6 Supersymmetric QCD

We now turn our attention to the quantum dynamics of supersymmetric gauge theories. Our focus will be on understanding the physics of super Yang-Mills and super QCD. There is, as we shall see, a wonderfully rich array of behaviour on display.

First, some basics. There are a number of facts that we've seen already in these lectures that we can combine to great effect in supersymmetric theories. First, we know that the gauge coupling runs

$$\frac{1}{g^2(\mu)} = \frac{1}{g_0^2} - \frac{b_0}{(4\pi)^2}\log\frac{\Lambda_{UV}^2}{\mu^2}$$

where g_0^2 is the coupling constant evaluated at the cut-off scale Λ_{UV}. The general expression for the 1-loop beta function in non-supersymmetric theories is (5.7)

$$b_0 = \frac{11}{6}I(\text{adj}) - \frac{2}{6}\sum_{\text{fermions}} I(R_f) - \frac{1}{6}\sum_{\text{scalars}} I(R_s)$$

In supersymmetric theories this simplifies. Gauge bosons are necessarily accompanied by an adjoint Weyl fermion and chiral multiplets come in fermion/boson pairs. The upshot is that

$$b_0 = \frac{3}{2}I(\text{adj}) - \frac{1}{2}\sum_{\text{chirals}} I(R) \tag{6.1}$$

In the quantum theory, the running gauge coupling is replaced by the dynamical scale Λ, below which the non-Abelian gauge theory is strongly coupled. For reasons that will become clear shortly, we will refer to this as $|\Lambda|$. (It was always a real, positive energy scale so there's nothing lost in doing this.) This was defined in (5.4) as

$$|\Lambda| = \mu\exp\left(-\frac{8\pi^2}{b_0 g^2(\mu)}\right)$$

It is RG invariant, meaning that Λ is independent of the scale μ.

Importantly, something novel happens in supersymmetric theories. This is because, as we have seen, the gauge coupling constant sits as the imaginary part of a complex coupling (4.18)

$$\tau(\mu) = \frac{\vartheta}{2\pi} + \frac{4\pi i}{g^2(\mu)} \tag{6.2}$$

The theta angle does not run, essentially because it is a periodic variable $\vartheta \in [0, 2\pi)$ and so has nowhere to go. This motivates us to define the *complexified strong coupling scale*

$$\Lambda = \mu \exp\left(\frac{2\pi i \tau(\mu)}{b_0}\right) = |\Lambda| e^{i\vartheta/b_0} \tag{6.3}$$

Recall from Section 3.3 that superpotentials are holomorphic in both fields and parameters. The complexified scale Λ is therefore crying out to sit in the superpotential. We'll see many examples of this as we proceed.

The complexified scale also ties together two other ideas that we've encountered previously. First, when discussing what kinds of superpotentials can arise in a quantum theory in Section 3.3, we found it useful to think of a larger class of symmetries under which parameters also transform as so-called "spurions". Of course, if a symmetry changes a parameter then it's not a true symmetry of the theory but nonetheless we saw that these spurious symmetries can prove useful in restricting the kind of behaviour that can occur in supersymmetric theories.

Second, when discussing chiral anomalies in Section 5.2, we saw that a symmetry of the classical theory can fail to be a symmetry of the quantum theory by shifting the theta angle (5.26). In the supersymmetric context, a transformation of theta angle manifests itself as a complex rotation of Λ. This means that Λ acts as a spurion for anomalous $U(1)$ symmetries. It also means that we can use anomalous symmetries to restrict the form of quantum corrections to a theory, just as we used other broken symmetries in Section 3.3. Again, we'll see many examples of this as we proceed.

A Comment on Exact Beta Functions

There is an interesting, and somewhat subtle, story about higher order corrections to the beta function. We can write the one-loop correction in a more revealing way by inverting (6.3),

$$\tau(\Lambda; \mu) = \frac{b_0}{2\pi i} \log\left(\frac{\Lambda}{\mu}\right) \tag{6.4}$$

Importantly, the periodicity of $\vartheta \in [0, 2\pi)$ is manifest on both sides of this equation through

$$\vartheta \to \vartheta + 2\pi \quad \Leftrightarrow \quad \tau \to \tau + 1 \quad \Leftrightarrow \quad \Lambda \to \Lambda e^{2\pi i/b_0}$$

Any corrections to (6.4) should retain this property. But that's tricky to achieve while retaining the holomorphy implied by supersymmetry. The most general form of holomorphic corrections, consistent with the periodicity of ϑ, is

$$\tau(\Lambda;\mu) = \frac{b_0}{2\pi i}\log\left(\frac{\Lambda}{\mu}\right) + \sum_{n=1}^{\infty} a_n \left(\frac{\Lambda}{\mu}\right)^{b_0 n} \tag{6.5}$$

for some unknown coefficients a_n. (The restriction to $n > 0$ comes from requiring that this is a weak coupling expansion and should not diverge as $\Lambda \to 0$.) But these additional terms are proportional to $e^{-8\pi^2 n/g^2}$ and are identified as instanton effects (5.32). We see that all higher perturbative contributions vanish and, as far as perturbation theory is concerned, the beta function is one-loop exact.

The fact that the beta function is one-loop exact in supersymmetric theories is a striking statement. It appears to be even more striking when you actually compute the two-loop contribution and find that it doesn't vanish! What's going on?

The resolution is that one should be careful about what quantity is actually being computed. The holomorphic gauge coupling τ originates in a superpotential term $\int d^2\theta\ \tau W^\alpha W_\alpha$ such that $1/g^2$ sits in front of the Yang-Mills action. The story that we told above assumes a renormalisation scheme in which this holomorphy is protected.

Meanwhile, the physical gauge coupling is computed after a rescaling $A_\mu \to gA_\mu$, so that the coupling now appears in vertices. But absorbing the gauge coupling into the gauge field in this way is not an entirely innocent thing to do and there is a price to pay in the form a Jacobian in the path integral. This means that while the holomorphic gauge coupling is one-loop exaxt, the physical gauge coupling can, and does, receive contributions at all loops[4]. (It's not dissimilar to our discussion in Section 3.3 where we saw that the physical parameters are renormalised even though the superpotential is not.)

Nonetheless, it turns out that the one-loop exactness of the holomorphic gauge coupling puts strong constraints on the beta function for the physical gauge coupling which is known as the NSVZ beta function (after Novikov, Shifman, Vainshtein, and Zakharov).

[4]You can read more about these issues in the paper by Nima Arkani-Hamed and Hitoshi Muryama.

6.1 Super Yang-Mills

We will start our study of quantum dynamics with pure super Yang-Mills. The theory consists of a non-Abelian gauge field coupled to a single, adjoint Weyl fermion,

$$S_{\text{SYM}} = \int d^4x \; \text{Tr}\left[\frac{1}{g^2}\left(-\frac{1}{2}F_{\mu\nu}F^{\mu\nu} - 2i\lambda\sigma^\mu\mathcal{D}_\mu\bar{\lambda}\right) + \frac{\vartheta}{16\pi^2}F_{\mu\nu}{}^\star F^{\mu\nu}\right]$$

We will work with gauge group $G = SU(N_c)$.

The one-loop beta function (6.1) is $b_0 = 3N_c$ and the theory flows to strong coupling at the scale $|\Lambda|$. The question that we want to answer is: what happens?

6.1.1 Confinement and Chiral Symmetry Breaking

Our first port of call is to understand the global symmetries of the theory. Classically the theory has a $U(1)_R$ symmetry, under which

$$U(1)_R: \; \lambda \to e^{i\alpha}\lambda$$

This symmetry does not survive quantisation: it suffers an anomaly which can be viewed as a transformation of the theta angle

$$U(1)_R: \; \vartheta \to \vartheta + I(\text{adj})\,\alpha = \vartheta + 2N_c\alpha \tag{6.6}$$

Equivalently, we can think of the strong coupling scale (6.3) transforming as

$$U(1)_R: \; \Lambda \to e^{2i\alpha/3}\Lambda$$

We say that Λ has R-charge $R[\Lambda] = \frac{2}{3}$. As we've stressed repeatedly, the shift of ϑ means that $U(1)_R$ is not a symmetry of the quantum theory.

However, all is not lost. We can see from (6.6) that a shift by $\alpha = 2\pi/2N_c$ transforms $\vartheta \to \vartheta + 2\pi$. This means that a discrete $\mathbb{Z}_{2N_c}$ subgroup of the R-symmetry survives, rotating the fermion as

$$\lambda \to \omega\lambda \quad \text{with} \quad \omega^{2N_c} = 1$$

We learn that $SU(N_c)$ super Yang-Mills has a discrete $\mathbb{Z}_{2N_c}$ R-symmetry.

Next we should start to understand the quantum dynamics. We don't have enough control over the strong coupling physics of $\mathcal{N} = 1$ supersymmetric theories to show from first principles that theory confines. (It turns out that we do have such control in theories with $\mathcal{N} = 2$ supersymmetry.) We assume that, as with pure Yang-Mills, the theory confines with a mass gap. There is little doubt that this is correct.

Furthermore, as in non-supersymmetric QCD, a fermion bilinear forms

$$\langle \, \mathrm{Tr}\lambda\lambda \, \rangle \sim \Lambda^3 \tag{6.7}$$

This time supersymmetry does help us get a handle on this. We'll see how as we proceed through this section and, in particular, will be able to pin down the dimensionless coefficient that sits in front of the right-hand side. But first let us understand the consequences of the condensate.

As in non-supersymmetric QCD, this condensate spontaneously breaks a symmetry. The difference is that it in super Yang-Mills the condensate breaks our discrete R-symmetry,

$$\langle \, \mathrm{Tr}\lambda\lambda \, \rangle \rightarrow \omega^2 \langle \, \mathrm{Tr}\lambda\lambda \, \rangle$$

This, however, is a spontaneous breaking rather than an explicit breaking: the theory is invariant under $\mathbb{Z}_{2N_c}$ but the ground state is not. The discrete R-symmetry is broken to

$$\mathbb{Z}_{2N_c} \rightarrow \mathbb{Z}_2$$

where the surviving $\mathbb{Z}_2$ acts as fermion parity $\lambda \rightarrow -\lambda$. This is subgroup of the $Spin(1,3)$ Lorentz group and, as such, cannot be spontaneously broken.

When a continuous symmetry is spontaneously broken, we get massless Goldstone modes. When a discrete symmetry is spontaneously broken, we get multiple ground states. These ground states are characterised by the phase of the gluino condensate (6.7) which, in general, can take the form

$$\langle \, \mathrm{Tr}\lambda\lambda \, \rangle = a\omega^{2k}\Lambda^3 \qquad k = 0, 1, \ldots, N_c - 1 \tag{6.8}$$

with $\omega = e^{\pi i/N_c}$ and $a \in \mathbb{R}$ an undetermined coefficient. The upshot is that $SU(N_c)$ super Yang-Mills has N_c distinct ground states that differ by the phase of the condensate (6.8)

Before we go on, it's worth pointing out that the condensate takes the form

$$\Lambda^3 \sim e^{-8\pi^2/g^2 N_c} e^{i\theta/N_c}$$

This *isn't* of the form (5.32) expected from an instanton contribution. Roughly, it looks like the contribution from $1/N_c$ of an instanton! But we should acknowledge that the condensate arises in the strongly coupled regime of the theory and instantons are not a good guide to what's going on.

So far we haven't managed to figure out the overall constant a in front of the condensate. In non-supersymmetric theories, the equivalent calculation is not possible. But in supersymmetric theories it can be done, albeit with a fairly technical computation. Conceptually the idea is to deform the theory so that it is weakly coupled. We then compute the gluino condensate in that regime and argue, using holomorphy, that it remains unchanged as we move back. The end result is

$$a = 16\pi^2 \tag{6.9}$$

There are (at least) two methods to get this result. One is to study the theory on $\mathbb{R}^3 \times \mathbf{S}^1$ rather than $\mathbb{R}^4$. It turns out that the theory can be made weakly coupled when the $\mathbf{S}^1$ has radius $R \ll 1/|\Lambda|$. Moreover, rather wonderfully, when placed on a circle instantons actually do fractionalise into N_c smaller objects and can be shown to generate the gluino condensate[5]. We'll see another method to determine $a = 16\pi^2$ later in these lectures.

6.1.2 The Witten Index

There is another way to see the existence of N_c supersymmetric ground states. This is to compute the Witten index, defined in Section 3.4.2 as

$$\mathrm{Tr}\,(-1)^F e^{-\beta H}$$

This counts the number of supersymmetric ground states of the theory, weighted with a sign.

The beauty of the Witten index is that it stays the same no matter what you do to the theory as long as you preserve supersymmetry. This means that if we can deform super Yang-Mills in some way so that the theory becomes weakly coupled, then we can just compute the Witten index using standard perturbative quantum field theory, safe in the knowledge that it can't then change as we deform back to the strongly coupled regime that we care about. So the question becomes: how can we make super Yang-Mills weakly coupled?

The way to do this is fairly dramatic. We consider the theory on a spatial torus $\mathbf{T}^3$ and take the radius of each circle to be R, so that the volume is $V = (2\pi R)^3$. We know

[5] This calculation can be found in the paper by Davies, Hollowood, Khoze and Mattis . Be warned: the computation of background determinants in this paper is incorrect, although the final answer is right.

that super Yang-Mills is weakly coupled in the UV, but flows to strong coupling at a scale $|\Lambda|$. If we take the spatial torus to be very small, so that

$$R \ll \frac{1}{|\Lambda|}$$

then the RG flow never reaches strong coupling. Of course, the physics of the theory on such a tiny spatial torus is very different from the physics that we might care about. In particular, the size of space is now much smaller than the Compton wavelength of any massive particle so this is not going to be any good to compute, say, the S-matrix. But there's one thing that we can compute and that's the Witten index.

When we compactify space in this way, nearly all states will have an energy set by $E \sim 1/R$. We can ignore these if we want to compute the number of ground states and focus only on those modes that, classically, have zero energy. These degrees of freedom come from both the gauge field and the fermions and we deal with each in turn.

On a torus $\mathbf{T}^3$, there are gauge configurations A_i that have vanishing field strength $F_{ij} = 0$, but are nonetheless not gauge equivalent to the vacuum. These are parameterised by mutually commuting holonomies around each of the three different cycles

$$U_i = \mathrm{Tr}\,\mathcal{P}\exp\left(i\oint A_i\right) \qquad i = 1, 2, 3$$

where $\mathcal{P}$ is path ordering. We can use an $SU(N_c)$ gauge transformation to diagonalise each of these, so that they read

$$U_i = \mathrm{diag}(e^{i\theta^i_1}, \ldots, e^{i\theta^i_{N_c}})$$

The zero energy modes are the coordinates θ^i_a, with $i = 1, 2, 3$ labelling the spatial directions and $a = 1, \ldots, N_c$ the gauge indices. Because $U_i \in SU(N_c)$, these coordinates are not all independent but are constrained to obey

$$\sum_{a=1}^{N_c} \theta^a_i = 0 \bmod 2\pi \tag{6.10}$$

We should quantise each of these periodic rotors θ^a_i, subject to this constraint. But this is essentially the same as the quantisation of a particle on a circle and we know that there is a unique ground state in which the wavefunction is independent of the θ's. Physically, this can be understood because a non-zero momentum for θ corresponds to non-Abelian electric field $F_{0i} \neq 0$. This means that there's no subtlety in quantising the gauge field and we get a unique ground state[6].

[6]A different way to count ground states can be found in Witten's original paper "Constraints on Supersymmetry Breaking".

We're left with the adjoint fermion. We impose periodic boundary conditions and the zero modes are simply the constant modes over the torus. We can again diagonalise the fermions by an $SU(N_c)$ gauge transformation and write

$$\lambda_\alpha = \text{diag}(\lambda_\alpha^1, \ldots, \lambda_\alpha^{N_c})$$

with $\alpha = 1, 2$ the spinor index. Each of these is a complex Grassmann mode. Because λ sits in the algebra $su(N_c)$, these are constrained to obey

$$\sum_{a=1}^{N_c} \lambda_\alpha^a = 0 \tag{6.11}$$

Let's first recall what usually happens with such modes in quantum mechanics. A single Grassmann mode ψ has anti-commutation relations $\{\psi, \psi^\dagger\} = 1$ and gives rise to a qubit. This arises by first defining a fiducial state $|0\rangle$ that obeys $\psi|0\rangle = 0$. The Hilbert space then consists of two states $|0\rangle$ and $\psi^\dagger|0\rangle$.

