill (5.7)

where the covariant derivative is defined by

$$D_\mu = \partial_\mu + i A_\mu.$$  \hfill (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

$$A_\mu = g \hat{A}_\mu, \quad q = g \hat{q}.$$  \hfill (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

$$\alpha_s(\mu) = \frac{g^2(\mu)}{4\pi}$$  \hfill (5.10)
the renormalized coupling constant in the \( \overline{\text{MS}} \) scheme, at the subtraction point \( \mu \). The \( \beta \)-function is written as
\[
\beta(\alpha_s) = \mu^2 \frac{\partial \alpha_s}{\partial \mu^2} = \beta_0 \alpha_s^2 + \beta_1 \alpha_s^3 + \ldots
\]
(5.11)

The \( \beta \)-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
\[
\beta_0 = \beta_{0g} + \beta_{0f} = -\frac{1}{4\pi} \left( \frac{11N_c}{3} - \frac{2N_f}{3} \right),
\]
(5.12)

where \( N_c \) is the number of colors and \( N_f \) the number of massless quark flavours. \( \beta_{0g} \) and \( \beta_{0f} \) denote respectively the gluon and fermion contribution to the one-loop \( \beta \)-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
\[
\Lambda^2 = \mu^2 e^{1/(\beta_0 \alpha_s(\mu))}
\]
(5.13)

is in fact independent of \( \mu \), 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 \( \theta \) vacua

In this subsection we follow [71, 63].

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
\[
Q = \int d^4x q(x),
\]
(5.14)

where
\[
q(x) = \frac{1}{32\pi^2} (F, \tilde{F}) = \frac{1}{64\pi^2} \epsilon_{\mu\nu\rho\sigma} (F^{\mu\nu}, F^{\rho\sigma}).
\]
(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
\[
\mathcal{L}_\theta = \frac{1}{4g^2} (F^{\mu\nu}, F_{\mu\nu}) - i\theta q(x),
\]
(5.16)
where $\theta$ 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 $\theta$ parameter. One such quantity is the ground-state energy density, computed at large, finite volume $V$ as

$$\exp(-VE(\theta)) = \int [DA]e^{-\int d^4 x L_\theta}.$$  

(5.17)

The ground state energy $E(\theta)$ can be expanded around $\theta = 0$ as

$$E(\theta) - E(0) = \frac{1}{2} \chi^V_t \theta^2 s(\theta), \quad s(\theta) = 1 + \sum_{n=1}^\infty b_{2n} \theta^{2n}. \quad (5.18)$$

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 $\theta$ appear in the expansion of $E(\theta)$. The coefficient $\chi^V_t$ is an important quantity and measures the leading dependence of $E(\theta)$ on the $\theta$ angle around $\theta = 0$. It is called the topological susceptibility and it can be written as

$$\chi^V_t = \left( \frac{d^2 E}{d\theta^2} \right)_{\theta=0} = \frac{\langle Q^2 \rangle}{V} = \int_V d^4x (q(x)q(0)). \quad (5.19)$$

The last equality follows from

$$\langle Q^2 \rangle = \int_V d^4x \int_V d^4y \langle 0 | q(x)q(y) | 0 \rangle = \int_V d^4x \int_V d^4y \langle 0 | q(x - y)q(0) | 0 \rangle = V \chi^V_t, \quad (5.20)$$

where translation invariance of the vacuum has been used. The infinite-volume limit of the quantity $\chi^V_t$ will be simply denoted by

$$\chi_t = \lim_{V \to \infty} \chi^V_t \quad (5.21)$$

Although we have said that observables in YM theory should be sensitive to the $\theta$ parameter, this dependence is very subtle. The reason is that (5.15) is a total divergence,

$$q(x) = \partial_\mu K^\mu, \quad (5.22)$$

where

$$K^\mu = \frac{1}{16\pi^2} \epsilon_{\mu\nu\rho\sigma}(A^\nu_\rho \partial_\sigma A_\sigma + \frac{2}{3} A^\rho_\rho A_\sigma). \quad (5.23)$$

The three–form appearing here is the so-called Chern–Simons term. This means, in particular, that

$$\tilde{q}(p) = \int d^4xe^{-ipx}q(x) \quad (5.24)$$

75
vanishes at zero momentum, since it is of the form $p^\mu \tilde K_\mu(p)$. But the topological susceptibility is given by

$$\chi_t = \lim_{k \to 0} U(k), \quad (5.25)$$

where

$$U(k) = \int d^4x e^{ikx} \langle q(x)q(0) \rangle. \quad (5.26)$$

We can write

$$U(k) = \int \frac{d^4p'}{(2\pi)^4} \langle \tilde q(-k)\tilde q(p') \rangle \quad (5.27)$$

Since $\tilde q(0) = 0$, this quantity vanishes order by order in perturbation theory. However, as noticed by Witten in [68], 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

$$Q = \int d\Sigma \mu K_\mu. \quad (5.28)$$

Let us take as surface of integration two spatial planes at $t = \pm \infty$, so that

$$Q = \int d^3\vec{x} K^0(t \to \infty, \vec{x}) - \int d^3\vec{x} K^0(t \to -\infty, \vec{x}) \equiv K_+ - K_. \quad (5.29)$$

These operators are Hermitian, and related to each other by time reversal, so their spectra coincide. Let $|n_\pm\rangle$ denote their eigenstates,

$$K_\pm |n_\pm\rangle = n |n_\pm\rangle. \quad (5.30)$$

We can now expand the physical vacuum as

$$|\theta\rangle = \sum_n c_n(\theta)|n_+\rangle = \sum_n c_n(\theta)|n_-\rangle. \quad (5.31)$$

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,

\[
\frac{i}{\partial \theta} \langle \theta | O | \theta \rangle = i \frac{\partial}{\partial \theta} \langle 0 | O e^{-f d^4x L_{\theta}} | 0 \rangle = \int d^4x \langle 0 | q(x) O e^{-f d^4x L_{\theta}} | 0 \rangle = \int d^4x \langle \theta | q(x) O | \theta \rangle,
\]

so the operator \( i \partial_{\theta} \) is equivalent to the insertion of \( Q \). But because of (5.29) we find

\[
\frac{i}{\partial \theta} \langle \theta | O | \theta \rangle = \langle \theta | K^+ O | \theta \rangle - \langle \theta | O K^- | \theta \rangle. \tag{5.33}
\]

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 (5.31), we find,

\[
\frac{i}{\partial \theta} \sum_{n,k} c^*_n(\theta)c_k(\theta) = \sum_{n,k} (n - k)c^*_n(\theta)c_k(\theta), \tag{5.34}
\]

which leads to

\[
c_n = Ce^{i\theta}, \tag{5.35}
\]

where \( C \) is an overall constant. In terms of the eigenstates of \( K_{\pm} \), we find that

\[
| \theta \rangle = \sum_n e^{i\theta} | n \rangle, \tag{5.36}
\]

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 Chiral symmetry and chiral symmetry breaking

Excellent references for this section are [71], Chapter 7, and [30]. 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,

\[
\mathcal{L} = i \sum_{f=1}^{N_f} \bar{q}_f D q_f - \sum_{f=1}^{N_f} m_f \bar{q}_f q_f + \cdots \tag{5.37}
\]
We can write this in terms of left-handed and right-handed components

\[
q_{L,f} = \frac{1 - \gamma_5}{2} q_f, \quad q_{R,f} = \frac{1 + \gamma_5}{2} q_f
\]  

(5.38)
as follows

\[
\mathcal{L} = \sum_{f=1}^{N_f} \left( \bar{q}_{L,f} D q_{L,f} + \bar{q}_{R,f} D q_{R,f} \right) - \sum_{f=1}^{N_f} m_f \left( \bar{q}_{R,f} q_{L,f} + \bar{q}_{L,f} q_{R,f} \right) + \cdots
\]

(5.39)

Let us consider the symmetry group

\[
U_L(N_f) \times U_R(N_f)
\]

(5.40)
acting as

\[
q_{L,f} \rightarrow \sum_{f'=1}^{N_f} L_{ff'} q_{L,f'}, \quad q_{R,f} \rightarrow \sum_{f'=1}^{N_f} R_{ff'} q_{R,f'}.
\]

(5.41)
Clearly, only the term for the quark masses is not invariant under this symmetry. Therefore, in a world of massless quarks, the above group is a symmetry of QCD. When masses are present, the only remaining symmetry is the flavour symmetry \(U(1)^f\) acting as

\[
q_f \rightarrow e^{i \theta_f} q_f.
\]

(5.42)
We will now restrict ourselves to the symmetry

\[
SU_L(N_f) \times SU_R(N_f)
\]

(5.43)
The corresponding conserved currents are

\[
J_{\mu}^{aL} = \sum_{f,f'} \bar{q}_{L,f} T^a_{f,f'} \gamma_\mu q_{L,f'}, \quad J_{\mu}^{aR} = \sum_{f,f'} \bar{q}_{R,f} T^a_{f,f'} \gamma_\mu q_{R,f'}.
\]

(5.44)
where \(T^a, a = 1, \ldots, N_f^2 - 1\) are generators of \(SU(N_f)\). Equivalently, we can consider axial and vector currents

\[
V^\mu_a = \sum_{f,f'} V^\mu_{f,f'} T^a_{f,f'}, \quad A^\mu_a = \sum_{f,f'} A^\mu_{f,f'} T^a_{f,f'},
\]

(5.45)
with

\[
V^\mu_{f,f'} = \bar{q}_{f'} \gamma_\mu q_f, \quad A^\mu_{f,f'} = \bar{q}_{f'} \gamma_\mu \gamma_5 q_f.
\]

(5.46)
Notice that the vectorial current corresponds to the diagonal of (5.43). One fundamental, nonperturbative aspect of Nature is that the chiral symmetry (5.43) is \textit{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 \(5.43\) is broken down to the diagonal,

\[
SU_L(N_f) \times SU_R(N_f) \to SU_V(N_f).
\]

In other words, the charges

\[
Q^5_a(t) = \int d^3 \vec{x} \ A^0_a(t, \vec{x})
\]

do not leave the vacuum invariant. This is called chiral symmetry breaking \(\chi SB\), in short.

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 \(N_f^2 - 1\) 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 \(\pi\), the three \(K\), the \(\eta_0\) and the \(\eta_8\).

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

\[
A^\mu_{ud}(x) = \bar{u} \gamma^\mu \gamma_5 d(x).
\]

This current is not conserved in the real world where \(u, d\) are massive, and its divergence is given by

\[
\partial_\mu A^\mu_{ud}(x) = i(m_u + m_d) \bar{u} \gamma_5 d
\]

This current has the same quantum numbers as \(\pi^+\), so we can use it as a composite pion field operator. In other words, if

\[
|\pi(p)\rangle
\]

is the state of a pion with momentum \(p\), we must have

\[
\langle 0| A^\mu_{ud}(x)|\pi(p)\rangle = ip^\mu C_\pi e^{-ip \cdot x}.
\]

where \(C_\pi\) is a constant. This constant is typically parametrized as

\[
C_\pi = \frac{\sqrt{2} F_\pi}{\sqrt{2} E_p},
\]

where \(F_\pi\) is called the pion decay constant. It can be determined experimentally from the weak decay \(\pi^+ \to \mu^+ \nu\), and one finds

\[
F_\pi \sim 93 \text{ MeV}.
\]

79
If we introduce the normalized pion field
\[ \langle 0 | \phi_\pi(x) | \pi(p) \rangle = \frac{1}{(2\pi)^{3/2} \sqrt{2E}} e^{-ip \cdot x} \] (5.55)
we can write
\[ \partial_\mu A_{ud}^\mu(x) = \sqrt{2} F_\pi m_\pi^2 \phi_\pi(x). \] (5.56)

Using now (5.56) and (5.50) we can obtain a formula for \( m_\pi^2 \) in terms of \( F_\pi \) and \( m_u, m_d \). The basic idea for the formula is the following. If we sandwich (5.56) with the vacuum and a pion state, we find
\[ \langle \pi(q) | \partial_\mu A_{\mu}(0) | 0 \rangle = \frac{\sqrt{2} F_\pi m_\pi^2}{(2\pi)^{3/2} \sqrt{2E}} \] (5.57)
where we used (5.55). On the other hand, this equals
\[ \langle \pi(q) | i(m_u + m_d) \bar{u} \gamma_5 d | 0 \rangle. \] (5.58)

The strategy is to evaluate this correlator to relate \( m_\pi^2 \) to \( m_u, m_d \). To calculate (5.58), we need the soft pion theorem. We follow the short treatment in [30], section IV-5. Chapter 2 of [23] contains a more detailed treatment. We consider the matrix element for the process
\[ \alpha \to \beta + \pi(q) \] (5.59)
where \( \pi(q) \) is a pion state. The LSZ reduction formula states that
\[ \langle \pi(q) | \beta \rangle | \mathcal{O}(0) | \alpha \rangle = \frac{i}{(2\pi)^{3/2} \sqrt{2E}} \int d^4x e^{ip \cdot x} (m_\pi^2 - q^2) \langle \beta | T\phi_\pi(x) \mathcal{O}(0) | \alpha \rangle. \] (5.60)

We can use (5.56) again to write (5.60) as
\[ \frac{i}{(2\pi)^{3/2} \sqrt{2E}} \frac{m_\pi^2 - q^2}{\sqrt{2F_\pi^2 m_\pi^2}} \int d^4x e^{ip \cdot x} \langle \beta | T\partial_\mu A_{\mu}(x) \mathcal{O}(0) | \alpha \rangle. \] (5.61)

We remind that
\[ T \mathcal{O}_1(x) \mathcal{O}_2(0) = \theta(x^0) \mathcal{O}_1(x) \mathcal{O}_2(0) + \theta(-x^0) \mathcal{O}_2(0) \mathcal{O}_1(x) \] (5.62)
therefore
\[ \partial_{\mu} T A_{\mu}(x) \mathcal{O}(0) = T(\partial_{\mu} A_{\mu}(x) \mathcal{O}(0)) + \delta(x^0) [A_0(x), \mathcal{O}(0)] \] (5.63)
which holds as an operator equality. We then find, after integrating by parts,

\[
\langle \pi(q)\beta |\mathcal{O}(0)|\alpha \rangle = \frac{i}{(2\pi)^{3/2}} \frac{m_{\pi}^2 - q^2}{2E\sqrt{2F_{\pi}^2}} \cdot \int d^4x e^{iq\cdot x} \left\{ -\delta(x^0) \langle [A_0(x), \mathcal{O}(0)] |\alpha \rangle - iq^\mu \langle [T_{A_\mu}(x)\mathcal{O}(0)] |\alpha \rangle \right\}.
\]

(5.64)

If we now take the limit as \( q \to 0 \) of this equation, we find

\[
\lim_{q \to 0} \langle \pi(q)\beta |\mathcal{O}(0)|\alpha \rangle = -\frac{i}{(2\pi)^{3/2}} \frac{m_{\pi}^2}{2E\sqrt{2F_{\pi}^2}} \int d^4x e^{iq\cdot x} \left\{ -\delta(x^0) \langle [A_0(x), \mathcal{O}(0)] |\alpha \rangle - iq^\mu \langle [T_{A_\mu}(x)\mathcal{O}(0)] |\alpha \rangle \right\}. \tag{5.65}
\]

where

\[
R_\mu = -\frac{i}{(2\pi)^{3/2}} \frac{1}{2E\sqrt{2F_{\pi}^2}} \int d^4x e^{iq\cdot x} \langle [T_{A_\mu}(x)\mathcal{O}(0)] |\alpha \rangle. \tag{5.66}
\]

In our case,

\[
\mathcal{O} = i(m_u + m_d)\bar{u}\gamma_5 d. \tag{5.67}
\]

To compute the commutator in (5.65) 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, [71]),

\[
\delta(x^0 - y^0)[A^0_{f'f}(x), q_{f'}(y)] = -\delta(x - y)\delta_{f'f}\gamma_5 q_f(x). \tag{5.68}
\]

Using this, we find,

\[
-m \delta(x^0)[A^0(x), \partial_\mu A^\mu(0)] = (m_u + m_d) \left\{ \delta(x^0)[A^0(x), \bar{u}(0)]\gamma_5 d(0) + \bar{u}(0)\gamma_5 \delta(x^0)[A^0(x), d(0)] \right\} \tag{5.69}
\]

\[
= -(m_u + m_d) \delta(x) \left\{ \bar{u}(0)u(0) + \bar{d}(0)d(0) \right\}.
\]

We finally obtain,

\[
\lim_{q \to 0} \langle \pi(q)|i(m_u + m_d)\bar{u}\gamma_5 d|0 \rangle = -\frac{1}{(2\pi)^{3/2}} \frac{m_u + m_d}{\sqrt{2F_{\pi}^2}} \langle 0 |\bar{u}(0)u(0) + \bar{d}(0)d(0) |0 \rangle. \tag{5.70}
\]

By chiral symmetry, and to leading order in the quark masses, we can set

\[
\langle 0 |\bar{u}(0)u(0) |0 \rangle = \langle 0 |\bar{d}(0)d(0) |0 \rangle = \langle 0 |\bar{q}q |0 \rangle \tag{5.71}
\]

81
and taking into account (5.57) we finally obtain the formula relating the mass of the pion to the masses of the quarks and the $\chi$SB order parameter $\langle 0|\bar{q}q|0\rangle$:

$$m_\pi^2 = -\frac{m_u + m_d}{F_\pi^2}\langle 0|\bar{q}q|0\rangle \quad (5.72)$$

An alternative, elegant derivation of this relation using chiral Lagrangians can be found in Appendix C.

