Supersymmetric Field Theory

University of Cambridge Part III Mathematical Tripos

David Tong

Department of Applied Mathematics and Theoretical Physics,
Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge, CB3 OBA, UK

http://www.damtp.cam.ac.uk/user/tong/susy.html
d.tong@damtp.cam.ac.uk

Recommended Books and Resources

Here is a collection of useful textbooks on supersymmetry.

- Wess and Bagger "*Supersymmetry*"

This is a strange little book, with chapters that are 2 pages long followed by several pages of key equations. It's not particularly good for learning the subject, but makes a remarkably useful reference guide.

- Bailin and Love "*Supersymmetric Gauge Field Theory and String Theory*"

Probably the best book covering the basics of supersymmetric Lagrangians.

- Dan Freedman and Toine Van Proeyen "*Supergravity*"

As the name suggests, this book is mostly focussed on supergravity rather than global supersymmetry. But it kicks off with a really excellent description of classical field theory. The section on spinors in various dimensions is particularly useful.

- Steven Weinberg "*The Quantum Theory of Fields, Volume III: Supersymmetry*"

The third volume of Weinberg's magnum opus covers supersymmetry. As always, it contains many important things that are difficult to find elsewhere. As always, these things are sometimes frustratingly buried in unconventional notation and dressed with more indices than you can shake a stick at.

- John Terning "*Modern Supersymmetry: Dynamics and Duality*"

This is one of the few books (possibly the only book) that describes the quantum dynamics of supersymmetric field theories, rather than just their classical action. (Weinberg has a chapter on the Seiberg-Witten solution, but it feels like his heart isn't in it and any mention of Seiberg duality is noticeably absent.) There are, fortunately, many lecture notes that make up for the deficiency. You can find links on the course webpage.

Contents

Acknowledgements

These lecture notes owe a debt to Graham Shore, from whom I first learned supersymmetry, and to Philip Argyres' wonderful notes on the subject. I'm also grateful to Ben Allanach and Fernando Quevedo, both previous lecturers of the supersymmetry course in Cambridge.

This is one of the more advanced courses in Part III. It assumes a familiarity with quantum field theory, in particular the renormalisation group. You will also need to be comfortable with some group theory.

Spinor Conventions

We work in Minkowski space with signature $(+,-,-,-)$. Spinor indices are raised and lowered with $\psi^\alpha = \epsilon^{\alpha\beta}\psi_\beta$ and $\bar{\psi}^{\dot{\alpha}} = \epsilon^{\dot{\alpha}\dot{\beta}}\bar{\psi}_{\dot{\beta}}$ where the invariant, anti-symmetric tensor is

$$\epsilon^{\alpha\beta} = \epsilon^{\dot{\alpha}\dot{\beta}} = -\epsilon_{\alpha\beta} = -\epsilon_{\dot{\alpha}\dot{\beta}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Left-handed spinors are contracted as $\psi\chi = \psi^\alpha\chi_\alpha$ and right-handed spinors are contracted as $\bar{\psi}\bar{\chi} = \bar{\psi}_{\dot{\alpha}}\bar{\chi}^{\dot{\alpha}}$. Sigma matrices are defined by

$$(\sigma^\mu)_{\alpha\dot{\alpha}} = (1, \sigma^i)_{\alpha\dot{\alpha}} \quad \text{and} \quad (\bar{\sigma}^\mu)^{\dot{\alpha}\alpha} = \epsilon^{\alpha\beta}\epsilon^{\dot{\alpha}\dot{\beta}}\sigma^\mu_{\beta\dot{\beta}} = (1, -\sigma^i)^{\dot{\alpha}\alpha}$$

and the generators of the Lorentz group in the left-handed and right-handed spinor representation are, respectively,

$$(\sigma^{\mu\nu})_\alpha{}^\beta = \frac{i}{4}\left(\sigma^\mu\bar{\sigma}^\nu - \sigma^\nu\bar{\sigma}^\mu\right)_\alpha{}^\beta \quad \text{and} \quad (\bar{\sigma}^{\mu\nu})^{\dot{\alpha}}{}_{\dot{\beta}} = \frac{i}{4}\left(\bar{\sigma}^\mu\sigma^\nu - \bar{\sigma}^\nu\sigma^\mu\right)^{\dot{\alpha}}{}_{\dot{\beta}}$$

1 Introduction

Supersymmetry is the name given to a novel symmetry that relates bosons and fermions. In many ways it is a surprise that such a symmetry could exist at all. This is because bosons and fermions are, to put it mildly, different.

Bosons are gregarious. Put many of them in a box and they huddle together to form a macroscopic quantum object called a Bose-Einstein condensate. In contrast, fermions are loners, an isolation enforced by the Pauli exclusion principle. Put many fermions in a box and you get a more familiar, but ultimately even stranger, state of quantum matter called a Fermi surface.

Within the framework of relativistic quantum field theories, the difference between fermions and bosons is even more stark. Fermions are matter particles. Bosons are force carriers. Any symmetry that relates the two must somehow entail a unification of matter and force.

Of course, we know from our earlier lessons on Quantum Field Theory that the distinction between bosons and fermion can be traced to something that is, in some sense, rather minor. They differ only by the simple matter of $\hbar/2$ in their angular momentum, with the spin-statistics theorem then doing the heavy lifting that ensures the resulting particles have such different properties. However, this too highlights just how unusual supersymmetry must be. The angular momentum of a particle is a property that follows from the symmetries of spacetime. Anything that relates particles with different angular momentum must involve some kind of extension of the symmetries of spacetime. And that sounds interesting!

All of this means that it's not at all obvious that something like supersymmetry can exist and we should, if nothing else, be curious about how it can come about. But why else should we care? In the rest of this introduction, I give three reasons why studying supersymmetric quantum field theories is worthwhile.

Reason 1: Strongly Interacting Quantum Field Theories

Quantum field theory is hard. This is particularly true when coupling constants, which specify the strength of interactions, are not small. This means that we can no longer understand the physics using the familiar methods of perturbation theory and Feynman diagrams. In this case, the word "hard" typically means "no one knows how to solve it".

Supersymmetric theories are not wildly different from other quantum field theories. They have a carefully curated collection of fields, with some interactions tuned to take certain values, but otherwise they exhibit many of the strongly coupled phenomena expected of any other quantum field theory. The magic of supersymmetry, however, is that in many cases we are able to make exact statements about the properties of the theory. This is because supersymmetry places certain restrictions on the kind of dynamics that can occur. Fortuitously, it turns out that these restrictions are not strong enough to stop interesting things happening, but are strong enough to allow us to solve certain aspects of the theory. In this way, supersymmetric field theories provide an important collection of toy models that allow us to understand what quantum field theory can do in regimes where we would otherwise have very little control.

Here is an example. The theory of the strong nuclear force, QCD, exhibits a remarkable property known as *confinement*. Quarks are always trapped inside hadrons and we never see isolated quarks on their own. There is no doubt that the theory of QCD has this property – we can see it clearly in numerical simulations – but we are a long way from being able to prove confinement from first principles. However, there are supersymmetric gauge theories, similar to QCD but with slightly different matter content, where confinement can be proven analytically. (This follows from the famous Seiberg-Witten solution of $\mathcal{N} = 2$ supersymmetric theories.) While the supersymmetric proof of confinement is not directly applicable to real-world QCD, it nonetheless gives us good intuition for how confinement might proceed in that context.

These lectures will very much be given in the spirt of using supersymmetry to tell us interesting things about strongly coupled quantum field theories. We will learn about topics that exist for real world QCD, such as confinement and chiral symmetry breaking, and see how these manifest themselves in more tractable supersymmetric theories. We will also learn about novelties that appear not to be of relevance for QCD but give us an insight into what strongly interacting quantum field theories can do. Foremost among these novelties is the concept of *duality*, the idea that two very different looking quantum field theories may, in fact, describe the same physics.

Reason 2: Mathematics

As our understanding of supersymmetric field theories grew, increasingly sophisticated mathematical constructs were found lurking within them. These are primarily, but not exclusively, ideas from geometry.

This link between supersymmetry and mathematics starts with some simple quantum mechanical models whose solutions give new perspectives on, among other things, Morse

theory and index theorems. But the real fun starts when we turn to supersymmetric field theories. Understanding supersymmetric field theories in $d = 1 + 1$ dimensions led to the discovery of *mirror symmetry*, a relationship between topologically distinct manifolds. As we move to higher dimensional quantum field theories, we find ever more elaborate structures, some of which are known to mathematicians and some of which are novel. It is clear that there is much more to uncover.

We won't have anything to say about the connection to mathematics in these lectures, although we will stumble upon the concept of *Kähler geometry* as we proceed which at least gives a feel for how interesting geometric concepts arise naturally from supersymmetry. The companion lectures on Supersymmetric Quantum Mechanics have more of an eye towards the mathematical aspects of supersymmetry, albeit without getting very deep into the subject.

Reason 3: Our World

The million dollar question is: does supersymmetry have anything to do with our world? The rather disappointing answer is: we don't know.

There is certainly no experimental evidence that supersymmetry is a symmetry of nature at the fundamental level. Moreover, it's not for want of trying. To fill in the details, I'll first explain what it would mean for our world to be supersymmetric. Then I'll explain what reasons we have (or had!) for thinking that this might be the case.

In any supersymmetric theory, particles come in pairs – one a boson, the other a fermion – and this pair of particles share many of their properties, such as their masses and the forces that they experience. You don't need to build an LHC to realise that our world most certainly does not have this property! There is no bosonic particle with the same mass and charge as an electron; no massless fermionic particle with the same properties as the photon. (No, the neutrino doesn't do it!) There is, in short, no supersymmetry.

However, not all symmetries are manifest in the world around us. This is because of the phenomenon of *symmetry breaking* in which the dynamics of the theory make a choice which masks the underlying symmetry. There are many examples of symmetry breaking that we know take place, some mundane and familiar, others more exotic. Here are two. In a magnet, all the spins align in a given direction, breaking the underlying rotation symmetry. In the Standard Model, electroweak symmetry is broken by the Higgs boson ensuring, among other things, that the (left-handed) electron and neutrino look very different to our low-energy eyes despite the fact that they are indistinguishable at high energies.

It may well be that supersymmetry is a symmetry of our world but is broken and so hidden at low-energies. If this is the case, the breaking comes with an energy scale that we will call $M_{\rm susy}$. All of the superpartners – the other half of each boson/fermion pair – would get a mass that sits somewhere around $M_{\rm susy}$. So to answer the question of whether supersymmetry exists in nature we must also address the partner question: what is the scale of $M_{\rm susy}$?

For many years, supersymmetry was viewed as the most promising candidate for physics beyond the Standard Model, with $M_{\rm susy} \approx 1$ TeV. At this scale, supersymmetry provides a compelling solution to the hierarchy problem (the question of why the Higgs mass is not driven to higher scales by quantum fluctuations). Furthermore, if you adopt this solution then it comes with a number of happy consequences, from the unification of coupling constants to enticing candidates for dark matter.

However, with the advent of the LHC we have now explored the TeV scale and there is no sign of the predicted superpartners. It's not quite game over: it may well be that these extra particles are lurking just around the corner, tantalisingly out of reach of our current accelerator and will be found as we go to higher energies. But it's certainly fair to say that the parameter space of allowed theories has shrunk dramatically, as have our reasons for believing in supersymmetry at the TeV scale. This means that if supersymmetry is a symmetry of our world, it now appears to be broken at some scale $M_{\rm susy} \gtrsim 1$ TeV. But where?

There is reason to think that supersymmetry might show up by the time we reach the Planck scale $M_{\rm pl} \approx 10^{15}$ TeV. This reason is string theory. Of course, we don't know that string theory is the right theory of quantum gravity but it is presently the only viable candidate where a microscopic quantum theory gives the Einstein equations emerging at large distances. And string theory appears to require supersymmetry. (I include the word "appears" here because there are some open questions about bosonic (i.e. non-supersymmetric) string theory that we don't have a good handle on and it may be premature to throw this out as a viable theory.)

So if you buy into string theory, then you'll most likely want supersymmetry to be manifest by the time you get $M_{\rm pl}$. And, as we've seen above, it looks like it should be broken at some scale $M_{\rm susy} \gtrsim 1$ TeV. But there are 15 orders of magnitude between the TeV scale and the Planck scale. Where in this range should we expect supersymmetry to be broken if not at the TeV scale, or just above it, to provide a solution to the hierarchy problem? Sadly, I don't think that we have any good idea, and there are

no hints from nature that it is more useful to have $M_{\rm susy}$ at some large scale $\gg$ TeV rather than another.

This leaves us with the current situation, one of no small befuddlement about what role, if any, supersymmetry has to play in our world. Given this, in these lectures we won't make any attempt to describe how supersymmetry may appear in our world. In particular, we will not devote effort to constructing supersymmetric versions of the Standard Model (the simplest is known as the MSSM where the first M stands for "minimal" and you can guess the rest) nor will we describe the many subtleties that come with how supersymmetry might be broken and how this manifests itself. Instead we will focus on places where supersymmetry has proved invaluable, viewing the theories as toy models to guide us in our understanding of quantum field theories.

1.1 A First Look at Supersymmetry

To motivate some of what lies ahead, we'll jump in with a particularly simple supersymmetric theory. The theory consists of a single, complex scalar ϕ together with a 2-component Weyl fermion ψ_α. (If you're unfamiliar with Weyl fermions, we'll describe their properties in detail in Section 2.1.)

The following action has kinetic terms for these two fields, together with some carefully tuned interactions

$$S = \int d^4x \ \left[\partial_\mu \phi^\dagger \partial^\mu \phi - i\psi\sigma^\mu \partial_\mu \bar{\psi} - \left| \frac{\partial W}{\partial \phi} \right|^2 - \frac{1}{2}\frac{\partial^2 W}{\partial \phi^2}\psi\psi - \frac{1}{2}\frac{\partial^2 W^\dagger}{\partial \phi^{\dagger\,2}}\bar{\psi}\bar{\psi} \right] \tag{1.1}$$

Here $\sigma^\mu = (1, \sigma^i)$ with σ^i the usual collection of three Pauli matrices. Note that there is a relation between the scalar potential $V(\phi) = |W'(\phi)|^2$ and the scalar-fermion interactions, both of which are dictated by a function $W(\phi)$ known as the *superpotential*. If we want a renormalisable theory, this function should be no more than cubic

$$W(\phi) = \frac{1}{2}m\phi^2 + \frac{1}{3}\lambda\phi^3$$

This ensures that the potential is a quartic polynomical, $V(\phi) = |m\phi + \lambda\phi^2|^2$, while the scalar-fermion interactions take the usual Yukawa form $\phi\psi\psi$. Crucially, the function $W(\phi)$ should be *holomorphic*: it depends only on ϕ and not on $\phi^\dagger$. This fact will take on increasing significance as these lectures progress, but for now we will just take this as given.

Even without doing any detailed calculations, we can see that there's something curious about the action (1.1): the boson ϕ and the fermion ψ have the same mass $|m|$. Usually in quantum field theory, we shouldn't ascribe too much meaning to such an observation since masses receive quantum corrections and there's no guarantee that the physical masses of two distinct particles will coincide just because the masses in the Lagrangian are equal. However, for the particular action (1.1), it turns out that the equality of bosonic and fermionic masses persists in the full quantum theory. This arises because the action enjoys a rather surprising symmetry, with the infinitesimal variation given by

$$\delta\phi = \sqrt{2}\epsilon\psi \quad \text{and} \quad \delta\psi = \sqrt{2}i\sigma^\mu\bar{\epsilon}\,\partial_\mu\phi - \sqrt{2}\epsilon\frac{\partial W^\dagger}{\partial\phi^\dagger} \tag{1.2}$$

This is our first example of *supersymmetry*. It is a symmetry that relates the bosonic field ϕ with the fermionic field ψ. Because ψ is a Grassmann field, while ϕ is not, the infinitesimal object ϵ, which parameterises the transformation, must also be a Grassmann-valued Weyl spinor.

You can't tell just by staring at the action (1.1) that it is invariant under the supersymmetry transformation (1.2). Instead, it takes a calculation, one that turns out to be a little bit of a headache. (Some balm for this headache will be offered in Section 3.2.3.)

The action (1.1) is the simplest supersymmetric theory in $d = 3 + 1$ dimensions. It is known as the *Wess-Zumino* model. The existence of such a symmetry opens up a number of questions. What, if anything, is the symmetry good for? Are there other theories that also exhibit such symmetry? What properties might they have? All of these will be answered as these lectures progress.

There is also another question that might have occurred to you: why is it such a pain to see that the action (1.1) is invariant under supersymmetry? Usually, the existence of symmetries in an action jumps out at you. Indeed, one of the main advantages of working with the Lagrangian approach, rather than the Hamiltonian approach, is that all symmetries are manifest. Typically you need do little more than ensure that various indices are contracted in the right way. This suggests that there may be a better way to write the action (1.1) that makes supersymmetry as obvious as any other symmetry. And there is. Our first task in these lectures – one that will carry us through much of Sections 2, 3 and 4 – is to better understand the structure behind supersymmetry and the corresponding supersymmetric actions.

2 The Supersymmetry Algebra

The purpose of this section is to describe, in mathematical terms, what supersymmetry actually is. Usually in physics, we think of symmetries as associated to groups. But, at least for continuous symmetries, these groups have an underlying algebra and often that contains all the information that we need. So it is with supersymmetry. We will describe the algebra that underlies supersymmetry and start to explore some of its representations.

I should warn you that this section will be a little dry in flavour. There will be few fields and certainly no dynamics. These will come in later sections. But this section lays the necessary groundwork for the stories that are to come.

2.1 The Lorentz Group

Minkowski space $\mathbb{R}^{1,3}$ is the stage for relativistic quantum field theory. This space comes equipped with the Minkowski metric

$$\eta_{\mu\nu} = \text{diag}(+1, -1, -1, -1)$$

The set of symmetries of Minkowski space include Lorentz transformations of the form $x^\mu \to \Lambda^\mu{}_\nu x^\nu$ where

$$\Lambda^T \eta \Lambda = \eta$$

Embedded among these are a couple of discrete transformations: parity with $\Lambda = \text{diag}(1, -1, -1, -1)$ and time reversal with $\Lambda = \text{diag}(-1, 1, 1, 1)$. The transformations that are continuously connected to the identity have $\det\Lambda = 1$ and $\Lambda^0{}_0 > 0$ and form the *Lorentz group* $SO(1,3)$. (The restriction to $\Lambda^0{}_0 > 0$ is sometimes written as $SO^+(1,3)$.)

Our main goal in this section is the spell out some properties of the spinor representations of the Lorentz group. In fact, strictly speaking the group $SO(1,3)$ doesn't have any spinor representations. However, there is a closely related group called $\text{Spin}(1,3)$ that does admit spinors. This is the double cover, in the sense that

$$SO(1,3) \cong \text{Spin}(1,3)/\mathbb{Z}_2$$

where that $\mathbb{Z}_2$ is the famous minus sign that spinors pick up under a 2π rotation, a minus sign that vectors like x^μ are oblivious to. The fact that there are spinors in our world is the statement that the true symmetry group is $\text{Spin}(1,3)$ rather than $SO(1,3)$.

When we introduced spinors in the Quantum Field Theory course, we did so by first looking at the algebra $so(1,3)$ that is shared by both groups Spin$(1,3)$ and $SO(1,3)$. A Lorentz transformation acting on a 4-vector can be written as

$$\Lambda = \exp\left(-\frac{i}{2}\omega_{\mu\nu}M^{\mu\nu}\right) \tag{2.1}$$

where $\omega_{\mu\nu}$ are six numbers that specify what Lorentz transformation we're doing, while $M^{\mu\nu} = -M^{\nu\mu}$ are a choice of six 4×4 anti-symmetric matrices that generate the different Lorentz transformations. The matrix indices are suppressed in the above expressions; in their full glory we would write $(M^{\mu\nu})^{\rho}{}_{\sigma}$. So, for example

$$(M^{01})^{\rho}{}_{\sigma} = i\begin{pmatrix} 0&1&0&0\\ 1&0&0&0\\ 0&0&0&0\\ 0&0&0&0\end{pmatrix} \quad \text{and} \quad (M^{12})^{\rho}{}_{\sigma} = i\begin{pmatrix} 0&0&0&0\\ 0&0&-1&0\\ 0&1&0&0\\ 0&0&0&0\end{pmatrix} \tag{2.2}$$

(Note that the generators differ by a factor of i from those defined in the Quantum Field Theory lectures. This is compensated by an extra factor of i in the exponent (2.1).) The matrices generate the algebra $so(1,3)$,

$$[M^{\mu\nu}, M^{\rho\sigma}] = i\left(\eta^{\nu\rho}M^{\mu\sigma} - \eta^{\nu\sigma}M^{\mu\rho} + \eta^{\mu\sigma}M^{\nu\rho} - \eta^{\mu\rho}M^{\nu\sigma}\right) \tag{2.3}$$

In the lectures on Quantum Field Theory, we then constructed the spinor representations by first looking at the Clifford algebra of gamma matrices, $\{\gamma^{\mu},\gamma^{\nu}\} = 2\eta^{\mu\nu}$ and, from these, constructing a new representation of the Lorentz algebra (2.3). Here, we'll take a slightly different path. It will be useful to first extract a little more information from the algebra (2.3).