We can quantise the zero modes λ_α^a in the same way, except we have to make sure that the end result is gauge invariant. Diagonalising λ has already exhausted much of the gauge symmetry, but we're still left with the Weyl group which permutes the λ_α^a. This means that any wavefunction must be invariant such permutations.

We begin by again introducing a fiducial state that obeys $\lambda_\alpha^a|0\rangle = 0$ for all $\alpha = 1, 2$ and $a = 1, \ldots, N_c$. We can build zero energy excited states by acting with $(\lambda_\alpha^a)^\dagger$, subject to the requirement of gauge invariance and (6.11). It's straightforward to see that there is no such state where we excite just a single $(\lambda_\alpha^a)^\dagger$: the requirement that it is invariant under permutations means that it has to take the form $\sum_a (\lambda_\alpha^a)^\dagger|0\rangle$ but this vanishes by virtue of (6.10).

There is a single state with two $(\lambda_\alpha^a)^\dagger$ excited. We first construct the gauge invariant combination

$$S = \text{Tr}\,\lambda\lambda = \sum_{a=1}^{N_c} \epsilon^{\alpha\beta}\lambda_\alpha^a\lambda_\beta^a$$

and then build a ground state $S^\dagger|0\rangle$. All gauge invariant states with more $\lambda^\dagger$ excitations then arise by acting with further copies of $S^\dagger$. The end result is that there are N_c ground states, given by

$$|k\rangle = (S^\dagger)^k|0\rangle \qquad k = 0, \ldots, N_c - 1$$

The series ends at $|N_c - 1\rangle$ because the Grassmann nature of λ_α^a, together with the constraint (6.10), means that $(S^\dagger)^{N_c} = 0$.

G	$SU(N)$	$Sp(N)$	$\mathrm{Spin}(2N+1)$	$\mathrm{Spin}(4N)$	$\mathrm{Spin}(4N+2)$	E_6	E_7	E_8	F_4	G_2
h	N	$N+1$	$2N-1$	$4N-2$	$4N$	12	18	30	9	4

Table 3. The dual Coxeter number h for all simply connected gauge groups.

Each of the states $|k\rangle$ contains an even number of Grassmann operators and so contributes to the Witten index with the same sign. We learn that in the regime $R \ll 1/|\Lambda|$, where the theory is weakly coupled, the Witten index of $SU(N_c)$ super Yang Mills is given by

$$\mathrm{Tr}\,(-1)^F e^{-\beta H} = N_c$$

But now we are at liberty to take R as large as we like, safe in the knowledge that the Witten index does not change. Indeed, the counting above agrees with the expectations from discrete chiral symmetry breaking (6.8), although the physics underlying these N_c states looks very different in the two regimes.

Other Gauge Groups

There is a similar story for other gauge groups G. The R-symmetry group of super Yang-Mills $\mathbb{Z}_{2h}$ where h is a group theoretic quantity known as the *dual Coxeter number*. The value of h is shown for various groups G in Table 3. The fermionic condensate (6.7) then spontaneously breaks

$$\mathbb{Z}_{2h} \rightarrow \mathbb{Z}_2$$

giving h distinct vacua. Similarly, one can compute the Witten index on $\mathbf{T}^3$ to find the same result[7]

$$\mathrm{Tr}\,(-1)^F e^{-\beta H} = h$$

In fact, there is a further subtlety in the computation on $\mathbf{T}^3$. It turns out that the Witten index depends on the global structure of the gauge group meaning that, for example, the number of supersymmetric ground states for $G = \mathrm{Spin}(N)$ and $G = SO(N)$ are different. You can read more about this in Yuji Tachikawa's lecture notes.

6.1.3 A Superpotential

Later in this section we will derive Wilsonian effective actions for light degrees of freedom. But for super Yang-Mills there are no light degrees of freedom. The theory has mass gap, with the lightest states having mass around $\sim |\Lambda|$.

[7]The original Witten index paper contains a subtle mistake for $\mathrm{Spin}(N)$ gauge groups that was corrected by Witten in a subsequent appendix, with further elaborations in this paper.

Nonetheless, there is an interesting effective action that we can write down. It doesn't involve any dynamical degrees of freedom and instead depends only the parameter Λ. We've already seen that the R-charge of this parameter is $R[\Lambda] = 2/3$ and the superpotential must have R-charge 2, which means that the only thing we can write down is

$$W_{\text{eff}} = c\Lambda^3 \tag{6.12}$$

for some, as yet, undetermined constant c.

What's the meaning of such an effective action when it doesn't contain any dynamical fields? In fact, it's just another way of capturing the gluino condensate (6.7). Here we explain why.

First, recall how we compute expectation values in the path integral. We add a source $J(x)$ for the operator of interest. We then compute the path integral in the presence of the source

$$Z[J] = \int D(\text{fields})\, e^{iS_{SYM}} \exp\left(i\int d^4x\ J\,\text{Tr}\lambda\lambda + \text{h.c.}\right) \tag{6.13}$$

The expectation value is then given by

$$\langle\, \text{Tr}\,\lambda\lambda\,\rangle = \left.\frac{\partial \log Z}{\partial J}\right|_{J=0}$$

Now let's go back to the original action for super Yang-Mills, written in terms of superfields (4.19)

$$S_{\text{SYM}} = -\int d^4x\ \left[\int d^2\theta\ \frac{i\tau}{8\pi}\,\text{Tr}\,W^\alpha W_\alpha + \text{h.c.}\right]$$

The lowest component of the chiral superfields is $W^\alpha W_\alpha = \lambda^\alpha\lambda_\alpha + \ldots$. But this means that a source for the gluino bilinear naturally arises if we promote the parameter τ to be a chiral superfield with its full complement of components

$$\tau = \tau + \sqrt{2}\theta\psi_\tau + \theta^2 F_\tau$$

The source appears as the F-term: $J = F_\tau/8\pi$.

The low-energy effective action is what we get when we do the path integral, so

$$Z[J] = e^{iS_{\rm eff}}$$

To write the effective action we again promote τ to a chiral superfield. There can be a complicated Kähler potential for τ but this doesn't concern us. (It will give terms proportional to $F_\tau F_\tau^\dagger$ but these will vanish when we set $J = 0$ in (6.13).) All we need for our purposes is the contribution to $S_{\rm eff}$ from an effective superpotential

$$S_{\rm eff} \supset \int d^4x\, d^2\theta\ W_{\rm eff} + \text{h.c.} = \int d^4x\ \frac{\partial W_{\rm eff}}{\partial \tau} F_\tau + \text{h.c.}$$

The goal is to write down a $W_{\rm eff}$ that captures the right physics. Repeating the steps above, we have

$$\langle\, \text{Tr}\, \lambda\lambda\, \rangle = 8\pi i \frac{\partial S_{\rm eff}}{\partial F_\tau} = 8\pi i \frac{\partial W_{\rm eff}}{\partial \tau}$$

In this way, the effective superpotential is simply a device to encode the value of the gluino condensate.

With these path integral gymnastics under our belt, let's now turn to the superpotential (6.12). As we've seen, it's the only thing that we can write down consistent with the (anomalous) R-symmetry. In terms of τ is is

$$W_{\rm eff} = c\mu^3 e^{2\pi i \tau/N_c} \quad \Rightarrow \quad \langle\, \text{Tr}\, \lambda\lambda\, \rangle = \frac{16\pi^2 c}{N_c} \Lambda^3$$

in agreement with our previous result (6.8). To match the normalisation (6.9), the coefficient c should be

$$c = N_c \tag{6.14}$$

Note that $W_{\rm eff}$ hasn't taught us anything new about the theory. In particular, there's nothing to fix the coefficient c and we will have some work to do to make sure that it's non-vanishing. However, it will turn out that $W_{\rm eff}$ will be useful in making contact with the results that we will derive from SQCD.

6.2 A First Look at SQCD

Now we add matter. We will consider supersymmetric QCD: $SU(N_c)$ gauge theory coupled to N_f massless flavours. In superspace, the Lagrangian is

$$\mathcal{L}_{SQCD} = \text{Tr}\left[\int d^2\theta\ \frac{i\tau}{8\pi} W^\alpha W_\alpha + \text{h.c.}\right] + \int d^4\theta \sum_{i=1}^{N_f} \left[\Phi_i^\dagger e^{2V} \Phi^i + \tilde{\Phi}^{i\,\dagger} e^{-2V} \tilde{\Phi}_i\right]$$

The action written in component fields can be found in (4.21) .

Each flavour consists of two chiral multiplets, Φ in the fundamental representation $\mathbf{N}_c$ and $\tilde{\Phi}$ in the conjugate representation $\bar{\mathbf{N}}_c$. The one-loop beta function (6.1) is

$$b_0 = 3N_c - N_f$$

For $N_f \geq 3N_c$, the theory is non-renormalisable and infra-red free. Here the low-energy physics is easy. We want to understand what happens when $N_f < 3N_c$.

6.2.1 Symmetries

The first step in understanding any quantum field theory is to get the symmetries nailed down. Let's start with the classical symmetries. These are:

	$SU(N_c)$	$SU(N_f)_L$	$SU(N_f)_R$	$U(1)_B$	$U(1)_A$	$U(1)_{R'}$
ϕ	$\square$	$\square$	$\mathbf{1}$	1	1	0
$\tilde{\phi}$	$\overline{\square}$	$\mathbf{1}$	$\overline{\square}$	-1	1	0
ψ	$\square$	$\square$	$\mathbf{1}$	1	1	-1
$\tilde{\psi}$	$\overline{\square}$	$\mathbf{1}$	$\overline{\square}$	-1	1	-1
λ	adj	$\mathbf{1}$	$\mathbf{1}$	0	0	1

Some obvious comments to make sure that we're all on the same page. The first column denotes the $SU(N_c)$ gauge symmetry; all others are flavour symmetries. For the non-Abelian symmetries, $\square$ denotes the fundamental, $\overline{\square}$ denotes the anti-fundamental, and $\mathbf{1}$ means that it is a singlet.

(As an aside: the symmetries above are actually incomplete for $N_c = 2$ because the fundamental $\mathbf{2}$ is pseudoreal and so equivalent to the $\bar{\mathbf{2}}$. This gives an enhanced $SU(2N_f)$ symmetry. We won't need this subtlety in what follows.)

Both $U(1)_B$ and $U(1)_A$ are flavour symmetries, as evidenced by the fact that the scalars and fermions in the same multiplet transform the same way. Meanwhile, $U(1)_{R'}$ is an R-symmetry, meaning that the component fields in a chiral multiplet transform as

$$R[\text{fermion}] = R[\text{boson}] - 1 \tag{6.15}$$

We've called this symmetry $U(1)_{R'}$ rather than $U(1)_R$ for a reason that will become clear shortly. The choice of $R[\phi] = 0$ is somewhat arbitrary since we could always define a new R-symmetry by combing it with any amount of the global A-symmetry. The important point is that the R-charge of the scalars ϕ and fermions ψ differ by 1. Note that the gluino λ always has charge $+1$ under the R-symmetry.

Not all the classical symmetries survive quantisation. $U(1)_B$ is left unscathed as it is vector-like, but both $U(1)_A$ and $U(1)_{R'}$ suffer chiral anomalies. As we saw in (5.22), the current conservation equation becomes

$$\partial_\mu j^\mu = \frac{\mathcal{A}}{32\pi^2}\,\mathrm{Tr}\,F_{\mu\nu}{}^\star F^{\mu\nu} \quad \text{with} \quad \mathcal{A} = \sum_{\text{fermions}} qI(R)$$

where q is the charge and R the representation under $SU(N_c)$. Again, it's worth stressing that the complex scalars ϕ and $\tilde{\phi}$ do not contribute to the anomaly. It is just the fermions that have this subtlety. For the two symmetries $U(1)_A$ and $U(1)_{R'}$, we have

$$\mathcal{A}_A = N_f \times 1 + N_f \times 1 = 2N_f \tag{6.16}$$

and

$$\mathcal{A}_{R'} = N_f \times (-1) + N_f \times (-1) + 2N_c \times 1 = 2(N_c - N_f)$$

However, we can form a linear combination of these currents that remains conserved. This is given by

$$R = R' + \frac{N_f - N_c}{N_f} A$$

This is an R-symmetry, rather than a flavour symmetry, because the chiral multiplet components still obey (6.15) and $R[\lambda] = 1$. (The convention of fixing the normalisation by insisting that $R[\lambda] = 1$ comes with the unhappy side effect that other charges are fractional.) We can now draw up a table of the true quantum symmetries of the theory:

	$SU(N_c)$	$SU(N_f)_L$	$SU(N_f)_R$	$U(1)_B$	$U(1)_R$
ϕ	$\square$	$\square$	$\mathbf{1}$	1	$\frac{N_f - N_c}{N_f}$
$\tilde{\phi}$	$\overline{\square}$	$\mathbf{1}$	$\overline{\square}$	-1	$\frac{N_f - N_c}{N_f}$
ψ	$\square$	$\square$	$\mathbf{1}$	1	$-\frac{N_c}{N_f}$
$\tilde{\psi}$	$\overline{\square}$	$\mathbf{1}$	$\overline{\square}$	-1	$-\frac{N_c}{N_f}$
λ	adj	$\mathbf{1}$	$\mathbf{1}$	0	1

However, this misses some crucial information. This is because, as we've seen previously, it's useful to keep the anomalous symmetry as a spurious symmetry. The full symmetry structure of the theory should be thought of as reinstating the anomalous $U(1)_A$, but with a transformation on Λ showing that it's not a true symmetry of the theory:

	$SU(N_c)$	$SU(N_f)_L$	$SU(N_f)_R$	$U(1)_B$	$U(1)_A$	$U(1)_R$
Φ	$\square$	$\square$	$\mathbf{1}$	1	1	$\frac{N_f-N_c}{N_f}$
$\tilde{\Phi}$	$\overline{\square}$	$\mathbf{1}$	$\overline{\square}$	-1	1	$\frac{N_f-N_c}{N_f}$
Λ^{b_0}	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{1}$	0	$2N_f$	0

Some of the previous information is hidden in this table. In particular, the R-symmetry charge is that of the scalar component of the chiral multiplet and you have to remember that $R[\text{fermion}] = R[\text{boson}] - 1$, together with the fact that $R[\lambda] = 1$. The final row shows how the anomalous symmetries act on $\Lambda^{b_0} \sim e^{i\vartheta}$. We see that Λ transforms only under the anomalous $U(1)_A$, with the charge given by (6.16). We'll have cause to return to this table a number of times in what follows.

6.2.2 Runaway for $N_f < N_c$

The dynamics of SQCD will depend crucially on the ratio N_f/N_c. We start with small number of colours

$$N_f < N_c$$

We already discussed the classical theory back in Section 4.3. The theory has a moduli space of vacua $\mathcal{M}$ parameterised by the N_f^2 gauge invariant, massless meson fields

$$M_j{}^i = \tilde{\Phi}_j \Phi^i$$

At a generic point on the moduli space $\mathcal{M}$, the gauge group is spontaneously broken to

$$SU(N_c) \to SU(N_c - N_f) \tag{6.17}$$

The mesons are neutral under $SU(N_c - N_f)$ (otherwise they would break it further) so, at the classical level, we have massless $SU(N_c - N_f)$ gauge bosons essentially decoupled from the massless mesons. We want to know what happens in the quantum theory.

We already know what will happen to the $SU(N_f - N_c)$ gauge bosons: they will confine and get a mass. That leaves us with the mesons. It's useful to start by asking: what could possibly happen? At the crudest level, the massless fields could remain massless, or they too could get a mass. If the latter happens, it would manifest itself in terms of a potential generated on the moduli space. And this potential would appear in the form of a superpotential. So we should check if it's possible that quantum corrections generate a superpotential that lifts the moduli space.

