11

Quantum Chromodynamics (QCD)

11.1 Definition of QCD

Quantum Chromodynamics is the theory of strong interactions. It is a non-Abelian gauge theory based on the gauge group $SU(3)$, which is called the colour gauge group. The gauge symmetry is preserved in this theory and, specifically, it is not spontaneously broken. The gauge bosons that carry the strong interaction are called gluons. The matter content of the theory consists of quarks, which are spin one-half fermions that transform according to the fundamental representation of $SU(3)$, that is a three-component, complex triplet. The quark model was proposed in the 1960s and 1970s and elaborated in its incorporation into the "standard model" of particle physics corresponding to a gauge-theoretic description of the strong, weak and electromagnetic interactions. This model is now at the level of a confirmed theory. An untold number of experimental data have shown the existence of quarks and gluons, in addition to the matter content corresponding to the non-strongly interacting particles, the leptons, and the corresponding gauge bosons of the weak and electromagnetic interactions, which are known as the W and Z gauge bosons, and the photon.

The strong interactions govern the interactions that give rise to nuclear forces. The matter that experiences these forces is generally called hadronic matter. The hadrons split into two categories: baryons, which correspond to the neutron, proton and atomic nuclei, which seem to be stable; and mesons, such as the pions, kaons and others, which all seem to be unstable. The fundamental building blocks of the hadrons are the quarks. The quarks interact directly with the gauge bosons of the colour $SU(3)$ gauge group, which are the gluons. The quarks have colour charges and couple directly to the gluons, which themselves have colour charges. However, it is believed that the QCD vacuum is such that colour charges are confined, that free colour charges correspond to states of infinite energy. Therefore, the observable hadrons must all be colour singlet states. The baryons

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

correspond to the bound states of three quarks, and a colour singlet in the three-fold tensor product of the fundamental representation $3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1$. The mesons correspond to bound states of quarks and anti-quarks, $3 \otimes \bar{3} = 8 \oplus 1$. There are many other possibilities for obtaining singlets, but these have not been experimentally observed.

11.1.1 The Quark Model and Chiral Symmetry

In the 1960s the quark model of hadrons was invented, with contributions from many different authors coming independently. It was understood that quarks come in many flavours, and these were named up, down, charm, strange, top, bottom, and more, if necessary. In our daily experience, we only encounter the up and down quarks. During the 1960s and 1970s, it was discovered how the quarks fit together to give rise to the observable hadrons, and also their interactions with the non-hadronic particles called generically leptons, the electron, muon, and taon, their neutrinos. The quark model seemed to indicate the existence of families of elementary particles, which bring together the strong, weak and electromagnetic interaction with gauge group $SU_c(3) \times SU(2) \times U(1)$, the gauge group of the standard model. Models of grand unification correspond to the inclusion of this group inside a single, semi-simple group, with symmetry breaking giving rise to the observed symmetry group of the standard model. The $SU_c(3)$ is the colour gauge group of QCD. The weak interactions are mediated by the $SU(2)$, while the $U(1)$ corresponds to what is called weak hypercharge. The weak $SU(2)$ is spontaneously broken to a $U(1)$ subgroup, the by now celebrated Higgs field and Higgs mechanism, and the actual electromagnetic $U(1)$ gauge group corresponds to a linear combination of this unbroken remnant of the weak $SU(2)$ and the $U(1)$ hypercharge gauge symmetry. We will not elaborate the full standard model here, it is out of our interest and there are many very good references that describe the standard model in all its detail. For us it will suffice to know that the left-handed quark fields and the leptons feel the weak interaction, which only acts on left-handed fields, and transform according to the doublet representation of the weak interaction gauge symmetry. All right-handed fields, quark or lepton, do not feel the weak interaction, and only feel the strong and electromagnetic interaction.

The first family comprises left-handed up and down quarks forming a doublet of the weak interactions based on the group $SU(2)$ and transforming individually according to a $U(1)$ charge called weak hypercharge, along with the left-handed electron and its neutrino, which also form a weak doublet with their respective weak hypercharges. The family is completed with the right-handed partners of the up and down quarks and the right-handed partner of the electron. The neutrino was not supposed to have a right-handed partner; however, this is no longer certain as it has been observed that the neutrinos must have mass. For the purposes of this book, we will not add a right-handed neutrino. The right-handed partners of all the particles did not experience the weak interaction but did experience the weak hypercharge, and each member had a corresponding value for the weak hypercharge. The second family comprises the charm and strange quarks and the muon and its neutrino; and the third family comprises the top and bottom quarks and the taon and its neutrino. Chiral symmetry corresponds to the notion that there is a complete symmetry under unitary rotation of the quarks amongst themselves. In principle, this would correspond to a "flavour" symmetry group of $SU_f(6)$.

Chiral symmetry is, explicitly, badly broken by the mass spectrum of the quarks. The best preserved subgroup is chiral $SU(2)$ (which is also, coincidentally, the weak interaction symmetry) corresponding to iso-rotations of the up and down quarks amongst themselves as these quarks have masses in the range of a few MeV, which is almost negligible at the scale of the strong interactions. Including the next lightest quark, the strange quarks gives rise to chiral $SU(3)$ symmetry, which is broken at a 10% level as the strange quark mass is around 100 MeV. This symmetry was named $SU_f(3)$, the three-dimensional unitary symmetry of flavour. Identification of this symmetry led to a great advance in the organization of the hadronic particle spectrum. This meant that the Lagrangian of the quarks was made up of three fermionic fields and it is invariant under the unitary rotation of the three fields into each other. The energy eigenstates then must form representations of this group of symmetry, much like the energy levels of the hydrogen atom form representations of the group of spatial rotations, $SO(3)$. Even though the $SU_f(3)$ is broken at the 10% level, the physical hadrons, which are the energy eigenstates of the theory, are easily identifiable as being members of various representations of this symmetry group. The baryons form the representations 8 and 10 of $SU_f(3)$, while the mesons fall into the 8, as shown in Figures 11.1, 11.2 and 11.3, which were created by [84, 83, 82].

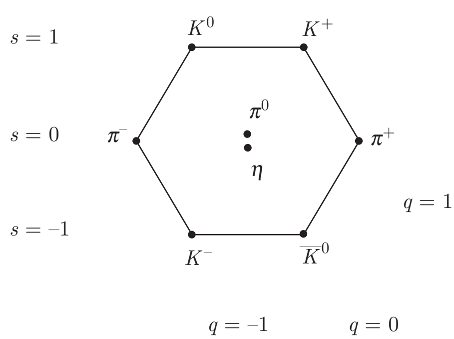

Figure 11.1. QCD flavour diagram of the meson octet

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

$SU_f(3)$, being a symmetry of the theory, cannot be responsible for the strong force between the hadrons. The strong force must be independent of the flavour symmetry, for the flavour symmetry to manifest itself as a symmetry of the mass spectrum. The charm quark mass is about 1.2 GeV and the top and bottom masses are over 150 GeV, hence invoking chiral symmetry including these quarks is quite unrealistic. But what was holding the quarks together?

11.1.2 Problems with Chiral Symmetry

1. Chiral $SU(3)$ symmetry implies the existence of multiplets of hadronic particle states, which have all been observed, and brings order to the chaos of the zoo of observed hadronic particles. However, there is a problem, as chiral symmetry predicts hadronic states such as the Δ^{++} which is made of three up quarks or the Δ^- the corresponding states of three down quarks or the Ω^- that of three strange quarks, each of them in a spin 3/2 state. The problem has to do with their wave functions. The three quarks should be in a spatially symmetric state as there is no additional angular momentum, a spin-symmetric state

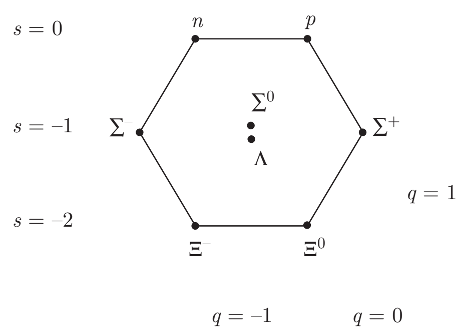

Figure 11.2. QCD flavour diagram of the baryon octet

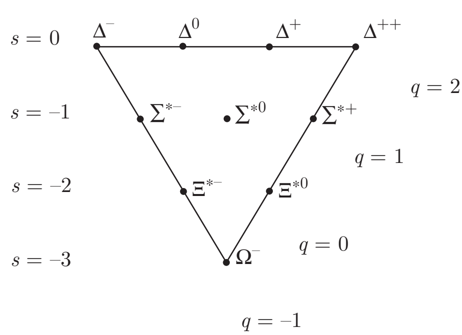

Figure 11.3. QCD flavour diagram of the baryon decouplet

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

giving rise to the spin 3/2 of the state and an iso-spin symmetric state as the iso-spin of each quark is identical. Such a state is not permitted for fermions by the Pauli exclusion principle, which requires that the wave function of identical fermions must be anti-symmetric under the exchange of any two. Therefore the quarks must have another, hidden quantum number, and the wave function of the state must be anti-symmetric under this hidden degree of freedom.

2. There exists a second experimental reason why the quarks should come in three colours. The ratio

$$R = \frac{\sigma(e^+e^- \to q\bar{q})}{\sigma(e^+e^- \to \mu^+\mu^-)} = \sum_i Q_i^2 \tag{11.1}$$

is simply predicted by perturbation theory, where Q_i is the electrical charge of the quark. This ratio is measured experimentally and gives a rising function of the incoming energy, with a few isolated peaks corresponding to resonances at the positions of particles. However, it reaches a first plateau with a numerical value of 2 when it crosses the threshold for production of the strange quark. Now the sum over the charges of the lightest quarks, up, down and strange, which are, respectively, $\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}$, is given by

$$\sum_{\text{lightest quarks}} Q_i^2 = \left(\frac{2}{3}\right)^2 + \left(-\frac{1}{3}\right)^2 + \left(-\frac{1}{3}\right)^2 = \frac{2}{3}. \tag{11.2}$$

Clearly if each quark came three times with three colours we get the required value 2. Increasing the energy of the scattering, once we pass the charm threshold at about 1.2 GeV, the value of R increases to a second plateau at $3\frac{1}{3}$. This corresponds exactly to the addition of the charge of the charm quark squared, $\left(\frac{2}{3}\right)^2 \times 3$. Finally after crossing the bottom quark threshold at an energy of about 4.2 GeV, the value of R again increases to a plateau at $3\frac{2}{3}$ corresponding to the charge of the bottom quark, appropriately $\left(\frac{1}{3}\right)^2 \times 3$.

3. Another experimental reason for three colours has to do with the decay rate of the neutral pion to two photons, $\pi^0 \to 2\gamma$. This decay is mediated by the so-called anomaly diagram. The amplitude for the decay predicted if only one quark is circulating in the triangle is exactly three times too small from the observed amplitude.

4. "Anomaly cancellation" gives another reason to believe that there must be three colours. As mentioned, part of the flavour symmetry group is actually also gauged and gives rise to the weak and electromagnetic interaction. Sometimes gauge symmetries are broken by quantization of chiral fermions. A gauged symmetry must be respected at the quantum level; it is necessary to prove the renormalizability of the theory. Invariance under gauge transformations for the quantum theory is used in an essential manner to prove renormalizability. Therefore, it is imperative that the weak and

electromagnetic gauge symmetries are anomaly-free. The anomalies of the corresponding gauge group all potentially reside in the weak hypercharge $U(1)$ symmetry. The weak hypercharge of the left-handed up and down quarks is $\frac{1}{3}$ while that of the right-handed up quark is $\frac{4}{3}$ and that of the right-handed down quark is $-\frac{2}{3}$. The left-handed leptons, the electron and its neutrino, have weak hypercharge -1 while the right-handed electron has weak hypercharge -2, and we are assuming that the right-handed neutrino does not exist. The anomaly is proportional to the sum of the cubes of all the left-handed hypercharges minus the same for the right-handed charges. We must not forget that the quarks each come in three colours, giving an additional factor of three, and then this gives

$$3\times\left(\frac{1}{3}\right)^3+3\times\left(\frac{1}{3}\right)^3+(-1)^3+(-1)^3-\left(3\times\left(\frac{4}{3}\right)^3+3\times\left(-\frac{2}{3}\right)^3+(-2)^3\right)$$
$$=\left(\frac{1}{9}\right)+\left(\frac{1}{9}\right)-2-\left(\frac{64}{9}\right)+\left(\frac{8}{9}\right)+8=0. \tag{11.3}$$

5. Finally, there has to be some mechanism by which the colour degree of freedom is not seen in hadronic states, and has to be confined. There is a good theoretical indication why a non-Abelian gauge theory could supply the correct interaction. First of all, the colour degree of freedom is flavour-blind, it is identical for each flavour. However, QCD being a renormalizable theory, we can perturbatively calculate the renormalization of the coupling constant. Non-trivial renormalization means that naive calculations of, say, the perturbative corrections to the coupling constant give infinite answers. However, by scaling the bare coupling constants of the theory appropriately, all the infinities can be absorbed into these inobservable, infinite, bare coupling constants, while the physically observed coupling constants are finite and defined at a chosen energy scale. However, then the value of the coupling constant at different energy scales is predicted by finite scaling, which is called the renormalization group. Perturbative calculations indicate that as the energy scale is increased the value of the coupling constant decreases (rendering, in fact, the perturbative calculations, which are valid for a small coupling constant, more and more precise). Evidently for lower and lower energies the coupling constant must increase. These properties are called asymptotic freedom at high energies and infrared slavery at low energies. Of course, the perturbative calculation becomes less and less reliable as the coupling constant increases, and hence actually only indications of infrared slavery are predictable via the perturbation theory. Nevertheless, the picture for quarks emerges, that when they are close together, at short distances which correspond to high energies, they are essentially free and non-interacting. However, as they try to separate from one another, at long distances, the force between them increases and, in principle, it would require infinite energy to

separate them infinitely far. The theoretical prediction of asymptotic freedom has been observed experimentally. When very high electrons impinge on a hadronic target and suffer deep inelastic scattering, they scatter off the individual quarks, which, because of the high energy of the electrons, are being probed at very short distances. The quarks then should behave as free particles. This is exactly what is observed. The deep inelastic scattering cross-section for electrons on hadrons exhibits the property of scaling, that the cross-section is simply that of an electron scattering from a free quark of momentum $x \times p_H$, where p_H is the total momentum of the hadron and x is the fraction of that momentum carried by the quark, multiplied by a factor that corresponds to the probability of finding a quark with momentum fraction x.

Thus the colour degree of freedom arose, and making it a local gauge degree of freedom gave the added bonus that it provided a means for obtaining interactions between the quarks that would in principle bind them together.

11.1.3 The Lagrangian of QCD

The Lagrangian density of N flavours of free quarks is given by

$$\mathcal{L} = \sum_{a=1}^{N} \bar{\psi}^a_\alpha \left(i\gamma^\mu \partial_\mu - m^a\right) \psi^a_\alpha. \tag{11.4}$$

The label a corresponds to the different flavours, while the label α corresponds to the colour and the summation over repeated colour, flavour and Lorentz indices is assumed.[1] Interaction terms involving just the fields ψ_α themselves, such as $\left(\bar{\psi}^a_\alpha \psi^a_\alpha\right)^2$ or $\left(\bar{\psi}^a_\alpha \gamma^\mu \psi^a_\alpha\right)\left(\bar{\psi}^b_\beta \gamma_\mu \psi^b_\beta\right)$ and any others, are not renormalizable. To have interactions between the quarks, we must add other fields such as gauge fields or scalar fields with which the quarks interact, and then with each other through the exchange of the additional particles. We will consider the idea of gauging the added $SU(3)$ colour symmetry, the symmetry in any case seems to be required for the existence of fermionic statistics of the quarks in some of the hadronic states.

The colour degree of freedom corresponds to the index α, which goes from 1 to 3, and we will now add gauge fields corresponding to making the gauge symmetry $SU(3)$ local,

$$\mathcal{L} = \sum_{i=a}^{N} \bar{\psi}^a_\alpha \left(i\gamma^\mu (\partial_\mu + A_\mu) - m^a\right) \psi^a_\alpha. \tag{11.5}$$

[1] The colour metric or the flavour metric are both simply the identity matrix so we will write the indices above or below depending on convenience. The Lorentz indices are summed with the Minkowski metric, thus for these we will rigorously only sum a raised index with a repeated lowered index.

The covariant derivative $D_\mu = \partial_\mu + A_\mu$ now appears with $A_\mu = iA^i_\mu \lambda_i$, $i = 1, \cdots, 8$, and the λ_i correspond to the 3×3 Gell-Mann matrices

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \ \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \ \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$
$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \ \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{11.6}$$

The Gell-Mann matrices satisfy the Lie algebra of $SU(3)$,

$$[\lambda_i, \lambda_j] = i f^{ijk} \lambda_k, \tag{11.7}$$

where f^{ijk} are the structure constants of $SU(3)$. The structure constants are completely anti-symmetric, $f^{ijk} = -f^{jik} = -f^{ikj}$ with $f^{123} = 1, f^{147} = -f^{156} = f^{246} = f^{257} = f^{345} = -f^{367} = 1/2, f^{458} = f^{678} = \sqrt{3}/2$. To this action, we add the Lagrangian for the gauge fields

$$\mathcal{L}_{gauge} = -\frac{1}{4g^2} F^i_{\mu\nu} F^{i\mu\nu}, \tag{11.8}$$

where, as previously defined, $F^i_{\mu\nu}$ is obtained from

$$[D_\mu, D_\nu] = i F^i_{\mu\nu} \lambda_i, \tag{11.9}$$

explicitly

$$F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\mu A^i_\nu - f^{ijk} A^j_\mu A^k_\nu. \tag{11.10}$$

Our aim in this book is to consider the importance of the classical solutions to the Euclidean equations of motion, the instantons. Thus we will write the Euclidean Lagrangian density as

$$\mathcal{L}_E = \frac{1}{4} F^i_{\mu\nu} F^i_{\mu\nu} = -\frac{1}{2} Tr\left(F_{\mu\nu} F_{\mu\nu}\right), \tag{11.11}$$

where now the Lorentz index becomes a Euclidean vectorial index and the metric in Euclidean space is just the identity, hence we change the sign in the first equality, and

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu], \tag{11.12}$$

which is an anti-hermitean matrix-valued field.

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

11.2 Topology of the Gauge Fields

We shall look for configurations of finite Euclidean action

$$S_E = \int d^4x \mathcal{L}_E. \tag{11.13}$$

We assume that for large radius r in four-dimensional Euclidean space, the gauge fields can be expanded in powers of $1/r$. For finite action then, $F_{\mu\nu}$ must decrease as $o(1/r^2)$, where $o(1/r^2)$ means faster than $1/r^2$. This implies that the gauge field must decrease at least as $o(1/r)$ up to pure gauge terms

$$A_\mu = o\left(\frac{1}{r}\right) + g(\Omega)\partial_\mu g^{-1}(\Omega), \tag{11.14}$$

where $g(\Omega)$ is a function only of the angular variables Ω at infinity. Then $g(\Omega)\partial_\mu g^{-1}(\Omega) \sim 1/r$, and this yields the required behaviour for $F_{\mu\nu}$.

But $g(\Omega)$ is defined essentially at infinity of Euclidean spacetime, $\mathbf{R}^4$, which is topologically the three-sphere S^3. Thus $g(\Omega)$ defines a mapping of the three-sphere at infinity into the gauge group $SU(3)$,

$$g(\Omega) : S^3 \to SU(3). \tag{11.15}$$

These fall into the homotopy classes of mappings which define the homotopy group $\Pi_3(SU(3))$. Gauge group configurations $g_1(\Omega)$ and $g_2(\Omega)$ can be continuously deformed one into the other only if they fall into the same homotopy class. We write $g_1(\Omega) \sim g_2(\Omega)$ if they are in the same homotopy class. The homotopy group is well known,

$$\Pi_3(SU(3)) = \mathbb{Z}, \tag{11.16}$$

where $\mathbb{Z}$ corresponds to the integers, and an integer corresponding to a homotopy class is called the winding number. This means that each configuration can be associated with a class of homotopically equivalent configurations, which have the same winding number. Configurations with different winding numbers cannot be continuously deformed one into another, since the winding number can only change discretely. Continuous changes cannot change the winding number. Consequently, different gauge field configurations of finite Euclidean action must also fall into topologically distinct homotopy classes. A gauge field configuration $A_1(x)$ with a limiting value defined by the asymptotic gauge group configuration $g_1(\Omega)$ cannot be continuously deformed into another gauge field configuration $A_2(x)$ with a limiting value defined by the asymptotic gauge group configuration $g_2(\Omega)$ unless $g_1(\Omega) \sim g_2(\Omega)$. If the asymptotic gauge group configurations are in different homotopy classes, the existence of a deformation of the gauge fields into each other continuously keeping the Euclidean action finite would be a contradiction, as it would provide a deformation of one asymptotic gauge group configuration into the other.

We might imagine that as the theory is invariant under local gauge transformations, we might be able to remove the asymptotic gauge dependence. Suppose we make a gauge transformation at infinity,

$$g \to hg \tag{11.17}$$

for some group element h. Then the gauge field transforms as

$$\begin{aligned} A_\mu &\to h(A_\mu + \partial_\mu)h^{-1} \\ &= h\left(g\partial_\mu g^{-1} + o(1/r) + \partial_\mu\right)h^{-1} \\ &= hg(\partial_\mu g^{-1})h^{-1} + h\partial_\mu h^{-1} + o(1/r) \\ &= hg(\partial_\mu (hg)^{-1}) + o(1/r). \end{aligned} \tag{11.18}$$

Thus if we chose $h = g^{-1}$, we could eliminate g. But this is impossible because the gauge transformation h should be a differentiable function defined over the whole space $\mathbf{R}^4$. At least h should be a continuous function over all of $\mathbf{R}^4$. Thus if we define $h = g^{-1}$ at infinity, we must be capable of continuing the definition of h throughout space, including the origin. This is clearly impossible since the origin is a degenerate sphere on which the mapping must be trivial. This implies that the gauge transformation h must be in a class of gauge transformations that can be continuously deformed to the trivial mapping. Hence h cannot satisfy $h = g^{-1}$ at infinity, as g is not in the class of trivial mappings. Thus any gauge transformation h can modify g at infinity, but only within its homotopy class, $g \to hg \sim g$. The integer invariant corresponding to the homotopy class of g is seen to be exactly the Chern number of the gauge field configuration.

