5. Chiral Symmetry Breaking

In this section, we discuss the following class of theories: $SU(N_c)$ gauge theory coupled to N_f Dirac fermions, each transforming in the fundamental representation of the the gauge group. A particularly important member of this class is QCD, the theory of the strong nuclear interactions, and we will consider this specific theory in some detail in Section 5.4. Furthermore, throughout this section we will adopt various terminology of QCD. For example, we will refer to the fermions throughout as *quarks*.

It turns out that the most startling physics occurs when we take the fermions to be massless. For this reason, we will start our discussion with this case, and delay consideration of massive fermions to Section 5.2.3. The Lagrangian of the theory is

$$\mathcal{L} = -\frac{1}{2g^2}\text{Tr}\, F_{\mu\nu}F^{\mu\nu} + \sum_{i=1}^{N_f} i\bar{\psi}_i \not{D}\psi_i \tag{5.1}$$

where $\not{D}\psi = \not{\partial}\psi - i\gamma^\mu A_\mu \psi$. Here $i = 1, \ldots, N_f$ labels the species of quark and is sometimes referred to as a *flavour* index. (Note that ψ also carries a colour index that runs from 1 to N_c and is suppressed in the expressions above.)

Much of what we have to say below will follow from the global symmetries of the theory (5.1). Indeed, the theory has a rather large symmetry group which is only manifest when we decompose the fermionic kinetic terms into into left-handed and right-handed parts

$$\sum_{i=1}^{N_f} i\bar{\psi}_i \not{D}\psi_i = \sum_{i=1}^{N_f} i\psi^\dagger_{+i}\bar{\sigma}^\mu \mathcal{D}_\mu \psi_{+i} + i\psi^\dagger_{-i}\sigma^\mu \mathcal{D}_\mu \psi_{-i}$$

Written in this way, we see that the classical Lagrangian has the symmetry

$$G_F = U(N_f)_L \times U(N_f)_R$$

which acts as

$$U(N_f)_L : \psi_{-i} \mapsto L_{ij}\psi_{-j} \quad \text{and} \quad U(N_f)_R : \psi_{+i} \mapsto R_{ij}\psi_{+j} \tag{5.2}$$

where L and R are both $N_f \times N_f$ unitary matrices. As we will see in some detail below, in the quantum theory different parts of this symmetry group suffer different fates.

Perhaps the least interesting is the overall $U(1)_V$, under which both ψ_- and ψ_+ transform in the same way: $\psi_{\pm,i} \rightarrow e^{i\alpha}\psi_{\pm,i}$. This symmetry survives and the associated conserved quantity counts the number of quark particles (minus the number of anti-quarks) of either handedness. In the context of QCD, this is referred to as baryon number.

The other Abelian symmetry is the axial symmetry, $U(1)_A$. Under this, the left-handed and right-handed fermions transform with an opposite phase: $\psi_{\pm,i} \rightarrow e^{\pm i\beta}\psi_{\pm,i}$. We already saw the fate of this symmetry in Section 3.1 where we learned that it suffers an anomaly.

This means that the global symmetry group of the quantum theory is

$$G_F = U(1)_V \times SU(N_f)_L \times SU(N_f)_R \tag{5.3}$$

In this section, our interest lies in what becomes of the two non-Abelian symmetries. These act as (5.2), but where L and R are now each elements of $SU(N_f)$ rather than $U(N_f)$.

5.1 The Quark Condensate

As we've seen in Section 2.4, the dynamics of our theory depends on the values of N_f and N_c. For low enough N_f, we expect that the low-energy physics will be dominated by two logically independent phenomena. We have met the first of these phenomena already: confinement. In this section, we will explore the second of these phenomena: the formation of a *quark condensate*.

The quark condensate – also known as a *chiral condensate* – is a vacuum expectation value of the composite operators $\bar{\psi}_{-i}(x)\psi_{+j}(x)$. (As usual in quantum field theory, one has to regulate coincident operators of this type to remove any UV divergences.) It turns out that the strong coupling dynamics of non-Abelian gauge theories gives rise to an expectation value of the form

$$\langle \bar{\psi}_{-i}\psi_{+j}\rangle = -\sigma\delta_{ij} \tag{5.4}$$

Here σ is a constant which has dimension of $[\text{Mass}]^3$ because a free fermion in $d = 3+1$ has dimension $[\psi] = \frac{3}{2}$. (An aside: in Section 2 we referred to the string tension as σ; it's not the same object that appears here.) The only dimensionful parameter in our theory is the strong coupling scale Λ_{QCD}, so we expect that parameterically $\sigma \sim \Lambda_{QCD}^3$, although they may differ by some order 1 number.

There are a couple of obvious questions that we can ask.

- Why does this condensate form?
- What are the consequences of this condensate?

The first of these questions is, like many things in strongly coupled gauge theories, rather difficult to answer with any level of precision, and a complete understanding is still lacking. In what follows, we will give some heuristic arguments. In contrast, the second question turns out to be surprisingly straightforward to answer, because it is determined entirely by symmetry. We will explore this in Section 5.2.

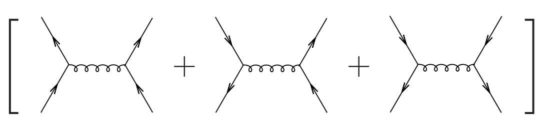
Why Does the Quark Condensate Form?

The existence of a quark condensate (5.4) is telling us that the vacuum of space is populated by quark-anti-quark pairs. This is analogous to what happens in a superconductor, where pairs of electrons condense.

In a superconductor, the instability to formation of an electron condensate is a result of the existence of a Fermi surface, together with a weak attractive force mediated by phonons. In the vacuum of space, however, things are not so easy. The formation of a quark condensate does not occur in weakly coupled theory. Indeed, this follows on dimensional grounds because, as we mentioned above, the only relevant scale in the game is Λ_{QCD}.

To gain some intuition for why a condensate might form, let's look at what happens at weak coupling $g^2 \ll 1$. Here we can work perturbatively and see how the gluons change the quark Hamiltonian. There are two, qualitatively different effects. The first is the kind that we already met in Section 2.5.1; a tree level exchange of gluons gives rise to a force between quarks. This takes the form

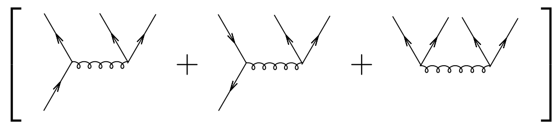
$$\Delta H_1 = g^2$$



As we saw in Section 2.5.1, the upshot of these diagrams is to provide a repulsive force between two quarks in the symmetric channel, and an attractive force in the anti-symmetric channel. Similarly, a quark-anti-quark pair attract when they form a colour singlet and repel when they form a colour adjoint.

The second term is more interesting for us. The relevant diagrams take the form

$$\Delta H_2 = g^2$$



The novelty of these terms is that they they provide matrix elements which mix the empty vacuum with a state containing a quark-anti-quark pair. In doing so, they change the total number of quarks + anti-quarks;

The existence of the quark condensate (5.4) is telling us that, in the strong coupling regime, terms like ΔH_2 dominate. The resulting ground state has an indefinite number of quark-anti-quark pairs. It is perhaps surprising that we can have a vacuum filled with quark-anti-quark pairs while still preserving Lorentz invariance. To do this, the quark pairs must have opposite quantum numbers for both momentum and angular momentum. Furthermore, we expect the condensate to form in the attractive colour singlet channel, rather than the repulsive adjoint.

The handwaving remarks above fall well short of demonstrating the existence of the quark condensate. So how do we know that it actually forms? Historically, it was first realised from experimental considerations since it explains the spectrum of light mesons; we will describe this in some detail in Section 5.4. At the theoretical level, the most compelling argument comes from numerical simulations on the lattice. However, a full analytic calculation of the condensate is not yet possible. (For what it's worth, the situation is somewhat better in certain supersymmetric non-Abelian gauge theories where one has more control over the dynamics and objects like quark condensates can be computed exactly.) Finally, there is a beautiful, but rather indirect, argument which tells us that the condensate (5.4) must form whenever the theory confines. We will give this argument in Section 5.6.

5.1.1 Symmetry Breaking

Although the condensate (5.4) preserves the Lorentz invariance of the vacuum, it does not preserve all the global symmetries of the theory. To see this, we can act with a chiral $SU(N_f)_L \times SU(N_f)_R$ rotation, given by

$$\psi_{-i} \mapsto L_{ij}\psi_{-j} \quad \text{and} \quad \psi_{+i} \mapsto R_{ij}\psi_{+j}$$

The ground state of the our theory is not invariant. Instead, the condensate transforms as

$$\langle \bar{\psi}_{-i}\psi_{+j}\rangle \mapsto \sigma(L^\dagger R)_{ij}$$

This is an example of spontaneous symmetry breaking which, in the present context, is known as *chiral symmetry breaking* (sometimes shortened to χSB). We see that the condensate remains untouched only when $L = R$. This tells us that the symmetry breaking pattern is

$$G_F = U(1)_V \times SU(N_f)_L \times SU(N_f)_R \;\rightarrow\; U(1)_V \times SU(N_f)_V \tag{5.5}$$

where $SU(N_f)_V$ is the diagonal subgroup of $SU(N_f)_L \times SU(N_f)_R$. The purpose of this chapter is to explore the consequences of this symmetry breaking. As we will see, the consequences are astonishingly far-reaching.

Other Symmetry Breaking Patterns

Throughout this chapter, we will only discuss the symmetry breaking pattern (5.5), since this is what is observed in QCD. But before we move on, it's worth briefly mentioning that other gauge theories can exhibit different symmetry breaking patterns.

For example, consider a $SO(N)$ gauge theory coupled to N_f Dirac fermions in the N-dimensional vector representation. In contrast to the $SU(N)$ gauge theory described above, the vector representation of $SO(N)$ is real. This means that we can equivalently describe the system as having $2N_f$ Weyl fermions, each of which transform in the same vector representation. Correspondingly, the global symmetry group of this theory is

$$G_F = SU(2N_f)$$

A chiral condensate of the form (5.4) will spontaneously break

$$G_F = SU(2N_f) \;\rightarrow\; O(2N_f)$$

Symmetry breaking patterns of this type are typical for fermions in real representations of the gauge group.

The other representative symmetry breaking pattern occurs for $Sp(N)$ gauge groups, again coupled to N_f Dirac fermions in the fundamental ($2N$-dimensional) representation. This representation is pseudo-real; if you take the complex conjugate you can turn it back into the original representation through the use of an anti-symmetric invariant tensor J^{ab}. (A familiar example is $SU(2) \equiv Sp(1)$ where you can turn a $\mathbf{2}$ representation into a $\bar{\mathbf{2}}$ representation by multiplying by the ϵ^{ab} invariant tensor.) This meanst that, once again, the global symmetry group is $G_F = SU(2N_f)$. However, this time when the chiral condensate (5.4) forms, it spontaneously breaks

$$G_F = SU(2N_F) \;\rightarrow\; Sp(N_f)$$

Symmetry breaking patterns of this type are typical for fermions in pseudo-real representations.

5.2 The Chiral Lagrangian

The existence of a spontaneously broken symmetry (5.5) immediately implies a whole slew of interesting phenomena. First, the vacuum of our theory is not unique. Instead, there is a manifold of vacua, parameterised by the condensate

$$\langle \bar{\psi}_{-i}\psi_{+j}\rangle = -\sigma\, U_{ij}$$

where $U \in SU(N_f)$. Next, Goldstone's theorem tells us that there are massless particles in the spectrum. These are bound states of the original quarks, but are now best thought of as long-wavelength ripples of the condensate, where it's value now varies in space and time: $U = U(x)$. Note that there are $N_f^2 - 1$ such Goldstone bosons, one for each broken generator in (5.5). We parameterise these excitations by writing

$$U(x) = \exp\left(\frac{2i}{f_\pi}\,\pi(x)\right) \quad \text{with} \quad \pi(x) = \pi^a(x)\,T^a \tag{5.6}$$

Here $\pi(x)$ is valued in the Lie algebra $su(N_f)$. The matrices T^a_{ij} are the generators of the $su(N_f)$ and the component fields $\pi^a(x)$, labelled by $a = 1, \ldots, N_f^2 - 1$ are called *pions*. (As we explain in Section 5.4, these are named after certain mesons in QCD.)

We have also introduced a dimensionful constant f_π in the definition (5.6). For now, this ensures that the pions have canonical dimensions for scalar fields in four dimensions. It is sometimes called the *pion decay constant*, although this name makes very little sense in our current theory because the pions are stable, massless excitations and don't decay. We'll see where the name comes from in Section 5.4.3 when we discuss how these ideas manifest themselves in the Standard Model.

The Low-Energy Effective Action

We would now like to understand the dynamics of the massless Goldstone modes. As we will see, at low-energies, the form of this action is entirely determined by the symmetries of the theory.

To proceed, we want to construct a theory of the Goldstone modes U. We will require that our theory is invariant under the full global chiral symmetry $G_F = U(1)_V \times SU(N_f)_L \times SU(N_f)_R$, under which

$$U(x) \to L^\dagger U(x) R$$

What kind of terms can we add to the action consistent with this symmetry? The obvious term is $\mathrm{tr}\, U^\dagger U$ but because $U \in SU(N_f)$, we have $\mathrm{tr}\, U^\dagger U = N_f$ and so cannot appear in the action. (Here the trace is over the N_f flavour indices.) Happily, this is consistent with the fact that U is a massless Goldstone field and it means that we need to look for terms which depend on the spacetime derivatives, $\partial_\mu U$. There are, of course, many such terms. However, our interest is in the low-energy dynamics which, since we have only massless particles, is the same thing as the long-wavelength physics. This means that the most important terms are those with the fewest derivatives.

The upshot of these arguments is that the low-energy effective Lagrangian can be written as a derivative expansion. The leading term has two derivatives. At first glance, it looks as if there are three different candidates:

$$(\mathrm{tr}\, U^\dagger \partial_\mu U)^2 \quad , \quad \mathrm{tr}\,(\partial^\mu U^\dagger \partial_\mu U) \quad , \quad \mathrm{tr}\,(U^\dagger \partial_\mu U)^2$$

However the first term vanishes because $U^\dagger \partial U$ is an $su(N)$ generator and, hence, traceless. Furthermore, we can use the fact that $U^\dagger \partial U = -(\partial U^\dagger)U$ to write the third term in terms of the second. This means that, at leading order, there is a unique action that describes the dynamics of pions,

$$\mathcal{L}_2 = \frac{f_\pi^2}{4}\, \mathrm{tr}\,(\partial^\mu U^\dagger\, \partial_\mu U) \tag{5.7}$$

This is the *chiral Lagrangian*. Although the Lagrangian is very simple, this is not a free theory because U is valued in $SU(N_f)$. In fact, this is an example of an important class of scalar field theories in which the fields are coordinates on some manifold which, in the present case, is the group manifold $SU(N_f)$. Theories of this type are called *non-linear sigma models* and arise in many different areas of physics.

Historically, the chiral Lagrangian was the first example of a non-linear sigma model, first introduced by Gell-Mann and Lévy in 1960. The origin of the name "sigma-model" is rather strange: the "sigma-particle" is a particular meson in QCD which, it turns out, is the one particle that is *not* captured by the sigma-model! We will explain this a little more in Section 5.4.

For now, the fact that U is valued in $SU(N_f)$ has a rather straightforward consequence: it means that we cannot set $U = 0$. Indeed, our sigma-model describes a degeneracy of ground states, but in each of them $U \neq 0$. This ensures that the chiral Lagrangian spontaneously breaks the $SU(N_f)_L \times SU(N_f)_R$ symmetry, as it must.

5.2.1 Pion Scattering

The beauty of the chiral Lagrangian is that it contains an infinite number of interaction terms, packaged in a simple form by the demands of symmetry. To see these interactions more explictly, we rewrite the chiral Lagrangian in terms of the pion fields defined in (5.6). Keeping only terms quadratic and quartic, the chiral Lagrangian $\mathcal{L}_2$ becomes

$$\mathcal{L}_2 = \text{tr}\,(\partial\pi)^2 - \frac{2}{3f_\pi^2}\text{tr}\,\left(\pi^2(\partial\pi)^2 - (\pi\partial\pi)^2\right) + \ldots \tag{5.8}$$

Note that if we use $\text{tr}\, T^a T^b = \frac{1}{2}\delta^{ab}$ for $su(N_f)$ generators, then the kinetic term has the standard normalisation for each pion field: $\text{tr}\,(\partial\pi)^2 = \frac{1}{2}\,\partial^\mu\pi^a\partial_\mu\pi^a$.