Such a superpotential should be written in the terms of the low-energy meson fields and must respect the various symmetries of the problem. The meson field itself transforms in the $(\Box, \overline{\Box})$ of $SU(N_f)_L \times SU(N_f)_R$, so to get something invariant we should consider $\det M$. Under the remaining $U(1)$ symmetries, the relevant charges are then

	$U(1)_B$	$U(1)_A$	$U(1)_R$
$\det M$	0	$2N_f$	$2(N_f - N_c)$
$\Lambda^{3N_c - N_f}$	0	$2N_f$	0

Recall that the superpotential should have R-charge $R[W] = 2$ and must be neutral under $U(1)_A$ and $U(1)_B$. There is a unique combination that is allowed by symmetries

$$W_{\text{eff}} = C\left(\frac{\Lambda^{3N_c - N_f}}{\det M}\right)^{\frac{1}{N_c - N_f}} \tag{6.18}$$

with some coefficient underdetermined coefficient $C = C(N_c, N_f)$.

We've learned that symmetries allow for a superpotential only of the specific form (6.18). But is it actually generated? In other words, is $C \neq 0$? There is a general rule of thumb in quantum field theory that anything that isn't prohibited by some symmetry or other principle always occurs. The superpotential (6.18) is constructed to be invariant under all symmetries. It is also physically sensible, with a positive power of Λ reflecting the fact that it could be generated by strong coupling effects. Indeed, it turns out that it is generated with the coefficient $C(N_c, N_f)$ given by

$$C(N_c, N_f) = N_c - N_f$$

The result (6.18) is known as the *Affleck-Dine-Seiberg*, or *ADS*, superpotential. We'll give an incomplete explanation of how to determine $C(N_c, N_f)$ in Section 6.2.4.

Note that if we set $N_f = 0$, then the ADS superpotential agrees with our previous result (6.12) that captures the gluino condensate. However, when $N_f \geq 1$, the superpotential W_{eff} is a function of dynamical fields M and tells us the fate of those fields.

First, let's understand the physics of the superpotential W_{eff}. The moduli space of vacua is a large dimensional space but we can get a sense for what happens if we think of $\det M \sim M^{N_f}$. The superpotential is then $W_{\text{eff}} \sim M^{-N_f/(N_f - N_c)}$. If we ignore the Kähler potential, then the scalar potential takes the form

$$V(M, M^\dagger) \sim \left|\frac{\partial W_{\text{eff}}}{\partial M}\right|^2 \to 0 \quad \text{as} \quad |M| \to \infty$$

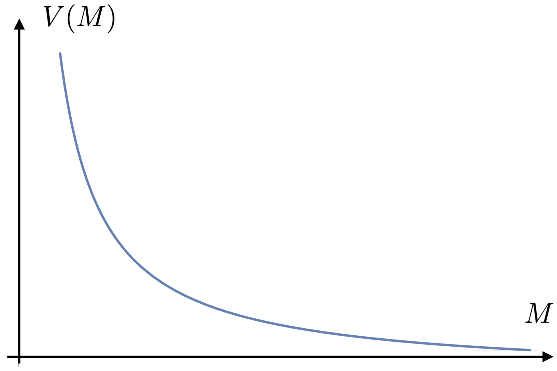

Figure 9. The runaway potential on the moduli space for $N_f < N_c$ massless flavours.

This is rather striking behaviour. Classically we had an infinite number of vacua, forming the moduli space $\mathcal{M}$. Quantum mechanically we have none! The potential is non-zero everywhere, asymptoting to $V \to 0$ only as $M \to \infty$ as shown schematically in Figure 9. This is known as a *runaway potential*. We have a quantum theory with no ground state. This is not something that we saw in non-supersymmetric QCD. Indeed, it should be clear that it arises in SQCD only because of the existence of massless scalars and their moduli space.

There are a number of caveats regarding the form of the potential, all deriving from the fact that we don't have good control over the Kähler potential which, as we know from (3.29), affects the actual potential $V(M)$. In some circumstances, it may well be possible that $V(M)$ does not increase monotonically towards the interior of the moduli space but has some local, non-supersymmetric, minima at $V(M) \neq 0$. If so, these would be metastable ground states, with some finite lifetime before tunnelling out and rolling down to infinity.

6.2.3 Adding Masses

The runaway behaviour arises for massless matter. What happens if we add a mass term? This arises from the addition of a superpotential to the our original theory,

$$W_{\text{mass}} = m^j{}_i \tilde{Q}_j Q^i$$

with m^i_j the mass matrix. (Sorry for the proliferation of "M" variables. To remind you, M is the meson, m is the mass, and $\mathcal{M}$ is the moduli space!) We can always use the $SU(N_f)$ symmetries to diagonalise the mass matrix

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

However, in what follows we won't lose anything by considering a general m.

We care about the low-energy physics. We can again play the same game to determine the superpotential using symmetries and holomorphy. In addition to M and Λ, we now also have the mass matrix m. The transformation properties of the fields and parameters are

	$SU(N_f)_L$	$SU(N_f)_R$	$U(1)_B$	$U(1)_A$	$U(1)_R$
M	$\square$	$\overline{\square}$	0	2	$\frac{2(N_f-N_c)}{N_f}$
$\Lambda^{3N_c-N_f}$	$\mathbf{1}$	$\mathbf{1}$	0	$2N_f$	0
m	$\overline{\square}$	$\square$	0	-2	$\frac{2N_c}{N_f}$

Again, we can ask: what possible superpotentials are consistent with the symmetry? The answer is that we can have any function

$$W_{\rm eff} = \left(\frac{\Lambda^{3N_c-N_f}}{\det M}\right)^{\frac{1}{N_c-N_f}} f(x)$$

where $f(x)$ is any holomorphic function of the unique holomorphic variable x that is invariant under all symmetries

$$x = {\rm Tr}(mM)\left(\frac{\det M}{\Lambda^{3N_c-N_f}}\right)^{\frac{1}{N_c-N_f}}$$

We can pin down the function $f(x)$ by taking various limits. In the limit $m \to 0$ and $\Lambda \to 0$, we must have $f(x) = C + x$ so the superpotential is just the sum of the mass term and the dynamically generated superpotential (6.18),

$$W_{\rm eff} = (N_c - N_f)\left(\frac{\Lambda^{3N_c-N_f}}{\det M}\right)^{\frac{1}{N_c-N_f}} + {\rm Tr}(mM) \tag{6.19}$$

But this limit encompasses all possible values of x, meaning that this is the exact superpotential.

What is the physics now? We can start by looking at the case $N_f = 1$ where there is a just a single complex meson $M = \tilde{\Phi}\Phi$. The superpotential now has a critical point,

$$\frac{\partial W_{\rm eff}}{\partial M} = 0 \quad \Rightarrow \quad M^{N_c} = \frac{\Lambda^{3N_c-1}}{m^{N_c-1}} \tag{6.20}$$

This is an interesting result. First, there is now a supersymmetric minimum, with the potential sketched in Figure 10. Moreover, there are actually N_c such minima coming

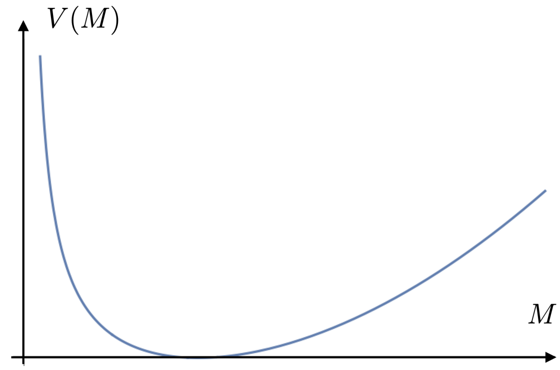

Figure 10. The rescued runaway, with a supersymmetric minimum when mass is added.

from taking the N_c^{th} root in (6.20). This is to be expected since it coincides with the Witten index for super-Yang Mills. As the mass $m \to 0$, the minima move off to infinity in field space. In the opposite regime, $|m| \gg |\Lambda|$, the flavour decouples and the theory reduces to super Yang-Mills.

Decoupling

We can look more closely at what happens in the limit $|m| \gg |\Lambda|$. For simplicity, we'll take m real in what follows. Clearly this theory should reduce to super Yang-Mills but, to make this precise, we need to be more careful about the strong coupling scales. In particular, when we try to decouple some heavy degrees of freedom like this, there are two strong coupling scales at play. This is because the running of the gauge coupling happens in two steps:

- $E > m$: Here the gauge coupling runs with the beta function $b_0 = 3N_c - 1$ that is appropriate for $N_f = 1$ flavours. We have

$$\frac{1}{g^2(\mu)} = \frac{1}{g_0^2} - \frac{b_0}{(4\pi)^2} \log\left(\frac{\Lambda_{UV}^2}{\mu^2}\right)$$

  If we continued this running to energies lower than m then we would hit strong coupling at a scale that we will call

$$\Lambda_{\text{old}} = \Lambda_{UV} e^{-8\pi^2/b_0 g_0^2} = m e^{-8\pi^2/b_0 g^2(m)}$$

  where, in the second equality, we've used the fact that Λ is an RG invariant. This Λ_{old} is the scale Λ that appears in the formulae (6.19) and (6.20) above. However, when the chiral multiplets have a mass, it is better thought of as something of a counterfactual scale. The RG running never gets as low as $\Lambda_{\text{old}} < m$ because something changes along the way . . .

- $E < m$: Now the massive chiral multiplets decouple and no longer contribute to the beta function which becomes that of pure super Yang-Mills, with $b_0' = 3N_c$. We can continue the running of the gauge coupling with this new beta function, now starting at the scale m

$$\frac{1}{g^2(\mu)} = \frac{1}{g^2(m)} - \frac{b_0'}{(4\pi)^2} \log\left(\frac{m^2}{\mu^2}\right)$$

  Now it hits strong coupling at a scale that we will call

$$\Lambda_{\text{new}} = m e^{-8\pi^2/b_0' g^2(m)}$$

  This is the actual scale at which the gauge coupling becomes strong.

Comparing the two results above, we have the matching condition

$$\left(\frac{\Lambda_{\text{old}}}{m}\right)^{b_0} = \left(\frac{\Lambda_{\text{new}}}{m}\right)^{b_0'} \tag{6.21}$$

In principle there can be additional multiplicative factors that arise from the matching at scale m at higher loops. These go by the name of *threshold effects*. One can always choose a regularisation scheme in which they vanish.

The result (6.21) can be used generally. For our specific purposes, we decouple from the theory with $N_f = 1$ to pure super Yang-Mills, and this equation reads

$$\Lambda_{\text{old}}^{3N_c - 1} m = \Lambda_{\text{new}}^{3N_c}$$

In this case, $\Lambda_{\text{new}} > \Lambda_{\text{old}}$. This is because the presence of matter slows the running of the coupling. When that matter is removed, the running speeds up and so raises the strong coupling scale.

We can now evaluate the formulae (6.19) and (6.20) in terms of the true, low-energy scale Λ_{new}. First we determine the expectation value M in the vacuum (6.20). Then we substitute this into the superpotential (6.19) at the vacuum. A short calculation shows that

$$W_{\text{eff}} = N_c \, \Lambda_{\text{new}}^3$$

This, of course, we've seen before. It is precisely the superpotential (6.12) for super Yang-Mills, now with the strong coupling scale Λ_{new}. Even the coefficient (6.14) comes out correctly. In this way, the Affleck-Dine-Seiberg superpotential correctly predicts the value of the gluino condensate in super Yang-Mills.

A General Mass Matrix

We can repeat the calculation above for N_f flavours and a general mass matrix m_{ij}. We just need to find the critical point

$$\frac{\partial W_{\rm eff}}{\partial M^{ij}} = 0$$

of the superpotential (6.19). To do so, we should Jacobi's formula

$$\delta(\det M) = {\rm tr}({\rm Adj}(M)\,\delta M) \tag{6.22}$$

with ${\rm Adj}(M)$ the adjugate matrix. If M is invertible then this coincides with the more familiar $\delta(\det M) = (\det M)\,{\rm tr}(M^{-1}\delta M)$. Assuming that M is indeed invertible, we find that the critical point obeys

$$M_j{}^i = (m^{-1})_j{}^i \left(\frac{\Lambda^{3N_c-N_f}}{\det M}\right)^{\frac{1}{N_c-N_f}} \tag{6.23}$$

We take the determinant of both sides to find

$$\det M = \frac{1}{\det m}\left(\frac{\Lambda^{3N_c-N_f}}{\det M}\right)^{\frac{N_f}{N_c-N_f}} \quad \Rightarrow \quad M_j{}^i = (m^{-1})_j{}^i \left(\det m\, \Lambda^{3N_c-N_f}\right)^{1/N_c}$$

Again, we see that the vacua sit at a position inversely proportional to the mass, ensuring that they move off to infinity as $m \to 0$. The $N_c^{\rm th}$ root on the right-hand side provides the phase ambiguity that gives rise to the N_c ground states expected from the Witten index.

6.2.4 The Potential at Weak Coupling

There is something special that happens when $N_f = N_c - 1$. This is because, with this number of flavours, at a generic point on the moduli space $\mathcal{M}$ the gauge group is generically completely broken.

This is important. For any $N_f < N_c - 1$, there is always a residual unbroken $SU(N_c - N_f)$ non-Abelian gauge group which means that the theory is necessarily strongly coupled. However, for $N_f = N_c - 1$ the theory can be weakly coupled.

However, weak coupling isn't guaranteed. For simplicity, let's consider the point on the moduli space where all scalars have the same expectation value (4.37),

$$\phi_a^i = \tilde{\phi}_a^{\dagger\, i} = \left(\begin{array}{ccc:c} v & \ldots & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \ldots & v & 0 \end{array}\right) \tag{6.24}$$

The Higgs mechanism halts the running of the gauge coupling at the scale v of breaking, so in the infra-red $g^2 = g^2(v)$. This is small provided that

$$v \gg \Lambda$$

In other words, we can trust our weakly coupled intuition when we are far out on the $N_f = N_c - 1$ moduli space, with $|M| \sim v^2 \gg \Lambda$. This means that, in this regime, we should be able to compute the Affeck-Dine-Seiberg superpotenial in some more traditional manner.

The form of the superpotential itself tells us where to look. When $N_f = N_c - 1$, (6.18) becomes

$$W_{\text{eff}} = C_\star \frac{\Lambda^{2N_c+1}}{\det M} \tag{6.25}$$

with $C_\star = C(N_c, N_c - 1)$. This is proportional to $\Lambda^{b_0} \sim e^{-8\pi^2/g^2 + i\vartheta}$, which, as we saw in (5.32), is the characteristic signature of an instanton .

This gives a window of opportunity. Until now, our results for the quantum dynamics have relied on symmetries and, crucially, holomorphy. Supersymmetry, of course, bought us the latter. But this approach can only get us so far and, as we have stressed, there is nothing to fix the overall constant C. In particular, we need to check that it doesn't vanish. This requires us to roll up our sleeves and do a weak coupling, instanton computation. And the theory with $N_f = N_c - 1$ is the place to do it. The calculation is rather technical and we won't describe it here[8]. But the result is

$$C_\star = 1$$

Decoupling: From Weak to Strong Coupling

The single coefficient $C_\star = 1$ for $N_f = N_c - 1$ is sufficient for us to derive the coefficient $C(N_c, N_f)$ for all other values of $N_f < N_c$. We do this by decoupling arguments.

Let's start with the theory with $N_f = N_c - 1$ flavours. We will give a large mass m to k of these flavours. We then expect to flow down to the theory with

$$N_f' = N_c - (k+1) \tag{6.26}$$

We want to derive the effective superpotential for this new theory.

[8]The instanton calculation was first done by Affleck, Dine and Seiberg who showed that $C_\star \neq 0$. The exact result $C_\star = 1$ was first derived by Finnell and Pouliot.