We can explicitly construct the gauge transformations that give rise to the different classes of gauge fields

$$\begin{aligned} g^{(0)}(x) &= 1 \\ g^{(1)}(x) &= \frac{x_4 + i\vec{x}\cdot\vec{\sigma}}{(x_4 + |\vec{x}|^2)^{1/2}} \\ &\cdot \\ &\cdot \\ &\cdot \\ g^{(\nu)}(x) &= \left(g^{(1)}\right)^{\nu} \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned} \tag{11.19}$$

defined over each S^3 that contains the origin. The gauge transformations are singular at the origin (except $g^{(0)}$).

11.2.1 Topological Winding Number

We can explicitly calculate the winding number of the gauge field configuration through the following analysis. Consider the integral

$$\nu = \frac{-1}{24\pi^2}\int d^3\theta\epsilon^{ijk}Tr\left((g\partial_i g^{-1})(g\partial_j g^{-1})(g\partial_k g^{-1})\right), \tag{11.20}$$

where the integral is over a three-sphere with local coordinates θ_i. For any local, infinitesimal transformation of g, $g \to g(1+\delta T)$ and $g^{-1} \to (1-\delta T)g^{-1}$ so that $gg^{-1} = 1$ is unchanged. This means that with $\delta g = g\delta T$ and $\delta g^{-1} = -\delta T g^{-1}$ we will show that ν is unchanged. We then find

$$\begin{aligned}\delta(g\partial_k g^{-1}) &= g\delta T\partial_k g^{-1} - g\partial_k(\delta T g^{-1})\\ &= g\delta T\partial_k g^{-1} - g(\partial_k\delta T)g^{-1} - g\delta T\partial_k g^{-1}\\ &= -g(\partial_k\delta T)g^{-1}.\end{aligned} \tag{11.21}$$

Thus the change in ν is

$$\begin{aligned}\delta\nu &= \frac{1}{24\pi^2}\int d^3x\epsilon^{ijk}Tr\left(g\partial_i\delta T)g^{-1}(g\partial_j g^{-1})(g\partial_k g^{-1})\right)\times 3\\ &= \frac{1}{8\pi^2}\int d^3\theta\epsilon^{ijk}Tr\left(\partial_i\delta T)(\partial_j g^{-1})g(\partial_k g^{-1})g\right)\\ &= \frac{-1}{8\pi^2}\int d^3\theta\epsilon^{ijk}Tr\left(\partial_i\delta T)(\partial_j g^{-1})(\partial_k g)\right)\\ &= \frac{-1}{8\pi^2}\int d^3\theta\epsilon^{ijk}\partial_i Tr\left(\delta T)(\partial_j g^{-1})(\partial_k g)\right) = 0,\end{aligned} \tag{11.22}$$

where in the first line the factor of 3 comes because the contribution from each of the three factors is the same, in the third line we use $g(\partial_k g^{-1})g = -\partial_k g$ and the last line vanishes as the integral is of a total derivate over a three-sphere that has no boundary.

We can evaluate ν explicitly for $g^{(1)}$. At the "north pole", $x^4 = 1, x^i \approx 0$ then we can take $\theta^i = x^i$

$$\begin{aligned}g^{(1)}\partial_i\left(g^{(1)}\right)^{-1}\Big|_{north\ pole} &= \left(\frac{x^4 + i\vec{x}\cdot\vec{\sigma}}{(x^4+|\vec{x}|^2)^{1/2}}\right)\partial_i\left(\frac{x^4 - i\vec{x}\cdot\vec{\sigma}}{(x^4+|\vec{x}|^2)^{1/2}}\right)\Big|_{x^4=1,x^i=0}\\ &= -i\sigma_i - \left(\frac{(x^4 - i\vec{x}\cdot\vec{\sigma})x^i}{(x^4+|\vec{x}|^2)^{3/2}}\right)\Big|_{x^4=1,x^i=0} = -i\sigma_i.\end{aligned} \tag{11.23}$$

However, the symmetry of the configuration means that the integrand is the same at all points on the sphere. Hence,

$$\begin{aligned}\epsilon^{ijk}Tr\left((g\partial_i g^{-1})(g\partial_j g^{-1})(g\partial_k g^{-1})\right) &= i\epsilon^{ijk}Tr\left(\sigma_i\sigma_j\sigma_k\right)\\ &= i\epsilon^{ijk}Tr\left(i\epsilon_{ijl}\sigma_l\sigma_k\right)\\ &= -\epsilon^{ijk}\epsilon_{ijl}2\delta_{lk} = -2\cdot 6 = -12\end{aligned} \tag{11.24}$$

using $\sigma_i \sigma_j = i\epsilon^{ijk}\sigma_k + \delta_{ij}$. Thus

$$\nu = -\frac{1}{24\pi^2}(-12)\int d^3\theta = \frac{1}{2\pi^2}\int d^3\theta = 1, \tag{11.25}$$

since the volume of the unit three-sphere is exactly $2\pi^2$. This is obtainable by integrating over the angular variables in $\mathbf{R}^4$ in the generalization of spherical coordinates. It is easy to see the $\nu(g_1 g_2) = \nu(g_1) + \nu(g_2)$, indeed, using form notation

$$\begin{aligned}
\nu(g_1 g_2) &= \frac{-1}{24\pi^2}\int Tr\left(g_1 g_2 d(g_1 g_2)^{-1}\right)^3 \\
&= \frac{-1}{24\pi^2}\int Tr\left(g_1 g_2 (dg_2^{-1})g_1^{-1} + g_1 dg_1^{-1}\right)^3 \\
&= \nu(g_1) + \nu(g_2) + 3\frac{-1}{24\pi^2} \\
&\quad \times \int Tr\left(g_1 g_2 (dg_2^{-1})g_1^{-1} g_1 (dg_1^{-1})(g_1 g_2 (dg_2^{-1})g_1^{-1} + g_1 dg_1^{-1})\right) \\
&= \nu(g_1) + \nu(g_2) + \frac{-1}{8\pi^2} \\
&\quad \times \int Tr\left(g_2 (dg_2^{-1})(dg_1^{-1})g_1 g_2 (dg_2^{-1}) + g_2 (dg_2^{-1})(dg_1^{-1})g_1 (dg_1^{-1})g_1)\right) \\
&= \nu(g_1) + \nu(g_2) + \frac{-1}{8\pi^2}\int d\left(Tr\left(g_2 (dg_2^{-1})(dg_1^{-1})g_1\right)\right) \\
&= \nu(g_1) + \nu(g_2)
\end{aligned} \tag{11.26}$$

where we have used $d(gd(g^{-1})) = -gd(g^{-1})gd(g^{-1})$.

We can define

$$G_\mu = 4\epsilon_{\mu\nu\lambda\sigma} Tr\left(A_\nu \partial_\lambda A_\sigma + \frac{2}{3}A_\nu A_\lambda A_\sigma\right) \tag{11.27}$$

then

$$\begin{aligned}
\partial_\mu G_\mu &= 4\epsilon_{\mu\nu\lambda\sigma} Tr\left(\partial_\mu A_\nu \partial_\lambda A_\sigma + \frac{2}{3}\left(\partial_\mu A_\nu A_\lambda A_\sigma + A_\nu \partial_\mu A_\lambda A_\sigma + A_\nu A_\lambda \partial_\mu A_\sigma\right)\right) \\
&= 4\epsilon_{\mu\nu\lambda\sigma} Tr\left(\partial_\mu A_\nu \partial_\lambda A_\sigma + 2\left(\partial_\mu A_\nu A_\lambda A_\sigma\right)\right) \\
&= 4\epsilon_{\mu\nu\lambda\sigma} Tr\left((\partial_\mu A_\nu + A_\mu A_\nu)(\partial_\lambda A_\sigma + A_\lambda A_\sigma)\right) \\
&= \epsilon_{\mu\nu\lambda\sigma} Tr\left((\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu])(\partial_\lambda A_\sigma - \partial_\sigma A_\lambda + [A_\lambda, A_\sigma])\right) \\
&= \epsilon_{\mu\nu\lambda\sigma} Tr\left(F_{\mu\nu} F_{\lambda\sigma}\right).
\end{aligned} \tag{11.28}$$

But

$$\int d^4x \partial_\mu G_\mu = \oint_{r\to\infty} dS_\mu G_\mu = \oint_{r\to\infty} dS_\mu 4\epsilon_{\mu\nu\lambda\sigma} Tr\left(A_\nu F_{\lambda\sigma} - \frac{1}{3}A_\nu A_\lambda A_\sigma\right) \tag{11.29}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

and since

$$F_{\lambda\sigma} = o\left(\frac{1}{r^2}\right) \Rightarrow A_\nu = g\partial_\nu(g^{-1)} + o\left(\frac{1}{r}\right), \tag{11.30}$$

then the first term in Equation (11.29) falls off too fast to contribute while the second term gives exactly the expression for ν in Equation (11.20). Hence

$$\begin{aligned}\nu &= \frac{1}{32\pi^2}\int d^4xTr\left(\epsilon_{\mu\nu\lambda\sigma}F_{\mu\nu}F_{\lambda\sigma}\right)\\ &= \frac{1}{16\pi^2}\int d^4xTr\left(F_{\mu\nu}\tilde{F}_{\mu\nu}\right)\end{aligned} \tag{11.31}$$

where the dual field strength is defined as $\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\lambda\sigma}F_{\lambda\sigma}$.

We can summarize our findings as follows.

1. Each gauge field configuration of finite Euclidean action is associated with an integer, called its Pontryagin number.
2. It is impossible to continuously deform one gauge field configuration into another with different Pontryagin numbers, keeping the Euclidean action finite.

For any other gauge group, $SU(3)$ in particular, there is a theorem by Bott [18] that says that any mapping of S^3 into a semi-simple Lie group G can be continuously deformed to a mapping into a $SU(2)$ subgroup of G. Hence everything that we have shown for $SU(2)$ is actually valid for any semi-simple Lie group G. The only thing that changes is the normalization in the formulae for the winding number. However, if we use the notion of the Cartan scalar product on the Lie algebra of G, defining

$$\langle T^a T^b\rangle = \delta^{ab} = \alpha Tr\left(T^a T^b\right) \tag{11.32}$$

then α depends on the representation of the T^a, but the formula for the Pontryagin number is universal

$$\nu = \frac{1}{32\pi^2}\int d^4x\langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle. \tag{11.33}$$

Now with the possibility of many inequivalent classical sectors in the space of field configurations, we expect the existence of the many different vacuum configurations, and of course the possibility of quantum tunnelling between them.

11.3 The Yang–Mills Functional Integral

We begin with the functional integral

$$\mathcal{I} = \mathcal{N}\int \mathcal{D}Ae^{\int d^4x\frac{1}{4g^2}\langle F_{\mu\nu}F_{\mu\nu}\rangle}. \tag{11.34}$$

We must fix the gauge, we will choose $A_3 = 0$. We then have the following observations:

1. It is easy to see that all gauge field configurations may be put into this gauge, simply take
$$h = \mathcal{P}\left(\exp\int_{-\infty}^{x_3} dx'^3 A_3(x^1, x^2, x'^3, x^4)\right). \tag{11.35}$$
Then
$$h\partial_3 h^{-1} = -hA_3 h^{-1} \tag{11.36}$$
hence
$$A'_3 = h(A_3 + \partial_3)h^{-1} = hA_3h^{-1} - hA_3h^{-1} = 0. \tag{11.37}$$
2. The Faddeev–Popov factor is just a constant.

11.3.1 Finite Action Gauge Fields in a Box

We will consider the theory in a finite spatial volume V, but always have in mind that $V \to \infty$ at the end. The same for the Euclidean time T. We must choose boundary conditions on the walls. We will choose the boundary conditions such that the bulk equations of motion are not modified because of them. The general variation of the action is

$$\begin{aligned}
\delta S &= \int d^4x \frac{\partial\mathcal{L}}{\partial A_\mu}\delta A_\mu + \frac{\partial\mathcal{L}}{\partial\partial_\nu A_\mu}\delta\partial_\nu A_\mu \\
&= \int d^4x \left(\frac{\partial\mathcal{L}}{\partial A_\mu} - \frac{\partial\mathcal{L}}{\partial\partial_\nu A_\mu}\right)\delta A_\mu + \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial\partial_\nu A_\mu}\delta A_\mu\right) \\
&= \int d^3s\, \hat{n}^\nu \frac{\partial\mathcal{L}}{\partial\partial_\nu A_\mu}\delta A_\mu + \int d^4x \left(\frac{\partial\mathcal{L}}{\partial A_\mu} - \frac{\partial\mathcal{L}}{\partial\partial_\nu A_\mu}\right)\delta A_\mu \\
&= \int d^3s\, \hat{n}^\nu F_{\nu\mu}\delta A_\mu + \int d^4x \left(\frac{\partial\mathcal{L}}{\partial A_\mu} - \frac{\partial\mathcal{L}}{\partial\partial_\nu A_\mu}\right)\delta A_\mu.
\end{aligned} \tag{11.38}$$

Therefore, to not have any contribution from the boundary we must impose

$$\hat{n}^\nu F_{\nu\mu}\delta A_\mu = 0 \tag{11.39}$$

on the boundary. We can decompose δA_μ into its normal and tangential components, $\delta A_\mu = (\delta A^{norm.})\hat{n}_\mu + \delta A_\mu^{tang.}$, where $\hat{n}_\mu \delta A_\mu^{tang.} = 0$. Then the boundary condition Equation (11.39) becomes

$$\hat{n}^\nu F_{\nu\mu}\left((\delta A^{norm.})\hat{n}_\mu + \delta A_\mu^{tang.}\right) = \hat{n}^\nu F_{\nu\mu}\delta A_\mu^{tang.} = 0, \tag{11.40}$$

since $F_{\nu\mu} = -F_{\mu\nu}$. Thus we are required to fix the tangential components of the gauge field on the boundary and, consequently, we impose that the tangential components may not be varied on the boundary, so that $\delta A_\mu^{tang.} = 0$. We must also respect the gauge-fixing condition, $A_3 = 0$, and we are only interested in field configurations whose action remains finite as the box size is taken to infinity. We

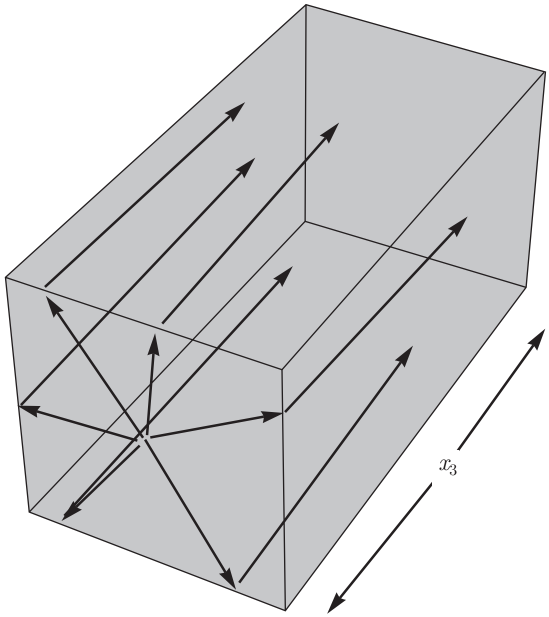

Figure 11.4. Paths over the boundary defining the gauge group element

will see that these conditions mean that the winding number inside the box must be a definite integer. We will show that the only vestige of the boundary conditions is that the winding number inside the box is a definite integer.

Indeed, inside a large box of dimensions (L_1, L_2, L_3, T), gauge fields that remain of finite action when the box is taken to infinite size must have the behaviour

$$A_\mu = g\partial_\mu g^{-1} + o\left(\frac{1}{r}\right) \tag{11.41}$$

on the boundary. $g\partial_\mu g^{-1}$ is obtained from the limiting values of the gauge field configuration, and hence must be continuously defined over the entire boundary. g is extracted by performing the path-ordered exponential integral, as shown in Figure 11.4, along a nest of paths that start at an initial point x_0^μ on the boundary at $x^3 = -L_3/2$ and move along and cover the boundary to all other points on the boundary x^μ

$$g(x^\mu) = \mathcal{P}\exp\left(-\int_{x_0^\mu}^{x^\mu} A_\nu(x'^\mu)dx'^\nu\right). \tag{11.42}$$

The integrability condition that the gauge group element obtained from the path-ordered exponential integral from two different paths is the same, and is exactly that the field strength vanishes on a surface whose boundary comprises the two paths. This condition can be easily verified for an infinitesimal loop. The field strength does indeed vanish for $A_\mu = g\partial_\mu g^{-1}$. Thus g is continuously defined over the entire boundary. g is the unique solution of the linear, first-order differential equation $\partial_\mu g^{-1} = g^{-1}A_\mu$ (or equivalently $\partial_\mu g = -A_\mu g$), up to an irrelevant, multiplicative, constant gauge group element. Equivalently, the actual gauge group element defined by Equation (11.41) is also ambiguously defined up to a constant gauge group element g_0; we can simply take $g \to gg_0$ as then the gauge field $A_\mu = g\partial_\mu g^{-1}$ is invariant. The constant gauge group element is irrelevant, it does not contribute to the action or any winding number. The integration paths are perpendicular to the x^3 direction on the two faces at $x^3 = \pm L_3/2$, hence g is necessarily independent of x^3. Along the other surfaces, we integrate along lines parallel to the x^3 direction, but since $A_3 = 0$ the gauge group element is unchanged. On the two surfaces at $x_3 = \pm L_3/2$, the gauge transformation is not necessarily the same.

Specifying g on the boundary fixes only the tangential components of A_μ since g only varies along the boundary that corresponds to the directions tangential to the boundary. The normal component of A_μ must also be given by the form given in Equation (11.41). However, these then depend on how g varies as we move away from the boundary into the bulk. The normal components of A_μ do not need to be specified, since all we insist on is that the boundary values do not contribute to the equations of motion. Thus we do not have to specify the variation of g as we move away from the boundary into the bulk. One thing is important, since the surface of the box is topologically S^3, the gauge group element g defined on the boundary can perfectly well be in a non-trivial homotopy class of $\Pi_3(G)$, and hence may not necessarily be continuously defined throughout the entire box. Indeed, g is only defined by the asymptotic behaviour of the gauge field on and near the boundary.

On the surfaces at $x^3 = \pm L_3/2$, the gauge group element depends, in principle, non-trivially on the three coordinates (x^1, x^2, x^4) and $g(x^1, x^2, -L_3/2, x^4) \neq g(x^1, x^2, +L_3/2, x^4)$ as in Figure 11.5. But on the surfaces that connect the boundaries of these two ends, since $A_3 = g\partial_3 g^{-1} = 0$ from the gauge condition, we must have that g is independent of x^3. Thus the values of g on the boundaries of the two end surfaces at $x^3 = \pm L_3/2$, *i.e.* for at least one of: $x^1 = \pm L_1/2$, $x^2 = \pm L_2/2$ or $x^4 = \pm T/2$, and $x^3 = \pm L_3/2$, are the same. Now we will perform a gauge transformation by $h(x^1, x^2, x^3, x^4)$, which is actually independent of x^3 and defined by the value of the gauge group element at the surface $x^3 = -L_3/2$, *i.e.*

$$h(x^1, x^2, x^3, x^4) = g^{-1}(x^1, x^2, -L_3/2, x^4). \tag{11.43}$$

Then

$$A_\mu \to h(A_\mu + \partial_\mu)h^{-1} = g^{-1}(x^1, x^2, -L_3/2, x^4)(A_\mu + \partial_\mu)g(x^1, x^2, -L_3/2, x^4). \tag{11.44}$$

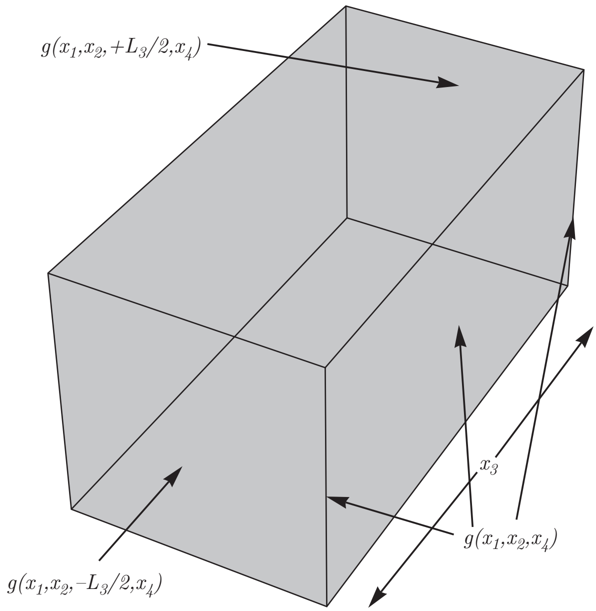

Figure 11.5. Boundary with the gauge group element

But since $A_\mu = g\partial_\mu g^{-1} + o\left(\frac{1}{r}\right)$ for large r, we get

$$A_\mu \to \left(g^{-1}\big|_{x^3=-L_3/2}\right)\left(g\partial_\mu g^{-1} + o\left(\frac{1}{r}\right) + \partial_\mu\right)\left(g\big|_{x^3=-L_3/2}\right). \tag{11.45}$$

We emphasize that the g appearing inside the middle bracket depends on x^3 while that on the outside is independent of x_3 and is equal to its value at $x^3 = -L_3/2$. Evidently we can then write

$$A_\mu = g_1\partial_\mu g_1^{-1} + o\left(\frac{1}{r}\right), \tag{11.46}$$

where

$$g_1(x^\mu) = \left(g^{-1}\big|_{x^3=-L_3/2}\right) g(x^\mu). \tag{11.47}$$

Evidently, g_1 is equal to the identity on the surface at $x^3 = -L_3/2$ and also on the surfaces where x^3 changes, $-L_3/2 \to L_3/2$. On the surface at $x^3 = L_3/2$, g_1 must then be the identity on the boundary of this surface (for $x^1 = \pm L_1/2$ and so on), but for interior values of x^1, x^2 and x^4 generally $g|_{x^3=L_3/2}(x^1,x^2,x^4)$, need not be equal to $g|_{x^3=-L_3/2}(x^1,x^2,x^4)$. Indeed, if the instanton number of the gauge field configuration is non-trivial, then $g_1|_{x^3=L_3/2} \neq \mathbb{I}$ since the gauge transformation h of Equation (11.43) cannot change the instanton number. The instanton number given by the integral Equation (11.31) is gauge-invariant. The surface at $x^3 = L_3/2$ with its boundary identified is also topologically S^3, and g_1 defined on this surface goes to the identity at its boundary. This means that g_1

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

is also defined on this surface with its boundary identified as topologically S^3. Thus g_1 defines a map from $S^3 \to G$, which contains all the topological winding number information of the gauge field defined in the entire box. We want to see what happens under a change of the boundary conditions. We will implement this by fractionally changing the size of the box.