The six different Lorentz transformations naturally decompose into three rotations J_i and three boosts K_i, defined by

$$J_i = \frac{1}{2}\epsilon_{ijk}M_{jk} \quad \text{and} \quad K_i = M_{0i}$$

where these $j,k = 1,2,3$ indices are summed over, and $\epsilon_{123} = +1$. The rotation matrices are Hermitian, with $J_i^{\dagger} = J_i$ while the boost matrices are anti-Hermitian with $K_i^{\dagger} = -K_i$. This ensures that the rotations in (2.1) give rise to a compact group while the boosts are non-compact. From the Lorentz algebra, we find that these generators obey

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad , \quad [J_i, K_j] = i\epsilon_{ijk}K_k \quad , \quad [K_i, K_j] = -i\epsilon_{ijk}J_k$$

The rotations form an $su(2)$ sub-algebra. That, of course, is to be expected and is related to the fact that $SO(3) \cong SU(2)/\mathbb{Z}_2$.

We can, however, find two mutually commuting $su(2)$ algebras sitting inside $so(1,3)$. For this we take the linear combinations

$$A_i = \frac{1}{2}(J_i + iK_i) \quad \text{and} \quad B_i = \frac{1}{2}(J_i - iK_i)$$

Both of these are Hermitian, with $A_i^\dagger = A_i$ and $B_i^\dagger = B_i$. They obey

$$[A_i, A_j] = i\epsilon_{ijk}A_k \quad , \quad [B_i, B_j] = i\epsilon_{ijk}B_k \quad , \quad [A_i, B_j] = 0 \tag{2.4}$$

But we know about representations of $SU(2)$: they are labelled by an integer or half-integer $j \in \frac{1}{2}\mathbb{Z}$ which, in the context of rotations, we call "spin". The dimension of the representation is then $2j+1$. The fact that we can find two $su(2)$ sub-algebras of the Lorentz algebra tells us that all representations must carry two such labels

$$(j_1, j_2) \quad \text{with} \quad j_1, j_2 \in \frac{1}{2}\mathbb{Z} \tag{2.5}$$

and has dimension $(2j_1+1)(2j_2+1)$. We'll flesh out the meaning of these representations more below. But for now, we can identify the simplest such representations just by counting: we have

$$\begin{aligned}
(0,0): &\quad \text{scalar} \\
(\tfrac{1}{2},0): &\quad \text{left-handed Weyl spinor} \\
(0,\tfrac{1}{2}): &\quad \text{right-handed Weyl spinor} \\
(\tfrac{1}{2},\tfrac{1}{2}): &\quad \text{vector} \\
(1,0): &\quad \text{self-dual 2-form} \\
(0,1): &\quad \text{anti-self-dual 2-form}
\end{aligned}$$

We see that the smallest representations of the Lorentz group are the left- and right-handed Weyl spinors. What we call the physical spin of a particle is the quantum number under rotations $\vec{J}$: this is $j = j_1 + j_2$.

There's something a little odd about the our discovery of two $su(2)$ sub-algebras. After all, it certainly isn't true that the Lorentz group is isomorphic to two copies of $SU(2)$. This is because $SU(2)$ is a compact group: keep doing a rotation and you will eventually get back to where you started. Indeed, two copies of the group $SU(2)$ give rotation group of Euclidean space $\mathbb{R}^4$.

$$\text{Spin}(4) \cong SU(2) \times SU(2) \quad \text{with} \quad SO(4) \cong \text{Spin}(4)/Z_2$$

In contrast, the Lorentz group is non-compact: keep boosting and you get further and further from where you started. How does this manifest itself in the two $su(2)$ algebras that we've found in (2.4)?

The answer is a little subtle and is to be found in the reality properties of the generators A_i and B_i. Recall that all integer, $j \in \mathbb{Z}$, representations of $SU(2)$ are real, while all half-integer spin, $j \in \mathbb{Z} + \frac{1}{2}$, are pseudoreal (which means that, while not actually real, the representation is isomorphic to its complex conjugate). However, the A_i and B_i in (2.4) do *not* have these properties. You can see in (2.2) that both J_i and K_i are pure imaginary. This, in turn, means that the generators A_i and B_i are complex conjugates of each other

$$(A_i)^\star = -B_i$$

This is where the difference lies that distinguishes $SO(4)$ from $SO(1,3)$. The Lie algebra $so(1,3)$ does not contain two, mutually commuting copies of the real Lie algebra $su(2)$, but only after a suitable complexification. This means that certain complex linear combinations of the Lie algebra $su(2) \times su(2)$ are isomorphic to $so(1,3)$. To highight this, the relationship between the two is sometimes written as

$$so(1,3) \cong su(2) \times su(2)^\star$$

For our purposes, it means that the complex conjugate of a representation (j_1, j_2) exchanges the two quantum numbers

$$(j_1, j_2)^\star = (j_2, j_1)$$

Both the scalar representation $(0,0)$ and the vector representation $(\frac{1}{2}, \frac{1}{2})$ are real, while the left- and right-handed Weyl spinors $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ are exchanged under complex conjugation. This last statement will be important as we proceed. In the context of quantum field theory, if a field appears in a theory then so too does its complex conjugate. This means that if you have a left-handed spinor, you also have a right-handed complex conjugated spinor.

2.1.1 Spinors and $SL(2,\mathbb{C})$

There is another way to discover spinors, this time one that doesn't involve going through the algebra. We will use the fact that there is an isomorphism between two groups

$$\text{Spin}(1,3) \cong SL(2,\mathbb{C}) \tag{2.6}$$

To see this, we first note that we can write a point x^μ in Minkowski space as a 2×2 Hermitian matrix,

$$X = x_\mu \sigma^\mu = \begin{pmatrix} x_0 + x_3 & x_1 - ix_2 \\ x_1 + ix_2 & x_0 - x_3 \end{pmatrix}$$

where we've introduced the 4-vector of 2×2 matrices,

$$\sigma^\mu = (1, \sigma^i) \quad \text{with} \quad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.7}$$

The σ^i are, of course, the Pauli matrices. The matrix X is Hermitian: $X = X^\dagger$. Moreover, there is clearly a one-to-one map between 4-vectors x^μ and 2×2 Hermitian matrices. The Minkowski inner product is particularly natural in this language: it is

$$\det X = (x_0)^2 - (x_1)^2 - (x_2)^2 - (x_3)^2 = x_\mu x^\mu$$

Now consider an $SL(2, \mathbb{C})$ transformation that acts as

$$X \;\rightarrow\; X' = SXS^\dagger \tag{2.8}$$

with $S \in SL(2, \mathbb{C})$. We have $(X')^\dagger = X'$ and $\det X' = \det X$ since $\det S = 1$. This means that the map (2.8) must be a Lorentz transformation.

In fact, it is not hard to see that we can implement all Lorentz transformations this way and we'll give an explicit construction of the generators shortly. For now, we can just do some simple counting. A general complex 2×2 matrix has 4 complex entries. The requirement that its determinant is 1 reduces this to 3 complex parameters, or 6 real parameters. This agrees with the dimension of the Lorentz group: $6 = 3$ rotations $+$ 3 boosts. Moreover, the $SL(2, \mathbb{C})$ transformation $S = -1$ does not act on X, which is the reason why $SL(2, \mathbb{C})$ coincides with the double cover (2.6).

It is clear that the fundamental representation of $SL(2, \mathbb{C})$ is not a 2×2 matrix: it is a 2-component, complex object $\psi_\alpha = (\psi_1, \psi_2)$ that transforms as

$$\psi_\alpha \;\rightarrow\; S_\alpha{}^\beta \psi_\beta \qquad \alpha, \beta = 1, 2$$

Clearly it is a complex two-dimensional representation. In terms of our previous classification (2.5), we take it to correspond to $(\frac{1}{2}, 0)$: it is what we call a *left-handed Weyl spinor*.

Given any complex representation of a Lie group, we can always form another representation by taking the conjugate. This is equivalent to the original if we can find a matrix C for which $S^\star = CSC^{-1}$. In the present case, no such C exists and the matrix S and its conjugate $S^\star$ are inequivalent representations. We denote the complex conjugate as

$$(\psi_\alpha)^\dagger = \bar{\psi}_{\dot{\alpha}}$$

We've adopted two notational flourishes to distinguish the two representations. First, we use different indices $\alpha, \beta = 1, 2$ and $\dot{\alpha}, \dot{\beta} = 1, 2$ for the two different representations. This is useful because the two indices are telling us that the objects transform in different ways. In addition, we also add a bar over any object, like $\bar{\psi}$, that transforms in the conjugate representation. This allows us to identify these objects even when we suppress the indices. (Note that a bar on a Weyl spinor simply means complex conjugation while, as we learned in the Quantum Field Theory lectures, a bar on a Dirac spinor means complex transpose together with multiplication by γ^0.) The complex conjugate spinor then transforms as

$$\bar{\psi}_{\dot{\alpha}} \;\rightarrow\; (S^\star)_{\dot{\alpha}}{}^{\dot{\beta}} \bar{\psi}_{\dot{\beta}} \qquad \dot{\alpha}, \dot{\beta} = 1, 2$$

In our previous classification (2.5) it is the representation $(0, \frac{1}{2})$. It is a *right-handed Weyl spinor*.

Some of the index conventions above (and below) differ from what you may have seen in other contexts and it's worth quickly explaining why. Suppose that we've got a vector u that transforms in the fundamental of $SU(N)$. We write the components as u_a with $a = 1, \ldots, N$. The vector $u^\dagger$ transforms in the conjugate representation and we would write these components as $(u^\dagger)^a$, with the index raised and no dots in sight. This reflects the fact that we can contract $u^\dagger$ and u to form a singlet: $(u^\dagger)^a u_a$. However, the representations of $SL(2, \mathbb{C})$ have a different structure and, as we'll see shortly, you can't contract a spinor and its conjugate to get a singlet. That's why we introduce the strange looking dotted indices, rather than raising the index, to distinguish the conjugate representation

Building Scalars from Spinors

The group $SL(2, \mathbb{C})$ has the following invariant tensors

$$\epsilon^{\alpha\beta} = \epsilon^{\dot{\alpha}\dot{\beta}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad \epsilon_{\alpha\beta} = \epsilon_{\dot{\alpha}\dot{\beta}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Note that the $\epsilon_{\alpha\beta}$ with indices lowered differs by a minus sign from $\epsilon^{\alpha\beta}$. This ensures that one is the inverse of the other: $\epsilon^{\alpha\beta}\epsilon_{\beta\gamma} = \delta^\alpha_\gamma$. This, in turn, means that when we use epsilon symbols to raise and lower indices (as we will below) then if we choose to raise an index and subsequently lower it again then we don't get a minus sign for our troubles.

Given, say, two left-handed Weyl fermions ψ and χ, we can use the epsilon tensors to form invariants. We define

$$\psi\chi := \epsilon^{\alpha\beta}\psi_\beta\chi_\alpha = \psi_2\chi_1 - \psi_1\chi_2$$

To see that these are, indeed, invariants under $SL(2,\mathbb{C})$, we just need to perform a transformation

$$\psi\chi \;\to\; S_\alpha{}^\gamma S_\beta{}^\delta \epsilon^{\alpha\beta}\psi_\delta\chi_\gamma = (\det S)\epsilon^{\gamma\delta}\psi_\delta\chi_\gamma = \psi\chi \tag{2.9}$$

where, in the first equality we've used the fact that $S_\alpha{}^\gamma S_\beta{}^\delta\epsilon^{\alpha\beta} = \det S\,\epsilon^{\gamma\delta}$, which you can confirm simply by checking all the cases $\gamma,\delta = 1,2$. In the second equality we've used the fact that $\det S = 1$.

In some ways, the ϵ symbols play a role for spinors that is akin to role played by the metric $\eta^{\mu\nu}$ for vectors. Of course, one key difference is that $\epsilon^{\alpha\beta}$ is anti-symmetric, but this tallies nicely with the fact that, in quantum field theory, spinors are anti-commuting Grassmann variables. We then have

$$\psi\chi = \psi_2\chi_1 - \psi_1\chi_2 = -\chi_1\psi_2 + \chi_2\psi_1 = \chi\psi$$

In particular, $\psi\psi = 2\psi_2\psi_1$ is non-vanishing.

We can do something similar for right-handed fermions. However, a fiddly minus sign rears its head. We define

$$\bar\psi\bar\chi := \epsilon^{\dot\alpha\dot\beta}\bar\psi_{\dot\alpha}\bar\chi_{\dot\beta} = \bar\psi_1\bar\chi_2 - \bar\psi_2\bar\chi_1 \tag{2.10}$$

With anti-commuting spinors, we again have $\bar\psi\bar\chi = \bar\chi\bar\psi$. Note that the ordering of the indices in (2.10) differs from (2.9). The reason for choosing this different ordering, resulting in a minus sign difference in the definitions, is that it ensures that $(\psi\chi)^\dagger = \bar\psi\bar\chi$, since

$$(\psi\chi)^\dagger = (\psi_2\chi_1 - \psi_1\chi_2)^\dagger = \bar\chi_1\bar\psi_2 - \bar\chi_2\bar\psi_1 = \bar\psi\bar\chi$$

We can use the ϵ symbols to raise and lower spinor indices, just as we use the Minkowski metric to raise and lower vector indices. We have

$$\psi^\alpha = \epsilon^{\alpha\beta}\psi_\beta \quad , \quad \psi_\alpha = \epsilon_{\alpha\beta}\psi^\beta \quad \text{and} \quad \bar\psi^{\dot\alpha} = \epsilon^{\dot\alpha\dot\beta}\bar\psi_{\dot\beta} \quad , \quad \bar\psi_{\dot\alpha} = \epsilon_{\dot\alpha\dot\beta}\bar\psi^{\dot\beta}$$

In this notation, the Lorentz scalars (2.10) become

$$\psi\chi = \psi^\alpha\chi_\alpha \quad \text{and} \quad \bar\psi\bar\chi = \bar\psi_{\dot\alpha}\bar\chi^{\dot\alpha}$$

Our fiddly minus sign difference between (2.9) and (2.10) has now transmuted into the following rule: for left-handed spinors we should contract (undotted) indices in the direction $\searrow$, while for right-handed spinors we should contract (dotted) indices in the direction $\nearrow$.

We can ask how these new objects ψ^α and $\bar{\psi}^{\dot{\alpha}}$ fare under Lorentz transformations. We have

$$\begin{aligned} \psi^\alpha &\rightarrow \epsilon^{\alpha\beta} S_\beta{}^\gamma \psi_\gamma = (S^{-1\,T})^\alpha{}_\beta \psi^\beta \\ \bar{\psi}^{\dot{\alpha}} &\rightarrow \epsilon^{\dot{\alpha}\dot{\beta}} (S^\star)_{\dot{\beta}}{}^{\dot{\gamma}} \bar{\psi}_{\dot{\gamma}} = (S^{-1\,\dagger})^{\dot{\alpha}}{}_{\dot{\beta}} \bar{\psi}^{\dot{\beta}} \end{aligned} \tag{2.11}$$

where the equality follows from the following algebra

$$S_\alpha{}^\gamma \epsilon^{\alpha\beta} S_\beta{}^\delta = \epsilon^{\gamma\delta} \quad \Rightarrow \quad (S^T)^\gamma{}_\alpha \epsilon^{\alpha\beta} S_\beta{}^\delta = \epsilon^{\gamma\delta} \quad \Rightarrow \quad \epsilon^{\alpha\beta} S_\beta{}^\delta = (S^{-1\,T})^\alpha{}_\gamma \epsilon^{\gamma\delta}$$

with similar manipulations for the right-handed spinor. The matrices $S^{-1\,T}$ don't form a new representation of $SL(2,\mathbb{C})$; they are equivalent to the fundamental representation since, from above, we have $\epsilon S \epsilon^{-1} = S^{-1\,T}$. This means that the covariant and contravariant left-handed spinors ψ_α and ψ^α transform in equivalent representations. Similarly, the right- handed spinors $\bar{\psi}_{\dot{\alpha}}$ and $\bar{\psi}^{\dot{\alpha}}$ transform in equivalent representations.

Building Vectors from Spinors

A key take-away from our discussion above is that if you want to form a Lorentz scalar then you need a pair of left-handed fermions or a pair of right handed fermions. Suppose that we instead have one object of each type, say a left-handed spinor ψ_α and a right-handed spinor $\bar{\chi}_{\dot{\alpha}}$. What kind of object can we then build? The answer is clear from the quantum numbers of these representations:

$$(\tfrac{1}{2}, 0) \otimes (0, \tfrac{1}{2}) = (\tfrac{1}{2}, \tfrac{1}{2})$$

This is the vector representation of the Poincaré group.

To explicitly construct the vector, we sandwich the Pauli matrices

$$(\sigma^\mu)_{\alpha\dot{\alpha}} = (1, \sigma^i)_{\alpha\dot{\alpha}}$$

between two spinors. We write

$$\psi \sigma^\mu \bar{\chi} = \psi^\alpha (\sigma^\mu)_{\alpha\dot{\alpha}} \bar{\chi}^{\dot{\alpha}}$$

Note that, as shown above, the Pauli matrices σ^μ should come with an index of each type – one undotted, and one dotted – and both subscripts. Taking the conjugate, we have $(\psi \sigma^\mu \bar{\chi})^\dagger = \chi \sigma^\mu \bar{\psi}$.

To see that the object does indeed transform as a 4-vector, we can contract this with any other 4- vector x^μ to give $\psi X\bar{\chi}$ with $X = x_\mu\sigma^\mu$. But we know from (2.8) and (2.11) how each of these transforms: we then have

$$\psi X\bar{\chi} = \psi^\alpha X_{\alpha\dot{\alpha}}\bar{\chi}^{\dot{\alpha}} \;\rightarrow\; (\psi^\beta (S^{-1})_\beta{}^\alpha)(S_\alpha{}^\delta X_{\delta\dot{\delta}} S^\star_{\dot{\alpha}}{}^{\dot{\delta}})(\bar{\chi}^{\dot{\beta}}(S^{\star\,-1})_{\dot{\beta}}{}^{\dot{\alpha}}) = \psi X\bar{\chi}$$

The fact that $\psi X\bar{\chi}$ forms a singlet shows that $\psi\sigma^\mu\bar{\chi}$ must transform as a vector. In fancy maths words, we say that the Pauli matrices act as the intertwiner between the different representations.

We can use the epsilon symbols to raise the spinor indices on the Pauli matrices $\sigma^\mu_{\alpha\dot{\alpha}}$. This gives us a closely related set of matrices that we denote

$$(\bar{\sigma}^\mu)^{\dot{\alpha}\alpha} = \epsilon^{\alpha\beta}\epsilon^{\dot{\alpha}\dot{\beta}}\sigma^\mu_{\beta\dot{\beta}}$$

The bar on $\bar{\sigma}$ doesn't denote anything to do with complex conjugation. The $\bar{\sigma}^\mu$ are simply a different set of 2×2 matrices from σ^μ. Note that the indices have not only been raised, but also switched: σ^μ has the undotted index first, while $\bar{\sigma}^\mu$ has the dotted index first. If we define $\epsilon = i\sigma^2$ then, viewed as matrix multiplication, we have $\bar{\sigma} = \epsilon\sigma^T\epsilon^T$. A quick calculation shows that

$$(\bar{\sigma}^\mu)^{\dot{\alpha}\alpha} = (1, -\sigma^i)^{\dot{\alpha}\alpha}$$

We can then similarly construct the vector

$$\bar{\chi}\bar{\sigma}^\mu\psi = \bar{\chi}_{\dot{\alpha}}(\bar{\sigma}^\mu)^{\dot{\alpha}\alpha}\psi_\alpha$$

This isn't a new object: you can check that $\psi\sigma^\mu\bar{\chi} = -\bar{\chi}\bar{\sigma}^\mu\psi$.

Generators of $SL(2,\mathbb{C})$

Finally we can give a description of the generators of $SL(2,\mathbb{C})$. We define the anti-symmetrised product of sigma matrices,

$$(\sigma^{\mu\nu})_\alpha{}^\beta = \frac{i}{4}\left(\sigma^\mu\bar{\sigma}^\nu - \sigma^\nu\bar{\sigma}^\mu\right)_\alpha{}^\beta$$

These are linearly independent and so can be taken as a generators of $SL(2,\mathbb{C})$. Because of the anti-symmetry in μ and ν, there are six such generators which is the dimension of the Lorentz group. Indeed, we can see explicitly that these generate the Lorentz group by computing the commutator

$$[\sigma^{\mu\nu}, \sigma^{\rho\sigma}] = i\left(\eta^{\nu\rho}\sigma^{\mu\sigma} - \eta^{\nu\sigma}\sigma^{\mu\rho} + \eta^{\mu\sigma}\sigma^{\nu\rho} - \eta^{\mu\rho}\sigma^{\nu\sigma}\right)$$

This reproduces the algebra of the Lorentz group (2.3) as promised. A left-handed spinor then transforms as

$$\psi_\alpha \;\to\; \exp\left(-\frac{i}{2}\omega_{\mu\nu}\sigma^{\mu\nu}\right)_\alpha^{\ \beta}\psi_\beta \tag{2.12}$$

where $\omega_{\mu\nu}$ are the same set of six numbers that specify the Lorentz transformation (2.1).