Our starting point is the superpotential (6.19) for $N_f = N_c - 1$

$$W = \frac{\Lambda_{\rm old}^{2N_c+1}}{\det M} + {\rm Tr}(mM) \tag{6.27}$$

where now the coefficient $C_\star = 1$ in front of the first term should be viewed as fixed by the weak-coupling instanton calculation. Note that we've added the subscript "old" to the strong coupling scale in anticipation of the fact that we will integrate out matter to flow to a new theory with N_f' flavours. We give a mass matrix of the form

$$m = m\begin{pmatrix} 0 & 0 \\ 0 & \mathbb{1}_k \end{pmatrix}$$

The critical point $\partial W/\partial M^i_j = 0$ solves, from (6.23),

$$mM = \frac{\Lambda_{\rm old}^{2N_c+1}}{\det M}\,\mathbb{1}_{N_f} \tag{6.28}$$

We should pause to understand what this is telling us. The meson matrix M takes the form

$$M = \begin{pmatrix} \tilde{M} & 0 \\ 0 & Z \end{pmatrix}$$

where Z is a $k \times k$ matrix and $\tilde{M}$ is a $(N_f - k) \times (N_f - k)$ matrix. Note that the off-diagonal terms in M must vanish by the equation of motion (6.28).

At first glance, it looks tricky to solve the matrix equation (6.28) because of all those zeroes in the upper left corner of m make it difficult for the left-hand side to be equal to the identity matrix $\mathbb{1}_{N_f}$. But the physics is actually clear. The massive k flavours in the matrix Z have an expectation value that's stabilised as $Z \sim 1/m$. Meanwhile, the remaining massless flavours in the matrix $\tilde{M}$ have a runaway behaviour $\tilde{M} \to \infty$ as we've seen before.

Here our interest is subtly different. We will integrate out the heavy degree of freedom Z. This means that we solve (6.28) only for Z and substitute it back in to get an effective action for $\tilde{M}$. This effective action will then tell us that $\tilde{M}$ suffers a runaway, which we knew anyway. But our goal is only to find the overall coefficient $C(N_c, N_f)$ in front of this runaway superpotential.

Focussing on the $k \times k$ part of (6.28) gives the matrix equation

$$mZ = \frac{\Lambda_{\text{old}}^{2N_c+1}}{\det \tilde{M} \det Z} \mathbb{1}_k$$

Taking traces and determinants gives

$$m\text{Tr}\, Z = \frac{k\Lambda_{\text{old}}^{2N_c+1}}{\det \tilde{M} \det Z} \quad \text{and} \quad (\det Z)^{k+1} = \left(\frac{\Lambda_{\text{old}}^{2N_c+1}}{m \det \tilde{M}}\right)^k$$

If we substitute this back into the original superpotential (6.27), then we get a superpotential purely for the $\tilde{M}$ mesons. It is

$$W = (k+1)\left(\frac{\Lambda_{\text{old}}^{2N_c+1} m^k}{\det \tilde{M}}\right)^{\frac{1}{k+1}}$$

From (6.26), we know that $k+1 = N_c - N_f'$. Meanwhile, the kind of RG matching arguments that led us to (6.21) reveal that the numerator is the strong coupling scale associated to $SU(N_c)$ with N_f' massless flavours

$$\Lambda_{\text{new}}^{3N_c - N_f'} = \Lambda_{\text{old}}^{2N_c+1} m^k$$

The upshot is that we reproduce the Affleck-Dine-Seiberg superpotential for the light meson fields as expected,

$$W = (N_c - N_f')\left(\frac{\Lambda_{\text{new}}^{3N_c - N_f'}}{\det \tilde{M}}\right)^{\frac{1}{N_c - N_f'}}$$

But with the added bonus that we've derived the long-promised coefficient $C(N_c, N_f) = N_c - N_f$.

6.3 A Second Look at SQCD

We've seen that the moduli space of vacua is lifted for $N_f < N_c$. Now we look at what happens for higher N_f.

Our first observation is that the superpotential (6.18)

$$W_{\text{eff}} = C\left(\frac{\Lambda^{3N_c - N_f}}{\det M}\right)^{\frac{1}{N_c - N_f}}$$

is the only one allowed by the symmetries, regardless of N_f. But it makes no sense for $N_f \geq N_c$. First, it clearly diverges when $N_f = N_c$. Moreover, for $N_f < N_c < 3N_c$ it has negative powers of Λ, which means that the superpotential scales as e^{+1/g^2} (with some coefficient). But this diverges as $g^2 \to 0$ and so isn't compatible with the weak coupling limit. In particular, we know that if we set $g^2 = 0$ then the theory is simply free and nothing can be going on. This rules out the possibility of a superpotential.

When $N_f < 3N_c$, the superpotential does have a positive power of Λ. But this corresponds to the situation where $b_0 < 0$ and the theory is infra-red free and no superpotential can be generated. (Another way of saying this is that the putative strong coupling scale scale Λ is actually bigger than the UV cut-off and shouldn't be trusted.) We'll look at this theory in more detail below.

All of this means that for $N_f \geq N_c$ there is no possible superpotential that can arise. The moduli space of vacua survives and, correspondingly, there are necessarily massless degrees of freedom. Our goal is to understand them.

We will start in this section by looking at two special cases: $N_f = N_c$ and $N_f = N_c{+}1$. Both exhibit interesting phenomena[9]. In later sections we'll then look at higher N_f.

6.3.1 A Deformed Moduli Space for $N_f = N_c$

Recall that for $N_f = N_c$, the moduli space is parameterised by mesons $M_j{}^i = \tilde{\Phi}_j \Phi^i$ and baryons

$$B = \phi^1_{a_1} \ldots \phi^{N_c}_{a_{N_c}} \epsilon^{a_1 \ldots a_{N_c}} \quad \text{and} \quad \tilde{B} = \tilde{\phi}^{a_1}_1 \ldots \tilde{\phi}^{a_{N_c}}_{N_c} \epsilon_{a_1 \ldots a_{N_c}}$$

These fields, gauge invariant composites, and parameters transform under the following symmetries:

	$SU(N_f)_L$	$SU(N_f)_R$	$U(1)_B$	$U(1)_A$	$U(1)_R$
Φ	$\square$	$\mathbf{1}$	1	1	0
$\tilde{\Phi}$	$\mathbf{1}$	$\overline{\square}$	-1	1	0
M	$\square$	$\overline{\square}$	0	2	0
B	$\mathbf{1}$	$\mathbf{1}$	N_c	N_c	0
$\tilde{B}$	$\mathbf{1}$	$\mathbf{1}$	$-N_c$	N_c	0
Λ^{2N_c}	$\mathbf{1}$	$\mathbf{1}$	0	$2N_c$	0

The classical moduli space is defined as an algebraic variety, with a single constraint (4.42) between the fields

$$\det M \; - \tilde{B}B = 0 \tag{6.29}$$

[9]The original paper is by Nati Seiberg, "Exact Results on the Space of Vacua of Four Dimensional SUSY Gauge Theories".

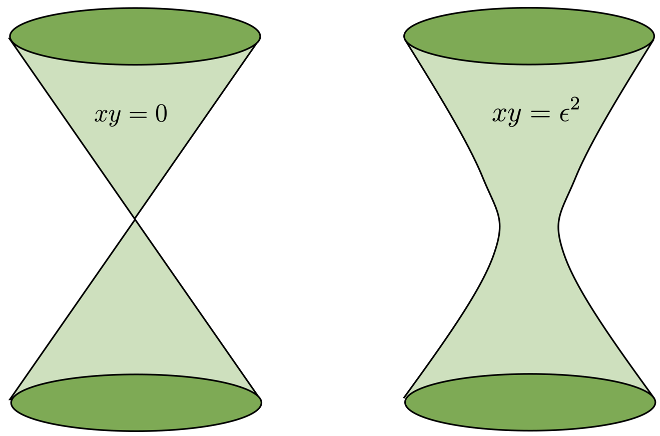

Figure 11. The singular space $xy = 0$ on the left and the smooth space $xy = \epsilon^2$ on the right. This is a cartoon for the moduli space of SQCD when $N_f = N_c$. On the left the classical, singular modular space; on the right, the smooth quantum moduli space.

We know that this can't be lifted by a superpotential. But it turns out that the space is deformed. The quantum moduli space satisfies the constraint

$$\det M \;\; - \tilde{B}B = \Lambda^{2N_c} \tag{6.30}$$

There are a number of questions that spring to mind. First, what is the meaning of this deformation? And second, how do we know that it happens?

Let's start by answering the first of these. The mathematics is all about of the singularities of the space, the physics all about their meaning. We can start by looking at a much simpler example. Consider the algebraic variety defined by

$$xy = 0$$

with $x, y \in \mathbb{C}$. This is obviously the intersection of two complex lines. (The complex line, or often just "line" is the name given by algebraic geometers to what you used to think of as a plane.) The space is obviously singular at the origin $x = y = 0$. The way to see this mathematically is to look a the tangent vectors, δx and δy. These obey

$$\delta x\, y + x\delta y = 0 \tag{6.31}$$

For any point other than the origin, there is a unique complex tangent vector. For example, if $x \neq 0$ then the tangent vector is δx since we necessarily have $\delta y = 0$. But at the origin there is no constraint on δx and δy which is telling us that tangent vector is ill-defined and, correspondingly, the space is singular.

We can compare this to the deformed variety

$$xy = \epsilon^2$$

Again, this is a space with one complex dimension and, far from the origin, looks much like $xy = 0$. But the origin $x = y = 0$ is no longer part of this space and this means that the singularity has now been removed. Tangent vectors must still obey (6.31) but now there is a unique tangent vector for each point obeying $xy = \epsilon^2$. The singular and deformed spaces are shown in Figure 11.

This simple example captures the key features of the moduli space $\mathcal{M}$. The classical moduli space (6.29) is singular. This is obviously true at the origin $M = \tilde{B} = B = 0$, but more generally it is singular on any submanifold where $\tilde{B} = B = 0$ and the meson matrix has $\text{rank}(M) \leq N_c - 2$. In contrast, the quantum moduli space (6.30) is smooth. All singularities have been removed. What is this telling us?

As we've seen in numerous examples in Section 4.3, singularities in the moduli space signify the existence of new massless degrees of freedom. In the present case, there is no mystery to this: the new massless degrees of freedom are gauge bosons. In particular, when $\text{rank}(M) = k \leq N_c - 2$, an $SU(k)$ gauge group is unbroken.

But these singularities are removed in the quantum theory. This tells us that the additional particles at the origin of moduli space that were classically massless have now gained a mass. This is the famous mass gap problem! Here we see that the a complicated quantum effect – namely the fact that gauge bosons get a mass through strong coupling – arises in a surprising geometric manner.

Now for the second question: how do we know that the quantum deformation of the moduli space takes place? The first thing to note is that it's consistent with the symmetries and, as we've noted before, anything that isn't prohibited typically occurs. Of course, you might be forgiven for not being aware that deforming the constraint through quantum effects was even something that could happen, but the discussion above about the meaning of removing singularities will hopefully serve to allay such doubts. However, we should strive to find more convincing evidence than this. And, indeed, there are two very compelling reasons to believe that the deformation happens.

6.3.2 't Hooft Anomaly Matching

Our picture of physics described by the quantum modified constraint assumes that the only massless degrees of freedom are the mesons and baryons. There are a number of interesting constraints that this picture must satisfy. These come from 't Hooft anomalies.

The original global symmetry of the theory is

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

The 't Hooft anomalies must be matched at each point on the quantum moduli space. At different points, the global symmetry is broken to some subgroup, $G_F \to H_F$ and this surviving subgroup changes as we move around $\mathcal{M}$. But importantly, the point $M = B = \tilde{B} = 0$ where the full global symmetry G_F would be completely unbroken has been removed by the quantum deformation (6.30). There are, however, two points where the surviving symmetry H_F is maximal and anomaly matching is most stringent. These are

- $B = \tilde{B} = 0$ with $M = \Lambda^2 \mathbb{1}_{N_c}$. At this point, the surviving global symmetry group is

$$H_F = SU(N_f)_{\text{diag}} \times U(1)_B \times U(1)_R \tag{6.32}$$

  This is not dissimilar to the chiral symmetry breaking pattern in non-supersymmetric QCD

- $M = 0$ with $\tilde{B} = B = \Lambda^{N_c}$. At this point, the surviving global symmetry group is

$$H_F = SU(N_f) \times SU(N_f) \times U(1)_R \tag{6.33}$$

  This is a symmetry breaking pattern that doesn't (we think!) occur in non-supersymmetric QCD. The non-Abelian chiral symmetry is unbroken but, in contrast, baryon number is broken.

We do anomaly matching at each of these points in turn. For what follows, we will need to frequently turn to the table of symmetries that we constructed at the beginning of this subsection.

The Point with $\tilde{B} = B = 0$

We need to match anomalies for symmetries, and any mixed anomalies between symmetries, for H_F given in (6.32). We'll do each in turn, starting with the non-Abelian $SU(N_f)_{\text{diag}}$ symmetry.

$\underline{SU(N_f)^3_{\text{diag}}}$: In the UV, we have the quarks ψ and $\tilde{\psi}$. But these cancel in their contribution to the anomaly, giving $\mathcal{A}_{UV} = 0$. In the infra-red, only the meson carries non-Abelian charge. Under the diagonal $SU(N_f)_{\text{diag}}$ it transforms in $\Box \otimes \overline{\Box} = \text{adj} \oplus \mathbf{1}$.

But the adjoint is a real representation and doesn't contribute to the anomaly, so we have $\mathcal{A}_{IR} = 0$.

$\underline{SU(N_f)^2_{\rm diag} \cdot U(1)_B}$: In the UV, the quarks ψ and $\tilde{\psi}$ carry opposite $U(1)_B$ charge and so cancel in their contribution, giving $\mathcal{A}_{UV} = 0$. In the IR, the mesonic fermions are uncharged under $U(1)_B$ so also give $\mathcal{A}_{IR} = 0$.

$\underline{SU(N_f)^2_{\rm diag} \cdot U(1)_R}$: This is more interesting. We need to remember that the charges listed in the table are for bosons in the chiral multiplet, with $R[\text{fermion}] = R[\text{boson}]-1$. In the UV, we have

$$\mathcal{A}_{UV} = N_c \times I(\Box) \times (-1) + N_c \times I(\overline{\Box}) \times (-1) = -2N_c$$

where the factors of N_c are because each quark has N_c colours. Meanwhile, in the IR, the contribution from the fermionic mesons is

$$\mathcal{A}_{IR} = I(\text{adj}) \times (-1) = -2N_f$$

Now there is no contribution from colour degrees of freedom because the mesons are confined. Instead there is only the $SU(N_f)_{\rm diag}$ group theory factor $I(\text{adj})$. Nonetheless, we have $\mathcal{A}_{UV} = \mathcal{A}_{IR}$ because we are working in the theory with $N_f = N_c$.

$\underline{U(1)^2_B \cdot U(1)_R}$: In the UV, the quarks contribute

$$\mathcal{A}_{UV} = N_c N_f \times (+1)^2 \times (-1) + N_c N_f \times (-1)^2 \times (-1) = -2N_c N_f$$

In the IR, only the fermionic baryons contribute. These give

$$\mathcal{A}_{IR} = (N_c)^2 \times (-1) + (-N_c)^2 \times (-1) = -2N_c^2$$

Again, $\mathcal{A}_{UV} = \mathcal{A}_{IR}$.