We imagine placing the original box in a larger box that is extended along the x^3 direction by Δ with the gauge field configuration also extended into the larger box. In the larger box, after the corresponding gauge transformation, there will also be a gauge group element, g_2, defined on the surface at $x^3 = L_3/2 + \Delta$, which is identity on its boundary (and identity on all the other surfaces of the box), just like g_1 on its respective boundary. We will extend the box in such a way that the fractional change in the volume is negligible. If we choose $\Delta = (L_3)^{1/2}$, the fractional change in the volume is negligible,

$$\frac{\delta V}{V} = \frac{L_1 L_2 L_4 \Delta}{L_1 L_2 L_3 L_4} = \frac{\Delta}{L_3} = (L_3)^{-1/2} \to 0. \tag{11.48}$$

Alternatively, the volume of the larger box is

$$V + \delta V = L_1 L_2 L_3 L_4 \left(1 + \frac{1}{\Delta}\right) = V\left(1 + \frac{1}{L_3^{1/2}}\right) \to V \quad \text{when} \quad L_3 \to \infty. \tag{11.49}$$

If g_1 and g_2 are in the same homotopy class, we will show that all gauge field configurations defined in the smaller box can be extended to gauge field configurations in the larger box, with negligible change in the action. If g_1 and g_2 are not in the same homotopy class, this is not the case: there has to be at least one more instanton outside the smaller box which implies an increase the action by at least $8\pi^2/g^2$, which is the minimum action of one instanton, as we will see in the next section[2]. This increase in the action is independent of the amount of the extension of volume of the box, even if the volume is only extended fractionally, negligibly. Thus for g_1 and g_2 in the same homotopy class, the action changing negligibly means that extending the box is simply equivalent to a changing boundary condition $g_1 \to g_2$. The action is invariant, but the only vestige of the boundary condition is the topological winding number encoded in g_1 or any other homotopically equivalent boundary gauge group element and the corresponding boundary condition.

Let $g(x^1, x^2, s, x^4)$ with $s \in [0,1]$ be a homotopy from g_1 to g_2:

$$g(s=0) = g_1, \;\; g(s=1) = g_2. \tag{11.50}$$

Then for

$$h = h(x^1, x^2, x^3, x^4) = g(x^1, x^2, (x^3 - L_3/2)/\Delta, x^4) \tag{11.51}$$

[2] See Equation (11.71).

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

for $x^3 \in [+L_3/2, L_3/2+\Delta]$, and with the gauge field extended as $A_\mu = h\partial_\mu h^{-1}$, evidently the action does not change. However, the gauge condition $A_3 = 0$ is not respected. The choice

$$A_\mu = \begin{cases} h\partial_\mu h^{-1} & \mu \neq 3 \\ 0 & \mu = 3 \end{cases} \tag{11.52}$$

does satisfy the gauge condition, but the action is slightly changed. The increase in the action is easily calculated, using gauge invariance. We transform the gauge field of Equation (11.52) by h^{-1}, which gives

$$A_\mu = \begin{cases} 0 & \mu \neq 3 \\ h^{-1}\partial_3 h & \mu = 3 \end{cases}. \tag{11.53}$$

Then, integrating only over the extension,

$$S_E = -\int d^4x \frac{1}{4g^2}\langle F_{\mu\nu}F_{\mu\nu}\rangle = -\int d^4x \frac{1}{4g^2}\langle F_{\mu 3}F_{\mu 3}\rangle. \tag{11.54}$$

Then $A_3 = g^{-1}\partial_s g \partial_3 \left(\frac{x^3 - L_3/2}{\Delta}\right) = \left(g^{-1}\partial_s g\right)\frac{1}{\Delta} \sim \frac{1}{\Delta}$, and consequently

$$F_{\mu 3} = \partial_\mu A_3 - \partial_3 A_\mu = \partial_\mu A_3 \sim \frac{1}{\Delta} \tag{11.55}$$

as the commutator term vanishes. Thus $\langle F_{\mu\nu}F_{\mu\nu}\rangle \sim \frac{1}{\Delta^2}$. But the integral

$$\int_{L_3/2}^{L_3/2+\Delta} d^4x \sim \Delta, \tag{11.56}$$

which implies that the action also changes by a negligible amount

$$\delta S_E \sim \frac{1}{\Delta} = (L_3)^{-1/2} \to 0. \tag{11.57}$$

Hence a change in the boundary conditions that preserves the winding number is just a surface effect, not a volume effect. The action is invariant. Therefore, only the winding number remains, which is defined by the gauge group element $g(x^1, x^2, L_3/2, x^4)$ which defines a map $S^3 \to G$.

Now suppose we decided to choose a different boundary condition, not the one that fixes the tangential components of the gauge field on the boundary but some arbitrary, other boundary condition. We will still work with the gauge condition $A_3 = 0$. The condition that the action be finite still imposes that

$$A_\mu = g\partial_\mu g^{-1} + o\left(\frac{1}{r}\right) \tag{11.58}$$

and nothing obstructs from performing the x^3 independent gauge transformation that gives $\tilde{g}(x^1, x^2, x^4)$ on the end surface at $x^2 = L_3/2$, in the same way as before. We will compare gauge field configurations in this gauge. But now $\tilde{g}$ will not correspond to any such gauge group element obtained when the boundary conditions fixed the tangential components of the gauge field, although the homotopy class of $\tilde{g}$ is fixed by the winding number. Thus gauge field configurations giving rise to $\tilde{g}$ would not be included in the subset of configurations that satisfy the boundary conditions on the tangential components of the gauge field that we have considered. However, the arguments given above show that, although we do not get the exact gauge field configurations with different boundary conditions, we do get configurations that are arbitrarily close, by small deformations near the end at $x^3 = L_3/2$. Indeed, any given $\tilde{g}$ can be obtained by making changes to the gauge field configuration only in the extended part of the box, as we did when defining the homotopy from g_1 to g_2, now we will simply consider the homotopy from g_1 to $\tilde{g}$. Thus all gauge field configurations apart from a small difference in the extended portion of the box are permitted by our boundary conditions, and this difference gives negligible change for a large enough box.

11.3.2 The Theta Vacua

Therefore, for a large enough box, we can simply forget the boundary conditions, but impose that all configurations in the functional integration correspond to those of a fixed winding number n.

$$F(V,T,n) \equiv \mathcal{N} \int \mathcal{D}A e^{-S_E} \delta_{\nu n}, \tag{11.59}$$

where $\mathcal{D}A = \mathcal{D}(A_1, A_2, A_4)$. $F(V,T,n)$ is a matrix element between an initial state and a final state that are determined by the boundary conditions. For T_1 and T_2 taken very large,

$$F(V,T_1+T_2,n) = \sum_{n_1+n_2=n} F(V,T_1,n_1)F(V,T_2,n_2). \tag{11.60}$$

This is because the winding number

$$\nu = \frac{1}{32\pi^2} \int d^4x \langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle \tag{11.61}$$

is an integral of a local density $\langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle$. This means that one way to put a configuration of winding number n into the box with Euclidean time length given by T_1+T_2 is to put $\nu = n_1$ into the first part of the box and $\nu = n_2$ into the second part. Such configurations neglect configurations with significant action on the border between the two parts; however, we expect that this contribution is negligible for large T_1 and T_2. Normally a matrix element for $T = T_1 + T_2$ that gets a contribution from only one energy state follows a multiplicative law. The convolutive law of combination of the matrix elements above, Equation (11.61), can be simply disentangled into the more familiar multiplicative law by a simple

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

Fourier transformation. Defining

$$F(V,T,\theta) = \sum_n e^{in\theta} F(V,T,n) \equiv \mathcal{N} \int \mathcal{D}A e^{-S_E} e^{i\nu\theta} \tag{11.62}$$

implies

$$F(V,T_1+T_2,\theta) = F(V,T_1,\theta) F(V,T_2,\theta). \tag{11.63}$$

This implies the existence of states such that

$$F(V,T,\theta) = N' \langle\theta| e^{-HT} |\theta\rangle, \tag{11.64}$$

where the states $|\theta\rangle$ are eigenstates of the Hamiltonian. Our field theory is now surprisingly separated into a family of sectors enumerated by θ. In each sector we use the same action except we add an extra term proportional to $\nu = \theta\langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle$.

We can obtain all of these results from the functional integral and from the possible instanton solutions to the Euclidean equations of motion. If there is a solution for $\nu = 1$, all the results follow. Suppose such a solution exists with action S_0. Then translation invariance gives at least four zero modes, and

$$\begin{aligned}\langle\theta| e^{-HT} |\theta\rangle &= N' \int \mathcal{D}A e^{-S_E} e^{i\nu\theta} \\ &= \sum_{n,\bar{n}} \left(\left(K e^{-S_0}\right)^{n+\bar{n}} (VT)^{n+\bar{n}} e^{i(n-\bar{n})\theta} \right) / n!\bar{n}! \\ &= e^{2KVTe^{-S_0}\cos\theta}\end{aligned} \tag{11.65}$$

where K is the usual determinantal factor and a sum has been done over n instantons and $\bar{n}$ anti-instantons. Then we can read off the energy of the $|\theta\rangle$ states,

$$E(\theta) = -2VK\cos\theta e^{-S_0}. \tag{11.66}$$

We can also compute the expectation value

$$\begin{aligned}\langle\theta|\langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle|\theta\rangle &= \frac{1}{VT} \int d^4x \langle\theta|\langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle|\theta\rangle \\ &= \frac{32\pi^2 \int \mathcal{D}A\, \nu e^{-S_E} e^{i\nu\theta}}{VT \int \mathcal{D}A e^{-S_E} e^{i\nu\theta}} \\ &= \frac{-32\pi^2 i}{VT} \frac{d}{d\theta} \ln\left(\int \mathcal{D}A e^{-S_E} e^{i\nu\theta} \right) \\ &= \frac{-32\pi^2 i}{VT} \left(-2K\cos\theta e^{-S_0}\right) VT \\ &= -64\pi^2 i K e^{-S_0} \sin\theta.\end{aligned} \tag{11.67}$$

The answer is imaginary, but this is correct. Since $\langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle = \langle F_{12}F_{34} + perm.\rangle$ in Euclidean space, analytic continuation to Minkowski space yields, for example, $F_{j4} \to iF_{j0}$. Hence the imaginary result in Euclidean space corresponds to the correct, real result in Minkowski space. Everything depends on θ, it is a physical parameter.

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

11.3.3 The Yang–Mills Instantons

The instantons do actually exist as solutions of the Euclidean equations of motion. We can prove that the action in the single instanton sector is bounded from below, and when the bound is saturated, the configuration must satisfy the equations of motion. Consider the inequality which is evidently satisfied

$$-\int d^4x \langle (F_{\mu\nu} \pm \tilde{F}_{\mu\nu})(F_{\mu\nu} \pm \tilde{F}_{\mu\nu}) \rangle \geq 0, \tag{11.68}$$

note that the minus sign is there because our gauge fields and field strengths are anti-hermitean. This implies

$$-\int d^4x \left(\langle F_{\mu\nu} F_{\mu\nu} \rangle + \langle \tilde{F}_{\mu\nu} \tilde{F}_{\mu\nu} \rangle \pm 2 \langle F_{\mu\nu} \tilde{F}_{\mu\nu} \rangle \right) \geq 0. \tag{11.69}$$

But the first two terms are equal, hence choosing the $\pm$ sign appropriately, we have

$$-\int d^4x \langle F_{\mu\nu} F_{\mu\nu} \rangle \geq \left| \int d^4x \langle F_{\mu\nu} \tilde{F}_{\mu\nu} \rangle \right|. \tag{11.70}$$

But the right-hand side is just the instanton number while the left-hand side is proportional to the action, thus we find

$$S_E \geq \frac{8\pi^2}{g^2} |\nu| \tag{11.71}$$

as we had promised to show earlier. The equality is attained for

$$F_{\mu\nu} = \pm \tilde{F}_{\mu\nu}, \tag{11.72}$$

where we get the $+$ sign for $\nu \geq 0$ and the minus sign for $\nu \leq 0$. If we can find the solutions of Equation (11.72), we automatically get solutions of the full equations of motion, as the action is minimal for such configurations and hence, must be stationary. A bonus is that Equation (11.72) as a differential equation is now a first-order equation, instead of a second-order equation, and is consequently easier to solve.

For $\nu = 1$ we will look for a solution that asymptotically behaves as

$$A_\mu = g^{(1)} \partial_\mu \left(g^{(1)} \right)^{-1} + o\left(\frac{1}{r} \right), \tag{11.73}$$

where $g^{(1)}$ is the gauge group element defined in Equation (11.19). $g^{(1)}$ is spherically symmetric, hence we make the ansatz

$$A_\mu = f(r) r^2 g^{(1)} \partial_\mu \left(g^{(1)} \right)^{-1} = -i A^i_\mu \sigma^i. \tag{11.74}$$

Using a double index notation, for the gauge group $SU(2)$ seen as the diagonal subgroup of $SO(4)$, we can write

$$A^i_\mu = \frac{1}{2} \left(A^{0i}_\mu + \frac{1}{2} \epsilon^{ijk} A^{jk}_\mu \right) \tag{11.75}$$

for anti-symmetric $SO(4)$-valued gauge fields $A_\mu^{\alpha\beta} = -A_\mu^{\beta\alpha}$. An easy calculation with g given in Equation (11.19) gives

$$A_\mu^{\alpha\beta} = f(r)(x_\alpha\delta_{\mu\beta} - x_\beta\delta_{\mu\alpha}). \tag{11.76}$$

Then a straightforward calculation gives

$$F_{\mu\nu}^{\alpha\beta} = (2f - r^2 f^2)(\delta_{\mu\alpha}\delta_{\nu\beta} - \delta_{\mu\beta}\delta_{\nu\alpha}) + \left(\frac{f'}{r} + f^2\right) \times (x_\alpha x_\mu \delta_{\nu\beta} - x_\alpha x_\nu \delta_{\mu\beta} + x_\beta x_\nu \delta_{\mu\alpha} - x_\beta x_\mu \delta_{\nu\alpha}). \tag{11.77}$$

The condition of self duality, $F_{\mu\nu} = \tilde{F}_{\mu\nu}$, is automatically satisfied for the first term $(\delta_{\mu\alpha}\delta_{\nu\beta} - \delta_{\mu\beta}\delta_{\nu\alpha}) = \epsilon_{\mu\nu\sigma\tau}\epsilon_{\sigma\tau\alpha\beta}$

$$\frac{1}{2}\epsilon_{\lambda\rho\mu\nu}(\delta_{\mu\alpha}\delta_{\nu\beta} - \delta_{\mu\beta}\delta_{\nu\alpha}) = \frac{1}{2}\epsilon_{\lambda\rho\mu\nu}\epsilon_{\mu\nu\sigma\tau}\epsilon_{\sigma\tau\alpha\beta} = \epsilon_{\lambda\rho\sigma\tau}\epsilon_{\sigma\tau\alpha\beta} \tag{11.78}$$

but not so for the second term

$$\frac{1}{2}\epsilon_{\lambda\rho\mu\nu}(x_\alpha x_\mu \delta_{\nu\beta} - x_\alpha x_\nu \delta_{\mu\beta} + x_\beta x_\nu \delta_{\mu\alpha} - x_\beta x_\mu \delta_{\nu\alpha}) = \epsilon_{\lambda\rho\mu\beta} x_\alpha x_\mu - \epsilon_{\lambda\rho\mu\alpha} x_\beta x_\mu. \tag{11.79}$$

Thus we obtain a self dual field strength by imposing

$$\frac{f'}{r} + f^2 = 0. \tag{11.80}$$

This integrates trivially as

$$f(r) = \frac{1}{r^2 + \lambda^2}, \tag{11.81}$$

where λ is an arbitrary integration constant. Thus

$$A_\mu = \frac{r^2}{r^2 + \lambda^2} g^{(1)} \partial_\mu \left(g^{(1)}\right)^{-1}. \tag{11.82}$$

We will find that there exist eight parameters corresponding to symmetries of the action that are broken by the solution. These correspond in principle to scale transformations, rotations, translations, special conformal transformations and global gauge transformations. Scale transformations correspond to changing λ. Note that the global gauge transformations preserve the gauge-fixing conditions $A_3 = 0$. Rotations and global gauge transformations are tied together, the solution is invariant under the diagonal subgroup of simultaneous rotations and global gauge transformations by the same amount. Note that the rotation group $SO(3)$ and the global gauge group $SU(2)$ are essentially the same group. Special conformal transformations can be obtained by translations and gauge transformations and hence do not give rise to new solutions. This in the end leaves eight parameters, coming from one scale transformation, four translations and three rotations (or equivalently global gauge transformations). For a configuration on n instantons and $\bar{n}$ anti-instantons, the number of parameters is simply $8(n+\bar{n})$.

11.4 Theta Vacua in QCD

The existence of the instanton solutions means that there exist different, inequivalent classical ground states, between which the instantons mediate quantum tunnelling. We have not explicitly seen these vacuum configurations; to uncover them, it is more convenient to use the temporal gauge, *i.e.* $A_0 = 0$. As, in principle, everything we do does not depend on the gauge choice, we are free to take any gauge that we want. The dynamical variables in this gauge are just the spatial components of the gauge field A_i.

In Minkowski space, the Hamiltonian is given by

$$\mathcal{H} = \int d^3x \frac{1}{2}\left((E_i^a)^2 + (B_i^a)^2 \right), \tag{11.83}$$

where the electric and magnetic fields are given by

$$\begin{aligned} E_i^a &= \dot{A}_i^a \\ B_i^a &= \frac{1}{2}\epsilon_{ijk}\left(\partial_j A_k^a - \partial_k A_j^a + f^{abc} A_j^b A_k^c\right). \end{aligned} \tag{11.84}$$

In this gauge, since there is no field A_0, the equation of motion that usually comes from varying with respect to it is missing. This is the Gauss law

$$\mathcal{G}^a = \partial_i \dot{A}_i^a + f^{abc} A_i^b \dot{A}_i^c \equiv (D_i E_i)^a = 0. \tag{11.85}$$

However, in this gauge the Hamiltonian is invariant under time-independent, spatial gauge transformations. The corresponding conserved charge is actually a local expression, exactly the Gauss law operator, $\dot{\mathcal{G}}^a = 0$, *i.e.*

$$[\mathcal{H}, \mathcal{G}^a] = 0. \tag{11.86}$$

Thus we must impose this constraint on the initial values of the fields, then since the time evolution preserves the constraint, the Gauss law operator will be preserved for all time.

Now in the quantum theory, eigenstates of the field operator $A_i^a(x)$ correspond to the states $|A_i^a(x)\rangle$ and the amplitude for a transition between two such states is given by the functional integral

$$\langle \tilde{A}_i^a(x)| e^{-i\mathcal{H}T} |A_i^a(x)\rangle = \int_{A_i^a(x)}^{\tilde{A}_i^a(x)} \mathcal{D}\left(A_i^a(\vec{x}(t))\right) e^{\frac{-i}{2g^2}\int_0^T d^t d^3x \left((E_i^a)^2 + (B_i^a)^2 \right)}, \tag{11.87}$$

where the functional integral is over all time histories $A_i^a(\vec{x}(t)$ that interpolate between the initial and final configurations. In the quantum theory, however, the Gauss law constraint is imposed as a constraint on the Hilbert space, a physical state in the Hilbert space of all states is one that is annihilated by the Gauss law operator

$$\mathcal{G}^a(x)|\Psi\rangle = (D_i E_i)^a |\Psi\rangle = 0, \tag{11.88}$$

the states $|A_i^a(x)\rangle$ do not satisfy this constraint. We wish to characterize the states that do satisfy the Gauss constraint.