### 5.4 The anomalous $U(1)$ and the $U(1)$ problem

In the world of massless quarks there are two classical $U(1)$ symmetries which are part of (5.40). The first one is a vectorial $U_V(1)$

$$q_f \rightarrow e^{i\theta} q_f,$$
$$\bar{q}_f \rightarrow e^{-i\theta} \bar{q}_f, \quad (5.73)$$

where $f = 1, \ldots, N_f$. The associated current

$$Q^\mu = \sum_{f=1}^{N_f} \bar{q}_f \gamma^\mu q_f \quad (5.74)$$

and it 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 $U_A(1)$

$$q_f \rightarrow e^{i\theta \gamma_5} q_f,$$
$$\bar{q}_f \rightarrow e^{-i\theta \gamma_5} \bar{q}_f, \quad (5.75)$$

with current

$$J^\mu = \frac{1}{g^2} \sum_{f=1}^{N_f} \bar{q}_f \gamma^\mu \gamma_5 q_f, \quad (5.76)$$

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, there would be a ninth Goldstone boson. In Appendix C we show that a chiral Lagrangian for a ninth Goldstone boson $\eta'$ 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 $\sqrt{3}m_\pi^2$ [64]. The $U(1)$ problem (reviewed in for example [21]) 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 $\eta'$, with a mass of almost 1 GeV. This is far too heavy to be the ninth Goldstone boson.

The resolution of the $U(1)$ problem is based on the fact that the $U_A(1)$ symmetry is anomalous. The divergence of $J^\mu$ can be computed to be

$$\partial_\mu J^\mu = 2N_f q(x) + \frac{2i}{g^2} \sum_{f=1}^{N_f} m_f \bar{q}_f \gamma_5 q_f.$$  \hspace{1cm} (5.77)

A particular elegant derivation of this relation, originally due to Fujikawa [35], can be obtained in the path integral formulation by studying the change in the path integral measure, see for example [30]. The first term in (5.77) is the anomaly, and leads to a non-vanishing of the divergence even for massless quarks.

We can in principle apply the same type of current algebra arguments that we used before to analyze pions. In other words, if

$$|\eta'(p)\rangle$$  \hspace{1cm} (5.78)

is the state of the $\eta'$ with momentum $p$, we must have

$$\langle 0| J^\mu(x) |\eta'(p)\rangle = ip^\mu \frac{F_{\eta'}}{(2\pi)^{3/2}\sqrt{2E_p}} e^{-ip\cdot x}.$$  \hspace{1cm} (5.79)

It follows that

$$\langle \eta'(q)| \partial_\mu J^\mu(0)|0\rangle = \frac{F_{\eta'} m_{\eta'}^2}{(2\pi)^{3/2}\sqrt{2E}}$$  \hspace{1cm} (5.80)

is equal to

$$\langle \eta'(q)| \left\{2N_f q(x) + \frac{2i}{g^2} \sum_{f=1}^{N_f} m_f \bar{q}_f \gamma_5 q_f \right\}|0\rangle.$$  \hspace{1cm} (5.81)

If there was no anomalous term here, one could deduce that, as for the true Goldstone bosons,

$$m_{\eta'}^2 = \mathcal{O}(m_f).$$  \hspace{1cm} (5.82)

The anomalous term, although being a total derivative, has nontrivial effects, and in principle gives a mass to the $\eta'$. 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

$$q'_f = q_f + i\alpha \gamma_5 q_f, \quad \bar{q}'_f = \bar{q}_f + \bar{q} f i\gamma_5 \alpha.$$  \hspace{1cm} (5.83)
which is just a chiral rotation with arbitrary angle $\alpha$. The fermion measure in the path integral changes as [35]

$$Dq\,D\bar{q} = Dq'\,D\bar{q}' \exp\left\{2N_f\alpha\int d^4x\,q(x)\right\}. \quad (5.84)$$

Precisely because the quarks are massless there is no other change in the path integral. Therefore, after this change of variables we have that

$$Z(\theta) = Z(\theta + 2N_f\alpha), \quad (5.85)$$

where $\alpha$ is an arbitrary angle. In other words, the partition function (and any other observable) will be independent of $\theta$.

6 Instantons and renormalons in gauge theory

6.1 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

$$S_E \sim \int dr\, r^3 F^2 \quad (6.1)$$

If we want this to be finite, the integrand has to go like $1/r^2$, i.e. we require

$$F \sim \frac{1}{r^3} \quad (6.2)$$

as $r \to \infty$. In principle, this leads to the following behavior for $A(r)$,

$$A(r) \sim \frac{1}{r^2} \quad (6.3)$$

but since $A$ is only well defined up to a gauge transformation, we can have the more general behavior

$$A_\mu \to g\partial_\mu g^{-1} + \mathcal{O}(r^{-2}), \quad r \to \infty. \quad (6.4)$$
Since the limit has to be well-defined, we can define the function $g$ on the boundary at infinity $S^3 \subset \mathbb{R}^4$. 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 $S^3$ to the gauge group, i.e.

$$g : S^3 \rightarrow G.$$  \hspace{1cm} (6.5)

Under gauge transformation, $g$ will change. Therefore, what is a gauge-invariant concept is the homotopy type of mappings from $S^3$ to $G$. As in the theory of solitons, these homotopy types are classified by

$$\pi_3(G).$$  \hspace{1cm} (6.6)

A toy example are instantons in Euclidean two–dimensional space with $U(1)$ gauge group. Here, the homotopy group is $\pi_1(S^1) = \mathbb{Z}$, and the nontrivial map is just the $\nu$ covering

$$g^{(\nu)}(\theta) = e^{i\nu \theta}.$$  \hspace{1cm} (6.7)

Let us consider in detail the case in which $G = SU(2)$. Any element of $SU(2)$ can be written as

$$g = a + i \vec{b} \cdot \vec{\sigma}, \quad |a|^2 + |b|^2 = 1,$$

hence $SU(2)$ is homeomorphic to $S^3$. The homotopy group is

$$\pi_3(S^3) = \mathbb{Z}.$$  \hspace{1cm} (6.9)

**Remark 6.1.** This homotopy group can be computed by using Hurewicz isomorphism theorem (see for example [14]), which holds in this case due to the fact that

$$\pi_1(S^3) = 0.$$  \hspace{1cm} (6.10)

This theorem relates homotopy groups to homology groups, which are typically much easier to calculate, and in this case it says that

$$\pi_2(S^3) = H_2(S^3), \quad \pi_3(S^3) = H_3(S^3) = \mathbb{Z}.$$  \hspace{1cm} (6.11)

We than have learned that every field configuration of finite action has an integer $n$ associated to it, at least when the gauge group is $SU(2)$. This integer comes from the homotopy group (6.9). This topological argument tells us that there is an infinite set of classical vacua enumerated by an integer $n$. This integer $n$ is called the winding number.

In fact, instanton gauge potentials are pure gauge at infinity,

$$A_\mu \rightarrow g \partial_\mu g^{-1}, \quad r \rightarrow \infty,$$  \hspace{1cm} (6.12)
where \( g \) belongs to the homotopy class labeled by \( n \). 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 [55] for further details.

We start from the expression (5.28), integrated over the boundary at infinity, which is a three-sphere \( S^3 \). For a gauge field satisfying (6.12), the field strength \( F_{\mu\nu} \) vanishes at infinity. Therefore, on \( S^3 \),

\[
\epsilon_{\mu\nu\alpha\beta} \partial^\mu A^\alpha = -\epsilon_{\mu\nu\alpha\beta} A^\alpha A^\beta,
\]

and one finds,

\[
Q = -\frac{1}{48\pi^2} \int d\Sigma^\mu \epsilon_{\mu\nu\alpha\beta} (A^\nu, A^\alpha A^\beta).
\]

By using the boundary behavior of the gauge potentials, one can also write this quantity as

\[
Q = \frac{1}{48\pi^2} \int d\theta_1 d\theta_2 d\theta_3 \epsilon^{ijk} (g^{-1} \partial_i g, g^{-1} \partial_j g, g^{-1} \partial_k g).
\]

This quantity is a homotopy invariant and gives the winding number associated to the homotopy class of \( g \).

**Example 6.2.** An explicit expression for a map

\[
g : S^3 \rightarrow SU(2)
\]

with winding number \( n \) is given by

\[
g^{(n)}(x) = \left( \frac{x_4 + i \vec{x} \cdot \vec{\sigma}}{r} \right)^n.
\]

For \( n = 0 \), this is the trivial map, while for \( n = 1 \) it is the identity. Let us use the integral expression (6.14) to verify that \( g^{(1)} \) indeed has \( n = 1 \). The inverse map is

\[
g^{-1} = \frac{x_4 - i \vec{x} \cdot \vec{\sigma}}{r}.
\]

One finds,

\[
Q = -\frac{1}{24\pi^2} \int d\Sigma^\mu \left( -\frac{12 x_4^\mu}{|x|^4} \right).
\]

Using now

\[
d\Sigma^\mu = x^\mu |x|^2 d\Omega_3,
\]

we obtain

\[
Q = \frac{1}{2\pi^2} \int d\Omega_3 = 1.
\]
So far, we have seen that, if there are field configurations of finite action, they will be classified by an integer winding number. We still have to construct explicitly such configurations. In order to do this, we first establish a BPS-type of bound.

Start with
\[ \int d^4x (F \pm \tilde{F})^2 \geq 0 \] (6.22)

From here we find
\[ \frac{1}{4g^2} \int d^4x (F, F) \geq \mp \frac{1}{4g^2} \int d^4x (F, \tilde{F}) \] (6.23)

Since the above is a topological invariant, we find
\[ S \geq \mp \frac{8\pi^2 \nu}{g^2}. \] (6.24)

Therefore, in each topological sector characterized by a winding number \( \nu \), the YM action is minimized for field configurations satisfying

\[ F = \tilde{F} \rightarrow S = \frac{8\pi^2 \nu}{g^2}, \]
\[ F = -\tilde{F} \rightarrow S = -\frac{8\pi^2 \nu}{g^2}. \] (6.25)

These conditions (self-duality or anti-self-duality of the gauge field strength) can be used to search for instantons. If any of these conditions holds, the corresponding gauge field minimizes the action for a fixed topological class, and in particular solves the EOM.

It is possible to write down explicitly the asymptotic expression of the gauge field for the instanton configuration with gauge group \( SU(2) \) and \( \nu = 1 \). To do this, we simply set

\[ A_\mu = -(\partial_\mu g) g^{-1}, \] (6.26)

where \( g = g^{(1)} \), and \( g^{(m)} \) is the the map (6.17). Since

\[ \partial_4 g = \frac{x_\mu}{r^2} g + \frac{1}{r}, \]
\[ \partial_i g = \frac{x_\mu}{r^2} g + \frac{i\sigma_i}{r}, \quad i = 1, 2, 3. \] (6.27)

One then finds

\[ A_4 = \frac{\vec{x} \cdot \vec{\sigma}}{r^2}; \]
\[ A_i = -\frac{1}{r^2} \left( x_4 \sigma_i + \epsilon_{ijk} x_j \sigma_k \right), \] (6.28)
where we used that
\[ \sigma_i \vec{x} \cdot \vec{\sigma} = x_i + i \epsilon_{ijk} x_j \sigma_k. \] (6.29)

If we write
\[ A_\mu = -\frac{i}{2} \sigma_a A^a_\mu, \] (6.30)
and we introduce the 't Hooft matrices \( \eta_{\mu \nu}^a \) by
\[ \eta_{ij}^a = \epsilon_{aij}, \ \ \eta_{i4}^a = \delta_{ai}, \ \ \eta_{4i}^a = -\delta_{ai}, \] (6.31)
where \( i, j = 1, 2, 3 \), we find that
\[ A^a_\mu = 2 \eta_{\mu \nu}^a \frac{x^\nu}{r^2}. \] (6.32)

This asymptotic form suggests the following ansatz for the exact form
\[ A^a_\mu = 2 \eta_{\mu \nu}^a \frac{x^\nu}{r^2} f(r^2), \] (6.33)
where
\[ f(r^2) \to 1, \quad r \to \infty. \] (6.34)

Also, regularity at the origin requires that
\[ f(r^2) = cr^2 + \mathcal{O}(r^4), \quad r \to 0, \] (6.35)
where \( c \) is a constant. It can be easily seen that the self-duality condition for the gauge field leads to the following first order equation for \( f \),
\[ f(1 - f) - r^2 \frac{df}{dr^2} = 0, \] (6.36)
which can be integrated immediately:
\[ f(r^2) = \frac{r^2}{r^2 + \rho^2}. \] (6.37)

Here, \( \rho \) is an integration constant which can be regarded as the size of the instanton. The function (6.37) gives an exact instanton solution for \( \nu = 1 \) which is centered at the origin. A more general solution can be written down,
\[ A^a_\mu = 2 \eta_{\mu \nu}^a \frac{(x - x_0)^\nu}{(x - x_0)^2 + \rho^2} \] (6.38)
where \( x_0 \) is the position of the center of the instanton.
6.2 Instantons and theta vacua

The physical interpretation of instantons is as tunneling configurations between different vacua, i.e. the instanton field in a sector with winding number \( \nu \) 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 (5.28) is given by an integral over an \( S^3 \) 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

\[
\nu = n_+ - n_-,
\]

where

\[
n_\pm = -\frac{1}{48\pi^2} \int d^3x \epsilon_{ijk} (A_i, A_j A_k) \bigg|_{\tau = \pm \infty}.
\]

The field configurations at \( \tau \to \pm \infty \) correspond to different vacua whose homotopy numbers differ by \( \nu \), the charge of the instanton. Furthermore, one can choose the gauge in such a way that \( n_- = 0 \). Therefore, we find an explicit realization of all the vacua \( |n\rangle \) that we introduced in (5.31), and we find that they are labeled by an integer.

In particular, the transition amplitude between two vacua is given by

\[
\langle n(+\infty)|m(-\infty) \rangle = \int DA_{n-m} \exp \left[ -\int d^4x \mathcal{L}(A) \right],
\]

where the measure \( DA_{n-m} \) means that we integrate over all gauge fields with fixed winding number \( n - m \). From this expression we also obtain,

\[
\langle \theta'(+\infty)|\theta(-\infty) \rangle = \sum_{n,m} e^{-i\theta + im\theta'} \int DA_{n-m} \exp \left[ -\int d^4x \mathcal{L}(A) \right]
\]

\[
= \sum_{n,\nu} e^{-i(\theta - \theta') + i\nu \theta'} \int DA_{\nu} \exp \left[ -\int d^4x \mathcal{L}(A) \right]
\]

\[
= \delta(\theta - \theta') \sum_{\nu} e^{i\nu \theta} \int DA_{\nu} \exp \left[ -\int d^4x \mathcal{L}(A) \right].
\]

This can be written as

\[
\langle \theta'(+\infty)|\theta(-\infty) \rangle = \delta(\theta - \theta') \int DA \exp \left[ -\int d^4x \mathcal{L}_\theta(A) \right]
\]

where we have introduced the Lagrangian with a \( \theta \) term (5.16) and we integrate now over all possible gauge fields (belonging to all possible homotopy classes). This confirms that,
indeed, the theta vacuum is the vacuum which is obtained by quantizing the theory with the Lagrangian (5.16).

We can try to compute the amplitude (6.42) (or, equivalently, the energy density (5.17)) by using instanton calculus (we follow the elegant treatment by Coleman in [23]). The leading contributions comes from the one-instanton as well as the one-anti-instanton, since both have

$$S_c = \frac{8\pi^2}{g^2}$$

but opposite $\nu = \pm 1$. Therefore, they contribute together

$$E(\theta) \sim 2 \cos \theta e^{-\frac{8\pi^2}{g^2}}.$$

Of course, this just comes from an evaluation of the classical action, and at one-loop we must compute the determinant of fluctuations around the (anti)-instanton. 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 (6.45) (remember from (5.17) that the l.h.s. involves originally $VE(\theta)$). Another collective coordinate is the size of the instanton $\rho$, and finally there are three extra parameters coming from gauge rotations. In total, we have 8 parameters that lead to a factor

$$S^4_c = \left(\frac{8\pi^2}{g^2}\right)^4.$$ 

The integral over gauge transformations leads to a constant factor. The integral over $\rho$ must be of the form

$$\int_0^\infty \frac{d\rho}{\rho^5} f(\rho \mu)$$

just based on dimensional reasons: recall that we are computing an energy density, therefore it has units of length$^{-4}$, while $\rho$ has dimensions of length. $f(\rho \mu)$ is a function which depends on $\mu$, the renormalization mass needed in 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 $g^2(\mu)$

$$e^{-\frac{8\pi^2}{g^2(\mu)}} = e^{-\frac{2\pi}{\alpha_s(\mu)}}.$$ 

In view of (5.13), this must combine with

$$\mu^{-4\pi\beta_0}$$
in order to produce a RG-invariant integrand, and this fixes the form of \( f(\rho \mu) \) at leading order as
\[
 f(\rho \mu) = (\rho \mu)^{-4\pi/\beta_0}. 
\]  
(6.50)

We can now write
\[
e^{-\frac{2\pi}{\alpha_s(\mu)}} \mu^{-4\pi/\beta_0} = e^{-\frac{2\pi}{\alpha_s(1/\rho)}} \rho^{4\pi/\beta_0} = \Lambda^{-4\pi/\beta_0}
\]  
(6.51)
because of RG invariance, and the integral becomes
\[
e^{-\frac{2\pi}{\alpha_s(\rho)}} \int_0^\infty \frac{d\rho}{\rho^5} (\rho \mu)^{-4\pi/\beta_0} = \int_0^\infty \frac{d\rho}{\rho^5} e^{-\frac{2\pi}{\alpha_s(1/\rho)}}
\]  
(6.52)
which is the RG-invariant way of writing it. At small \( \rho \) we can use asymptotic freedom and the one-loop beta function to write the integral as
\[
\int_0^\infty \frac{d\rho}{\rho^5} (\rho \Lambda)^{11N_c/3}
\]  
(6.53)
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 (6.53) is not really the right answer. As the instanton size becomes large, the running coupling constant \( \alpha_s(1/\rho) \) in (6.52) 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 at small, finite volume (as in [47]) or if we have a Higgs-like field with a large VEV which sets the scale (as in supersymmetric gauge theories). Otherwise, instanton calculations in QCD are doubtful. Indeed, the \( \theta \) dependence in (6.45), which seems to be a universal feature of instanton-based approaches to the topological susceptibility, is currently disfavoured by lattice calculations [40].