An Example: $N_f = 2$

For concreteness, we work with $N_f = 2$ and take the $su(2)$ generators to be proportional to the Pauli matrices: $T^a = \frac{1}{2}\sigma^a$. The interaction terms then read

$$\mathcal{L}_{\text{int}} = -\frac{1}{6f_\pi^2}\left(\pi^a\pi^a\partial\pi^b\partial\pi^b - \pi^a\partial\pi^a\pi^b\partial\pi^b\right)$$

From this we can read off the tree-level $\pi\pi \to \pi\pi$ scattering amplitude using the techniques that we described in the Quantum Field Theory lectures. We label the two incoming momenta as p_a and p_b and the two outgoing momenta as p_c and p_d. The amplitude is

$$\begin{aligned} i\mathcal{A}^{abcd} = \frac{i}{6f_\pi^2}\Big[&\delta^{ab}\delta^{cd}\Big(4(p_a\cdot p_b + p_c\cdot p_d) + 2(p_a\cdot p_c + p_a\cdot p_d + p_b\cdot p_c + p_b\cdot p_d)\Big) \\ &+ (b\leftrightarrow c) + (b\leftrightarrow d)\Big] \end{aligned}$$

Momentum conservation, $p_a + p_b = p_c + p_d$, ensures that some of these terms cancel. This is perhaps simplest to see using Mandelstam variables which, because all particles are massless, are defined as

$$\begin{aligned} s &= (p_a + p_b)^2 = 2p_a\cdot p_b = 2p_c\cdot p_d \\ t &= (p_a - p_c)^2 = -2p_a\cdot p_c = -2p_b\cdot p_d \\ u &= (p_a - p_d)^2 = -2p_a\cdot p_d = -2p_b\cdot p_c \end{aligned}$$

Using the relation $s + t + u = 0$, the amplitude takes the particularly simple form,

$$i\mathcal{A}^{abcd} = \frac{i}{f_\pi^2}\left[\delta^{ab}\delta^{cd}s + \delta^{ac}\delta^{bd}t + \delta^{ad}\delta^{bc}u\right]$$

Above we have worked at tree level, keeping only the two-derivative terms. We can try to improve our results in two ways: we can include higher derivative terms in the chiral Lagrangian, and we can try to calculate diagrams at one-loop level and higher.

At the next order in the derivative expansion, there are three independent terms. We have $\mathcal{L} = \mathcal{L}_2 + \mathcal{L}_4$ with

$$\mathcal{L}_4 = a_1 \left(\mathrm{tr}\, \partial^\mu U^\dagger \, \partial_\mu U\right)^2 + a_2 \left(\mathrm{tr}\, \partial_\mu U^\dagger \, \partial_\nu U\right) \left(\mathrm{tr}\, \partial^\mu U^\dagger \, \partial^\nu U\right)$$
$$+ a_3 \mathrm{tr} \left(\partial_\mu U^\dagger \, \partial^\mu U \partial_\nu U^\dagger \, \partial^\nu U\right) \tag{5.9}$$

Here a_i are dimensionless coupling constants. These terms will provide corrections to pion-pion scattering that are suppressed at low energy by powers of E/f_π.

Next: loops. The chiral Lagrangian (5.7) is non-renormalisable which means that we need an infinite number of counterterms to regulate divergences. However, this shouldn't be viewed as any kind of obstacle; the theory is designed only to make sense up to a UV cut-off of order f_π. As long as we restrict our attention to low-energies, the theory is fully predictive.

In fact, there is a slightly more interesting story here which I will not describe in detail. If you compute the one-loop correction to pion scattering from $\mathcal{L}_2$, you will find that it scales as $p^4 \log p^2$. The presence of the logarithm means that this term cannot be generated by a tree graph from higher order terms in the chiral Lagrangian and, indeed, at low-energies is enhanced relative to the contributions from $\mathcal{L}_4$.

Furthermore, it turn out that there is a term more important than $\mathcal{L}_4$ that we've missed. This is known as the Wess-Zumino-Witten term. It doesn't contribute to pion scattering, so we can neglect it for the purposes above. However, it plays a key role in the overall structure of the theory. We will discuss this term in detail in Section 5.5.

5.2.2 Currents

We started our discussion with the microscopic non-Abelian gauge theory (5.1) and have ended up, at low-energies, with a very different looking theory (5.7). In general, it is useful to know how operators in the UV get mapped to operators in the IR. There is one class of operators for which this map is particularly straightforward: these are the currents associated to the $SU(N_f)_L \times SU(N_f)_R$ chiral symmetry.

In the microscopic theory, the flavour currents are written most simply in terms of the vector and axial combinations: $J^a_{V\mu} = J^a_{L\mu} + J^a_{R\mu}$ and $J^a_{A\mu} = J^a_{L\mu} - J^a_{R\mu}$, with the familiar expressions

$$J^a_{V\mu} = \bar{\psi}_i T^a_{ij} \gamma_\mu \psi_j \quad \text{and} \quad J^a_{A\mu} = \bar{\psi}_i T^a_{ij} \gamma_\mu \gamma^5 \psi_j \tag{5.10}$$

where T^a_{ij} are $su(N_f)$ generators. What are the analogous expressions in the chiral Lagrangian?

To answer this, let's start with $SU(N_f)_L$. Consider the infinitesimal transformation

$$L = e^{i\alpha^a T^a} \approx 1 + i\alpha^a T^a$$

Under this $U \to L^\dagger U$ so, infinitesimally,

$$\delta_L U = -i\alpha^a T^a U$$

We can now compute the current using the standard trick: elevate $\alpha^a \to \alpha^a(x)$. The Lagrangian is no longer invariant, but now transforms as $\delta\mathcal{L} = \partial_\mu \alpha^a J^a_{L\,\mu}$; the function $J^a_{L\,\mu}$ is the current that we're looking for. Implementing this, we find

$$J^a_{L\,\mu} = \frac{if_\pi^2}{4}\,\mathrm{tr}\Big(U^\dagger T^a \partial_\mu U - (\partial_\mu U^\dagger) T^a U\Big) \tag{5.11}$$

We can also expand this in pion fields (5.6). To leading order we have simply

$$J^a_{L\,\mu} \approx -\frac{f_\pi}{2}\partial_\mu \pi^a$$

Similarly, under $SU(N_f)_R$, we have $\delta U = i\alpha^a U T^a$ and

$$J^a_{R\,\mu} = \frac{if_\pi^2}{4}\Big(- T^a U^\dagger \partial_\mu U + (\partial_\mu U^\dagger) U T^a\Big) \approx +\frac{f_\pi}{2}\partial_\mu \pi^a \tag{5.12}$$

Note that both currents have non-vanishing matrix elements between the vacuum $|0\rangle$ and a one-particle pion state $|\pi^a(p)\rangle$. For example

$$\langle 0| J^a_{L\,\mu}(x) |\pi^b(p)\rangle = -i\frac{f_\pi}{2}\delta^{ab}\, p_\mu e^{-ix\cdot p} \tag{5.13}$$

Historically, the approach to chiral symmetry breaking was known as *current algebra*, and this equation plays a starring role. It is telling us that the chiral $SU(N_f)_L \times SU(N_f)_R$ is spontaneously broken, and acting on the vacuum gives rise to the particles that we call pions.

Although the chiral symmetry is broken, the diagonal combination $SU(N_f)_V$ survives, and

$$\langle 0| J^a_{V\,\mu} |\pi^b\rangle = \langle 0| J^a_{L\,\mu} + J^a_{R\,\mu} |\pi^b\rangle = 0$$

5.2.3 Adding Masses

Our discussion so far has been for massless quarks. We now consider the effect of turning on masses. The Lagrangian is:

$$\mathcal{L} = -\frac{1}{2g^2}\mathrm{Tr}\, F_{\mu\nu}F^{\mu\nu} + \sum_{i=1}^{N_f}\left(i\bar{\psi}_i \not{D}\psi_i - m_i \bar{\psi}_i\psi_i\right)$$

If the masses are large compared to Λ_{QCD}, then the quarks play no role in the low-energy physics. Here we will be interested in the situation where the masses are small, $m_i \ll \Lambda_{QCD}$.

It is a general rule – and a deep fact about quantum field theory – that turning on a mass for fermions always breaks some global symmetry. In the present case, the masses explicitly break the chiral symmetry. If all the masses are equal, then there remains a non-Abelian $U(N_f)_V$ flavour symmetry. In contrast, if all the masses are different, we have only the Cartan subalgebra $U(1)^{N_f}$.

In the previous section, we saw that we can derive powerful statements about the low-energy physics due to the spontaneous breaking of the chiral symmetry. Now this symmetry is explicitly broken by the masses themselves, but all is not lost. For $m_i \ll \Lambda_{QCD}$, we still have an *approximate* chiral symmetry. The quark condensate is still associated to the scale Λ_{QCD}, and the masses give only a small correction. This means that we can still write

$$\langle \bar{\psi}_{-i} \psi_{+j} \rangle \approx -\sigma \, U_{ij}$$

with $U \in SU(N_f)$. We can then incorporate the masses in the chiral Lagrangian by introducing the $N_f \times N_f$ mass matrix,

$$M = \text{diag}(m_1, \ldots, m_{N_f})$$

In the presence of masses, the leading order chiral Lagrangian is

$$\mathcal{L}_2 = \int d^4x \; \frac{f_\pi^2}{4} \, \text{tr} \, (\partial^\mu U^\dagger \, \partial_\mu U) + \frac{\sigma}{2} \text{tr} \, \left(MU + U^\dagger M^\dagger \right)$$

This lifts the vacuum manifold of the theory. It can be thought of as adding a potential to the $SU(N_f)$ vacuum moduli space, resulting in a unique ground state. To see the effect in terms of pion fields, we can again expand $U = e^{2i\pi/f_\pi}$, to find

$$\mathcal{L}_2 = \text{tr} \, (\partial \pi)^2 - \frac{\sigma}{f_\pi^2} \, \text{tr}(M + M^\dagger) \pi^2 + \ldots \tag{5.14}$$

and we see that we get a mass term for the pions as expected.

5.3 Miraculously, Baryons

The purpose of the chiral Lagrangian is to describe the low-energy dynamics of pions. These are the massless Goldstone bosons that arise after spontaneous symmetry breaking which, in terms of the original quarks take the schematic form $\bar{\psi}_i \psi_j$. These particles are all neutral under the $U(1)_V$ vector symmetry.

There are also bound states of quarks which carry quantum numbers under $U(1)_V$. These are the *baryons* that arise by contracting the $a = 1, \ldots, N_c$ colour indices. Schematically these take the form

$$\epsilon_{a_1 \ldots a_{N_c}} \psi_{i_1}^{a_1} \ldots \psi_{i_{N_c}}^{a_{N_c}} \tag{5.15}$$

where we have neglected the spinor indices. The baryons are bosons when N_c is even and fermions when N_c is odd. With our normalisation, they have charge $+N_c$ under the vector symmetry $U(1)_V$. Often one rescales the charges of the quarks to have $U(1)_V$ charge $1/N_c$ so that the baryon has charge $+1$; this re-scaled symmetry is then referred to simply as *baryon number*.

Assuming that our theory confines, the baryons are expected to have mass $\sim \Lambda_{QCD}$. Nonetheless, they are the lightest particles carrying $U(1)_V$ charge and so are stable.

There is no reason to expect that the chiral Lagrangian knows anything about the baryons. Indeed, to construct the chiral Lagrangian we intentionally threw out all but the massless excitations. It is therefore something of a wonderful surprise to learn that the baryons do arise in the chiral Lagrangian: they are solitons.

The Topological Charge

Let's first show that the chiral Lagrangian has a hidden conserved current. Static field configurations in the chiral Lagrangian are described by a map from spatial $\mathbf{R}^3$ to the group manifold $SU(N_f)$. If we insist that the field asymptotes to the same vacuum state so, for example,

$$U(\mathbf{x}) \to 1 \quad \text{as } |\mathbf{x}| \to \infty$$

then we effectively compactify $\mathbf{R}^3$ to $\mathbf{S}^3$. Now static configurations can be thought of as a map

$$U(\mathbf{x}) : \mathbf{S}^3 \mapsto SU(N_f)$$

Such configurations are characterised by their winding

$$\Pi_3(SU(N_f)) = \mathbf{Z}$$

This winding number — which we denote by $B \in \mathbf{Z}$ — is computed by the integral

$$B = \frac{1}{24\pi^2} \int d^3x \; \epsilon_{ijk} \mathrm{tr}\left(U^\dagger (\partial_i U)\, U^\dagger (\partial_j U)\, U^\dagger\, \partial_k U\right) \tag{5.16}$$

In fact, we can go further and write down a local current

$$B^\mu = \frac{1}{24\pi^2}\epsilon^{\mu\nu\rho\sigma}\text{tr}\,(U^\dagger(\partial_\nu U)\,U^\dagger(\partial_\rho U)\,U^\dagger\,\partial_\sigma U)$$

which obeys $\partial_\mu B^\mu = 0$ by virtue of the anti-symmetric tensor. The winding number is then given by $B = \int d^3x\ B^0$.

It is natural to search for an interpretation of this conserved current B^μ, it terms of the microscopic theory. The only candidate is $U(1)_V$, strongly suggesting that we should identify B^μ with the baryon number current and, correspondingly, the solitons with baryons. This appears to be magic. We tried to throw away everything that wasn't massless. But if you treat the pions correctly, the baryons reappear as solitons.

A First Attempt at Solutions

What do these soliton solutions look like? Let's start with the two-derivative chiral Lagrangian. The associated energy functional for static field configurations is

$$E = \frac{f_\pi^2}{4}\int d^3x\ \text{tr}\,\partial_i U^\dagger\cdot\partial_i U$$

where now $i = 1,2,3$ runs over spatial indices only. Solutions to the equations of motion are minima (or, more generally, saddle points) of this energy functional. A simple scaling argument tell us that these don't exist. To see this, consider a putative solution $U_\star(\mathbf{x})$ with energy $E_\star$. Then the new configuration $U_\lambda(\mathbf{x}) = U_\star(\lambda\mathbf{x})$ has energy

$$E_\lambda = \frac{f_\pi^2}{4}\int d^3x\ \text{tr}\,\partial_i U_\star^\dagger(\lambda\mathbf{x})\cdot\partial_i U_\star(\lambda\mathbf{x}) = \frac{1}{\lambda}E_\star$$

We see that we can always lower the energy of any configurations simply by rescaling its size. This simple observation — which goes by the name of *Derrick's theorem* — means that although the chiral Lagrangian has the topology to support solitons, no static solutions exist. The reason for this is that the classical theory is scale invariant so there is nothing to set the size of the soliton. (The only dimensionful quantity, f_π, multiplies the whole action and so doesn't affect the classical equations of motion.)

5.3.1 The Skyrme Model

The situation improves when we include higher derivative terms. These will scale differently with λ, and may result in a minimum of the energy functional.

We saw previously that there are three possible terms with four derivatives (5.9),

$$\mathcal{L}_4 = a_1 \left(\mathrm{tr}\, \partial^\mu U^\dagger\, \partial_\mu U\right)^2 + a_2 \left(\mathrm{tr}\, \partial_\mu U^\dagger\, \partial_\nu U\right) \left(\mathrm{tr}\, \partial^\mu U^\dagger\, \partial^\nu U\right)$$
$$+a_3 \mathrm{tr} \left(\partial_\mu U^\dagger\, \partial^\mu U \partial_\nu U^\dagger\, \partial^\nu U\right)$$

and we expect that the effective action contains all three terms with some choice of coefficients a_1, a_2 and a_3. However, it turns out to be much easier to discuss solitons if we take a particular linear combination of these terms. We take the effective action to be

$$\mathcal{L} = \frac{f_\pi^2}{4}\, \mathrm{tr}\, \left(\partial^\mu U^\dagger \partial_\mu U\right) + \frac{1}{32g^2} \mathrm{tr}\, \left([U^\dagger \partial^\mu U, U^\dagger \partial^\nu U][U^\dagger \partial_\mu U, U^\dagger \partial_\nu U]\right)$$

This is called the *Skyrme model*.

There is no first-principles justification for this particular 4-derivative term although it's worth mentioning that it is the unique term which contains no more than two time derivatives, making it more straightforward to interpret the classical equations of motion. Here g^2 is a dimensionless coupling constant that will ultimately determine the scale of the soliton relative to f_π.

To simplify our notation, we introduce the $su(N_f)_L$ current.

$$L_\mu = U^\dagger \partial_\mu U$$

After massaging the four-derivative terms, you can check that the static energy can be written as

$$E = \frac{f_\pi^2}{4} \int d^3x\ \mathrm{tr}\, \left(L_i L_i^\dagger - \frac{1}{4g^2 f_\pi^2} (\epsilon_{ijk} L_i L_j)(\epsilon_{lmk} L_l^\dagger L_m^\dagger)\right)$$

We now use the Bogomolnyi trick that we already employed in Section 2 for instantons, vortices and monopoles: we write the energy functional as a total square,

$$E = \frac{f_\pi^2}{4} \int d^3x\ \mathrm{tr}\, \left|L_i \mp \frac{1}{2g f_\pi} \epsilon_{ijk} L_j L_k\right|^2 \pm \frac{f_\pi}{4g} \int d^3x\ \epsilon_{ijk} L_i L_j L_k$$

The first term is clearly positive definite. But the second term is something that we've seen before: it is the topological winding (5.16) that we identified with the baryon number B. We learn that the energy is bounded below by the baryon number

$$E \geq \frac{6\pi^2 f_\pi}{g} |B| \tag{5.17}$$

This now looks more promising: the energy of multiple baryons grows at least linearly with B. Soliton configurations with non-trivial winding are called *Skyrmions* and are identified with baryons in the theory.