Under a gauge transformation

$$A_i \to \mathcal{U}_\lambda A_i \mathcal{U}_\lambda^{-1} + i\mathcal{U}_\lambda \partial_i \mathcal{U}_\lambda^{-1} \tag{11.89}$$

with $\mathcal{U}_\lambda = e^{i\lambda^a T^a}$ where $\lambda(\vec{x})$ is independent of t. Defining the corresponding conserved charge

$$\begin{aligned} Q_\lambda &= \int d^3x \dot{A}_i^a \left(\partial_i \lambda^a + F^{abc} A_i^b \lambda^c\right) \\ &= -\int d^3x \lambda^a(\vec{x})(D_i E_i)^a \end{aligned} \tag{11.90}$$

integrating by parts and assuming $\lambda^a \to 0$ for $|\vec{x}| \to \infty$. Then the gauge transformation can be effected by Q_λ as

$$\begin{aligned} A_i \vec{x}, t) \to A_i' \vec{x}, t) &= e^{-iQ_\lambda} A_i \vec{x}, t) e^{iQ_\lambda} \\ &= A_i \vec{x}, t) - i[Q_\lambda, A_i \vec{x}, t)] \quad \text{for infinitesimal } \lambda \\ &= A_i \vec{x}, t) - i \int d^3y \\ &\quad \times \left[\dot{A}_i^a(\vec{y}, t) \left(\partial_i \lambda^a(\vec{y}, t) + f^{abc} A_i^b(\vec{y}, t) \lambda^c(\vec{y}, t)\right), A_i(\vec{x}, t)\right] \\ &= A_i \vec{x}, t) - (D_i \lambda)^a \vec{x}, t) T^a \end{aligned} \tag{11.91}$$

using the equal time canonical commutator $[\dot{A}_i^a(\vec{y}, t)), A_j^b(\vec{x}, t))] = \delta^{ab}\delta^3(\vec{x} - \vec{y})$ and that the time variable in the integral expression for Q_λ can be chosen arbitrarily since it is in fact time-independent and here conveniently chosen equal to the time variable t in $A_i(\vec{x}, t)$. Thus Q_λ generates the infinitesimal gauge transformation corresponding to λ, and physical states should be invariant under the action of this gauge transformation, *i.e.*

$$e^{iQ_\lambda}|\Psi\rangle = |\Psi\rangle \tag{11.92}$$

for λ falling off sufficiently fast as $|\vec{x}| \to \infty$. But

$$e^{iQ_\lambda}|A_i^a(x)\rangle = |\mathcal{U}_\lambda(A_i + i\partial_i)\mathcal{U}_\lambda^{-1}\rangle, \tag{11.93}$$

where $\mathcal{U}_\lambda = e^{i\lambda^a(\vec{x})T^a}$. Therefore, a physical state will be obtained if we sum over all states that are related by a gauge transformation

$$|A_i(\vec{x})\rangle_{physical} = \int \mathcal{D}\lambda'^a(\vec{x}) e^{iQ_{\lambda'}} |A_i^a(x)\rangle \tag{11.94}$$

integrating over a dummy field variable λ'. This is obvious since the integration measure is invariant under translation, hence

$$\begin{aligned} e^{iQ_\lambda}|A_i(\vec{x})\rangle_{physical} &= e^{iQ_\lambda} \int \mathcal{D}\lambda'^a(\vec{x}) e^{iQ_{\lambda'}} |A_i^a(x)\rangle = \int \mathcal{D}\lambda'^a(\vec{x}) e^{iQ_{\lambda+\lambda'}} |A_i^a(x)\rangle \\ &= \int \mathcal{D}\lambda'^a(\vec{x}) e^{iQ_{\lambda'}} |A_i^a(x)\rangle = |A_i(\vec{x})\rangle_{physical}. \end{aligned} \tag{11.95}$$

Now what are the possible classical ground-state configurations? Such configurations must have $\mathcal{H}=0$. This requires $E_i^a=0$ and $B_i^a=0$. A vanishing magnetic field means that the gauge field is pure gauge $A_i = g\partial_i g^{-1}$ and a vanishing electric field means $\dot{A}_i=0$, which implies $\dot{g}=0$. Thus for a classical vacuum the gauge field must be $A_i = g(\vec{x})\partial_i(g(\vec{x}))^{-1}$, and then we must implement gauge invariance as in Equation (11.94). Writing a corresponding state as $|g(\vec{x})\rangle$, for a gauge transformation

$$|g(\vec{x})\rangle \to e^{iQ_\lambda}|g(\vec{x})\rangle = |\mathcal{U}_\lambda g(\vec{x})\rangle = |\tilde{g}(\vec{x})\rangle. \tag{11.96}$$

Since $\lambda\to 0$ for $|\vec{x}|\to\infty$, $\tilde{g}$ and g must be homotopically equivalent, *i.e.* $\mathcal{U}_\lambda$ is homotopically trivial (evidently, just switch $\lambda\to 0$). Hence a potential vacuum state $|0\rangle$ is given by

$$|0\rangle = \sum_{g\,\in\,\text{one homotopy class}} |g(\vec{x})\rangle. \tag{11.97}$$

Without loss of generality, for the state $|0\rangle$ we choose the equivalence class corresponding to all gauge group elements that are homotopically trivial, *i.e.* in the same homotopy class as the constant, identity gauge transformation $g=\mathbb{I}$ and are generated by multiplication by $\mathcal{U}_\lambda$.

$$|0\rangle = \sum_{g\,\in\,\text{trivial homotopy class}} |g(\vec{x})\rangle. \tag{11.98}$$

But what if we define a different vacuum, obtained from $|\bar{g}(\vec{x})\rangle$

$$\overline{|0\rangle} = \sum_{g\,\in\,\text{ homotopy class of}\,\bar{g}} |g(\vec{x})\rangle \tag{11.99}$$

for some other gauge group element $\bar{g}$ which is not in the trivial homotopy class. $\bar{g}$ must go to identity at infinity. If $\bar{g}$ does not go to identity at infinity, and instead goes to some other constant gauge group element, g_0, then such a state is irrelevant. The matrix element between the state so defined and the state $|0\rangle$ must necessarily vanish since we must integrate over configurations, in Equation (11.87), which are spatially constant at infinity but change in time from $\mathbb{I}$ to g_0. Such configurations will have a non-zero $\dot{A}_i$ over an infinite spatial volume (at infinity), for which the action is infinite and consequently the transition amplitude vanishes. Thus $\bar{g}$ defines an element of the homotopy classes that need not be the trivial class. Evidently the state $\overline{|0\rangle}$ is also a vacuum state; the corresponding classical field configurations that we integrate over $A_i = g\partial_i g^{-1}$ in Equation (11.94), are all of zero energy.

All gauge group elements that we are considering here satisfy $\lim_{|\vec{x}|\to\infty} g(\vec{x}) \to \mathbb{I}$. Thus, topologically, all gauge group elements are defined on $\mathbf{R}^3$ with the point at infinity identified, topologically S^3. Each $g(\vec{x})$ defines a mapping from $S^3\to G$, which fall into the homotopy classes of $\Pi_3(G)=\mathbb{Z}$. We can index the homotopy

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

classes by an integer n,

$$g^{(0)} = \mathbb{I}$$
$$g^{(1)} = e^{i\pi \frac{\vec{x}\cdot\vec{\sigma}}{(|\vec{x}|^2+\lambda^2)^{1/2}}}$$
$$g^{(n)} = \left(g^{(1)}\right)^n. \tag{11.100}$$

Correspondingly, we can enumerate the classical vacua with the integer n

$$|n\rangle = \sum_{g\,\text{of winding number}\,n} |g\rangle \tag{11.101}$$

and the winding number is given by the formula

$$n = -\frac{1}{24\pi^2}\int d^3x\epsilon_{ijk}Tr\left(g\partial_i g^{-1} g\partial_j g^{-1} g\partial_k g^{-1}\right). \tag{11.102}$$

If we denote by $\mathcal{R}$, the operator that implements the gauge transformation $g^{(1)}$, then

$$\mathcal{R}|n\rangle = |n+1\rangle. \tag{11.103}$$

Note that $g^{(1)} \neq e^{i\lambda^a T^a}$ with $\lim_{|\vec{x}|\to\infty}\lambda^a \to 0$, hence gauge invariance does not impose that the vacuum state be invariant under action of $\mathcal{R}$. But physically we would imagine that gauge invariance would require at least that

$$\mathcal{R}|\Psi\rangle = e^{i\theta}|\Psi\rangle, \tag{11.104}$$

since we cannot physically measure an overall phase factor. A vacuum state that satisfies Equation (11.104) is called a theta vacuum, and is denoted by $|\theta\rangle$.

There is no physical principle that can predict θ. However, θ must be time-independent since $[\mathcal{H},\mathcal{R}] = 0$. θ must also be gauge-invariant. We say that θ labels the superselection sectors of the Hilbert space and the Hamiltonian is diagonal, block by block, for each superselection sector indexed by θ. We can explicitly construct the state labelled by θ by a simple Fourier sum

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{in\theta}|n\rangle. \tag{11.105}$$

These are the physical vacua of QCD, gauge invariant under trivial gauge transformations, and invariant up to an overall phase under topologically non-trivial gauge transformations. In the next section we will see how instantons give rise to these theta vacua.

11.4.1 Instantons: Specifics

In this section we will complete the specifics of the instanton solutions, some of which we have already used. The solution from Equation (11.75) is given by the gauge field configuration for gauge group $SU(2)$ (we will put all indices down for convenience)

$$A_\mu^{\alpha\beta} = \frac{1}{r^2+\lambda^2}(x_\alpha \delta_{\mu\beta} - x_\beta \delta_{\mu\alpha}). \tag{11.106}$$

The corresponding field strength is

$$F_{\mu\nu}^{\alpha\beta} = \frac{r^2+2\lambda^2}{(r^2+\lambda^2)^2}(\delta_{\mu\alpha}\delta_{\nu\beta} - \delta_{\mu\beta}\delta_{\nu\alpha}). \tag{11.107}$$

Equivalently in matrix form

$$A_\mu = \frac{r^2}{r^2+\lambda^2} g^{(1)} \partial_\mu \left(g^{(1)}\right)^{-1}. \tag{11.108}$$

Obviously, as $r \to \infty$, the field strength vanishes quadratically as $\sim 1/r^2$. Thus the action of the instanton is located in an essentially compact region of Euclidean spacetime. This was the reason for the name "instanton"; if we scan up through Euclidean time, there is nothing at the beginning, then, for an instant, the instantons turn on and off in a localized spatial region, and then, again, there is nothing.

The instanton solution is rotationally invariant when compensated by a global gauge transformation. This is evident for the field strength $F_{\mu\nu}^{\alpha\beta}$. For the gauge field we must have the same. A rotation is defined by

$$x_\mu \to x'_\mu = R_{\mu\nu} x_\nu \tag{11.109}$$

with

$$R_{\mu\nu} R_{\mu\sigma} = \delta_{\nu\sigma}, \tag{11.110}$$

since $x_\mu x_\mu \to R_{\mu\nu} x_\nu R_{\mu\sigma} x_\sigma = R_{\mu\nu} R_{\mu\sigma} x_\nu x_\sigma$. For infinitesimal transformations, we have $R_{\mu\nu} = \delta_{\mu\nu} + \Lambda_{\mu\nu}$, where $\Lambda_{\mu\nu}$ is infinitesimal. Then

$$x_\mu \to x'_\mu = x_\mu + \Lambda_{\mu\nu} x_\nu \tag{11.111}$$

and

$$(\delta_{\mu\nu} + \Lambda_{\mu\nu})(\delta_{\mu\sigma} + \Lambda_{\mu\sigma}) = \delta_{\nu\sigma} + \Lambda_{\sigma\nu} + \Lambda_{\nu\sigma} = \delta_{\nu\sigma}, \tag{11.112}$$

which requires

$$\Lambda_{\sigma\nu} + \Lambda_{\nu\sigma} = 0 \tag{11.113}$$

or equivalently, $\Lambda_{\sigma\nu} = -\Lambda_{\nu\sigma}$. Then the gauge field transforms as

$$A_\mu(x) \to A'_\mu(x') = A_\mu(x) + (\Lambda_{\mu\nu} + \delta_{\mu\nu}\Lambda_{\sigma\tau} x_\sigma \partial_\tau) A_\nu(x) \tag{11.114}$$

thus the gauge field $A_\mu(x) = \frac{r^2}{r^2+\lambda^2} g \partial_\mu g^{-1}$ will transform as

$$\begin{aligned} A_\mu(x) \to A'_\mu(x') &= A_\mu(x) + \frac{r^2}{r^2+\lambda^2}\big(\Lambda_{\mu\nu} g \partial_\nu g^{-1} + (\Lambda_{\sigma\tau} x_\sigma \partial_\tau g)\partial_\mu g^{-1} \\ &\quad + g\partial_\mu(\Lambda_{\sigma\tau} x_\sigma \partial_\tau g^{-1}) - g\Lambda_{\mu\tau}\partial_\tau g^{-1}\big) \\ &= \frac{r^2}{r^2+\lambda^2}\left(g(x')\partial_\mu g^{-1}(x')\right) \end{aligned} \tag{11.115}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

using that r is invariant and $g(x') = g(x) + \Lambda_{\sigma\tau}x_\sigma\partial_\tau g$ to first order. Then explicitly with $g(x) = g^1(x) = \frac{1}{r}(x^4 + ix^i\sigma^i)$

$$g(x') = \frac{1}{r}(x^4 + ix^i\sigma^i) + \frac{1}{r}\Lambda_{\sigma\tau}x_\sigma\partial_\tau(x^4 + ix^i\sigma^i). \tag{11.116}$$

Then with $\Lambda_{i4} = \lambda_i$ and $\Lambda_{ij} = \epsilon_{ijk}\alpha_k$ we get

$$\begin{aligned}
g(x') &= g(x) + \frac{1}{r}\lambda_i(x_i\partial_4 - x_4\partial_i)(x_4 + ix_j\sigma_j) \\
&\quad + \frac{1}{r}\epsilon_{ijk}\alpha_k x_i\partial_j i x_l\sigma_l \\
&= g(x) + \frac{1}{r}\lambda_i(x_i - ix_4\sigma_i) + \frac{i}{r}\epsilon_{ijk}\alpha_k x_i\sigma_j \\
&= g(x) + \frac{1}{r}\left(x_4(-i\lambda_i\sigma_i) + x_i(\lambda_i + i\epsilon_{ijk}\sigma_j\alpha_k)\right) \\
&= \frac{1}{r}\left(x_4(1 - i\lambda_i\sigma_i) + x_i(i\sigma_i + \lambda_i + i\epsilon_{ijk}\sigma_j\alpha_k)\right) \\
&= \frac{1}{r}(1 - i\gamma_i\sigma_i)(x_4 + ix_j\sigma_j)(1 + i\beta_k\sigma_k)
\end{aligned} \tag{11.117}$$

where $\lambda_i = \gamma_i - \beta_i$ and $\alpha_i = \gamma_i + \beta_i$, since, continuing the algebra to first order we get

$$\begin{aligned}
g(x') = &= \frac{1}{r}(x_4(1 - i(\gamma_i - \beta_i)\sigma_i) + x_i(i\sigma_i + \gamma_j\sigma_j\sigma_i - \beta_j\sigma_i\sigma_j)) \\
&= \frac{1}{r}(x_4(1 - i(\gamma_i - \beta_i)\sigma_i) + x_i(i\sigma_i + (\gamma_i - \beta_i) + i\epsilon_{ijk}(\gamma_j + \beta_j)\sigma_k)
\end{aligned} \tag{11.118}$$

confirming

$$g(x') = (1 - i\gamma_i\sigma_i)g(x)(1 + i\beta_i\sigma_i) = a^{-1}g(x)b, \tag{11.119}$$

where $a = (1 + \gamma_i\sigma_i)$ while $b = (1 + i\beta_i\sigma_i)$ to first order. Then from Equation (11.115)

$$A'_\mu(x') = \frac{r^2}{r^2 + \lambda^2}a^{-1}gb\partial_\mu(b^{-1}g^{-1}a) = a^{-1}A_\mu(x)a. \tag{11.120}$$

Thus the solution is clearly invariant under an arbitrary choice of b. Thus the instanton is invariant under an arbitrary choice of b, but it is not invariant under a.

This should give rise to three zero modes, it does, but in a slightly more complicated way. The important point is that rotations can be compensated by global gauge transformations. The rotation group $SO(4) = SO_a(3) \times SO_b(3)$ is explicitly broken to the $SO_a(3)$ subgroup. The instanton is invariant under the $SO_b(3)$ subgroup and it does not give rise to new solutions or equivalently to zero modes. In principle, this subgroup can be used to characterize the representations under which the physical states of the theory transform, exactly as, for example,

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

the invariance of a physical system under spatial rotations tells us that the physical states of the system must transform according to representations of the rotation group. The broken subgroup is $SO_a(3)$, which should give rise to new solutions and zero modes, but its transformation can be exactly compensated by a transformation of the group of global gauge transformation $SO_{gl.}(r)$, which is also $SU(2) = SO_{gl.}(3)$. Under a global gauge transformation

$$A'_\mu(x') \to hA'_\mu(x')\; h^{-1} = ha^{-1}A_\mu(x)\; ah^{-1} \tag{11.121}$$

hence $h = a$ is a symmetry of the solution. Therefore, the symmetry group $SO_{gl.}(3) \times SO_a(3)$ is in fact broken to the diagonal subgroup $SO_d(3)$, corresponding to $h = a$, which remains a symmetry of the instantons. The anti-diagonal subgroup $SO_{ad}(3)$, with $h = a^{-1}$, is broken by the instanton configuration, and gives rise to exactly three zero modes. Thus the rotation group is broken to $SO_a(3)$

$$SO(4) = SO_a(3) \times SO_b(3) \to SO_a(3), \tag{11.122}$$

while the group of global gauge transformations $SU(2) = SO_{gl.}(3)$ is mixed with the rotation group

$$SO_{gl.}(3) \times SO_a(3) = SO_d(3) \times SO_{ad}(3) \tag{11.123}$$

and the diagonal subgroup remains an symmetry of the solution, while the anti-diagonal subgroup gives rise to three zero modes.

11.4.2 Transitions Between Vacua

The instantons are perfectly suited to describing quantum tunnelling transitions between the $|n\rangle$ vacua. The solution can be put into the gauge $A_4 = 0$ by the straightforward gauge transformation $A_\mu \to h(A_\mu + \partial_\mu)h^{-1}$ with

$$h = \mathcal{P}\left(e^{i\int_{-\infty}^{x_4} dx'_4 A_4(x'_4)}\right). \tag{11.124}$$

Then at $\tau = -\infty$ the gauge field is

$$A_i|_{\tau=-\infty} = 0 \tag{11.125}$$

but at $\tau = \infty$ we have

$$A_i|_{\tau=\infty} = g\partial_i g^{-1}, \tag{11.126}$$

where

$$g = e^{-i\pi\frac{\vec{x}\cdot\vec{\sigma}}{(x^2+\lambda^2)^{1/2}}}. \tag{11.127}$$

The gauge field is given in slightly different notation by 't Hooft [112],

$$A^a_\mu = 2\frac{x_\nu}{x^2+\lambda^2}\eta_{a\mu\nu}, \tag{11.128}$$

where $\eta_{a\mu\nu}$ is the 't Hooft tensor

$$\eta_{a\mu\nu} = \epsilon_{4a\mu\nu} + \frac{1}{2}\epsilon_{abc}\epsilon_{bc\mu\nu} \tag{11.129}$$

with corresponding field strength

$$F^a_{\mu\nu} = 4\frac{\lambda^2}{(x^2+\lambda^2)^2}\eta_{a\mu\nu}. \tag{11.130}$$

This configuration corresponds to a change in the winding number between $\tau = -\infty$ and $\tau = \infty$

$$\begin{aligned}\Delta n &= \frac{-1}{24\pi^2}\int d^3x \epsilon_{ijk} Tr\left(g\partial_i g^{-1} g\partial_j g^{-1} g\partial_k g^{-1}\right)\Big|_{\tau=-\infty}^{\tau=\infty} \\ &= \frac{1}{32\pi^2}\int d^3x\, G_4\Big|_{\tau=-\infty}^{\tau=\infty} \\ &= \frac{1}{32\pi^2}\int d^4x\, \partial_\tau G_4 \\ &= \frac{1}{32\pi^2}\int d^4x\, (\partial_\mu G_\mu - \partial_i G_i)\end{aligned} \tag{11.131}$$

where G_μ was defined in Equation (11.27). The spatial components G_i are, using $A_4 = 0$,

$$G_i \sim \epsilon_{i\mu 0\lambda}(A_\mu \partial_\nu A_\lambda + \frac{2}{3}A_\mu A_\nu A_\lambda) = \epsilon_{i\mu 0\lambda} A_\mu \dot{A}_\lambda = \epsilon_{ijk} A_j \dot{A}_k, \tag{11.132}$$

which vanish as $|\vec{x}| \to \infty$ since the electric field $\dot{A}_k = E_k \to 0$ so that the subtracted spatial divergence gives no contribution from the surface at infinity. Thus we find

$$\Delta n = \frac{1}{32\pi^2}\int d^4x\, \partial_\mu G_\mu = \frac{1}{32\pi^2}\int d^4x\, \langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle = 1. \tag{11.133}$$

Thus the instanton mediates transitions with the vacua $|0\rangle \to |1\rangle$.