### 6.3 Renormalons

As we saw in section 3, 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
Figure 18: 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.

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 [12, 2], 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

\[
(-i) \int d^4x e^{-ixq} \langle 0 | T(j_\mu(x)j_\nu(0)) | 0 \rangle = (q_\mu q_\nu - q^2 g_{\mu\nu}) \Pi(Q^2) \tag{6.54}
\]

with \( Q^2 = -q^2 \). We now compute the contribution of the fermion bubble diagrams shown in Fig. 18 to the Adler function

\[
D(Q^2) = 4\pi^2 \frac{d\Pi(Q^2)}{dQ^2}. \tag{6.55}
\]

The set of selected diagrams is gauge-invariant, but it is not the only set of diagrams that contributes to renormalon divergence. It is selected here for illustration.

The Adler function requires no additional subtractions beyond those contained in the renormalized QCD Lagrangian. Therefore no regularization is needed, provided the fermion loop insertions are renormalized. The renormalized fermion loop is given by

\[
-\beta_0 f \alpha_s \left[ \ln \left( \frac{k^2}{\mu^2} \right) + C \right] \tag{6.56}
\]

with a scheme-dependent constant \( C \). In the \( \overline{\text{MS}} \) scheme \( C = -5/3 \). \( \alpha_s \) is the running coupling constant (5.10). Let us consider the type of diagrams contributing to (6.56) shown in Fig. 18. To calculate their contribution, 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

\[
D = \sum_{n=0}^\infty \alpha_s \int_0^{\hat{k}_0^2} \frac{d\hat{k}^2}{\hat{k}^2} F(\hat{k}^2) \left[ \beta_0 f \alpha_s \ln \left( \frac{\hat{k}^2 Q^2 e^{-5/3}}{\mu^2} \right) \right]^n. \tag{6.57}
\]

92
where we have plugged in the leading logarithmic term of the fermion loop in each blob. Here we simply have a geometric series in the logs that can be summed up, and we obtain
\[
D = \int_0^\infty \frac{d\hat{k}^2}{\hat{k}^2} F(\hat{k}^2) \frac{\alpha_s}{1 - \beta_0 f \alpha_s \log\left(\frac{\hat{k}^2 Q^2 e^{5/3}}{\mu^2}\right)}.
\]  
(6.58)

It remains to perform the integral over \(\hat{k}^2\). The dominant contributions to the integral come from \(k \gg Q\) and \(k \ll Q\), because of the large logarithmic enhancements in these regions. Hence, it is sufficient to know the small-\(\hat{k}\) and large-\(\hat{k}\) behaviour of \(F\). In fact, one can expand \(F\) around \(\hat{k}^2 = 0\) in power series of \(\hat{k}^2\), and near \(\hat{k}^2 = \infty\) in inverse powers \((\hat{k}^2)^{-1}\). Precise equations can be found in [12].

We then split the integral in two regions
\[
(0, \frac{\mu^2}{(Q^2 e^{-5/3})}, \frac{\mu^2}{(Q^2 e^{-5/3})}, \infty).
\]  
(6.59)

For the first integral in the IR region we find terms with the generic form
\[
\int_{\mu^2/(Q^2 e^{-5/3})}^{\infty} \frac{d\hat{k}^2 (\hat{k}^2)^{h-1}}{1 - \beta_0 f \alpha_s \ln\left(\frac{\hat{k}^2 Q^2 e^{5/3}}{\mu^2}\right)} \frac{\alpha_s}{1 - \beta_0 f \alpha_s \ln\left(\frac{\hat{k}^2 Q^2 e^{5/3}}{\mu^2}\right)}.
\]  
(6.60)

Let us introduce the variable
\[
t = \log \frac{\mu^2}{(Q^2 e^{-5/3})}.
\]  
(6.61)

The integral reads,
\[
\left(\frac{\mu^2}{Q^2 e^{-5/3}}\right)^h \int_0^\infty dt e^{-th} \frac{\alpha_s}{1 + \beta_0 f \alpha_s t}.
\]  
(6.62)

A further normalization \(t \rightarrow t h\) leads to
\[
\frac{1}{h} \left(\frac{\mu^2}{Q^2 e^{-5/3}}\right)^h \int_0^\infty dt e^{-t} \frac{\alpha_s}{1 + \beta_0 f \alpha_s t}.
\]  
(6.63)

This has precisely the structure of the Borel transform (4.40), with poles at
\[
t = -\frac{h}{\beta_0 f}.
\]  
(6.64)

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 \(-2/\beta_0 f, -3/\beta_0 f, \cdots\). These are called IR renormalons. We can now repeat the procedure with the UV integral,
\[
\int_{\mu^2/(Q^2 e^{-5/3})}^{\infty} d\hat{k}^2 (\hat{k}^2)^{-1-r} \frac{\alpha_s}{1 - \beta_0 f \alpha_s \ln\left(\frac{\hat{k}^2 Q^2 e^{5/3}}{\mu^2}\right)}.
\]  
(6.65)
Here we introduce the variable
\[ t = \log \frac{\hat{k}^2}{\mu^2/(Q^2e^{-5/3})}, \] (6.66)
and we find
\[ \left( \frac{Q^2 e^{-5/3}}{\mu^2} \right)^r \int_0^\infty \frac{e^{-rt}}{1 + \beta_0 \alpha_s t} \frac{\alpha_s}{1 + \beta_0 \alpha_s t}. \] (6.67)
A further normalization \( t \to tr \) leads to
\[ \frac{1}{r} \left( \frac{Q^2 e^{-5/3}}{\mu^2} \right)^r \int_0^\infty dt e^{-rt} \frac{\alpha_s}{1 + \beta_0 \alpha_s t}. \] (6.68)
Now the poles are at
\[ t = \frac{r}{\beta_0 f}. \] (6.69)
These are called \( UV \) renormalons. An explicit computation shows that \( r \geq 1 \), hence there are poles at \( 1/\beta_0 f, 2/\beta_0 f, \cdots \).

One can argue that, after including other effects, the coefficient \( \beta_0 f \) becomes \( \beta_0 \), the full coefficient of the beta function at one loop [12]. 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
\[ \left( \frac{\mu^2}{Q^2} \right)^h e^{h/(\beta_0 \alpha_s(\mu))} = \left( \frac{\Lambda^2}{Q^2} \right)^h, \] (6.70)
where \( \Lambda \) is the dynamically generated scale of QCD. These are power corrections due to nonperturbative effects.

The fact that IR renormalons are located at (6.64) but with \( \beta_0 \) instead of \( \beta_0 f \) can be argued by using RG arguments, as we will now show following [43] (the original argument is due to Parisi [53]). A generic correlation function in QCD can be written as
\[ F(\alpha_s) = F_p(\alpha_s) + F_{np}(\alpha_s) \] (6.71)
where
\[ F_p(\alpha_s) = \sum_{n=0}^\infty f_n \alpha_s^{n+1} \] (6.72)
is the perturbative contribution, and \( F_{np}(\alpha_s) \) is a nonperturbative contribution. If \( F(\alpha_s) \) is given by the vev of the product of two currents, as we considered above, we can calculate it through an operator product expansion (OPE), and \( F_{np}(\alpha_s) \) can be calculated as a
A condensate is given by the vev of an operator of dimension $d$, and it satisfies the RG equation

$$
\left( \mu^2 \frac{\partial}{\partial \mu^2} + \beta(\alpha_s) \frac{\partial}{\partial \alpha_s} + \gamma_n(\alpha_s) \right) F_{np}(\alpha_s) = 0 \tag{6.73}
$$

where

$$
\gamma_n(\alpha_s) = \gamma_1 \alpha_s + \cdots \tag{6.74}
$$
is the anomalous dimension of the operator. Using this equation, one can determine $F_{np}(\alpha_s)$ up to an overall constant,

$$
F_{np}(\alpha_s) = C \mu^d \alpha_s^{\delta} \exp \left[ \frac{d}{2 \beta_0 \alpha_s} \right] \left( 1 + \mathcal{O}(\alpha_s) \right). \tag{6.75}
$$

This can be easily checked

$$
\mu^2 \frac{\partial}{\partial \mu^2} F_{np}(\alpha_s) = \frac{d}{2} F_{np}(\alpha_s), \tag{6.76}
$$

$$
\beta(\alpha_s) \frac{\partial}{\partial \alpha_s} F_{np}(\alpha_s) = \left( \beta_0 \alpha_s^2 + \beta_1 \alpha_s^3 + \cdots \right) \left[ \frac{\delta}{\alpha_s} - \frac{d}{\beta_2 \alpha_s^2} + \cdots \right] F_{np}(\alpha_s) \tag{6.77}
$$

$$
= \left( -\frac{d}{2} + \delta \beta_0 \alpha_s - \frac{d \beta_1}{2 \beta_0 \alpha_s} + \cdots \right) F_{np}(\alpha_s),
$$

so consistency requires

$$
\delta = \frac{d \beta_1}{2 \beta_0^2} - \frac{\gamma_1}{\beta_0}. \tag{6.78}
$$

Now, the perturbative part $F_p(\alpha_s)$ is typically not Borel summable. If we define the Borel transform

$$
B_F(z) = \sum_{n=0}^{\infty} \frac{f_n}{n!} z^n, \tag{6.79}
$$

we obtain a representation

$$
F_p(\alpha_s) = \int_0^\infty dz \exp^{-z/\alpha_s} B_F(z) \tag{6.80}
$$

(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 $\alpha_s$). Assuming now a divergence of the type (4.45), one obtains an ambiguity in the Borel representation leading to an imaginary part of the form

$$
\text{Im } F_p(\alpha_s) \sim \alpha_s^{1-b} e^{-A/\alpha_s}. \tag{6.81}
$$
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 (6.75) should have an imaginary part which cancels against the imaginary part of the Borel resummation (6.81). 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 (6.81), equals the corresponding exponent in (6.75). We then obtain

$$A = -\frac{d}{2\beta_0}.$$  

(6.82)

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

$$1 - b = \delta.$$  

(6.83)

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 (6.82) should be identified with those found in (6.64) after setting $h = d/2$.

To check the consistency of this picture, we point out that (6.70) 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 (6.63). Conversely, one can use renormalons to obtain hints about nonperturbative effects in the computation of correlation functions, see [12].

What about the role of instantons in QCD and their effect on the large order behavior of perturbation theory? It has been argued [17] 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 ($\nu = 2$ in (6.25)):

$$S = \frac{16\pi^2}{g^2} = \frac{4\pi}{\alpha_s}.$$  

(6.84)

This would lead to a singularity in the Borel plane at

$$z_{\text{inst}} = 4\pi.$$  

(6.85)
IR renormalons at $t = -h/\beta_0$, $h \geq 2$

UV renormalons at $t = h/\beta_0$, $h \geq 1$

Figure 19: The conjectural structure of the Borel plane for the current-current correlation function in QCD.

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

$$ |z_{\text{ren}}| \leq \frac{12\pi}{11 N_c} < 4\pi $$

and corresponds to the UV renormalon. We depict in Fig. 19 the conjectural structure of the Borel plane for the current-current correlation function in QCD.

7 The $\mathbb{P}^{N-1}$ sigma model

Our first example of the $1/N$ expansion, and of the type of nonperturbative methods associated to resumming an infinite number of diagrams, will be the a two-dimensional toy model: the $\mathbb{P}^{N-1}$ sigma model. This model can be exactly solved at large $N$ and illustrates many important aspects of the $1/N$ expansion and of nonperturbative aspects of quantum field theories. In particular, we will be able to obtain a purely nonperturbative result: a nonzero value for a two-dimensional analogue of the topological susceptibility.

We follow the original references [25, 67].
7.1 Defining the model

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:

\[
z_1(x), \cdots, z_N(x), \quad \sum_{i=1}^{N} |z_i|^2 = 1.
\]  

(7.1)

There is also a \( U(1) \) gauge symmetry

\[
z_i \rightarrow e^{\text{i} \alpha(x)} z_i
\]  

(7.2)

We can cook up a gauge field out of the \( z_i \), since the composite field

\[
A_\mu = \frac{\text{i}}{2} \left( \bar{z}_i \partial_\mu z_i - (\partial_\mu z_i) \bar{z}_i \right),
\]  

(7.3)

which is real, transforms as

\[
A_\mu \rightarrow A_\mu - \partial_\mu \alpha(x).
\]  

(7.4)

Let us check this:

\[
\partial_\mu z_i \rightarrow e^{\text{i} \alpha(x)} \left( \text{i} \partial_\mu \alpha + \partial_\mu z_i \right)
\]  

(7.5)

Therefore

\[
\bar{z}_i \partial_\mu z_i \rightarrow \bar{z}_i \partial_\mu z_i + \text{i} \partial_\mu \alpha \bar{z}_i z_i = \bar{z}_i \partial_\mu z_i + \text{i} \partial_\mu \alpha.
\]  

(7.6)

The dynamics of this field is described by the gauge invariant action

\[
S = \frac{1}{g^2} \int d^2 x \bar{D_\mu z} D^\mu z, \quad D_\mu = \partial_\mu + \text{i} A_\mu.
\]  

(7.7)

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 (7.7)

\[
\mathcal{L} = \bar{D_\mu z} D^\mu z
\]  

(7.8)

we find

\[
\mathcal{L} = \partial^\mu \bar{z}_i \partial_\mu z_i - \text{i} A_\mu \bar{z}_i \partial^\mu z_i + \text{i} A_\mu \partial^\mu \bar{z}_i z_i + A_\mu A^\mu \bar{z}_i z_i
\]  

(7.9)

which is

\[
\mathcal{L} = \partial^\mu \bar{z}_i \partial_\mu z_i + A_\mu^2 - A_\mu \text{i} \left( \bar{z}_i \partial^\mu z_i - (\partial^\mu \bar{z}_i) z_i \right).
\]  

(7.10)

The classical EOM for \( A_\mu \) gives precisely the definition (7.3). Notice that

\[
z_i \bar{z}_i = 1 \Rightarrow (\partial^\mu \bar{z}_i) z_i + \bar{z}_i \partial^\mu z_i = 0
\]  

(7.11)
therefore we can write
\[ A^\mu = i\bar{z}_i \partial^\mu z_i = -i(\partial^\mu \bar{z}_i)z_i, \]  
(7.12)
and
\[ \mathcal{L} = \partial^\mu \bar{z}_i \partial_\mu z_i - A^2_\mu = \partial^\mu \bar{z}_i \partial_\mu z_i - (\bar{z}_i \partial^\mu z_i)(\bar{z}_j \partial_\mu \bar{z}_j) \]  
(7.13)
or equivalently
\[ \mathcal{L} = \partial^\mu \bar{z}_i \partial_\mu z_i + (\bar{z}_i \partial^\mu z_i)(\bar{z}_j \partial_\mu \bar{z}_j). \]  
(7.14)

### 7.2 Instantons

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 [25] as well as in section 4.5 of [55]. These instantons are topologically nontrivial configurations with finite action. Notice that finite action means here that
\[ D_\mu z_i = 0, \quad \text{at } |x| \to \infty, \quad i = 1, \cdots, n, \]  
(7.15)
therefore, at infinity, \( z_i \) is covariantly constant, i.e. it must be a constant vector up to a phase. We write
\[ z_i = n_i e^{i\sigma(x)}, \quad |x| \to \infty, \quad n_i \bar{n}_i = 1. \]  
(7.16)
This can be seen in detail by spelling out the condition (7.15). It means that
\[ -iA_\mu = \frac{\partial_\mu z_i}{z_i} = \frac{\partial|z_i|}{|z_i|} + i\partial_\mu \phi_i, \]  
(7.17)
where \( \phi_i \) is the phase of \( z_i \). Since \( iA_\mu \) is pure imaginary and indepedent of the index \( i \), we deduce that, at infinity,
\[ \partial_\mu |z_i| = 0, \quad \phi_i = \sigma(\theta), \quad i = 1, \cdots, N, \]  
(7.18)
which is precisely (7.16).

The topological charge classifying instantons is given by
\[ Q = \frac{1}{2\pi} \int d^2x \epsilon_{\mu\nu} \partial_\mu A_\nu. \]  
(7.19)
Since
\[ \epsilon_{\mu\nu} \partial_\mu A_\nu = i\epsilon_{\mu\nu} \partial_\mu \bar{z}_i \partial_\nu z_i \]  
(7.20)
this can be rewritten as
\[ Q = \frac{1}{2\pi i} \int d^2x \epsilon_{\mu\nu} \partial_\nu (\bar{z}_i \partial_\mu z_i). \]  
(7.21)
In order to be able to talk about instantons, we have to show that the topological charge (7.19) is quantized. We follow the discussion in [67]. Using Stokes theorem, we can write (7.21) as an integral at the boundary, i.e. at infinity

$$Q = \frac{1}{2\pi i} \oint d^2x \bar{z}_i \partial_\mu z_i. \quad (7.22)$$

Plugging in here the boundary behavior (7.16), we obtain

$$Q = \frac{1}{2\pi} \oint d^2x \partial_\mu \sigma = \frac{1}{2\pi} \Delta \sigma, \quad (7.23)$$

where $\Delta \sigma$ is just the change of $\sigma$ as we go around a circle at infinity. Since a phase is defined up to an integer multiple of $2\pi$, 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

$$q(x) = \frac{1}{2\pi} \epsilon_{\mu\nu} \partial_\mu A_\nu = \frac{1}{2\pi} \epsilon_{\mu\nu} \overline{D_\mu z} \cdot D_\nu z. \quad (7.24)$$

To see this, notice that the last term equals

$$\frac{i}{2\pi} \epsilon_{\mu\nu}(\partial_\mu \bar{z}_i - iA_\mu \bar{z}_i)(\partial_\nu z_i + iA_\nu z_i) \quad (7.25)$$

and due to antisymmetry of $\epsilon_{\mu\nu}$ we only have to check that

$$-i\epsilon_{\mu\nu}(A_\mu \bar{z}_i \partial_\nu z_i - A_\nu \bar{z}_i \partial_\mu z_i) \quad (7.26)$$

vanishes. Using (7.11) we can write it as

$$-i\epsilon_{\mu\nu}(A_\mu \bar{z}_i \partial_\nu z_i + A_\nu \bar{z}_i \partial_\mu z_i) = 0, \quad (7.27)$$

therefore the topological charge can be written as

$$Q = \frac{1}{2\pi} \int d^2x \epsilon_{\mu\nu} \overline{D_\mu z} \cdot D_\nu z. \quad (7.28)$$

From the obvious inequality

$$|D_\mu z \mp i\epsilon_{\mu\nu} \overline{D_\nu z}|^2 \geq 0 \quad (7.29)$$

we find

$$\overline{D_\mu z} \cdot D_\mu z + \epsilon_{\mu\rho} \epsilon_{\mu\sigma} \overline{D_\rho z} \cdot D_\sigma z \mp 2i\epsilon_{\mu\nu} \overline{D_\mu z} \cdot D_\nu z \geq 0, \quad (7.30)$$

and since $\epsilon_{\mu\nu} \epsilon_{\mu\sigma} = \delta_{\nu\sigma}$ we get at the end of the day

$$\overline{D_\mu z} \cdot D_\mu z \geq i\epsilon_{\mu\nu} \overline{D_\mu z} \cdot D_\nu z, \quad (7.31)$$
after integration one finds,
\[
\frac{1}{g^2} \int d^2 x \overline{D}_\mu z \cdot D_\mu z \geq \frac{i}{g^2} \int d^2 x \epsilon_{\mu\nu} \overline{D}_\mu z \cdot D_\nu z, \tag{7.32}
\]
i.e.
\[
S \geq \frac{2\pi}{g^2} |Q|. \tag{7.33}
\]
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:
\[
D_\mu z \mp i \epsilon_{\mu\nu} \overline{D}_\nu z = 0. \tag{7.34}
\]
The ± signs give instanton and anti-instanton solutions respectively. These are the analogues of the (anti) self-duality conditions for instantons in QCD.