$\underline{U(1)^3_R}$: This time we have to remember that there are $N_c^2 - 1$ gluinos with charge $R[\lambda] = +1$ in the UV. These didn't contribute to any of the anomalies above, but they do now. Including both gluinos and quarks, we have

$$\mathcal{A}_{UV} = (N_c^2 - 1) \times (+1)^3 + N_c N_f \times (-1)^3 + N_c N_f \times (-1)^3 = N_c^2 - 2N_f N_c - 1$$

In the IR, both mesons and baryons contribute to the anomaly, all with R-charge -1. This is the first time that all the IR fields contributed and this means that it's the first time we need to take into account the constraint (6.30). This is a constraint not just on the expectation values, but also on the fluctuations of the fields. This means that the number of massless IR fields is $\dim \mathcal{M} = N_f^2 + 2 - 1$ with the $+2$ the baryons B and $\tilde{B}$ and the -1 coming from the constraint. The upshot is that the IR anomaly is

$$\mathcal{A}_{IR} = \dim \mathcal{M} \times (-1)^3 = -N_f^2 - 1$$

Again, we see the anomaly matches with the UV.

There are two remaining anomalies, $U(1)_B^3$ and $U(1)_R^2 \cdot U(1)_B$. You can check that both have $\mathcal{A}_{UV} = \mathcal{A}_{IR} = 0$ because $U(1)_B$ is vector-like.

In addition, we can match mixed $U(1)$-gravitational anomalies. This simply means that the sum of $U(1)$ charges must be the same in the UV and IR. However, in the present case these don't really give anything new. For $U(1)_B$, we have $\sum q_B = 0$ in both UV and IR. For $U(1)_R$ all charges are $q_R = \pm 1$ so $\sum q_R = \sum q_R^3$ and this reduces the $U(1)_R^3$ calculation that we did above. When we consider other theories the matching of mixed gauge-gravitational anomalies will give more compelling results.

The Point with $M = 0$

We now need to match anomalies for H_F given in (6.33). The only real difference from the calculation above lies in the $SU(N_f)_L^3$ anomaly. In the UV. In the UV, just the quarks ψ contribute and give

$$\mathcal{A}_{UV} = N_c \times A(\Box) = N_c$$

In the IR, the N_f^2 mesons contribute. We have

$$\mathcal{A}_{IR} = N_f \times A(\Box) = N_f$$

Again, $\mathcal{A}_{UV} = \mathcal{A}_{IR}$ because we're working in the theory with $N_f = N_c$. The anomaly matching for $SU(N_f)_L^2 \cdot U(1)_R$ works in much the same way, giving $\mathcal{A}_{UV} = \mathcal{A}_{IR} = -N_c$. The anomaly matching for $U(1)_R^3$ works in the same way as we saw above.

The calculations of anomaly matching are straightforward. But the agreement is not entirely trivial. In particular, it's clear that it works only when $N_f = N_c$. As we proceed, we'll see anomaly matching working in more intricate ways.

Decoupling

There is a second way to see the need for the quantum deformation of the moduli space. This uses a trick that we've seen before: we look at the fate of the theory when we give one flavour a mass and decouple it.

It's not immediately obvious how to do this since, as we saw above, we don't have a superpotential to start with! The trick is to view the constraint (6.30) itself as a superpotential

$$W = X\left(\det M \;\; - \tilde{B}B - \Lambda^{2N_c}\right)$$

where we've introduced a new chiral superfield X whose sole role is to act as a Lagrange multiplier, imposing the constraint. We now add a mass for just one flavour. The superpotential is

$$W = X\left(\det M \;\; - \tilde{B}B - \Lambda_{\rm old}^{2N_c}\right) + {\rm Tr}(mM) \tag{6.34}$$

We've added the superscript "old" appears because we're playing an integrating out game. We're going to look at what happens when $|m| \gg |\Lambda_{\rm old}|$ so that we have one massive flavour and $N_f = N_c - 1$ massless flavours. In this case, we should be able to re-derive the appropriate Affleck-Dine-Seiberg superpotential. Let's see how it works.

The rest of the calculation is very similar to the decoupling that we saw in previous sections. The critical point for the mesons sits at $\partial W/\partial M_j{}^i = 0$, or

$$mM = -X \det M\, \mathbb{1}_{N_f} \tag{6.35}$$

If we turn on a mass term for just the final $N_f^{\rm th}$ flavour, with $m = {\rm diag}(0,\ldots,0,m)$. The meson fields take the form

$$M = \begin{pmatrix} \tilde{M} & 0 \\ 0 & Z \end{pmatrix}$$

with $Z = M_{N_f}{}^{N_f}$ the final flavour and the off-diagonal terms set to zero at the critical point (6.35). The equation arising from $\partial W/\partial Z$ in (6.35) tells us that

$$X = -\frac{m}{\det \tilde{M}}$$

Meanwhile, the critical points for B and $\tilde{B}$ are

$$\frac{\partial W}{\partial B} = -X\tilde{B} = 0 \quad \text{ and } \quad \frac{\partial W}{\partial \tilde{B}} = -BX = 0$$

which, since $X \neq 0$, means that we must have $\tilde{B} = B = 0$. So far Z is undetermined, but this is fixed by the equation of motion for X which, of course, is simply the constraint itself. It now reads

$$Z \det \tilde{M} = \Lambda_{\rm old}^{2N_c}$$

We now substitute this back into the superpotential (6.34). Only the final $\mathrm{Tr}(mM) = mZ$ term contributes and gives

$$W = \frac{\Lambda_{\rm old}^{2N_c} m}{\det \tilde{M}} = \frac{\Lambda_{\rm new}^{2N_c+1}}{\det \tilde{M}}$$

with the now familiar RG matching giving $\Lambda_{\rm new}^{2N_c+1} = \Lambda_{\rm old}^{2N_c} m$. This we recognise as the Affleck-Dine-Seiberg superpotential (6.25) in the case $N_f = N_c - 1$ (with even the coefficient correct). Notice that the quantum deformation of the constraint was necessary for us to reproduce the known physics when we integrate out massive flavours. This is our first piece of evidence (beyond the symmetries) that the deformation actually occurs.

6.3.3 Confinement Without χSB for $N_f = N_c + 1$

The case of $N_f = N_c + 1$ also exhibits some rather startling behaviour and is worth exploring in some detail. Recall from Section 4.3 that, in addition to the mesons $M_j{}^i$, we now have N_f baryons of each type

$$B_j = \epsilon_{j i_1 \ldots i_{N_c}} B^{i_1 \ldots i_{N_c}} \quad \text{and} \quad \tilde{B}^j = \epsilon^{j i_1 \ldots i_{N_c}} \tilde{B}_{i_1 \ldots i_{N_c}}$$

This satisfy the constraints (4.43)

$$\mathrm{Adj}(M)^i{}_j = B^i \tilde{B}_j \quad \text{and} \quad M_j{}^i B^j = M_j{}^i \tilde{B}_i = 0 \qquad (6.36)$$

Recall that if the adjugate matrix $\mathrm{Adj}(M)$ is invertible then it is given by $\mathrm{Adj}(M) = (\det M) M^{-1}$. We can gather the various gauge fields together to list their symmetries in a now-familiar table

	$SU(N_f)_L$	$SU(N_f)_R$	$U(1)_B$	$U(1)_A$	$U(1)_R$
Φ	$\square$	$\mathbf{1}$	1	1	$\frac{1}{N_f}$
$\tilde{\Phi}$	$\mathbf{1}$	$\overline{\square}$	-1	1	$\frac{1}{N_f}$
M	$\square$	$\overline{\square}$	0	2	$\frac{2}{N_f}$
B	$\overline{\square}$	$\mathbf{1}$	N_c	N_c	$\frac{N_c}{N_f}$
$\tilde{B}$	$\mathbf{1}$	$\square$	$-N_c$	N_c	$\frac{N_c}{N_f}$
Λ^{2N_c-1}	$\mathbf{1}$	$\mathbf{1}$	0	$2N_f$	0

As we've already seen, there can be no superpotential generated on the moduli space. But, this time, there can be no quantum deformation of the constraints either! There is no possibility consistent with the symmetries and various weakly coupled limits. Our quantum moduli space has singularities.

What are we to make of this? As we've seen in several examples, the singularities signify new massless degrees of freedom. Classically, these degrees of freedom are gauge bosons. It's tempting to conclude that the singularities in the quantum theory are telling us that the gauge bosons are free at the origin of the moduli space. However, it turns out that this is not the case. Instead, the quantum interpretation of the singularities is rather different.

In fact an obvious quantum interpretation suggests itself if we assume that the theory confines. This means that the low-energy fields are necessarily mesons and baryons which, in general, are constrained by (6.36). Geometrically, the singularities of $\mathcal{M}$ arise when the fluctuations of M, B and $\tilde{B}$ are no longer restricted to lie on $\mathcal{M}$. Physically, this translates into the suggestion that the singularities of $\mathcal{M}$ might be due to *unconstrained* mesons and baryons. In particular, it would suggest that at the origin of moduli space $M = B = \tilde{B}$, we should think of the physics as described by free, massless mesons and baryons.

This interpretation of the singularity is rather remarkable, not least because we would have confinement *without* the accompanying chiral symmetry breaking. At the origin of moduli space, the full chiral symmetry

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

is unbroken. Famously, confinement without chiral symmetry breaking is *not* possible in QCD. (We sketched the argument in Section 5.2.3.) The suggestion is that this does happen in SQCD with $N_f = N_c + 1$.

The phenomenon of confinement without chiral symmetry breaking in SQCD sometimes goes by the name of *s-confinement*. It's a rubbish name. Here "s" can stand for "smooth" or perhaps "screening" depending on taste.

More 't Hooft Anomaly Matching

There is a fairly stringent test that any proposal for confinement without chiral symmetry breaking must pass. This is 't Hooft anomaly matching. Let's see how we do.

$\underline{SU(N_f)_L^3}$: In the UV, we have the quarks contributing to give $\mathcal{A}_{UV} = N_c$. In the IR, we have both mesons M, which contribute N_f and the baryons B which contribute -1 as they sit $\overline{\Box}$. Together they give $\mathcal{A}_{IR} = N_f - 1 = N_c$.

$\underline{SU(N_f)_L^2 \cdot U(1)_B}$: The quarks give $\mathcal{A}_{UV} = N_c$. In the infra-red, the mesons don't contribute while the baryon B gives $\mathcal{A}_{IR} = N_c$.

$\underline{SU(N_f)_L^2 \cdot U(1)_R}$: Now things get more fiddly, largely because of the fractional R-charges. In the UV, the quarks give

$$\mathcal{A}_{UV} = N_c \left(\frac{1}{N_f} - 1 \right) = -\frac{N_c^2}{N_c + 1}$$

In the IR, both the meson and baryon contribute:

$$\mathcal{A}_{IR} = N_f \left(\frac{2}{N_f} - 1 \right) + \left(\frac{N_c}{N_f} - 1 \right)$$

A little algebra reassuringly shows that $\mathcal{A}_{UV} = \mathcal{A}_{IR}$.

The remaining anomaly matching involving $U(1)_R$ gets a little messy. For example, we have

$\underline{U(1)_R}$: The mixed $U(1)_R$ gravitational anomaly simply requires that we add up the R-charges. Including the gluinos, we have

$$\mathcal{A}_{UV} = (N_c^2 - 1) + 2N_c N_f \left(\frac{1}{N_f} - 1 \right) = -N_f^2 + 2N_f - 2$$

Meanwhile,

$$\mathcal{A}_{IR} = N_f^2 \left(\frac{2}{N_f} - 1 \right) + 2N_f \left(\frac{N_c}{N_f} - 1 \right) = \mathcal{A}_{IR}$$

$\underline{U(1)_R^3}$: The calculation is the same as above, but with R^3 instead of R. We have

$$\mathcal{A}_{UV} = (N_c^2 - 1) + 2N_c N_f \left(\frac{1}{N_f} - 1 \right)^3 = -\frac{N_f^4 - 6N_f^3 + 12N_f^2 - 8N_f + 2}{N_f^2}$$

Meanwhile,

$$\mathcal{A}_{IR} = N_f^2 \left(\frac{2}{N_f} - 1 \right)^3 + 2N_f \left(\frac{N_c}{N_f} - 1 \right)^3$$

Again, we find $\mathcal{A}_{UV} = \mathcal{A}_{IR}$.

By now, you won't be surprised to hear that all other 't Hooft anomalies also match. The messier the computation, the more compelling the evidence. It certainly feels like there is something deep going on when these complicated algebraic expressions are found to agree.

Decoupling

For $N_f < N_c$, we built up an impressive pattern of consistency, understanding how our new results can be used to imply our earlier ones. We can do this again here. But there's a curious lesson awaiting us.

You might think that we should impose the constraints (6.36) by introducing a bunch of Lagrange multipliers. This, it turns out, doesn't work. Instead the constraints arise in a slightly different way. To see this, note that the symmetries allow us to introduce the superpotential

$$W = -\frac{1}{\Lambda^{2N_c-1}}\left(\det M - BM\tilde{B}\right) \tag{6.37}$$

Using Jacobi's formula (6.22), equations of motion from this superpotential are (ignoring the overall factor of Λ^{2N_c-1} for now)

$$\frac{\partial W}{\partial B} = M\tilde{B} = 0 \ , \quad \frac{\partial W}{\partial \tilde{B}} = BM = 0 \ , \quad \frac{\partial W}{\partial M_j{}^i} = -\mathrm{Adj}(M)^i{}_j + B^i\tilde{B}_j = 0$$

The upshot is that the superpotential (6.37) gives the constraints (6.36) as the equations of motion, rather than through a Lagrange multiplier. This, it turns out, is the way the constraints should be imposed when $N_f = N_c + 1$.

This is a much softer way to implement constraints. A Lagrange multiplier imposes a constraint absolutely in the path integral. In contrast, the classical equations of motion are merely a gentle suggestion that, at weak coupling, certain configurations carry more weight in the path integral. Presumably this is related to the fact that the unconstrained mesons and baryons manifest themselves at the origin.

There is one further unusual aspect of (6.37) and that's the negative power of Λ. In previous sections, we discarded some possible superpotentials on the grounds that they scale as e^{+1/g^2} (with some appropriate exponent) and so didn't reproduce our weak coupling needs. But in this case the constraints are classical constraints and the classical limit $g \to 0$ simply imposes them more strenuously. So there's nothing to be concerned about.

We know the deal by now. We introduce a mass for the last flavour, so the superpotential reads

$$W = -\frac{1}{\Lambda_{\text{old}}^{2N_c-1}}\left(\det M - BM\tilde{B}\right) + \mathrm{Tr}(mM)$$

with $m = \text{diag}(0, \ldots, 0, m)$. The critical point of the meson now sits at

$$\det M - BM\tilde{B} = \Lambda_{\text{old}}^{2N_c-1} mM \tag{6.38}$$

The meson and baryon fields can be shown to take the form,

$$M = \begin{pmatrix} \tilde{M} & 0 \\ 0 & Z \end{pmatrix} \quad , \quad B^i = \begin{pmatrix} 0 \\ B \end{pmatrix} \quad , \quad \tilde{B}_j = \begin{pmatrix} 0 \\ \tilde{B} \end{pmatrix}$$

with $Z = M_{N_f}^{N_f}$ the final flavour. The constraints $BM = M\tilde{B} = 0$ tell us that $Z = 0$ if $B, \tilde{B} \neq 0$. But we should still impose the equation of motion. And, indeed, Z drops out of the equation (6.38) which becomes

$$\det \tilde{M} - \tilde{B}B = m\Lambda_{\text{old}}^{2N_c-1} = \Lambda_{\text{new}}^{2N_c}$$

This, of course, is the quantum modified constraint (6.30) of the theory with $N_f = N_c$.

6.4 A Peek in the Conformal Window

At this point, we will jump to the other end of the flavour spectrum. We know that SQCD is no longer asymptotically free when $N_f \geq 3N_c$. In this situation, the low-energy physics is easy: it is just weakly interacting gluons, gluinos and massless (s)quarks.

What if we now lower N_f slightly below the asymptotic freedom bound. Here, too, the physics is well understood. This is for the same reason that we saw in non-supersymmetric QCD: there is a zero of the beta function at weak coupling where we trust the calculation. This is the Banks-Zaks fixed point. The argument holds for SQCD just as it does for normal QCD.

Now let's lower N_f still further. The expectation is that we will continue to flow to an interacting conformal field theory for some range of N_f, presumably with a different CFT for each N_c and N_f. The question is: how low can N_f go?