11.5 Instantons and Confinement

We will now consider quantum corrections that come from the Gaussian functional integral that should be performed about the instanton configuration. We must first extract the zero modes. For the single instanton, there exist, in fact, eight. In principle the amplitude $\langle 1|e^{-HT}|0\rangle$ is given by

$$\langle 1|e^{-HT}|0\rangle \equiv Z(T) = e^{-S_0}\int \mathcal{D}(Q^a_\mu) e^{-\frac{1}{2}\int d^4x\, Q\cdot\left(\frac{\delta^2}{\delta Q^2}\mathcal{L}(A^a_\mu)\right)\cdot Q}, \tag{11.134}$$

where Q^a_μ is the fluctuation that gives rise to the quantum corrections. The zero modes coming from translations and scale transformations can be eliminated by

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

using a Faddeev–Popov method. We insert one into the integral

$$1 = S_0^5 \int d^4R\delta^4\left(\int d^4x\,(\mathcal{L}(A_\nu^a(x))(x_\mu - R_\mu))\right)\int_0^\infty d(\lambda^2)$$
$$\delta\left(\int d^4x(\mathcal{L}(A_\nu^a(x))((x-R)^2-\lambda^2))\right). \tag{11.135}$$

The delta functions choose the values of R and λ as

$$\delta^4\left(\int d^4x\,(\mathcal{L}(A_\nu^a(x))(x_\mu - R_\mu))\right) = \delta^4\left(\left(\int d^4x\,\mathcal{L}(A_\nu^a(x))\right)(\bar{R}_\mu - R_\mu) + \cdots\right)$$
$$= \frac{\delta^4(R_\mu - \bar{R}_\mu)}{\left(\int d^4x\,\mathcal{L}(A_\nu^a(x))\right)^4} = \frac{\delta^4(R_\mu - \bar{R}_\mu)}{S_0^4} \tag{11.136}$$

with the obvious definition

$$\bar{R}_\mu = \frac{\int d^4x\,(\mathcal{L}(A_\nu^a(x))x_\mu)}{\int d^4x\,\mathcal{L}(x)}. \tag{11.137}$$

Furthermore,

$$\delta\left(\int d^4x(\mathcal{L}(A_\nu^a(x))((x-R)^2-\lambda^2))\right) = \frac{\delta(\lambda^2 - \bar{\lambda}^2)}{\int d^4x\,\mathcal{L}(x)} = \frac{\delta(\lambda^2 - \bar{\lambda}^2)}{S_0} \tag{11.138}$$

with

$$\bar{\lambda}^2 = \frac{\int d^4x\,\mathcal{L}(A_\nu^a(x))(x-R)^2}{\int d^4x\,\mathcal{L}(x)} \tag{11.139}$$

and here R_μ could be replaced with $\bar{R}_\mu$ because of the first delta function. Evidently, $\bar{R}$ depends on what gauge field $A_\nu^a(x)$ is chosen: it should correspond to an instanton, and it contains the data on where the instanton is and its scale, λ. Then starting with $Z(T)$ slightly differently

$$Z(T) = S_0^5 \int d(\lambda^2)\int d^4R \int \mathcal{D}(Q_\mu^a)e^{-S_E} \times$$
$$\times\delta^4\left(\int d^4x\,\big(\mathcal{L}(Q_\mu^a(x))(x_\mu - R_\mu)\big)\right)\delta\left(\int d^4x(\mathcal{L}(Q_\mu^a(x))((x-R)^2-\lambda^2))\right). \tag{11.140}$$

First we perform a translation and a conformal scaling

$$x_\mu \to x'_\mu = \lambda x_\mu + R_\mu$$
$$Q_\mu^a(x_\nu) \to \lambda Q_\mu^a(x'_\nu). \tag{11.141}$$

The action is invariant under a translation and conformal scaling (actually under all special conformal transformations) with

$$\mathcal{L}(x) = -\frac{1}{4g^2}Tr\left(\partial_\mu Q_\nu(x) - \partial_\nu Q_\mu(x) + [Q_\mu(x), Q_\nu(x)]\right)^2 \tag{11.142}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

we have

$$
\begin{aligned}
d^4x\mathcal{L}(\lambda Q(\lambda x - R)) &= -\lambda^{-4} d^4x' Tr\left(\lambda\partial'_\mu \lambda Q_\nu(x') - \lambda\partial'_\nu \lambda Q_\mu(x')\right.\\
&\quad \left. +\lambda^2 [Q_\mu(x'), Q_\nu(x')]\right)^2 /4g^2\\
&= -d^4x' Tr\left(\partial'_\mu Q_\nu(x') - \partial'_\nu Q_\mu(x') + [Q_\mu(x'), Q_\nu(x')]\right)^2 /4g^2\\
&= d^4x' \mathcal{L}(x')
\end{aligned}
\tag{11.143}
$$

and we will simply rename $x' \to x$. Then we will change the functional integration variable

$$
Q^a_\mu \to (A^a_\mu + Q^a_\mu)(x_\nu), \tag{11.144}
$$

where A^a_μ is a classical field, an instanton solution. Expanding the action to second order in Q^a_μ (the first-order variation vanishes after one integration by parts, as A^a_μ satisfies the equations of motion), we get

$$
\begin{aligned}
Z(T) &= S_0^5 \int d(\lambda^2) \int d^4R \int \mathcal{D}(Q^a_\mu) e^{-\int d^4x\, \mathcal{L}(A^a_\mu(x)) + \frac{1}{2} Q\cdot \frac{\delta^2}{\delta Q^2} \mathcal{L}(A^a_\mu(x))\cdot Q} \times\\
&\quad \times \delta^4\left(\int d^4x\, \left(\mathcal{L}(A_\mu + Q_\mu)(\lambda x_\mu)\right)\right) \delta\left(\int d^4x (\mathcal{L}(A_\mu + Q_\mu)(\lambda^2 x^2 - \lambda^2))\right)\\
&= S_0^5 \int \frac{d(\lambda^2)}{\lambda^2} \int \frac{d^4R}{\lambda^4} \int \mathcal{D}(Q^a_\mu) e^{-\int d^4x\, \mathcal{L}(A^a_\mu(x)) + \frac{1}{2} Q\cdot \frac{\delta^2}{\delta Q^2} \mathcal{L}(A^a_\mu(x))\cdot Q} \times\\
&\quad \times \delta^4\left(\int d^4x\, \left(\mathcal{L}(A_\mu + Q_\mu)(x_\mu)\right)\right) \delta\left(\int d^4x (\mathcal{L}(A_\mu + Q_\mu)(x^2 - 1))\right).
\end{aligned}
\tag{11.145}
$$

Now we choose the instanton configuration to be centred on the origin and of unit scale size, *i.e.*

$$
\int d^4x \mathcal{L}(A_\nu) x_\mu = \int d^4x \mathcal{L}(A_\nu)(x^2 - 1) = 0. \tag{11.146}
$$

Then expanding in the first-order Taylor expansion of the action in Q_ν, and using the notation D^A_σ to be the covariant derivative with respect to the classical field A_μ, $D^A_\sigma \cdot = \partial_\sigma \cdot + [A_\sigma, \cdot]$ and the corresponding field strength $F^A_{\sigma\tau}$, we have

$$
\mathcal{L}(A_\nu + Q_\nu) = \mathcal{L}(A_\nu) - \frac{1}{g^2} Tr\left(F^A_{\sigma\tau} D^A_\sigma Q_\tau\right) \tag{11.147}
$$

we get in the delta functions

$$
\delta^4\left(\int d^4x \frac{1}{g^2} Tr\left(F^A_{\sigma\tau} D^A_\sigma Q_\tau x_\mu\right)\right) = \delta^4\left(\int d^4x \frac{1}{g^2} Tr\left(F^A_{\mu\tau} Q_\tau\right)\right) \tag{11.148}
$$

and

$$
\delta\left(\int d^4x \frac{1}{g^2} Tr\left(F^A_{\sigma\tau} D^A_\sigma Q_\tau (x^2 - 1)\right)\right) = \delta\left(-2\int d^4x \frac{1}{g^2} Tr\left(F^A_{\sigma\tau} Q_\tau x_\sigma\right)\right). \tag{11.149}
$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

Thus the delta functions impose the conditions

$$\int d^4x Tr\left(F^A_{\mu\tau}Q_\tau\right) = 0$$
$$\int d^4x Tr\left(x_\sigma F^A_{\sigma\tau}Q_\tau\right) = 0. \tag{11.150}$$

But these conditions are exactly the conditions that the quantum fluctuation Q_ν be orthogonal to the zero modes corresponding to translations and scale transformations, respectively. Indeed, for translations we simply transform the classical solution A_τ by the broken symmetry, translation in the x_σ direction generated by a simple partial derivative in that direction. However, this is not gauge-invariant, hence we also perform a gauge transformation, $\delta_\sigma A_\tau = -D^A_\tau(A_\sigma)$, for an infinitesimal gauge transformation, to give

$$\psi^{tr.}_{\sigma\tau} = \partial_\sigma A_\tau - D^A_\tau(A_\sigma) = F^A_{\sigma\tau}. \tag{11.151}$$

The normalized zero mode is $\hat{\psi}^{tr.}_{\sigma\tau} = \frac{1}{\sqrt{N^{tr.}}}\psi^{tr.}_{\sigma\tau}$, with $N^{tr.}$ defined by

$$N^{tr.} = -\int d^4x Tr(\psi^{tr.}_{\sigma\tau})^2 = -\int d^4x \sum_{\substack{\tau \\ \sigma \text{ fixed}}} Tr\left(F^A_{\sigma\tau}F^A_{\sigma\tau}\right)$$
$$= -\frac{1}{4}\int d^4x \sum_{\sigma,\tau} Tr\left(F^A_{\sigma\tau}F^A_{\sigma\tau}\right) = g^2 S_0. \tag{11.152}$$

We wish to keep track of powers of g and hence we note that the normalization factor does not have any powers of g since $S_0 \sim 1/g^2$. For the scale transformation, the infinitesimal generator is $(1 + x_\sigma\partial_\sigma)A_\tau$ and then the infinitesimal variation of the gauge field, made gauge-invariant, gives the zero mode

$$\psi^{sc.}_\tau = (1 + x_\sigma\partial_\sigma)A_\tau - D^A_\tau(x_\sigma A_\sigma) = x_\sigma F^A_{\sigma\tau}. \tag{11.153}$$

A simple analysis also shows that the normalized zero mode, $\hat{\psi}^{sc.}_\tau = \psi^{sc.}_\tau/\sqrt{N^{sc.}}$, will not have have any net powers of g.

The delta functions impose that the integration over Q should be restricted to the function space that is orthogonal to these zero modes. But in writing Q in a sum over normal modes there is a subtlety involved. We should add a factor of g in the expansion

$$Q_\tau = C^{tr.}_\sigma g\frac{F^A_{\sigma\tau}}{\sqrt{N^{tr.}}} + C^{sc.}g\frac{x_\sigma F^A_{\sigma\tau}}{\sqrt{N^{sc.}}} + \sum_\xi C^\xi g\hat{\psi}^\xi_\tau, \tag{11.154}$$

where $\hat{\psi}^\xi_\tau$ are the non-zero modes. The integration measure is the usual infinite product

$$\mathcal{D}(Q) = \prod_\sigma \frac{dC^{tr.}_\sigma}{\sqrt{2\pi}}\frac{dC^{sc.}}{\sqrt{2\pi}}\prod_\xi \frac{dC^\xi}{\sqrt{2\pi}} \tag{11.155}$$

as long as the same conventions are followed in the numerator and the denominator for the functional integral. The reason for the extra factor of g comes from the exponent in the integrand. We examine the exponent more carefully,

$$\int d^4x\,\frac{1}{2}Q\cdot\frac{\delta^2}{\delta Q^2}\mathcal{L}(A^a_\mu(x))\cdot Q = -\frac{1}{g^2}\frac{1}{2}\int d^4x\,Tr\big(Q_\mu\left(D^A_\nu D^A_\mu - D^A_\sigma D^A_\sigma\delta_{\mu\nu} - F^A_{\mu\nu}\right)Q_\nu\big). \tag{11.156}$$

The using the expansion in Equation (11.154) with its extra factor of g, and where $\hat\psi^\xi_\mu$ explicitly are the normalized non-zero eigenfunctions of the operator of the second-order variation $\left(D^A_\nu D^A_\mu - D^A_\sigma D^A_\sigma\delta_{\mu\nu} - F^A_{\mu\nu}\right)$ with eigenvalues ϵ_ξ, we get

$$\int d^4x\,\frac{1}{2}Q\cdot\frac{\delta^2}{\delta Q^2}\mathcal{L}(A^a_\mu(x))\cdot Q = \frac{1}{2}\sum_\xi \epsilon_\xi\left(C^\xi\right)^2. \tag{11.157}$$

Notice that ϵ_ξ contains no powers of g, nor does the determinant that ensues after integrating over the C^ξ and the zero modes simply drop out in the exponent and the delta functions control the integration over their coefficients. Then the delta function for translations gives

$$\delta^4\left(\frac{1}{g^2}\int d^4x F^A_{\sigma\tau}g\frac{F^A_{\sigma\tau}}{\sqrt{N^{tr.}}}C^{tr.}_\sigma\right) = \delta^4\left(\frac{\sqrt{N^{tr.}}}{g}C^{tr.}_\sigma\right) = \left(\frac{g}{\sqrt{N^{tr.}}}\right)^4\delta^4(C^{tr.}_\sigma). \tag{11.158}$$

The delta function for scale transformation also gives a factor of $\frac{g}{\sqrt{N^{sc.}}}$, giving an overall factor of g^5. Thus writing a prime as usual to indicate the restriction, we get

$$Z(T) = e^{-S_0}S_0^5 g^5\frac{1}{(\sqrt{N^{tr.}})^4}\frac{1}{(\sqrt{N^{sc.}})}\int\frac{d(\lambda^2)}{\lambda^2}\int\frac{d^4R}{\lambda^4}\int \mathcal{D}'(Q^a_\mu)e^{-\int d^4x\,\frac{1}{2}Q\cdot\frac{\delta^2}{\delta Q^2}\mathcal{L}(A^a_\mu(x))\cdot Q} \tag{11.159}$$

with

$$S_0 = \frac{8\pi^2}{g^2}. \tag{11.160}$$

The overall power of g is then g^{-5} due to the five zero modes that we have treated. There are in fact more zero modes associated with the global gauge transformations and the rotation group [112, 98, 24, 12, 65]. As we have seen for $SU(2)$, the diagonal subgroup of these two is unbroken, but the anti-diagonal subgroup is broken with three broken generators. We will not analyse these explicitly, it will suffice to say that they give exactly another power of g^{-3}, for the gauge group $SU(2)$ giving the total, overall factor of g to be g^{-8}. In the case of interest, QCD, the gauge group is $SU(3)$ with eight generators. We imagine putting the instanton solution in an $SU(2)$ subgroup of $SU(3)$, but then one generator always commutes with the $SU(2)$ subgroup. For example, if we generate the subgroup with $\lambda_1,\lambda_2,\lambda_3$, then λ_8 commutes with these three

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

matrices. Thus we find there are only seven broken generators. The $SU(2)$ subgroup mixes in an identical way with the rotation group as in the case when the entire group was $SU(2)$. Thus the upshot is there are seven additional zero modes, which give a factor of g^{-7} and an overall factor of g^{-12}.

We have computed the contribution of a single instanton to the transition between the two vacua $|0\rangle$ and $|1\rangle$. As usual, multi-instanton configurations are negligible except for those corresponding to well-separated single instantons.

$$\begin{aligned} Z(|0\rangle \to |1\rangle) &= \sum_{n,\bar{n}=-\infty}^{\infty} \delta_{n-\bar{n},1} \frac{1}{n!\bar{n}!} \left(\int d^4R \int_0^\infty \frac{d\lambda}{\lambda^5} g^{-12} e^{-\frac{8\pi^2}{g^2}} K \right)^{n+\bar{n}} \\ &= \sum_{n,\bar{n}=-\infty}^{\infty} \delta_{n-\bar{n},1} \frac{1}{n!\bar{n}!} \left(VT \int_0^\infty \frac{d\lambda}{\lambda^5} g^{-12} e^{-\frac{8\pi^2}{g^2}} K \right)^{n+\bar{n}} \end{aligned} \tag{11.161}$$

where K is the determinantal factor including various other normalization factors and constants independent of g. This result directly generalizes to the amplitude

$$Z(|m\rangle \to |\tilde{m}\rangle) = \sum_{n,\bar{n}=-\infty}^{\infty} \delta_{n-\bar{n},m-\tilde{m}} \frac{1}{n!\bar{n}!} \left(VT \int_0^\infty \frac{d\lambda}{\lambda^5} g^{-12} e^{-\frac{8\pi^2}{g^2}} K \right)^{n+\bar{n}} \tag{11.162}$$

and finally to

$$\begin{aligned} Z(\theta) &= \langle\theta|e^{-HT}|\theta\rangle \\ &= \sum_{m,\tilde{m}=-\infty}^{\infty} e^{i\theta(m-\tilde{m})} \langle\tilde{m}|e^{-HT}|m\rangle \\ &= \sum_{m,\tilde{m}-\infty}^{\infty} e^{i\theta(m-\tilde{m})} Z(|m\rangle \to |\tilde{m}\rangle) \\ &= \left(\sum_{\tilde{m}=-\infty}^{\infty} \right) \exp\left(2\cos\theta VT \int_0^\infty \frac{d\lambda}{\lambda^5} g^{-12} e^{-\frac{8\pi^2}{g^2}} K \right). \end{aligned} \tag{11.163}$$

The infinite, constant prefactor is simply a consequence of the plane wave normalization of the theta vacuum states.

Then from Equation (11.65) and Equation (11.66) we get the energy of the ground state

$$E(\theta)/V = -\cos\theta \int_0^\infty \frac{d\lambda}{\lambda^5} g^{-12} e^{-\frac{8\pi^2}{g^2}} K. \tag{11.164}$$

We have purposely left the g-dependent factors inside the integration over λ for a reason, and we have absorbed all constant factors into K. This is because evaluation of the determinants requires renormalization of the coupling constant, and renormalization inserts a scale dependence into g and K. The infinite product of eigenvalues of the operator corresponding to the second variation of the action in the presence of the instanton, Equation (11.156), is not rendered finite when divided by the same infinite product but in the absence of the instanton. We have to add counterterms with infinite coefficients so that the divergences are absorbed. Adding the counterterm proportional to $\sim F_{\mu\nu}F_{\mu\nu}$ exactly renormalizes the value of the coupling constant g. However, the renormalization inserts a dimensionful mass scale, M, into the theory, which fixes the physical, finite, observed value of the coupling constant at that scale. The coupling constant obeys the equation

$$\frac{1}{g^2(\lambda)} = \frac{1}{g^2(M)} - \frac{11}{8\pi^2}\ln(\lambda M) + o(g^2), \tag{11.165}$$

where the λ dependence comes from the simple fact that g is a dimensionless coupling constant, since the only dimensionful parameter that exists, apart from the renormalization scale M, is the instantons scale λ, the two must come together in the dimensionless combination λM. The factor $-\frac{11}{8\pi^2}$ is the famous result of asymptotic freedom for the beta function of QCD, which is a long, hard calculation in perturbation theory [57, 102], which we will not describe here. Asymptotic freedom means that as the scale λ gets smaller $\lambda \ll 1/M$, and the instanton size goes to zero, the coupling constant $g^2(\lambda)$ becomes smaller as $\ln\lambda M$ is negative and the right-hand side becomes larger. Indeed, replacing the solution Equation (11.165) in the expression for the energy gives

$$\begin{aligned}
&E(\theta)/V \\
&= -\cos\theta \int_0^\infty \frac{d\lambda}{\lambda^5}\left(\frac{1}{g^2(M)} - \frac{11}{8\pi^2}\ln(\lambda M)\right)^6 e^{-8\pi^2\left(\frac{1}{g^2(M)} - \frac{11}{8\pi^2}\ln(\lambda M)\right)} K(\lambda M) \\
&= -\cos\theta \int_0^\infty \frac{d\lambda}{\lambda^5} g^{-12}(M) e^{-\frac{8\pi^2}{g^2(M)}}\left(1 + o\left(g^2(M)\ln(\lambda M)\right)\right) e^{11\ln(\lambda M)} K(\lambda M) \\
&= -\cos\theta \int_0^\infty d\lambda \lambda^6 M^{11} g^{-12}(M) e^{-\frac{8\pi^2}{g^2(M)}}\left(1 + o\left(g^2(M)\ln(\lambda M)\right)\right) K(\lambda M).
\end{aligned} \tag{11.166}$$

Thus for small λ, in the ultraviolet, the integral is perfectly convergent; however, in the infrared, as $\lambda \to \infty$ the integral is obviously divergent. Thus the integral is well-behaved in the region where we trust our calculations, when $g \to 0$, but does not make sense in the regions where $g \gg 1$, where we do not trust our calculations. Indeed, we expect new, non-perturbative (not instanton effects, which are also non-perturbative but only valid for small g) strong coupling effects to kick in as g becomes large, effects which we have made no pretence to be able to compute. Thus we stop the calculation at this point, content with the expectation that large coupling, confinement-related effects cure the behaviour of this integral.