### 7.3 The effective action at large \( N \)

The (Euclidean) action of the \( \mathbb{P}^{N-1} \) model is given by
\[
S = \int d^2 x \left[ \frac{1}{g^2} \overline{D}_\mu z \cdot D_\mu z - i \lambda (z_i \overline{z}_i - 1) + \frac{i \theta}{2\pi} \epsilon_{\mu\nu} \partial_\mu A_\nu \right], \tag{7.35}
\]
and includes the analogue of a theta term. We will introduce the ’t Hooft parameter
\[
t = g^2 N. \tag{7.36}
\]
Here we treat \( A_\mu \) and \( \lambda \) as auxiliary gauge fields. When we integrate them out we obtain the action for the fields \( z_i \) together with the constraint (7.1). But since the action is quadratic in \( z_i \), of the form
\[
\int d^2 x \overline{z}_i \Delta z_i, \tag{7.37}
\]
where
\[
\Delta = -\frac{N}{t} D_\mu D^\mu - i \lambda, \tag{7.38}
\]
we can integrate out the \( N \) bosonic, complex variables \( z_i \). Each of them gives a factor
\[
\frac{1}{\det \Delta} \tag{7.39}
\]
and since we have \( N \) of them, we obtain, after writing the determinant as the exponential of a trace of a log,
\[
\exp \left[ -N \text{Tr} \log \left( -\left( \delta_\mu + iA_\mu \right)^2 - i \frac{t \lambda}{N} \right) \right]. \tag{7.40}
\]
This leads to the effective action

\[ S_{\text{eff}} = N \text{Tr} \log \left( -\partial_{\mu} + iA_{\mu} \right)^2 - i\frac{t\lambda}{N} + i\lambda - \frac{i\theta}{2\pi} \epsilon_{\mu\nu} \partial_{\mu} A_{\nu} \]  

(7.41)

which depends on the fields \( A_{\mu} \) and \( \lambda \). We will often Fourier-transform the fields as

\[ \tilde{\lambda}(p) = \int d^2x e^{-ipx} \lambda(x). \]  

(7.42)

Notice that in this effective action \( N \) plays the role of \( \frac{1}{\hbar} \). For large \( N \) it makes sense to evaluate the path integral by looking at stationary points of the form

\[ A_{\mu} = 0, \quad \lambda = \text{constant}. \]  

(7.43)

This is of course what one expects from Lorentz invariance. The EOM for \( \lambda \) is obtained from

\[ \frac{\delta}{\delta \lambda} \left[ i \int d^2x \lambda + N \text{Tr} \log \left( -\partial_{\mu} + iA_{\mu} \right)^2 - i\frac{t\lambda}{N} \right] = 0 \]  

(7.44)

or

\[ i - it \text{Tr} \frac{1}{\partial_{\mu}^2 - i \frac{t\lambda}{N}} = 0 \]  

(7.45)

since we are taking \( A_{\mu} = 0 \). Evaluating the trace in momentum space we find

\[ 1 - t \int \frac{d^2k}{(2\pi)^2} \frac{1}{k^2 - i\frac{t\lambda}{N}} = 0. \]  

(7.46)

This is a divergent integral and we introduce a cutoff \( \Lambda \) for \( |k| \). Going to polar variables we have to evaluate

\[ \int \frac{d^2k}{(2\pi)^2} \frac{1}{k^2 - i\frac{t\lambda}{N}} = \int \frac{dk}{2\pi} \frac{k}{k^2 - \frac{t\lambda}{N}} = \frac{1}{4\pi} \log \left( k^2 - \frac{t\lambda}{N} \right) \bigg|_0^\Lambda \]

(7.47)

\[ = \frac{1}{4\pi} \log \left( \frac{\Lambda^2}{t\lambda} + 1 \right) \approx \frac{1}{4\pi} \log \frac{\Lambda^2}{m^2}, \]

where we have assumed that the solution to the equation (7.46) is of the form

\[ \lambda = i\frac{Nm^2}{t}, \quad m^2 > 0, \]  

(7.48)

and in the last line we have neglected the subleading term in \( N \). We then find,

\[ \frac{\Lambda^2}{m^2} = e^{4\pi/t} \Rightarrow m^2 = \Lambda^2 e^{-4\pi/t}, \]  

(7.49)
confirming our ansatz (7.48). If we then look back at (7.35), we see that the saddle-point for \( \lambda \) generates a mass for the \( z_i \) field. This is the first dynamical effect that can be obtained at large \( N \). Notice that the dependence on \( t \) is non-perturbative.

For simplicity, we will denote the fluctuation of the \( \lambda \) field around its vev (7.48) by \( \lambda \) as well. It turns out that the natural normalizations for \( A_\mu \) and the fluctuation \( \lambda \) are

\[
A_\mu \to \frac{1}{\sqrt{N}} A_\mu, \quad \lambda \to \sqrt{N} \lambda / t
\]  
(7.50)

as we will see. We can now perform a systematic expansion of the effective action around the saddle point that we have just found in inverse powers of \( N \). We expand

\[
\text{Tr} \log \left( - (\partial_\mu + i A_\mu / \sqrt{N})^2 + m^2 - \frac{\lambda}{\sqrt{N}} \right)
\]  
(7.51)

around \( A_\mu = 0, \lambda = 0 \), and we find

\[
N \log(-\partial^2 + m^2) + N \log \left[ 1 + \Delta \left( \frac{A^2}{N} - \frac{\lambda}{\sqrt{N}} - i \{ A, \partial \} / \sqrt{N} \right) \right]
\]  
(7.52)

where

\[
\Delta = (-\partial^2 + m^2)^{-1}.
\]  
(7.53)

We expand in inverse powers of \( N \) and at leading order we find, schematically,

\[
\Delta A^2 + \frac{1}{2} \Delta^2 (\partial A + 2A \partial)^2 + \frac{1}{2} \Delta^2 \lambda^2.
\]  
(7.54)

Let us write the last term in a more detailed way. After taking the trace we find

\[
\int d^2 x \Delta(x, y) \lambda(y) \int d^2 y \Delta(y, x) \lambda(x),
\]  
(7.55)

where

\[
\Delta(x, y) = \int \frac{d^2 p}{(2\pi)^2} \frac{e^{i p(x-y)}}{p^2 + m^2}.
\]  
(7.56)

In terms of Fourier-transformed fields, we can write this as

\[
\frac{1}{2} \int \frac{d^2 p}{(2\pi)^2} \tilde{\lambda}(p) \tilde{\Gamma}^\lambda(p) \tilde{\lambda}(-p),
\]  
(7.57)

where

\[
\tilde{\Gamma}^\lambda(p) = \int \frac{d^2 q}{(2\pi)^2} \frac{1}{(q^2 + m^2)((p + q)^2 + m^2)}.
\]  
(7.58)
Similarly, for the quadratic term in the $A_\mu$ fields, we find,

$$\frac{1}{2} \int \frac{d^2p}{(2\pi)^2} \tilde{A}^\mu(p) \tilde{\Gamma}^A_{\mu\nu}(p) \tilde{A}^\nu(-p),$$  \hspace{1cm} (7.59)

where

$$\tilde{\Gamma}^A_{\mu\nu}(p) = 2\delta_{\mu\nu} \int \frac{d^2q}{(2\pi)^2} \frac{1}{(q^2 + m^2)} - \int \frac{d^2q}{(2\pi)^2} \frac{(p_\mu + 2q_\mu)(p_\nu + 2q_\nu)}{(q^2 + m^2)((p + q)^2 + m^2)}. \hspace{1cm} (7.60)$$

The computation of (7.58) and (7.60) is a standard but somewhat long exercise in Feynman calculus. The details can be found in Appendix D. The results are the following:

$$\tilde{\Gamma}^\lambda(p) = f(p) = \frac{1}{2\pi \sqrt{p^2 + 4m^2}} \log \frac{\sqrt{p^2 + 4m^2} - \sqrt{b^2}}{\sqrt{p^2 + 4m^2} + \sqrt{b^2}},$$

$$\tilde{\Gamma}^A_{\mu\nu}(p) = \left( \delta_{\mu\nu} - \frac{p_\mu p_\nu}{p^2} \right) \left\{ (p^2 + 4m^2)f(p) - \frac{1}{\pi} \right\}. \hspace{1cm} (7.61)$$

Notice that the quadratic term in $\tilde{A}$ is of the form

$$(\delta_{\mu\nu} - \frac{p_\mu p_\nu}{p^2})(c + \mathcal{O}(p^2)) \hspace{1cm} (7.62)$$

where

$$c = \frac{1}{12\pi m^2}. \hspace{1cm} (7.63)$$

This structure is a consequence of gauge invariance, and leads to the standard gauge field kinetic energy

$$(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 \hspace{1cm} (7.64)$$

written in momentum space. In other words, the quantum corrections have generated a kinetic energy for $A_\mu$. 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 $A_\mu$. 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.
7.4 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 $\chi_t$ is given by the limit (5.25). We will then compute

$$U(p) = \int d^2x \, e^{ipx} \langle q(x)q(0) \rangle = \int \frac{d^2p'}{(2\pi)^2} \langle \tilde{q}(-p)\tilde{q}(p') \rangle$$

(7.65)

where $q(x)$ is the topological density defined in (7.24). This quantity has now a factor $1/N$ which comes from the normalization of $A_{\mu}$. The Fourier transform of $q(p)$ is given by

$$\tilde{q}(p) = -\frac{i}{2\pi} \epsilon_{\mu\nu} p_{\mu} \tilde{A}_{\nu}.$$  

(7.66)

Therefore,

$$\langle \tilde{q}(-p)\tilde{q}(p') \rangle = \frac{1}{4\pi^2} \epsilon_{\mu\nu} \epsilon_{\rho\sigma} p_{\mu} p'_{\rho} (\tilde{A}_{\nu}(-p)\tilde{A}_{\sigma}(p')).$$  

(7.67)

To calculate the two-point function of the gauge field we first choose the Lorentz gauge

$$\partial_{\mu} A_{\mu} = 0.$$  

(7.68)

In this gauge the two-point function can be immediately deduced from (7.61), and one finds

$$\langle \tilde{A}_{\nu}(p)\tilde{A}_{\sigma}(-p') \rangle = (2\pi)^2 \delta(p - p') \left( \delta_{\mu\nu} - \frac{p_{\mu} p_{\nu}}{p^2} \right) D_A(p),$$  

(7.69)

where

$$D_A(p) = \left\{ (p^2 + 4m^2) f(p) - \frac{1}{\pi} \right\}^{-1}.$$  

(7.70)

The $(2\pi)^2$ factor in (7.69) comes from the kinetic term (7.59) in momentum space. Since

$$\epsilon_{\mu\nu} \epsilon_{\rho\sigma} p_{\mu} p_{\rho} \left( \delta_{\sigma\sigma'} - \frac{p_{\mu} p_{\sigma}}{p^2} \right) = (\delta_{\mu\nu} \delta_{\sigma\sigma'} - \delta_{\mu\sigma} \delta_{\nu\rho} p_{\mu} p_{\rho} \left( \delta_{\nu\sigma} - \frac{p_{\nu} p_{\sigma}}{p^2} \right) = p^2$$

(7.71)

we find

$$\langle \tilde{q}(-p)\tilde{q}(p') \rangle = \frac{p^2}{4\pi^2} (2\pi)^2 D_A(p) \delta(p - p').$$  

(7.72)

Therefore,

$$\int \frac{d^2p}{(2\pi)^2} \langle \tilde{q}(p)\tilde{q}(p') \rangle = p^2 D_A(p) = \frac{3m^2}{\pi N} + \mathcal{O}(p^2),$$  

(7.73)

and the topological susceptibility reads, at leading order in the $1/N$ expansion,

$$\chi_t = \frac{3m^2}{\pi N}.$$  

(7.74)
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 (5.27). 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 [68].

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.

- 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).
- The field $A_\mu$, which started its life as an auxiliary variable, becomes a dynamic gauge field which leads to quark confinement.
- 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 [59]. Classic reviews of this topic are [23, 69]. A more modern reference is [49].

8.1 Fatgraphs

We will write down the QCD Lagrangian (5.7) as

$$\mathcal{L} = \frac{N}{t} \left[ \frac{1}{4} (F_{\mu\nu}, F^{\mu\nu}) + \sum_{f=1}^{N_f} \bar{q}_f (i \gamma \vec{D} - m_f) q_f \right]$$

(8.1)

where we have introduced the ’t Hooft parameter as

$$t = g^2 N.$$  

(8.2)

The large $N$ limit is defined as

$$N \to \infty, \quad g^2 \to 0, \quad t \text{ fixed.}$$

(8.3)
In this way the theory is still nontrivial. A first indication of this is the one-loop $\beta$ function of QCD, (5.11)--(5.12), which can be written as

$$\frac{d g}{d \mu} = -\left(\frac{11}{3} N - \frac{2}{3} N_f\right) \frac{g^3}{16\pi^2},$$

and becomes, after multiplying by $N^\frac{1}{2}$,

$$\frac{d t}{d \mu} = -\left(\frac{11}{3} N^\frac{1}{2} - \frac{2}{3} N^\frac{1}{2} N_f\right) \frac{t^3/N^2}{16\pi^2} = -\left(\frac{11}{3} - \frac{2}{3} N_f/N\right) \frac{t^3}{16\pi^2},$$

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,

$$A_\mu = \frac{t}{\sqrt{N}} \hat{A}_\mu, \quad q = \frac{t}{\sqrt{N}} \hat{q}.$$  

(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 [61]. Let us see how this works.

The quark propagator is

$$\langle \psi^i(x) \bar{\psi}^j(y) \rangle = \frac{t}{N} \delta^{ij} S(x-y), \quad i, j = 1, \cdots, N.$$  

(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 $\delta^{ij}$ in eq. (8.7). The gluon propagator is

$$\langle A^a_\mu(x) A^b_\nu(y) \rangle = \frac{t}{N} \delta^{ab} D_{\mu\nu}(x-y),$$  

(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.

\[(A_\mu)^i_j = A_\mu^a (T_a)^i_j \quad (8.9)\]

Here, $(T_a)^i_j$ is a basis of the Lie algebra which satisfies the normalization condition

\[\text{Tr} (T_a T_b) = \delta_{ab}, \quad a, b = 1, \cdots, N^2. \quad (8.10)\]

They also satisfy,

\[\sum_a (T_a)^i_j (T_a)^k_l = \delta^i_i \delta^k_j \quad (8.11)\]

for $U(N)$, and

\[\sum_a (T_a)^i_j (T_a)^k_l = \delta^i_i \delta^k_j - \frac{1}{N} \delta^i_j \delta^k_l \quad (8.12)\]

for $SU(N)$. Therefore, we can rewrite the $U(N)$ gluon propagator as

\[\langle A^i_{\mu j} (x) A^k_{\nu l} (y) \rangle = \frac{t}{N} D^{\mu \nu}_{\mu \nu} (x - y) \delta^i_i \delta^j_j. \quad (8.13)\]

The group structure of this propagator can be represented by a double line, as in Fig. 20.

![Figure 20: The gluon propagator in the double line notation.](image)

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

\[[T_a, T_b] = f_{abc} T_c. \quad (8.14)\]

By multiplying by $T_d$ and taking a trace, we find the relation

\[f_{abc} = \text{Tr} (T_a T_b T_c) - \text{Tr} (T_b T_a T_c). \quad (8.15)\]

The trace of three generators of the Lie algebra can be interpreted as a cubic vertex. Indeed, it comes from

\[\text{Tr}(A_\mu A_\nu A_\rho) = A_\mu^a A_\nu^b A_\rho^c \text{Tr}(T_a T_b T_c) \quad (8.16)\]
but in the double line notation it leads to the index structure
\[
\sum_{i,j,k} (A_\mu)^i_j (A_\nu)^j_k (A_\rho)^k_i
\]
(8.17)
which can be depicted as in (21). Since we have a commutator, we get an additional term
\[
-\sum_{i,j,k} (A_\nu)^i_j (A_\mu)^j_k (A_\rho)^k_i,
\]
(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. 22. In summary, we obtain a rule translating the group structure of the three-gluon vertex into the sum of two double-line structures. In

Figure 22: The twisted vertex (8.18) in the double line notation.