5.3.2 Skyrmions

Let's see what Skyrmion solutions look like. The usual way to proceed with bounds like (5.17) is to try to saturate them. For $B > 0$, this occurs when the fields obey the first order differential equation

$$L_i = \frac{1}{2gf_\pi}\epsilon_{ijk}L_jL_k \tag{5.18}$$

While this is usually a sensible approach, it turns out that it doesn't help in the present case. One can show that there are no solutions to (5.18). Instead, we must turn to the full, second order, equations of motion and solve

$$\partial_\mu L^\mu = \frac{1}{4f_\pi^2 g^2}\partial_\mu[L_\nu, [L^\mu, L^\nu]] \tag{5.19}$$

We will solve this for the simplest case of

$$N_f = 2$$

Here, the target space = group manifold $SU(2) = \mathbf{S}^3$. For a single Skyrmion, the field $U(\mathbf{x})$ must wrap once around the $\mathbf{S}^3$ target space as we move around the spatial $\mathbf{R}^3$. This is achieved by the so-called hedgehog ansatz,

$$U_{\text{Skyrme}}(\mathbf{x}) = \exp\left(if(r)\boldsymbol{\sigma}\cdot\hat{\mathbf{x}}\right) = \cos f(r) + i\boldsymbol{\sigma}\cdot\hat{\mathbf{x}}\sin f(r) \tag{5.20}$$

This field configuration has winding number $B = 1$ if we pick the function $f(r)$ to have boundary conditions

$$f(r) \to \begin{cases} 0 & \text{at } r = 0 \\ \pi & \text{as } r \to \infty \end{cases}$$

The equation of motion (5.19) then becomes an ordinary differential equation on $f(r)$,

$$(r^2 + 2\sin^2 f)f'' + 2rf' + \sin 2f\ f'^2 - \sin 2f - \sin^2 f \sin 2f = 0$$

which can be solved numerically; it is a monotonically increasing function whose exact form is not needed for our purposes. The energy of this solution turns out to be about 25% higher than the bound (5.17).

Our Skyrme model is built around symmetries. For $N_f = 2$, the symmetry group is $SU(2)_L \times SU(2)_R$, but if we insist (as we did above) that the field tends towards its vacuum value asymptotically, $U(\mathbf{x}) \to 1$, then it leaves us only with the diagonal

$SU(2)_V$ as a global symmetry. Including the group of spatial rotations, we have the symmetry group

$$SU(2)_{\rm rot} \times SU(2)_V \tag{5.21}$$

The single Skyrmion (5.20) is not invariant under either of these $SU(2)$ groups separately. However, it is invariant under the diagonal $SU(2)$ which acts simultaneously as a spatial and flavour rotation.

The subgroup of (5.21) which acts non-trivially on the Skyrmion solution (5.20) can be used to generate new solutions. These are trivially related to the original, and just change its embedding in the target space. Nonetheless, they have important consequences. After quantisation, they endow the Skyrmion with quantum numbers under $SU(2)_V$. For example, one can show that the simplest Skyrmion described above sits in a doublet of $SU(2)_V$. In QCD, viewed as having two light quarks, this is interpreted as the proton and neutron.

The Skyrme model has spawned a mini-industry, and there is much more to say about its quantisation, and its utility in describing both nucleons and higher nuclei. We won't say this here.

There, however, is one important aspect of Skyrmions that we have not yet understood: their quantum statistics. Since the baryon (5.15) contains N_c quarks, we would hope that the Skyrmion is a boson when N_c is even and a fermion when N_c is odd. Yet, so far, the chiral Lagrangian knows nothing about the number of colours N_c. It turns out that we have missed a rather subtle term in the effective action, known as the Wess-Zumino-Witten term. This will be introduced in section 5.5, and in section 5.5.3 we will see that it indeed makes the Skyrmion fermionic or bosonic depending on the number of colours N_c.

5.4 QCD

Until now, we have kept our discussion general. However, there is one example of the class of theories that we have been discussing whose importance dwarfs all others. This is QCD, the theory of the strong nuclear interaction.

QCD is an $SU(3)$ gauge theory coupled to $N_f = 6$ Dirac fermions that we call quarks. However, for many questions concerning the low-energy behaviour of the theory, only two — or sometimes three – of these quarks are important. To see why, we need to look at their masses. (I've included their electromagnetic charge Q for convenience.)

Quark	Charge	Mass (in MeV)
d = down	-1/3	4
u = up	+2/3	2
s = strange	-1/3	95
c = charm	+2/3	1250
b = bottom	-1/3	4200
t = top	+2/3	170,000

Note that the up quark is lighter than the down, an inversion of the hierarchy relative to the other two generations. We can compare these quark masses to the strong coupling scale,

$$\Lambda_{QCD} \approx 300 \text{ MeV}$$

We see that the masses of the two lightest quarks $m_u, m_d \ll \Lambda_{QCD}$ while the strange quark has mass $m_s < \Lambda_{QCD}$, although there is not a large separation of scales. Meanwhile, the other three quarks are clearly substantially heavier than Λ_{QCD} and play no role in the low-energy physics. This means that, for many purposes we can consider QCD to have $N_f = 3$ quarks while, for some purposes, we may want to take $N_f = 2$.

When we take $N_f = 3$, we have several different $SU(3)$ groups floating around. The gauge group is $SU(3)$ and the global symmetry group is $SU(3)_L \times SU(3)_R$, which is spontaneously broken down to $SU(3)_V$ by the chiral condensate. In this section, it is these global symmetries that are of interest.

The global flavour symmetries are not exact because they are broken explicitly by the quark masses. The fact that $m_u \approx m_d$ means that the $SU(2)_V \subset SU(3)_V$ subgroup which rotates only up and down quarks is a rather better symmetry of Nature than the full $SU(3)_V$. This approximate $SU(2)_V$ symmetry was first noticed by Heisenberg in 1932 and is called *isospin*. Heisenberg viewed this symmetry as rotating protons and neutrons, rather than quarks.

Confinement of quarks means that the particles we observe are either mesons (comprising a quark + anti-quark) or baryons (comprising three quarks). These excitations must arrange themselves in representations of the unbroken symmetries of the theory. As we noted, the global symmetries are not exact due to the different quark masses but, as we describe below, are nonetheless visible in the observed spectrum. The fact that mesons and baryons arrange themselves into approximate multiplets of $SU(3)_V$ was first noticed by Gell-Mann, who referred to this classification as the *eightfold way*.

Meson	Quark Content	Mass (in MeV)	Lifetime (in s)
Pion π^+	$u\bar{d}$	140	10^{-8}
Pion π^0	$\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$	135	10^{-16}
Eta η	$\frac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2s\bar{s})$	548	10^{-19}
Eta Prime η'	$\frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s})$	958	10^{-21}
Kaon K^+	$u\bar{s}$	494	10^{-8}
Kaon K^0	$d\bar{s}$	498	$10^{-8} - 10^{-11}$

5.4.1 Mesons

Many hundreds of mesons are observed in Nature[10]. A simple model of a meson views it as a bound state of a quark and an anti-quark, or some linear combination of these states. Each quark is a fermion, so mesons are bosons and, as such, have integer spin. Here we will describe some of the lightest mesons with spin 0 and 1, containing only up, down and strange quarks.

Let's start with the spin 0 mesons. These are all pseudoscalars, with parity -1. A number of these have masses that are lighter or comparable to the proton (which weighs in at 938 MeV). These are shown in the table above.

The $\pm$ and 0 superscripts tell us the electromagnetic charge of the meson. The charged mesons, π^+ and K^+ both have anti-particles, π^- and K^- respectively. The neutral mesons π^0, η and η' are all their own anti-particles; each is described by a real scalar field. Finally, the neutral K^0 is described by a complex scalar field and its anti-particle is denoted $\bar{K}^0$. The list therefore contains, in total, nine different particles + anti-particles.

All mesons are unstable, decaying via the weak force. We will describe this briefly in Section 5.4.3 but, for now, our interest lies in understanding how these mesons arise in the first place. In particular, we would like to understand why this particular pattern of masses emerges.

First, an obvious comment: the masses of the mesons are not equal to the sum of the masses of their constituent quarks! This gets to the heart of what it means to be a strongly coupled quantum field theory. The mesons – and, indeed the baryons – are complicated objects, consisting of a bubbling sea of gluons, quarks and anti-quarks. This is what gives mesons and baryons mass, and also makes these particles hard to

[10]All the properties of all the particles in the universe can be found in the Particle Data Group website http://pdg.lbl.gov/.

understand. Thankfully, for a subset of the mesons, we have the chiral Lagrangian to help us.

Let's see what we would expect based on chiral symmetry. If we consider QCD with just two light quarks – the up and the down – then the spontaneous symmetry breaking of $SU(2)_L \times SU(2)_R$ symmetry should give us three light almost-Goldstone modes. These are the three pions, π^+, π^- and π^0.

The fact that the pions are both bound states of fundamental fermions, and yet can also be viewed as Goldstone bosons, was first suggested by Yoichiro Nambu in the early 1960s. His vision is all the more remarkable given that it came 10 years before the formulation of QCD, and several years before Gell-Mann and Zweig introduced the idea of quarks. Nambu made many further ground-breaking contributions to theoretical physics, including the realisation that quarks carry three colours (not to mention writing down one of the key equations of string theory). He had to wait until 2008 for his Nobel prize.

Suppose now that we consider $N_f = 3$ light quarks. We expect $N_f^2 - 1 = 8$ almost Goldstone-modes. These are usually referred as *pseudo-Goldstone bosons*. And, indeed, there are eight mesons which are substantially lighter than the others: these are the pions, kaons and the η. They sit inside our 3×3 matrix π like this:

$$\pi = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{\pi^0}{\sqrt{2}} + \frac{\eta}{\sqrt{6}} & \pi^+ & K^+ \\ \pi^- & -\frac{\pi^0}{\sqrt{2}} + \frac{\eta}{\sqrt{6}} & K^0 \\ K^- & \bar{K}^0 & -\frac{2\eta}{\sqrt{6}} \end{pmatrix} \tag{5.22}$$

This is not an obvious arrangement. How do we figure out which particles go where? The answer, as with everything in this game, is symmetry. Our theory has a $SU(3)_V$ symmetry, which allows us to assign two Cartan charges $U(1) \times U(1) \subset SU(3)_V$ to each element of the the matrix π. These charges are called "isospin" and "strangeness" and coincide with almost-conserved quantities of the particles that can be determined experimentally.

The eight Goldstone modes that sit in π would be exactly massless if the $SU(3)_L \times SU(3)_R$ were exact. However, chiral symmetry is broken by the quark mass matrix

$$M = \text{diag}(m_u, m_d, m_s)$$

Since we're now dealing with a low-energy effective theory, the masses that appear here should be the renormalised masses, rather than the bare quark masses quoted in the earlier table. Equation (5.14) then gives us the pion masses. Expanding this out, we find

$$\mathcal{L}_{\text{mass}} = \frac{-\sigma}{f_\pi^2}\left[\frac{1}{2}(m_u+m_d)\left((\pi^0)^2+2\pi^+\pi^-\right)+(m_u+m_s)K^-K^+ \right. \tag{5.23}$$
$$\left. +(m_d+m_s)\bar{K}^0K^0+\frac{1}{2}\left(\frac{m_u}{3}+\frac{m_d}{3}+\frac{4m_s}{3}\right)\eta^2+\frac{1}{\sqrt{3}}(m_u-m_d)\pi^0\eta\right]$$

Note that there is mixing between π^0 and η, albeit one that disappears when $m_u = m_d$ so that isospin is restored. There is lots of interesting information in this equation. First note that we cannot directly relate the quark masses to the meson masses; they depend on the unknown ratio σ/f_π^2. Nonetheless, there are a number of simple relations between meson masses, quark masses and the chiral condensate that we can extract. For example, the mass of π^0 is given by

$$m_\pi^2 = \frac{2\sigma}{f_\pi^2}(m_u+m_d)$$

We learn that the square of the pion mass scales linearly with the quark masses. This is known as the *Gell-Mann-Oakes-Renner relation*.

By taking ratios, we can relate meson and quark masses directly. For example, we have

$$\frac{m_{K^+}^2 - m_{K^0}^2}{m_\pi^2} = \frac{m_u - m_d}{m_u + m_d} \tag{5.24}$$

Finally, we can also derive expected relationships between the meson masses. For example, we have $3m_\eta^2 + m_\pi^2 = \frac{2\sigma}{f_\pi^2}\left(2(m_u+m_d)+4m_s\right)$. If we accept that $m_u \approx m_d$, then we get the relation

$$m_K^2 \approx \frac{3}{4}m_\eta^2 + \frac{1}{4}m_\pi^2$$

This is known as the *Gell-Mann-Okubo relation*. Comparing against the experimentally measured masses, we have $\frac{1}{2}\sqrt{3m_\eta^2+m_\pi^2} \approx 480$ MeV, which is not far off the measured value of $m_K \approx 495$ MeV.

The η' Meson

There is one meson listed in the table that is not a Goldstone boson. This is the η' which, despite having similar quark content to the η, has almost twice the mass. Note that, in contrast to the other eight mesons, $\eta' = \frac{1}{\sqrt{3}}(u\bar{u}+d\bar{d}+s\bar{s})$ is a singlet under $SU(3)_V$. This is actually the would-be Goldstone boson associated to the $U(1)_A$ axial symmetry. However, as we have seen, this symmetry suffers from an anomaly, which means that the η' meson is not massless in the chiral limit, and is not particularly light in the real world.

The Mysterious Sigma

There is one light scalar meson listed in the particle data book that I have not yet mentioned. It goes by the catchy name of $f_0(500)$ and has a mass which is listed as somewhere between 400 - 550 MeV. The reason that it's so difficult to pin down is that it decays very quickly – via the strong force rather than weak force – to two pions. Moreover, it has vanishing quantum numbers (angular momentum, parity, isospin and strangeness).

Experimentally, its probably best not to refer to this resonance as a particle at all. However, theoretically it has played a very important role, for this is the "sigma" after which the sigma-model is named. It can be thought of as the excitation that arises from ripples in the value of the quark condensate, $\sigma = \bar{\psi}\psi$, rather than rotations in the quark condensate U.

5.4.2 Baryons

We will briefly describe the baryon spectrum in QCD. In the non-relativistic quark model, with $G = SU(3)$ gauge group, each baryon contains three quarks. As with the mesons, this is a caricature of a baryon which, in reality, is a complicated object that contains many hundreds of gluons, quarks and anti-quarks, but with three more quarks than anti-quarks. This caricature sometimes goes by the name of the non-relativistic quark model.

If we work with $N_f = 3$ species of light quarks, each transforms in the $\mathbf{3}$ of $SU(3)_V$. We have

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{10}$$

A little bit of group theory, combined with the Pauli exclusion principle, shows that those baryons which have spin 1/2 must lie in the $\mathbf{8}$ of $SU(3)_V$. Indeed, there is an octuplet of baryons whose mass differ from each other by about 30%. These are shown in the table on the next page.

Similarly, one can show that baryons with spin 3/2 lie in the $\mathbf{10}$ of $SU(3)_V$. Such a decuplet of baryons also exists: they go by the names Δ (with charges 0, ±1 and 2), $U^\star$ (with charges 0 and ±1), $\Xi^\star$ (with charges -1 and 0) and Ω^- with charge -1.

The fact that the baryons sit nicely into representations of $SU(3)_V$ was first noticed by Gell-Mann who dubbed it the *eightfold way.* At the time the Ω^- baryon — which has quark content sss — had not been discovered. Gell-Mann (and, independently, Ne'eman) used the representation properties to predict the mass, charge and decay products of this particle.

Baryon	Quark Content	Mass (in MeV)	Lifetime (in s)
Proton p	uud	938	stable
Neutron n	udd	940	10^3
Lambda Λ^0	uds	1115	10^{-10}
Sigma Σ^+	uus	1189	10^{-10}
Sigma Σ^0	uds	1193	10^{-19}
Sigma Σ^0	dds	1197	10^{-10}
Xi Ξ^0	uss	1315	10^{-10}
Xi Ξ^-	dss	1321	10^{-10}

For the pions, we showed how the mass splitting can be explained from the chiral Lagrangian. We will not do this for baryons, although with some work one can show that the Skyrmion spectrum indeed gives reasonable agreement.

5.4.3 Electromagnetism, the Weak Force, and Pion Decay

It's not just the quark masses that explicitly break the $SU(3)_V$ flavour symmetry of the Standard Model; the symmetry is also broken by the coupling to the other forces.

At low energies, the relevant force is electromagnetism. The $U(1)_{EM}$ of electromagnetism is a subgroup of $SU(3)_V$, generated by

$$Q = \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & -\frac{1}{3} \end{pmatrix} \tag{5.25}$$

This is enough to tell us how to couple photons to the chiral Lagrangian. We simply need to replace the derivatives in (5.7) with covariant derivatives,

$$S = \int d^4x \, \frac{f_\pi^2}{4} \mathrm{tr}\left(\mathcal{D}^\mu U^\dagger \, \mathcal{D}_\mu U\right) \tag{5.26}$$

where

$$\mathcal{D}_\mu U = \partial_\mu U - ieA_\mu [Q, U]$$

with e the electric charge of an electron.