We don't know the answer in the non-supersymmetric case. But it turns out, we do know the answer for SQCD. We flow to an interacting conformal field theory in the regime

$$\frac{3N_c}{2} < N_f < 3N_c \tag{6.39}$$

This is the *conformal window*.

Obviously we should ask how we know the lower bound of the conformal window. This, it turns out, follows from certain properties of supersymmetric conformal field theories. In the rest of this section we will state these properties, although we won't derive them. Then, in Section 6.5, we'll turn to the outstanding question of what happens in the gap between $N_f = N_c + 1$ and the conformal window at $N_f > 3N_c/2$.

6.4.1 Facts About Conformal Field Theories

A conformal field theory (or CFT) describes the dynamics of interacting massless particles. Its defining feature is that it is invariant under scale transformations, also known as *dilatations*,

$$x^\mu \to \lambda x^\mu$$

Such a scaling would be broken by any dimensionful parameter, such as a mass, which is one way of seeing that conformal field theories can only describe massless excitations.

Any relativistic, scale invariant theory appears to also enjoy a more dramatic additional symmetry known as *special conformal transformations*. This acts as

$$x^\mu \to \frac{x^\mu - a^\mu x^2}{1 - 2a\cdot x + a^2 x^2}$$

In $d = 1 + 1$ dimensions, there is a proof that scale invariance implies conformal invariance. In higher dimensions, the proofs are not complete but, nonetheless, it is thought to be true in any interacting conformal field theory.

The generators of dilatations D and of special conformal transformations K_μ take the form

$$D = -ix^\mu \partial_\mu \quad , \quad K_\mu = -i(2x_\mu x^\nu \partial_\nu - x^2 \partial_\mu)$$

They combine with the usual generators of the Poincaré algebra to form the conformal algebra, which has the additional commutation relations

$$\begin{gathered}
[D, K^\mu] = -iK_\mu \quad , \quad [D, P^\mu] = iP^\mu \\
[K^\mu, P^\nu] = 2i(D\eta^{\mu\nu} - M^{\mu\nu}) \\
[M^{\mu\nu}, K^\sigma] = i\left(K^\nu \eta^{\mu\sigma} - K^\mu \eta^{\nu\sigma}\right)
\end{gathered}$$

The kinds of questions that we want to ask about conformal field theories are somewhat different from what we're used to. We no longer care about the masses of particles because they're all zero. Nor do we usually care about the S-matrix which is challenging to define in a theory of massless particles where there can be arbitrarily low energy excitations of increasingly long wavelengths.

Instead, in a CFT we care about correlation functions. In particular, we care about *scaling dimensions*. This means that we want to find operators $\mathcal{O}(x)$ that have the nice property

$$\mathcal{O}(\lambda x) = \lambda^{-\Delta}\mathcal{O}(x)$$

with Δ the scaling dimension. If we then look at the two-point function of these operators, we necessarily have

$$\langle\, \mathcal{O}^\dagger(x)\mathcal{O}(0)\,\rangle \sim \frac{1}{|x|^{2\Delta}}$$

These scaling dimensions are closely related to the critical exponents that were the focus in the lectures on Statistical Field Theory.

It's useful to look to a free, massless scalar field as an example of a trivial CFT. Here the theory is described by the action

$$S = \int d^dx\ \frac{1}{2}\partial_\mu\phi\partial^\mu\phi$$

The scaling dimension of ϕ coincides with what we often call the "engineering dimension", or sometimes just "dimension". It is

$$\Delta[\phi] = \frac{d-2}{2}$$

We don't have Lagrangian descriptions for interacting CFTs. The closest we can get is to write down the Lagrangian for a field theory in the UV that flows, in the IR, to an interacting CFT. This, for example, is what happens in massless (S)QCD with a suitable number of flavours. It may be that the resulting CFT is weakly coupled, such as for a Banks-Zaks fixed point, in which case we can compute the scaling dimensions Δ perturbatively. Or it may be that resulting CFT is strongly coupled, in which case we need to turn to some other method. Other methods on the table include numerics, the ϵ expansion that we met in Statistical Field Theory, an approach known as the bootstrap and, as we will see, supersymmetry.

There is one important result that we will need. The interactions always serve to increase the scaling dimension. Or, said more precisely, the dimension of any scalar operator in a unitary, interacting CFT is bounded by

$$\Delta[\mathcal{O}] \geq \frac{d-2}{2}$$

This is known as the *unitarity bound*[10]. In the language of perturbative quantum field theory, this is telling us that the anomalous dimensions of operators are always positive.

[10]It is not too difficult to derive this bound. They key step is to quantise the theory on $\mathbf{S}^3\times\mathbb{R}$ where we get to use the so-called *state-operator map* that relates local operators to states in the Hilbert space.

In addition, any operator that saturates the bound corresponds to a free field. This means that it must decouple from everything else that's going on in the theory.

Conformal field theories are of interest in many dimensions d. But our interests lie strictly in $d = 3 + 1$. The unitarity bound reads

$$\Delta[\mathcal{O}] \geq 1 \tag{6.40}$$

Any operator with $\Delta[\mathcal{O}] = 1$ is free.

Perturbing Conformal Field Theories

Suppose that you sit at a conformal fixed point. As we mentioned above, typically there's no action that can describe these dynamics directly but, for the sake of discussion, it will be useful to pretend. So lets call it S_{CFT}. (If you're worried about this, it's better to think in terms of a partition function in the presence of sources.)

Now we perturb the CFT. We do this by adding an extra term to the action. This extra term is some operator $\mathcal{O}(x)$ which, if you're in the setting of Lagrangian field theory, would be some combination of fields. The new action is

$$S = S_{\text{CFT}} + \lambda \int d^dx \; \mathcal{O}(x)$$

with λ the coefficient that governs the perturbation. The question is: what happens next?

The answer to this depends on the dimension $\Delta[\mathcal{O}]$. Roughly speaking, there are three possibilities

- $\Delta < d$: Such perturbations are called *relevant*. They change the dynamics in the infra-red and should be thought of as initiating an RG flow from our original CFT to somewhere else. An example is a mass term for a free, massless scalar field. In this case, the end point is a gapped theory. However, it's not true that a relevant deformation always pushes us to a gapped phase. We may, instead, flow to a different CFT.

- $\Delta > d$: These perturbations are *irrelevant*. They don't change the low-energy dynamics of the CFT. An example is a ϕ^6 interaction in $d = 3+1$ dimensions: it is important at high energies, but is insignificant at low energies.

Then you simply require the positivity of an arbitrary state $|P_\mu P^\mu|\phi\rangle|^2 > 0$ and the unitary bound follows after a few commutation relations using the conformal algebra. What is more challenging is to show that there is not a more stringent bound coming from some other requirement. You can find details in the excellent Lectures on Conformal Field Theory by Joshua Qualls.

- $\Delta = d$: These perturbations are called *marginal*. This arises when the parameter λ is dimensionless.

  Now things are a little more subtle. Typically, once you deform the theory by an arbitrarily small, marginal perturbation then the dimension of λ changes and runs under RG. It may become smaller as you flow to the IR and such perturbations are said to be *marginally irrelevant*. This happens, for example, for a ϕ^4 deformation or Yukawa terms in $d = 3+1$. Alternatively, the perturbation may grow stronger as you flow towards the IR as is the case for the coupling constant of Yang-Mills. Such perturbations are said to be *marginally irrelevant*.

  Alternatively, it may be that λ doesn't run at all under RG. In this case it is said to be *exactly marginal* and it means that we have a line of different conformal field theories, parameterised by λ. This situation is rare, but does occur for certain supersymmetric conformal field theories.

6.4.2 Facts About Superconformal Field Theories

When a theory with $\mathcal{N} = 1$ supersymmetry flows to an interacting conformal fixed point, it gives rise to a superconformal field theory (or SCFT). In addition to the supercharges Q_α and $\bar{Q}_{\dot{\alpha}}$ there are now superconformal charges S_α and $\bar{S}_{\dot{\alpha}}$.

Importantly, SCFTs necessarily have a $U(1)_R$ symmetry. Recall that this was somewhat optional in ordinary quantum field theories. For example, $U(1)_R$ is anomalous in super Yang-Mills and this is reflected in the transformation of the strong coupling scale Λ. But in an SCFT $U(1)_R$ is not an option. These theories always have an R-symmetry.

The $\mathcal{N} = 1$ superconformal algebra augments the conformal algebra with the Grassmann generators. There are commutators

$$\begin{aligned}
[D, Q_\alpha] &= \frac{1}{2} Q_\alpha \quad , \quad [D, S_\alpha] = -\frac{1}{2} S_\alpha \\
[R, Q_\alpha] &= Q_\alpha \quad , \quad [R, S_\alpha] = -S_\alpha \\
[K^\mu, Q_\alpha] &= i\sigma^\mu_{\alpha\dot{\alpha}} \bar{S}^{\dot{\alpha}} \quad , \quad [P^\mu, S_\alpha] = i\sigma^\mu_{\alpha\dot{\alpha}} \bar{Q}^{\dot{\alpha}}
\end{aligned}$$

and anti-commutators

$$\begin{aligned}
\{Q_\alpha, \bar{Q}_{\dot{\alpha}}\} &= \sigma^\mu_{\alpha\dot{\alpha}} P^\mu \quad , \quad \{S_\alpha, \bar{S}_{\dot{\alpha}}\} = 2\sigma^\mu_{\alpha\dot{\alpha}} K^\mu \\
\{Q_\alpha, S_\beta\} &= M_{\mu\nu} \sigma^\mu_{\alpha\dot{\alpha}} (\sigma^\nu)^{\dot{\alpha}}{}_\beta - i\left(D - \tfrac{3}{2}R\right)\epsilon_{\alpha\beta}
\end{aligned}$$

Now there is a slight twist to the unitarity bound. The fact that the R-symmetry and dilatation operator sit within the same algebra means that there is a rather remarkable relation between them. It can be shown that the dimension of any operator is bounded by its R-charge

$$\Delta[\mathcal{O}] \geq \frac{3}{2}\Big|R[\mathcal{O}]\Big|$$

Furthermore, chiral operators necessarily saturate this bound. Any chiral superfield Φ has

$$\Delta[\Phi] = \frac{3}{2}R[\Phi] \tag{6.41}$$

while any anti-chiral superfield $\bar{\Phi}$ has

$$\Delta[\bar{\Phi}] = -\frac{3}{2}R[\bar{\Phi}]$$

This is an extraordinarily powerful result. Usually in conformal field theories (at least in dimension $d > 2$) the scaling dimensions are extremely difficult to compute. And this remains true for most operators in a superconformal field theory. But there are a special class of operators – those described by chiral superfields – where the scaling dimension is trivial to compute. We just need to know its R-charge.

There is a way to get a feel for the factor of 3/2 in (6.41). Consider the Wess-Zumino model with $W(\Phi) = \lambda\Phi^3$, which leads to a $V(\phi) \sim |\phi|^4$ potential. This potential is classically marginal but one can show that it is marginally irrelevant at one-loop. This is the statement that $\lambda \to 0$ in the infra-red, so that the theory becomes free at low energies. Nonetheless, the classical potential fixes the R-charge to be $R[\Phi] = 2/3$ so that $R[W] = 2$ as it should. Correspondingly, $\Delta[\Phi] = 1$ in the infra-red which is indeed the right result for a free chiral multiplet.

The powerful result (6.41) also makes life easier in another way. If we have two chiral superfields Φ_1 and Φ_2 then $\Phi_1\Phi_2$ is also a chiral superfield. Their R-charges simply add: $R[\Phi_1\Phi_2] = R[\Phi_1] + R[\Phi_2]$. But so too do their dimensions: $\Delta[\Phi_1\Phi_2] = \Delta[\Phi_1] + \Delta[\Phi_2]$. This is unusual in a conformal field theory. Typically if you multiply operators together then you get divergences as their positions come close and regulating these divergences changes the dimension of the composite. But for chiral superfields, things are much easier. We say that the chiral operators form the *chiral ring*.

There is, however, a small fly in the ointment. You've got to be able to identify the correct R-symmetry that appears in the superconformal algebra. For example, suppose that your theory has an R-symmetry R and a global symmetry F. Then there's nothing to stop us from saying that $R + \alpha F$ is also a valid R-symmetry for any $\alpha \in \mathbb{R}$. How do we know that this isn't the thing that we should use when computing dimensions?! This loophole threatens to make the wondrous relation (6.41) completely toothless.

Happily, there is a procedure for figuring out what combination of symmetries forms the correct R-symmetry. This procedure is known as *a-maximization*. This is important for understanding many theories and we will describe the procedure in Section 7.2.4. However, as we'll now see, it is not needed for SQCD.

6.4.3 The Conformal Window for SQCD

We determined the symmetries of SQCD back in Section 6.2. The charges of the chiral superfields under the non-anomalous R-symmetry are

$$R[\Phi] = R[\tilde{\Phi}] = \frac{N_f - N_c}{N_f}$$

This means the R-charge of the meson $M = \tilde{\Phi}\Phi$ is

$$R[M] = \frac{2(N_f - N_c)}{N_f} \tag{6.42}$$

Given the discussion above, one might wonder if we should worry about mixing of $U(1)_R$ with $U(1)_B$. Happily, the meson M is neutral under $U(1)_B$ so it's not something that we have to worry about. We can say immediately that the dimension of the meson operator is

$$\Delta[M] = \frac{3(N_f - N_c)}{N_f} \tag{6.43}$$

Let's first test drive this formula by looking at what happens when $N_f \geq 3N_c$ where SQCD is infra-red free. At the edge, we have

$$N_f = 3N_c \quad \Rightarrow \quad \Delta[M] = 2 \tag{6.44}$$

But this is precisely what we expect. The theory is effectively free in the infra-red, so the fields ϕ and $\tilde{\phi}$ both have their canonical dimension $\Delta[\phi] = \Delta[\tilde{\phi}] = 1$ which agrees with the result (6.44). The result (6.44) is telling us that the scalar fields ϕ and $\tilde{\phi}$ (together with their fermionic partners) are free at $N_f = 3N_c$.

Note that there's already something a little surprising here. We knew that the theory was infra-red free at $N_f = 3N_c$, but only by computing the beta function. In contrast, the result above uses only the non-anomalous R-charge! Yet the two coincide. It's a sign that all these things are interconnected in SQCD in a way that doesn't happen in the absence of supersymmetry.

What happens if we now change N_f? We can start by looking at $N_f > 3N_c$ where, at first glance it appears that we become a little unstuck. Here the theory remains free and so we should still have $\Delta[M] = 2$. But that's not what the formula (6.41) seems to be telling us. However, since the theory is free in the IR, the anomalous $U(1)_A$ symmetry is reincarnated and can now mix with the R-symmetry, changing the answer. This is a salutary warning: there can be subtleties in blindly following (6.41).

Now let's look at what happens as we decrease N_f below the asymptotic freedom bound of $N_f = 3N_c$. We know that when $N_f = 3N_c - \epsilon$, for some small ϵ, we're sitting in a weakly coupled Banks-Zaksesque superconformal field theory. The formula (6.43) tells us that the meson has dimension

$$\Delta[M] = 2 - \frac{1}{3}\frac{\epsilon}{N_c} + \ldots$$

In other words, it's slightly less than two. You should think of the meson as describing a loosely bound state of ϕ and $\tilde{\phi}$. But as N_f decreases, so too does the dimension $\Delta[M]$. This is telling us that the state is becoming more and more tightly bound. At some point, the Banks-Zaks superconformal field theory becomes strongly coupled and we lose control over its dynamics. But, by the magic of supersymmetry, we remarkably keep control over the dimension of the chiral meson field! Eventually, the dimension of the meson his the bound (6.40). This occurs when

$$N_f = \frac{3}{2}N_c \quad \Rightarrow \quad \Delta[M] = 1$$

But, as we mentioned above, any scalar operator that has dimension 1 is necessarily a free scalar field. This equation is telling us that the binding between ϕ and $\tilde{\phi}$ has become so strong that the composite meson operator M is actually no longer composite! It is acting just like a fundamental scalar field. Moreover, it is now decoupled and is free.