The conjugate representation is generated by

$$(\bar\sigma^{\mu\nu})^{\dot\alpha}_{\ \dot\beta} = \frac{i}{4}\left(\bar\sigma^\mu\sigma^\nu - \bar\sigma^\nu\sigma^\mu\right)^{\dot\alpha}_{\ \dot\beta}$$

These too satisfy the algebra of the Lorentz group. Correspondingly, a right-handed spinor transforms as

$$\bar\psi^{\dot\alpha} \;\to\; \exp\left(-\frac{i}{2}\omega_{\mu\nu}\bar\sigma^{\mu\nu}\right)^{\dot\alpha}_{\ \dot\beta}\bar\psi^{\dot\beta} \tag{2.13}$$

Note that, from the positioning of the indices of $\bar\sigma^{\mu\nu}$, these act naturally as generators on $\bar\psi^{\dot\alpha}$, with the index raised.

2.1.2 Lagrangians for Spinors

We can now describe how to construct Lagrangians from a Weyl spinor. Suppose that we have just a single left-handed Weyl spinor ψ to play with. This necessarily comes with its conjugate, a right-handed spinor $\bar\psi = \psi^\dagger$. We can then form a kinetic term

$$\mathcal{S}_{\rm Weyl} = -\int d^4x\; i\bar\psi\bar\sigma^\mu\partial_\mu\psi \tag{2.14}$$

Upon quantisation, this theory gives a single massless, left-handed fermion of helicity $-\frac{1}{2}$ and massless right-handed anti-particle of helicity of $+\frac{1}{2}$. The theory has a global $U(1)$ symmetry under which $\psi \to e^{i\alpha}\psi$; if the left-handed fermion has charge $+1$ then the right-handed fermion has charge -1, as befits an anti-particle.

We can add a mass term for a single Weyl fermion. This is known as a *Majorana mass*,

$$\mathcal{S}_{\rm Maj} = \int d^4x\; \frac{m}{2}\psi\psi + \frac{m^\star}{2}\bar\psi\bar\psi \tag{2.15}$$

In general, we can take $m \in \mathbb{C}$ although any complex phase of m can be absorbed into ψ and, upon quantisation, the resulting particle has mass $|m|$. Importantly, the Majorana mass explicitly breaks the global $U(1)$ symmetry, so there is no quantum number to distinguish particle from anti-particle. Upon quantisation, the theory consists of a single massive spin $\frac{1}{2}$ particle that is now its own anti-particle.

Because the Majorana mass term explicitly breaks the $U(1)$ symmetry, it is not allowed if the $U(1)$ is gauged. Relatedly, it's not possible to write down such a term for any fermion ψ that transforms in a complex representation of a gauge group. It is, however, possible to write down such terms for fermions in real representations.

Recovering Dirac Spinors

All this discussion of spinors and, so far, not a gamma matrix or Clifford algebra in sight! Yet these played a central role in the discussion of spinors that we met in the Quantum Field Theory lectures. What's going on?

The Dirac spinor is *not* an irreducible representation of the Lorentz group in $d = 3+1$ dimensions. Instead, it consists of independent left- and right-handed spinors. In our earlier notation:

$$(\tfrac{1}{2}, 0) \oplus (0, \tfrac{1}{2}) : \quad \text{Dirac spinor}$$

We write a Dirac spinor as a 4-component object, consisting of a left-handed Weyl fermion ψ_α and a right-handed Weyl fermion $\bar{\chi}^{\dot{\alpha}}$ (note the index up),

$$\Psi = \begin{pmatrix} \psi_\alpha \\ \bar{\chi}^{\dot{\alpha}} \end{pmatrix}$$

We also introduce the chiral basis of gamma matrices

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix} \tag{2.16}$$

These obey the Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$. In the Quantum Field Theory lectures, we showed that the generators of Lorentz transformations for a Dirac spinor are

$$S^{\mu\nu} = \frac{i}{4}[\gamma^\mu, \gamma^\nu] = \begin{pmatrix} \sigma^{\mu\nu} & 0 \\ 0 & \bar{\sigma}^{\mu\nu} \end{pmatrix}$$

(As with our earlier definition of $M^{\mu\nu}$, this differs by a factor of i from the conventions in the Quantum Field Theory lectures.) Under a Lorentz transformation, a Dirac spinor transforms as $\Psi \to \exp(-\frac{i}{2}\omega_{\mu\nu}S^{\mu\nu})\Psi$. This reproduces the transformations of Weyl spinors that we saw in (2.12) and (2.13).

The Dirac action that we met in our Quantum Field Theory lectures is

$$S_{\rm Dirac} = -\int d^4x \;\; i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - M\bar{\Psi}\Psi$$

where, for a Dirac spinor (but not a Weyl spinor!) the bar notation means $\bar{\Psi} = \Psi^\dagger\gamma^0$. Decomposed in terms of Weyl fermions, it becomes

$$S_{\rm Dirac} = -\int d^4x \;\; i\bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi + i\chi\sigma^\mu\partial_\mu\bar{\chi} - M(\chi\psi + \bar{\psi}\bar{\chi}) \tag{2.17}$$

The first term coincides with the kinetic term (2.14) for a left-handed fermion. The second term is simply a different way of writing this, with the derivative now acting on a right- handed fermion; if you play around lowering and raising indices then the second term can be massaged to look like the first.

The mass term in (2.17) is *not* of the Majorana type (2.15). First, the mass is necessarily real, $M \in \mathbb{R}$, although it can be positive or negative. Second, because the mass term involves two distinct Weyl fermions it preserves a $U(1)$ symmetry, under which the phase of ψ and χ rotate oppositely. The result is that, upon quantisation, the action (2.17) gives a particle of spin $+\frac{1}{2}$ and charge $+1$, together with a distinct anti-particle of spin $+\frac{1}{2}$ and charge -1, both with mass $|M|$.

It is possible to restrict the Dirac fermion Ψ to have the same content as a single Weyl fermion. In a general basis of gamma matrices, we do this by introducing a charge conjugation matrix. But in the chiral basis (2.16), it's particularly simple: we just restrict $\bar{\chi} = \bar{\psi} \equiv \psi^\dagger$. A Dirac spinor with such a restriction is called a *Majorana spinor*.

Throughout these lectures, we will have no need to resort to 4-component spinors. We will write everything in terms of 2-component Weyl fermions.

2.1.3 The Poincaré Group and its Extensions

The continuous symmetries of Minkowski space comprise of Lorentz transformations together with spacetime translations. Combined, these form the *Poincaré group*. Spacetime translations are generated, as usual, by the momentum 4-vector P^μ. Their commutation relations with themselves and with the Lorentz generators $M^{\mu\nu}$ are given by

$$[P^\mu, P^\nu] = 0 \quad \text{and} \quad [M^{\mu\nu}, P^\sigma] = i\left(P^\mu\eta^{\nu\sigma} - P^\nu\eta^{\mu\sigma}\right) \tag{2.18}$$

The latter of these is equivalent to the statement that P^μ transforms as a 4-vector under Lorentz transformations. These commutation relations should be considered in conjunction with the Lorentz algebra (2.3),

$$[M^{\mu\nu}, M^{\rho\sigma}] = i\left(\eta^{\nu\rho}M^{\mu\sigma} - \eta^{\nu\sigma}M^{\mu\rho} + \eta^{\mu\sigma}M^{\nu\rho} - \eta^{\mu\rho}M^{\nu\sigma}\right) \tag{2.19}$$

Together, (2.18) and (2.19) form the algebra of the Poincaré group.

It's not unusual for quantum field theories to exhibit further continuous symmetries. Say, a global $U(1)$ symmetry that rotates the phase of a complex field, or perhaps a non-Abelian $SU(N)$ symmetry under which a multiplet of fields transforms. The generators of these symmetries – which we'll denote collectively as T – correspond to some conserved charge or isospin and are always Lorentz scalars. This means that they necessarily commute with the Poincaré generators,

$$[P^\mu, T] = [M^{\mu\nu}, T] = 0$$

One could ask: is it possible for something less trivial to happen, with the new generators transforming in some interesting fashion under the Poincaré group? For example, this would happen if the additional generators T themselves carried some spacetime index. If this were possilble, the Poincaré group would be subsumed into a larger group. And that sounds interesting.

A theorem due to Coleman and Mandula greatly restricts this possibility. Roughly speaking, the theorem states that, in any spacetime dimension greater than $d = 1 + 1$, the symmetry group of any interacting quantum field theory must factorise as

$$\text{Poincaré} \times \text{Internal} \tag{2.20}$$

We won't prove the Coleman-Mandula theorem here[1]. The gist of the proof is that Poincaré invariance already greatly restricts what can happen in, say, 2 to 2 scattering, with only the scattering angle left undetermined. Any internal symmetries that factorise, as in (2.20), put restrictions on the kinds of interactions that are allowed, for example enforcing conservation of electric charge. But if the generators T were to carry a spacetime index then they would put further constraints on the scattering angle itself and that would be overly restrictive, at best allowing scattering to occur only at discrete angles. But if one assumes that the scattering amplitudes are analytic functions of the angle then the amplitude must vanish for all angles and the theory is free.

[1]The original Coleman-Mandula paper is from 1967 and entitled "All Possible Symmetries of the S-matrix". Witten's "Introduction to Supersymmetry" lectures give a clear intuitive explanation of the theorem. A full proof can be found Weinberg vol III.

Like all no-go theorems in physics, the Coleman-Mandula theorem comes with a number of underlying assumptions. Some of these are eminently reasonable, such as locality and causality. But it may be possible to relax other assumptions to find interesting loopholes to the Coleman-Mandula theorem. Two such loopholes have proven to be extremely important.

- **Conformal Invariance:** The Coleman-Mandula theorem assumes that the theory has a mass gap, meaning that all particles are massive. Indeed, it studies symmetries of the S-matrix which is really only well defined for massive particles where we don't have to worry about IR divergences. For theories of massless particles something interesting can, and often does, happen.

  The first interesting thing is that interacting massless theories typically exhibit scale invariance. This means that physics is unchanged under the symmetry $x^\mu \to \lambda x^\mu$. The associated symmetry generator is called D for "dilatation". This can only be a symmetry of a theory that has no dimensionful parameters. In particular, no masses.

  The second interesting thing is more surprising. For reasons that are not entirely understood, theories that exhibit scale invariance also exhibit a further symmetry known as *special conformal transformations* of the form

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

  This transformation depends on a vector parameter a^μ and the associated generator is a 4-vector K^μ. The resulting conformal algebra extends the Poincaré algebra (2.18) and (2.19) with the non-trivial commutators

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

  Interacting conformal field theories crop up in many places in physics. In their Euclidean incarnation, they describe critical points, or second order phase transitions, that were the focus of our lectures on Statistical Field Theory. In $d = 1+1$ dimensions the conformal group has rather more structure and a detailed introduction can be found in the lectures on String Theory. We'll meet examples of supersymmetric conformal field theories later in Section 6.4 when we discuss the low-energy physics of certain gauge theories.

- **Supersymmetry:** The second loophole to the Coleman-Mandula theorem is supersymmetry. As you may by now have guessed, exploiting this loophole will be the topic of the rest of these lectures.

2.2 The Supersymmetry Algebra

Supersymmetry evades the Coleman-Mandula no-go theorem because it is a different kind of symmetry. In contrast to the symmetries discussed above, it is not characterised by a Lie algebra. Instead it is characterised by a mathematical structure known as a $\mathbf{Z}_2$- graded Lie algebra. For our purposes, this simply means that the algebra contains both commutation and anti-commutation relations.

A generalisation of the Coleman-Mandula theorem to graded Lie algebras was given by Haag, Lopuszanski and Sohnius. Roughly speaking, it says that the only possibility is supersymmetry. We will now, finally, explain what this means.

Supersymmetric theories have a new conserved charge that is a left-handed Weyl spinor Q_α, together with its right-handed counterpart $\bar{Q}_{\dot\alpha}$. This is known as the *supercharge*. It is possible to have multiple supercharges, a situation known as extended supersymmetry. We will discuss this in Section 2.4 and, for now, stick to just a single complex supercharge. This is known as $\mathcal{N} = 1$ supersymmetry.

At the heart of the supersymmetry algebra is the anti-commutation relation

$$\{Q_\alpha, \bar{Q}_{\dot\alpha}\} = 2\sigma^\mu_{\alpha\dot\alpha} P_\mu \tag{2.21}$$

It is no surprise that a spinor should have an anti-commutator. But the structure of this relation is interesting: it tells us that the supercharges should be viewed as the square-root of spacetime translations! Our goal in these lectures is to understand what, exactly, this means.

The full supersymmetry algebra comprises of commutation relations (2.18) and (2.19) of the Poincaré group, which remain unchanged, together with the (anti)-commutation relations of the supercharges. The first of these is

$$[M^{\mu\nu}, Q_\alpha] = (\sigma^{\mu\nu})_\alpha{}^\beta Q_\beta \quad \text{and} \quad [M^{\mu\nu}, \bar{Q}^{\dot\alpha}] = (\bar\sigma^{\mu\nu})^{\dot\alpha}{}_{\dot\beta} \bar{Q}^{\dot\beta} \tag{2.22}$$

This is simply the statement that the supercharges transform under a Lorentz transformation in the manner expected of operators that are Weyl fermions. To see this, first recall from (2.12) that any spinor like Q_α transforms as $Q_\alpha \to U_\alpha{}^\beta Q_\beta$ where $U = \exp(-\frac{i}{2}\omega_{\mu\nu}\sigma^{\mu\nu})$. But Q_α is also an operator acting on a Hilbert space and, viewed through this lens, we get a different expression for how it transforms. Any

state in the Hilbert space transforms as $|\phi\rangle \to V|\phi\rangle$ with $V = \exp(-\frac{i}{2}\omega_{\mu\nu}M^{\mu\nu})$. Here, $M^{\mu\nu}$ is the abstract generator of Lorentz transformations and its action on any state depends on the quantum number of that state. Correspondingly, operators $\mathcal{O}$ transform as $\mathcal{O} \to V\mathcal{O}V^\dagger$ since this ensures that the matrix elements $\langle\phi'|\mathcal{O}|\phi\rangle$ remains unchanged. Equating these two ways in which the supercharge transforms, we have $VQ_\alpha V^\dagger = (UQ)_\alpha$. The algebra (2.22) is the infinitesimal version of this transformation law.

The remaining commutation relations are somewhat less interesting, although no less important

$$[Q_\alpha, P^\mu] = \{Q_\alpha, Q_\beta\} = 0 \tag{2.23}$$

There are, however, reasons why these commutators take this boring form.

First, why do we necessarily have $[Q_\alpha, P^\mu] = 0$? Clearly the right-hand side should be something with α and μ indices so that the commutator is covariant under Lorentz transformations. But that leaves the option for $[Q_\alpha, P^\mu] = c(\sigma^\mu)_{\alpha\dot\alpha}\bar{Q}^{\dot\alpha}$ for some $c \in \mathbf{C}$. What forces us to have $c = 0$?

The answer to this lies in the Jacobi identity

$$[P^\mu, [P^\nu, Q_\alpha]] + [P^\nu, [Q_\alpha, P^\mu]] + [Q_\alpha, [P^\mu, P^\nu]] = 0$$

Clearly the last term vanishes, as $[P^\mu, P^\nu] = 0$. If we choose $[Q_\alpha, P^\mu] = c(\sigma^\mu)_{\alpha\dot\alpha}\bar{Q}^{\dot\alpha}$ and, correspondingly, $[\bar{Q}^{\dot\alpha}, P^\mu] = c^\star(\bar\sigma^\mu)^{\dot\alpha\beta}Q_\beta$ then the Jacobi identity becomes

$$-c\sigma^\nu_{\alpha\dot\alpha}[P^\mu, \bar{Q}^{\dot\alpha}] + c\sigma^\mu_{\alpha\dot\alpha}[P^\nu, \bar{Q}^{\dot\alpha}] = |c|^2(\sigma^\nu\bar\sigma^\mu - \sigma^\mu\bar\sigma^\nu)_\alpha{}^\beta Q_\beta = 0$$

This requires $c = 0$.

There is a similar reason for why we must have $\{Q_\alpha, Q_\beta\} = 0$. Once again, there is an alternative since if we just try to pair up indices then we might think that $\{Q_\alpha, Q^\beta\} = c'(\sigma^{\mu\nu})_\alpha{}^\beta M_{\mu\nu}$ would be acceptable for any $c' \in \mathbf{R}$. But if we take the commutator with P^ρ then, from the argument above, the left-hand-side must vanish which, because $[P^\rho, M^{\mu\nu}] \neq 0$, tells us that $c' = 0$.

(An aside: there's actually a subtlety in this last discussion. While it is true that $\{Q_\alpha, Q_\beta\} = 0$ when sandwiched between any finite energy states, some supersymmetric theories have multiple ground states and it turns out that $\{Q_\alpha, Q_\beta\}$ can be non-vanishing when evaluated on the infinite energy domain walls that interpolate between these ground states. This subtlety is interesting, at least if you care about domain walls, but somewhat beyond the scope of these lectures.)

2.2.1 R-Symmetry

We started this section by noting that all internal symmetries must commute with the spacetime symmetries of the Poincaré group. But must they also commute with the supercharge Q_α? The answer is: almost.

All internal symmetries must commute with Q_α with one exception: it may be that theories admit an internal $U(1)$ symmetry that acts as

$$Q_\alpha \to e^{-i\lambda} Q_\alpha \quad \text{and} \quad \bar{Q}_{\dot{\alpha}} \to e^{i\lambda} \bar{Q}_{\dot{\alpha}} \tag{2.24}$$

This $U(1)$ symmetry is known as an *R-symmetry* and is sometimes denoted $U(1)_R$. If we denote the generator as R then it has commutation relations

$$[R, Q_\alpha] = -Q_\alpha \quad \text{and} \quad [R, \bar{Q}_{\dot{\alpha}}] = +\bar{Q}_{\dot{\alpha}} \tag{2.25}$$

When we turn to theories of extended supersymmetry in Section 2.4, we'll see different R-symmetry groups arising. But for theories with $\mathcal{N} = 1$ symmetry we have only $U(1)_R$. Nonetheless, this will play an important role when we come to analyse the dynamics of supersymmetric theories in later sections. We'll see this, for example, in Section 3.3.

This, then, is the supersymmetry algebra: it comprises of the algebra of the Poincaré group (2.18) and (2.19), together with the algebra of the supercharges (2.21), (2.22) and (2.23) and, finally, the R-symmetry (2.25). The next question is: what can we do with it?

2.2.2 A Consequence: Energy is Positive

Even before we write down any field theories, we can derive one feature of supersymmetric theories from the algebra alone. This follows from the key algebraic relation (2.21),

$$\{Q_\alpha, \bar{Q}_{\dot{\alpha}}\} = 2\sigma^\mu_{\alpha\dot{\alpha}} P_\mu \tag{2.26}$$

If we compute the expectation of the left-hand side in any state $|\phi\rangle$ then we find that it is necessarily positive

$$\langle\phi|Q_\alpha \bar{Q}_{\dot{\alpha}} + \bar{Q}_{\dot{\alpha}} Q_\alpha|\phi\rangle = |(Q_\alpha)^\dagger|\phi\rangle|^2 + |Q_\alpha|\phi\rangle|^2 \geq 0 \tag{2.27}$$

The same must be true of the right-hand side

$$\sigma^\mu_{\alpha\dot{\alpha}} \langle\phi|P_\mu|\phi\rangle \geq 0$$

If we set $\alpha = \dot{\alpha}$ and sum over $\alpha = 1, 2$ then we make use of the fact that $\mathrm{tr}\,\sigma^0 = 2$ and $\mathrm{tr}\,\sigma^i = 0$. This then reduces to the statement that the energy of any state in a supersymmetric theory is necessarily positive

$$\langle\phi|P_0|\phi\rangle \geq 0$$

This is curious. Usually in physics, we don't care about the overall value of the energy: if you add an overall constant to all energies, then physics remains unchanged. There are two places where this state of affairs no longer holds. The first is in gravity where the energy of the vacuum contributes as a cosmological constant. The second is, as we've seen above, in supersymmetric theories where energies are necessarily positive definite.

Physically, it's far from clear if there is any deep relation between these two ideas. In fact, as we will see later in these lectures, the energy of the ground state acts as an order parameter for the breaking of supersymmetry. This means that the ground state energy is zero if supersymmetry is exact, otherwise it is non-zero. In our world, it's clear that there is no supersymmetry visible at the TeV scale, while the cosmological constant is many of orders of magnitude smaller, at 10^{-3} eV. This makes it difficult to see how supersymmetry can help alleviate the cosmological constant problem.

However, at the formal mathematical level, the relationship between supersymmetry and gravity has proven rather useful. For example, there exists a greatly simplified proof of the positive energy theorem in general relativity, due to Witten, that uses ideas of supersymmetry.