Fig. 23 we show a diagram with a quark line and one of the fatgraphs that it generates.

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

109
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

\[
N^h g^{2(E-V)}, \tag{8.19}
\]

but in terms of the ’t Hooft parameter this is

\[
N^{V-E+h}t^{V-E}. \tag{8.20}
\]

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 write

\[
h + V - E = \chi = 2 - 2g - b \tag{8.21}
\]

where \( g \) is the genus of the Riemann surface and \( b \) the number of boundaries. Therefore the factor of \( N \) in (8.19) is

\[
N^{2-2g-b} = N^\chi \tag{8.22}
\]

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

Figure 23: A typical QCD diagram (left) and one of the fatgraphs it generates (right).
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. Each of these diagrams will be a fatgraph $\Sigma$, which we will weight by $N$
to the power $h(\Sigma)$, the number of closed loops. We then have

$$r(G) = \sum_\Sigma N^{h(\Sigma)}.$$  \hspace{1cm} (8.23)

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 24: A planar graph obtained by contracting two cubic vertices.](image)

**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: the “untwisted” graph shown in Fig. 24, with multiplicity 2, and
the “twisted” graph shown in Fig. 25, also with multiplicity 2. Therefore, we find

$$r(\text{untwisted}) = 2N^3 - 2N.$$  \hspace{1cm} (8.24)

The first graph is planar, while the second one has $g = 1$.

### 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 (8.22). The largest values of $\chi$ 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

- The leading connected vacuum-to-vacuum graphs are of order $N^2$. They are planar graphs made out of gluons.
- 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

$$F(N, t) = \sum_{g=0}^{\infty} F_g(t)N^{2-g},$$

where

$$F_g(t) = \sum_{n \geq 0} a_{g,n}t^n$$

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

$$\text{Tr} F_{\mu\nu}F^{\mu\nu}, \quad \text{Tr}_R U_\gamma,$$

(8.27)
where
\[ U_\gamma = \text{P exp} \oint_\gamma A \]  
(8.28)
is a Wilson line operator around the closed loop \( \gamma \). We add to the action
\[ S \to S + N \sum_i J_i G_i \]  
(8.29)
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,
\[ \langle G_1 \cdots G_r \rangle^{(c)} = \frac{1}{N^r} \frac{\partial^r \Gamma(J)}{\partial J_1 \cdots \partial J_r} \big|_{J=0}. \]  
(8.30)
Since the leading contribution to this generating functional is again of order \( N^2 \), we conclude that
\[ \langle G_1 \cdots G_r \rangle^{(c)} \sim N^{2-r} \]  
(8.31)
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
\[ W^{(r)}(N, t) = \langle G_1 \cdots G_r \rangle^{(c)} = \sum_{g=0}^{\infty} W^{(r)}_g(t) N^{2-2g-r} \]  
(8.32)
where
\[ W^{(r)}_g(t) = \sum_{n \geq 0} W^{(r)}_{n,g} t^n \]  
(8.33)
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
\[ \bar{\psi} \psi, \quad \bar{\psi}(y) \text{P e}^{\int_x A \psi(x)}, \]  
(8.34)
and so on. We now perturb the action as
\[ S \to S + N \sum_i J_i M_i \]  
(8.35)
where \( b_i \) are sources, and
\[ \langle M_1 \cdots M_r \rangle^{(c)} = \frac{1}{N^r} \frac{\partial^r \Gamma(b)}{\partial J_1 \cdots \partial J_r} \big|_{J=0}. \]  
(8.36)
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. 26. We conclude that

$$\langle M_1 \cdots M_r \rangle^{(c)} \sim N^{1-r}. \quad (8.37)$$

Figure 26: A graph with $g = 0$, $b = 1$ where quark bilinear operators are inserted at the quark loop.

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

$${G_i|0\rangle \sim |G_i\rangle}, \quad {M_i|0\rangle \sim |M_i\rangle}. \quad (8.38)$$

To look for appropriately normalized states, we notice that

$${\langle G_1|G_2\rangle \sim \langle G_1G_2\rangle^{(c)} \sim \mathcal{O}(N^0)}, \quad (8.39)$$

therefore $G_i$ creates glueball states with unit amplitude. Similarly,

$${\langle M_1|M_2\rangle \sim \langle M_1M_2\rangle^{(c)} \sim \mathcal{O}(1/N)}, \quad (8.40)$$

therefore the appropriately normalized meson state is

$$\sqrt{N}M_i|0\rangle \quad (8.41)$$

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

$$\langle \sqrt{N}M_1 \cdots \sqrt{N}M_r \rangle^{(c)} \sim N^{1-r/2} \quad (8.42)$$
and each additional meson adds a $1/\sqrt{N}$ suppression. Finally, mixed glueball-meson correlators will be suppressed as

$$\langle G_1 \cdots G_s \sqrt{N} M_1 \cdots \sqrt{N} M_r \rangle^{(c)} \sim N^{1-s-r/2} \quad (8.43)$$

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 8.2.** Consider for example

$$\langle 0| \text{Tr}(FF)|M \rangle, \quad \langle 0| \text{Tr}(FF)|G \rangle \quad (8.44)$$

Using the rules above we find

$$\langle 0| \text{Tr}(FF)|M \rangle \sim \frac{1}{\sqrt{N}}, \quad \langle 0| \text{Tr}(FF)|G \rangle \sim \mathcal{O}(1). \quad (8.45)$$

In terms of rescaled fields, we have $\text{Tr}(\hat{F}\hat{F}) \sim \sqrt{N} \text{Tr}(FF)$, therefore

$$\langle 0| \text{Tr}(\hat{F}\hat{F})|M \rangle \sim \sqrt{N}, \quad \langle 0| \text{Tr}(\hat{F}\hat{F})|G \rangle \sim N. \quad (8.46)$$

We will use these results later one, when analyzing the $U(1)$ problem from the viewpoint of the $1/N$ expansion.

**Example 8.3.** *Large N scaling of $F_\pi$.* The pion decay constant is defined by (5.52)-(5.53). This has the structure

$$\langle 0|M_1 |M_2 \rangle \sim 1/\sqrt{N}. \quad (8.47)$$

Since $\hat{q} \sim \sqrt{N}q$, it follows that $\hat{A}_{ud} \sim NA_{ud}$, and we finally obtain

$$F_\pi \sim \sqrt{N}. \quad (8.48)$$

### 8.3 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 (8.26) 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 [16], \( N = 4 \) super Yang–Mills theory [13] and Chern–Simons theory [37].

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. 27. 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 [29] 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 \( j = 1, 2, \ldots, k \) (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 \( c_{j-1} c_{k-j} \). Summing over all the possible positions gives the recursion relation

\[
c_k = \sum_{j=1}^{k} c_{j-1} c_{k-j} \quad c_0 = 1
\]  

(8.49)
which is solved by the Catalan numbers

\[ c_k = \frac{(2k)!}{(k + 1)!k!}. \]  

(8.50)

This number, in contrast to (2.14), grows only exponentially

\[ c_k \sim 4^k. \]  

(8.51)

This is an indication that, if we sum over fatgraphs with the same topology, we might obtain amplitudes with a finite radius of convergence.

An example of this type of analyticity is discussed in Appendix D, where we promote the variable \( z \) in the quartic integral (2.25) to an \( N \)-vector variable, and the large \( N \) limit is now an analytic function of \( g \). This does not involve fatgraphs, but one can promote \( z \) to an \( N \times N \) Hermitian matrix and obtain instead of (2.25) a matrix integral with a quartic potential. This matrix integral can be studied, from the diagrammatic point of view, with the same tools that we have developed in QCD. The planar diagrams of this theory can be summed exactly, and in particular one can obtain a closed expression for \( F_0(t) \) in (8.26) which is indeed analytic at \( t = 0 \) [16]. We will indeed analyze in detail a matrix version of the anharmonic oscillator and we will show that the large \( N \) contribution to the ground state energy is an analytic function of the coupling constant.

### 8.4 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 [69], 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 (5.52). 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

$$\langle J(k)J(-k) \rangle = \sum_n \frac{a_n^2}{k^2 - m_n^2}. \quad (8.52)$$

Here the sum is over one-particle meson states $|n\rangle$ with masses $m_n$, and

$$a_n = \langle 0|J|n\rangle \quad (8.53)$$

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. 34 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 (8.52) 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 (8.52) is well-defined at large $N$. For example, if we normalize the currents as in (8.42), 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.

### 8.5 Baryons at large $N$

Baryons are color singlet hadrons made up of quarks. The $SU(N)$ invariant $\epsilon$-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 $\epsilon$-symbol must be different for it to be non-zero. Quarks obey Fermi statistics, and the $\epsilon$-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 [49] and references therein for an update on more recent results, and here we will content ourselves with some basic results from [69].

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. 28(a) is Fig. 28(b).

![Figure 28: A baryon interaction and its corresponding connected component](image)

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. 28(b) is given by cutting Fig. 23 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. 29 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.

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 $O(N^n)$ 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

$$H = Nm + \sum_i \frac{p_i^2}{2m} + \frac{1}{N} \sum_{i \neq j} V(x_i - x_j) + \frac{1}{N^2} \sum_{i \neq j \neq k} V(x_i - x_j, x_i - x_k) + \ldots$$  \hspace{1cm} (8.54)$$

Each term contributes $O(N)$ to the total energy. The interaction terms in the Hamiltonian eq. (8.54) 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

$$\psi_0(x_1, \ldots, x_N) = \prod_{i=1}^N \phi_0(x_i),$$  \hspace{1cm} (8.55)$$

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

$$\psi_1(x_1, \ldots, x_N) = \frac{1}{\sqrt{N}} \sum_{k=1}^N \phi_k(x_k) \prod_{i=1, i \neq k}^N \phi_0(x_i).$$  \hspace{1cm} (8.56)$$

Further details about this approach can be found in [23, 69].
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

$$L_M = \text{Tr} \left[ \frac{1}{2} \dot{M}^2 + V(M) \right],$$

where $V(M)$ is a polynomial in $M$. Notice that this problem has a symmetry

$$M \rightarrow U M U^\dagger$$

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

$$V(M) = \frac{1}{2} M^2 + V_{\text{int}}(M)$$

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.

![Feynman rules for matrix quantum mechanics.](image)

Figure 30: Feynman rules for matrix quantum mechanics.
The propagator of MQM is
\[ e^{-|\tau|} \frac{1}{2} \delta_{ik} \delta_{jl}. \] (9.4)

For a theory with a quartic interaction
\[ V_{\text{int}}(M) = \frac{g}{N} M^4 \] (9.5)
the Feynman rules are illustrated in Fig. 30. 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 to 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
\[ g^V N^{h-V} = g^V N^{2-2g}, \] (9.6)
since the number of edges is twice the number of vertices, \( E = 2V \) (this is a quartic interaction!) and
\[ h + E - V = h - V. \] (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 [16]). These numbers can be checked by taking into account that the total symmetry factor for connected diagrams with \( n \) quartic vertices is given by the Gaussian average
\[ \frac{1}{n!} \left\langle (\text{Tr} M^4)^n \right\rangle^{(c)}. \] (9.8)
where \( M \) is a Hermitian \( N \times N \) matrix. For example,
\[ \left\langle \text{Tr} M^4 \right\rangle = 2N^3 + N, \]
\[ \frac{1}{2!} \left\langle (\text{Tr} M^4)^2 \right\rangle^{(c)} = \frac{1}{2} \left( \left\langle (\text{Tr} M^4)^2 \right\rangle - \langle \text{Tr} M^4 \rangle^2 \right) = 18N^4 + 30N^2, \] (9.9)

Table 2: Symmetry factors of the planar quartic diagrams shown in Fig. 3.
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

$$
\begin{align*}
1 & : \frac{1}{4} \cdot 2 \\
2a & : -\frac{1}{16} \cdot 1 \cdot 16, \\
2b & : -\frac{1}{16} \cdot \frac{1}{2} \cdot 2, \\
3a & : \frac{1}{64} \cdot \frac{3}{2} \cdot 256 \\
3b & : \frac{1}{64} \cdot \frac{3}{8} \cdot 32 \\
3c & : \frac{1}{64} \cdot \frac{5}{8} \cdot 64 \\
3d & : \frac{1}{64} \cdot 1 \cdot 128
\end{align*}
$$

We then find

$$
E_0(N) = N^2 \mathcal{E}_0(g) + \mathcal{E}_1(g) + \cdots
$$

where

$$
\mathcal{E}_0(g) = \frac{1}{2} + \frac{1}{2}g - \frac{17}{16}g^2 + \frac{75}{16}g^3 + \cdots
$$

### 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 [16], which we now explain.

After quantization of the system we obtain a Hamiltonian operator

$$
H = \text{Tr} \left[ -\frac{1}{2} \frac{\partial^2}{\partial M^2} + V(M) \right],
$$

where

$$
\text{Tr} \frac{\partial^2}{\partial M^2} = \sum_{ab} \frac{\partial^2}{\partial M_{ab} \partial M_{ba}}
$$

In order to study the spectrum of this Hamiltonian, it is useful to change variables

$$
M = U \Lambda U^\dagger
$$
where

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_N)$$  \hspace{1cm} (9.16)

is a diagonal matrix. It can be shown that

$$\text{Tr} \frac{\partial^2}{\partial M^2} = \frac{1}{\Delta(\lambda)} \sum_{a=1}^{N} \left( \frac{\partial}{\partial \lambda_a} \right)^2 \Delta(\lambda) + \sum_{a < b} \frac{\mathcal{F}_{ab}}{(\lambda_a - \lambda_b)^2},$$  \hspace{1cm} (9.17)

where

$$\Delta(\lambda) = \prod_{a < b} (\lambda_a - \lambda_b)$$  \hspace{1cm} (9.18)

is the Vandermonde determinant, and $\mathcal{F}_{ab}$ are differential operators w.r.t. the angular coordinates in $U$ (see [50] for a statement of this result). Notice that the first term in (9.17) can be written as

$$\sum_{a=1}^{N} \left( \frac{\partial}{\partial \lambda_a} \right)^2 + \frac{2}{\Delta} \sum_{a=1}^{N} \frac{\partial \Delta}{\partial \lambda_a} \frac{\partial}{\partial \lambda_a} + \frac{1}{\Delta} \sum_{a=1}^{N} \frac{\partial^2 \Delta}{\partial \lambda_a^2}.$$  \hspace{1cm} (9.19)

Since

$$\log \Delta = \sum_{a < b} \log(\lambda_a - \lambda_b)$$  \hspace{1cm} (9.20)

we find

$$\frac{1}{\Delta} \frac{\partial \Delta}{\partial \lambda_a} = \frac{\partial \log \Delta}{\partial \lambda_a} = \sum_{b \neq a} \frac{1}{\lambda_a - \lambda_b}.$$  \hspace{1cm} (9.21)

We also have

$$\sum_{a=1}^{N} \frac{\partial^2 \Delta}{\partial \lambda_a^2} = 0.$$  \hspace{1cm} (9.22)

Therefore, we find, acting on singlet states (i.e., states that are invariant under the full $U(N)$ group)

$$-\frac{1}{2} \text{Tr} \frac{\partial^2}{\partial M^2} = -\frac{1}{2} \sum_{a=1}^{N} \frac{\partial^2}{\partial \lambda_a^2} + \sum_{b \neq a} \frac{1}{\lambda_b - \lambda_a} \frac{\partial}{\partial \lambda_a}. $$  \hspace{1cm} (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,

$$\Psi(\lambda_i),$$  \hspace{1cm} (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 \textit{antisymmetric} wavefunction
\[ \Phi(\lambda) = \Delta(\lambda)\Psi(\lambda) \] (9.25)

The equation
\[ H\Psi = E\Psi \] (9.26)
can be written as
\[ \left( \sum_{i=1}^{N} h(\lambda_i) \right) \Phi(\lambda_j) = E\Phi(\lambda_j) \] (9.27)
where \( h(\lambda) \) is the Hamiltonian
\[ h(\lambda) = -\frac{\hbar^2}{2} \frac{\partial^2}{\partial\lambda^2} + V_N(\lambda). \] (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 Schrödinger equation for a single particle of unit mass,
\[ h(\lambda)\phi_n(\lambda) = E_n\phi_n(\lambda) \] (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
\[ E(N) = \sum_{n=1}^{N} E_n \] (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
\[ V_N(\lambda) = NV(\lambda/\sqrt{N}). \] (9.31)

In this case, one can see that \( \lambda \) is of order \( N^{\frac{1}{2}} \), so we can redefine
\[ \lambda \rightarrow N^{\frac{1}{2}}\lambda, \] (9.32)
and \( \lambda \) is now of order 1. In this way we find a Hamiltonian where the potential only depends on \( \lambda \)
\[ \frac{1}{N} h(\lambda) = -\frac{\hbar^2}{2N^2} \frac{\partial^2}{\partial\lambda^2} + V(\lambda). \] (9.33)
and the problem to be solved is
\[ \left\{-\frac{\hbar^2}{2N^2} \frac{d^2}{d\lambda^2} + V(\lambda)\right\}\phi_n(\lambda) = \epsilon_n\phi_n(\lambda). \] (9.34)
where we denoted
\[ e_n = \frac{1}{N} E_n \]  
(9.35)

For example,
\[ V_N(\lambda) = \frac{1}{2} \lambda^2 + \frac{g}{N} \lambda^4, \]
(9.36)

has the right scaling properties. This can be interpreted as saying that
\[ g = N \gamma \]  
(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
\[ E_0(N) = \sum_{k=1}^{N} E_k = N \sum_{k=1}^{N} e_k = N^2 \mathcal{E}_0 + \cdots, \]
(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
\[ NJ(e_n) = n - \frac{1}{2}, \quad n \geq 1, \]
(9.39)

where
\[ J(e) = \frac{1}{\pi \hbar} \int_{\lambda_{1}(e)}^{\lambda_{2}(e)} d\lambda \sqrt{2(e - V(\lambda))} \]
(9.40)

and \( \lambda_{1,2}(e) \) are the turning points of the potential. If we denote
\[ \xi = \frac{n - \frac{1}{2}}{N}, \]
(9.41)

we see that (9.39) defines implicitly a function \( e(\xi) \). The total ground state energy can then be written as
\[ E_0(N) = N \sum_{n=1}^{N} e(\xi). \]
(9.42)