How should we think of this? The proposal is that the meson becoming free signifies the end of the conformal window (6.39). In fact, we will argue shortly that the theory at $N_f = 3N_c/2$ is a completely free theory in the IR with a whole bunch of other fields joining M in the sense that they become non-interacting at low energies.

To argue this, we will turn to a new description of the physics that holds throughout the conformal window and, also, for $N_f < 3N_c/2$. This is known as the *dual description*.

6.5 Seiberg Duality

Throughout this section, our interest has been in massless SQCD, defined as

$$SU(N_c) \text{ gauge theory to coupled to } N_f \text{ flavours } \Phi \text{ and } \tilde{\Phi}$$

We've found a plethora of interesting physics as N_f is varied. But we haven't yet understood what happens when $N_f + 2 \leq N_c \leq 3N_c/2$. Moreover, at the lower end of conformal window, where we might expect a strongly interacting CFT, we've seen that the meson becomes free. It would certainly be good to understand this better.

Some light comes from a rather remarkable direction. Consider the following theory

$$SU(\tilde{N}_c) \text{ gauge theory to coupled to } N_f \text{ flavours } q \text{ and } \tilde{q} \text{ and } N_f^2 \text{ singlets } M$$

In the absence of the singlets, this clearly coincides with our earlier theory just with the number of colours renamed as $\tilde{N}_c$. However, we arrange the singlets as a matrix $M_j{}^i$ with $i, j = 1, \ldots, N_f$ which is subsequently coupled to the squark superfields through the superpotential

$$W = \lambda \tilde{q} M q \tag{6.45}$$

with λ a dimensionless coupling. This is now a slight twist on our original SQCD and its dynamics may differ. We'll see how below. Note that we've given the singlets the name M. You may recall that this is the also the name that we gave to the meson in our original theory. This is what writers call foreshadowing.

For our purposes, it's particularly interesting to consider the case where the number of colours in the two theories are related by

$$\tilde{N}_c = N_f - N_c \tag{6.46}$$

This second theory is known as *magnetic SQCD* (or mSQCD). We'll also at time refer to the original $SU(N_c)$ SQCD as the *electric* theory and we'll elucidate the reasons behind these names as we go along. We now make the following, somewhat astonishing, claim:

$$\boxed{SU(N_c) \text{ SQCD and } SU(N_f - N_c) \text{ mSQCD have the same low-energy physics}}$$

This relationship is known as *Seiberg duality*[11]. The purpose of this section is to give evidence for the claim and to understand its consequences.

[11]This was first proposed by Seiberg in the paper "Electric-Magnetic Duality in Supersymmetric Non-Abelian Gauge Theories".

6.5.1 Matching Symmetries

First let's look at some evidence. Given that the one of the two theories is always strongly coupled, it is challenging to do any direct calculations. The simplest thing that we can check is agreement of the symmetries.

Gauge Symmetries are Redundancies

First, the elephant in the room. The gauge symmetries are not the same! Should we care? The answer is no. Gauge symmetries are not true symmetries of a theory: they are merely a redundancy in the way we choose to describe the theory.

These are easy words to wheel out, but they also grate with other things we know about physics. The theory of electromagnetism is synonymous with $U(1)$ gauge theory. The Standard Model of particle physics is defined as having gauge group $SU(3) \times SU(2) \times U(1)$. If the gauge symmetry is something that isn't actually inherent to a theory, but just a redundancy in our choice of description, why do we hang so much on it elsewhere?

The reason is that gauge symmetry is an extraordinarily useful redundancy when theories are weakly coupled. In that situation, attempting to describe the physics in terms of anything other than the gauge field, with particular gauge group, is so ridiculously complicated that it borders on the absurd. You could, for example, choose to describe quantum Maxwell theory in terms of the field strengths $F_{\mu\nu}$ and *all* possible Wilson line operators $\exp\left(i \oint A\right)$ which carry the gauge invariant information. But that's certainly not easier than our usual gauge dependent description in terms of A_μ.

This means that when gauge theories are weakly coupled, the description in terms of the gauge symmetry G is indispensable. But when things become strongly coupled, the story is very different. In this case, the gauge symmetry reveals itself for what it is: a redundancy. Seiberg duality makes this stark. You can describe the same physics using two very different gauge theories. Sometimes one formulation is best suited to the problem at hand because the physics is weakly coupled in those variables. Sometimes the other formulation is easiest. But neither formulation is ever wrong and the fact that the gauge symmetries don't match in the two dual theories is a feature, not a bug.

Global Symmetries

The story is different for global symmetries. These must match. Moreover, as both theories are claimed to flow to the same infra-red physics, their UV 't Hooft anomalies must match as well. Let's see how we do.

It's useful to list, one last time, how the various fields transform. In the electric theory, we have

	$SU(N_c)$	$SU(N_f)_L$	$SU(N_f)_R$	$U(1)_B$	$U(1)_A$	$U(1)_R$
Φ	$\square$	$\square$	$\mathbf{1}$	1	1	$\frac{N_f-N_c}{N_f}$
$\tilde{\Phi}$	$\overline{\square}$	$\mathbf{1}$	$\overline{\square}$	-1	1	$\frac{N_f-N_c}{N_f}$
Λ^{b_0}	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{1}$	0	$2N_f$	0

with $b_0 = 3N_c - N_f$. For the magnetic theory, we have

	$SU(N_f - N_c)$	$SU(N_f)_L$	$SU(N_f)_R$	$U(1)_B$	$U(1)_A$	$U(1)_R$
q	$\square$	$\overline{\square}$	$\mathbf{1}$	$\frac{N_c}{N_f-N_c}$	-1	$\frac{N_c}{N_f}$
$\tilde{q}$	$\overline{\square}$	$\mathbf{1}$	$\square$	$-\frac{N_c}{N_f-N_c}$	-1	$\frac{N_c}{N_f}$
M	$\mathbf{1}$	$\square$	$\overline{\square}$	0	2	$\frac{2(N_f-N_c)}{N_f}$
$\tilde{\Lambda}^{\tilde{b}_0}$	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{1}$	0	$-2N_f$	0

Here $\tilde{\Lambda}$ is the strong coupling scale of the magnetic theory with $\tilde{b}_0 = 3(N_f - N_c) - N_f = 2N_f - 3N_c$ the 1-loop beta function.

The normalisation of the non-anomalous $U(1)_R$ charge is fixed, as usual, by the requirement that the (magnetic) gluinos have charge $+1$. (This, in turn, follows from the fact that the superspace coordinate has $R[\theta] = -1$.) This, in turn, fixes the R-charge for the dual squarks which came be written as $R[q] = R[\tilde{q}] = N_c/N_f = (N_f - \tilde{N}_c)/N_f$, where we see that it mimics the form in the original theory. The requirement that the superpotential has $R[W] = 2$ then fixes the R-charge of the singlet M.

$$R[M] = \frac{2(N_f - N_c)}{N_f}$$

But this is the same as the R-charge as the meson $\tilde{\Phi}\Phi$ in the original electric theory. Moreover, because these are chiral fields, if their R-charges match then so too do their dimensions. This provides our first, and most important, entry in the dictionary relating the electric and magnetic theories: the singlet fields M in the magnetic theory correspond to the meson in the electric theory.

$$M \sim \tilde{\Phi}\Phi$$

This matching provides an opportunity to reiterate a lesson from above. We have not attempted to match individual quarks and gluons on the two sides of the duality.

This is because these are not gauge invariant objects and so have no physical meaning on their own. However, gauge invariant observables or fields should match across the duality.

Next the $U(1)_B$ charges. We want to identify $U(1)_B$ in the two theories but there's an ambiguity in the normalisation. We've fixed this ambiguity in the table above by ensuring that the dual baryons $b \sim q^{N_f - N_c}$ and $\tilde{b} \sim \tilde{q}^{N_f - N_c}$ have the same $U(1)_B$ charges as their electric counterparts B and $\tilde{B}$. Crucially, their R-charges also match. This then provides the second entry in our dictionary between the two theories: $B \sim b$ and $\tilde{B} \sim \tilde{b}$. We will look a little closer at the identification of these operators shortly.

't Hooft Anomaly Matching

Now we can play the increasingly familiar 't Hooft anomaly game. We denote the 't Hooft anomalies in the original theory as $\mathcal{A}_{\rm el}$ and those in the dual as $\mathcal{A}_{\rm mag}$. We have

$\underline{SU(N_f)^3_L}$: The quarks contribute $\mathcal{A}_{\rm el} = N_c$ while the dual quarks and mesons give $\mathcal{A}_{\rm mag} = -(N_f - N_c) + N_f$. Note that it was important that the dual quarks sit in the $\overline{\Box}$ of $SU(N_f)_L$ while the quarks sit in the $\Box$. This was also need to ensure that the meson fields M have the same quantum numbers.

$\underline{SU(N_f)^2_L \cdot U(1)_B}$: We have $\mathcal{A}_{\rm el} = \mathcal{A}_{\rm mag} = N_c$.

$\underline{SU(N_f)^2_L \cdot U(1)_R}$: We have $\mathcal{A}_{\rm el} = -N_c^2/N_f$ and $\mathcal{A}_{\rm mag} = (N_f - N_c) \times \frac{N_c - N_f}{N_f} + N_f \times \frac{N_f - 2N_c}{N_f}$ which agree. This same counting essentially ensures that the mixed $U(1)_R$-gravitational anomaly also matches.

The 't Hooft anomalies for $U(1)_B$ and '$U(1)^3_B$ trivially vanish in both the electric and magnetic theories because $U(1)_B$ is a vector-like symmetry. However, we do have the mixed anomaly

$\underline{U(1)^2_B \cdot U(1)_R}$: $\mathcal{A}_{\rm el} = 2N_f N_c \times \left(-\frac{N_c}{N_f}\right) = -2N_c^2$. The magnetic theory has $\mathcal{A}_{\rm mag} = 2(N_f - N_c)N_f \left(\frac{N_c}{N_f - N_c}\right)^2 \times \left(\frac{N_c - N_f}{N_f}\right) = -2N_c^2$

For the final matchings involving just $U(1)_R$, we need to remember the existence of the gluinos.

$\underline{U(1)_R}$: We have $\mathcal{A}_{\rm el} = (N_c^2-1) + 2N_cN_f(-N_c/N_f) = -(N_c^2+1)$. And

$$\mathcal{A}_{\rm mag} = ((N_f-N_c)^2-1) + 2(N_f-N_c)N_f\left(\frac{N_c-N_f}{N_f}\right) + N_f^2\left(\frac{N_f-2N_c}{N_f}\right) = -(N_c^2+1)$$

$\underline{U(1)_R^3}$: Now

$$\mathcal{A}_{\rm el} = (N_c^2-1) + 2N_cN_f\left(-\frac{N_c}{N_f}\right)^3$$

and

$$\mathcal{A}_{\rm mag} = ((N_f-N_c)^2-1) + 2(N_f-N_c)N_f\left(\frac{N_c-N_f}{N_f}\right)^3 + N_f^2\left(\frac{N_f-2N_c}{N_f}\right)^3$$

Both are equal. We see that all the anomalies do indeed match and seemingly in a non-trivial fashion.

6.5.2 Completing the Phase Diagram for SQCD

Next, let's look at some of the more immediate consequences of the duality. Clearly magnetic SQCD, as defined in (6.46), only makes sense when $N_f \geq N_c + 2$ so the claim of Seiberg duality is that it has something to tell us about the original theory in this regime. Moreover, we know that mSQCD is no longer asymptotically free when

$$N_f \geq 3\tilde{N}_c \quad \Rightarrow \quad N_f \leq \frac{3}{2}N_c$$

But this is precisely the regime $N_c + 2 \leq N_f \leq 3N_c/2$ that was left unresolved by our previous methods.

If Seiberg duality is correct (and we have every reason to believe that it is!) then it gives a very surprising answer for what happens in this regime: the original $SU(N_c)$ gauge theory becomes strongly coupled and flows, in the infra-red, to an entirely different $SU(N_f - N_c)$ gauge theory, coupled to the the matter q, $\tilde{q}$ and M. This is known as the *free magnetic phase*.

Note that there is no suggestion that $SU(N_f - N_c)$ is a subgroup of $SU(N_c)$, one that perhaps arises through a Higgs mechanism. The gluons of $SU(N_f - N_c)$ are *not* the gluons of $SU(N_c)$! Instead they are new, composite spin 1 particles that emerge at strong coupling, presumably some complicated bound states of all the degrees of freedom of the original electric theory. . We will have more to say about how the two gauge groups are related in Section 6.5.4.

Let's now increase N_f for fixed N_c. When the electric theory sits in the conformal window, so too does the magnetic dual

$$\frac{3}{2}N_c < N_f < 3N_c \quad \Leftrightarrow \quad 3\tilde{N}_c > N_f > \frac{3}{2}\tilde{N}_c$$

However, crucially, when one theory is weakly coupled, the other is necessarily strongly coupled. For example, at the far end of the conformal window, $N_f = 3N_c - \epsilon$, the original electric theory is at a Banks-Zaks fixed point and under control, while the magnetic theory is something strongly coupled. In contrast, at the lower end of the conformal window, $N_f = \frac{3}{2}N_c + \epsilon$, it is the other way around: the dual magnetic theory sits at (something like) a Banks-Zaks fixed point, while the electric theory is strongly coupled.

To understand the fate of the magnetic theory, we also need to take into account the effect of the superpotential

$$W \sim \tilde{q}Mq$$

Viewed from the perspective of the UV, this superpotential gives Yukawa terms between various fermions and scalars in the magnetic theory. The parameter λ is dimensionless, so this appears to be a marginal operator. But, a one-loop calculation shows that λ initially decreases as we flow towards the infra-red. The superpotential is a marginally irrelevant operator of the free, UV fixed point.

However, this story is different when viewed from the infra-red. Suppose that we first flow to the fixed point within the conformal window of mSQCD and then add the superpotential (6.45). What now happens? To understand this, we need to compute the dimension of the superpotential W at the IR fixed point.

Happily, supersymmetry gives us a handle on this because W is a chiral and so its dimension is related to its R-charge. As we've seen above, the R-charges of the dual squarks are $R[q] = R[\tilde{q}] = N_c/N_f$. That leaves us with the meson field M. And here there's something of a subtlety.

We already listed the R-charge of M in the table above but we need to revisit this. That R-charge was determined by assuming that $R[W] = 2$ which is pre-judging the answer! This is not what we want for the present calculation. Instead, we need to remember that before we add the superpotential, M is just a free field, decoupled from everything else. This means that it has dimension $\Delta[M] = 1$ and, correspondingly,

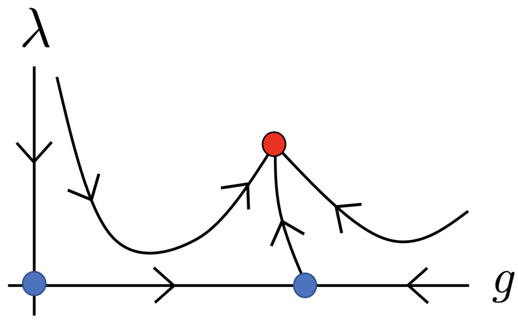

Figure 12. The RG flow in mSQCD The free fixed point and the fixed point in the conformal window are shown as black points. The superpotential induces a further flow to the red point. This is conjectured to coincide with the fixed point of SQCD.

$R[M] = 2/3$. This means that, from the perspective of the IR, the superpotential $W = \tilde{q}Mq$ has dimension

$$\Delta[W] = \frac{3}{2}R[W] = \frac{3}{2}\left(\frac{2}{3} + \frac{2N_c}{N_f}\right) = 1 + \frac{3N_c}{N_f}$$

When we first enter the lower bound of the conformal window, we have

$$N_f > \frac{3}{2}N_c \quad \Rightarrow \quad \Delta[W] < 3$$

But this means that the superpotential is *always* a relevant deformation in the conformal window! (The measure in the action is $\int d^4x\, d^2\theta$ and $[d^4x] = -4$ while $[d^2\theta] = +1$ which is the why the the bound for a relevant superpotential is $\Delta[W] < 3$.)