There is one further piece of physics hiding in (2.26). For any other symmetry in field theory, we can think about gauging it. This means that we try to construct theories in which the symmetry is realised locally. Supersymmetry is no different. One can construct theories in which the associated infinitesimal parameter for supersymmetry transformations depends on x^μ. From (2.26), we see that such theories necessarily enjoy a symmetry in which you do different translations at different points in space. But such transformations are diffeomorphisms and are the characteristic feature of general relativity. In other words, theories of local supersymmetry are necessarily theories of gravity! Such theories are known as *supergravity*, usually shortened to the ugly acronym "sugra". We will mention supergravity only very briefly in this section. In subsequent sections our interest will be entirely on theories with global supersymmetry.

2.3 Representations on Particle States

Given an algebra, our next task is to explore its representations. There are different ways that we could approach this. Ultimately, we will be interested in quantum field theories that enjoy supersymmetry and this means understanding the way supersymmetry acts on fields. This we will do in later sections. Here, to build some intuition, we will understand how supersymmetry acts on single particle states in the Hilbert space.

Without doing any work, we can guess that something interesting is going on. The supercharge Q_α is a fermionic operator, both in the sense that it carries spin $\frac{1}{2}$ and in the sense that it is naturally anti-commuting as in (2.21). This means that, schematically, we must have

$$Q|\text{fermion}\rangle = |\text{boson}\rangle \quad \text{and} \quad \text{Q}|\text{boson}\rangle = |\text{fermion}\rangle \tag{2.28}$$

This is the defining feature of supersymmetry.

In fact, it is straightforward to show that any representation of the supersymmetry algebra must have an equal number of bosonic and fermionic states. To this end, we introduce the *fermionic number operator* $(-1)^F$. This acts on bosonic states as

$$(-1)^F|B\rangle = |B\rangle \quad \text{and} \quad (-1)^F|F\rangle = -|F\rangle$$

Because Q_α swaps a bosonic state for a fermionic state, we necessarily have

$$(-1)^F Q_\alpha = -Q_\alpha(-1)^F \quad \Rightarrow \quad \{(-1)^F, Q_\alpha\} = 0$$

The result that we now want follows straightforwardly from the algebra $\{Q_\alpha, \bar{Q}_{\dot\alpha}\} = 2\sigma^\mu_{\alpha\dot\alpha}P_\mu$. Suppose that we have a finite collection of one-particle states that form a representation of the supersymmetry algebra. We can take the folllowing trace over elements of this multiplet

$$\begin{aligned}\text{tr}\left[(-1)^F\{Q_\alpha, \bar{Q}_{\dot\alpha}\}\right] &= \text{tr}\left[(-1)^F Q_\alpha\bar{Q}_{\dot\alpha} + (-1)^F\bar{Q}_{\dot\alpha}Q_\alpha\right] \\ &= \text{tr}\left[-Q_\alpha(-1)^F\bar{Q}_{\dot\alpha} + (-1)^F\bar{Q}_{\dot\alpha}Q_\alpha\right] = 0\end{aligned}$$

Here the second equality we've uses the fact that $\{(-1)^F, Q_\alpha\} = 0$ while the final equality uses the cyclicity of the trace. The supersymmetry algebra then tells us that

$$\sigma^\mu_{\alpha\dot\alpha}\,\text{tr}\left[(-1)^F P_\mu\right] = 0$$

Note that $\sigma^\mu_{\alpha\dot\alpha}$ sits outside the trace over states: it's just a bunch of numbers as far as the trace is concerned. Meanwhile P_μ sits inside the trace because it is an operator

acting on states. We can choose these states to be momentum eigenstates, so that $P_\mu|\text{any state}\rangle = p_\mu|\text{any state}\rangle$. We then simply have

$$\sigma^\mu_{\alpha\dot\alpha} p_\mu \,\text{tr}\,(-1)^F = 0$$

But $\text{tr}(-1)^F$ simply counts the number of bosonic states n_B minus the number of fermionic states n_F,

$$\text{tr}(-1)^F = n_B - n_F = 0$$

The number of such states must be equal. The quantity $\text{tr}(-1)^F$ is called the *Witten index*.

There's actually a loophole in the discussion above. It may be that Q_α and $\bar{Q}_{\dot\alpha}$ annihilate states in the supersymmetry multiplet. From the supersymmetry algebra (and the positivity conditions (2.27) that follows from it) this can only happen for states of zero energy which are necessarily the ground states of the system. This means that there may be a mismatch between the number of bosonic and fermionic ground states of a system. It is in studying such ground states that the *Witten index* really finds it teeth and we'll revisit this in Section 3.4.2. More sophisticated examples can be found in the lectures on Supersymmetric Quantum Mechanics.

We now know that supersymmetry requires an equal number of bosonic and fermionic states. The next step is to understand exactly what kind of fermion is paired with what kind of boson.

2.3.1 Representations of the Poincaré Group

To set the scene, let's first recall how we construct the irreducible representations of the Poincaré group. In fact, let's start even more simply: how do we construct irreducible representations of the rotation group?

We work with the algebra $so(3) \cong su(2)$ rather than the group. This is, of course, defined by the familiar commutation relations

$$[J_i, J_j] = i\epsilon_{ijk} J_k$$

To construct representations, the first thing we do is look to the Casimirs. These are operators that commute with all generators of the group. For $su(2)$, there is just a single Casimir,

$$C = \sum_{i=1}^{3} J_i^2$$

Irreducible representation are labelled by their eigenvalue of the Casimir. For $su(2)$, the eigenvalue of J^2 is $j(j+1)$ with the spin j taking values in $j = 0, \frac{1}{2}, 1, \ldots$. Each representation has dimension $2j+1$, with the states within a multiplet identified by their eigenvalue under, say, J_3 whose eigenvalue lies in $|j_3| \leq j$. The result is the familiar one from quantum mechanics: states are labelled by $|j, j_3\rangle$

Now let's turn to the Poincaré group. The irreducible representations are what we call "particles". Again, they are characterised by the Casimirs. I won't tell you how to construct Casimirs, but will instead just present you the result: the Poincaré group has two Casimirs, given by

$$C_1 = P_\mu P^\mu \quad \text{and} \quad C_2 = W_\mu W^\mu$$

Here $W^\mu = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} P_\nu M_{\rho\sigma}$ is the *Pauli-Lubański vector*. It can be thought of as a relativistic version of angular momentum.

Representations of the Poincaré group are then labelled by the eigenvalues of C_1 and C_2. The first of these is simply the mass m of a particle: $C_1 = m^2$. What happens next is a little different depending on whether the particles are massive or massless.

- Massive Particles: In this case, we can always boost to the rest frame of the particle so that $P^\mu = (m, 0, 0, 0)$. In this frame, the Pauli-Lubański vector is

  $$W^0 = 0 \quad \text{and} \quad W^i = -mJ^i$$

  with J^i the generators of rotations. This means that $C_2 = -m^2 J^2$ and so is specified by the eigenvalue of J^2. We find the familiar fact that massive particles are characterised by their mass m and spin j.

- Massless Particles: Now $C_1 = m^2 = 0$. There are some subtleties that we sweep under the rug here, but it turns out that the most interesting representations also have $C_2 = W^2 = 0$, so both Casimirs vanish. To characterise the representation, we choose a frame such that, say, $P^\mu = (E, 0, 0, E)$. There, we have $W^\mu = M_{12} P^\mu$, so the constant of proportionality between W and P is determined by the eigenvalue of the $U(1)$ rotation in the (x^1, x^2)-plane. The eigenvalue of this rotation is the *helicity*, $h = 0, \frac{1}{2}, 1, \ldots$. We learn that massless particles are characterised by (obviously) $m = 0$ and their helicity h.

Although the results are different for $m = 0$ and $m \neq 0$, the strategy is the same. In each case, we boost to a preferred frame of the particle which is then characterised by how it transforms under the surviving symmetry group. This surviving symmetry — $SU(2)$ for a massive particle, $U(1)$ for a massless one — is called the *little group*.

There is a slight twist to the story when it comes to realising these representations on the Hilbert space of single particle states. For massive particles, the states take the form

$$|p_\mu; j, j_3\rangle \tag{2.29}$$

where the momentum is restricted to obey $p_\mu p^\mu = m^2$ while the azimuthal angular momentum takes values in $j_3 \leq |j|$. This fills out the $2j+1$ dimensional set of spin sets. However, for massless particles, there is just a single state $|p_\mu; h\rangle$. This is because the helicity describes the representation of the Abelian group $U(1)$ generated by M^{12} rather than the non-Abelian group $SU(2)$ and irreducible representations of Abelian groups are one-dimensional.

The problem is that we know that massless particles also have internal degrees of freedom. For example, the photon necessarily has two polarisation states. Clearly we're missing something. What we're missing is the additional requirement that the spectrum of states is invariant under CPT. For massive particles, this doesn't buy us anything new: the set of states (2.29) is already invariant under CPT. However, for massless particles CPT flips $h \mapsto -h$ and tells us that massless states must come in pairs

$$|p_\mu; h\rangle \quad \text{and} \quad |p_\mu, -h\rangle$$

This is the origin of the two polarisation states of the photon or graviton, or the two helicities of a massless Weyl spinor. Note that a massless scalar has helicity $h = 0$ and so is CPT self-conjugate. This means that there's no requirement from CPT to add an additional degree of freedom in this case.

2.3.2 Massless Representations

We now turn to the representations of the $\mathcal{N} = 1$ supersymmetry algebra. The simple observation (2.28) tells us that we should expect representations to contain particles of different spin and this will turn out to be true. Once again we need to treat massless and massive particles separately.

The supersymmetry algebra also has two Casimirs. The first is familiar:

$$C_1 = P_\mu P^\mu$$

The fact that this is a Casimir tells us that all particles in a supersymmetric multiplet must have the same mass, $C_1 = m^2$.

In contrast, the other Casimir of the Poincaré group, $W_\mu W^\mu$, is *not* a Casimir of the supersymmetry algebra. This is because $[W_\mu, Q_\alpha] \neq 0$ which, in turn, can be traced to the commutation relation $[M_{\mu\nu}, Q_\alpha] \neq 0$. But it was $W_\mu W^\mu$ that told us that representations of the Poincaré group are characterised by the spin of a particle. The fact that $W_\mu W^\mu$ is no longer a Casimir means that representations of the supersymmetry algebra can contain particles of different spin.

It is possible to construct a new Casimir. First define

$$Y_\mu = W_\mu - \frac{1}{4}\bar{Q}_{\dot\alpha}\bar{\sigma}_\mu^{\dot\alpha\beta}Q_\beta$$

Then the second Casimir of the supersymmetry algebra turns out to be

$$\tilde{C}_2 = (Y_\mu P_\nu - Y_\nu P_\mu)(Y^\mu P^\nu - Y^\nu P^\mu)$$

However, in what follows we won't need this result. Instead we will build up a representation of the supersymmetry algebra more directly. Our strategy is to start from a particle (i.e. a representation of the Poincaré group) and then act on it with successive supersymmetry generators until we build up a representation of the full algebra.

It turns out that things are slightly simpler for massless representations. Consider a state $|p_\mu, h\rangle$ of a massless particle of helicity h. We can again boost to a frame in which $p_\mu = (E, 0, 0, E)$. Restricted to act on such states, the supersymmetry algebra becomes

$$\{Q_\alpha, \bar{Q}_{\dot\alpha}\} = 2\sigma^\mu_{\alpha\dot\alpha}P_\mu = 2E(1+\sigma^3)_{\alpha\dot\alpha} = 4E\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

From the positivity condition (2.27), we see that Q_2 and $\bar{Q}_2$ necessarily annihilate this state,

$$\langle p_\mu, h|\{Q_2, \bar{Q}_2\}|p_\mu, h\rangle = 0 \quad \Rightarrow \quad Q_2|p_\mu, h\rangle = \bar{Q}_2|p_\mu, h\rangle = 0$$

To build a representation of the full supersymmetry algebra, we only need consider the action of Q_1 and $\bar{Q}_1$. But these act just like fermionic creation and annihilation operators. Specifically, if we rescale the operators to become

$$a = \frac{Q_1}{\sqrt{4E}} \quad \text{and} \quad a^\dagger = \frac{\bar{Q}_1}{\sqrt{4E}} \quad \Rightarrow \quad \{a, a^\dagger\} = 1 \quad \text{and} \quad \{a, a\} = \{a^\dagger, a^\dagger\} = 0$$

The representations of this algebra are straightforward: they consist of two states $|0\rangle$ and $|1\rangle$ such that $a|0\rangle = 0$ and $|1\rangle = a^\dagger|0\rangle$. This ensures that $a^\dagger|1\rangle = 0$. For us, this means that we can start by taking a state which, by assumption, is annihilated by a,

$$a|p_\mu, h\rangle = 0$$

The full supersymmetry multiplet then consists of $|p_\mu, h\rangle$ and $a^\dagger|p_\mu, h\rangle$. The question is: what is the helicity of this second state? This follows from the commutation relation (2.22)

$$[M^{\mu\nu}, Q_\alpha] = (\sigma^{\mu\nu})_\alpha{}^\beta Q_\beta \quad \text{and} \quad [M^{\mu\nu}, \bar{Q}^{\dot\alpha}] = (\bar\sigma^{\mu\nu})^{\dot\alpha}{}_{\dot\beta}\bar{Q}^{\dot\beta} \tag{2.30}$$

Restricting to rotations in the (x^1, x^2) plane, which is what we mean by helicity, we have

$$[M^{12}, Q_1] = \frac{1}{2}Q_1 \quad \text{and} \quad [M^{12}, Q_2] = -\frac{1}{2}Q_2$$
$$[M^{12}, \bar{Q}^1] = \frac{1}{2}\bar{Q}^1 \quad \text{and} \quad [M^{12}, \bar{Q}^2] = -\frac{1}{2}\bar{Q}^2$$

The first equation tells us that Q_1 raises the helicity by $\frac{1}{2}$. This suggests that the adjoint $\bar{Q}_1$ lowers the helicty by $\frac{1}{2}$. To see that this is the case, we need to remember that, after lowering an index, $\bar{Q}_1 = -\bar{Q}^2$ so we have

$$[M^{12}, \bar{Q}_1] = -\frac{1}{2}\bar{Q}_1$$

So $\bar{Q}_1$ does indeed lower the helicity by $\frac{1}{2}$ as anticipated. We learn that the massless representations of the supersymmetry algebra consist of just two states:

$$|p_\mu, h\rangle \quad \text{and} \quad |p_\mu, h - \frac{1}{2}\rangle = \frac{\bar{Q}_1}{\sqrt{4E}}|p_\mu, h\rangle$$

As we saw above, for massless states we must also add their CPT conjugates. The different representations of the supersymmetry algebra then arise by picking different starting helicities h. There are three representations that are most important:

- If we start with $h = \frac{1}{2}$ then we have

h	$-\frac{1}{2}$	0	$+\frac{1}{2}$
multiplicity	1	2	1

  This is the matter content that we get from quantising a single Weyl spinor together with a *complex* scalar. This is known as a *chiral multiplet*.

  The chiral multiplets should be thought of as matter particles. We will devote Section 3 to studying field theories associated to chiral multiplets. Here we make a quick comment. The fact that any other internal symmetry generator must commute with Q_α means that the fermion and scalar in a given chiral multiplet must experience the same force. In particular, if one is charged under a gauge group then so is the other. We'll see this explicitly when we construct supersymmetry gauge theories in Section 4.

- If we start with $h = 1$ then we have

h	-1	$-\frac{1}{2}$	$+\frac{1}{2}$	$+1$
multiplicity	1	1	1	1

  This is the matter content of a photon together with a single Weyl spinor. It is known as the *gauge multiplet* or *vector multiplet*.

  We will devote Section 4 to the study of vector multiplets. There we will see that we can construct supersymmetric versions of Yang-Mills theory with gauge group G by taking $\dim G$ vector multiplets. As usual, the $h = 1$ gauge bosons transform in the adjoint of the gauge group. But now, so too, must its fermionic supersymmetric partner. In this context, the fermion is called a *gaugino*.

- If we start with $h = 2$ then we have

h	-2	$-\frac{3}{2}$	$+\frac{3}{2}$	$+2$
multiplicity	1	1	1	1

  This is the matter content of a graviton together with a helicity $\frac{3}{2}$ spinor, sometimes known as a *Rarita-Schwinger* field or, in this context, the *gravitino*. They combine to form the *supergravity multiplet*.

If we keep going, we get massless fields with helicity $h > 2$. But there are strong restrictions that prohibit the existence of interacting theories with massless fields of such high helicity. (This statement is true in Minkowski spacetimes; there are remarkable "higher spin" theories that include an infinite tower of massless states in de Sitter or anti de Sitter spacetimes.) We also skipped the $h = \frac{3}{2}$ multiplet for similar reasons; it turns out that the existence of a massless helicity $\frac{3}{2}$ particle implies the existence of a local supersymmetry which, in turn, requires that the theory is coupled to gravity.

2.3.3 Massive Representations

We next turn to massive representations of the supersymmetry algebra. In the rest frame of a particle we have $p_\mu = (m, 0, 0, 0)$. Acting on such states, the supersymmetry algebra becomes

$$\{Q_\alpha, \bar{Q}_{\dot{\alpha}}\} = 2\sigma^\mu_{\alpha\dot{\alpha}} P_\mu = 2m\sigma^0_{\alpha\dot{\alpha}} = 2m \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.31}$$

This time, after rescaling, both Q_1 and Q_2 act as fermionic creation/annihilation operators

$$a_\alpha = \frac{Q_\alpha}{\sqrt{2m}} \quad \text{and} \quad a^\dagger_{\dot\alpha} = \frac{\bar{Q}_{\dot\alpha}}{\sqrt{2m}} \quad \Rightarrow \quad \{a_\alpha, a^\dagger_{\dot\alpha}\} = \delta_{\alpha\dot\alpha}$$

with $\{a_\alpha, a_\beta\} = \{a^\dagger_{\dot\alpha}, a^\dagger_{\dot\beta}\} = 0$. We start with a state $|\Omega\rangle = |p_\mu; j, j_3\rangle$ that we assume to be annihilated by $a_\alpha|\Omega\rangle = 0$. Then the full supermultiplet consists of four states

$$\begin{array}{c} |\Omega\rangle \\ a^\dagger_1|\Omega\rangle \quad \text{and} \quad a^\dagger_2|\Omega\rangle \\ a^\dagger_1 a^\dagger_2|\Omega\rangle \end{array}$$

Again, the question is: what is the spin of these other states. We could use the commutation relations (2.30) to understand how the new states transform under the $SU(2)$ little group but it's a little fiddly while the end result is intuitive and straightforward. The initial state $|\Omega\rangle$ has spin j. The states $a^\dagger_\alpha|\Omega\rangle$ then sit in the tensor product of representations $j \otimes \frac{1}{2} = (j + \frac{1}{2}) \oplus (j - \frac{1}{2})$. The final state can be written as $a^\dagger_1 a^\dagger_2|\Omega\rangle = \frac{1}{2}\epsilon^{\alpha\beta} a^\dagger_\alpha a^\dagger_\beta|\Omega\rangle$, where the $\epsilon^{\alpha\beta}$ now contracts the creation operators to be a spin singlet. This means that the state $a^\dagger_1 a^\dagger_2|\Omega\rangle$ once again has spin j.

The upshot is that a massive supermultiplet contains two particles of spin j, a particle of spin $j - \frac{1}{2}$ and a particle of spin $j + \frac{1}{2}$. Note that the degeneracy of the two particles of spin j is precisely equal to the degeneracies of the other two particles:

$$2 \times (2j + 1) = \left[2\left(j + \frac{1}{2}\right) + 1\right] + \left[2\left(j - \frac{1}{2}\right) + 1\right]$$

This is simply that statement that we saw previously: a supermultiplet must have an equal number of bosonic and fermionic degrees of freedom.

There are just two massive supermultiplets that will be of interest

- If we start with $j = 0$, we have

j	0	$\frac{1}{2}$
multiplicity	2	1

  This is the matter content of a massive complex scalar with a single massive Weyl fermion. We recognise it as the same matter content as the chiral multiplet that we met previously, now of course with all particles having a mass.

- If we start with $j = \frac{1}{2}$, we have

j	0	$\frac{1}{2}$	1
multiplicity	1	2	1

  In other words, we have a massive spin 1 particle, two massive Weyl fermions, and a massive spin 0 particle. This is now more states than we found in the massless gauge multiplet. In fact, this collection of states is equivalent to a massless gauge multiplet *and* a massless chiral multiplet. But that makes sense. In quantum field theory, a massless gauge boson can become massive only through the Higgs mechanism, in which the gauge boson "eats" a scalar. The supersymmetric extension of this is that a massless vector multiplet "eats" a chiral multiplet to become the massive vector multiplet described above.