11.6 Quarks in QCD

We will next consider the question of quarks in QCD. The quarks come in the fundamental representation of $SU(3)$, which is generated exactly by the 3×3

Gell-Mann matrices of Equation (11.6) and in six flavours, up, down, strange, charm, top and bottom, which we will denote by a label a, and correspond to Dirac fields $\psi_a(x)$. The colour index is suppressed but takes on three values, $1, 2$ and 3, thus the Dirac field is a three-component column for each flavour index. The Lagrangian density in Minkowski spacetime is then given by

$$\mathcal{L} = \frac{1}{4g^2} Tr[F_{\mu\nu}(x)F^{\mu\nu}(x)] + \sum_a \left(\bar{\psi}^a(x) i\gamma^\mu(\partial_\mu + A_\mu)\psi^a(x) - m^a\bar{\psi}^a(x)\psi^a(x)\right), \tag{11.167}$$

where the gauge field is a 3×3 anti-hermitean matrix, $A_\mu = iA^a_\mu\lambda_a$ $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$ is also anti-hermitean. The γ^μ are the usual Dirac matrices. The masses m^α are quite small for the up and down quarks, less than 10 MeV. Thus the massless limit is a reasonably good approximation when considering processes that largely imply only the up and down quarks. This limit has a higher symmetry, called chiral symmetry which is spontaneously broken, and can be treated with chiral perturbation theory. The strange quark mass is a little more, around 95 MeV, but still within the purview chiral symmetry and chiral perturbation theory. The charm mass is about 1.3 GeV, the bottom mass is 4.2 GeV, and the top mass is 173 GeV. Neglecting these masses is not a good approximation. In what follows, we will restrict our considerations to the up and down quarks and neglect their masses, which is a rather good approximation. Then the fermionic part of the Lagrangian density is

$$\mathcal{L} = \sum_{a=\text{up,down}} \bar{\psi}^a(x) i\gamma_\mu(\partial_\mu + A_\mu)\psi^a(x). \tag{11.168}$$

The Lagrangian in this case has a symmetry $SU_L(2) \times SU_R(2)$, called chiral symmetry. The subscripts L and R correspond to independent $SU(2)$ transformations on the left-handed and right-handed components of the Dirac spinor. If we write the spinor fields as

$$\psi = \begin{pmatrix} \psi_u \\ \psi_d \end{pmatrix} \tag{11.169}$$

then the chiral transformation corresponds to

$$\psi \to e^{i\left(\frac{1-\gamma_5}{2}\right)\vec{\alpha}_L\cdot\vec{\sigma}} e^{i\left(\frac{1+\gamma_5}{2}\right)\vec{\alpha}_R\cdot\vec{\sigma}}\psi, \tag{11.170}$$

where $\vec{\alpha}_{L,R}$ are independent parameters of the two $SU(2)$ transformations and the $\vec{\sigma}$ are the Pauli matrices. The chiral projection operators, $\frac{1\pm\gamma_5}{2}$, project onto the left-handed and right-handed components of the Dirac spinor

$$\psi = \psi_L + \psi_R = \left(\frac{1-\gamma_5}{2}\right)\psi + \left(\frac{1+\gamma_5}{2}\right)\psi. \tag{11.171}$$

In the chiral representation of the Dirac matrices,

$$\gamma_5 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{11.172}$$

so that

$$\frac{1-\gamma_5}{2} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \frac{1+\gamma_5}{2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{11.173}$$

The Lagrangian Equation (11.169) is also invariant under two $U(1)$ symmetries, $U_V(1) \times U_A(1)$,

$$\psi \to e^{i\theta} e^{i\gamma_5 \theta'} \psi \tag{11.174}$$

for independent parameters θ and θ'. The corresponding conserved currents, via Noether's theorem, are denoted

$$j^\mu = \bar{\psi}\gamma^\mu\psi \quad \text{and} \quad j_5^\mu = \bar{\psi}\gamma^\mu\gamma_5\psi. \tag{11.175}$$

The chiral symmetry group $SU_L(2) \times SU_R(2)$ is spontaneously broken to the diagonal subgroup $SU_D(2)$, which is identified as the isospin group, in this case with just two flavours, up and down. Due to this spontaneous symmetry-breaking, the Goldstone theorem [56] implies the existence of three Goldstone bosons, massless scalar fields, which are then identified with the pions. The pions are not massless, however; the mass terms for the up and down quarks softly but explictly break the chiral symmetry. The consequence of this explict breaking is to give the putative Goldstone bosons, the pions, a small mass. This analysis is called chiral perturbation theory [117]. The $U_V(1)$ symmetry corresponds to the baryonic charge and is presumed to be conserved. The one question that remains is what happens to the $U_A(1)$, how does this symmetry manifest itself? If it is not broken, spontaneously or explicitly, then it should be associated with a conserved quantum number. We do not see any such additional conserved quantum number. If it is spontaneously broken, then we should see another corresponding massless Goldstone boson. It can be shown that this does not correspond to the η, Weinberg has shown [118] that the mass of such a putative Goldstone bosons must satisfy the inequality $m_{G.B.} \leq \sqrt{m}_\pi$. This lack of understanding of how the $U_A(1)$ symmetry manifests itself is called the $U(1)$ problem.

The $U(1)$ problem is related to instantons, the theta vacua and the chiral anomaly, which we will explain in this section and the next. The upshot is that the $U_A(1)$ symmetry is actually explicitly broken, due to a quantum effect, called the chiral anomaly. To understand the chiral anomaly it is easiest to work in a

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

function integral formulation for the fermionic fields, the subject to which we will now turn.

11.6.1 Quantum Fermi Fields

Canonical quantization of fermionic fields demands that the fermions satisfy equal time anti-commutation relations

$$\left\{\psi(\vec{x},t), i\psi^\dagger(\vec{y},t)\right\} = -i\hbar\delta^3(\vec{x}-\vec{y}), \tag{11.176}$$

where $\{A,B\} = AB + BA$. Why does the anti-commutator arise? For a free field with equation of motion

$$(i\gamma^\mu\partial_\mu - m)\psi = 0 \tag{11.177}$$

we can construct the solution by simple Fourier transformation

$$\psi = \int \frac{d^3k}{(2\pi)^3}\frac{m}{(\vec{k}^2+m^2)^{1/2}}\sum_{\alpha=1,2}\left(b_\alpha(k)u^\alpha(k)e^{-ik\cdot x} + d_\alpha^\dagger(k)v^\alpha(k)e^{ik\cdot x}\right), \tag{11.178}$$

where $k \equiv ((\vec{k}^2+m^2)^{1/2}, \vec{k})$, $u^\alpha(k)$ and $v^\alpha(k)$ are specific, orthonormalized spinor solutions of the Dirac equation (11.177) of positive and negative energy, respectively, while $b_\alpha(k)$ and $d^\dagger(k)$ are arbitrary, operator valued coefficients. The expression for the Hamiltonian (energy) then becomes

$$\mathcal{H} = \int \frac{d^3k}{(2\pi)^3}(\vec{k}^2+m^2)^{1/2}\sum_\alpha\left(b_\alpha^\dagger(k)b_\alpha(k) - d_\alpha(k)d_\alpha^\dagger(k)\right), \tag{11.179}$$

where we have not changed the order of the operators in the expression for the Hamiltonian. The order of the d_α and the $d_\alpha^\dagger$ occurs because we expanded ψ with $d_\alpha^\dagger$ rather than d_α but the minus sign occurs because the v^αs correspond to negative energy solutions of the Dirac equation. If we had used d_α in the expansion of ψ we would have arrived at the expression in Equation (11.179) with the d_α and the $d_\alpha^\dagger$ interchanged; however, the minus sign would still be there. Now if we want to have $\mathcal{H} \geq 0$, up to a constant, we need

$$d_\alpha(k)d_\alpha^\dagger(k) = -d_\alpha^\dagger(k)d_\alpha(k) + 1, \tag{11.180}$$

where we have chosen the constant to be 1 for the case of discrete k. For a continuum of k's we get

$$\{d_\alpha(k), d_\alpha^\dagger(k')\} = \delta^4(k-k') \tag{11.181}$$

and the Hamiltonian, up to a constant (which can very well be an infinite constant!) is

$$\mathcal{H} = \int \frac{d^3k}{(2\pi)^3}(\vec{k}^2+m^2)^{1/2}\sum_\alpha\left(b_\alpha^\dagger(k)b_\alpha(k) + d_\alpha^\dagger(k)d_\alpha(k)\right), \tag{11.182}$$

a positive semi-definite form. As the b_αs and the d_αs are equivalent, we must choose anti-commutation relations for both.

11.6.2 Fermionic Functional Integral

The limit $\hbar \to 0$ in the canonical anti-commutation relations Equation (11.176) does not yield ordinary, commuting c-number fields in the classical limit. The fields become anti-commuting fields, so-called Grassmann number-valued fields whose anti-commutator, rather than commutator, vanishes

$$\{\psi(\vec{x},t), \psi^\dagger(\vec{y},t)\} = 0. \tag{11.183}$$

Thus the classical limit gives fields that are elements of an infinite dimensional Grassmann algebra from an infinite dimensional Clifford algebra in the quantum domain. Then, if there is to be a Feynman path-integral description of fermions, the integral should be defined over the classical space of fields, fields that are Grassmann algebra-valued. Such an integral can be formally defined. For free theories, perhaps all such formalism is rather unnecessary. However, for interacting theories of fermions, the functional integral description must at least be able to generate the perturbative expansion. In fact, we can almost think that the fermionic functional integral representation for the amplitudes of a quantum field theory with fermions is simply a very compact and efficient notation that can and does serve as a means of generating the perturbative expansion.

Abstractly, an integral is a linear map that takes a space of functions to the real numbers. We will define the functional integral over a Grassmann number in this way, first for a finite set of Grassmann numbers, and then generalize to the infinite limit. A Grassmann number θ satisfies

$$\{\theta, \theta\} = \theta\theta + \theta\theta = 2\theta\theta = 2\theta^2 = 0. \tag{11.184}$$

We define a differential operator $\frac{d}{d\theta}$ by the very reasonable rules for any other anti-commuting number β and for a c-number a,

$$\frac{d}{d\theta}\theta = 1, \quad \frac{d}{d\theta}\beta = \frac{d}{d\theta}a = 0. \tag{11.185}$$

The derivative operator should be thought of as a Grassmann-valued operator; it should anti-commute with other Grassmann numbers. A general function of $f(\theta)$, *i.e.* a commutative function, can be expanded in two terms

$$f(\theta) = a + \beta\theta, \tag{11.186}$$

where a is real while β is Grassmannian. Then

$$\frac{d}{d\theta}f(\theta) = \frac{d}{d\theta}a + \frac{d}{d\theta}\beta\theta = -\frac{d}{d\theta}\theta\beta = -\beta. \tag{11.187}$$

The idea that f is a commutative function means that it is composed of an even number of Grassmann numbers, 0 and 2 in this case. β is a Grassmann number, hence

$$\beta^2 = 0, \quad \{\beta, \theta\} = 0 \tag{11.188}$$

then it is easy to verify

$$[f,\beta]=[f,\theta]=0. \tag{11.189}$$

Then clearly

$$\frac{d}{d\theta}\frac{d}{d\theta}=\frac{d^2}{d\theta^2}=0. \tag{11.190}$$

This means that the integral can in no way be the inverse of differentiation, the derivative is a nilpotent operator. However, we will define it, following Berezin [13] to be a linear map from the space of Grassmann numbers to the real numbers, we define

$$\int d\theta\, 1=0 \quad \int d\theta\,\theta=1, \tag{11.191}$$

which implies

$$\int d\theta\, f(\theta)=\int d\theta\,(a+\beta\theta)=0+\int d\theta\,\beta\theta=-\int d\theta\,\theta\beta=-\beta. \tag{11.192}$$

For N Grassmann numbers we have the algebra

$$\begin{aligned}\{\theta_i,\theta_j\}&=0\\ \left\{\frac{d}{d\theta_i},\theta_j\right\}&=\delta_{ij}\\ \left\{\frac{d}{d\theta_i},\frac{d}{d\theta_j}\right\}&=0,\end{aligned} \tag{11.193}$$

for $i,j=1,\cdots N$. Then a general, commutative function is expanded as

$$f(\theta_i)=a+c_i\theta_i+c_{ij}\theta_i\theta_j+\cdots+c\theta_1\theta_2\cdots\theta_N \tag{11.194}$$

and we notice it has a finite number of terms. $c_{ijkl\cdots}$ is Grassmannian if the number of indices is odd but a real number if the number of indices is even. The integration rules generalize as

$$\int d\theta_i\, 1=0, \quad \int d\theta_i\,\theta_j=\delta_{ij} \tag{11.195}$$

and by convention and consistency

$$\int d\theta_1 d\theta_2\,\theta_1\theta_2\equiv\int d\theta_1\left(\int d\theta_2(-\theta_2)\right)\theta_1=-\int d\theta_1\,\theta_1=-1. \tag{11.196}$$

Then it is easy to see for an anti-symmetric matrix M (clearly any symmetric part of M will not contribute)

$$I_N(M)=\int d\theta_1\cdots d\theta_N e^{-\sum_{ij}\theta_i M_{ij}\theta_j}=\begin{cases}2^{N/2}\sqrt{det(M)}, & \text{for } N \text{ even}\\ 0 & \text{for } N \text{ odd}\end{cases}. \tag{11.197}$$

For an invertible, anti-symmetric M and a set of Grassmann parameters χ_i, $i=1\cdots N$ and $\{\chi_i,\chi_j\}=\{\chi_i,\theta_j\}=0$, we can compute

$$\mathcal{I}_N(M;\chi)=\int d\theta_1\cdots d\theta_N e^{-\sum_{ij}\theta_i M_{ij}\theta_j+\sum_j\chi_j\theta_j} \tag{11.198}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

as follows. Translating the integration variable $\theta_i = \theta'_i - \frac{1}{2}M^{-1}_{ij}\chi_j$, we get, using matrix notation

$$\begin{aligned}
\theta^T M\theta - \chi^T\theta &= \left(\theta' - \frac{1}{2}M^{-1}\chi\right)^T M\left(\theta' - \frac{1}{2}M^{-1}\chi\right) - \chi^T\left(\theta' - \frac{1}{2}M^{-1}\chi\right)\\
&= \theta'^T M\theta' - \frac{1}{2}(\chi^T {M^{-1}}^T M\theta' + \theta'^T\chi) + \frac{1}{4}\chi^T {M^{-1}}^T\chi\\
&\quad -\chi^T\theta' + \frac{1}{2}\chi^T M^{-1}\chi\\
&= \theta'^T M\theta' - \frac{1}{2}(-\chi^T\theta' - \chi^T\theta') - \frac{1}{4}\chi^T M^{-1}\chi\\
&\quad -\chi^T\theta' + \frac{1}{2}\chi^T M^{-1}\chi\\
&= \theta'^T M\theta' + \frac{1}{4}\chi^T M^{-1}\chi
\end{aligned} \tag{11.199}$$

using ${M^{-1}}^T M = (M^T M^{-1})^T = (-MM^{-1})^T = -\mathbb{I}^T = -\mathbb{I}$ since M is anti-symmetric and the fact that the χ is also anti-commuting. Then

$$\begin{aligned}
\mathcal{I}_N(M;\chi) &= \int d\theta'_1 \cdots d\theta'_N e^{-\sum_{ij}\left(\theta'_i M_{ij}\theta'_j - \frac{1}{4}\chi_i M^{-1}\chi_j\right)}\\
&= \begin{cases} 2^{N/2}\sqrt{detM}\, e^{-\sum_{ij}\frac{1}{4}\chi_i M^{-1}\chi_j}, & \text{for } N \text{ even}\\ 0 & \text{for } N \text{ odd}\end{cases}.
\end{aligned} \tag{11.200}$$

For complex fields, we have the equivalent of complex Grassmann numbers

$$\eta = \frac{\theta_1 + i\theta_2}{\sqrt{2}}, \quad \eta^* = \frac{\theta_1 - i\theta_2}{\sqrt{2}}. \tag{11.201}$$

Considering the 2×2 case, we impose $-\sum_{i,j}\theta_i M_{ij}\theta_j = i\eta^*\tilde{M}\eta$ which gives $\tilde{M} = 2M_{12}$ and we have

$$\int d\eta^* d\eta e^{i\eta^*\tilde{M}\eta} = det(\tilde{M}), \tag{11.202}$$

where the integration is done by treating η and η^* as completely independent Grassmann variables. Dropping the tilde, the integration formula generalizes as

$$\int \prod_{i,j} d\eta_i^* d\eta_j e^{i\sum_{i,j}\eta_i^* M_{ij}\eta_j} = det(M) \tag{11.203}$$

and with sources, suppressing the indices and summation signs,

$$\int \prod d\eta^* d\eta e^{i\eta^* M\eta + i\xi^*\eta + i\eta^*\xi} = det(M)e^{-i\xi^* M^{-1}\xi}. \tag{11.204}$$

Then boldly generalizing to infinite dimensional integrals we get for the fermionic field

$$\int \mathcal{D}(\psi,\bar{\psi})e^{i\int d^4x\bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi} = det(i\gamma^\mu\partial_\mu - m) \tag{11.205}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

and including sources

$$\int \mathcal{D}(\psi,\bar{\psi})e^{i\int d^4x\bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi+\bar{\psi}\zeta+\bar{\zeta}\psi} = det(i\gamma^\mu\partial_\mu - m)e^{-i\int d^4x\bar{\zeta}(i\gamma^\mu\partial_\mu - m)^{-1}\zeta}$$
$$= \mathcal{N}' e^{-i\int d^4p\bar{\zeta}(p)\frac{1}{\not p - m}\zeta(p)}. \tag{11.206}$$

Then for a general gauge interaction the usual perturbative expansion ensues from the coupling (where we have expressly put the coupling constant e)

$$\mathcal{L}' = ie\bar{\psi}\gamma^\mu A_\mu \psi \tag{11.207}$$

then

$$Z(\zeta, A) = \int \mathcal{D}(\psi,\bar{\psi})e^{i\int d^4x\mathcal{L}+\bar{\zeta}\psi+\bar{\psi}\zeta+ie\bar{\psi}\gamma^\mu A_\mu\psi}$$
$$= \mathcal{N}' e^{i\int d^4x\left(\frac{\delta}{\delta\zeta(x)}ie\gamma^\mu A_\mu\frac{\delta}{\delta\bar{\zeta}(x)}\right)} e^{-i\int d^4p\bar{\zeta}(p)\frac{1}{\not p - m}\zeta(p)}. \tag{11.208}$$

The derivatives with respect to $\zeta(x)$ and $\bar{\zeta}(x)$ can be trivially converted into derivatives with respect to $\zeta(p)$ and $\bar{\zeta}(p)$ by Fourier transformation. This gives rise to the usual perturbation expansion expressed in Feynman diagrams [101].

Reverting back to Euclidean space, the action is

$$S_E = -\int d^4x\bar{\psi}(i\gamma_\mu\partial_\mu - im)\psi, \tag{11.209}$$

where the γ_μ matrices satisfy the Clifford algebra

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}. \tag{11.210}$$

The fields $\bar{\psi}$ and ψ are no longer related to each other, but are in fact completely independent Grassmann-valued fields. We can infer this from many points of view. First, and most importantly, if the formula Equation (11.203) is to work, the integration variables η and η^* are completely independent. First of all, η and η^* satisfy

$$\{\eta_i, \eta_j\} = \{\eta_i^*, \eta_j^*\} = \{\eta_i^*, \eta_j\} = 0. \tag{11.211}$$

Then if η^* were the adjoint of η, *i.e.* $\eta^* = \eta^\dagger C$, where C is a fixed matrix akin to a charge conjugation matrix, then the last relation would imply (multiplying by C^{-1}) and contracting together

$$\sum_i (\eta_i\eta_i^\dagger + \eta_i^\dagger\eta_i) = 0. \tag{11.212}$$

This says that the sum of two positive operators vanishes, requiring the operators to be zero. Additionally, the Euclidean Dirac fields transform according to the $(\frac{1}{2},0)\oplus(0,\frac{1}{2})$ representation of the four-dimensional Euclidean rotation group $SO(4) = SU(2)\times SU(2)$. The two $SU(2)$ subgroups are totally independent of one another, hermitean conjugation does not take one into the other, as is the case in Minkowski spacetime.

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

We decompose ψ and $\bar{\psi}$ with a complete set of orthonormal spinor solutions of the Dirac equation

$$\psi = \sum_r a_r \psi_r \quad \bar{\psi} = \sum_r \bar{a}_r \bar{\psi}_r, \tag{11.213}$$

where the coefficients a_r and $\bar{a}_r$ are independent Grassmann numbers and

$$\int d^4x \psi_r^\dagger \psi_s = \int d^4x \bar{\psi}_r \bar{\psi}_s^\dagger = \delta_{rs}. \tag{11.214}$$

Then we define the functional integration measure as

$$\mathcal{D}(\psi, \bar{\psi}) = \prod_r da_r d\bar{a}_r. \tag{11.215}$$

Then

$$S_E = -\int d^4x \bar{\psi}(i\gamma_\mu \partial_\mu)\psi = -\sum_r \lambda_r \bar{a}_r a_r, \tag{11.216}$$

where

$$(i\gamma_\mu \partial_\mu)\psi_r = \lambda_r \psi_r. \tag{11.217}$$

Then the integral

$$\int \mathcal{D}(\psi, \bar{\psi}) e^{-S_E} = \int \prod_s da_s d\bar{a}_s e^{\sum_r \lambda_r \bar{a}_r a_r} = \prod_r \lambda_r = det(i\gamma_\mu \partial_\mu). \tag{11.218}$$

In the massless limit, $m \to 0$, the action Equation (11.209) is invariant under global chiral transformations, decomposed as vector and axial transformations

$$\psi \to e^{i(\alpha + \beta\gamma_5)}\psi \quad \bar{\psi} \to \bar{\psi} e^{i(\alpha + \beta\gamma_5)}. \tag{11.219}$$

The chiral anomaly corresponds to the fact that it is impossible to define the functional integral while simultaneously keeping the axial gauge symmetry and the vector gauge symmetry. The full chiral symmetry of two-flavour QCD is

$$SU_V(2) \times SU_A(2) \times SU_V(1) \times U_A(1), \tag{11.220}$$

but the $SU_A(2)$ is spontaneously broken, giving rise to three massless Goldstone bosons, the pions,

$$SU_V(2) \times SU_A(2) \times U_V(1) \times U_A(1) \to SU_V(2) \times U_V(1) \times U_A(1). \tag{11.221}$$

The anomaly results from the impossibility to preserve the remaining chiral symmetry in the quantum theory. Fundamentally, the anomaly results because of divergences in the naive, original theory. Then to make sense of the theory these divergences must be removed; this is done in a rather brutal fashion and is called regularization. The brutality of the regularization means that it seems necessary to explicitly break at least some of the chiral symmetry of the original Lagrangian. Indeed, there is no known regularization that can preserve all of the chiral symmetry and it is understood that no such regularization exists. The hope and expectation was that once the regularization is removed, the full chiral symmetry of the theory would return. The anomaly corresponds to the fact that this is not the case. In fact, upon removing the regularization it is not possible to preserve both the $U_V(1)$ and the $U_A(1)$ symmetries.