The RG flows are shown in Figure 12. There are three fixed points in the magnetic theory: the free theory at $g = \lambda = 0$ that can be thought of as the starting point in the UV; the fixed point without a superpotential in the conformal window with $\lambda = 0$ and $g \neq 0$; and the final fixed point with $g, \lambda \neq 0$. The claim of Seiberg duality is that this final fixed point of the dual theory, shown as the red dot, coincides with the fixed point in the conformal window of the electric theory.

By the time we reach our final fixed point, shown by the red dot in the figure, we should now take $R[W] = 2$. This gives us the R-charge $R[M]$ that we listed in the table with the corresponding dimension

$$R[M] = \frac{2(N_f - N_c)}{N_f} \quad \Rightarrow \quad \Delta[M] = \frac{3(N_f - N_c)}{N_f}$$

It's only when we reach this fixed point that the R-charge and dimension of M in the magnetic theory coincides with those of the meson in the original theory.

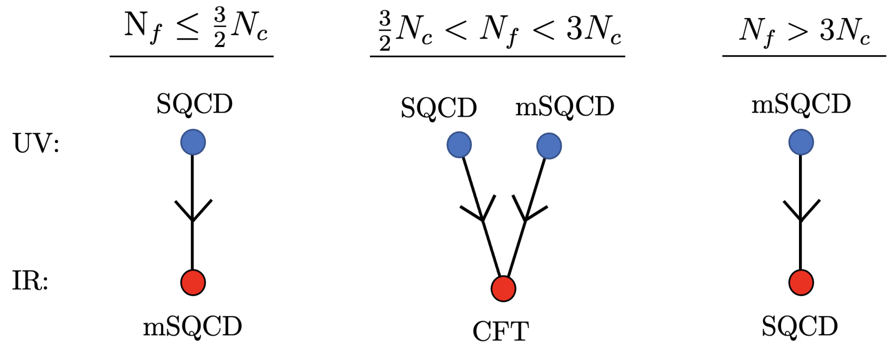

Figure 13. Seiberg duality is a statement about RG flows, although the precise statement changes as we vary N_f/N_c.

As we increase $N_f \ge 3N_c$, there is no mystery about our electric theory: it is free in the infra-red. In contrast, the magnetic theory flows to strong coupling but now becomes the weakly interacting $SU(N_c)$ theory in the infra-red. We see again that Seiberg duality is an example of a strong-weak coupling duality. When one theory is strongly coupled, the other may be weakly coupled and vice versa. This makes it useful.

Of course there are also regimes – notably in the middle of the conformal window – when both theories are strongly coupled. So the duality isn't a magic bullet, solving all our woes. But it is a dramatic and unexpected step forward.

All of this means that the exact interpretation of Seiberg duality depends on the value of N_f/N_c. For small N_f, the electric theory flows to the weakly coupled magnetic theory. For large N_f, the opposite happens: the magnetic theory flows to a weakly coupled electric theory. While for N_f in the conformal window, both theories flow to the same infra-red fixed point. This is summarised in Figure 13. However, in all cases Seiberg duality is a statement about RG flows. This should be distinguished from other "exact dualities" of quantum field theories or many body systems, where there are two very different descriptions that hold at any energy scale. Examples of exact dualities includes the high/low temperature duality of the Ising model, or electromagnetic dualities of $\mathcal{N} = 2$ and $\mathcal{N} = 4$ supersymmetric theories.

6.5.3 Deformations of the Theories

So far we've focussed on the fixed point. But both theories also have a moduli space of vacua, and this too should match. However, showing this isn't straightforward because,

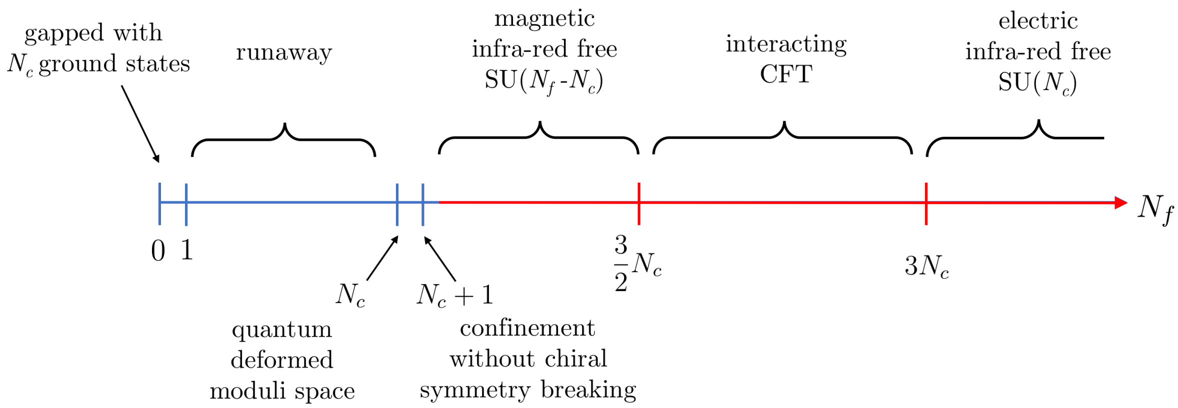

Figure 14. The phases of massless SQCD. For the values of N_f shown in red, we have a dual description in terms of mSQCD. This dual description is weakly coupled from $N_f = N_c + 2$ to $N_f = 3N_c/2 + \epsilon$.

as we saw in Section 4.3, there are some non-trivial constraints between the mesons and baryons.

Nonetheless, we can see roughly how things work. We've already seen that the singlets M are dual to the mesons in the electric theory

$$\tilde{\Phi}\Phi \sim M \tag{6.47}$$

The symmetries also allow us to match the baryon degrees of freedom

$$B^{i_1 \ldots i_{N_c}} \sim \epsilon^{i_1 \ldots i_{N_c} j_1 \ldots j_{\tilde{N}_c}} b_{j_1 \ldots j_{\tilde{N}_c}}$$
$$\tilde{B}_{i_1 \ldots i_{N_c}} \sim \epsilon_{i_1 \ldots i_{N_c} j_1 \ldots j_{\tilde{N}_c}} b^{j_1 \ldots j_{\tilde{N}_c}}$$

Each transforms in the $\binom{N_c}{N_f}$-antisymmetric representation of $SU(N_f)$ which, of course, is equivalent to the $\binom{N_f - N_c}{N_f}$-antisymmetric representation.

The magnetic theory also has its own meson fields $\tilde{m} = \tilde{q}q$ and you might wonder what becomes of these. But the equation of motion for the singlets M is simply $\tilde{m} = 0$ so these dual mesons don't give us any further light degrees of freedom.

Masses and Expectation Values

We can now perform some simple tests of the duality. Suppose that we turn on the electric meson fields to move out on the moduli space. To start we just turn on a single entry

$$\tilde{\phi}\phi = \begin{pmatrix} v & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}$$

This breaks the gauge symmetry $SU(N_c) \to SU(N_c - 1)$, now with $N_f - 1$ flavours. We would like to see this behaviour in the dual theory. In fact, this is straightforward. Giving the singlet M the same expectation value, we have

$$W_{\rm mag} \sim \tilde{q}Mq = v\tilde{q}_1 q_1$$

This is just a mass term for the dual squark and we can integrate it out, giving us $SU(\tilde{N}_c)$ with $N_f - 1$ flavours. This is the expected dual.

Alternatively, we could give a mass to one of the quarks in the electric theory by adding the superpotential

$$W_{\rm el} = m\tilde{\Phi}_1\Phi_1$$

After integrating out this massive flavour, we're left with $SU(N_c)$ with $N_f - 1$ flavours.

In the magnetic theory, this same mass deformation gives

$$W_{\rm mag} = \tilde{q}Mq + mM_{11}$$

The equation of motion for the singlet M then induces an expectation value for the dual squark

$$\tilde{q}_1 q_1 = -m$$

This, in turn, breaks the dual gauge group $SU(\tilde{N}_c) \to SU(\tilde{N}_c - 1)$. The upshot is that we're left with the dual theory of an $SU(N_f - N_c - 1)$ gauge group coupled to $N_f - 1$ flavours. This is the expected result.

We see that these simple deformations respect the duality, with a mass term on one side mimicked by a Higgs effect on the other.

Matching RG Scales

There's a slight subtlety that we've brushed under the carpet so far. The key element in our dictionary relating mesons $\tilde{\Phi}\Phi \sim M$ can't quite be right. This is because the quarks on the left-hand side are defined in the UV of SQCD and each have dimension 1 so $\tilde{\Phi}\Phi$ has dimension 2. Meanwhile the singlet M is a free field in the dual theory so has dimension 1. So our dimensional analysis is amiss.

This should be straightforward to patch up: we just need some invariant RG scale to take up the slack. But this scale should be holomorphic and, moreover, we don't want it to mess up the symmetries on the two sides. Either the electric RG scale Λ or magnetic scale $\tilde{\Lambda}$ change the (admittedly spurious) $U(1)_A$ charge. But we can introduce a new scale μ which is some geometric mean of the two

$$\Lambda^{3N_c-N_f}\tilde{\Lambda}^{3(N_f-N_c)-N_f} = (-1)^{N_f-N_c}\mu^{N_f} \tag{6.48}$$

The scale μ is, by construction, invariant under all symmetries, spurious or otherwise. A better characterisation of the dictionary is then

$$\frac{\tilde{\Phi}\Phi}{\mu} = M$$

The strange looking minus sign in (6.48) is largely a convention, but it can be shown to ensure that the dual of the dual theory brings us back to the original.

The Theory $N_f = N_c + 1$ Again

We've advertised Seiberg duality as holding for $N_f \geq N_c + 2$. But it also gives the right answer for $N_f = N_c + 1$, at least if we include the additional term $\det M$ in the superpotential so that (6.45) becomes

$$W \sim \det\, M + \tilde{q}Mq$$

This is the expected superpotential (6.37) for the $N_f = N_c + 1$ theory, with the dual quarks q and $\tilde{q}$ identified with the baryons B and $\tilde{B}$.

A Glimpse of the Superconformal Index

Until now, we've given no more than plausible evidence for Seiberg duality. The symmetries and 't Hooft anomalies match and it passes some simple tests as we deform the theory. It turns out that there is a much more quantitative test that the duality passes. This comes from computing an object known as the *superconformal index*.

The superconformal index is an extension of the Witten index. While the Witten index receives contributions only from the ground states, the superconformal index receives contributions from a much larger, but still restricted class of states. Moreover, it can be reliably computed for theories even at weak coupling.

The superconformal index is defined for superconformal theories on $\mathbf{S}^3 \times \mathbb{R}$. It is a function of two variables, p and q, by tracing over all states

$$\mathcal{I}(p,q) = \text{Tr}\,(-1)^F p^{j_1+j_2-\frac{1}{2}R} q^{j_1-j_2-\frac{1}{2}R}$$

Here R is the R-charge of the state while j_1 and j_2 are the two angular momenta associated to the rotation group $SO(4) \cong SU(2) \times SU(2)$.

The formulae for the superconformal indices are fairly complicated and, at first glance, look very different for SQCD and mSQCD. It is a highly non-trivial mathematical fact that these formulae do, in fact, coincide[12].

6.5.4 Why Seiberg Duality is Electromagnetic Duality

There is one feature of Seiberg duality that perhaps remains mysterious: why have we called the dual theory "magnetic" and the original theory "electric"? The answer to this gets to the heart of how to think about Seiberg duality and other related phenomena.

The basic idea goes back to Maxwell theory. The equations of motion are usually written as

$$\partial_\mu F^{\mu\nu} = J^\mu \quad \text{and} \quad \partial_\mu {}^\star F^{\mu\nu} = 0$$

with J^μ the electric current. If there are no charged particles in the theory then $J^\mu = 0$ and the Maxwell equations exhibit a surprising symmetry in which we exchange $F^{\mu\nu} \rightarrow {}^\star F^{\mu\nu}$. In terms of the underlying electric and magnetic fields, this means

$$\mathbf{E} \rightarrow \mathbf{B} \quad \text{and} \quad \mathbf{B} \rightarrow -\mathbf{E}$$

This is *electromagnetic duality*. It is broken in electromagnetism because our world has electric sources, but no magnetic sources.

However, one could imagine a theory in which there are particles carrying both electric and magnetic charges. The latter are called *magnetic monopoles*. In this case, Maxwell's equations should be replaced by

$$\partial_\mu F^{\mu\nu} = J^\mu_e \quad \text{and} \quad \partial_\mu {}^\star F^{\mu\nu} = J^\mu_m$$

with J^μ_e and J^μ_m the electric and magnetic currents respectively. In such a theory, electromagnetic duality may be restored, now with the electric and magnetic particles interchanged. However, there is a consistency condition between electric charges $q_{\rm el}$ and magnetic charges $q_{\rm mag}$: they can be shown to obey the *Dirac quantisation condition*

$$\frac{q_{\rm el} q_{\rm mag}}{2\pi} \in \mathbb{Z}$$

A derivation of this can be found in the lectures on Gauge Theory. This has an interesting consequence. The electric charge is a measure of the strength of the electromagnetic force. (For example, the fine structure constant is $\alpha = q_{\rm el}^2/4\pi\epsilon_0\hbar c$.) The Dirac quantisation conditions tells us that if the electric charges are weakly coupled, then magnetic charges will necessarily be strongly coupled.

[12]For more information about the superconformal index, see the lectures by Yuji Tachikawa or by Abhijit Gadde.

It's not so easy to write down versions of QED that include both electric and magnetic charges. This is because we must work with the gauge field A_μ, and the resulting Bianchi identity $\partial_\mu {}^\star F^{\mu\nu} = 0$ immediately implies that there are no magnetic monopoles. However, the story becomes richer in certain non-Abelian gauge theories. It turns out that some non-Abelian gauge theories necessarily have magnetic monopoles arising as solitons. This means that although we start by writing a theory purely of electric charges, the actual theory includes both electric and magnetic charges. Examples of theories with solitonic magnetic monopoles include $\mathcal{N} = 2$ and $\mathcal{N} = 4$ super Yang-Mills.

However, the $\mathcal{N} = 1$ SQCD theories that we've been considering in this Section do not obviously contain magnetic monopoles. There are certainly no classical soliton solutions that one can construct that have magnetic charge. On the other hand, the theories are strongly coupled and it's not at all clear what properties their excitations have. Part of the claim of Seiberg duality is that the dual description should really be thought of as a kind of electromagnetic duality, with the $SU(N_f - N_c)$ gauge group related to the original $SU(N_c)$ gauge group by something morally equivalent to swapping electric and magnetic fields. Correspondingly, the dual baryons b and $\tilde{b}$ should be viewed as some kind of magnetic excitation from the perspective of the original theory.

You may have noticed that I'm saying a lot of words here and not writing down any formulae! That's because it's difficult to make the above claims precise. There are, however, some hints that this is the right way to think about things. For example, the relationship (6.48) between the scales

$$\Lambda^{3N_c - N_f}\tilde{\Lambda}^{2N_f - 3N_c} \sim \text{constant}$$

This formalises something that we've already seen: Seiberg duality is a strong-weak duality. As the gauge coupling in one theory gets smaller, the coupling in the other gets larger. This is reminiscent of the behaviour in electromagnetic duality.

However, the best evidence that Seiberg duality should be viewed as electromagnetic duality comes from exploring other theories. In particular, $\mathcal{N} = 2$ and $\mathcal{N} = 4$ theories both exhibit a form of electromagnetic duality where both electric and magnetic degrees of freedom can be made manifest. The existence of a duality means that there are two formulations of the theory, one in which the electric objects are viewed as fundamental particles and the other in which magnetic objects are fundamental particles. In either of these descriptions, the other particles arise as solitons. Its only when Seiberg duality is viewed within this larger context as one of many dualities among quantum field theories, that it becomes clearer that it is, indeed, a version of electromagnetic duality.