There's one further subtlety that is worth flagging up. This is how parity acts on the two scalars in the massive chiral multiplet. It turns out that one of them is a scalar and the other a pseudoscalar. Here, the meaning of a "pseudoscalar" is that it picks up a minus sign under parity. This statement follows, like everything else in this section, from the supersymmetry algebra. We denote the parity operator as $\hat{\mathcal{P}}$ to distinguish it from the momentum operator P^μ. By definition, we must have

$$\hat{\mathcal{P}} P^\mu \hat{\mathcal{P}}^{-1} = (P^0, -P^i)$$

Meanwhile, parity also exchanges left-handed and right-handed spinors. This means that parity must exchange some combination of Q_α and $\bar{Q}_{\dot\alpha}$. One can check that the supersymmetry algebra remains unchanged if we take

$$\hat{\mathcal{P}} Q_\alpha \hat{\mathcal{P}}^{-1} = (\sigma^0)_{\alpha\dot\alpha}\bar{Q}^{\dot\alpha} \quad \text{and} \quad \hat{\mathcal{P}}\bar{Q}^{\dot\alpha}\hat{\mathcal{P}}^{-1} = -(\sigma^0)^{\dot\alpha\alpha}Q_\alpha$$

(More generally one can include a complex phase in these relations but it will not affect our discussion here.)

Now our two scalar states in the massive chiral multiplet are $|\Omega\rangle$ and $|\Omega'\rangle = a_1^\dagger a_2^\dagger|\Omega\rangle \sim \bar{Q}_1\bar{Q}_2|\Omega\rangle$. They obey $Q_\alpha|\Omega\rangle = \bar{Q}_{\dot\alpha}|\Omega'\rangle = 0$. Since parity exchanges Q_α and $\bar{Q}_{\dot\alpha}$, it must also exchange $|\Omega\rangle$ and $|\Omega'\rangle$. This means that the parity eigenstates are

$$\hat{\mathcal{P}}\left(|\Omega\rangle \pm |\Omega'\rangle\right) = \pm\left(|\Omega\rangle \pm |\Omega'\rangle\right)$$

and we have one scalar (with the + sign) and one pseudoscalar (with the - sign) as advertised.

2.4 Extended Supersymmetry

It is possible for theories to exhibit more than one supersymmetry. This means that there is a collection of $\mathcal{N}$ supercharges

$$Q^I_\alpha \quad \text{and} \quad \bar{Q}^I_{\dot\alpha} \quad I = 1, \ldots, \mathcal{N}$$

Each of these supercharges retains the same commutation relations with the generators of the Poincaré group,

$$[M^{\mu\nu} Q^I_\alpha] = (\sigma^{\mu\nu})_\alpha{}^\beta Q^I_\beta \quad \text{and} \quad [P^\mu, Q^I_\alpha] = 0$$

and the key part of the supersymmetry algebra holds for each generator separately

$$\{Q^I_\alpha, \bar{Q}^J_{\dot\alpha}\} = 2\sigma^\mu_{\alpha\dot\alpha} P_\mu \, \delta^{IJ}$$

However, there are two novelties. The first is that the anti-commutator of the supercharges with themselves can be more interesting

$$\{Q^I_\alpha, Q^J_\beta\} = \epsilon_{\alpha\beta} Z^{IJ} \quad \text{and} \quad \{\bar{Q}^I_{\dot\alpha}, \bar{Q}^J_{\dot\beta}\} = \epsilon_{\dot\alpha\dot\beta} (Z^\dagger)^{IJ} \tag{2.32}$$

Here $Z^{IJ} = -Z^{JI}$ is a *central charge*, meaning that it commutes with all other elements of the algebra. The exact nature of these central charges depends on the precise theory that we consider, but they must be constructed from other conserved quantities that are at hand. We'll see the role that these central charges play shortly.

The second novelty is the R-symmetry group. Recall that for $\mathcal{N} = 1$ we had a $U(1)_R$ symmetry (2.24) that rotates the phase of the supercharge. For $\mathcal{N} > 1$, the R-symmetry rotates the supercharges among themselves. For reasons that will become clear shortly, our primary interest will be in $\mathcal{N} = 2$ and $\mathcal{N} = 4$ supersymmetry. Here the R-symmetries are:

- $\underline{\mathcal{N} = 2}$: The R-symmetry group is $U(2)_R \cong U(1)_R \times SU(2)_R$.
- $\underline{\mathcal{N} = 4}$: A priori, the R-symmetry group is $U(4)$. However, it turns out that only $SU(4)$ is realised on fields. This is equivalent to $SU(4) \cong \text{Spin}(6)$. (This is sometimes written, a little inaccurately, as $SO(6)$ but the supercharges transform in the spinor representation of Spin(6) which is not a representation of $SO(6) = \text{Spin}(6)/\mathbf{Z}_2$.)

Theories with extended supersymmetry are a subset of those theories with $\mathcal{N} = 1$ supersymmetry. This means that the representations of theories with $\mathcal{N} > 1$ must be constructed by joining together the $\mathcal{N} = 1$ supermultiplets that we described above. In the rest of this section, we explain how this works.

2.4.1 Massless Representations

For representations on states $|p^\mu, h\rangle$ of massless particles, we proceed as before. We boost to a frame with $p_\mu = (E, 0, 0, E)$ and restrict attention to the algebra on such states. We then have

$$\{Q^I_\alpha, \bar{Q}^J_{\dot{\alpha}}\} = 4E \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \delta^{IJ}$$

As previously, we have $Q^I_2|p^\mu, h\rangle = \bar{Q}^I_2|p^\mu, h\rangle = 0$. From (2.32), we then have $Z^{IJ}|p^\mu, h\rangle = 0$ which tells us that the central charges play no role for the massless states. We're left, as before, just with the Q^I_1 and $\bar{Q}^I_1$ operators to deal with. These now form a collection of $\mathcal{N}$ fermionic creation and annihilation operators

$$a^I = \frac{Q^I_1}{\sqrt{4E}} \quad \text{and} \quad a^{I\dagger} = \frac{\bar{Q}^I_1}{\sqrt{4E}} \quad \Rightarrow \quad \{a^I, a^{J\dagger}\} = \delta^{IJ} \quad \text{and} \quad \{a^I, a^J\} = \{a^{I\dagger}, a^{J\dagger}\} = 0$$

We now start with some fiducial state $|\Omega\rangle = |p^\mu, h\rangle$ satisfying $a^I|\Omega\rangle = 0$ and build up the full representation by acting with successive creation operators. The end result is a collection of states

$$\begin{gathered} |\Omega\rangle \\ a^{I\dagger}|\Omega\rangle \\ a^{I\dagger}a^{J\dagger}|\Omega\rangle \\ \cdots \\ a^{1\dagger}\ldots a^{\mathcal{N}\dagger}|\Omega\rangle \end{gathered}$$

Our initial state $|\Omega\rangle$ has helicity h. If we act with p of the $a^\dagger$ excitation operators then there are $\binom{\mathcal{N}}{p}$ different states, each of which has helicity $h - p/2$. The full multiplet consists of $2^\mathcal{N}$ different states. If we add the CPT conjugate states then we have $2^{\mathcal{N}+1}$ states overall. Let's now look at some specific examples.

$\mathcal{N} = 2$ Supersymmetry

Again, the different multiplets arise by considering initial states $|\Omega\rangle$ with different helicities. We'll deal with each in turn.

- If we start with $h = \frac{1}{2}$ then there are two states in the first level, $a^{I\dagger}|\Omega\rangle$, each with $h = 0$, and a single state in the final level, $a^{1\dagger}a^{2\dagger}|\Omega\rangle$, with $h = -\frac{1}{2}$. After adding the CPT conjugate we end up with

h	$-\frac{1}{2}$	0	$+\frac{1}{2}$
multiplicity	2	4	2

This is called a *hypermultiplet*. In consists of two chiral multiplets or, equivalently, two complex scalars and a Dirac fermion (i.e. two Weyl fermions).

You might wonder why we needed to add the CPT conjugate in this case. After all, starting with $h = +\frac{1}{2}$ gave a single chiral multiplet which is already CPT self-conjugate. The answer to this is buried in the details of the $SU(2)_R$ symmetry which acts on the scalars $a^{I\dagger}|\Omega\rangle$ as a doublet. But this means that each of these scalars must be complex and that, in turn, requires that we add the CPT conjugate.

- If we start with $h = 0$ then we get two additional states with $h = -\frac{1}{2}$ and one with $h = -1$. Adding the CPT conjugate gives

h	-1	$-\frac{1}{2}$	0	$+\frac{1}{2}$	$+1$
multiplicity	1	2	2	2	1

This is the $\mathcal{N} = 2$ vector multiplet, comprising of an $\mathcal{N} = 1$ vector multiplet and $\mathcal{N} = 1$ chiral multiplet.

- If we start with $h = 2$ then, after adding the CPT conjugate, we end up with

h	-2	$-\frac{3}{2}$	-1	$+1$	$+\frac{3}{2}$	$+2$
multiplicity	1	2	1	1	2	1

This is the $\mathcal{N} = 2$ supergravity multiplet. It comprises of an $\mathcal{N} = 1$ supergravity multiplet together with an $\mathcal{N} = 1$ vector multiplet.

There's one important feature of the spectrum above that is worth highlighting. The fermions now come in pairs, meaning that they can be viewed as Dirac fermions rather than Weyl fermions. This puts restrictions on the kind of supersymmetric theories that we can build. In particular, it's not possible to construct a *chiral gauge theory* with $\mathcal{N} > 1$ supersymmetry. Here a chiral theory is one in which left- and right-handed fermions experience different forces, like in the Standard Model. Such theories are possible with $\mathcal{N} = 1$ supersymmetry (or, indeed, $\mathcal{N} = 0$ supersymmetry as in our world!). But any extended supersymmetry forces the theories to be vector-like.

$\mathcal{N} = 4$ Supersymmetry

We can play the same game with $\mathcal{N} = 4$ supersymmetry.

- If we start with $h = 1$ then we get the following multiplet

h	-1	$-\frac{1}{2}$	0	$+\frac{1}{2}$	$+1$
multiplicity	1	4	6	4	1

  This consists of an $\mathcal{N} = 2$ vector multiplet with an $\mathcal{N} = 2$ hypermultiplet and is the unique $\mathcal{N} = 4$ multiplet that does not include gravity. Note that there is now no longer a distinction between forces and matter: once you specify the gauge group, all matter content is also fixed. Furthermore, all matter fields necessarily transform in the adjoint representation of the gauge group.

  For once, we did not need to add the CPT conjugate to the above multiplet: it's already CPT self- conjugate. As we saw above, it was almost possible to achieve this for the $\mathcal{N} = 2$ matter representation but we fell at the last hurdle when we considered how the $SU(2)_R$ symmetry acts on the scalars. But now we have no such concern. The scalars are the set of 6 states $a^{I\dagger}a^{J\dagger}|\Omega\rangle$ and transform in the **6** of the $SU(4)$ R-symmetry. But this is a real representation and there is no need to add the CPT conjugate.

- If we start with $h = 2$ then, after adding the CPT conjugate multiplet, we have

h	-2	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$+\frac{1}{2}$	$+1$	$+\frac{3}{2}$	$+2$
multiplicity	1	2	2	2	2	2	2	2	1

  This is the $\mathcal{N} = 4$ supergravity multiplet, comprising of an $\mathcal{N} = 2$ supergravity multiplet and $\mathcal{N} = 2$ vector multiplet.

You may have noticed that we jumped straight from $\mathcal{N} = 2$ to $\mathcal{N} = 4$, missing out $\mathcal{N} = 3$ in the middle. If you try to build a multiplet of single particle states with $\mathcal{N} = 3$ supersymmetry starting from, say, $h = \frac{1}{2}$ or $h = 1$ then you'll find that you're obliged to add the CPT conjugate representation and you just end up with $\mathcal{N} = 4$ supersymmetry after all. This observation is the key element of a proof that says any *perturbative* theory with $\mathcal{N} = 3$ global supersymmetry necessarily has $\mathcal{N} = 4$ supersymmetry.

The word "perturbative" is important in the above statement. This means that the theory is weakly coupled and the single particle states that we're considering here are a good approximation to the spectrum of the theory. It turns out $\mathcal{N} = 3$ supersymmetry can be realised in strongly coupled, interacting quantum field theories, with no perturbative regime.

$\mathcal{N} = 8$ Supersymmetry

If we go beyond $\mathcal{N} = 4$ supersymmetry then we no longer have multiplets with helicities $h \leq 1$. This means that we are now necessarily in the realm of local supersymmetry and supergravity. Furthermore, by the time we get beyond $\mathcal{N} = 8$ supersymmetry the multiplets have particles with helicity $h > 2$. As we mentioned before, such theories are always free in Minkowski space and therefore of limited interest. In this sense, $\mathcal{N} = 8$ is the maximum number of supersymmetries possible. The theory has a unique supergravity multiplet with the following degeneracies

h	-2	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$+\frac{1}{2}$	$+1$	$+\frac{3}{2}$	$+2$
multiplicity	1	8	28	56	70	56	28	8	1

$\mathcal{N} = 8$ supergravity has some interesting properties and plays a role in string theory. However, we won't discuss it further in this course.

2.4.2 Massive Representations and BPS Bounds

Rather than repeating the whole story for massive representations, we will instead just focus on the novelty. This arises from the central charges Z^{IJ} that appear in the supersymmetry algebra

$$\{Q^I_\alpha, Q^J_\beta\} = \epsilon_{\alpha\beta} Z^{IJ}$$

For reasons that we now explain, this is where much of the power of extended supersymmetry comes from.

Our goal is to understand representations of this algebra, in conjunction with the original supersymmetry algebra which, in the rest frame of the particle, reads (2.31)

$$\{Q^I_\alpha, \bar{Q}^J_{\dot\alpha}\} = 2m \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \delta^{IJ}$$

We'll illustrate the story with $\mathcal{N} = 2$ supersymmetry, although the general idea holds for any theory with extended supersymmetry. With $\mathcal{N} = 2$, the anti-symmetric central charge is necessarily just a complex number Z

$$Z^{IJ} = 2\epsilon^{IJ} Z$$

For simplicity, we take Z to be real. (Typically it's not but we'll dodge this issue for now and state the full result below.) We then define the following combination of creation and annihilation operators

$$a_\alpha = \frac{1}{\sqrt{2}}\begin{pmatrix} Q_1^1 + \bar{Q}_2^2 \\ Q_2^1 - \bar{Q}_1^2 \end{pmatrix} \quad \text{and} \quad b_\alpha = \frac{1}{\sqrt{2}}\begin{pmatrix} Q_1^1 - \bar{Q}_2^2 \\ Q_2^1 + \bar{Q}_1^2 \end{pmatrix}$$

Note that we've mixed up α and $\dot{\alpha}$ indices. This is acceptable because we're working in the rest frame of the particle and so have already broken Lorentz invariance. The choice of a and b operators is designed to disentangle the mass and central charge Z, so their commutation relations read

$$\{a_\alpha, a_\beta^\dagger\} = 2(m+Z)\delta_{\alpha\beta} \quad \text{and} \quad \{b_\alpha, b_\beta^\dagger\} = 2(m-Z)\delta_{\alpha\beta}$$

with all other anti-commutators vanishing. The $\{a_\alpha, a_\beta^\dagger\}$ and $\{b_\alpha, b_\beta^\dagger\}$ are both positive definite, so the corresponding right-hand sides must be too. But this is only true if the masses are bounded by the central charges,

$$m \geq |Z|$$

This formula also holds if Z is complex; we just need to redefine the operators a and b using a phase to derive the same result. This formula is interesting. Although we haven't seen yet any specific examples, recall that the central charge Z is some combination of conserved charges in the quantum field theory. We learn that the masses of particles is bounded by the charges. This is known as the *BPS bound* although in the present context the name *Witten-Olive bound* would be more appropriate.

What about the representation theory of the algebra? Crucially, this depends on whether $m > |Z|$ or $m = |Z|$.

If $m > |Z|$, then we are in a situation very similar to the massive representation theory that we saw before. Both $a_\alpha^\dagger$ and $b_\alpha^\dagger$ act as creation operators and the result is that we have a multiplet comprising of 16 states. This is known as a *long multiplet*. We can also repeat this story with $\mathcal{N}$ supersymmetries to find that long multiplets have $2^{2\mathcal{N}}$ states.

More interesting is what happens when $m = |Z|$. In this case, half of the creation operators do nothing. For example, when $m = Z$, the b_α operators must just vanish on all states in the multiplet. Now we're back to the situation we met when discussing massless representations, with only $a_\alpha^\dagger$ acting as creation operators. The result is the hypermultiplet or vector multiplet that we saw above, each with 8 states, but now with a mass $m = Z$. This is known as a *short multiplet*.

The existence of short multiplets, whose mass is fixed to be $m = |Z|$, turns out to be a wonderfully powerful tool in the study of quantum field theories with extended supersymmetry. The basic idea is that one can usually solve quantum field theories at weak coupling. There we can identify the various states and understand the spectrum of long and short multiplets. As one moves into the strong coupling realm, we typically lose control over the dynamics. However, the short multiplets are special because their mass is pinned to be $m = |Z|$. The mass can't deviate from $|Z|$ because this would need there to be extra states in the Hilbert space and these can't magically appear from nowhere as some parameter, like a coupling constant, is varied. The only way that the short multiplets can free themselves from this constraint is if two or more short multiplets become degenerate and then combine to become a long multiplet whose mass is no longer protected. By understanding when this can (or, better yet, can't) happen we get a precious handle on the strong coupling dynamics of certain quantum field theories.

In this way, the study of short BPS multiplets shines a rare light into what happens at strong coupling. It allows us to effectively solve the dynamics of $\mathcal{N} = 2$ and $\mathcal{N} = 4$ gauge theories. It also allows us to understand the strong coupling limits of string theory, including the existence of M-theory, and to compute the microscopic entropy of certain BPS black hole solutions. It is, in short, a very useful tool.

The BPS trick is not available for $\mathcal{N} = 1$ theories and so we won't be wielding it for much of these lectures. (Actually, it can be used to compute the tension of domain walls and vortex strings in certain $\mathcal{N} = 1$ theories, but not the masses of particle states.)

2.4.3 Supersymmetry in Other Dimensions

Throughout these lectures, we will restrict ourselves to supersymmetric theories in $d = 3 + 1$ spacetime dimensions. There are, however, many interesting things to say about supersymmetric theories in other dimensions. Here we merely make a few very simple comments.

Supersymmetric Gauge Theories in Different Dimensions

We've seen that the vector multiplet of $\mathcal{N} = 1$ supersymmetry has a photon paired with a single massless Weyl spinor. This works because both have two internal degrees of freedom in $d = 3 + 1$ dimensions. We can ask: in what other spacetime dimensions might we be able to pair a photon with a fermion?

The number of polarisation states of a photon is $d - 2$. So the question really is: in what dimensions does a spinor have $d - 2$ degrees of freedom? We will see that we

can have a supersymmetric theory in which a photon pairs with a single fermion in $d = 3, 4, 6$ and 10 Lorentzian spacetime dimensions.

The story is simplest in $d = 3 + 1$ and $d = 5 + 1$. In even spacetime dimension d, a Dirac spinor has $2^{d/2}$ complex components. But the irreducible representations of the Lorentz group are Weyl spinors with $2^{(d-2)/2}$ complex components. While a complex scalar has two degrees of freedom, a complex spinor has the same number of degrees of freedom as the number of components. This is because the Dirac equation (or Weyl equation) is first order so these components include both "position" and "momentum". This means that if we want the number of degrees of freedom of a Weyl spinor to match those of a photon then we need to solve the equation

$$2^{(d-2)/2} = d - 2$$

The solutions are $d = 4$ and $d = 6$ as advertised.

In $d = 3 + 1$ dimensions we can choose to impose *either* a Majorana condition *or* a chiral projection to a Weyl fermion. However in $d = 2 \bmod 8$ spacetime dimensions, it is possible to impose both a Majorana and Weyl condition. This halves the number of degrees of freedom of a Weyl fermion. Attempting to match the degrees of freedom of a Majorana-Weyl fermion to a photon we have

$$2^{(d-4)/2} = d - 2 \quad \text{with } d = 2 \bmod 8$$

The unique solution is $d = 10$.

Finally we're left searching solutions in odd spacetime dimensions. It is not hard to see that there is just one possibility. In $d = 2 + 1$ dimensions, a photon has just a single polarisation state. Meanwhile, a Dirac spinor in $d = 2 + 1$ has two complex components. However we can impose a Majorana condition to make the spinor real. (For example, we can take the real Clifford algebra $\gamma^0 = i\sigma^2$, $\gamma^1 = \sigma^2$ and $\gamma^2 = \sigma^3$.) So a Majorana spinor in $d = 2 + 1$ has two real components and, correspondingly, one degree of freedom, matching that of the photon.

If we're not in the magic spacetime dimension $d = 3, 4, 6$ or 10 then we can still have supersymmetric theories that relate a photon to a fermion. But now we need to include extra scalar degrees of freedom as well to make up the numbers.

The fact that the number of fermion degrees of freedom increases exponentially with d, while the number of bosonic degrees of freedom increases only linearly, suggests that there may be a maximum spacetime dimension in which supersymmetry is possible.

Indeed this is the case. If we don't wish to get our hands dirty with supergravity then $d = 9 + 1$ dimensions is the highest we can go. If we're happy to include gravity in the mix then there is a unique supersymmetry theory in $d = 10 + 1$ dimensions known, reasonably enough, as *eleven dimensional supergravity*. It is extremely interesting and describes the low-energy behaviour of M-theory.