At large \( N \), the spectrum becomes denser and denser, and the variable \( \xi \) becomes a continuous variable
\[ \xi \in [0,1]. \]
(9.43)
At large $N$, the sum in (9.42) becomes an integral through the rule
\[
\sum_{n=1}^{N} \rightarrow N \int_0^1 d\xi \tag{9.44}
\]
and we find
\[
E_0(N) \rightarrow N^2 \int_0^1 d\xi e(\xi), \tag{9.45}
\]
in other words,
\[
\mathcal{E}_0 = \int_0^1 d\xi e(\xi). \tag{9.46}
\]
To evaluate this integral, we change variables from $\xi$ 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,
\[
\mathcal{E}_0 = \int_0^1 d\xi e(\xi) = \int_{\min V(\lambda)}^{e_F} de \, eJ'(e), \tag{9.48}
\]
where
\[
J'(e) = \frac{1}{\pi \hbar} \int_{\lambda_1(e)}^{\lambda_2(e)} d\lambda \frac{e}{\sqrt{2(e - V(\lambda))}}. \tag{9.49}
\]
An easy calculation gives,
\[
\mathcal{E}_0 = \frac{1}{\pi \hbar} \int_{\min V(\lambda)}^{e_F} de \int_{\lambda_1(e)}^{\lambda_2(e)} d\lambda \frac{e}{\sqrt{2(e - V(\lambda))}}
= \frac{1}{\pi \hbar} \int_{\lambda_1(e_F)}^{\lambda_2(e_F)} d\lambda \int_{V(\lambda)}^{e_F} de \frac{e}{\sqrt{2(e - V(\lambda))}} \tag{9.50}
= \frac{1}{3\pi \hbar} \int_{\lambda_1(e_F)}^{\lambda_2(e_F)} d\lambda (2V(\lambda) + e_F) \sqrt{2(e_F - V(\lambda))}
\]
and the final expression we obtain is
\[
\mathcal{E}_0 = e_F - \frac{1}{3\pi \hbar} \int_{\lambda_1(e_F)}^{\lambda_2(e_F)} d\lambda \left[2(e_F - V(\lambda))\right]^{3/2}. \tag{9.51}
\]
The previous development suggests to introduce a density of eigenvalues $\rho(\lambda)$. Using (9.47) we find that
\[
\frac{1}{\pi \hbar} \int_{\lambda_1(e_F)}^{\lambda_2(e_F)} d\lambda \sqrt{2(e_F - V(\lambda))} = 1, \tag{9.52}
\]
127
therefore
\[ \rho(\lambda) = \frac{1}{\pi\hbar} \sqrt{2(e_F - V(\lambda))} \quad (9.53) \]
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 [50].

The first excited state can be obtained by exciting the last fermion in the Fermi sea, i.e.
\[ E_1(N) = E(N) + E_{N+1} - E_N, \quad (9.54) \]
therefore
\[ E_1(N) - E(N) = E_{N+1} - E_N = N(e_{N+1} - e_N) = N\left( e(\xi + 1/N) - e(\xi) \right)_{\xi=1} \quad (9.55) \]
which at leading order in \(1/N\) is given by
\[ E_1(N) - E(N) = \frac{d}{d\xi} \bigg|_{\xi=1} \omega, \quad (9.56) \]
where
\[ \omega = \frac{1}{J'(e_F)}. \quad (9.57) \]
Notice that \(\omega\) 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
\[ 0 \leq h_1 < h_2 < \cdots < h_r, \quad 1 \leq p_1 < p_2 < \cdots < p_r, \quad (9.58) \]
and its energy is
\[ E_{h,p}(N) - E_0(N) \sim \omega \sum_{i=1}^{r} (h_i + p_i). \quad (9.59) \]

### 9.4 Some examples

**Example 9.1. Harmonic oscillator.** A simple case occurs for
\[ V_N(\lambda) = \frac{1}{2} \omega^2 \lambda^2. \quad (9.60) \]
The exact answer for the ground state energy is

\[ E_0(N) = N \sum_{n=1}^{N} \frac{\hbar}{N} \omega \left( n - \frac{1}{2} \right) = \frac{\hbar \omega}{2} N^2, \]  

(9.61)

therefore

\[ E_0 = \frac{\hbar \omega}{2}. \]  

(9.62)

Let us now compute this with the formulae above. First of all, we have that

\[ J(\theta) = \frac{\theta}{\hbar \omega} \Rightarrow e(\phi) = \hbar \omega \phi, \]  

(9.63)

therefore

\[ e_F = e(1) = \hbar \omega. \]  

(9.64)

One also finds,

\[ E_0 = e_F - \frac{1}{2} \frac{e_F^2}{\hbar^2 \omega^2} = \frac{1}{2} \hbar \omega, \]  

(9.65)

which agrees with the direct computation above.

**Example 9.2.** *The quartic potential.* This is the potential originally considered in [16]. The potential is given by

\[ V(\lambda) = \frac{1}{2} \lambda^2 + g\lambda^4. \]  

(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

\[ 2e - \lambda^2 - 2g\lambda^4 = 2g(a^2 - \lambda^2)(b^2 + \lambda^2), \]  

(9.67)

where

\[ a^2 = \frac{\sqrt{16eg + 1} - 1}{4g}, \quad b^2 = \frac{\sqrt{16eg + 1} + 1}{4g}. \]  

(9.68)

We introduce the elliptic modulus

\[ k^2 = \frac{a^2}{a^2 + b^2}. \]  

(9.69)

Then, we have that

\[ J(e) = \frac{2}{3\pi \hbar} (2g)^{\frac{1}{2}} (a^2 + b^2)^{\frac{1}{2}} \left[ b^2 K(k) + (a^2 - b^2) E(k) \right]. \]  

(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)
\[
e_F(g) = 1 + \frac{3g}{2} - \frac{17g^2}{4} + \frac{375g^3}{16} + \mathcal{O}(g^4).
\] (9.71)

The planar free energy is given by
\[
\mathcal{E}_0(g) = e_F(g) - \frac{1}{3\pi} I(g, e_F(g)).
\] (9.72)

which involves the integral
\[
I(g, e) = \int_{-a}^{a} dt \left[ (a^2 - t^2)(b^2 + t^2) \right]^{\frac{1}{2}} = \frac{2}{35} \sqrt{a^2 + b^2} \left\{ 2(a^2 - b^2)(a^4 + 6a^2b^2 + b^4)E(k) + b^2(2b^4 + 9a^2b^2 - a^4)K(k) \right\}. \tag{9.73}
\]

This can be also be computed in perturbation theory in $g$, and it gives
\[
\mathcal{E}_0 = \frac{1}{2} + \frac{g}{2} - \frac{17g^2}{16} + \frac{75g^3}{16} + \mathcal{O}(g^4)
\] (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. 31.
Figure 31: 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.

Figure 32: The Fermi energy $e_F$ as a function of $g$. For $g = 0$, one has $e_F = 1$.

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 Fig. 32. Using this result, we evaluate the planar free energy, which we plot it in Fig. 33 (red line). In the same figure we show 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 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 [1].

To start with, we have to add fermions to the model. For this, we add to (9.1) the fermionic piece

\[
L_F = q_t^i \left[ \frac{d}{d\tau} + m\sigma_3 + \frac{gF}{\sqrt{N}} M\sigma_1 \right] q.
\]  

(9.77)

Here, \( q_t^i,\alpha \) are two-component Fermi fields, with \( \alpha = 1, 2 \) and \( i, f \) color and flavor indices, respectively. The standard vacuum is defined by

\[
q_t^i,1|0\rangle = q_t^i,2|0\rangle = 0,
\]  

(9.78)

and the canonical commutation relations are

\[
\{q_t^i,\alpha, q_t^{i',\beta}\} = \delta^{ij}\delta_{ff'}\delta_{\alpha\beta}.
\]  

(9.79)

The fermion propagator is given by

\[
\langle 0|T(q_t^i,\alpha(\tau)q_t^{i',\beta}(\tau))|0\rangle = \delta^{ij}\delta_{ff'} \int\frac{dp}{2\pi}e^{ip(\tau-\tau')}(\frac{1}{ip + m\sigma_3})_{\alpha\beta}
\]  

(9.80)
This is computed as follows

\[ \int_{-\infty}^{\infty} \frac{dp}{2\pi} e^{ip(\tau-\tau')} = \int_{-\infty}^{\infty} \frac{dp}{2\pi} e^{ip(\tau-\tau') - i p + m\sigma_3} \]

\[ = \frac{m\sigma_3}{2\pi} \int_{-\infty}^{\infty} \frac{dp}{p^2 + m^2} - i \int_{-\infty}^{\infty} \frac{dp}{2\pi} e^{ip(\tau-\tau')} \]

\[ = \frac{m\sigma_3}{2m} e^{-m|\tau-\tau'|} - \frac{i 2\pi}{2\pi} e^{m|\tau-\tau'|} \]

\[ = \frac{1}{2} e^{-m|\tau-\tau'|} \left[ \sigma_3 + e(\tau - \tau') \right] . \] 

Therefore,

\[ \langle 0 | T(q^i_{f,\alpha}(\tau)q^{i\dagger}_{f,\beta}(\tau)) | 0 \rangle = \frac{1}{2} e^{-m|\tau-\tau'|} \left[ \sigma_3 + e(\tau - \tau') \right] \alpha \beta . \] 

In order to study the spectrum of the problem, we diagonalize \( M \) as in (9.15) and change variables to

\[ q^i \to \tilde{q}^i = U^i_j q^j \] 

so that the coupling between \( M \) and \( q, q^\dagger \) in (9.77) is diagonal. It is easy to compute that

\[ \frac{\partial \tilde{q}_a}{\partial M_{ij}} = \frac{\partial U_{ab}}{\partial M_{ij}} \frac{q^b}{\lambda_a - \lambda_b}. \]

We can write this as

\[ \frac{\partial \tilde{q}_a}{\partial M_{ij}} = [O_{ij}, \tilde{q}_a], \]

where

\[ O_{ij} = \sum_{a \neq b, \alpha, \beta} \frac{q^a_{\beta,\alpha} U_{\alpha,b} q^b_{\beta,\alpha}}{\lambda_a - \lambda_b}. \]

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 \([1, 48]\), and leads to a quartic term in fermion fields. When acting on singlets it reads (we set \( \tilde{q} = q \) from now on)

\[ H_F = \sum_{i=1}^{N} q^{i\dagger} \left( m\sigma_3 + \frac{g_F}{\sqrt{N}} \lambda_i \right) q^i + \frac{1}{2} \sum_{i \neq j} q^{i\dagger}_{\beta,\alpha} q^i_{\beta,\alpha} q^{j\dagger}_{\beta,\alpha} q^j_{\beta,\alpha} \frac{(\lambda_i - \lambda_j)^2}{(\lambda_i - \lambda_j)^2}. \]
To find the vacuum, we define a new set of Fermi operators $a^i_j, b^i_j$ which annihilate by definition the true ground state:

$$a^i_j|\theta\rangle = b^i_j|\theta\rangle = 0.$$  

(9.88)

The new operators are obtained from the old ones by a rotation of the form

$$\begin{pmatrix} q^i_{j,1} \\ q^i_{j,2} \end{pmatrix} = \exp\left\{ \frac{i}{2} \sigma_2 \theta_i \right\} \begin{pmatrix} a^i_j \\ b^i_j \end{pmatrix},$$

(9.89)

and this leads to a Hamiltonian $[1, 48]$

$$H_F = \frac{1}{2} \sum_{j \neq i} \frac{\cos^2 \left[ \frac{1}{2} (\theta_i - \theta_j) \right]}{(\lambda_i - \lambda_j)^2} \left( a^i_j a^i_j + b^i_j b^i_j + a^i_j a^i_j a^i_j a^i_j + 2 a^i_j b^i_j a^i_j b^i_j + b^i_j b^i_j b^i_j b^i_j \right)$$

$$+ \frac{1}{2} \sum_{j \neq i} \frac{\sin^2 \left[ \frac{1}{2} (\theta_i - \theta_j) \right]}{(\lambda_i - \lambda_j)^2} \left( 1 - a^i_j a^i_j - b^i_j b^i_j - a^i_j a^i_j b^i_j b^i_j - 2 a^i_j b^i_j a^i_j b^i_j - a^i_j a^i_j b^i_j b^i_j \right)$$

$$+ \frac{1}{2} \sum_{j \neq i} \frac{\sin (\theta_i - \theta_j)}{(\lambda_i - \lambda_j)^2} \left( a^i_j b^i_j - a^i_j b^i_j - a^i_j a^i_j a^i_j b^i_j - a^i_j b^i_j a^i_j a^i_j - a^i_j b^i_j a^i_j b^i_j - a^i_j b^i_j b^i_j a^i_j \right)$$

$$+ m \sum_i \cos \theta_i (a^i_j a^i_j + b^i_j b^i_j - 1) + m \sum_i \sin \theta_i (a^i_j b^i_j - a^i_j b^i_j)$$

$$- g_F \sum_i \frac{\lambda_i}{\sqrt{N}} \sin \theta_i (a^i_j a^i_j + b^i_j b^i_j - 1) + g_F \sum_i \frac{\lambda_i}{\sqrt{N}} \cos \theta_i (a^i_j b^i_j - a^i_j b^i_j).$$

(9.90)

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 $1/N$ [1]. For simplicity we will consider the case in which there is one single flavour, as in [1]. 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 [5] in the context of the ’t Hooft model for QCD$_2$. The theta angles are then fixed by the condition

$$\frac{1}{2} \sum_{j \neq i} \frac{\sin (\theta_i - \theta_j)}{(\lambda_i - \lambda_j)^2} + m \sin \theta_i + \frac{g_F}{\sqrt{N}} \lambda_i \cos \theta_i = 0.$$  

(9.91)

We now assume that at large $N$ the angles $\theta_i$ become functions of the eigenvalue $\lambda$, 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,

$$\theta_i \rightarrow \theta(\lambda)$$  

(9.92)
and
\[ \sum_i h(\lambda_i) \to N \int d\lambda \rho(\lambda) h(\lambda). \tag{9.93} \]

At large \( N \) the equation for the angle (9.91) becomes an integral equation
\[ \frac{1}{2} P \int d\lambda' \rho(\lambda') \frac{\sin[\theta(\lambda) - \theta(\lambda')]}{(\lambda - \lambda')^2} + m \sin \theta(\lambda) + g_F \lambda \cos \theta(\lambda) = 0 \tag{9.94} \]

where we rescaled the eigenvalues \( \lambda_i \) 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
\[ E_F = \langle \theta | H_F | \theta \rangle. \tag{9.95} \]

This is trivial to compute from the normal-ordered Hamiltonian (9.90), since only the constant terms contribute. In the large \( N \) limit we obtain
\[ \frac{E_F}{N} = \frac{1}{2} \int d\lambda d\lambda' \rho(\lambda) \rho(\lambda') \frac{\sin^2[\theta(\lambda) - \theta(\lambda')]}{(\lambda - \lambda')^2} + \int d\lambda \rho(\lambda) \left[ -m \cos \theta(\lambda) + g_F \lambda \sin \theta(\lambda) \right]. \tag{9.96} \]

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 (9.94) is solved by the trivial solution
\[ \theta(\lambda) = 0 \tag{9.97} \]

therefore
\[ E_F = -mN, \tag{9.98} \]

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 (9.94), see the Appendix of [1].