11.6.3 The Axial Anomaly

Conceptually, the clearest method for seeing this was discovered by Fujikawa [50][3]. He considered the fermionic functional integral

$$\mathcal{I} = \int \mathcal{D}(\psi, \bar{\psi}) e^{\int d^4x \bar{\psi}(i\gamma_\mu D_\mu)\psi} \tag{11.222}$$

and realized that the anomaly comes from the inability to define the functional integration measure in a manner that is invariant under all the chiral transformations. Here we generalize further, allowing the covariant derivative to include gauge fields, in principle, for all the global symmetries. However, we will find that some global symmetries are not preserved in the quantum theory, and then adding the gauge fields corresponding to those symmetries is inopportune. Their quantization makes no sense as renormalizability requires gauge-invariance. Thus we imagine adding gauge fields for all symmetries that can be preserved at the quantum level. For the case of QCD, this corresponds to gauge fields for the colour gauge symmetry $SU_c(3)$ and the $U_V(1)$ symmetry. Gauging the chiral $SU_V(2) \times SU_A(2)$ actually corresponds to part of the gauge group of the weak interactions, but we shall not develop this theory here. We will expand the fields slightly differently from Equation (11.213) as

$$\psi = \sum_r a_r \varphi_r \quad \bar{\psi} = \sum_r \varphi_r^\dagger \bar{a}_r \tag{11.223}$$

with

$$i\not{D}\varphi_r = \lambda_r \varphi_r \quad \int d^4x \varphi_r^\dagger \varphi_s = \delta_{rs} \tag{11.224}$$

and

$$\mathcal{D}(\psi, \bar{\psi}) = \prod_r d\bar{a}_r da_r. \tag{11.225}$$

For a local axial transformation $\psi(x) \to e^{i\beta(x)\gamma_5}\psi(x) \approx \psi(x) + i\beta(x)\gamma_5\psi(x)$ and $\bar{\psi}(x) \to \bar{\psi}(x)e^{i\beta(x)\gamma_5} \approx \bar{\psi}(x) + i\beta(x)\bar{\psi}(x)\gamma_5$ the Lagrangian is not invariant

$$\mathcal{L}(x) \to \mathcal{L}(x) - (\partial_\mu \beta(x))\bar{\psi}(x)\gamma_\mu\gamma_5\psi(x). \tag{11.226}$$

However,

$$\psi(x) \to \psi'(x) = \sum_r a'_r \varphi_r(x) = \sum_r a_r e^{i\beta(x)\gamma_5}\varphi_r(x). \tag{11.227}$$

[3] We note that Fujikawa used anti-hermitean Euclidean Dirac matrices. We stick with hermitean Euclidean Dirac matrices.

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

Then

$$a'_r = \sum_s \int d^4x \varphi_r^\dagger e^{i\beta(x)\gamma_5} \varphi_s(x) a_s \equiv \sum_s C_{rs} a_s \tag{11.228}$$

and

$$\prod_r da'_r = (detC)^{-1} \prod_s da_s. \tag{11.229}$$

Interestingly, the power of the determinant is -1. This is because the Grassmann integration actually behaves a lot like differentiation. Indeed,

$$1 = \int d(\lambda a)\,(\lambda a) = \int da\, J(\lambda a) = J\lambda \int da\, a = J\lambda, \tag{11.230}$$

thus $J = 1/\lambda$ and $d(\lambda a) = da/\lambda$. The determinant of the matrix C_{rs} for an infinitesimal transformation is

$$\begin{aligned}
&= det(C_{rs})^{-1}\ det\left(\delta_{rs} + i\int d^4x \beta(x)\varphi_r^\dagger(x)\gamma_5\varphi_s(x)\right)^{-1} \\
&= \exp\left(-Tr \ln\left(\delta_{rs} + \int d^4x\, \beta\varphi_r^\dagger\gamma_5\varphi_s\right)\right) \\
&= \exp\left(-i\int d^4x \beta(x)\sum_r \varphi_r^\dagger\gamma_5\varphi_r\right)
\end{aligned} \tag{11.231}$$

using the expansion of $\ln(1+\epsilon) \approx \epsilon$. We must not forget that an equal contribution will come from the variation of $\prod_r d\bar{a}_r$. We wish to evaluate $A(x) = \sum_r \varphi_r^\dagger(x)\gamma_5\varphi_r(x)$; however, the sum is surely hopelessly divergent. We regularize it with the eigenvalues of the Dirac operator, taking

$$A(x) \equiv \lim_{M\to\infty} \sum_r \varphi_r^\dagger(x)\gamma_5 e^{-(\lambda_r/M)^2}\varphi_r(x). \tag{11.232}$$

It is this choice of regulator that puts the anomaly in the axial symmetry, preserving the vector symmetry in the quantum theory. Another choice can preserve the axial symmetry but not the vector. We can choose which symmetry we wish to preserve; however, we cannot preserve both. Writing $\varphi_r(x) = \langle x|r\rangle$, (note the ket $|r\rangle$ must span the matrix indices of the coordinate wave function $\varphi(x)$) we have

$$\begin{aligned}
A(x) &= \lim_{M\to\infty}\sum_r \langle r|x\rangle\gamma_5 e^{-(\lambda_r/M)^2}\langle x|r\rangle = \lim_{M\to\infty}\sum_r Tr\left(\gamma_5\langle x|e^{-(\lambda_r/M)^2}|r\rangle\langle r|x\rangle\right) \\
&= \lim_{M\to\infty}\sum_r Tr\left(\gamma_5\langle x|e^{-(i\not{D}/M)^2}|r\rangle\langle r|x\rangle\right) = \lim_{M\to\infty}\lim_{x\to y} Tr\left(\gamma_5\langle x|e^{-(i\not{D}/M)^2}|y\rangle\right) \\
&= \lim_{M\to\infty}\lim_{x\to y} Tr\left(\gamma_5\langle x|e^{-(i\not{D}/M)^2}\int\frac{d^4k}{(2\pi)^4}|k\rangle\langle k|y\rangle\right) \\
&= \lim_{M\to\infty}\lim_{x\to y} Tr\left(\gamma_5 e^{-(i\not{D}(x)/M)^2}\langle x|\int\frac{d^4k}{(2\pi)^4}|k\rangle\langle k|y\rangle\right) \\
&= \lim_{M\to\infty}\lim_{x\to y} Tr\left(\gamma_5\int\frac{d^4k}{(2\pi)^4}e^{-(i\not{D}(x)/M)^2}e^{ik\cdot x}e^{-ik\cdot y}\right)
\end{aligned} \tag{11.233}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

and we should be aware that the Tr is over Dirac and internal indices. Now $i\not{D}(x)e^{ik\cdot x} = e^{ik\cdot x}(-\not{k} + i\not{D}(x))$, thus

$$\begin{aligned} A(x) &= \lim_{M\to\infty}\lim_{x\to y} Tr\left(\gamma_5 \int \frac{d^4k}{(2\pi)^4} e^{-ik\cdot y} e^{ik\cdot x} e^{-((-\not{k}+i\not{D}(x))/M)^2}\right) \\ &= \lim_{M\to\infty} Tr\left(\gamma_5 \int \frac{d^4k}{(2\pi)^4} e^{-((-\not{k}+i\not{D}(x))/M)^2}\right) \\ &= \lim_{M\to\infty} Tr\left(\gamma_5 \int \frac{d^4k}{(2\pi)^4} e^{-(\frac{\{\gamma_\mu,\gamma_\nu\}}{2}+\frac{[\gamma_\mu,\gamma_\nu]}{2})(-k_\mu+iD_\mu(x))(-k_\nu+iD_\nu(x))/M^2}\right) \\ &= \lim_{M\to\infty} Tr\left(\gamma_5 \int \frac{d^4k}{(2\pi)^4} e^{-(\delta_{\mu\nu}+\frac{[\gamma_\mu,\gamma_\nu]}{2})(-k_\mu+iD_\mu(x))(-k_\nu+iD_\nu(x))/M^2}\right) \\ &= \lim_{M\to\infty} Tr\left(\gamma_5 \int \frac{d^4k}{(2\pi)^4} e^{-\left((-k+iD(x))^2-\frac{[\gamma_\mu,\gamma_\nu]}{2}\frac{F_{\mu\nu}}{2}\right)/M^2}\right) \\ &= \lim_{M\to\infty} Tr\left(\gamma_5 \int \frac{d^4k}{(2\pi)^4} e^{-k^2/M^2} e^{\left(2ik\cdot D(x)+D^2(x)+\frac{1}{4}[\gamma_\mu,\gamma_\nu]F_{\mu\nu}\right)/M^2}\right) \\ &= \lim_{M\to\infty} Tr\left(\gamma_5 \int \frac{d^4k}{(2\pi)^4} e^{-k^2/M^2}\left(1+\cdots+\frac{1}{2}\left(\frac{1}{4}[\gamma_\mu,\gamma_\nu]F_{\mu\nu}\right)^2/M^4+\cdots\right)\right). \end{aligned} \tag{11.234}$$

The first term in the expansion of the exponential that survives the Dirac trace is shown and, although there will be other terms in the higher orders that survive this trace, they will have higher powers of M in the denominator. The Gaussian integral only gives a factor of M^4, hence in the limit $M \to \infty$ this is the only term that survives. Thus we get

$$A(x) = \lim_{M\to\infty}\int \frac{d^4k}{(2\pi)^4} Tr\left(\frac{1}{2}\frac{1}{M^4}\left(\gamma_5\frac{1}{4}[\gamma_\mu,\gamma_\nu]\frac{1}{4}[\gamma_\sigma,\gamma_\tau]\right)F_{\mu\nu}F_{\sigma\tau}\right)e^{-k^2/M^2}. \tag{11.235}$$

Using that $Tr(\gamma_5\gamma_\mu\gamma_\nu\gamma_\sigma\gamma_\tau) = 4\epsilon_{\mu\nu\sigma\tau}$ and that the Gaussian integral is

$$\int \frac{d^4k}{(2\pi)^4} e^{-k^2/M^2} = \frac{M^4}{16\pi^2} \tag{11.236}$$

gives

$$A(x) = \frac{1}{32\pi^2}\epsilon_{\mu\nu\sigma\tau} Tr\left(F_{\mu\nu}F_{\sigma\tau}\right). \tag{11.237}$$

Therefore, the fermionic functional integration measure is not invariant under axial transformations and transforms as

$$\mathcal{D}(\psi,\bar{\psi}) \to \mathcal{D}(\psi,\bar{\psi})e^{-i\frac{1}{16\pi^2}\int d^4x\,\beta(x)\epsilon_{\mu\nu\sigma\tau}Tr\left(F_{\mu\nu}F_{\sigma\tau}\right)}, \tag{11.238}$$

where we get twice the variation since both ψ and $\bar{\psi}$ contribute to the measure.

11.6.4 The U(1) Problem

With the chiral anomaly, we understand that, at the quantum level, we cannot preserve all of the classical symmetry of the action. Only the subgroup $U_A(1)$ is explicitly broken by the chiral anomaly and the remaining subgroup is spontaneously broken $SU_V(2)\times SU_A(2)\times U_V(1)\times SU_c(3)\rightarrow SU_V(2)\times U_V(1)\times SU_c(3)$. Under a local axial $U_A(1)$ transformation, $-S_E$, minus the action (that appears in the exponent) transforms as, keeping only terms to first order,

$$\begin{aligned}\int d^4x\bar{\psi}\left(i\gamma_\mu D_\mu\right)\psi \rightarrow &\int d^4x\bar{\psi}(1+i\beta(x)\gamma_5)\left(i\gamma_\mu D_\mu\right)(1+i\beta(x)\gamma_5)\psi\\ &=\int d^4x\bar{\psi}\left(i\gamma_\mu D_\mu\right)\psi-(\partial_\mu\beta(x))\bar{\psi}\gamma_\mu\gamma_5\psi.\end{aligned}\tag{11.239}$$

The functional integral under a change of variables must be invariant. If we transform to the field $\psi'=(1+i\beta(x)\gamma_5)\psi$, we get

$$\begin{aligned}\mathcal{I}&=\int\mathcal{D}(\psi,\bar{\psi})e^{\int d^4x\bar{\psi}i\not{D}\psi}=\int\mathcal{D}(\psi',\bar{\psi}')e^{\int d^4x\bar{\psi}i\not{D}\psi-(\partial_\mu\beta(x))\bar{\psi}\gamma_\mu\gamma_5\psi}\\ &=\int\mathcal{D}(\psi,\bar{\psi})e^{-i\int d^4x\beta(x)\left(\frac{1}{16\pi^2}\epsilon_{\mu\nu\sigma\tau}Tr\left(F_{\mu\nu}F_{\sigma\tau}\right)+i\partial_\mu\bar{\psi}\gamma_\mu\gamma_5\psi\right)}e^{\int d^4x\bar{\psi}i\not{D}\psi}.\end{aligned}\tag{11.240}$$

Then invariance requires

$$\partial_\mu\langle\bar{\psi}\gamma_\mu\gamma_5\psi\rangle^A=i\frac{1}{16\pi^2}\epsilon_{\mu\nu\sigma\tau}Tr\left(F_{\mu\nu}F_{\sigma\tau}\right)=i\frac{C}{8\pi^2}\langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle,\tag{11.241}$$

where the matrix element on the left-hand side signifies the fermionic expectation value of the axial current operator $\bar{\psi}\gamma_\mu\gamma_5\psi$ in the presence of the background gauge fields. The latter equality is easily obtained for an arbitrary multiplet of fermions in a representation with hermitean generators T^a of $SU(n)$, and then C is the constant in $Tr\left(T^aT^b\right)=C\delta^{ab}$. The i on the right-hand side is expected and disappears upon Wick rotation back to Minkowski space.

We can demonstrate the so-called chiral Ward–Takahashi identities, which have to do with symmetries, and will be useful in our analysis later. Consider the m point function

$$\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle^A\equiv\frac{\int\mathcal{D}(\psi,\bar{\psi})\,e^{-S_E(A)}\phi^1(x_1)\cdots\phi^m(x_m)}{\int\mathcal{D}(\psi,\bar{\psi})\,e^{-S_E(A)}},\tag{11.242}$$

where the $\phi^i(x_i)$ are local, multi-linear functions of the fermionic fields where S_E is given in Equation (11.209). With the variations (taken in the opposite sense to Fujikawa as in Equation (11.239), to stay with the conventions of Coleman)

$$\delta\psi=-i\gamma_5\psi\delta\alpha(x)\quad\delta\bar{\psi}=-i\bar{\psi}\gamma_5\delta\alpha(x)\tag{11.243}$$

we have

$$\delta\phi^i=\frac{\partial\phi^i}{\partial\alpha}\delta\alpha(x).\tag{11.244}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

For example,

$$\delta(\bar{\psi}\psi) = -2i\bar{\psi}\gamma_5\psi\delta\alpha(x). \tag{11.245}$$

But changing the variables in the functional integral must not make a difference, it must be invariant. Thus we get

$$0 = \delta\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle^A = \langle -\delta S_E \phi^1(x_1)\cdots\phi^m(x_m)\rangle \\ + \left\langle \sum_r \phi^1(x_1)\cdots\delta\phi^r(x_r)\cdots\phi^m(x_m)\right\rangle. \tag{11.246}$$

Then, since

$$-\delta S_E = \int d^4x\, \bar{\psi} i\gamma_\mu(-i\partial_\mu\delta\alpha)\gamma_5\psi = \int d^4x\,(\partial_\mu\delta\alpha) j_{5\mu} = -\int d^4x\,\delta\alpha\partial_\mu j_{5\mu} \tag{11.247}$$

we get

$$0 = \langle(-\partial_\mu j_{5\mu(x)})\phi^1(x_1)\cdots\phi^m(x_m)\rangle + \sum_r \delta(x-x_r)\langle\phi^1(x_1)\cdots\frac{\partial\delta\phi^r(x_r)}{\partial\alpha}\cdots\phi^m(x_m)\rangle \\ -2M\langle\bar{\psi}(x)\gamma_5\psi(x)\phi^1(x_1)\cdots\phi^m(x_m)\rangle, \tag{11.248}$$

where the last term is there if we add a mass term that breaks the chiral symmetry explicitly and we will write $j_5(x) = \bar{\psi}(x)\gamma_5\psi(x)$. Then using Equation (11.241) and integrating over x, we get

$$2M\left\langle \int d^4x\, j_5(x)\phi^1(x_1)\cdots\phi^m(x_m)\right\rangle^A \\ = \frac{\partial}{\partial\alpha}\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle^A \\ -i\frac{C}{8\pi^2}\int d^4x\,\langle F_{\mu\nu}\tilde{F}_{\mu\nu}\rangle\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle^A \\ = \frac{\partial}{\partial\alpha}\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle^A - 4iC\nu\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle^A. \tag{11.249}$$

But what is the effect of the fermions on the instanton? The instantons must still be solutions of the equations of motion

$$D_\mu F_{\mu\nu} = j_\nu \quad \text{with} \quad D_\nu j_\nu = 0 \quad \text{and} \quad i\gamma_\mu D_\mu\psi = 0. \tag{11.250}$$

These equations have a perfectly good solution, $\psi = 0$ and $D_\mu F_{\mu\nu} = 0$. The latter equation is satisfied by the instantons' and hence the instatons' configuration is unchanged by the fermions. All the previously found formulae must still be valid,

$$E(\theta)/V = -2K\cos\theta\, e^{-S_0} \tag{11.251}$$

and

$$\langle\theta|F\tilde{F}|\theta\rangle = -64\pi^2 iK\, e^{-S_0}\sin\theta. \tag{11.252}$$

The only change that occurs is that the Gaussian integral about the instanton configuration over the gauge fields is appended with a functional integral over the fermion fields (which is in a sense also Gaussian as the fermion fields only enter quadratically) in the presence of the instanton background

$$K \to K\frac{det(i\gamma_\mu(\partial_\mu + A_\mu))}{det(i\gamma_\mu(\partial_\mu))}. \tag{11.253}$$

The consequences of this change are profound. The fermionic determinant in the presence of the instanton vanishes exactly, giving

$$E(\theta)/V = 0 \quad \langle\theta|F\tilde{F}|\theta\rangle = 0. \tag{11.254}$$

This means that all the theta vacua become degenerate in energy, that the $U_A(1)$ symmetry is spontaneously broken. The $U(1)$ problem corresponds to the question, "Why is there no corresponding massless Goldstone boson?" What we will find is that the massless boson never contributes to gauge-invariant matrix elements and therefore is not physically manifested.

Why does the fermion determinant vanish? It is because in the presence of an instanton, there is necessarily a zero energy mode to the Dirac equation. Evidently

$$\int d\theta d\bar{\theta}\, e^{0\times\bar{\theta}\theta} = \int d\theta d\bar{\theta}\, 1 = 0, \tag{11.255}$$

which is quite unlike the bosonic case

$$\int d\varphi\, e^{-0\times\varphi^2} = \int d\varphi\, 1 = \infty(=\frac{1}{0}). \tag{11.256}$$

The zero mode follows from a deep theorem, the Atiyah–Singer index theorem [8]. However, we can quite easily establish the existence of the zero mode directly using the simplest chiral Ward–Takahashi relation. We will work with an $SU(2)$ gauge group with one doublet of fermions for simplicity. The Dirac equation for eigenmode of energy λ_r is

$$i\not{D}\psi_r = \lambda_r\psi_r \tag{11.257}$$

but then

$$i\not{D}\gamma_5\psi_r = -\gamma_5 i\not{D}\psi_r = -\lambda_r\gamma_5\psi_r. \tag{11.258}$$

Thus for each mode ψ_r of energy λ_r there is a matching eigenmode $\gamma_5\psi_r$ of energy $-\lambda_r$. But what happens if $\lambda_r = 0$? Let ψ_r^0 be a zero mode, $i\not{D}\psi_r^0 = 0$, but then obviously $i\not{D}\gamma_5\psi_r^0 = 0$. We can choose the zero mode to be an eigenmode of γ_5: with $\psi_r^{0\pm} = \frac{1\pm\gamma_5}{2}\psi_r^0$ we have $\gamma_5\psi_r^{0\pm} = \pm\psi_r^{0\pm}$ with $i\not{D}\psi_r^{0\pm} = 0$ and $\frac{1\pm\gamma_5}{2}\psi_r^{0\mp} = 0$. The eigenvalue of γ_5 of the zero mode is called its chirality, which we will call χ_r for zero mode ψ_r^0. We do not know if $\psi_r^{0+} = 0$ or perhaps $\psi_r^{0-} = 0$, or possibly neither vanishes (in which case there are two zero modes, of chirality ± 1, respectively); however, both cannot vanish if $\psi_r^{0+} + \psi_r^{0-} = \psi_r^0 \neq 0$.