Extended Supersymmetry and Higher Dimensions

There is a close relationship between supersymmetric theories in higher dimensions and extended supersymmetry. In particular, theories with $\mathcal{N} = 2$ supersymmetry naturally descend from $d = 5 + 1$ dimensions while those with $\mathcal{N} = 4$ supersymmetry come from $d = 9 + 1$ dimensions. (This statement, taken at face value, is true only at the classical level. But there are also a myriad of subtle and wonderful connections at the quantum level, none of which will be touched upon in these lectures.)

To see this, we will briefly jump ahead of ourselves slightly and use the language of fields, rather than the language of single particle quantum states that we've invoked until now. The relationship between theories in different dimensions involves a process known as *dimensional reduction*. This means that we take the fields in a higher dimension and state, by fiat, that they are independent of certain spatial coordinates. For example, consider a gauge field A_M in, say, $d = 5 + 1$ dimensions. This means that $M = 0, 1, \ldots, 5$. Upon dimensional reduction, we insist that this gauge field only depends on x^μ with $\mu = 0, 1, 2, 3$. The gauge field itself then decomposes as

$$A_M \to (A_\mu, \phi_4, \phi_5)$$

That is, we get a $d = 3+1$ dimensional gauge field A_μ together with two real scalars ϕ_4 and ϕ_5. But this is precisely the bosonic content of the $\mathcal{N} = 2$ vector multiplet that we found above. A $d = 5 + 1$ Weyl fermion decomposes into two $d = 3 + 1$ Weyl fermions in a similar fashion (although you have to work a little harder playing around with the gamma matrices to see this).

Playing the same game with a $d = 9 + 1$ gauge field, we find a $d = 3 + 1$ gauge field together with $10 - 4 = 6$ scalars. This is the bosonic content of the $\mathcal{N} = 4$ vector multiplet that we found above. Decomposing a $d = 9 + 1$ Majorana-Weyl fermion completes the story, giving four $d = 3 + 1$ Weyl fermions.

Finally, if you dimensionally reduce eleven dimensional supergravity you find $\mathcal{N} = 8$ supergravity in $d = 3 + 1$ dimensions.

Counting Supersymmetries

The way in which we count supersymmetries in different dimensions can be rather bewildering when you first meet it. In $d = 3 + 1$ we count supersymmetries by the number of Weyl spinor supercharges Q^I_α with $I = 1, \ldots, \mathcal{N}$. But this is clearly specific to 4d. In other dimensions the counting depends on what kinds of minimal spinors we can construct. Moreover, if we dimensionally reduce then what is a minimal supersymmetry in a higher dimension typically becomes an extended supersymmetry in a lower dimension.

To avoid this confusion, it can be useful to count the number of *components* of the supercharges. We count these as N (rather than the calligraphic $\mathcal{N}$.) These components are, sadly, also referred to a supercharges! Because spinors can be real in some dimensions, we count the number of real components or, equivalently, twice the number of complex components. This means that, in $d = 3+1$ dimensions, $\mathcal{N} = 1$ supersymmetry has four supercharges, $\mathcal{N} = 2$ has eight supercharges, and so on.

To orient you, here is a list of some of the most interesting classes of supersymmetric theories and how they are labelled in various dimensions. The list is by no means complete but gives some sense of the more compelling supersymmetric stories out there.

The maximum number of supercharges is $N = 32$. These are all supergravity theories and can exist in any dimension $d = 10 + 1$ and below. Upon dimensional reduction, the number of minimal spinor supercharges $\mathcal{N}$ in various dimensions is

N=32 supercharges:

Dimension d	11	10	6	4
Supersymmetry $\mathcal{N}$	1	(1,1)	(2,2)	8

This is not an exhaustive list: supersymmetric theories with $N = 32$ supercharges exist in all dimension $d \leq 11$. But the dimensions listed above are, for various reasons, the most interesting and well studied.

Note the strange (n, n) notation in $d = 5 + 1$ and $d = 9 + 1$. This is because of one more subtlety of representations of the Clifford algebra. When $d = 2$ mod 4, the two types of Weyl spinor are *not* related by complex conjugation in Lorentzian signature. This means that you can have a spinor of one chirality without necessarily having the other. In contrast, when $d = 0$ mod 4 (including, as we saw in great detail, in $d = 3+1$) the complex conjugate of a left-handed spinor is a right-handed spinor, so if you have one then you always have the other. The notation (n, n) tells us how many left- and right-handed spinor supercharges we have.

There is another supergravity theory in $d = 9 + 1$ dimension which has also 32 supercharges but with $\mathcal{N} = (2, 0)$ supersymmetry. This is more commonly known as Type IIB supergravity, with the $\mathcal{N} = (1, 1)$ theory known as Type IIA. They are the low-energy descriptions of Type IIA and IIB string theories.

Theories with $N = 16$ supercharges can exist in dimensions $d = 9 + 1$ and below. Upon dimensional reduction, the associated supersymmetry is:

N=16 supercharges:

Dimension d	10	6	4	3	2
Supersymmetry $\mathcal{N}$	(1,0)	(1,1)	4	8	(8,8)

The most famous and well studied of these is the Yang-Mills theory associated to the $\mathcal{N} = 4$ vector multiplet in $d = 3 + 1$. It has many remarkable properties, including electromagnetic duality and the fact that, at strong coupling, it is can be viewed as a theory of quantum gravity through the AdS/CFT correspondence. There are also interesting stories to tell about the quantum dynamics of the theories in $d = 2 + 1$ and $d = 1 + 1$ dimensions.

There is one further interesting theory with 16 supercharges. This is a strongly interacting superconformal quantum field theory in $d = 5 + 1$ dimensions with $\mathcal{N} = (2, 0)$ supersymmetry. In some ways, it can be viewed as the grandfather of all quantum field theories. Given its importance, it has a remarkably rubbish name: it is simply called the $(2, 0)$ theory.

Theories with $N = 8$ supercharges exist in $d = 5 + 1$ dimensions and below. Upon dimensional reduction, the names of the supersymmetries that one finds are

N=8 supercharges:

Dimension d	6	4	3	2
Supersymmetry $\mathcal{N}$	(1,0)	2	4	(4,4)

Again, the theories with $\mathcal{N} = 2$ supersymmetry in $d = 3 + 1$ dimensions are the best studied and were first solved by Seiberg and Witten.

Theories with $N = 4$ supercharges exist in $d = 3 + 1$ dimensions and below. Upon dimensional reduction, this becomes

N=4 supercharges:

Dimension d	4	3	2
Supersymmetry $\mathcal{N}$	1	2	(2,2)

Much of the focus of these lectures notes will be on understanding the dynamics of $\mathcal{N} = 1$ theories in $d = 3 + 1$ dimensions. But there are many beautiful stories in lower

dimensions as well. In particular, the study of superconformal $\mathcal{N} = (2,2)$ theories in $d = 1+1$ dimensions is where one can first find the mathematical study of mirror symmetry. There are also interesting 2d theories with $\mathcal{N} = (0,4)$ supersymmetry.

Finally, theories with $N = 2$ supercharges exist in $d = 2+1$ dimensions and below. The dimensional reduction to $d = 1+1$ gives

N=2 supercharges:

Dimension d	3	2
Supersymmetry $\mathcal{N}$	1	(1,1)

There are also $\mathcal{N} = (0,2)$ theories that do not descend from $d = 2+1$ dimensions. Note that these are usually written as $(0,2)$ rather than $(2,0)$ to give an extra hint that we're talking about 2d theories rather than the 6d theory mentioned above.

I've not included $d = 0+1$ theories in the above list, also known as quantum mechanics, but it's not for want of things to say. You can read about supersymmetric quantum mechanics in the companion lecture notes.

3 Chiral Superfields

In the previous section we've understood how supersymmetry acts on single particles states in the Hilbert space. But, ultimately, we want to write down field theories that are invariant under supersymmetry. Part of this requires understanding how supersymmetry acts on fields.

We've already seen a taster of this in the introduction. The action (1.1) was given by

$$S = \int d^4x \left[\partial_\mu \phi^\dagger \partial^\mu \phi - i\bar{\psi}\bar{\sigma}^\mu \partial_\mu \psi - \left| \frac{\partial W}{\partial \phi} \right|^2 - \frac{1}{2}\frac{\partial^2 W}{\partial \phi^2}\psi\psi - \frac{1}{2}\frac{\partial^2 W^\dagger}{\partial \phi^{\dagger\,2}}\bar{\psi}\bar{\psi} \right] \tag{3.1}$$

This involves a complex scalar ϕ and a single Weyl fermion ψ_α. After our discussion in the last section, we now recognise this as the fields corresponding to a chiral multiplet. We claimed in the introduction that this action is invariant under the transformation

$$\delta\phi = \sqrt{2}\epsilon\psi \quad \text{and} \quad \delta\psi_\alpha = \sqrt{2}i\sigma^\mu_{\alpha\dot{\alpha}}\bar{\epsilon}^{\dot{\alpha}}\,\partial_\mu\phi - \sqrt{2}\epsilon_\alpha \frac{\partial W^\dagger}{\partial \phi^\dagger} \tag{3.2}$$

There are a few questions that we'd like to ask. First: how can we construct actions like (3.1)? After all, it's not like we can just stare at the action and see that it's invariant under the transformations (3.2). It takes a bit of work to show this. Secondly, how are the transformations (3.2) related to the supercharges and supersymmetry algebra that we met in the previous section.

The purpose of this section is to answer these questions. In particular, we'll see how we can rewrite the action (3.1) in a way that the supersymmetry is manifest. The trick to doing this is to combine the bosonic field ϕ and the femionic field ψ_α into a single object known as a *superfield*.

3.1 Superspace

Usually, fields are functions of x^μ, the coordinates of Minkowski space. But, as we've seen, supersymmetry is an extension of the Poincaré group. Correspondingly, superfields live not on Minkowski space, but on an extension of Minkowski space known as *superspace*.

The coordinates of superspace are

$$x^\mu \ , \quad \theta_\alpha \ , \quad \bar{\theta}^{\dot{\alpha}}$$

Here x^μ, with $\mu = 0, 1, 2, 3$ are the coordinates of Minkowski space. In superspace these are augmented with Grassmann-valued spinors θ_α and $\bar{\theta}^{\dot{\alpha}}$. In other words, superspace is not a regular manifold of the kind that we know and love from courses on differential geometry. Instead it is an example of a supermanifold, with both commuting and anti-commuting dimensions.

3.1.1 The Geometry of Superspace

In what follows, we'll explore the idea of fields on superspace and see how they encapsulate a collection of fields that transform into each other under supersymmetry. However, we could reasonably ask: how did we come up with the idea of superspace in the first place? There is, it turns out, a group theoretic answer to this.

In general, if we're given a Lie group G, we might want to know what manifolds $\mathcal{M}$ accommodate a natural action of G.

One obvious choice is to take the manifold to be the group itself: $\mathcal{M} = G$. In this case, each element $g \in G$ gives us natural map $\mathcal{M} \mapsto \mathcal{M}$ given by $g' \in \mathcal{M} \mapsto g \cdot g'$.

A slightly less obvious choice is to take a *coset* space. This is the manifold $\mathcal{M} = G/H$ where $H \subset G$ is a subgroup of G. A point $\{g\}$ in the coset G/H is defined by the equivalence relation among elements of G

$$g \equiv g \cdot h \quad \text{for all } h \in H$$

Again, any element $g \in G$ gives us a natural map $\mathcal{M} = G/H \mapsto G/H$ defined by $\{g'\} \in \mathcal{M} \mapsto \{g \cdot g'\}$.

For example, the group $G = SU(2)$ is, as a manifold, $G = \mathbf{S}^3$. We can consider the subgroup $H = U(1) \subset SU(2)$ to get the coset $SU(2)/U(1) \cong \mathbf{S}^2$. (Mathematically, this is known as the *Hopf fibration*.) Obviously there is a natural action of $SO(3) \cong SU(2)/\mathbf{Z}_2$ on $\mathbf{S}^2$.

This, somewhat abstract, way of thinking gives us a new perspective on Minkowski space itself. It can be viewed as the coset

$$\mathbb{R}^{1,3} = G/H = \frac{\text{Poincaré Group}}{\text{Lorentz Group}}$$

Here a general element of the Poincaré group G is comprised of Lorentz boosts, generated by $M^{\mu\nu}$, and translations generated by P^μ. We write this as

$$g(\omega, a) = \exp\left(-\frac{i}{2}\omega_{\mu\nu}M^{\mu\nu} + ia_\mu P^\mu\right)$$

Meanwhile, the Lorentz group H consists only of Lorentz boosts. This means that coset space can be parameterised just by a^μ which we can equivalently think of as coordinates $x^\mu = a^\mu$ on Minkowski space. The fact that Minkowski space can be viewed as a coset merely confirms something that we knew already: there is an action of the Poincaré group on Minkowski space.

Now, however, we would like to construct a space on which the group of supersymmetry transformations naturally acts. These are given by

$$g(\omega, a, \theta, \bar{\theta}) = \exp\left(-\frac{i}{2}\omega_{\mu\nu}M^{\mu\nu} + ia_\mu P^\mu + i\theta^\alpha Q_\alpha + i\bar{\theta}_{\dot{\alpha}}\bar{Q}^{\dot{\alpha}}\right) \tag{3.3}$$

with Q_α and $\bar{Q}^{\dot{\alpha}}$ the supersymmetry generators that we met in the previous section. The spinors θ^α and $\bar{\theta}_{\dot{\alpha}}$ should be viewed as parameterising the "amount" of supersymmetry transformation that we're doing, albeit with the "amount" now somewhat harder to quantify as it's a Grassmann valued object. With Grassmann elements of this kind, g is an element of a *super Lie group* which, in this case, is known as the *super-Poincaré group*. The coset construction continues to work in the same way and we define superspace to be

$$\text{Superspace} = G/H = \frac{\text{Super-Poincaré Group}}{\text{Lorentz Group}}$$

A point in superspace is now parameterised by $x^\mu = a^\mu$ and the Grassmann-valued spinors θ_α and $\bar{\theta}^{\dot{\alpha}}$ as advertised above.

Before we go on, a quick comment on nomenclature. The Lorentz group is, of course, $SO(1,3)$. (Actually, strictly speaking if we want to include spinor representations it is $SL(2,\mathbb{C}) = \text{Spin}(1,3)$ but we'll ignore this double cover subtlety.) The Poincaré group is the semi-direct product $ISO(1,3) = SO(1,3) \ltimes \mathbb{R}^4$ and Minkowski space is $\mathbb{R}^{1,3} = ISO(1,3)/SO(1,3)$. Meanwhile, the super-Poincaré group is usually written as $ISO(1,3|1)$ with the additional "bar 1" or "slash 1" telling us that we have $\mathcal{N} = 1$ supersymmetry. Superspace is then the "4+4" dimensional supermanifold $\mathbb{R}^{1,3|4} = ISO(1,3|1)/SO(1,3)$. We'll have no need for any of this notation in these lectures.

The Action on Superspace

The whole point of the coset construction of superspace is that it tells us how the supergroup acts. This will be important in what follows so let's flesh it out a little. First, we write the general element of the supergroup (3.3) as

$$g(\omega, x, \theta, \bar{\theta}) = \tilde{g}(x, \theta, \bar{\theta})h(\omega)$$

where $h(\omega)$ is a Lorentz transformation and $\tilde{g}(a,\theta,\bar{\theta})$ is the representative of the coset

$$\tilde{g}(x,\theta,\bar{\theta}) = \exp\left(ix_\mu P^\mu + i\theta^\alpha Q_\alpha + i\bar{\theta}_{\dot{\alpha}}\bar{Q}^{\dot{\alpha}}\right)$$

This specifies a point $(x,\theta,\bar{\theta})$ in superspace,

We now want to see how the momentum operator P and supercharges Q and $\bar{Q}$ shift the point $(x,\theta,\bar{\theta})$ in superspace. Let's start with the momentum operator. We introduce the supergroup element

$$U(a) = \exp\left(ia_\mu P^\mu\right)$$

Then we have

$$U(a)\,\tilde{g}(x,\theta,\bar{\theta}) = e^{iaP}e^{ixP+i\theta Q+i\bar{\theta}\bar{Q}} = e^{i(x+a)P+i\theta Q+i\bar{\theta}\bar{Q}} = \tilde{g}(x+a,\theta,\bar{\theta})$$

This gives us a familiar result: momentum generates translations,

$$x^\mu \to x^\mu + a^\mu$$

Now we do the same for the supercharges. This time we will find a small twist to the story. We introduce the supergroup element

$$V(\epsilon,\bar{\epsilon}) = \exp\left(i\epsilon^\alpha Q_\alpha + i\bar{\epsilon}_{\dot{\alpha}}\bar{Q}^{\dot{\alpha}}\right)$$

Note that ϵ^α and $\bar{\epsilon}_{\dot{\alpha}}$ are Grassmann-valued spinors. They shouldn't be confused with the anti-symmetric $\epsilon^{\alpha\beta}$ matrices that we met earlier. (Sorry!) Now the action on superspace is given by

$$V(\epsilon,\bar{\epsilon})\,\tilde{g}(x,\theta,\bar{\theta}) = e^{i\epsilon Q+i\bar{\epsilon}\bar{Q}}e^{ixP+i\theta Q+i\bar{\theta}\bar{Q}} \tag{3.4}$$

The small twist is that Q and $\bar{Q}$ do not anti-commute with each other. In fact, now that we've multiplied the supercharges with anti-commuting spinors ϵ and θ, we can talk about commutation relations rather than anti-commutation relations. We have

$$\begin{aligned} Q_\alpha\bar{Q}_{\dot{\alpha}} + \bar{Q}_{\dot{\alpha}}Q_\alpha = 2\sigma^\mu_{\alpha\dot{\alpha}}P_\mu \quad &\Rightarrow \quad \epsilon^\alpha\left(Q_\alpha\bar{Q}_{\dot{\alpha}} + \bar{Q}_{\dot{\alpha}}Q_\alpha\right)\bar{\theta}^{\dot{\alpha}} = 2(\epsilon^\alpha\sigma^\mu_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}})P_\mu \\ &\Rightarrow \quad [\bar{\theta}^{\dot{\alpha}}\bar{Q}_{\dot{\alpha}}, \epsilon_\alpha Q^\alpha] = 2(\epsilon\sigma^\mu\bar{\theta})P_\mu \end{aligned} \tag{3.5}$$

where the Grassmann nature of $\bar{\theta}$, ϵ, Q and $\bar{Q}$ means that we pick up a minus sign in going from the first line to the second, turning $\{\ ,\ \}$ into $[\ ,\ ]$.

We now evaluate (3.4) using the BCH formula

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\ldots}$$

The commutator (3.5), together with the fact that the higher commutator terms ... in the BCH formula all vanish in the present case, gives us the result

$$\begin{aligned} V(\epsilon,\bar\epsilon)\,\tilde g(x,\theta,\bar\theta) &= e^{ixP+i(\theta+\epsilon)Q+i(\bar\theta+\bar\epsilon)\bar Q+(\epsilon\sigma\bar\theta)P-(\theta\sigma\bar\epsilon)P} \\ &= \tilde g(x+i\theta\sigma\bar\epsilon-i\epsilon\sigma\bar\theta,\theta+\epsilon,\bar\theta+\bar\epsilon) \end{aligned}$$

Here we see the twist. The supercharges shift the Grassmann coordinate in superspace as we might have anticipated. But, at the same time, they also shift the point in Minkowski space by a Grassmann bilinear

$$\begin{aligned} x^\mu &\to x^\mu + i\theta\sigma^\mu\bar\epsilon - i\epsilon\sigma^\mu\bar\theta \\ \theta &\to \theta+\epsilon \\ \bar\theta &\to \bar\theta+\bar\epsilon \end{aligned} \tag{3.6}$$

Note that the shift in x^μ due to the Grassmann bilinear can't be thought of as normal translation by some number. Instead, it's a more formal expression. Ultimately, we'll see how this manifests itself in terms of the superfields and their more familiar components.

3.1.2 Superfields

A superfield is a function on superspace, $Y = Y(x,\theta,\bar\theta)$. To start, we take this to be a complex-valued function on superspace.

In principle, the superfield could transform in some non-trivial representation of the Lorentz group. For example it could carry a vector index μ or a spinor index α. However, rather remarkably, we will find all the fields that we need – scalar, spinor and vector – lurking within the simplest scalar superfield. (We will, however, come across superfields carrying spinor indices in Section 4.)