At the classical level, this coupling preserves a $(U(1) \times SU(2))_L \times (U(1) \times SU(2))_R$ subgroup of the $SU(3)_L \times SU(3)_R$ chiral symmetry. This means that, if all quark masses vanish, the four neutral mesons π^0, η, K^0 and $\bar{K}^0$ would still be Goldstone bosons, and massless even when we include the effects of electromagnetism. In contrast, the charged pions $\pi^\pm$ and $K^\pm$ are massless only at tree level. One-loop effects give a contribution to their mass of the form $\delta m^2_{EM} \sim e^2 \text{tr}(QUQU)$. The charged pion masses in (5.23) then become

$$m^2_{\pi^\pm} = \frac{2\sigma}{f_\pi^2}(m_u + m_d) + \delta m^2_{EM} \quad \text{and} \quad m^2_{K^\pm} = \frac{2\sigma}{f_\pi^2}(m_d + m_s) + \delta m^2_{EM}$$

By taking ratios of these meson masses, we can cancel the factors of σ/f_π^2 and δm^2_{EM} and learn about the quark masses. For example, taking into account electromagnetic corrections, we can generalise (5.24) to

$$\frac{(m^2_{K^\pm} - m^2_{K^0}) - (m^2_{\pi^\pm} - m^2_{\pi^0})}{m^2_{\pi^0}} = \frac{m_u - m_d}{m_u + m_d}$$

From the measured masses of the mesons, we then get that $m_d/m_u \approx 2$.

Charged Pion Decay

Although certain pions are relatively long lived – most notably the $\pi^\pm$ and the kaons – none are absolutely stable. They decay through the weak force. Happily, this too is rather straightforward to calculate using the chiral Lagrangian, because the weak gauge group coincides with $SU(2)_L$ isospin.

For example, the charged pion $\pi^+ = u\bar{d}$ has a lifetime of $\sim 10^{-8}$ seconds, decaying almost always to

$$\pi^+ \to \bar{\mu} + \nu_\mu$$

The decay is mediated by the W-boson. If we integrate out the W-boson, we can equally well describe the decay using Fermi's four-fermion interaction,

$$\mathcal{L}_{Fermi} = \frac{G_F}{\sqrt{2}} \left[\bar{u}\gamma^\mu(1-\gamma^5)d\right]\left[\bar{\mu}\gamma_\mu(1-\gamma^5)\nu_\mu\right]$$

where $G_F \approx 10^{-5}$ GeV$^{-}2$ is the Fermi constant. The computation of the decay rate now factorises into two pieces: the leptonic part $\langle \bar{\mu}\nu_\mu | \bar{\mu}\gamma_\mu(1-\gamma^5)\nu_\mu|0\rangle$ can be computed perturbatively. However, the piece involving the quarks involve strongly interacting physics, $\langle 0|\bar{u}\gamma^\mu(1-\gamma^5)d|\pi^+\rangle$. Thankfully we can compute this using the currents that we introduced in Section 5.2.2. The operator coincides with the $SU(2)_L$ current (5.10),

$$\bar{u}\gamma_\mu(1-\gamma^5)d = 2(J^1_{L\,\mu} + iJ^2_{L\,\mu})$$

We can then use our result (5.13),

$$\langle 0|J^a_{L\,\mu}(x)|\pi^b(p)\rangle = -i\frac{f_\pi}{2}\delta^{ab}\,p_\mu e^{-ix\cdot p}$$

We simply need to remember that $\pi^+ = \frac{1}{\sqrt{2}}(\pi^1 + i\pi^2)$ to find that the matrix element is determined by f_π,

$$\langle 0|\bar{u}\gamma^\mu(1-\gamma^5)d|\pi^+\rangle = -i\sqrt{2}f_\pi p^\mu e^{-ip\cdot x}$$

Recall that when we first introduced f_π in Section 5.2, we mentioned that it is called the *pion decay constant*, even though that name made little sense in the theory we were considering. Now we see why: it is the scale which directly determines the decay width of the pion.

To compute the lifetime of the pion, we must square the matrix element and integrate over the phase space of $\bar{\mu}$ and ν_μ. The end result for the rate of decay is then given by

$$\Gamma(\pi^+ \to \bar{\mu} + \nu_\mu) = \frac{G_F^2 f_\pi^2}{4\pi} m_\pi m_\mu^2 \left(1 - \frac{m_\mu^2}{m_\pi^2}\right)^2$$

Neutral Pion Decay

The neutral pion, $\pi^0 = \frac{1}{\sqrt{2}}(\bar{u}u - \bar{d}d)$ has a substantially shorter lifespan than its charged cousin. It lasts only around $\sim 10^{-16}$ seconds, decaying primarily to

$$\pi^0 \to \gamma\gamma$$

There is an interesting story associated to this. Indeed, it was the effort to understand why this decay occurs at all that first led to the discovery of the anomaly.

The full history is, as with many things in this subject, rather convoluted. The pion decay was first computed in the 1940s, by assuming a coupling to the nucleons $N = (p, n)$ of the form $G_{\pi N}\pi^a \bar{N}\gamma^5\sigma^a N$. This gives a result which is pretty close to the observed value. Unfortunately, this calculation is wrong. As we've seen, the pion is really a Goldstone boson and so has only derivative couplings, at least in the limit $m_\pi \to 0$. Indeed, one can show that in a theory with an unbroken $SU(2)_L \times SU(2)_R$ chiral symmetry, the decay $\pi^0 \to \gamma\gamma$ would be forbidden. What's going on?

The answer is that we've missed something. Gauging a subgroup $U(1)_{EM} \subset SU(2)_V$ introduces an anomaly for the axial currents. We can import our calculation of the chiral anomaly from Section 3.1. For two quarks, up and down, each with $N_c = 3$ colours, we have

$$\partial^\mu J^a_{A\,\mu} = \frac{N_c}{16\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}\,\mathrm{tr}\left(\frac{\sigma^a}{2}Q^2\right)$$

where here $F_{\mu\nu}$ denotes the electromagnetic field strength. In contrast to (5.25), we now take $U(1)_{EM} \subset SU(2)_V$ to be generated by

$$Q = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & -\frac{1}{3} \end{pmatrix}$$

Only the $a = 3$ component of the current is non-vanishing, with

$$\partial^\mu J^3_{A\,\mu} = \frac{N_c}{96\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}$$

But this is precisely the current which, from (5.13), creates the neutral pion π^0, with $\langle 0|J^3_{A\,\mu}|\pi^0\rangle = -if_\pi\, p_\mu e^{-ix\cdot p}$. The anomaly equation then gives an amplitude for $\pi^0 \to \gamma\gamma$. This amplitude is the same as that which would arise from the coupling in the Lagrangian

$$\mathcal{L} = \frac{N_c e^2}{96\pi^2 f_\pi}\pi^0\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \tag{5.27}$$

Note that the decay amplitude is proportional to N_c, the number of colours. Comparing to the experimental data provides a way to determine $N_c = 3$. (Actually, this is a little bit quick because the $U(1)$ charge assignments above are fixed, in part, by anomaly cancellation which, as we saw in Section 3.4.4, changes if we change N_c.) Above we have used just two quarks, $N_f = 2$, but we get the same results using $N_f = 3$ if we correctly identify the current producing π^0 from within the matrix (5.22).

We have argued that the anomaly means there must be an effective coupling of the form (5.27). Yet there's something odd in this, because if we expand out the action (5.26), no such term arises. Indeed, naively this term appears to contradict the ethos of this whole section, because the Goldstone boson π^0 isn't obviously derivatively coupled, which seems very unGoldstonelike. Nonetheless, it would be nice to be able to write down a low-energy effective action that correctly captures the anomaly, rather than adding it in by hand. It turns out that there is a beautiful way to achieve this.

5.5 The Wess-Zumino-Witten Term

We have argued that, at low-energies, the dynamics of the Goldstone modes is captured by the chiral Lagrangian

$$S = \frac{f_\pi^2}{4} \int d^4x \; \mathrm{tr}(\partial_\mu U^\dagger \partial^\mu U) \tag{5.28}$$

We also briefly discussed in Section 5.2.1 the higher order terms that we could add to this action to improve its accuracy as we go to higher energies. It turns out, however, that this misses one very important term, one which, among other things, accounts for the anomaly. This is known as the Wess-Zumino-Witten term.

To motivate the need for an extra term, let's look more closely at the discrete symmetries of the chiral Lagrangian (5.28). They are:

- Charge conjugation, $C : U \mapsto U^*$.
- "Naive parity", $P_0 : \mathbf{x} \to -\mathbf{x}$ with $t \mapsto t$ and $U \mapsto U$.
- An extra symmetry: $U \to U^\dagger$. In terms of the pion fields (5.6)

$$U = \exp\left(\frac{2i}{f_\pi}\,\pi^a T^a\right) = 1 + \frac{2i}{f_\pi}\pi^a T^a + \ldots \tag{5.29}$$

  this symmetry acts as $\pi^a \mapsto -\pi^a$. In other words, it counts pions mod 2. For this reason, we denote the symmetry as $(-1)^{N_\pi}$ where N_π is the number of pions.

However, these are not all symmetries of the underlying QCD-like gauge theory. Indeed, the pions and other Goldstone bosons in QCD are pseudoscalars, meaning that they are odd under parity. The correct parity transformation should be

$$P = P_0(-1)^{N_\pi}$$

It is unusual – although not unheard of – to have a low-energy theory which enjoys more symmetries than its high-energy parent. It might lead us to suspect that we've missed something. Are we really sure that there are no terms that we can add to (5.28) which violate both P_0 and $(-1)^{N_\pi}$, leaving only P as a symmetry?

It is simple to look through the higher derivative terms (5.9) that we met before and convince yourself that they all preserve both P_0 and $(-1)^{N_\pi}$. Indeed, the way to get something that violates P_0 is to use the anti-symmetric tensor $\epsilon^{\mu\nu\rho\sigma}$. But if we try to form a four-derivative term in the action from this, we would have

$$\epsilon^{\mu\nu\rho\sigma}\;\mathrm{tr}\left(U^\dagger\,(\partial_\mu U) U^\dagger\,(\partial_\nu U) U^\dagger\,(\partial_\rho U) U^\dagger\,(\partial_\sigma U)\right) = 0 \tag{5.30}$$

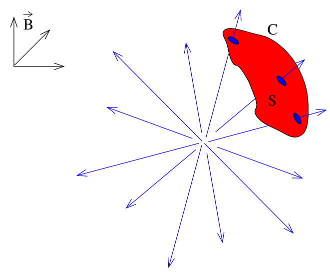


Figure 45: Integrating over S...

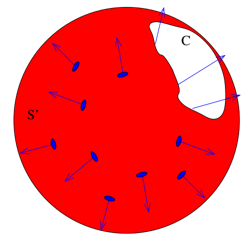


Figure 46: ...or over S'.

and, as shown, this vanishes by anti-symmetry. You can also consider higher derivative terms and see that they too preserve all these discrete symmetries. There's no way to construct terms in the action that violate P_0.

However, the story is rather different if we work with the equation of motion. The equation of motion arising from (5.28) is

$$\frac{1}{2}f_\pi^2\,\partial_\mu(U^\dagger\partial^\mu U) = 0$$

We could add to this the term

$$\frac{1}{2}f_\pi^2\,\partial_\mu(U^\dagger\partial^\mu U) = \frac{k}{48\pi^2}\epsilon^{\mu\nu\rho\sigma}\,U^\dagger\,(\partial_\mu U)U^\dagger\,(\partial_\nu U)U^\dagger\,(\partial_\rho U)U^\dagger\,(\partial_\sigma U) \qquad (5.31)$$

where k is some constant which we will fix shortly and the normalisation of $48\pi^2$ is for later convenience. This is the famous Wess-Zumino-Witten term, first introduced in this context by Witten. Despite our feeble attempts above, it turns out that there is a way to write an action for this term, but not if we restrict ourselves to actions in four-dimensions!

5.5.1 An Analogy: A Magnetic Monopole

A useful analogy can be found in Dirac monopoles. This is a story that we've already met in Section 1.1. Consider a particle of mass m and unit charge moving in $\mathbf{R}^3$ in the background of a Dirac monopole. The equation of motion is

$$m\ddot{x}_i = \lambda\epsilon_{ijk}x_j\dot{x}_k$$

with λ a constant which determines the strength of the monopole. This system shares some similarities with our discussion above. First, the left-hand side is invariant under two discrete symmetries: time reversal $t \mapsto -t$ and parity $x_i \mapsto -x_i$. However, the term on the right-hand side is not separately invariant under both of these, but only if we do both at once. Furthermore, the equation of motion is invariant under $SO(3)$ rotations.

Can we construct an action for this equation of motion? If we try to do so preserving the $SO(3)$ rotational invariance, we run into trouble because the obvious term that we might try to write down to reproduce the right-hand-side is $\epsilon^{ijk}x_i x_j \dot{x}_k = 0$ by anti-symmetry. This, of course, is analogous to (5.30). However, this doesn't mean that no action exists. In fact, there are two possibilities. One is to introduce a gauge potential $A_i(x)$ and write down the action

$$S = \int_C dt\ \frac{1}{2}m\dot{x}_i^2 + \lambda\, A_i(x)\dot{x}^i$$

where C is the worldline of the particle. An example of such a gauge potential was given in (1.5). This approach has two problems: the gauge potential necessarily breaks the $SO(3)$ symmetry, which is no longer manifest in the action; and the gauge potential necessarily suffers from a Dirac string singularity.

We can circumvent both of these problems simply by using Stokes' theorem. Suppose that we take C to be a closed path. We then write

$$\int_C dt\ A_i(x)\dot{x}^i = \int_S dS^{ij}F_{ij}(x) \tag{5.32}$$

where S is a two-dimensional disc, with boundary $\partial S = C$, as shown in the figure. Now things are much nicer. The field strength $F_{ij} = \epsilon_{ijk}x^k/|x|^3$ is both $SO(3)$ invariant and, away from the origin, non-singular. However, the price that we paid is that the action is written in terms of a two-dimensional surface, rather than the one-dimensional worldline.

There is one further problem with the action (5.32) because, as we saw in Section 1.1, there is an ambiguity in the choice of surface S. There is another surface S', with the opposite orientation, that also does the job. For the path integral to be well-defined, we require that these two options give the same answer. We must have

$$\exp\left(i\lambda\int_C dt\ A_i\dot{x}^i\right) = \exp\left(i\lambda\int_S dS^{ij}F_{ij}\right) = \exp\left(-i\lambda\int_{S'} dS^{ij}F_{ij}\right)$$

Stitching together the two discs gives the closed two sphere $\mathbf{S}^2$. The condition can then be written as the requirement

$$\exp\left(i\lambda\int_{S\cup S'} dS^{ij}F_{ij}\right) = \exp\left(i\lambda\int_{\mathbf{S}^2} dS^{ij}F_{ij}\right) = 1 \tag{5.33}$$

However, the magnetic flux through any closed surface is quantised, with the minimum flux given by $\int_{\mathbf{S}^2} dS^{ij}F_{ij} = 4\pi$. We see that the path integral is consistent only if

$$\lambda \in \frac{1}{2}\mathbf{Z}$$

This is simply a restatement of the Dirac quantisation condition that we already met in Section 1.1.

5.5.2 A Five-Dimensional Action

With the discussion of the magnetic monopole fresh in our minds, let's now return to the chiral Lagrangian. We would like to ask if there is some action which respects the $SU(N_f)_L \times SU(N_f)_R$ symmetry of the chiral Lagrangian and reproduces the term on the right-hand-side of (5.31). The answer is yes, but it can only be written by invoking a fifth dimension.

We will work in the Euclidean path integral and the argument is simplest if we take our spacetime to be $\mathbf{S}^4$. We introduce a five-dimensional ball, D, such that $\partial D = \mathbf{S}^4$. We extend the fields $U(x)$ over $\mathbf{S}^4$ to $U(y)$, where y are coordinates on the ball D. We can then reproduce the equation of motion (5.31) from the action

$$S = \frac{f_\pi^2}{4}\int d^4x \ \mathrm{tr}(\partial_\mu U^\dagger \partial^\mu U) + k\int_D d^5y \ \omega \tag{5.34}$$

where

$$\omega = -\frac{i}{240\pi^2}\,\epsilon^{\mu\nu\rho\sigma\tau}\mathrm{tr}\left(U^\dagger\frac{\partial U}{\partial y^\mu}\,U^\dagger\frac{\partial U}{\partial y^\nu}\,U^\dagger\frac{\partial U}{\partial y^\rho}\,U^\dagger\frac{\partial U}{\partial y^\sigma}\,U^\dagger\frac{\partial U}{\partial y^\tau}\right) \tag{5.35}$$

This is the Wess-Zumino-Witten (WZW) term. There are a few things to say about this. First, it is manifestly invariant under the $SU(N_f)_L \times SU(N_f)_R$ chiral symmetry. Second, it naively appears to depend on the choice of extension of $U(x)$ to the five-dimensional space $U(y)$, but this is an illusion. The equations of motion computed from the action depend only on $U(x)$ restricted to the boundary $\mathbf{S}^4$. There are a couple of ways to see this. A somewhat involved calculation shows that the variation of ω is indeed a boundary term. Alternatively, we can expand U in the pion fields as in (5.29),

$$U^\dagger\partial_\mu U = \frac{2i}{f_\pi}\partial_\mu\pi + \mathcal{O}(\pi^2)$$

Then

$$\begin{aligned}\int_D d^5y\ \omega &= \frac{2}{15\pi^2 f_\pi^5}\int_D d^5y\ \epsilon^{\mu\nu\rho\sigma\tau}\partial_\mu\,\mathrm{tr}\left(\pi\partial_\nu\pi\partial_\rho\pi\partial_\sigma\pi\partial_\tau\pi\right) + \mathcal{O}(\pi^6)\\ &= \frac{2}{15\pi^2 f_\pi^5}\int_{\mathbf{S}^4} d^4x\ \epsilon^{\nu\rho\sigma\tau}\,\mathrm{tr}\left(\pi\partial_\nu\pi\partial_\rho\pi\partial_\sigma\pi\partial_\tau\pi\right) + \mathcal{O}(\pi^6)\end{aligned}$$

Written in this form, the $SU(N_f)_L \times SU(N_f)_R$ symmetry is no longer manifest. This is entirely analogous to the lack of manifest rotation symmetry in the Dirac monopole connection. Nonetheless, since it came from the term (5.34), the symmetry must be there, albeit hidden.