Let us now study mesons. Their wavefunctions have the structure
\[ |\phi_M\rangle = \sum_i f_i a^i b^i |\theta\rangle, \tag{9.99} \]

since they correspond to \( \bar{q}q \) states. At leading order in \( 1/N \) we find the equation
\[ (H_F - E_F) |\phi_M\rangle = E_M |\phi_M\rangle \tag{9.100} \]

Again, the dynamics for the mesons takes place in the background of the master field for the pure “glue” theory.
To analyze (9.100) we consider the fermion-number conserving terms in the Hamiltonian $H_F$. After subtracting $E_F$, the quadratic part, proportional to $a^i a_i^\dagger + b^i b_i^\dagger$, can be written as

$$\sum_i \left[ \frac{1}{2} \sum_{j \neq i} \frac{\cos(\theta_i - \theta_j)}{(\lambda_i - \lambda_j)^2} + m \cos \theta_i - g_F \frac{\lambda_i}{\sqrt{N}} \sin \theta_i \right] (a_i^\dagger a_i^i + b_i^\dagger b_i^j) \tag{9.101}$$

where we have combined the terms appearing in the first and second lines of (9.90) as

$$\cos^2 \left[ \frac{1}{2} (\theta_i - \theta_j) \right] - \sin^2 \left[ \frac{1}{2} (\theta_i - \theta_j) \right] = \cos(\theta_i - \theta_j). \tag{9.102}$$

The operator (9.101) acting on (9.99) leads to

$$\sum_i \left[ \sum_{j \neq i} \cos^2 \left[ \frac{1}{2} (\theta_i - \theta_j) \right] \right] \left( \lambda_i - \lambda_j \right)^2 f_i a_i^\dagger b_i^j |\theta\rangle \tag{9.103}$$

while the quartic term

$$\sum_i \left[ \sum_{j \neq i} \cos^2 \left[ \frac{1}{2} (\theta_i - \theta_j) \right] \right] \left( \lambda_i - \lambda_j \right)^2 f_i a_i^\dagger b_i^j \tag{9.104}$$

gives

$$- \sum_i \left[ \sum_{j \neq i} \cos^2 \left[ \frac{1}{2} (\theta_i - \theta_j) \right] \right] \left( \lambda_i - \lambda_j \right)^2 f_i a_i^\dagger b_i^j |\theta\rangle, \tag{9.105}$$

where the $-$ sign arises from anticommutation. The Schrödinger equation for the mesons becomes

$$\left[ \sum_{j \neq i} \cos(\theta_i - \theta_j) \left( \lambda_i - \lambda_j \right)^2 + 2m \cos \theta_i - 2g_F \frac{\lambda_i}{\sqrt{N}} \sin \theta_i \right] f_i - \sum_{j \neq i} \cos^2 \left[ \frac{1}{2} (\theta_i - \theta_j) \right] \left( \lambda_i - \lambda_j \right)^2 f_j = E_M f_i. \tag{9.106}$$

In the large $N$ limit we have

$$f_i \to f(\lambda) \tag{9.107}$$

and we obtain the integral equation

$$P \int \frac{d\lambda' \rho(\lambda')}{(\lambda - \lambda')^2} \left[ \cos(\theta(\lambda) - \theta(\lambda'))(f(\lambda) - f(\lambda')) - \sin^2 \frac{\theta(\lambda) - \theta(\lambda')}{2} f(\lambda') \right] \tag{9.108}$$

$$+ \left[ 2m \cos \theta(\lambda) - 2g_F \lambda \sin \theta(\lambda) \right] f(\lambda) = E_M f(\lambda).$$

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

$$\tilde{f}(\lambda) = \rho(\lambda) f(\lambda).$$

The integral equation (9.108) becomes

$$-\rho(\lambda) \frac{d\lambda}{(\lambda - \lambda')^2} \left[ 1 - \sin^2(\theta(\lambda) - \theta(\lambda'))/2 \right] + \tilde{f}(\lambda) \tilde{q}(\lambda) = E \tilde{f}(\lambda),$$

where

$$\tilde{q}(\lambda) = \int \frac{d\lambda'}{(\lambda - \lambda')^2} \cos \left[ \theta(\lambda) - \theta(\lambda') \right] + 2m \cos \theta(\lambda) - 2gF_\lambda \sin \theta(\lambda).$$

Let us assume that

$$\tilde{f}(\lambda) \sim e^{iE g(\lambda) + \cdots}$$

for large $E$. The integral

$$I(\lambda) = P \int \frac{d\lambda' e^{iE g(\lambda')}}{(\lambda - \lambda')^2}$$

can be evaluated by the saddle-point method, and the largest contribution comes from $\lambda \sim \lambda'$. The computation gives

$$I(\lambda) \sim -e^{iE g(\lambda)} \pi E|g'(\lambda)| + O(1/E).$$

The term $\sin^2(\theta(\lambda) - \theta(\lambda'))/2$ is subleading in this expansion, and (9.110) becomes

$$\pi E \rho(\lambda)|g'(\lambda)| + \tilde{q}(\lambda) = E + O(1/E).$$

This is solved by

$$g(\lambda) = \pm \frac{1}{\pi} \int_{\lambda_1(e_F)}^{\lambda} \frac{d\lambda'}{\rho_0(\lambda')} \left[ 1 - \frac{\tilde{q}(\lambda')}{E} \right] + \text{constant.}$$

We then find the real solution

$$f(\lambda) \sim \frac{1}{\rho(\lambda)} \sin \left\{ \frac{E}{\pi} \int_{\lambda_1(e_F)}^{\lambda} \frac{d\lambda'}{\rho_0(\lambda')} \left[ 1 - \frac{\tilde{q}(\lambda')}{E} \right] + \phi \right\}.$$ 

Since $\rho(\lambda_2(e_F)) = 0$, the sine function must vanish at the endpoint of the distribution, and this gives

$$E_n = \left[ \int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \frac{d\lambda}{\rho(\lambda)} \right]^{-1} \left[ \pi^2 n + \int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \frac{d\lambda}{\rho(\lambda)} \tilde{q}(\lambda) - \pi \phi \right] + O(1/n)$$

(9.118)
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

\[
\pi^2 \left[ \int_{\lambda_1(e_F)}^{\lambda_2(e_F)} \frac{d\lambda}{\rho(\lambda)} \right]^{-1} = \omega
\]

(9.119)

where \( \omega \) 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 [69], see [48].

**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 [60]. This theory is defined by the Hamiltonian density

\[
\mathcal{H} = \frac{1}{2} \text{Tr} E^2 + \overline{\Psi}(i \gamma_1 \partial_1 + m) \Psi
\]

(9.120)

and the constraint

\[
\partial_1 E = J(x) = -g \Psi^\dagger \Psi.
\]

(9.121)

One can integrate out \( E \) to obtain a quartic Hamiltonian

\[
H = -\frac{1}{4} \int dx \, dy \, \text{Tr}(J(x)|x - y|J(y)) + \int dx \, \overline{\Psi}(i \gamma_1 \partial_1 + m) \Psi.
\]

(9.122)

This fermionic Hamiltonian can be analyzed with the same techniques we have used, see [1] and specially [5] for a detailed study.

**10 Applications in QCD**

**10.1 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 \( \eta' \) is not a Goldstone boson. To understand this more quantitatively, we should understand how the anomaly gives a mass to the \( \eta' \). This was solved by Witten [68] (and subsequently by Veneziano [62]) by using large \( N \) techniques, who obtained a remarkable expression for the mass of the \( \eta' \) at leading order in \( 1/N \) known as the Witten–Veneziano formula. This formula has been spectacularly confirmed by lattice gauge theory calculations [28].
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 (5.77) that
\[ \partial_{\mu} J^\mu = \frac{2Nf}{N} \frac{t}{64\pi^2} \varepsilon_{\mu\nu\rho\sigma} (\hat{F}^{\mu\nu}, \hat{F}^{\rho\sigma}) \] when expressed in terms of normalized fields (8.6). Therefore, at large $N$ the $\eta'$ 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 (5.57)-(5.58). To compute $m^2_{\eta'}$ we have to compute
\[ \langle \eta'(q)|q(x)|0 \rangle. \] (10.2)
To have a handle on this, we study the two point function (5.26), evaluated in the theory with quarks. This function can be organized as
\[ U(k) = \sum_{L \geq 0} U_L(k), \] (10.3)
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 (8.39))
\[ U_0(k) \sim O(N^0), \quad U_1(k) \sim O(N^{-1}). \] (10.4)
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,
\[ U(k) = \sum_{G_i} \frac{a_i^2}{k^2 - m_i^2} + \sum_{M_i} \frac{c_i^2}{N(k^2 - m_i^2)} \] (10.5)
where the first sum is over glueball states, and the second sum is over meson states. In this equation,
\[ a_i = (2\pi)^{3/2} \sqrt{2E}\langle 0|q(0)|G_i \rangle, \quad \frac{c_i}{\sqrt{N}} = (2\pi)^{3/2} \sqrt{2E}\langle 0|q(0)|M_i \rangle, \] (10.6)
see for example (A.10). We have already extracted the leading $N$ dependence as it follows from (8.45), so that $c_i, a_i$ are of order one. We have that, at leading order in the $1/N$ expansion,
\[ U_0(k) \sim \sum_{G_i} \frac{a_i^2}{k^2 - m_i^2}, \quad U_1(k) \sim \sum_{M_i} \frac{c_i^2}{N(k^2 - m_i^2)}. \] (10.7)

We know that, in a world of massless quarks, there is no $\theta$ dependence, and the topological susceptibility vanishes. But
\[ \chi_t = U(0). \] (10.8)
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 $\eta'$. If this is the case, the term

$$\frac{c^2_{\eta'}}{N(k^2 - m^2_{\eta'})}$$

in the sum over meson resonances becomes at $k = 0$

$$- \frac{c^2_{\eta'}}{Nm^2_{\eta'}} \sim \mathcal{O}(N^0)$$

and can kill the glueball contribution. Notice that this contribution is precisely

$$U_0(0)$$

at leading order in $N$. We deduce

$$\frac{c^2_{\eta'}}{Nm^2_{\eta'}} = U_0(0).$$

We can now put everything together to deduce a formula for $m^2_{\eta'}$. We have from (5.80) that

$$\frac{F_{\eta'} m^2_{\eta'}}{(2\pi)^{3/2}\sqrt{2E}} = \langle \eta'(p)|\partial_\mu J^\mu(0)|0\rangle = 2N_f \langle \eta'(p)|q(0)|0\rangle = \frac{1}{(2\pi)^{3/2}\sqrt{2E}} \frac{2N_f c_{\eta'}}{\sqrt{N}},$$

in other words,

$$c_{\eta'} = \frac{\sqrt{N}}{2N_f} F_{\eta'} m^2_{\eta'}.$$

Plugging this into (10.12) we find

$$m^2_{\eta'} = \frac{4N_f^2}{F^2_{\eta'}} U_0(0).$$

After an appropriate normalization, $F_{\eta'}$ equals $F_\pi$ at leading order in the $1/N$ expansion,

$$F_{\eta'} = \sqrt{2N_f F_\pi}.$$
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

$$m_{\eta'}^2 = \frac{2N_f}{F_\pi^2} \chi_t^{YM}$$

(10.17)

where $\chi_t^{YM}$ is the topological susceptibility in pure gluodynamics. In principle, $\chi_t^{YM}$ 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 [67, 25]. The consistent picture of the $\eta'$ developed by Witten in [68] requires that this is also the case in QCD.

The formula (10.17) for the mass of the $\eta'$ 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 [62]. The mass matrix $\mathcal{M}_{\eta-\eta'}$ for the $\eta, \eta'$ in the approximation $m_u = m_d$ is written down in (B.55). The diagonalization of this matrix leads to the masses $m_\eta^2, m_{\eta'}^2$. Since the trace is a unitary invariant, we find

$$\text{Tr } \mathcal{M}_{\eta-\eta'} = m_\eta^2 + m_{\eta'}^2 = 2m_K^2 + \frac{6}{F_\pi^2} \chi_t^{YM}$$

(10.18)

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,

$$\chi_t^{YM} = \frac{F_\pi^2}{6} \left( m_{\eta'}^2 + m_\eta^2 - 2m_K^2 \right).$$

(10.19)

Interestingly, the Witten–Veneziano solution of the $U(1)$ problem in 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, recent lattice calculations have been able to determine $\chi_t^{YM}$ for $N = 3$. One finds [28]

$$\chi_t^{YM} = (191 \pm 5 \text{ MeV})^4$$

(10.20)

On the other hand, by plugging the experimental values of the pion masses in (10.19) we get

$$\frac{F_\pi^2}{6} \left( m_{\eta'}^2 + m_\eta^2 - 2m_K^2 \right) \approx (180 \text{ MeV})^4.$$

(10.21)
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 [40].

A Polology and spectral representation

Let us consider a general correlation function in momentum space

\[ G(q_1, \cdots, q_n) = \int d^4x_1 \cdots d^4x_n e^{-iq_1 \cdot x_1} \cdots e^{-iq_n \cdot x_n} \langle A_1(x_1) \cdots A_n(x_n) \rangle. \] (A.1)

The analytic structure of this function in momentum space is quite complicated. Following [65], let us consider this as a function of \( q_2 \), where

\[ q = q_1 + \cdots + q_r = -q_{r-1} - \cdots - q_n \] (A.2)

and \( 1 \leq r \leq n - 1 \). A general nonperturbative result in QFT says that \( G \) has a pole at

\[ q^2 = -m^2 \] (A.3)

where \( m^2 \) is the mass of any one-particle state that has nonvanishing matrix elements with the states

\[ A_1^\dagger \cdots A_r^\dagger |0\rangle, \quad A_{r+1} \cdots A_n |0\rangle. \] (A.4)

The pole has the structure

\[ \frac{2i\sqrt{q_2^2 + m^2}}{q_2^2 + m^2 - i\epsilon} (2\pi)^7 \delta^4(q_1 + \cdots + q_n) \sum_\sigma M_0 |q,\sigma \rangle M_{q,\sigma} |0(q_{r+2}, \cdots, q_n) \] (A.5)

where the \( M \)s are defined by

\[ (2\pi)^4 \delta^4(q_1 + \cdots + q_r - p) M_0 |q_2, \cdots, q_r \rangle \]
\[ = \int d^4x_1 \cdots d^4x_r e^{-iq_1 \cdot x_1} \cdots e^{-iq_r \cdot x_r} \langle 0 | A_1(x_1) \cdots A_r(x_r) | \vec{p}, \sigma \rangle, \] (A.6)

\[ (2\pi)^4 \delta^4(q_{r+1} + \cdots + q_n - p) M_{\vec{p},\sigma} |0(q_{r+2}, \cdots, q_n) \]
\[ = \int d^4x_{r+1} \cdots d^4x_n e^{-iq_{r+1} \cdot x_{r+1}} \cdots e^{-iq_n \cdot x_n} \langle \vec{p}, \sigma | A_{r+1}(x_{r+1}) \cdots A_n(x_n) | 0 \rangle. \]

Therefore, we can read the residue at the pole from (A.5). The factors

\[ (2\pi)^{3/2} \left[ 2 \sqrt{\vec{k}^2 + m^2} \right]^{1/2} \] (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. 34, 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.

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
\[
\langle 0|\Phi(0)\rangle. \tag{A.9}
\]
Therefore, we have that
\[
\langle 0|\Phi(0)|\vec{p},\sigma\rangle = (2\pi)^{-3/2}(2\sqrt{\vec{p}^2 + m^2})^{-1/2}N_\sigma \tag{A.10}
\]
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
\[
\langle 0|\Phi(x)|\vec{p},\sigma\rangle = e^{ip\cdot x}\langle 0|\Phi(0)|\vec{p},\sigma\rangle \tag{A.11}
\]
therefore in this case
\[ M_{0|\vec{p},\sigma} = (2\pi)^{-3/2} (2\sqrt{\vec{p}^2 + m^2})^{-1/2} N_{\sigma}, \quad M_{\vec{p}|0} = (2\pi)^{-3/2} (2\sqrt{\vec{p}^2 + m^2})^{-1/2} N_{\sigma}^*, \]
(A.12)
and the above result implies that the momentum space function
\[ -i\Delta(q) = \langle \Phi(q)\Phi^\dagger(-q) \rangle = \int d^4x \exp^{-iq(x-y)} \langle 0|\Phi(x)\Phi^\dagger(y)|0 \rangle \]
(A.13)
has a pole at
\[ q^2 = -m^2 \]
(A.14)
with residue
\[ Z = |N_{\sigma}|^2 \]
(A.15)
This result can be rephrased in yet another way by using the Käller-Lehman representation of the propagator,
\[ \Delta'(p) = \int_0^\infty \rho(\mu^2) \frac{d\mu^2}{\mu^2 + \mu^2 - i\epsilon}. \]
(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
\[ \rho(\mu^2) = Z\delta(\mu^2 - m^2) + \sigma(\mu^2), \]
(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 [66] 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
\[ q \rightarrow \exp\left[i \sum_a \theta^V_a T^a + \theta^A_a T^a \gamma_5\right] q. \]
(B.18)
In order to write a Lagrangian for Goldstone bosons, we rewrite the quark fields as
\[ q = \exp\left(-i\gamma_5 \sum_a \xi_a T^a\right) \bar{q}, \]
(B.19)
where $\xi_a$ 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.

$$\tilde{q}' = \exp\left( i \sum_a \theta_a T^a \right) \tilde{q}, \quad \text{(B.20)}$$

and this imposes on the Goldstone bosons the transformation rule

$$\exp\left[ i \sum_a \theta^V_a T^a + \theta^A_a T^a \gamma_5 \right] \exp\left( -i \gamma_5 \sum_a \xi_a T^a \right) = \exp\left( -i \gamma_5 \sum_a \xi'_a T^a \right) \exp\left( i \sum_a \theta_a T^a \right). \quad \text{(B.21)}$$

In terms of left and right moving angles, we find

$$\exp\left( i \sum_a \theta^L_a T^a \right) \exp\left( -i \sum_a \xi_a T^a \right) = \exp\left( -i \sum_a \xi'_a T^a \right) \exp\left( i \sum_a \theta_a T^a \right),$$
$$\exp\left( i \sum_a \theta^R_a T^a \right) \exp\left( i \sum_a \xi_a T^a \right) = \exp\left( i \sum_a \xi'_a T^a \right) \exp\left( i \sum_a \theta_a T^a \right), \quad \text{(B.22)}$$

where

$$\theta^L_a = \theta^V_a + \theta^A_a, \quad \theta^R_a = \theta^V_a - \theta^A_a. \quad \text{(B.23)}$$

This means that

$$U = \exp\left( -2i \sum_a \xi_a T^a \right) \quad \text{(B.24)}$$

transforms as

$$U' = \exp\left( i \sum_a \theta^L_a T^a \right) U \exp\left( -i \sum_a \theta^R_a T^a \right). \quad \text{(B.25)}$$

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

$$\mathcal{L} = \frac{F^2}{4} \text{Tr} \left[ \partial_\mu U \partial^\mu U \right] \quad \text{(B.26)}$$

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:

$$\mathcal{L}_{\text{QCD}}(\ell, r, s, p) = \frac{1}{N_f} \sum_{f=1}^{N_f} \left( \bar{q}_{L,f} D q_{L,f} + \bar{q}_{R,f} D q_{R,f} \right) - \bar{q}_L (s + ip) q_L - \bar{q}_R (s + ip) q_R - \bar{q}_R \gamma_\mu (s + ip) q_L - \bar{q}_L \gamma_\mu (s + ip) q_R $$
$$+ \bar{q}_R \gamma_\mu \frac{1 + \gamma_5}{2} \ell^\mu q_L - \bar{q}_L \gamma_\mu \frac{1 + \gamma_5}{2} \ell^\mu q_R, \quad \text{(B.27)}$$

145
where $\ell_\mu, r_\mu, s, p$ are $N_f \times N_f$ matrices. The standard QCD Lagrangian is obtained by setting

$$\ell_\mu = r_\mu = p = 0, \quad s = m$$

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

$$\langle \bar{q}q \rangle = -\frac{\delta \mathcal{L}_{\text{eff}}}{\delta s^0}$$