To see this, we Taylor expand the superfield in θ and $\bar\theta$. But this is easy because θ and $\bar\theta$ are Grassmann valued objects obeying, for example,

$$\theta_\alpha\theta_\beta = -\theta_\beta\theta_\alpha$$

This means that the Taylor expansion truncates after some finite length. In particular we have $\theta_\alpha\theta_\beta\theta_\gamma = 0$. So the Taylor expansion of $Y(x,\theta,\bar\theta)$ stops after terms quadratic

in θ and $\bar{\theta}$. Expanding the superfield out in this way then reveals a bunch of more familiar fields lurking within,

$$\begin{aligned} Y(x,\theta,\bar{\theta}) = \phi(x) + \theta^\alpha \psi_\alpha(x) + \bar{\theta}_{\dot{\alpha}} \bar{\chi}^{\dot{\alpha}}(x) + \theta^2 M(x) + \bar{\theta}^2 N(x) \\ + \theta^\alpha \bar{\theta}^{\dot{\alpha}} V_{\alpha\dot{\alpha}}(x) + \theta^2 \bar{\theta}_{\dot{\alpha}} \bar{\lambda}^{\dot{\alpha}}(x) + \bar{\theta}^2 \theta^\alpha \rho_\alpha(x) + \theta^2 \bar{\theta}^2 D(x) \end{aligned} \tag{3.7}$$

Here $\theta^2 = \theta^\alpha \theta_\alpha$ and $\bar{\theta}^2 = \bar{\theta}_{\dot{\alpha}} \bar{\theta}^{\dot{\alpha}}$.

There are a few things to say about this. First, note that the superfield does indeed contain all the fields that we usually care about: there are four complex scalars ϕ, M, N and D, two left-handed spinors ψ and ρ, two right-handed spinors $\bar{\chi}$ and $\bar{\lambda}$ and a vector $V_{\alpha\dot{\alpha}} = \sigma^\mu_{\alpha\dot{\alpha}} V_\mu$.

Second, note that it contains many more fields that we might have thought from our analysis in the previous section! The representations on single particle states suggested that there should be a chiral multiplet containing a single complex scalar and a Weyl fermion and a vector multiplet containing a gauge field and a Weyl fermion. Yet the superfield Y contains a plethora of such fields. We will shortly see how we can impose further restrictions on Y that truncate the number of fields lying within to match our earlier expectation.

Our next task is to understand how superfields transform under supersymmetry transformations. We'll again start with translations $x^\mu \to x^\mu + a^\mu$ which, as we have seen, are generated by the unitary operator

$$U = \exp\left(i a^\mu P_\mu\right)$$

Previously, we viewed this as a group element acting on superspace. But in quantum field theory, it has another avatar as an operator acting on the Hilbert space. The fields in quantum field theory are, of course, also operators and the superfield is no different. The action of U on such operators enacts the translation, meaning

$$U Y(x,\theta,\bar{\theta}) U^\dagger = Y(x+a,\theta,\bar{\theta})$$

For infinitesimal a^μ, we expand $U = e^{iaP} = 1 + i a_\mu P^\mu + \mathcal{O}(a)^2$. We also Taylor expand the field, $Y(x+a) = Y(x) + a^\mu \partial_\mu Y(x) + \mathcal{O}(a^2)$. Equating the terms linear in a we see that the translations are captured in the commutation relation on fields

$$[P_\mu, Y] = -i\partial_\mu Y \tag{3.8}$$

We can treat the action of the supercharges in a similar fashion. We again have the unitary operator

$$V(\epsilon,\bar{\epsilon}) = \exp\left(i\epsilon^\alpha Q_\alpha + i\bar{\epsilon}_{\dot{\alpha}} \bar{Q}^{\dot{\alpha}}\right)$$

Acting on superfields, this gives

$$VY(x,\theta,\bar\theta)V^\dagger = Y(x + i\theta\sigma^\mu\bar\epsilon - i\epsilon\sigma^\mu\bar\theta, \theta+\epsilon, \bar\theta+\bar\epsilon)$$

where we've invoked the transformation of the superspace coordinate (3.6). If we now treat ϵ_α as an infinitesimal spinor and work to leading order in ϵ, we find the commutation relations

$$[Q_\alpha, Y] = \left(-i\frac{\partial}{\partial\theta^\alpha} - \sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu\right) Y \tag{3.9}$$

$$[\bar Q_{\dot\alpha}, Y] = \left(+i\frac{\partial}{\partial\bar\theta^{\dot\alpha}} + \theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu\right) Y \tag{3.10}$$

In this expression, the derivatives with respect to Grassmann coordinates are defined by

$$\partial_\alpha = \frac{\partial}{\partial\theta^\alpha} \quad \text{with} \quad \partial_\alpha\theta^\beta = \delta^\beta_\alpha \ \text{ and } \ \partial_\alpha\bar\theta_{\dot\beta} = 0$$
$$\bar\partial_{\dot\alpha} = \frac{\partial}{\partial\bar\theta^{\dot\alpha}} \quad \text{with} \quad \bar\partial_{\dot\alpha}\bar\theta^{\dot\beta} = \delta^{\dot\beta}_{\dot\alpha} \ \text{ and } \ \bar\partial_{\dot\alpha}\theta_\beta = 0$$

These Grassmann derivatives are themselves Grassmann. This means that they pick up a minus sign when they pass through other Grassmann variables. So, for example, if you wish to differentiate $\chi^\beta\theta^\gamma$, where both χ and θ are Grassmann variables, then you have

$$\frac{\partial}{\partial\chi^\alpha}(\chi^\beta\theta^\gamma) = \delta^\beta_\alpha\theta^\gamma \quad \text{and} \quad \frac{\partial}{\partial\theta^\alpha}(\chi^\beta\theta^\gamma) = -\delta^\gamma_\alpha\chi^\beta$$

where that extra minus sign in the second expression comes from dragging the $\partial/\partial\theta^\alpha$ through the χ^β before it gets to attack its prey.

It's useful to define differential operators associated to the right-hand sides of (3.8), (3.9) and (3.10). To this end, we write

$$\begin{aligned}\mathcal{P}_\mu &= -i\partial_\mu \\ \mathcal{Q}_\alpha &= -i\partial_\alpha - \sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu \\ \bar{\mathcal{Q}}_{\dot\alpha} &= +i\bar\partial_{\dot\alpha} + \theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu\end{aligned} \tag{3.11}$$

Be warned: these differ from the operators P_μ, Q_α and $\bar Q_{\dot\alpha}$ only by the use of curly calligraphic script. You can check that anti-commutation relation of these differential operators is something familiar

$$\{\mathcal{Q}_\alpha, \bar{\mathcal{Q}}_{\dot\alpha}\} = 2\sigma^\mu_{\alpha\dot\alpha}\mathcal{P}_\mu$$

together with $\{\mathcal{Q}_\alpha, \mathcal{Q}_\beta\} = \{\bar{\mathcal{Q}}_{\dot\alpha}, \bar{\mathcal{Q}}_{\dot\beta}\} = 0$. This is telling us that $\mathcal{P}$, $\mathcal{Q}_\alpha$ and $\bar{\mathcal{Q}}_{\dot\alpha}$ also furnish a representation of the supersymmetry algebra, now acting on fields on superspace

Supersymmetry Transformation of Fields

We can unpack the supersymmetry transformations (3.9) and (3.10) to see how it acts on the individual fields sitting with Y. The infinitesimal change of the superfield is defined to be

$$\delta Y = i[\epsilon Q + \bar{\epsilon}\bar{Q}, Y] = i(\epsilon\mathcal{Q} + \bar{\epsilon}\bar{\mathcal{Q}})Y \tag{3.12}$$

Expanding out Y in terms of the components (3.7), the operators $\mathcal{Q}$ and $\bar{\mathcal{Q}}$ act on each term. $\mathcal{Q}$ removes a θ (where there is one) and adds a $\bar{\theta}\partial_\mu$ (where there aren't too many $\bar{\theta}$'s already) Obviously $\bar{\mathcal{Q}}$ is the conjugate. We then compare the various θ and $\bar{\theta}$ and terms.

For example, the lowest term in Y is the scalar $\phi(x)$. To compute its variation, we look for the term in δY with neither θ's nor $\bar{\theta}$'s. This comes from ∂_α acting on the term $\theta\psi$ and $\bar{\partial}_{\dot{\alpha}}$ acting on $\bar{\theta}\bar{\chi}$. The result is

$$\delta\phi = \epsilon\psi + \bar{\epsilon}\bar{\chi} \tag{3.13}$$

Meanwhile, the highest term in Y is the scalar $D(x)$. To compute its variation, we find the term in δY that comes with the full complement of $\theta^2\bar{\theta}^2$. This happens comes from the $\bar{\theta}\partial_\mu$ term in $\mathcal{Q}$ and the $\theta\partial_\mu$ term in $\bar{\mathcal{Q}}$. The net effect is that the variation of $D(x)$ is a total derivative

$$\delta D = \frac{i}{2}\partial_\mu(\epsilon\sigma^\mu\bar{\lambda} - \rho\sigma^\mu\bar{\epsilon}) \tag{3.14}$$

This will prove to be part of the story as we proceed.

It takes a bit of work to get the transformation of all the remaining component fields in (3.7). You'll have the pleasure of doing this work in the first examples sheet. The answer turns out to be

$$\begin{aligned}
\delta\psi &= 2\epsilon M + (\sigma^\mu\bar{\epsilon})(i\partial_\mu\phi + V_\mu) \\
\delta\bar{\chi} &= 2\bar{\epsilon}N - (\epsilon\sigma^\mu)(i\partial_\mu\phi - V_\mu) \\
\delta M &= \bar{\epsilon}\bar{\lambda} - \frac{i}{2}\partial_\mu\psi\sigma^\mu\bar{\epsilon} \\
\delta N &= \epsilon\rho + \frac{i}{2}\epsilon\sigma^\mu\partial_\mu\bar{\chi} \\
\delta V_\mu &= \epsilon\sigma_\mu\bar{\lambda} + \rho\sigma_\mu\bar{\epsilon} + \frac{i}{2}\left(\partial^\nu\psi\sigma_\mu\bar{\sigma}_\nu\epsilon - \bar{\epsilon}\bar{\sigma}_\nu\sigma_\mu\partial^\nu\bar{\chi}\right) \\
\delta\bar{\lambda} &= 2\bar{\epsilon}D + \frac{i}{2}\bar{\sigma}^\nu\sigma^\mu\bar{\epsilon}\,\partial_\mu V_\nu + i\bar{\sigma}^\mu\epsilon\,\partial_\mu M \\
\delta\rho &= 2\epsilon D - \frac{i}{2}\sigma^\nu\bar{\sigma}^\mu\epsilon\,\partial_\mu V_\nu + i\sigma^\mu\bar{\epsilon}\,\partial_\mu N
\end{aligned} \tag{3.15}$$

The variation of each has at least two terms, one with a derivative ∂_μ and one without.

3.1.3 Constraining Superfields

As we already commented, the superfield Y is too big. It has way more fields than we expect from the representation theory of Section 2.3. This is because Y is not an irreducible representation. It can be reduced to something smaller. The question is: how?

We want to impose constraints on Y such that it remains a superfield. That means that whatever object we have after the constraint should also transform as (3.9) and (3.10) under supersymmetry transformations. So our first step to understanding the possible constraints is to figure out what kind of operations we can perform on superfields that keep them as superfields.

There are some obvious operations, albeit ones that won't help with our constraint. If we have two superfields Y_1 and Y_2 then αY_1 is a superfield for any $\alpha \in \mathbb{C}$, as is $Y_1 + Y_2$ and $Y_1 Y_2$. For example, to see that $Y_1 Y_2$ is a superfield, we need to note that

$$[Q_\alpha, Y_1 Y_2] = [Q_\alpha, Y_1]Y_2 + Y_1[Q_\alpha, Y_2] = (\mathcal{Q}_\alpha Y_1)Y_2 + Y_1(\mathcal{Q}_\alpha Y_2) = \mathcal{Q}_\alpha(Y_1 Y_2)$$

as required.

More pertinent for our purposes, if Y is a superfield then so too is $\partial_\mu Y$. However, crucially, neither $\partial_\alpha Y$ nor $\bar{\partial}_{\dot\alpha} Y$ are superfields. Algebraically, this is because

$$[\epsilon^\alpha Q_\alpha, \bar{\partial}_{\dot\alpha}] = \epsilon^\alpha \sigma^\mu_{\alpha\dot\alpha} \partial_\mu \neq 0$$

To build some intuition for what's going on, note that $\bar{\partial}_{\dot\alpha} Y$ doesn't include, for example, the highest component $\theta^2\bar\theta^2 D$ term; there was such a term in Y but one of the $\bar\theta$'s is removed after acting with $\bar{\partial}_{\dot\alpha}$. However, acting with a supercharge Q_α will generate such a term. In other words, it's not consistent with supersymmetry to simply state by fiat that the last term vanishes, $D(x) = 0$. Act with a supersymmetry transformation and this will no longer be true. It's analogous to setting $A_3 = 0$ in a vector field A_μ and thinking that you've found an object with just three components, only to realise that A_3 gets resurrected after a rotation.

However, there is a way forward. We define the *covariant derivatives*

$$\begin{aligned} \mathcal{D}_\alpha &= \partial_\alpha + i\sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu \\ \bar{\mathcal{D}}_{\dot\alpha} &= -\bar{\partial}_{\dot\alpha} - i\theta^\alpha \sigma^\mu_{\alpha\dot\alpha}\partial_\mu \end{aligned}$$

These are very similar to the $\mathcal{Q}_\alpha$ and $\bar{\mathcal{Q}}_{\dot\alpha}$ differential operators defined in (3.11), but with a relative minus sign difference (and an overall factor of i difference). Their key

property is that they anti-commute with $\mathcal{Q}$ and $\bar{\mathcal{Q}}$

$$\{\mathcal{D}_\alpha, \mathcal{Q}_\beta\} = \{\mathcal{D}_\alpha, \bar{\mathcal{Q}}_{\dot\beta}\} = \{\bar{\mathcal{D}}_{\dot\alpha}, \mathcal{Q}_\beta\} = \{\bar{\mathcal{D}}_{\dot\alpha}, \bar{\mathcal{Q}}_{\dot\beta}\} = 0 \tag{3.16}$$

The covariant derivatives also obey

$$\{\mathcal{D}_\alpha, \bar{\mathcal{D}}_{\dot\alpha}\} = 2\sigma^\mu_{\alpha\dot\alpha}\mathcal{P}_\mu \tag{3.17}$$

together with $\{\mathcal{D}_\alpha, \mathcal{D}_\beta\} = \{\bar{\mathcal{D}}_{\dot\alpha}, \bar{\mathcal{D}}_{\dot\beta}\} = 0$.

From (3.16), we have

$$[\epsilon\mathcal{Q} + \bar\epsilon\bar{\mathcal{Q}}\,,\, \mathcal{D}_\alpha] = [\epsilon\mathcal{Q} + \bar\epsilon\bar{\mathcal{Q}}\,,\, \bar{\mathcal{D}}_{\dot\alpha}] = 0$$

This tells us that both $\mathcal{D}_\alpha Y$ and $\bar{\mathcal{D}}_{\dot\alpha} Y$ are superfields. For example, under the supersymmetry transformation (3.12), we have

$$\delta Y = i(\epsilon\mathcal{Q} + \bar\epsilon\bar{\mathcal{Q}})Y \quad \Rightarrow \quad \delta(\mathcal{D}_\alpha Y) = i(\epsilon\mathcal{Q} + \bar\epsilon\bar{\mathcal{Q}})\mathcal{D}_\alpha Y$$

Now we can discuss the various constraints that we can place on a superfield Y. There are four of interest (of which, only three will play a major role in these lectures).

- A *chiral superfield* Φ is defined by the constraint

  $$\bar{\mathcal{D}}_{\dot\alpha}\Phi = 0$$

- An *anti-chiral superfield* Ψ is defined by the constraint

  $$\mathcal{D}_\alpha\Psi = 0$$

  Note that you can't impose both chiral and anti-chiral conditions since the anti-commutator (3.17) would then require that the superfield is actually constant. Moreover, if Φ is a chiral superfield then $\bar\Phi = \Phi^\dagger$ is an anti-chiral superfield. (I give a simple way to see this at the end of Section 3.1.4.) The fact that we can't impose both conditions simultaneously means that we can't take Φ to be real: chiral superfields are necessarily complex. We will see that chiral superfields correspond to the chiral multiplets that we met in Section 2.3.

- A *real superfield* V is defined by the simple requirement that

  $$V = V^\dagger$$

  We will postpone our discussion of real superfields to Section 4. There we will see that the real superfields correspond to the vector multiplet that we met in Section 2.3.

- Finally, a *linear superfield* J is defined

$$J = J^\dagger \quad \text{and} \quad \mathcal{D}^2 J = \bar{\mathcal{D}}^2 J = 0$$

These play a slightly less prominent role than the (anti)-chiral and real superfields. In particular, we won't build supersymmetry actions out of linear superfields. However, it turns out that they are useful homes for certain composite operators in quantum field theory, most notably Noether currents associated to global symmetries.

We will spend the rest of this section studying the properties of chiral superfields.

3.1.4 Chiral Superfields

A chiral superfield obeys the constraint

$$\bar{\mathcal{D}}_{\dot\alpha}\Phi = 0 \tag{3.18}$$

We will first solve this equation to understand what it means for the superfield Φ.

There's a useful trick here. We introduce the coordinate

$$y^\mu = x^\mu + i\theta\sigma^\mu\bar\theta$$

The advantage of this coordinate is that we have

$$\bar{\mathcal{D}}_{\dot\alpha} y^\mu = \left(-\bar\partial_{\dot\alpha} - i\theta^\alpha \sigma^\nu_{\alpha\dot\alpha}\partial_\nu\right)\left(x^\mu + i\theta^\beta\sigma^\mu_{\beta\dot\beta}\bar\theta^{\dot\beta}\right) = -i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha} - i\bar\partial_{\dot\alpha}(\theta^\beta\sigma^\mu_{\beta\dot\beta}\bar\theta^{\dot\beta}) = 0$$

where to see that the two terms cancel, you have to remember that you pick up an extra minus sign as the $\bar\partial_{\dot\alpha}$ passes through the θ^β. In addition, we have

$$\bar{\mathcal{D}}_{\dot\alpha}\theta_\beta = 0$$

This means that if we view a general superfield as a function of $\Phi = \Phi(y, \theta, \bar\theta)$ then, of the three arguments, only $\bar{\mathcal{D}}_{\dot\alpha}\bar\theta_{\dot\beta} \neq 0$ and the condition (3.18) tells us

$$\bar{\mathcal{D}}_{\dot\alpha}\Phi = 0 \quad \Rightarrow \quad \Phi = \Phi(y, \theta)$$

In other words Φ is almost a function only of θ and not of $\bar\theta$, the "almost" because there is in fact a $\bar\theta$ buried in the y^μ. This means that we can expand in components

$$\Phi(y, \theta) = \phi(y) + \sqrt{2}\theta\psi(y) + \theta^2 F(y)$$

where the $\sqrt{2}$ is a convention. We can then further Taylor expand the y^μ to get the expression for a chiral superfield in components

$$\begin{aligned}\Phi(x,\theta,\bar\theta) = \phi(x) + \sqrt{2}\theta\psi(x) + \theta^2 F(x) \\ + i\theta\sigma^\mu\bar\theta\partial_\mu\phi(x) - \frac{i}{\sqrt{2}}\theta^2\partial_\mu\psi(x)\,\sigma^\mu\bar\theta - \frac{1}{4}\theta^2\bar\theta^2\Box\phi(x)\end{aligned} \tag{3.19}$$

with $\Box = \partial_\mu\partial^\mu$. We see that the chiral superfield contains just three component fields: a complex scalar ϕ, a Weyl spinor ψ and another complex scalar F. The higher components of $\Phi(x)$ are simply derivatives of the first two fields.

This is much closer to what we expected based on our analysis in Section 2.3. There we found a chiral multiplet consists of single particle states associated to a complex scalar ϕ and a Weyl fermion ψ. However, we've also got a second complex scalar F. We will see later that this is an object known as an *auxiliary field*. For now it's worth noticing that, in contrast to ϕ and ψ, there are no terms in the chiral superfield with ∂F. This will be important as we proceed.

The supersymmetry transformations of the chiral multiplet are

$$\begin{aligned}\delta\phi &= \sqrt{2}\epsilon\psi \\ \delta\psi &= \sqrt{2}i\sigma^\mu\bar\epsilon\,\partial_\mu\phi + \sqrt{2}\epsilon F \\ \delta F &= \sqrt{2}i\bar\epsilon\bar\sigma^\mu\partial_\mu\psi\end{aligned} \tag{3.20}$$

Note that F transforms as a total derivative, just like D in the original unconstrained superfield (3.14). We'll see the relevance of this shortly.