We see that the new term gives a five-point interaction between Goldstone modes. In the context of QCD, this mediates the decay $K^+ + K^- \to \pi^+ + \pi^- + \pi^0$, which explicitly breaks the $(-1)^{N_\pi}$ symmetry of the original chiral Lagrangian.

Quantisation of the Coefficient

Just as for the Dirac monopole, there is an ambiguity in our choice of five-dimensional ball D with $\partial D = \mathbf{S}^4$. We could just as well take a ball D', also with $\partial D' = \mathbf{S}^4$ but with the opposite orientation. We can now make the same kind of arguments that, in (5.33), gave us Dirac quantisation. We have

$$\exp\left(ik\int_D d^5y\ \omega\right) = \exp\left(-ik\int_{D'} d^5y\ \omega\right)$$

Stitching together the two five-dimensional balls now makes a five-sphere: $D\cup D' = \mathbf{S}^5$. For our path integral to make sense, we must have

$$\exp\left(ik\int_{\mathbf{S}^5} d^5y\ \omega\right) = 1 \tag{5.36}$$

By now it's probably no surprise to learn that there's some pretty topology that underlies this formula! The integrand provides a map from $\mathbf{S}^5$ to the group manifold $SU(N_f)$, parameterised by $U(y)$. Such maps are characterised by the fifth homotopy group,

$$\Pi_5(SU(N)) = \mathbf{Z} \quad \text{for } N \geq 3$$

This means that as long as we have $N_f \geq 3$ flavours, each map can be assigned a winding $n \in \mathbf{Z}$. It turns out that this winding is computed by

$$\int_{\mathbf{S}^5} d^5y\ \omega = 2\pi n$$

The quantisation condition (5.36) is then satisfied providing

$$k \in \mathbf{Z}$$

This leads us to our next question. What is k?

Rediscovering the Anomaly

The Wess-Zumino-Witten term is closely related to the chiral anomaly. This, it turns out, will give us a strategy to determine the integer k.

Here is the plan. We will gauge a $U(1)$ subgroup of $SU(N_f)_{\rm diag} \subset SU(N_f)_L \times SU(N_f)_R$. To do this, we introduce a charge matrix Q, as in (5.25), and promote the derivatives in the chiral Lagrangian to covariant derivatives

$$S = \int d^4x \, \frac{f_\pi^2}{4} {\rm tr}\,(\mathcal{D}^\mu U^\dagger \, \mathcal{D}_\mu U) + S_{WZW}$$

with $\mathcal{D}_\mu U = \partial_\mu U - ieA_\mu[Q, U]$. However, we also need to find a way to make the Wess-Zumino-Witten term gauge invariant. It's tempting to just do the same trick, and promote $\partial_\mu U$ to $\mathcal{D}_\mu U$ in (5.35). But this isn't allowed because the resulting action now depends on what's going on in five dimensions. Any gauging must take place only in four dimensions.

To proceed, we first look at how the WZW term changes under an infinitesimal transformation $\delta U = i\alpha(x)[Q, U]$, where $\alpha(x)$ depends only on the four-dimensional coordinates. We have, schematically,

$$\delta(U^\dagger \partial U) = i\alpha[Q, U^\dagger \partial U] + i\partial\alpha \, U^\dagger[Q, U]$$

The variation of the 5-form ω defined in (5.35) has terms of order $\partial\alpha^n$, with $n = 0, 1, \ldots, 5$. Of these the $n = 0$ term vanishes by cyclicity of the trace, while the $n = 2, 3, 4, 5$ terms vanish by the anti-symmetry of the $\epsilon^{\mu\nu\rho\sigma\lambda}$ symbol. After judicious use of the identity $U^\dagger \partial U = -(\partial U^\dagger)U$, we find

$$\delta\omega = (\partial_\mu \alpha)\hat{J}^\mu$$

where the current $\hat{J}^\mu$ is given by

$$\begin{aligned} \hat{J}^\mu &= \frac{1}{48\pi^2} \, \epsilon^{\mu\nu\rho\sigma\lambda} {\rm tr}\left(\{Q, \partial_\nu U^\dagger\} \, \partial_\rho U \, U^\dagger \partial_\sigma U \, U^\dagger \partial_\lambda U\right) \\ &= \frac{1}{48\pi^2} \, \epsilon^{\mu\nu\rho\sigma\lambda} \partial_\nu {\rm tr}\left(\{Q, U^\dagger\} \, \partial_\rho U \, U^\dagger \partial_\sigma U \, U^\dagger \partial_\lambda U\right) \end{aligned}$$

where you need to work a little bit to check that the extra terms that you get from acting with ∂_ν vanish by anti-symmetry. Because the current is a total derivative (and because $\partial\alpha$ depends only on the four-dimensional coordinates), the variation of $\int_D d^5x \; \omega$ reduces to a boundary term and, at leading order, can be cancelled by the variation of the four-dimensional gauge field $\delta A_\mu = \partial_\mu \alpha / e$. This means that we can introduce the gauged WZW term

$$S_{WZW} = k\left[\int_D d^5x \; \omega - e \int d^4x \; A_\mu(x) J^\mu\right]$$

with the four-dimensional current given by

$$J^\mu = \frac{1}{48\pi^2}\,\epsilon^{\mu\rho\sigma\tau}\text{tr}\left(\{Q,U^\dagger\}\,\partial_\rho U\,U^\dagger\partial_\sigma U\,U^\dagger\partial_\lambda U\right)$$

However, it turns out that we're still not done. To get a fully gauge-invariant action, we need to work to one higher order in the gauge coupling e. Here we simply quote the result: the fully gauge invariant WZW term is given by

$$S_{WZW} = k\left[\int_D d^5x\;\omega - e\int d^4x\;A_\mu(x)J^\mu \right.$$
$$\left. + \frac{ie^2}{24\pi^2}\int d^4x\;\epsilon^{\mu\nu\rho\sigma}(\partial_\mu A_\nu)A_\rho\text{tr}\left(\{Q^2,U^\dagger\}\,\partial_\sigma U + U^\dagger QUQU^\dagger\partial_\sigma U\right)\right]$$

How does this help us determine k? To see this, we need to expand out this action in terms of pion fields. For simplicity, let's do this for $N_f = 3$ quarks, with the charge matrix (5.25) appropriate for QCD. Among the order e^2 terms from above, there sits

$$\mathcal{L} = \frac{ke^2}{96\pi^2 f_\pi}\pi^0\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}$$

But we've seen this before: this is the term which captures the anomaly (5.27). To agree with the anomaly, the integer k must be equal to the number of colours

$$k = N_c$$

This is a beautiful result. Until now the chiral Lagrangian has appeared to be independent of the gauge group $SU(N_c)$; all that was needed was for the gauge dynamics to initiate chiral symmetry breaking and then it seemed that it could be forgotten. We see that this isn't quite true: a memory of the underlying gauge group survives as the coefficient of the WZW term.

5.5.3 Baryons as Bosons or Fermions

We saw in section 5.3 that the chiral Lagrangian provides a lovely and surprising new perspective on baryons: they are solitons, constructed from topologically twisted pion fields. The conserved baryon current is identified with the topological current

$$B^\mu = \frac{1}{24\pi^2}\epsilon^{\mu\nu\rho\sigma}\text{tr}\left(U^\dagger(\partial_\nu U)\,U^\dagger(\partial_\rho U)\,U^\dagger\,\partial_\sigma U\right)$$

This winding number — which we denote by $B \in \mathbf{Z}$ — is computed by the integral

$$B = \frac{1}{24\pi^2}\int d^3x\;\epsilon_{ijk}\text{tr}\left(U^\dagger(\partial_i U)\,U^\dagger(\partial_j U)\,U^\dagger\,\partial_k U\right)$$

However, there was something lacking in our previous discussion. From the underlying quarks, we know that baryons should be bosons when N_c is even and fermions when N_c is odd. How is this basic fact reproduced in the chiral Lagrangian? Here we show that, for $N_f \geq 3$, the Wess-Zumino-Witten term is exactly what we need.

We focus on $N_f = 3$. (The story is basically unchanged for higher N_f.) Consider a static Skyrmion of the form (5.20) embedded in the $SU(3)$ matrix U as

$$U_0(\mathbf{x}) = \begin{pmatrix} U_{\rm Skyrme}(\mathbf{x}) & 0 \\ 0 & 1 \end{pmatrix}$$

We wish to compare the amplitude for two different processes to occur over some long time T. In the first process, the soliton simply sits stationary in space. In the second process, we rotate the soliton by 2π slowly about its origin. The first process has amplitude e^{iET}, where E is the energy of the soliton. We have to work a little harder to compute the amplitude for the second process. There are two contributions from the two different terms in the chiral Lagrangian (5.34). The first of these comes from the usual kinetic term. Since this involves two time derivatives, it will contribute a piece of order $\sim 1/T$ which can be ignored in the $T \to \infty$ limit. In contrast, the WZW term is linear in time derivatives and will contribute a constant piece. This is what we want.

Here we sketch the calculation. We saw in section 5.3 that the Skyrmion is invariant under a simultaneous spatial and isospin rotation. This means that we can swap our rotation in space for a flavour rotation. A suitable configuration is given by

$$U(\mathbf{x},t) = \begin{pmatrix} e^{i\pi t/T} & & \\ & e^{-i\pi t/T} & \\ & & 1 \end{pmatrix} U_0(\mathbf{x}) \begin{pmatrix} e^{-i\pi t/T} & & \\ & e^{i\pi t/T} & \\ & & 1 \end{pmatrix}$$

We must then extend this configuration over the 5-dimensional ball D and compute the integral

$$\Gamma = -\frac{i}{240\pi^2}\int_D d^5y\ \epsilon^{\mu\nu\rho\sigma\tau}{\rm tr}\left(U^\dagger\frac{\partial U}{\partial y^\mu}\,U^\dagger\frac{\partial U}{\partial y^\nu}\,U^\dagger\frac{\partial U}{\partial y^\rho}\,U^\dagger\frac{\partial U}{\partial y^\sigma}\,U^\dagger\frac{\partial U}{\partial y^\tau}\right)$$

One finds

$$\Gamma = \pi$$

This is what we needed. It means that the amplitude for a soliton which rotates by 2π is not e^{iET} but is instead

$$e^{iET}e^{iN_c\pi} = (-1)^{N_c}e^{iET}$$

The factor of $(-1)^{N_c}$ is telling us that these solitons are bosons when N_c is even and fermions when N_c is odd.

Baryons when $N_f = 2$

When $N_f = 2$ there is no WZW term. This means that the chiral Lagrangian does not know about the underlying number of colours N_c. Nonetheless, there is a new ingredient. This arises because

$$\Pi_4(SU(2)) = \mathbf{Z}_2 \tag{5.37}$$

while $\Pi_4(SU(N)) = 0$ for $N \geq 3$. Note that this is the same homotopy group that arose in the non-perturbative anomaly described in section 3.4.3.

If we work in compactified Euclidean spacetime, then any field configuration in the chiral Lagrangian is a map from $\mathbf{S}^4$ to $SU(2)$ and so is labelled by $\nu = \pm 1$. This gives us different options for the path integral. We could either weight all configurations equally, or weight them with a factor of $(-1)^\nu$. These should be thought of as two different theories which, in analogy with section 2.2, could be said to be distinguished by a "discrete theta parameter" $\theta = 0$ or π.

Here is an example of a field configuration with $\nu = -1$: create a soliton-anti-soliton pair from the vacuum, rotate one around the other, and then annihilate them again. In the theory with $\theta = 0$ this configuration is not weighted any differently and the solitons are bosons. In the theory with $\theta = \pi$, this configuration is weighted with an extra factor of -1. Here the solitons are fermions.

We learn that in the theory with $N_f = 2$, we have a choice: we can either quantise the solitons as a boson or as a fermion. This choice arises as an extra discrete parameter which we must stipulate to fully define the path integral.

5.6 't Hooft Anomaly Matching

Until now, our strategy has been to assume that the quark condensate (5.4) forms and then explore the consequences. Our justification for the condensate itself was rather flimsy. In this section we will improve slightly on this state of affairs. While we will not give a proof that the condensate forms, we will show that it is implied by another, well-known effect of strongly coupled gauge theories: confinement. To show this, we will use the 't Hooft anomaly matching arguments of Section 3.5.

5.6.1 Confinement Implies Chiral Symmetry Breaking

By now the global symmetry group of $G = SU(N_c)$ gauge theory with N_f quarks should be very familiar: it is

$$G_F = U(1)_V \times SU(N_f)_L \times SU(N_f)_R$$

This group has a 't Hooft anomaly which, at high energies, arises from the quarks. If the theory confines, this anomaly must be reproduced by massless bound state fermions in the infra-red. The essence of the argument is that no such bound states can exist.

Let's first compute the 't Hooft anomalies in the ultra-violet, where the quarks contribute. There is no anomaly for $[U(1)_V]^3$, but there are anomalies for both $[SU(N_f)]^3_L$ and $[SU(N_f)_L]^2 \times U(1)_V$, together with the corresponding anomalies for $SU(N_f)_R$. We have

$$\begin{aligned} [SU(N_f)_L]^3 : \quad & A_1 = N_c \\ [SU(N_f)_L]^2 \times U(1)_V : \quad & A_2 = N_c \end{aligned} \tag{5.38}$$

where, in both cases, $A = N_c$ is counting the number of colours of the quarks.

What about in the infra-red? Confinement means that the quarks bind to form colour singlets. Our task is to figure out how the resulting states transform under the flavour symmetry G_F. Here the details depend on the choice of gauge group. When N_c is even, both mesons and baryons are bosons so there are no solutions to the 't Hooft anomaly conditions. This is a striking result. It tells us that there is no way to form massless bound states which match the anomaly. For the theory to be consistent, it must be that G_F is spontaneously broken in the infra-red. The simplest possibility is that the symmetry is broken down to its vector-like subgroup which is free from anomalies. This, of course, is the pattern of chiral symmetry breaking (5.5) that arises from the quark condensate.

Fermionic Baryons

When the number of colours N_c is odd the baryons are fermions. Now we have to work a little harder. Is it possible that these baryons are massless and match the anomaly? To proceed, we will restrict attention to the simplest case of

$$N_c = 3$$

The arguments that follow can be generalised to arbitrary $SU(N_c)$ gauge group.

If the gauge group confines, then any massless fermion must be a colour singlet. The only possibility is baryons, comprised of three quarks. Each constituent quark can be either left-handed or right-handed. Under $SU(N_f)_L \times SU(N_f)_R \subset G_F$, the left-handed fermions transform as $(\mathbf{N}_f, \mathbf{1})$, while the right-handed fermions transform as $(\mathbf{1}, \mathbf{N}_f)$. Both of these Weyl fermions have charge $+1$ under $U(1)_V$. The putative massless baryons therefore transform under the G_F flavour symmetry in representations given by the Young diagrams,

$$\begin{array}{|c|c|c|}\hline l & l & l \\ \hline \end{array} \quad , \quad \begin{array}{|c|}\hline l \\ \hline l \\ \hline l \\ \hline \end{array} \quad , \quad \begin{array}{|c|}\hline l \\ \hline \end{array} \otimes \begin{array}{|c|c|}\hline r & r \\ \hline \end{array} \quad , \quad \begin{array}{|c|}\hline l \\ \hline \end{array} \otimes \begin{array}{|c|}\hline r \\ \hline r \\ \hline \end{array} \quad , \quad \begin{array}{|c|c|}\hline l & l \\ \hline l & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \tag{5.39}$$

What are the helicities of these baryons? We can take a pair of left- or right-handed fermions and form a Lorentz scalar $\epsilon^{\alpha\beta}\psi_{+\alpha}\psi_{+\beta}$ where, for once, we've explicitly written the α, β spinor indices. This means that it's possible to contract the spinor indices such that each baryon above is left-handed. Similarly, if we replace $\begin{array}{|c|}\hline l \\ \hline \end{array}$ with $\begin{array}{|c|}\hline r \\ \hline \end{array}$ then we have the possible set of right-handed baryons

$$\begin{array}{|c|c|c|}\hline r & r & r \\ \hline \end{array} \quad , \quad \begin{array}{|c|}\hline r \\ \hline r \\ \hline r \\ \hline \end{array} \quad , \quad \begin{array}{|c|}\hline r \\ \hline \end{array} \otimes \begin{array}{|c|c|}\hline l & l \\ \hline \end{array} \quad , \quad \begin{array}{|c|}\hline r \\ \hline \end{array} \otimes \begin{array}{|c|}\hline l \\ \hline l \\ \hline \end{array} \quad , \quad \begin{array}{|c|c|}\hline r & r \\ \hline r & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \tag{5.40}$$

These have opposite helicity of the representations in (5.39). The $[U(1)_V]^3$ anomaly remains trivially satisfied if the spectrum of massless baryons is vector-like so we will assume that if a massless baryon of the type (5.39) arises, then its counterpart in (5.40) also arises.