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

Let n_+ be the number of zero modes with positive chirality and n_- be the number of zero modes with negative chirality. The Atiyah–Singer index theorem states that $n_+ - n_- = \nu$. We can prove this theorem using the chiral Ward–Takahashi identities. Consider the simplest identity, without any fields ϕ^r, for a single doublet of fermions in $SU(2)$, $C = 1/2$, we have,

$$-2i\nu = 2M \int d^4x\, \langle\bar{\psi}\gamma_5\psi\rangle^A = 2M \frac{\int \mathcal{D}(\psi,\bar{\psi})\, e^{-S_E} \int d^4x\, \bar{\psi}\gamma_5\psi}{\int \mathcal{D}(\psi,\bar{\psi})\, e^{-S_E}} \tag{11.259}$$

with $S_E = \int d^4x\, \bar{\psi}\left(i\not{D} - iM\right)\psi$ and the solutions of the Dirac equation are unchanged from the massless case, only the energy eigenvalues are shifted

$$i(\not{D} - M)\psi_r = (\lambda_r - iM)\psi_r. \tag{11.260}$$

Clearly,

$$i(\not{D} - M)\gamma_5\psi_r = (-\lambda_r - iM)\gamma_5\psi_r, \tag{11.261}$$

but the eigenmodes ψ_r and $\gamma_5\psi_r$ must be orthogonal if $\lambda_r \neq 0$ as they are actually eigenmodes of the hermitean operator $i\not{D}$. We observe

$$\int d^4x\bar{\psi}_s\gamma_5\psi_s = 0 \quad \text{if} \quad \lambda_s \neq 0, \tag{11.262}$$

but for the zero modes

$$\int d^4x\bar{\psi}^0_s\gamma_5\psi^0_s = \chi_s, \tag{11.263}$$

thus

$$\int d^4x\, \bar{\psi}\gamma_5\psi = \sum_{s,\lambda_s=0} \chi_s\bar{b}_s a_s. \tag{11.264}$$

The ψ_r are a complete and orthonormal basis of the space of fermion fields, hence we can write

$$\psi = \sum_r a_r\psi_r \quad \bar{\psi} = \sum_r \psi^\dagger_r\bar{b}_r \tag{11.265}$$

with Grassmann coefficients a_r and $\bar{b}_r$. Then the functional integral is given by

$$\begin{aligned} -2i\nu &= 2M \frac{\int \prod_r da_r d\bar{b}_r\, e^{\sum_r(\lambda_r - iM)\bar{b}_r a_r} \int d^4x\bar{\psi}\gamma_5\psi}{\prod_r(\lambda_r - iM)} \\ &= 2M \frac{\int \sum_{s,\lambda_s=0} \prod_{r\neq s}(\lambda_r - iM)\chi_s}{\prod_r(\lambda_r - iM)} \end{aligned} \tag{11.266}$$

as the fermionic integral gives $(\lambda_r - iM)$ for all the non-zero modes but a factor of 1 for the zero mode in the sum $\sum_{s,\lambda_s=0}\chi_s\bar{b}_s a_s$. The infinite product cancels between numerator and denominator for all the non-zero modes, and therefore the chiral Ward identity gives

$$-2i\nu = 2i\sum_s \chi_s = 2i(n_+ - n_-). \tag{11.267}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

Thus $\nu = n_- - n_+$, which cannot be satisfied unless there are at least ν zero modes.

For the case $\nu = 1$ we can easily show that there is one zero mode with negative chirality and no zero modes of positive chirality. We note that the instanton configuration is self dual, $F_{\mu\nu} = \tilde{F}_{\mu\nu}$. We assume that there is a positive chirality zero mode $\not{D}\psi^{0+} = 0$ and $\gamma_5\psi^{0+} = \psi^{0+}$. Then

$$\begin{aligned} 0 = (\not{D})^2\psi^{0+} &= \left(\frac{1}{2}\{\gamma_\mu, \gamma_\nu\} + \frac{1}{2}[\gamma_\mu, \gamma_\nu]\right) D_\mu D_\nu \psi^{0+} \\ &= D_\mu D_\mu \psi^{0+} + \frac{1}{4}[\gamma_\mu, \gamma_\nu] F_{\mu\nu}\psi^{0+} \\ &= D^2\psi^{0+} + \frac{1}{4}[\gamma_\mu, \gamma_\nu] F_{\mu\nu}\psi^{0+}. \end{aligned} \tag{11.268}$$

However,

$$\begin{aligned} F_{\mu\nu}\frac{1}{2}[\gamma_\mu, \gamma_\nu]\psi^{0+} F_{\mu\nu}\frac{1}{2}[\gamma_\mu, \gamma_\nu]\gamma_5\psi^{0+} &= F_{\mu\nu}\left(-\frac{1}{2}\epsilon_{\mu\nu\sigma\tau}\gamma_\sigma\gamma_\tau\right)\psi^{0+} \\ &= -\tilde{F}_{\mu\nu}\gamma_\mu\gamma_\nu\psi^{0+} = -F_{\mu\nu}\frac{1}{2}[\gamma_\mu, \gamma_\nu]\psi^{0+}. \end{aligned} \tag{11.269}$$

Therefore, $F_{\mu\nu}\frac{1}{2}[\gamma_\mu, \gamma_\nu]\psi^{0+} = 0$ and consequently $D^2\psi^{0+} = 0$. Then

$$0 = \int d^4x (\psi^{0+})^\dagger(-D^2\psi^{0+}) = \int d^4x (D_\mu\psi^{0+})^\dagger(D_\mu\psi^{0+}), \tag{11.270}$$

which is positive unless $D_\mu\psi^{0+} = 0$ identically. Then, in the gauge $A_3 = 0$, this requires $\partial_3\psi^{0+} = 0$. However, this is inconsistent for a normalizable wave function except if $\psi^{0+} = 0$. Therefore, in fact no positive chirality zero mode can exist. Of course the analysis fails for a negative chirality solution, we cannot conclude $F_{\mu\nu}\frac{1}{2}[\gamma_\mu, \gamma_\nu]\psi^{0+} = 0$ for a negative chirality zero mode, and there has to be exactly one negative chirality zero mode so that $\nu = n_- - n_+$ is satisfied.

Therefore, the fermionic functional integral simply makes the contribution from all non-zero instanton sectors vanish. Thus the theta vacua are all degenerate in energy, and the chiral symmetry is certainly spontaneously broken.

11.6.5 Why is there no Goldstone Boson?

To see the non-existence of a Goldstone boson we must modify our chiral Ward identities. The following matrix element no longer makes sense in the non-zero instanton sector as the denominator vanishes,

$$\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle^A = \frac{\int \mathcal{D}(\psi,\bar{\psi})e^{-S_E(\psi,\bar{\psi})}\phi^1(x_1)\cdots\phi^m(x_m)}{\int \mathcal{D}(\psi,\bar{\psi})e^{-S_E(\psi,\bar{\psi})}}; \tag{11.271}$$

however, if we consider just the numerator

$$\langle\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle\rangle^A \equiv \int \mathcal{D}(\psi,\bar{\psi})e^{-S_E(\psi,\bar{\psi})}\phi^1(x_1)\cdots\phi^m(x_m) \tag{11.272}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

then formally the symmetry properties are identical, and we find

$$\left(\frac{\partial}{\partial\alpha} - 2i\nu\right)\langle\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle\rangle^A = 0. \tag{11.273}$$

Now the matrix element in a theta vacuum is given by

$$\begin{aligned}
&\langle\theta|\phi^1(x_1)\cdots\phi^m(x_m)|\theta\rangle^A \\
&\quad = \frac{\int\mathcal{D}(A)e^{-S_E(A)}e^{i\nu\theta}\int\mathcal{D}(\psi,\bar{\psi})e^{-S_E(\psi,\bar{\psi})}\phi^1(x_1)\cdots\phi^m(x_m)}{\int\mathcal{D}(A)e^{-S_E(A)}e^{i\nu\theta}\int\mathcal{D}(\psi,\bar{\psi})e^{-S_E(\psi,\bar{\psi})}} \\
&\quad = \frac{\int\mathcal{D}(A)e^{-S_E(A)}e^{i\nu\theta}\langle\langle\phi^1(x_1)\cdots\phi^m(x_m)\rangle\rangle^A}{\int\mathcal{D}(A)e^{-S_E(A)}e^{i\nu\theta}\langle\langle 1\rangle\rangle^A}
\end{aligned} \tag{11.274}$$

and now the denominator does not vanish for $\int\mathcal{D}(\psi,\bar{\psi})e^{-S_E(\psi,\bar{\psi})}1 \neq 0$ for the sector $\nu = 0$. Thus clearly

$$\left(\frac{\partial}{\partial\alpha} - 2\frac{\partial}{\partial\theta}\right)\langle\theta|\phi^1(x_1)\cdots\phi^m(x_m)|\theta\rangle = 0. \tag{11.275}$$

This is quite interesting. It means that the $U_A(1)$ transformation corresponds equivalently to a change in θ, *i.e.* a $U_A(1)$ transformation changes one theta vacuum into another. The chiral symmetry is therefore spontaneously broken, and the degenerate set of vacua are exactly the theta vacua.

In summary, we have first found the degenerate, classical vacua and their quantum counterparts, $|n\rangle$. Then instantons have the effect of breaking the degeneracy obtained by quantum tunnelling between the different $|n\rangle$ vacua, and the new combinations $|\theta\rangle = \sum_n e^{in\theta}|n\rangle$ are the new energy eigenstates with spectrum

$$E(\theta)/V = -2K\cos\theta e^{-S_0}, \tag{11.276}$$

where S_0 is the classical Euclidean action of one instanton. The parameter θ has nothing to do with chiral symmetry; indeed, there are no fermions yet. But once massless fermions are added to the theory, all the effects of the instantons disappear, due to the appearance of a fermionic zero mode. The $|\theta\rangle$ states suddenly become degenerate, and a chiral transformation corresponds exactly to a transformation of θ. The chiral symmetry is spontaneously broken as there exist infinitely many vacua which are transformed into each other by the action of a chiral transformation. There is one possible way that the system could escape these conclusions, if $\partial/\partial\alpha\langle\theta|\phi^1(x_1)\cdots\phi^m(x_m)|\theta\rangle = \partial/\partial\theta\langle\theta|\phi^1(x_1)\cdots\phi^m(x_m)|\theta\rangle = 0$, *i.e.* nothing depends on α or θ. This would mean that chiral symmetry is manifest and not spontaneously broken, and the vacua $|\theta\rangle$ are just copies of a single, unique vacuum state. It is easy to dispose of this possibility. If we calculate

$$\langle\theta|\sigma_\pm|\theta\rangle, \tag{11.277}$$

where $\sigma_\pm = \bar{\psi}\left(\frac{1\pm\gamma_5}{2}\right)\psi$ then

$$\frac{\partial}{\partial\alpha}\sigma_\pm = \pm 2i\sigma_\pm. \tag{11.278}$$

Then if $\langle\theta|\sigma_\pm|\theta\rangle \neq 0$ we have $\frac{\partial}{\partial\alpha}\sigma_\pm \neq 0$, which then requires that the symmetry is spontaneously broken. We will calculate $\langle\theta|\sigma_\pm|\theta\rangle$ next and show that it cannot vanish.

We have already done the Gaussian functional integral about the classical critical point, the instanton solution, up to a final integral over the scale size

$$K = \frac{2}{g^8}\int_o^\infty \frac{d\lambda}{\lambda^5} f(\lambda M), \tag{11.279}$$

where M is a renormalization point scale (represented by a mass), but now we should append this result with a fermionic functional integral. For a fermion in a background of a configuration of n well-separated instantons and anti-instantons, there are n fermionic zero modes to the operator $i\not{D}$. Then the corresponding fermionic functional integral over the corresponding Grassmann coefficients vanishes,

$$\int d\bar{a}_r da_r e^{0\times\bar{a}_r a_r}\phi(x) = 0, \tag{11.280}$$

unless ϕ contains exactly the bilinear $\bar{a}_r a_r$, for each zero mode. This requires $2n$ fermionic fields. We are interested in the bilinear $\sigma_\pm$, which contains two fermionic fields. Hence the fermionic functional integral vanishes in all sectors of the gauge field except for the sector with $n=1$. Indeed, we must have only exactly one instanton or one anti-instanton so that there is exactly one zero mode. We cannot have a configuration of n_+ instantons and n_- anti-instantons with $n_+ - n_- = \pm 1$, since this configuration will have $n = n_+ + n_- > 1$ fermionic zero modes and the fermionic functional integral will vanish.

Then in the sector with just one anti-instanton with $n_- = 1$, $n_+ = 0$ and self dual instanton fields, $F_{\mu\nu} = \tilde{F}_{\mu\nu}$ the fermionic functional integral will have just one term

$$\begin{aligned}\int d\bar{a}_r da_r e^{\sum_{r,\lambda_r\neq 0}\lambda_r\bar{a}_r a_r}(\psi^{0-})^\dagger\left(\frac{1-\gamma_5}{2}\right)\psi^{0-} &= (\psi^{0-})^\dagger\left(\frac{1-\gamma_5}{2}\right)\psi^{0-}\prod_{\lambda_r\neq 0}\lambda_r \\ &= (\psi^{0-})^\dagger\psi^{0-}\left(det' i\not{D}\right)\end{aligned} \tag{11.281}$$

as the zero mode is a chirality -1, $\left(\frac{1-\gamma_5}{2}\right)\psi^{0-} = \psi^{0-}$. The fermionic zero modes satisfy $i\not{D}\psi^{0-}(x) = i\gamma_\mu(\partial_\mu + A_\mu(x))\psi^{0-}(x) = 0$. If we move the position of the anti-instanton, we change $A_\mu(x) \to A_\mu(x+X)$, then evidently $\psi^{0-}(x) \to \psi^{0-}(x+X)$ and $i\gamma_\mu(\partial_\mu + A_\mu(x+X))\psi^{0-}(x+X) = 0$. For the case of an instanton, $\nu = 1$ with $n_+ = 1, n_- = 0$, which can in principle also contribute, we immediately get a vanishing contribution since $(\psi^{0+})^\dagger(x)\left(\frac{1-\gamma_5}{2}\right)\psi^{0+}(x) = 0$ as the zero mode has chirality $+1$ as $\left(\frac{1-\gamma_5}{2}\right)\psi^{0+}(x) = 0$. In the denominator only $\nu = 0$ can contribute, and the Gaussian integral is done around the configuration $A_\mu = 0$.

We perform the fermionic functional integral first as a functional of the gauge fields. In the sector $\nu = 1$ we must integrate over the position of the anti-instanton since nothing depends on the position of the anti-instanton, which gives a factor of TV. For σ_+ only the sector $\nu = -1$ contributes. The action remains the same and $e^{\nu\theta} = e^{\pm\theta}$ for the two sectors $\nu = \pm 1$. Hence finally we get

$$\langle\theta|\sigma_\pm|\theta\rangle = \int_0^\infty \frac{d\lambda}{\lambda^5} e^{\frac{-8\pi^2}{g^2}} e^{\mp i\theta} g^{-8} f(\lambda M) \frac{det'(i\not{D})}{det(i\not{\partial})}. \tag{11.282}$$

Dimensional analysis gives

$$\frac{det'(i\not{D})}{det(i\not{\partial})} = \lambda h(\lambda M) \tag{11.283}$$

for some dimensionless function $h(\lambda M)$, thus

$$\langle\theta|\sigma_\pm|\theta\rangle = \int_0^\infty \frac{d\lambda}{\lambda^4} e^{\frac{-8\pi^2}{g^2}} e^{\mp i\theta} g^{-8} f(\lambda M) h(\lambda M) \neq 0. \tag{11.284}$$

This amplitude also satisfies the chiral Ward identity

$$\frac{\partial}{\partial\alpha}\langle\theta|\sigma_\pm|\theta\rangle = \langle\theta| \pm 2i\sigma_\pm|\theta\rangle = -2\frac{\partial}{\partial\theta} e^{\mp i\theta} \int_0^\infty \frac{d\lambda}{\lambda^4} e^{\frac{-8\pi^2}{g^2}} e^{\mp i\theta} g^{-8} f(\lambda M) h(\lambda M) \tag{11.285}$$

$$\left(\frac{\partial}{\partial\alpha} + 2\frac{\partial}{\partial\theta}\right)\langle\theta|\sigma_\pm|\theta\rangle = 0. \tag{11.286}$$

Thus the symmetry transformation, corresponding to a change in α,

$$\frac{\partial}{\partial\alpha}\langle\theta|\sigma_\pm|\theta\rangle \neq 0 \tag{11.287}$$

and the symmetry is spontaneously broken. But there is no Goldstone boson. Such a boson must be a pseudoscalar and should give a pole at $p^2 = 0$ in any matrix element (now continued back to Minkowski space) such as

$$\begin{aligned}
\langle\theta|\sigma_+(x)\sigma_-(x)|\theta\rangle &= \sum_n \langle\theta|\sigma_+(x)|n\rangle\langle n|\sigma_-(0)|\theta\rangle \\
&= \int \frac{d^3p}{(2\pi)^3}\langle\theta|\sigma_+(x)|GB\,\vec{p}\rangle\langle GB\,\vec{p}|\sigma_-(0)|\theta\rangle\cdots \\
&= \int \frac{d^3p}{(2\pi)^3}\langle\theta|e^{i\hat{P}_\mu\cdot x^\mu}\sigma_+(0)e^{-i\hat{P}_\mu\cdot x^\mu}|GB\,\vec{p}\rangle\langle GB\,\vec{p}|\sigma_-(0)|\theta\rangle\cdots \\
&= \int \frac{d^3p}{(2\pi)^3} e^{-ip_\mu\cdot x^\mu}\langle\theta|\sigma_+(0)|GB\,\vec{p}\rangle\langle GB\,\vec{p}|\sigma_-(0)|\theta\rangle\cdots \\
&= \int \frac{d^4p}{(2\pi)^4} e^{-ip_\mu\cdot x^\mu}\delta(p^2)\langle\theta|\sigma_+(0)|GB\,\vec{p}\rangle\langle GB\,\vec{p}|\sigma_-(0)|\theta\rangle\cdots
\end{aligned} \tag{11.288}$$

https://doi.org/10.1017/9781009291248.012 Published online by Cambridge University Press

where $p_0 = |\vec{p}|$ and $|GB\,\vec{p}\rangle$ is a state with one Goldstone boson of momentum p_μ. Then using (note $p^\mu p_\mu \equiv p^2$)

$$\delta(p^\mu p_\mu) = \frac{1}{\pi}\Im m\left(\frac{1}{p^2+i\epsilon}\right) \tag{11.289}$$

we get

$$\langle\theta|\sigma_+(x)\sigma_-(x)|\theta\rangle = \Im m\int\frac{d^4p}{(2\pi)^4}e^{-ip_\mu\cdot x^\mu}\left(\frac{1}{p^2+i\epsilon}\right)$$
$$\langle\theta|\sigma_+(0)|GB\,\vec{p}\rangle\langle GB\,\vec{p}|\sigma_-(0)|\theta\rangle\cdots \tag{11.290}$$

and the singularity appears because of the existence of a massless particle. However, in the calculation, there is only a contribution from the sector $n_+ + n_- = 2$, which can be from $\nu = 0, \nu = \pm 2$. Because there is one σ_+ operator and one σ_-, the sectors with $\nu = \pm 2$ corresponding to two instantons or to two anti-instantons simply vanish. Only the sector $\nu = 0$ remains. Here there are two possible contributions, one is the normal, perturbative contribution without any instantons. It is straightforward to verify that there is no massless pole in the perturbative calculation. The only non-trivial configurations come in the sector $\nu = 0$ that correspond to a well-separated pair of one instanton and one anti-instanton. This contribution will simply be a constant

$$\begin{aligned}\langle\theta|\sigma_+(x)\sigma_-(0)|\theta\rangle &= \langle\theta|\sigma_+(x)|\theta\rangle\langle\theta|\sigma_-(0)|\theta\rangle\\ &= \left(\int_0^\infty\frac{d\lambda}{\lambda^4}e^{\frac{-8\pi^2}{g^2}}e^{\mp i\theta}g^{-8}f(\lambda M)h(\lambda M)\right)^2\end{aligned} \tag{11.291}$$

and certainly will not contain a massless pole. Indeed, if we check any matrix element of a set of gauge-invariant operators, we will find no massless pole.

However, if we consider a gauge-variant operator, for example

$$G_\mu = 4\epsilon_{\mu\nu\lambda\sigma}Tr\left(A_\nu\partial_\lambda A_\sigma + \frac{2}{3}A_\nu A_\lambda A_\sigma\right) \tag{11.292}$$

then matrix elements with G_μ inserted will contain a massless pole. This is because $\partial_\mu G_\mu = \epsilon_{\mu\nu\lambda\sigma}Tr\left(F_{\mu\nu}F_{\lambda\sigma}\right)$. Hence any matrix element with G_μ inserted must have no pole when contracted with p_μ. This implies that the original gauge-variant matrix element must have a structure of the form $\frac{p_\mu}{p^2+i\epsilon}$, *i.e.* exactly a massless pole. For example, consider

$$\langle\theta|G_\mu(x)\sigma_-(0)|\theta\rangle = \int d^4p e^{ip_\mu x_\mu}p_\mu\Sigma_-(p) \tag{11.293}$$

from Lorentz invariance. Then the divergence

$$\int d^4x\langle\theta|\partial_\mu G_\mu(x)\sigma_-(0)|\theta\rangle = \int d^4x\, d^4p\, e^{ip_\mu x_\mu}ip^2\Sigma_-(p) \tag{11.294}$$

must have a pole at $p^2 = 0$ if it does not vanish. This would then require

$$\Sigma_-(p) = \frac{C}{p^2 - i\epsilon} + \cdots . \tag{11.295}$$

However, $\int d^4x \partial_\mu G_\mu = 32\pi^2\nu$, thus

$$\int d^4x \langle\theta|\partial_\mu G_\mu(x)\sigma_-(0)|\theta\rangle = 32\pi^2\langle\theta|\sigma_-(0)|\theta\rangle \neq 0, \tag{11.296}$$

where the contribution to the matrix element of σ_- is only from the sector with $\nu = 1$. Thus $\Sigma_-(p)$ must have a pole at zero momentum.