To see how the sources appear in the effective Lagrangian, we gauge the chiral symmetry and we promote the sources to gauge fields [38]. Under the transformations

$$q_L \to L(x)q_L, \quad q_R \to R(x)q_R$$

we have, as we just have seen,

$$U \to L(x)UR^\dagger(x).$$

The sources $\ell_\mu, r_\mu$ behave as gauge potentials for $L, R$, respectively, and they transform as

$$\ell_\mu \to L(x)\ell_\mu L^\dagger(x) + i(\partial_\mu L)(x)L^\dagger(x),$$

$$r_\mu \to R(x)r_\mu R^\dagger(x) + i(\partial_\mu R)(x)R^\dagger(x).$$

The Lagrangian (B.27) 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

$$D_\mu U = \partial_\mu U + i\ell_\mu U - iUr_\mu,$$

and transforms covariantly

$$D_\mu U \to L(x)D_\mu UR^\dagger(x).$$

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

$$\mathcal{L} = \frac{F^2}{4} \text{Tr}(D_\mu UD^\mu U^\dagger) + \frac{F^2}{4} \text{Tr}(\chi U^\dagger + U\chi),$$

where

$$\chi = 2B_0(s + ip)$$

and $B_0$ is a constant. If we now evaluate this for (B.28) we can read off the masses of the pions. First we write

$$2 \sum_a \xi_a T_a = \frac{\sqrt{2}}{F_\pi} B,$$
where

\[ \mathcal{B} = \begin{pmatrix} \frac{1}{\sqrt{2}} \pi^0 + \frac{1}{\sqrt{6}} \eta_8 & \pi^+ & K^+ \\ \pi^- & -\frac{1}{\sqrt{2}} \pi^0 + \frac{1}{\sqrt{6}} \eta_8 & K^0 \\ K^- & K^0 & -\sqrt{\frac{2}{3}} \eta_8 \end{pmatrix} \]  

(B.38)

With this representation, the kinetic term of the Lagrangian is canonically normalized

\[ \frac{F^2}{4} \text{Tr}(D_\mu U D^{\mu} U^\dagger) = \frac{1}{2} \partial_\mu \pi^0 \partial^\mu \pi^0 + \partial_\mu \pi^+ \partial^\mu \pi^- + \partial_\mu K^+ \partial^\mu K^- + \partial_\mu K^0 \partial^\mu \overline{K}^0 + \frac{1}{2} \partial_\mu \eta_8 \partial^\mu \eta_8. \]  

(B.39)

The quark mass matrix reads

\[ \mathcal{M} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{pmatrix} \]  

(B.40)

and

\[ U + U^\dagger = 2 - \frac{2}{F_\pi^2} \mathcal{B}^2 + \ldots \]  

(B.41)

The mass term in the effective Lagrangian is

\[ -B_0 \text{Tr}(\mathcal{B}^2 \mathcal{M}) = -B_0 \left\{ m_u \left( \left( \frac{1}{\sqrt{2}} \pi^0 + \frac{1}{\sqrt{6}} \eta_8 \right)^2 + \pi^+ \pi^- + K^+ K^- \right) \\
+m_d \left( \left( -\frac{1}{\sqrt{2}} \pi^0 + \frac{1}{\sqrt{6}} \eta_8 \right)^2 + \pi^+ \pi^- + \overline{K}^0 \overline{K}^0 \right) \\
+m_s \left( K^- K^+ + \overline{K}^0 \overline{K}^0 + \frac{2}{3} \eta_8^2 \right) \right\} \]  

(B.42)

and from here we can read the masses of the Goldstone bosons:

\[ m_{\pi}^2 = B_0 (m_u + m_d), \]
\[ m_{K^\pm}^2 = B_0 (m_u + m_s), \]
\[ m_{K^0}^2 = B_0 (m_d + m_s), \]
\[ m_{\eta_8}^2 = B_0 \frac{m_u + m_d + 4m_s}{3}. \]  

(B.43)

There is also a mixing term between the \( \eta_8 \) and the \( \pi^0 \),

\[ m_{\pi_0}^2 = B_0 \frac{m_u - m_d}{\sqrt{3}}. \]  

(B.44)
One particular prediction of chiral symmetry, following from (B.43), is that
\[ m_{\eta_8}^2 = \frac{1}{3}(2m_{K^\pm}^2 + 2m_{K^0}^2 - m_{\pi}^2) \] (B.45)
which is called the *Gell-Mann–Okubo mass formula*. Taking as data the masses of the kaons and the pions, it predicts
\[ m_{\eta_8}^2 = 566 \text{ MeV} \] (B.46)
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.29). We have
\[ -\frac{\partial \mathcal{L}}{\partial s_{ff'}^0} = -\frac{F_\pi^2}{4} \text{Tr}(\chi U^\dagger + U \chi^\dagger) = -\frac{F_\pi^2 B_0}{2}(U^\dagger_{ff'} + U_{ff'}) \] (B.47)
Evaluating this in the vacuum \( U = 1 \), we obtain
\[ \langle 0|\bar{q}_f q_f|0 \rangle = -F_\pi^2 B_0 \delta_{ff'} \] (B.48)
Expressing \( B_0 \) in terms of the quark condensate, we recover from (B.43) the relation (5.72).

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 \( B + 1 \sqrt{3} \eta_0 \).

Under the axial \( U(1) \), \( \zeta \) transforms as required,
\[ \zeta \rightarrow e^{i\theta^L} \eta_8 e^{-i\theta^R} = e^{2i\theta^A} \eta_0. \] (B.50)

The new mass term in the effective Lagrangian is
\[ -B_0 \left\{ m_u \left( \frac{1}{\sqrt{2}} \eta^0 + \frac{1}{\sqrt{6}} \eta_8 + \frac{1}{\sqrt{3}} \eta_0 \right)^2 + \pi^+ \pi^- + K^+ K^- \right\} \\
+ m_d \left( -\frac{1}{\sqrt{2}} \eta^0 + \frac{1}{\sqrt{6}} \eta_8 + \frac{1}{\sqrt{3}} \eta_0 \right)^2 + \pi^+ \pi^- + K^0 \bar{K}^0 \right\} \\
+ m_s \left( K^- K^+ + K^0 \bar{K}^0 + \left( \frac{1}{\sqrt{3}} \eta_0 - \sqrt{\frac{2}{3}} \eta_8 \right)^2 \right) \right\} \] (B.51)
This leads to a mixing matrix for the neutral mesons \( \pi^0, \eta_8, \eta_0 \) (listed in this order) given by
\[ B_0 \left( \begin{array}{ccc}
m_u + m_d & \frac{1}{\sqrt{3}}(m_u - m_d) & \sqrt{\frac{2}{3}}(m_u - m_d) \\
\frac{1}{\sqrt{3}}(m_u - m_d) & \frac{1}{3}(m_u + m_d + 4m_s) & \frac{\sqrt{2}}{3}(m_u + m_d - 2m_s) \\
\sqrt{\frac{2}{3}}(m_u - m_d) & \frac{\sqrt{2}}{3}(m_u + m_d - 2m_s) & \frac{2}{3}(m_u + m_d + m_s) \end{array} \right) \] (B.52)
Let us continue the analysis assuming for simplicity that \( m_u = m_d \). This eliminates all mixings but \( \eta_8 - \eta_0 \), which leads to a matrix

\[
\begin{pmatrix}
\frac{4}{3}m_K^2 - \frac{1}{3}m_\pi^2 & -\frac{2\sqrt{2}}{3}(m_K^2 - m_\pi^2) \\
-\frac{2\sqrt{2}}{3}(m_K^2 - m_\pi^2) & \frac{2}{3}m_K^2 + \frac{1}{3}m_\pi^2
\end{pmatrix}
\]

(B.53)

The eigenvalues of this matrix are

\[
m_\pi^2, \quad 2m_K^2 - m_\pi^2.
\]

(B.54)

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 \( \sqrt{3}m_\pi^2 \) [64].

Using however the Witten–Veneziano formula, we see that the above matrix gets an extra contribution due to the anomaly [62],

\[
\mathcal{M}_{\eta-\eta'} = \begin{pmatrix}
\frac{4}{3}m_K^2 - \frac{1}{3}m_\pi^2 & -\frac{2\sqrt{2}}{3}(m_K^2 - m_\pi^2) \\
-\frac{2\sqrt{2}}{3}(m_K^2 - m_\pi^2) & \frac{2}{3}m_K^2 + \frac{1}{3}m_\pi^2 + \epsilon
\end{pmatrix}
\]

(B.55)

where

\[
\frac{\epsilon}{N_c} = \frac{2N_f}{F_\pi^2} \chi_{ YM}.
\]

(B.56)

\section*{C Effective action for the \( \mathbb{P}^{N-1} \) sigma model}

Here we compute the large \( N \), effective propagators (7.58) and (7.60).

We start by calculating (7.58). Using the standard trick of introducing Feynman parameters (see for example [54], p. 189), we find

\[
\tilde{\Gamma}\lambda(p) = \int \frac{d^2q}{(2\pi)^2} \int_0^1 dx \frac{1}{x(m^2 + q^2 + (1 - x)((p + q)^2 + m^2))}
\]

\[
= \int \frac{d^2q}{(2\pi)^2} \int_0^1 dx \frac{1}{m^2 + q^2 + xp^2 + 2xp \cdot q}
\]

\[
= \int \frac{d^2\ell}{(2\pi)^2} \int_0^1 dx \frac{1}{m^2 + \ell^2 + x(1 - x)p^2}
\]

where we have introduced

\[
\ell = q + xp \Rightarrow q^2 + 2xp \cdot q = \ell^2 - x^2p^2.
\]

(C.1)
We end up with
\[ \tilde{\Gamma}^\lambda(p) = \int_0^1 dx \int \frac{d^2 \ell}{(2\pi)^2} \frac{1}{(\ell^2 + \Delta)^2}, \quad \Delta = m^2 + x(1-x)p^2. \] (C.3)

We now recall the standard formula in dimensional regularization (see for example [54], p. 250),
\[ \int \frac{d^d \ell}{(2\pi)^d (\ell^2 + \Delta)^2} = \frac{1}{(4\pi)^{d/2}} \Gamma(2 - d/2) \Delta^{d/2 - 2}. \] (C.4)
In our case \( d = 2 \) and the integral is convergent, and we simply find
\[ \tilde{\Gamma}^\lambda(p) = \int_0^1 dx \frac{1}{4\pi \Delta} = \frac{1}{4\pi} \int_0^1 \frac{dx}{m^2 + x(1-x)p^2}. \] (C.5)
This integral is elementary. We write the denominator as
\[ -p^2(x-a)(x-b), \] (C.6)
where
\[ a = \frac{1}{2} - \frac{1}{2} \sqrt{1 + \frac{4m^2}{p^2}}, \] (C.7)
\[ b = \frac{1}{2} + \frac{1}{2} \sqrt{1 + \frac{4m^2}{p^2}}. \]

The integral reads now
\[ \frac{1}{4\pi} \int_0^1 \frac{dx}{m^2 + x(1-x)p^2} = -\frac{1}{4\pi p^2} \frac{1}{a-b} \left[ \log(x-a) - \log(b-x) \right]_0^1 = \frac{1}{4\pi p^2} \frac{1}{b-a} \log \frac{b(1-a)}{a(1-b)} = \frac{1}{4\pi p^2} \frac{1}{b-a} \log \left( \frac{-b}{a} \right) \] (C.8)
where we used \( 1-a = b, 1-b = a \). Therefore, we find
\[ \tilde{\Gamma}^\lambda(p) = f(p) \equiv \frac{1}{2\pi \sqrt{p^2(p^2 + 4m^2)}} \log \frac{\sqrt{p^2 + 4m^2} - \sqrt{p^2}}{\sqrt{p^2 + 4m^2} + \sqrt{p^2}}. \] (C.9)

We now compute \( \tilde{\Gamma}_{\mu\nu}^A(p) \). Both integrals appearing in (7.60) 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,
\[
\int \frac{d^2 q}{(2\pi)^2} (p_\mu + 2q_\mu)(p_\nu + 2q_\nu) = \int_0^1 dx \int \frac{d^2 \ell}{(2\pi)^2} \frac{p_\mu p_\nu (1 - 4x + 4x^2) + 4\ell_\mu \ell_\nu}{[m^2 + \ell^2 + x(1-x)p^2]^2}
\]
\[
= p_\mu p_\nu \int_0^1 dx \int \frac{d^2 \ell}{(2\pi)^2} \frac{1 - 4x + 4x^2}{[m^2 + \ell^2 + x(1-x)p^2]^2}
+ \frac{4\delta_{\mu\nu}}{d} \int_0^1 dx \int \frac{d^2 \ell}{(2\pi)^2} \frac{\ell^2}{[m^2 + \ell^2 + x(1-x)p^2]^2},
\]
(C.10)

where in the first line we have set to zero the integrals over linear terms in \(\ell_\mu\), and in the last line we have set
\[
\ell_\mu \ell_\nu \rightarrow \frac{\delta_{\mu\nu}}{d} \ell^2
\]
(C.11)
since this is the only contribution to the integral (see again [54], p. ). In total, we find that (7.60) has two contributions. One has the tensorial structure of \(\delta_{\mu\nu}\), with coefficient
\[
2 \int_0^1 \frac{d^2 q}{(2\pi)^2} \frac{1}{(q^2 + m^2)} - \frac{4}{d} \int_0^1 dx \int \frac{d^2 \ell}{(2\pi)^2} \frac{\ell^2}{[m^2 + \ell^2 + x(1-x)p^2]^2},
\]
(C.12)

while the other one has the tensorial structure of \(p_\mu p_\nu\), and coefficient
\[
- \int_0^1 dx \int \frac{d^2 \ell}{(2\pi)^2} \frac{1 - 4x + 4x^2}{[m^2 + \ell^2 + x(1-x)p^2]^2}.
\]
(C.13)

Let us compute (C.12). Using dimensional regularization and the integral
\[
\int \frac{d^d \ell}{(2\pi)^d (\ell^2 + \Delta)^2} = \frac{1}{(4\pi)^d/2} \frac{d}{d} \Gamma(1 - d/2) \Delta^{d/2 - 1}.
\]
(C.14)

we obtain
\[
2 \int_0^1 dx \left[ \frac{1}{(4\pi)^{d/2}} \frac{\Gamma(1 - d/2)}{\Delta_1^{1-d/2}} - \frac{2}{d} \frac{1}{(4\pi)^{d/2}} \frac{d \Gamma(1 - d/2)}{\Delta_1^{1-d/2}} \right],
\]
(C.15)

where
\[
\Delta_1 = m^2, \quad \Delta_2 = m^2 + x(1-x)p^2.
\]
(C.16)

This can be written as
\[
\frac{2\Gamma(1 - d/2)}{(4\pi)^{d/2}} \int_0^1 dx \left[ \Delta_1^{d/2 - 1} - \Delta_2^{d/2 - 1} \right].
\]
(C.17)
We now expand around 
\[ d = 2 - \epsilon \Rightarrow d/2 - 1 = -\epsilon/2. \] (C.18)

Since 
\[ \Delta_1^{d/2-1} - \Delta_2^{d/2-1} = e^{-\epsilon \log \Delta_1/2} - e^{-\epsilon \log \Delta_2/2} = \frac{\epsilon}{2} \log \frac{\Delta_2}{\Delta_1} + \cdots, \] (C.19)

and 
\[ \Gamma(\epsilon/2) = \frac{2}{\epsilon} - \gamma + O(\epsilon), \] (C.20)

the total result is finite as \( \epsilon \to 0 \) and given by
\[ 2 \int_0^1 dx \log \left[ 1 + x(1 - x) \frac{p^2}{m^2} \right]. \] (C.21)

\[ \textbf{D} \quad \text{A simple example of analyticity at large } N \]

We have seen that various series that appear naturally in perturbation theory have a zero radius of convergence, and that one way to get a finite radius of convergence for perturbation theory (barring renormalons) is to consider simultaneously a \( 1/N \) expansion. The easiest way to see this is to consider the \( N \)-dimensional quartic integral
\[ I_N = \int \frac{d^N x}{(2\pi)^{N/2}} \exp \left[ -\frac{1}{2} x^2 - \frac{g}{N} x^4 \right], \] (D.1)

where here we understand
\[ x^k = \sum_{i=1}^N x_i^k. \] (D.2)

The large \( N \) behavior of this integral is analyzed in detail in [44]. After integrating over angular variables, this reads
\[ I_N = \frac{2^{1-N/2}}{\Gamma(N/2)} \int_0^\infty dx x^{N-1} \exp \left[ -\frac{1}{2} x^2 - \frac{g}{N} x^4 \right]. \] (D.3)

After changing variables
\[ x = N^{1/2} e^{t/2} \] (D.4)

we find
\[ I_N = \frac{2(N/2)^{N/2}}{\Gamma(N/2)} \int_{-\infty}^\infty dt e^{-NF(t)} \] (D.5)

where
\[ F(t) = \frac{1}{2} e^t + ge^{2t} - \frac{1}{2} t. \] (D.6)
The $1/N$ expansion can be now recovered by performing a saddle-point evaluation of the above integral. The saddle is obtained as

$$t_* = \log \frac{\sqrt{1 + 16g} - 1}{8g}. \quad (D.7)$$

We find,

$$I_N = \frac{2(N/2)^{N/2}}{\Gamma(N/2)} e^{-NF(t_*)} \left( \frac{2\pi}{NF''(t_*)} \right)^{\frac{1}{2}} \sum_{k=0}^{\infty} a_k(g) N^{-k}. \quad (D.8)$$

One has $a_0 = 1$ and

$$a_1(g) = \frac{2g \left( 13 + 128g + 11\sqrt{1 + 16g} \right)}{3 \left( 1 + 16g \right)^{3} \left( 1 + \sqrt{1 + 16g} \right)}. \quad (D.9)$$

We can expand $a_k(g)$ in power series in $g$ around $g = 0$, and we see that we now have functions with a finite radius of convergence. In this case, the first singularity appears at

$$g_c = -\frac{1}{16}. \quad (D.10)$$

References


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.