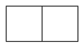
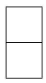
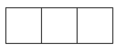
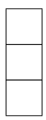
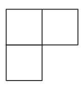
Since we're dealing with a strongly coupled theory, how can we be sure that the indices are contracted so that (5.39) are left-handed and (5.40) are right-handed? First, there is a theorem by Weinberg and Witten which says that one cannot form massless bound states with helicity $\lambda \geq 1$. So if the massless baryons above do indeed form then they must have helicity $\pm\frac{1}{2}$. But is it possible to dress these baryons with gluons which shift their helicity by ± 1?

To be on the safe side, we associate an *index*, $p_\alpha \in \mathbf{Z}$, with $\alpha = 1, \ldots, 5$ to each of the five baryons in (5.39). The magnitude $|p_\alpha|$ denotes the number of species of baryon that arise in the massless spectrum. If these baryons are left-handed then we take $p_\alpha > 0$; if they are right-handed then we take $p_\alpha < 0$. Our task is to find which values of p_α will satisfy anomaly matching and reproduce (5.38).

Next, we need a little group theory. For a representation $\mathbf{R}$ of $SU(N_f)$, we will need to know the dimension $\dim(\mathbf{R})$, the anomaly coefficient $A(\mathbf{R})$, as well as the Dynkin index $\mu(\mathbf{R})$,

$$\operatorname{tr} T^a T^b = \frac{1}{2}\mu(\mathbf{R})\delta^{ab}$$

For each of the representations of interest, we have

$\mathbf{R}$	$\dim(\mathbf{R})$	$\mu(\mathbf{R})$	$A(\mathbf{R})$
	N_f	1	1
	$\frac{1}{2}N_f(N_f+1)$	N_f+2	N_f+4
	$\frac{1}{2}N_f(N_f-1)$	N_f-2	N_f-4
	$\frac{1}{6}N_f(N_f+1)(N_f+2)$	$\frac{1}{2}(N_f+2)(N_f+3)$	$\frac{1}{2}(N_f+3)(N_f+6)$
	$\frac{1}{6}N_f(N_f-1)(N_f-2)$	$\frac{1}{2}(N_f-2)(N_f-3)$	$\frac{1}{2}(N_f-3)(N_f-6)$
	$\frac{1}{3}N_f(N_f^2-1)$	N_f^2-3	N_f^2-9

We can now compute the infra-red anomalies, assuming that we have p_α massless baryons of each type. For $[SU(N_f)_L]^3$ with $N_f \geq 3$, the anomaly is

$$A_1 = \frac{1}{2}(N_f+3)(N_f+6)p_1 + \frac{1}{2}(N_f-3)(N_f-6)p_2 + \left(\frac{1}{2}N_f(N_f+1) - N_f(N_f+4)\right)p_3$$
$$+ \left(\frac{1}{2}N_f(N_f-1) - N_f(N_f-4)\right)p_4 + (N_f^2-9)p_5$$

Note that the baryons with numbers p_3 and p_4 arise from tensor products and have two terms. For example, for p_3 the first term comes from the left-handed baryon $\boxed{l}\otimes\boxed{r\,|\,r}$, and the second — with the minus sign — from the right-handed baryon $\boxed{r}\otimes\boxed{l\,|\,l}$.

Meanwhile, for the $[SU(N_f)^2 \times U(1)_V]$ anomaly, each baryon has charge 3 under the $U(1)_V$. Dividing through by this, we get a contribution proportional to the Dynkin index $\mu(\mathbf{R})$,

$$\frac{A_2}{3} = \frac{1}{2}(N_f+2)(N_f+3)p_1 + \frac{1}{2}(N_f-2)(N_f-3)p_2 + \left(\frac{1}{2}N_f(N_f+1) - N_f(N_f+2)\right)p_3$$
$$+ \left(\frac{1}{2}N_f(N_f-1) - N_f(N_f-2)\right)p_4 + (N_f^2-3)p_5$$

To match the anomalies, we need to find p_α such that $A_1 = A_2 = 3$.

To start, let's look at $N_f = 3$. Anomaly matching gives

$$A_1 = 27p_1 - 15p_3 = 3 \quad \text{and} \quad \frac{A_2}{3} = 15p_1 - 9p_3 + 6p_5 = 1$$

We can immediately see that there can be no solutions to the second of these equations since $A_2/3$ in the infra-red theory is necessarily a multiple of 3 and cannot reproduce the ultra-violet anomaly $A_2/3 = 1$. We learn that $G = SU(3)$ gauge theory with $N_f = 3$ massless fermions must spontaneously break the G_F flavour symmetry, as long as the theory confines. You can check that the same argument works whenever N_f is a multiple of 3.

Decoupling Massive Quarks

When N_f is not a multiple of 3, things are not quite so simple. Indeed, we will need one further ingredient to complete the argument. To see this, let's look at the anomaly matching conditions for $G = SU(3)$ gauge theory with $N_f = 4$ flavours. They are:

$$\begin{aligned} A_1 &= 35p_1 - p_2 - 22p_3 + 6p_4 + 7p_5 = 3 \\ \frac{A_2}{3} &= 21p_1 + p_2 - 14p_3 - 2p_4 + 13p_5 = 1 \end{aligned}$$

Now there are solutions. For example $p_2 = 3$ and $p_4 = 1$ with $p_1 = p_3 = p_5 = 0$ does the job. This corresponds to four massless baryons in the representations

$$[3(\bar{\mathbf{4}}, \mathbf{1}) \oplus (\mathbf{4}, \mathbf{6})]_L \oplus [3(\mathbf{1}, \bar{\mathbf{4}}) \oplus (\mathbf{6}, \mathbf{4})]_R \tag{5.41}$$

where the L and R subscripts denote the chirality of these Weyl spinors. Note that the left-handed baryons now transform under both $SU(4)_L$ and $SU(4)_R$ of the chiral flavour symmetry.

Naively, the existence of the solution (5.41) suggests that there is a phase with massless baryons and the chiral symmetry left unbroken. In fact, this cannot happen. The problem comes when we think about giving one of the quarks a mass. We will make the following assumption: when we give a quark a mass, any baryon that contains this quark will also become massive. It is not obvious that this happens, and we will have to work harder below to justify this. But, for now, let's assume that this is true and see where it leads us.

If we give one of the quarks a mass, then the symmetry group is explicitly broken to

$$G_F = U(1)_V \times SU(4)_L \times SU(4)_R \;\longrightarrow\; G'_F = U(1)_V \times SU(3)_L \times SU(3)_R$$

What happens to our putative massless spectrum (5.41)? A little group decomposition tells us that under G'_F, the left-handed baryons transform as

$$3(\bar{\mathbf{4}}, \mathbf{1}) \rightarrow 3(\bar{\mathbf{3}}, \mathbf{1}) \oplus 3(\mathbf{1}, \mathbf{1}) \quad \text{and} \quad (\mathbf{4}, \mathbf{6}) \rightarrow (\mathbf{3}, \bar{\mathbf{3}}) \oplus (\mathbf{3}, \mathbf{3}) \oplus (\mathbf{1}, \bar{\mathbf{3}}) \oplus (\mathbf{1}, \mathbf{3})$$

The right-handed baryons have their $SU(3)_L \times SU(3)_R$ representations reversed. Of these, the $(\mathbf{1}, \mathbf{1})$ and the $(\mathbf{3}, \bar{\mathbf{3}})$ do not contain the massive fourth quark. By our assumption above, the remainder should become massive.

There is a further constraint however: all of the baryons that contain the fourth quark should become massive while leaving the surviving symmetry G'_F intact. This is because as the mass becomes large, we should return to the theory with $N_f = 3$ flavours and the symmetry group G'_F. Although we now know that G'_F will ultimately be spontaneously broken by the strong coupling dynamics, this should happen at the scale Λ_{QCD} and not at the much higher scale of the fourth quark mass.

So what G'_F-singlet mass terms can we write for the baryons that contain the fourth quark? The left-handed spinors transform as $3(\bar{\mathbf{3}},\mathbf{1})\oplus(\mathbf{3},\mathbf{3})\oplus(\mathbf{1},\bar{\mathbf{3}})\oplus(\mathbf{1},\mathbf{3})$. Of these, $(\mathbf{3},\mathbf{3})$ can happily pair up with its right-handed counterpart. Further, one of the $(\bar{\mathbf{3}},\mathbf{1})$ representations can pair up with the right-handed counterpart of $(\mathbf{1},\bar{\mathbf{3}})$. But that still leaves us with $2(\bar{\mathbf{3}},\mathbf{1})\oplus(\mathbf{1},\mathbf{3})$ and these have nowhere to go. Any mass term will necessarily break the remaining G'_F chiral symmetry and, as we argued above, this is unacceptable.

The result above should not be surprising. Any baryon that can get a mass without breaking G'_F does not change the 't Hooft anomaly for G'_F. If it were possible for all the baryons containing the massive quark to get a mass without breaking G'_F then the remaining massless baryons should satisfy anomaly matching. Yet we've seen that no such solution is possible for N_f.

The upshot of this argument is that there exists no solution to anomaly matching for $N_f = 4$ which is consistent with the decoupling of massive quarks. It is simple to extend this to all N_f and, indeed, to all N_c. 't Hooft anomaly matching then tells us that the chiral symmetry must be broken for all $N_c \geq 2$ and all $N_f \geq 3$.

5.6.2 Massless Baryons when $N_f = 2$?

There is one situation where it is possible to satisfy the anomaly matching: this is when $N_f = 2$. Since there is no triangle anomaly for $SU(2)$, we need only worry about the mixed $[SU(2)_L]^2 \times U(1)_V$ 't Hooft anomaly. We can import our results from earlier, although we should be a little bit careful: the anti-symmetric representation with two vertical boxes is just the singlet of $SU(2)$, while the representation with three boxes arranged as a dogleg is the same as the fundamental. Meanwhile, $\begin{array}{|c|}\hline r \\ \hline r \\ \hline\end{array}$ is the singlet of $SU(2)$. Meanwhile, the representation with three vertical boxes does not exist. The 't Hooft matching condition for gauge group $SU(3)$ now gives

$$\frac{A_2}{3} = 10p_1 - 5p_3 + p_4 = 1$$

This has many solutions. The simplest possibility $p_1 = p_3 = 0$ and $p_4 = 1$. This means that we can match the anomaly if there are massless baryons which transform under $SU(2)_L \times SU(2)_R \times U(1)_V$ as

$$(\mathbf{2}, \mathbf{1})_3 \oplus (\mathbf{1}, \mathbf{2})_3 \tag{5.42}$$

So for $N_f = 2$ we cannot use 't Hooft anomaly matching to rule out the existence of massless baryons. But it does not mean that they actually arise. To understand what happens, we need to look more carefully at the actual dynamics. The only real tool we have at our disposal is the lattice and this strongly suggests that even for $N_f = 2$ the chiral symmetry is broken and there are no massless baryons.

But what if...

Although the lattice tells us that the chiral symmetry is broken for $N_f = 2$, it is nonetheless an interesting exercise to understand better how we could have ended up with a massless baryon. The story that we will find has a nice twist and — as we will see in Section 5.6.4 — turns out to be realised in other contexts.

To start, let's return to our calculation of the classical force between quarks. We saw in Section 2.5.1 that a quark and anti-quark attract in the singlet channel and repel in the adjoint. This played a role in our initial discussion in Section 5.1 of why a quark condensate $\langle\bar{\psi}\psi\rangle$ might form in the first place.

However, we also saw in Section 2.5.1 that the two quarks attract in the anti-symmetric channel and repel in the symmetric channel. We might wonder if it's possible to form a condensate of quark pairs, rather than quark-anti-quark pairs. Such a condensate would break the gauge group.

In more detail, for $N_c = 3$ and $N_f = 2$ the initial gauge and global group of the theory is $G = SU(3)_{\rm gauge} \times SU(2)_L \times SU(2)_R \times U(1)_V$. The quarks transform as

$$\psi_- : \ (\mathbf{3}, \mathbf{2}, \mathbf{1})_1 \quad \text{and} \quad \psi_+ : \ (\mathbf{3}, \mathbf{1}, \mathbf{2})_1 \tag{5.43}$$

For $N_f = 2$, a condensate of quarks can take the form

$$\langle \psi^a_{+\,i} \psi^b_{+\,j} \rangle = \langle \psi^a_{-\,i} \psi^b_{-\,j} \rangle = -\epsilon^{abc} \epsilon_{ij} \sigma_c \tag{5.44}$$

Here the spinor indices are contracted so that the condensate is Lorentz invariant. The use of ϵ^{ij} means that the condensate is also invariant under the global $SU(2)_L \times SU(2)_R$ chiral symmetry. However, since the condensate σ_a transforms in the $(\mathbf{3} \otimes \mathbf{3})_{\rm anti-sym} = \bar{\mathbf{3}}$ of $SU(3)$, it breaks the gauge symmetry

$$G = SU(3)_{\rm gauge} \to SU(2)_{\rm gauge}$$

where we've added the "gauge" label because the number of different $SU(2)$ groups is about to get confusing. Naively it looks like the condensate (5.44) also breaks the $U(1)_V$ symmetry, but this can be restored by combining it with a suitable $U(1) \subset SU(3)_{\rm gauge}$. For example, if we take $\sigma_c = \sigma\delta_{1c}$ then the generator

$$Q'_V = Q_V + \text{diag}(2, -1, -1)_{\rm gauge}$$

is unbroken and commutes with $SU(2)_{\rm gauge}$. This means that, at low-energies, our theory has the symmetry

$$G' = SU(2)_{\rm gauge} \times SU(2)_L \times SU(2)_R \times U(1)'_V$$

How do the quarks (5.43) transform under G'? A little bit of representation decomposition shows

$$\psi_- : \ (\mathbf{1}, \mathbf{2}, \mathbf{1})_3 \oplus (\mathbf{2}, \mathbf{2}, \mathbf{1})_0 \quad \text{and} \quad \psi_+ : \ (\mathbf{1}, \mathbf{1}, \mathbf{2})_3 \oplus (\mathbf{2}, \mathbf{1}, \mathbf{2})_0$$

The existence of the condensate can be thought of as giving mass to the fermions that sit in the $\mathbf{2}$ of $SU(2)_{\rm gauge}$. (Note that, as in the condensate (5.44), we can form a singlet from $\mathbf{2}\otimes\mathbf{2}$ so there's no problem with either gauge invariance nor chiral symmetry.) But those fermions that are singlets under $SU(2)_{\rm gauge}$ are protected from getting a mass by the surviving $U(1)'_V$ chiral symmetry. The curious fact is that these massless fermions sit in precisely the representations (5.42) which satisfy 't Hooft anomaly matching.

There's something rather odd about this. In the 't Hooft anomaly matching argument, we assumed that the theory confines and looked for massless baryons – composites of three underlying quarks. In the analysis above, however, we proposed that the quark condensate Higgses the gauge group and the massless fermion is just a single quark, albeit with $U(1)'_V$ charge +3.

In fact, these are two different ways of looking at the same underlying physics. In the presence of the condensate (5.44), the vacuum is filled with pairs of quarks which can mix with the lone massless quark to form the composite baryon. Moreover, as we saw in Section 2.7.3, when we have a scalar in the fundamental representation — here played by the condensate $\psi\psi$ — there is no distinction between the Higgs and confining phases. The two descriptions — in terms of massless baryons or in terms of a condensate Higgs field — use different words, but are telling us the same thing. This situation sometimes goes by the rather pretentious name of *complementarity* (a much overused word in physics, and one which is possibly better saved for other, more subtle, phenomena).

Figure 47: Two possible behaviours for the baryon mass. The Vafa-Witten theorem rules out the second option.

As we mentioned above, it appears that the scenario sketched here doesn't occur for QCD-like theories with $N_f = 2$, presumably because the condensate which breaks chiral symmetry is preferred for more subtle, dynamical reasons. Nonetheless, something similar does happen for chiral gauge theories.

5.6.3 The Vafa-Witten-Weingarten Theorems

To invoke the full power of 't Hooft anomaly matching, we needed to assume that any baryon that contains a massive quark is itself massive. This is not at all obvious in a strongly interacting theory of the kind we're dealing with. When the mass of the quark is very large, $m \gg \Lambda_{\rm QCD}$, it is certainly true that the baryon must be massive. But for small quark masses $m \ll \Lambda_{\rm QCD}$, we could well imagine a situation where the binding energy cancels the quark mass, resulting in a massless bound state that contains massive constituents.

Two possibilities are depicted in Figure 47. The first shows the mass of the baryon increasing monotonically with the constituent quark mass. This is the scenario that we assumed above. The second figure shows another plausible scenario: the baryon remains massless for some finite value of the quark mass, before the theory undergoes some kind of phase transition at $m = m_\star$. If this were to happen, it would nullify our previous conclusions.

Fortunately, the second scenario cannot happen. It is ruled out by a theorem due to Vafa and Witten. In fact, there are a number of such theorems, all of which have rather similar proofs. We prove four such theorems below, the first two due to Vafa and Witten, the second two due to Weingarten. As we will see, the second Vafa-Witten theorem can be invoked to rule out the scenario shown in the second figure.

A Positive Definite Measure

Our setting is the QCD-like theories discussed throughout this chapter. All the theorems that we will prove rely on the same property of the path integral: a positive definite measure.

When computing correlation functions of gauge invariant operators, say $\mathcal{O}(x)$, we need to do the path integral. In Euclidean space, this takes the form

$$\langle \mathcal{O}(x) \ldots \mathcal{O}(y) \rangle = \frac{1}{Z} \int \mathcal{D}A \prod_{i=1}^{N_f} \mathcal{D}\bar{\psi}_i \, \mathcal{D}\psi_i \, e^{-S_{YM} + \sum \bar{\psi}_i (\not{D} + m)\psi_i} \, \mathcal{O}(x) \ldots \mathcal{O}(y)$$

Here S_{YM} is the usual Yang-Mills action. For simplicity, we've given each quark a common mass, m which we take to be positive: $m > 0$. Clearly it would be simple to generalise this. For some applications below, we'll explore the chiral limit by taking $m \to 0$. In practice, we should also include gauge fixing terms in this expression, but these don't affect the discussion below so we omit them for simplicity.

It is straightforward to do the fermionic path integral, leaving us with the path integral over the gauge fields. This takes the form

$$\langle \mathcal{O}(x) \ldots \mathcal{O}(y) \rangle = \frac{1}{Z} \int \mathcal{D}A \, e^{-S_{YM}} \left[\det(\not{D} + m) \right]^{N_f} \mathcal{O}(x) \ldots \mathcal{O}(y)$$

For many applications of interest, the operators $\mathcal{O}$ will also depend on the fermions. In this case, any fermion bi-linear should be replaced by its propagator in the usual manner. We'll see examples below.

We see that the effect of the fermions is to change the measure of the path integral over the gauge field. We write the correlation functions as

$$\langle \mathcal{O}(x) \ldots \mathcal{O}(y) \rangle = \int d\mu \, \mathcal{O}(x) \ldots \mathcal{O}(y)$$

where all the trickiness has now been absorbed in the measure

$$d\mu = \frac{1}{Z} \int \mathcal{D}A \, e^{-S_{YM}} \left[\det(\not{D} + m) \right]^{N_f} \tag{5.45}$$

The key observation is that this measure is positive definite. This is clearly true for the Yang-Mills part of the action, with $S_{YM} = \frac{1}{2g^2} \int \mathrm{tr}\, F^{\mu\nu} F_{\mu\nu}$. But it's also true for the Dirac operator. This is because QCD is a vector-like theory. Suppose that, for a choice of gauge field A_μ, the Dirac operator has a non-zero eigenvalue $\lambda \in \mathbf{R}$, so there is an eigenspinor

$$i\not{D}\psi = \lambda\psi$$

Then we also have an eigenvalue $-\lambda$. This follows because $\{\gamma^5, \not{D}\} = 0$, so

$$i\not{D}(\gamma^5\psi) = -i\gamma^5\not{D}\psi = -\lambda\gamma^5\psi$$

Of course, there may also be some number, n, of zero modes of $\not{D}$. The general form of the determinant is then

$$\det(\not{D} + m) = m^n \prod_\lambda (m - i\lambda)(m + i\lambda) = m^n \prod_\lambda (m^2 + \lambda^2) \tag{5.46}$$

which is manifestly positive definite providing $m > 0$. Before we go on, it's worth pausing to make a couple of comments.

- It's important that we set the theta angle to zero, $\theta = 0$, for the following arguments. This is because the theta term comes with an $\epsilon^{\mu\nu\sigma\rho}$ symbol,

$$S_\theta = \frac{\theta}{32\pi^2}\int \text{tr}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \tag{5.47}$$

  and so, when Wick rotated to Euclidean space, appears in the path integral as e^{iS_θ}. That extra factor of i means, when $\theta \neq 0$, the measure is not positive definite.

- Relatedly, the mass should be positive $m > 0$. We can see this explicitly in the contribution from the n zero modes in (5.46). But it's simpler to note that, from the chiral anomaly discussed in Section 3.3.3, a negative mass can be viewed as a non-zero θ angle.

- Clearly we needed the fermions to sit in a vector-like representation of the gauge group to argue for a positive definite measure. This means that many of the arguments we will make below fail in chiral gauge theories.

Let's now see what a positive definite measure buys us.

Theorem 1: Parity is not Spontaneously Broken

Parity is a symmetry of QCD. One might wonder if it remains a symmetry of the ground state. The spontaneous breaking of parity would show up as a non-zero expectation value for some parity odd scalar operator, $\mathcal{O}(x)$ which plays the role of an order parameter. We will argue that, in QCD, we necessarily have

$$\langle\mathcal{O}\rangle = 0$$

for any parity odd scalar.

To see this, consider the QCD Lagrangian deformed by the addition of this parity odd scalar.

$$\mathcal{L}(\alpha) = \mathcal{L}_{QCD} + \alpha\mathcal{O}$$

To leading order in α, the energy density of the ground state is

$$E(\alpha) = E(0) + \alpha\langle\mathcal{O}\rangle$$

If parity is spontaneously broken in QCD, then there are two ground states and $\langle\mathcal{O}\rangle$ is positive in one ground state and negative in the other. This means that spontaneous breaking of parity implies that $E(\alpha) < E(0)$ for arbitrarily small α.

Let's now calculate $E(\alpha)$ in the path integral. We have

$$e^{-VE(\alpha)} = \int d\mu \; e^{i\alpha\int d^4x\,\mathcal{O}}$$

where V is the volume of (Euclidean) spacetime. The important point is the factor of i in the exponent. This arises in Euclidean space only for parity odd operators because they necessarily come with an odd number of $\epsilon_{\mu\nu\rho\sigma}$. Indeed, we already saw an example of this with the θ term (5.47). We learn that adding a parity odd operator to the action changes the path integral by a phase. Because the measure is positive definite, this phase can only decrease the value of the path integral, so

$$e^{-V(\alpha)} \leq e^{-VE(0)} \quad \Rightarrow \quad E(\alpha) \geq E(0)$$

We learn that the energy density has a minimum at $\alpha = 0$ which, in turn, tells us that parity is not spontaneously broken in vector-like theories.

As a side remark, if we apply the argument above to the theta term itself (5.47), we learn that the addition of a theta term necessarily increases the energy of the vacuum: $E(\theta) \geq E(0)$. This observation sits at the heart of axion attempts to explain why the QCD theta angle is so small in our world.

Theorem 2: A Bound on Current-Current Correlation Functions

We now turn to the promised result: a relation between the masses of bound states and the bare masses of the underlying quarks. To proceed, we're going to consider two point functions of currents. We will take

$$J^a_\mu = \bar{\psi}_i\gamma_\mu(T^a)^{ij}\psi_j$$

where T^a is some $SU(N_f)$ flavour generator. In terms of the path integral, the two point function can be written as

$$\langle J^a_\mu(x)\, J^{b\,\dagger}_\nu(y)\rangle = \int d\mu \ \mathrm{tr}\ \big(\gamma_\mu T^a S(x,y)\gamma_\nu T^b S(y,x)\big)$$

where the trace is over spinor and flavour indices, and the propagator takes the form

$$S(x,y) = \langle x|\frac{1}{\not{D}+m}|y\rangle \tag{5.48}$$

Note that this is the propagator evaluated in the background of a fixed gauge field A_μ. The hard part is to then integrate over all gauge configurations, a procedure that is swept into the innocuous looking $\int d\mu$. Of course, we're not going to be able to do this integral. But we will be able to make remarkable progress simply from the knowledge that the measure is positive definite.

We will first give a slightly rough outline of the result, together with an explanation of why it shows what we want. We will then proceed with the proof and, along the way, see a number of further subtleties that we have to address.

The basic idea is to first fix A_μ. We would then like to invoke an inequality along the lines of

$$|S(x,y)| \leq C\, e^{-m|x,y|} \tag{5.49}$$

where m is the bare mass of the quark, C is some constant, and $|S(x,y)|$ refers to the matrix norm with respect to spinor and flavour indices. Crucially, this inequality must hold for any background gauge field A_μ, with the constants C and m independent of A_μ. In other words, it should be a *uniform* bound.

Such a uniform bound survives when averaged over all gauge fields with a positive definite measure. This then gives us a bound on the correlation function that we're interested in,

$$\langle J^a_\mu(x)\, J^{b\,\dagger}_\nu(y)\rangle < C' e^{-2m|x-y|} \tag{5.50}$$

What is the interpretation of such a bound? Suppose that the lightest particle carrying the flavour quantum numbers of the current has mass M. Then, at large distances, we would expect the current-current correlation function to be dominated by exchange of this particle, meaning that

$$\langle J^a_\mu(x)\, J^{b\,\dagger}_\nu(y)\rangle \sim e^{-M|x-y|}$$

The bound above tells us that the physical mass of the particle is bounded from below

$$M \geq 2m \tag{5.51}$$

where m is the bare mass of the quarks.

There are two immediate consequences of this result:

- First, it rules out the possibility of massless bosons. This is important because an equal bare mass for all the quarks breaks the axial symmetry, but leaves behind the vector $SU(N_f)_V$ flavour symmetry. The result (5.51) tells us that this vector flavour symmetry cannot be spontaneously broken, for it it were we would have massless Goldstone bosons with $M = 0$.

  We learn that the vector flavour symmetry is not spontaneously broken when the quarks have a bare mass. But if it's not spontaneously broken for any $m > 0$, then it can only become spontaneously broken in the limit $m \to 0$ if there is some miraculous accidental degeneracy, where a Lorentz invariant excited state decreases its energy, becoming exactly degenerate with the ground state at $m = 0$. This seems implausible. Under the assumption that no accidental degeneracy of this kind occurs, the Vafa-Witten theorem shows that vector-like symmetries are not spontaneously broken.

- Secondly, the Vafa-Witten theorem rules out the existence of massless fermions when the bare mass of the quarks are non-vanishing. This, of course, was what we wanted to prove.

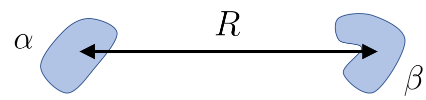
  Here, however, things are a little less straightforward and there is a subtlety that should be stressed. The calculation above holds in the presence of a finite UV cut-off, Λ. This was left implicit in the derivation, but the presence of the bare masses m in the inequality (5.51) is the hint that there is an underlying cut-off in the game. The Vafa-Witten theorem tells us that, for $m \neq 0$, there can be no massless composite fermion carrying flavour quantum numbers for any finite Λ. However, it does not rule out the possibility that a massless fermion emerges as $\Lambda \to \infty$ and the cut-off is removed. Indeed, we know that it is only in this limit that anomalies kick in so, strictly speaking, 't Hooft anomaly matching only requires the existence of massless fermions in the $\Lambda \to \infty$ limit. For QCD, it is not believed that such behaviour happens. But the Vafa-Witten theorem isn't as watertight as we might hope in showing this.

A Proof of Theorem 2

Let's now prove the Vafa-Witten theorem (5.50). The trick is not to work with the propagator (5.48) between position eigenstates $|x\rangle$ and $|y\rangle$, but instead to work with a smeared propagator

$$S(\alpha,\beta) = \langle\alpha|\frac{1}{\not{D}+m}|\beta\rangle$$

where $|\alpha\rangle$ and $|\beta\rangle$ are wavepackets that have support only in localised regions, separated by a distance R as shown below:



The "localised support" means that $A_\mu(x)|\alpha\rangle = 0$ for x outside of the region α, and similar for $|\beta\rangle$. We'll soon see the advantage of working with these smeared propagators.

To proceed, we use a standard trick to rewrite the propagator as

$$S(\alpha,\beta) = \int_0^\infty dt\ \langle\alpha|e^{-(\not{D}+m)t}|\beta\rangle = \int_0^\infty dt\ e^{-mt}\langle\alpha|e^{-i(-i\not{D})t}|\beta\rangle$$

Here t is just an artificial parameter that we've used to rewrite the integral. The next step is the clever one: we reinterpret t as a genuine time direction for a theory in $d = 4+1$ dimensions with Hamiltonian $H = -i\not{D}$. By causality, we know that a signal from region α takes at least time $t = R$ to reach the region β. This means that we must have

$$\langle\alpha|e^{-iHt}|\beta\rangle = 0 \quad \text{ for } 0 \leq t \leq R$$

Furthermore, at later times we can simply use the Cauchy-Schwarz inequality to bound

$$|\langle\alpha|e^{-iHt}|\beta\rangle| \leq \sqrt{\langle\alpha|\alpha\rangle}\,\sqrt{\langle\beta e^{+iHt}|e^{-iHt}\beta\rangle} = \sqrt{\langle\alpha|\alpha\rangle}\,\sqrt{\langle\beta|\beta\rangle}$$

where, in the second equality, we have used the unitarity of e^{iHt}. This then gives us the promised uniform bound on the propagator

$$|S(\alpha,\beta)| \leq\ |\alpha|\,|\beta| \int_R^\infty dt\ e^{-mt} = \frac{|\alpha|\,|\beta|}{m}e^{-mR} \tag{5.52}$$

This is more or less the result that we wanted. It's not quite the advertised bound on the propagator (5.49) because it uses wavepackets rather than position eigenstates. Nonetheless, it's just as good for the purposes of proving what we want. The derivation also makes it clear why we needed smeared wavepackets; it's because their norm $|\alpha|$ and $|\beta|$ appear explicitly in the bound. In contrast, position eigenstates aren't normalisable and so don't work for our purposes.

Theorem 3: The Pion is the Lightest Meson

There are yet more applications of the positive definite measure. These are inequalities between the masses of various physical particles, first introduced by Weingarten. The first of these says that the pion is the lightest meson.

We start by introducing the pseudoscalar meson field

$$\pi = \bar{\psi}_i \gamma^5 \psi_j$$

where we have picked some $i \neq j$. In QCD, we would most naturally pick $i =$ up quark and $j =$ down quark, so that π is identified with the genuine pion. Here, we'll refer to π as the pion for any i and j. We give all quarks the same mass, $m > 0$, so that the $SU(N_f)_V$ vector symmetry is unbroken. The propagator of the pion is then

$$\langle \pi(x) \pi^\dagger(y) \rangle = \int d\mu \; \text{tr} \left[S(x,y)\, \gamma^5\, S(y,x)\, \gamma^5 \right]$$

where $d\mu$ is the usual positive-definite measure (5.45), $S(x,y)$ is the fermion propagator introduced in (5.48) and where the trace is over spinor and colour indices. Now, because $\{\gamma^5,\, \not{D}\} = 0$, we have

$$\gamma^5 S(y,x) \gamma^5 = \gamma^5 \langle y | (\not{D} + m)^{-1} | x \rangle\, \gamma^5 = \langle y | (-\not{D} + m)^{-1} | x \rangle = \langle x | (\not{D} + m) | y \rangle^\dagger \quad (5.53)$$

where the final equality follows because $\not{D}$ is anti-Hermitian. Note that the $|x\rangle$ and $|y\rangle$ labels got swapped as part of taking the Hermitian conjugate; the remaining $\dagger$ acts on colour and spinor indices. But this means that we have

$$\langle \pi(x) \pi^\dagger(y) \rangle = \int d\mu \sum_{\text{colour,spinor}} |S(x,y)|^2 \geq 0 \quad (5.54)$$

We learn that the pion propagator is positive definite. Note that the presence of the γ^5 matrix was crucial to make this claim, since it gave us the Hermitian conjugate in (5.53).

Let's now contrast this with the propagator for a scalar meson that doesn't include the γ^5. We have

$$\sigma = \bar{\psi}_i \psi_j$$

where we take $i \neq j$ to be the same indices as those carried by the pion. Repeating the arguments above, we now get

$$\langle \sigma(x) \sigma^\dagger(y) \rangle = \int d\mu \; \text{tr} \left[S(x,y)\, S(y,x) \right] = \int d\mu \; \text{tr} \left[S(x,y) \gamma^5 S(x,y)^\dagger \gamma^5 \right]$$

This means that we're again summing over $|S(x,y)|^2$, but this time with different plus and minus signs for different spinor indices, coming from the presence of γ^5 matrices in the final expression. We learn that we necessarily have

$$\langle \sigma(x)\sigma^\dagger(y)\rangle \;\leq\; \langle \pi(x)\pi^\dagger(y)\rangle$$

But, at large distances, we expect each of these correlation functions to be dominated by the mass of the corresponding meson (or the mass of the lightest particle carrying the same quantum numbers). This means that the inequality above becomes, for large $|x-y|$,

$$e^{-m_\sigma|x-y|} \leq e^{-m_\pi|x-y|} \quad \Rightarrow \quad m_\sigma \geq m_\pi$$

This, of course, holds in our world because the pion is a Goldstone boson for broken chiral symmetry. Indeed, the mass inequality above can, like the Vafa-Witten theorem, be used to argue against the vector-like symmetry being broken, for then the σ meson would be a massless Goldstone boson. The result above says that this can't happen at finite m where the pion is massive, and so the sigma meson must also be massive.

It is straightforward to repeat the arguments above with a different gamma matrix structure. For example, we could look at vector mesons of the form $\rho = \bar{\psi}\gamma^\mu\psi$ and show that these too are heavier than the pions.

Theorem 4: Baryons are Heavier than Pions

The second Weingarten inequality bounds the mass of the baryon. For QCD, with three colours, the baryon takes the form

$$B = \epsilon^{abc}\psi^i_a\psi^j_b\psi^k_c$$

Here i,j and k are flavour indices and a,b and c are colour indices. The spinor indices are left implicit; they could be contracted to form a spin-$\frac{1}{2}$ baryon, or uncontracted for a spin-$\frac{3}{2}$ baryon. As we've seen, two-point correlation function takes the form

$$\langle B(x)B^\dagger(y)\rangle \sim e^{-m_B|x-y|}$$

where m_B is the mass of the lightest baryon sharing the quantum numbers of B. As previously, we take the bare masses of all quarks to be $m > 0$. We then have the expression

$$\langle B(x)B^\dagger(y)\rangle = \epsilon^{abc}\epsilon^{a'b'c'}\int d\mu\ \mathrm{tr}\left[S(x,y)_{aa'}S(x,y)_{bb'}S(x,y)_{cc'}\right]$$

where $d\mu$ is again the positive-definite measure (5.45) and this time we've kept the colour indices explicit. First, we use the Cauchy-Schwarz inequality to bound

$$\langle B(x)B^\dagger(y)\rangle \leq \int d\mu \left(\sum_{a,a',\text{spinor}} |S(x,y)_{aa'}|^2\right)^{3/2}$$

Suppose that we could argue that, for any choice of background gauge field,

$$|S(x,y)| \leq C' e^{-m|x-y|} \tag{5.55}$$

with C' a constant, independent of the choice of gauge field, and m the bare mass of the quark. In this case, we would immediately have

$$\langle B(x)B^\dagger(y)\rangle \;\leq\; C' e^{-m|x-y|}\int d\mu \sum_{\text{colour,spinor}} |S(x,y)|^2 = C' e^{-m|x-y|}\,\langle \pi(x)\pi^\dagger(y)\rangle$$

where, in the second equality, we've used our previous expression for the pion propagator (5.54). Now, recall from the proof of the Vafa-Witten theorem that we don't quite have (5.55), but we have something almost as good: we need to replace the position eigenstates $|x\rangle$ and $|y\rangle$ with smeared wavepackets $|\alpha\rangle$ and $|\beta\rangle$ and we can then derive the uniform bound (5.52). This will do for our purposes; we therefore come to the conclusion that $m_B \geq m_\pi + m$. In the limit that the bare mass vanishes, so $m \to 0$, we learn that

$$m_B \geq m_\pi$$

Of course, this is hardly groundbreaking information given what we know about particle physics. But here it is derived from first principles, with no assumption of chiral symmetry breaking. Moreover, it tells us that if we wish chiral symmetry to be unbroken, with the 't Hooft anomaly saturated by massless baryons, then the pion must also be massless. But this seems very unlikely, since if the pion is massless then there is nothing to stop it condensing and breaking the chiral symmetry after all.

5.6.4 Chiral Gauge Theories Revisited

The existence of a global symmetry with a 't Hooft anomaly guarantees the existence of massless particles in the spectrum. If the symmetry is spontaneously broken, we have Goldstone bosons. If the symmetry is unbroken, we have massless fermions whose presence is needed to reproduce the anomaly.

So far, we have discussed situations in which 't Hooft anomaly matching ensures the existence of massless bosons (together with the case of $N_f = 2$ where anomaly matching is ambivalent, but bosons arise anyway). Here we describe situations where massless fermions arise. Perhaps unsurprisingly, this typically happens in chiral gauge theories where tree-level fermion masses are prohibited by the gauge symmetry.

We will focus on one of the simplest chiral gauge theories,

$$G = SU(5) \text{ with two Weyl spinors: } \psi_a \text{ in the } \bar{\mathbf{5}} \text{ and } \chi^{ab} \text{ in the } \mathbf{10}$$

Here $a, b = 1, \ldots, 5$ are the gauge group indices. The classical theory has two global symmetries: $U(1)_\psi$ and $U(1)_\chi$, each rotating the phase of a single fermion. One combination of these suffers a mixed anomaly with $SU(5)$. The surviving generator is

$$Q = 3Q_\psi - Q_\chi$$

This has a 't Hooft anomaly

$$A = \sum_{\text{fermions}} Q^3 = 5 \times 3^3 + 10 \times (-1)^3 = 125$$

Let us now suppose that the theory confines, leaving the $U(1)_Q$ unbroken. The simplest colour singlet is the 3-fermion bound state

$$\psi_a \psi_b \chi^{ab} \tag{5.56}$$

This has charge $Q = 5$, giving an infra-red contribution to the 't Hooft anomaly

$$A = 5^3 = 125$$

We see that it is plausible that this fermion bound state does indeed remain massless.

A Different Perspective

We can reach the same conclusion through a rather different argument. Suppose that a fermion bi-linear forms a condensate. Since any such bilinear is necessarily charged under the gauge group, the condensate will partially Higgs the gauge symmetry. What symmetry breaking patterns occur?

This is not completely straightforward. We can make a number of different fermion bilinears, each decomposing into some number of channels. Based on the computation of the classical force between quarks described in section 2.5.1, some of these channels will be attractive and some repulsive. It seems likely that the condensate forms in an attractive channel, but there are several of these.

At this point, we need to use a little guesswork. The most naive approach is to determine which quark pair has the most attractive force and assume that the condensate forms in this channel. This is clearly optimistic — after all, we're dealing with a strongly coupled theory and the classical force calculation is unlikely to provide quantitative guidance — but does give sensible answers in many cases. It is known as the *maximally attractive channel* criterion. More generally in these situations, one tries different possibilities and sees which outcomes seem the least baroque. Note that, in contrast to the QCD-like theories, we cannot turn to the lattice for help because there are various obstacles to discretising chiral fermions.

For the problem in hand, it is thought that the naive, most-attractive channel hypothesis does give rise to the correct physics. In fact, there are two channels which are equally attractive. These are:

$$\mathbf{5} \subset \bar{\mathbf{5}} \otimes \mathbf{10} \quad \text{and} \quad \bar{\mathbf{5}} \subset \mathbf{10} \otimes \mathbf{10}$$

We therefore postulate the existence of two quark condensates

$$\langle \psi_a \chi^{ab} \rangle = \sigma^b \quad \text{and} \quad \langle \chi^{ab} \chi^{cd} \rangle = \epsilon^{abcde} \Delta_e \tag{5.57}$$

These two condensates are not gauge invariant. Between them, they could break the $SU(5)$ gauge group to either $SU(4)$ (if they lie parallel to each other) or $SU(3)$. Again, we have to engage in a little guesswork. We will assume that they line up, with $\sigma^a = \sigma\delta^{a1}$ and $\Delta_a = \Delta\delta_{a1}$. The gauge group is then broken to

$$G = SU(5)_{\text{gauge}} \to SU(4)_{\text{gauge}}$$

Naively, each of the condensates breaks the non-anomalous $U(1)$ global symmetry, with $Q(\sigma) = 2$ and $Q(\Delta) = -2$. However, as in the previous section, we can define a new, unbroken global symmetry by mixing the $U(1)$ with a suitable generator of the $SU(5)$ gauge symmetry,

$$Q' = Q - \frac{1}{2}\text{diag}(4, -1, -1, -1, -1)$$

At low-energies, the gauge and global symmetry groups are

$$G = SU(4)_{\text{gauge}} \times U(1)'$$

Decomposing each fermion into representations of this new group, we have

$$\psi : \ \bar{\mathbf{5}}_3 \ \to \ \bar{\mathbf{4}}_{5/2} \oplus \mathbf{1}_5 \quad \text{and} \quad \chi : \ \mathbf{10}_{-1} \ \to \ \mathbf{6}_0 \oplus \mathbf{4}_{-5/2}$$

The $\langle \psi\chi \rangle$ condensate in (5.57) gives mass to $\bar{\mathbf{4}}_{5/2} \otimes \mathbf{4}_{-5/2}$, while the $\langle \chi\chi \rangle$ condensate gives mass to $\mathbf{6}_0 \otimes \mathbf{6}_0$. This leaves us with the gauge singlet $\mathbf{1}_5$. This has the same quantum numbers as the massless composite fermion (5.56) that we anticipated by 't Hooft anomaly matching.

Although we've had to engage in some guesses along the way, we end up with a plausible situation: the low energy dynamics of the chiral $SU(5)$ theory consists of a single, free Weyl fermion. This can either be viewed as a composite fermion (5.56) in a confining theory, or as a fundamental fermion in a theory with quark condensates (5.57): the end result is the same.

We could also ask if there are other possibilities which look equally plausible. For example, is it possible that the global $U(1)_Q$ is spontaneously broken, resulting in a massless boson instead of a massless fermion? For this to happen, we need to construct a bosonic, gauge invariant condensate. The simplest contains six fermions — $\psi_a\psi_b\chi^{ab}\psi_c\psi_d\chi^{cd}$ — and it seems unlikely that such a condensate would form.

More Chiral Gauge Theories

The $SU(5)$ gauge theory described above is not the only one which is thought to confine, giving massless composite fermions. Indeed, the same behaviour is thought to occur for the two classes of chiral gauge theories introduced in Section 3.4.2. The first of these is:

$$G = SU(N) \text{ with a } \begin{smallmatrix}\square\\[-2pt]\square\end{smallmatrix} \text{ and } N-4\ \overline{\square} \text{ Weyl fermions}$$

We denote the $N-4$ fermions in the anti-fundamental representation as ψ and the fermion in the anti-symmetric as χ. The theory has a $SU(N-4)\times U(1)$ global symmetry, where the $SU(N-4)$ factor rotates the ψ fields, while the non-anomalous $U(1)$ charges are given by

$$Q_\psi = N-2 \quad \text{and} \quad Q_\chi = 4-N$$

Assuming that this theory confines, the question is: what becomes of this global symmetry. As we have seen, if it is to survive unscathed then there must be a massless, composite fermion that reproduces the 't Hooft anomaly. A candidate is the collection of 3-fermion bound states that, schematically, take the form $\lambda = \psi\chi\psi$. Displaying all the indices, this is

$$(\lambda_\alpha)_{ij} = \psi_{\beta a i}\,\epsilon^{\beta\gamma}\,\chi_\gamma{}^{ab}\psi_{\alpha b j} \tag{5.58}$$

where $\alpha,\beta,\gamma = 1,2$ are spinor indices, $i,j = 1,\ldots,N-4$ are $SU(N-4)$ flavour indices, and $a,b = 1,\ldots,N$ are $SU(N)$ gauge indices. If you track through all the symmetry properties, you'll find that λ_{ij} is symmetric in ij, so this spinor transforms in the symmetric $\square\!\square$ representation of the $SU(N-4)$ global symmetry group. It also has charge $Q = N$. It is not hard to check that these massless fermions λ do indeed saturate the 't Hooft anomalies, and therefore provide a good candidate for the infra-red physics of this theory.

As with the $SU(5)$ model, there is also a complementary approach to deriving the same result in which one first assumes that fermi bilinears $\langle\chi\psi\rangle$ and $\langle\chi\chi\rangle$ condense, breaking the gauge group $SU(N) \to SU(4)$, with all fermions pairing up except for a lone λ with the same quantum numbers that we saw above.

The second chiral gauge theory that we met earlier is similar, but has quarks in the symmetric rather than anti-symmetric representation

$$G = SU(N) \text{ with a } \square\!\square \text{ and } N+4\ \overline{\square}$$

This time there is a global $SU(N+4)$ symmetry, together with a single non-anomalous $U(1)$ under which the anti-fundamental fermions ψ and the symmetric fermion χ have charges

$$Q_\psi = N+2 \quad \text{and} \quad Q_\chi = -(N+4)$$

Once again, it seems plausible that the theory confines without breaking the $SU(N+4)\times U(1)$ global symmetry, with the 't Hooft anomalies saturated by a fermion (5.58). Tracking through the symmetrisation, this time λ sits in the anti-symmetric $\begin{smallmatrix}\square\\\square\end{smallmatrix}$ representation of the global symmetry group $SU(N+4)$, again with charge $Q = N$. A few short calculations show that the 't Hooft anomalies do indeed match.

5.7 Further Reading

Spontaneous symmetry breaking is a powerful and unifying idea, explaining disparate phenomena in both particle physics and condensed matter physics. It is responsible for the existence of phonons in a solid and, as we have seen, the existence of pions in the strong force. When implemented in gauge theories, it provides a unified explanation for superconductivity and the electroweak vacuum.

Jeffery Goldstone was the first to realise that a spontaneously broken global symmetry gives rise to a massless particle – what we now call the Goldstone boson. He made this conjecture, and provided examples, in a 1961 paper whose title – "Field theories with Superconductor Solutions" – reveals the early cross-fertilisation between condensed matter and particle physics [78]. The general proof of the theorem followed soon afterwards in a paper with Salam and Weinberg [79].

Goldstone's theorem was initially viewed with some dismay in particle physics. The existence of strictly massless bosons was ruled out by experiment, suggesting that spontaneous symmetry breaking had little role to play at the fundamental level. This, of course, was too hasty. Subsequent work by Higgs and others, exploring symmetry breaking in gauge theories, provided the underpinning for the Standard Model. Meanwhile, it was realised that an approximate global symmetry could be spontaneously broken, resulting in an approximate Goldstone boson. (The name pseudo-Goldstone boson was coined by Weinberg, apparently to Jeffrey's annoyance.)

The discovery of what we would now call chiral symmetry was actually made slightly before Goldstone's insight. In 1960, Yoichiro Nambu explained that an exact axial-vector current in beta decay would imply the existence of a massless pion field [140]. Like many papers of the time, it avoids the language of field theory and instead focusses on the "current algebra", in which one works with commutation relations between currents and their matrix elements. This somewhat masks the connection to spontaneous symmetry breaking, which is not emphasised in the paper. This was one of the (many!) contributions for which Nambu was awarded the 2008 Nobel prize.

A more modern formulation of the chiral Lagrangian came only in the mid-1960s. Gell-Mann and Levy introduced the sigma model [72]. In fact, they introduced two versions: the first is what we might call a "linear sigma model" and includes the field σ, related to the pion fields by a constraint $\sigma^2 + \vec{\pi}^2 = 1$. Embarrassed by the new field which had not been observed in experiments, they subsequently integrated out to derive the "non-linear sigma model", now named after a particle that does not exist and does not appear in anywhere in the theory. The group-theoretic formulation of the non-linear sigma model that we used here is due to Weinberg [206], and was extended to general groups in [22].

The idea that baryons could arise as solitons in the chiral Lagrangian was proposed by Tony Skyrme, in a remarkably prescient pair of papers written in 1960 and 1961 and [184, 185]. These papers were apparently written without any awareness of the work described above, and were essentially ignored for more than a decade while the story of chiral symmetry breaking unfolded. The papers came to prominence only in the 1980s when it was realised that they played an important role in the story. The term "skyrmion" was coined in a 1984 meeting in honour of Tony Skyrme. (In a cute twist, the second paper thanks "Mr A. J. Leggatt" for performing the calculations as an undergraduate student. This mis-spelled student went to win the Nobel prize.)

The WZW term was introduced by Witten in 1983 [226]. The arguments in Section 5.5 are largely taken from this paper. (Many of Witten's papers from this time are masterclasses in clarity; the best way to learn much of modern physics is simply to read Witten's papers.) As we saw, for $N_f = 2$ there is no WZW term, but the fact that topology can determine the quantum statistics of the Skyrmion was noted by Finkelstein and Rubinstein, back in 1968 [60].

More on the history of chiral symmetry breaking can be found in the article by Weinberg [209]. More details about the physics of chiral symmetry breaking can be found in the lecture notes of Scherer and Schindler [173] and Peskin [153].

The idea that anomalies place severe constraints on the spectrum of strongly interacting gauge theories was first emphasised by 't Hooft in the lectures [105], with the application to chiral symmetry breaking that we described in these lectures. This was elaborated on by Frishman, Schwimmer, Banks and Yankielowicz, [64]. The "persistent mass condition", prohibiting the formation of massless bound states using massive constituents, was framed by Preskill and Weinberg [163] and found a more rigorous grounding in the Vafa-Witten theorems [195, 196]. The mass inequalities, which also make use of the positive definite measure, were first introduced by Weingarten (very) slightly before the Vafa-Witten theorem [210]. The idea that the Higgs and confining phases provide complementary, but equivalent, viewpoints on the dynamics of chiral gauge theories was first enunciated in [189].