There is a very similar story for the anti-chiral superfields. As we mentioned previously, these can be viewed as the complex conjugate of a chiral superfield. To see this, note that if a chiral superfield $\Phi(y,\theta)$ is function of y^μ and θ, then its conjugate $\Phi^\dagger(\bar y,\bar\theta)$ is a function of $\bar y^\mu = x^\mu - i\theta\sigma^\mu\bar\theta$ and $\bar\theta$. But it's simple to check that $\mathcal{D}_\alpha\bar y^\mu = \mathcal{D}_\alpha\bar\theta^{\dot\alpha} = 0$ and so $\Phi^\dagger$ is indeed an anti-chiral superfield obeying $\mathcal{D}_\alpha\Phi^\dagger = 0$. In components, we have

$$\Phi^\dagger(\bar y,\bar\theta) = \phi^\dagger(\bar y) + \sqrt{2}\bar\theta\bar\psi(\bar y) + \bar\theta^2 F^\dagger(\bar y)$$

We can then further expand out $\bar y$ further if we wish to get an expression analogous to (3.19),

$$\begin{aligned}\Phi^\dagger(x,\theta,\bar\theta) = \phi^\dagger(x) + \sqrt{2}\bar\theta\bar\psi(x) + \bar\theta^2 F^\dagger(x) \\ - i\theta\sigma^\mu\bar\theta\partial_\mu\phi^\dagger(x) + \frac{i}{\sqrt{2}}\bar\theta^2\theta\sigma^\mu\partial_\mu\bar\psi(x) - \frac{1}{4}\theta^2\bar\theta^2\Box\phi^\dagger(x)\end{aligned}$$

3.2 And... Action

To construct actions that are invariant under Poincaré group, we take suitable Lagrangian densities of fields and integrate them over spacetime. Analogously, to construct actions that are invariant under supersymmetry, we take suitable Lagrangian densities of superfields and integrate them over superspace.

3.2.1 Integrating Over Superspace

First, let's remind ourselves how Grassmann integration works. (It is, happily, much easier than normal integration!) If we have a single Grassmann variable θ then

$$\int d\theta\ 1 = 0 \quad \text{and} \quad \int d\theta\ \theta = 1$$

This means that if we have a function $f(x,\theta) = f_0(x) + \theta f_1(x)$, then Grassmann integration picks out the component multiplying θ,

$$\int d\theta\ f(x,\theta) = f_1(x)$$

In this manner, integration over Grassmann variables is the same thing as differentiation: $\int d\theta = \partial/\partial\theta$. In particular, we have a Grassmann version of the fundamental theorem of calculus

$$\int d\theta\ \frac{\partial f}{\partial\theta} = \int d\theta\ f_0(x) = 0 \tag{3.21}$$

Here we will need to integrate over superspace, parameterised by θ_α and $\bar\theta^{\dot\alpha}$. We define

$$\int d^2\theta = \frac{1}{2}\int d\theta^1 d\theta^2 \quad \text{and} \quad \int d^2\bar\theta = -\frac{1}{2}\int d\bar\theta^1 d\bar\theta^2$$

Those strange factors of $\frac{1}{2}$ are because $\theta^2 = \theta^\alpha\theta_\alpha = -2\theta^1\theta^2$. We then have

$$\int d^2\theta\ \theta^2 = -\int d\theta^1 d\theta^2\ (\theta^1\theta^2) = 1$$

where the minus sign disappears when $d\theta^2$ moves past θ^1. Note that the measure $d^2\bar\theta$ comes with an extra minus sign but this cancels the corresponding minus sign in $\bar\theta^2 = \bar\theta_{\dot\alpha}\bar\theta^{\dot\alpha} = +2\bar\theta^1\bar\theta^2$. Once again, we have $\int d^2\bar\theta\ \bar\theta^2$. Finally, we also use the (not entirely logical) notation

$$\int d^4\theta = \int d^2\theta\, d^2\bar\theta$$

Now suppose that we build an action out of some function of superfields. That function will itself be a superfield that we will call $K(x, \theta, \bar{\theta})$ but, in contrast to what we've discussed so far, we'll view K as a composite superfield whose component are functions of other fields. We the construct the action of the form

$$S = \int d^4x\, d^4\theta\ K(x, \theta, \bar{\theta}) \tag{3.22}$$

The action is real if K is a real superfield, obeying $K = K^\dagger$. As we saw above, this is a valid constraint on a superfield. Under a supersymmetry transformation, we have

$$\delta S = \int d^4x\, d^4\theta\ \delta K$$

where any superfield K must change as (3.12). This means that we have

$$\delta K = \epsilon^\alpha(\partial_\alpha K - i\sigma^\mu_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}\partial_\mu K) + (-\bar{\partial}_{\dot{\alpha}}K + i\theta^\alpha\sigma^\mu_{\alpha\dot{\alpha}}\partial_\mu K)\bar{\epsilon}^{\dot{\alpha}}$$

But each of these terms involves a derivative. Those terms that are differentiated with respect to a Grassmann coordinate automatically vanish when integrated over superspace by virtue of (3.21). Meanwhile, those terms that involve a differential ∂_μ give at most a boundary term which, if fields drop off suitably quickly asymptotically, also vanishes. We learn that any action of the form (3.22) is necessarily invariant under supersymmetry:

$$\delta S = 0$$

In fact, we can give an expression for the action. The superfield K has an expansion

$$K(x, \theta, \bar{\theta}) = K_{\text{first}}(x) + \ldots + \theta^2\bar{\theta}^2\, K_{\text{last}}(x)$$

The action (3.22) simply picks up the last of these terms

$$S = \int d^4x\ K_{\text{last}}(x)$$

We refer to terms in the action that come from integrating over all of superspace as *D-terms*. The name isn't a great one but comes from the fact that the last component in a real superfield is usually denoted D.

In anticipation of this, in the general expansion of the superfield (3.7) we called the final term D. We also saw that it transforms as a total derivative under a supersymmetry transformation (3.14). This gives another way of seeing the result above: any Lagrangian given by a D-term transforms as a total derivative and so the action is invariant.

3.2.2 The Action for Chiral Superfields

What does this mean for our chiral superfield Φ? As with any other field, we have a choice of what action to build. But, typically in quantum field theory, the simplest possibilities are the most interesting.

Because Φ is complex, we also necessarily have the anti-chiral superfield $\Phi^\dagger$ to play with. Multiplying these together gives a real superfield $\Phi^\dagger\Phi$ that we can integrate over superspace to get the action,

$$S_{\text{chiral}} = \int d^4x\, d^4\theta\ \Phi^\dagger\Phi$$

This means that the action is given by the D-term of $\Phi^\dagger\Phi$. A short calculation, and some integration by parts, shows that the action becomes

$$S_{\text{chiral}} = \int d^4x\ \left[\partial_\mu\phi^\dagger\partial^\mu\phi - i\bar\psi\bar\sigma^\mu\partial_\mu\psi + F^\dagger F\right]$$

where we have thrown away some total derivatives. These are just the standard kinetic terms for a complex scalar ϕ and Weyl fermion ψ. But now we see that there's something special about F: it doesn't have any kinetic terms. Moreover, this will continue to be true as we write down further supersymmetric interactions. This is what it means to be an auxiliary field.

Because there are no kinetic terms for F, it has no propagating degrees of freedom and, when quantised, doesn't give rise to any particle states. That's why it didn't appear in our representation theory analysis of Section 2.3. Nonetheless, there is a good reason that F appears in the chiral superfield.

When looking at single particle states, we previously argued that there have to be equal number of bosonic and fermionic degrees of freedom. And there are. But now we're looking at the action, we can ask two variants of this question. First, we can insist that the number of physical propagating degrees of freedom match. In the context of field theory, these are said to be "on-shell" degrees of freedom. This means that we count the degrees of freedom *after* imposing the equations of motion. The complex scalar field ϕ has two degrees of freedom, while the non-propagating scalar F has none. Meanwhile, the Weyl fermion ψ_α has two complex components but obeys a first order, rather than second order equation of motion which means that ψ_α counts both "position" and "momentum". So the equation of motion cuts the number of on-shell degrees of freedom, giving two. This, of course, matches the degrees of freedom of ϕ.

However, we require the action to be invariant under supersymmetry for all field configurations, not just those that obey the equations of motion. And this motivates us to count the "off-shell" degrees of freedom, meaning the number of fields before equations of motion are imposed. The two complex scalars ϕ and F have two each, while the Weyl spinor ψ_α has four off-shell degrees of freedom because it contains two complex components. The presence of the auxiliary field F is required to match these off-shell degrees of freedom.

Next we want to write down supersymmetric masses and Yukawa-type interactions for these fields. These don't arise from D-terms. Indeed, you could try writing down a more general function $K(\Phi, \Phi^\dagger)$ and integrating over $\int d^4\theta$ but you'll find that it doesn't generate the kind of interactions we want. (We'll see what it does generate in Section 3.2.4.) Instead we have to do something different.

This something different is an option that arises only for chiral superfields. Roughly speaking, because a chiral superfield depends on only half of superspace, we can get a supersymmetric action by integrating it over only half of superspace.

More precisely, given a chiral superfield Φ the function $W(\Phi)$ is also a chiral superfield. In components it reads

$$W(\Phi) = W(\phi) + \sqrt{2}\frac{\partial W}{\partial \phi}\,\theta\psi + \theta^2\left(\frac{\partial W}{\partial \phi}F - \frac{1}{2}\frac{\partial^2 W}{\partial \phi^2}\psi\psi\right) + \ldots$$

where the $+\ldots$ are the extra terms on the second line of (3.19) that include a $\bar\theta$ term. But, as you can see in (3.19), each of these is a total derivative and so will not contribute to the action. This means that, for the purposes of building an action, we can think of $W(\Phi)$ as a function only of θ and not of $\bar\theta$. This means that we can construct a supersymmetric action by integrating over only half of superspace

$$S_W = \int d^4x \left[\int d^2\theta\; W(\Phi) + \int d^2\bar\theta\; W^\dagger(\Phi^\dagger)\right]$$

where the second term is the Hermitian conjugate of the first and is needed to make the action real. This action picks out the θ^2 term in $W(\Phi)$ and is known as an *F-term*, so named because the auxiliary field in a chiral multiplet is usually called F.

We see in (3.20) that the F field (and, by extension any F term that multiplies θ^2 in a chiral multiplet) transforms as a total derivative under supersymmetry. This gives us another way to see that the action S_W is indeed invariant under supersymmetry.

Putting together the D-term and F-term contributions, we get our final supersymmetric action

$$S = S_{\text{chiral}} + S_W = \int d^4x \ \left[\partial_\mu\phi^\dagger\partial^\mu\phi - i\bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi + F^\dagger F + \left(F\frac{\partial W}{\partial\phi} - \frac{1}{2}\frac{\partial^2 W}{\partial\phi^2}\psi\psi + \text{h.c.}\right)\right]$$

This is known as the *Wess-Zumino action*. The function $W(\Phi)$ is called the *superpotential*.

(An aside: There is a completely different object that is also called the Wess-Zumino action, or sometimes the Wess-Zumino-Witten or WZW action. This is a topological term that involves an integral over a higher dimensional space. It has nothing to do with supersymmetry. You can read about it in the lectures on Gauge Theory.)

As promised, the auxiliary field F appears only algebraically in the action. For such fields, it is legitimate to eliminate it by the equation of motion which, in this case, reads simply

$$F + \frac{\partial W^\dagger}{\partial\phi^\dagger} = 0 \quad \text{and} \quad F^\dagger + \frac{\partial W}{\partial\phi} = 0$$

Putting this back into the action gives us an action just in terms of those fields that have propagating degrees of freedom,

$$S = \int d^4x \ \left[\partial_\mu\phi^\dagger\partial^\mu\phi - i\bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi - \left|\frac{\partial W}{\partial\phi}\right|^2 - \frac{1}{2}\frac{\partial^2 W}{\partial\phi^2}\psi\psi - \frac{1}{2}\frac{\partial^2 W^\dagger}{\partial\phi^{\dagger\,2}}\bar{\psi}\bar{\psi}\right]$$

This is the form of the action that we met back in the introduction in (1.1). We see that the scalar potential is positive definite and takes the form

$$V(\phi,\phi^\dagger) = \left|\frac{\partial W}{\partial\phi}\right|^2$$

We still have to specify the form of the superpotential. In general, this can be any holomorphic function of ϕ. If want to restrict ourselves to theories that are renormalisable then we should take a superpotential that is no greater than cubic. For example, we could take

$$W(\Phi) = \frac{m}{2}\Phi^2 + \frac{\lambda}{3}\Phi^3 \tag{3.23}$$

In general, both m and λ can be complex. This gives the potential

$$V = \left|m\phi + \lambda\phi^2\right|^2$$

After expanding this out, the mass of the scalar field is $|m|$. Note that, in addition to the $|\phi|^4$ term, there are also cubic terms $\phi^2\phi^\dagger$ and $\phi^{\dagger\,2}\phi$. These give Feynman diagrams in which a single ϕ particle splits into two others which means that particle number is not conserved in the Wess-Zumino model and, relatedly, there is no way to distinguish particles from anti-particles. This is related to the fact the theory does *not* have a $U(1)$ global symmetry in the presence of the general superpotential (3.23) with $m, \lambda \neq 0$.

With a cubic superpotential, the equation of motion for the Weyl fermion is

$$i\bar{\sigma}^\mu\partial_\mu\psi + m^\star\bar{\psi} = -2\lambda^\star\phi^\dagger\bar{\psi}$$

The fermion also has mass $|m|$. There is no $U(1)$ symmetry associated to this fermion and the mass is an example of a Majorana mass. Note also that the Yukawa term on the right-hand side specifies the interaction between the fermion and scalar and is characterised by the same coupling λ that determines the self-interaction of the scalar. This will have important consequences when we turn to the quantum theory.

Multiple Chiral Superfields

There is a straightforward generalisation of the Wess-Zumino action to multiple chiral superfields Φ_i. We now take the action

$$S = \int d^4x\, d^4\theta\ \sum_i \Phi_i^\dagger\Phi_i + \int d^4x \left[\int d^2\theta\ W(\Phi) + \text{h.c.}\right] \tag{3.24}$$

where if we wish the theory to be renormalisable we should again restrict to a cubic superpotential

$$W(\Phi) = \frac{1}{2}m_{ij}\Phi_i\Phi_j + \frac{1}{3}\lambda_{ijk}\Phi_i\Phi_j\Phi_k$$

The resulting potential is

$$V(\phi) = \sum_i \left|\frac{\partial W}{\partial\phi_i}\right|^2$$

Again, this is positive definite as it must be in a supersymmetric theory since the energy is necessarily positive.

As we have seen, for a single massive chiral multiplet the Weyl fermion necessarily has a Majorana mass. With two chiral multiplets, we may have a Dirac mass. Let's call the chiral multiplets Φ and $\tilde{\Phi}$. Then the simple superpotential

$$\mathcal{W} = m\tilde{\Phi}\Phi$$

gives rise to two Weyl equations, each of which mixes the spinors ψ and $\tilde{\psi}$,

$$i\bar{\sigma}^\mu\partial_\mu\psi + m^\star\bar{\tilde{\psi}} = 0 \quad \text{and} \quad i\bar{\sigma}^\mu\partial_\mu\tilde{\psi} + m^\star\bar{\psi} = 0$$

This is the Dirac equation, decomposed into two Weyl pieces. (Sorry for the ugliness of piling a bar on top of a tilde.) Note that it now has a $U(1)$ symmetry, under which ψ and $\tilde{\psi}$ (or, equivalently the superfields Φ and $\tilde{\Phi}$) rotate with opposite charges.

3.2.3 Supersymmetry of the Wess-Zumino Model Revisited

It's worth pausing for a recap. We've derived the Wess-Zumino model which, for a single chiral superfield, before integrating out F, is given by

$$S = \int d^4x \ \left[\partial_\mu\phi^\dagger\partial^\mu\phi - i\bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi + F^\dagger F + \left(F\frac{\partial W}{\partial\phi} - \frac{1}{2}\frac{\partial^2 W}{\partial\phi^2}\psi\psi + \text{h.c.}\right)\right]$$

Our arguments involving superspace have told us that this action is invariant under the supersymmetry transformations (3.20).

$$\begin{aligned}\delta\phi &= \sqrt{2}\epsilon\psi \\ \delta\psi &= \sqrt{2}i\sigma^\mu\bar{\epsilon}\,\partial_\mu\phi + \sqrt{2}\epsilon F \\ \delta F &= \sqrt{2}i\bar{\epsilon}\bar{\sigma}^\mu\partial_\mu\psi\end{aligned}$$

together with the hermitian conjugate transformations

$$\begin{aligned}\delta\phi^\dagger &= \sqrt{2}\bar{\epsilon}\bar{\psi} \\ \delta\bar{\psi} &= -\sqrt{2}i\epsilon\sigma^\mu\partial_\mu\phi^\dagger + \sqrt{2}\bar{\epsilon}F^\dagger \\ \delta F^\dagger &= \sqrt{2}i\epsilon\sigma^\mu\partial_\mu\bar{\psi}\end{aligned}$$

But this is something that we can just check. It's a little tedious but, given the importance of this result, it's worth doing. From our discussion above, we know that the kinetic terms and the superpotential terms should be independently invariant. We can check each in turn. First the kinetic terms. We have

$$\delta S_{\text{chiral}} = \int d^4x \ \left[\partial^\mu\phi^\dagger\partial_\mu\delta\phi - i\delta\bar{\psi}\,\bar{\sigma}^\mu\partial_\mu\psi + F^\dagger\delta F + \text{h.c.}\right]$$

We've kept only half the terms, the other half buried in the hermitian conjugate. (Admittedly, there was some forethought involved in which terms to keep to ensure

that they cancel among themselves.) Using the supersymmetry transformations above, we have

$$\delta S_{\rm chiral} = \sqrt{2}\int d^4x\ \left[\partial^\mu\phi^\dagger\epsilon\partial_\mu\psi - \partial_\nu\phi^\dagger\epsilon\sigma^\nu\bar\sigma^\mu\partial_\mu\psi - iF^\dagger\bar\epsilon\bar\sigma^\mu\partial_\mu\psi + iF^\dagger\bar\epsilon\bar\sigma^\mu\partial_\mu\psi + \text{h.c.}\right]$$

We see that the two terms with $F^\dagger$ cancel immediately. For the other two terms we have a little bit of work to do. Note that, by integrating by parts twice, we can symmetrise over $(\mu\nu)$ in the second term. But you can check that $\sigma^{(\nu}\bar\sigma^{\mu)} = \eta^{\mu\nu}$ which then ensures that the first two terms also cancel and $\delta S_{\rm chiral} = 0$.

For the superpotential terms we have

$$\delta S_W = \int d^4x\ \left[\delta F\frac{\partial W}{\partial\phi} + F\frac{\partial^2 W}{\partial\phi^2}\delta\phi - \frac{\partial^2 W}{\partial\phi^2}\psi\,\delta\psi - \frac{1}{2}\frac{\partial^3 W}{\partial\phi^3}\psi\psi\,\delta\phi + \text{h.c.}\right]$$

The final $\partial^3 W/\partial\phi^3$ term multiplies ψ^3 and so vanishes because ψ is a 2-component Grassmann field. We're then left with

$$\delta S_W = \sqrt{2}\int d^4x\ \left[i\bar\epsilon\bar\sigma^\mu\partial_\mu\psi\frac{\partial W}{\partial\phi} + F\frac{\partial^2 W}{\partial\phi^2}\epsilon\psi - i\frac{\partial^2 W}{\partial\phi^2}\psi\sigma^\mu\bar\epsilon\partial_\mu\phi - \frac{\partial^2 W}{\partial\phi^2}F\epsilon\psi + \text{h.c.}\right]$$

The $F\epsilon\psi$ terms cancel immediately. The other two cancel after an integration by parts, together with the fact that $\psi\sigma^\mu\bar\epsilon = -\bar\epsilon\bar\sigma^\mu\psi$. We then have $\delta S_W = 0$ as promised.

There is also a version of this calculation after we have integrated out the auxiliary field F, replacing it with its equation of motion $F = -\partial W^\dagger/\partial\phi^\dagger$. As we've seen, the Wess-Zumino action becomes

$$S = \int d^4x\ \left[\partial_\mu\phi^\dagger\partial^\mu\phi - i\bar\psi\bar\sigma^\mu\partial_\mu\psi - \left|\frac{\partial W}{\partial\phi}\right|^2 - \frac{1}{2}\frac{\partial^2 W}{\partial\phi^2}\psi\psi - \frac{1}{2}\frac{\partial^2 W^\dagger}{\partial\phi^{\dagger\,2}}\bar\psi\bar\psi\right]$$

We can also replace F in the supersymmetry transformations. These become

$$\delta\phi = \sqrt{2}\epsilon\psi \quad \text{and} \quad \delta\psi = \sqrt{2}i\sigma^\mu\bar\epsilon\,\partial_\mu\phi - \sqrt{2}\epsilon\frac{\partial W^\dagger}{\partial\phi^\dagger}$$

The calculation described above goes through with only minor modifications (although you can no longer treat the kinetic and superpotential terms independently). This is the supersymmetry invariance of the Wess-Zumino model that we promised back in the introduction.

3.2.4 Non-Linear Sigma Models

The restriction to a cubic superpotential above is motivated by the requirement that the theory be renormalisable. But for theories of scalars, this requirement isn't always at the top of our list. The reason is that these theories may arise as the low-energy description of something more interesting. In this situation, there's no reason to think that the low-energy description should be valid at arbitrarily high-energy scales and so no reason to impose renormalisability.

An illustrative analogy can be found in QCD. At high energies this is a theory of quarks and gluons but at low energies, after confinement has imposed itself on the dynamics, it is a theory of light scalar particles called pions. We denote these fields as $\pi^i(x)$ with i labelling the different pion fields. (For what it's worth, $i = 1, \ldots, 8$ in QCD if we include mesons that contain up, down and strange quarks.) The low-energy dynamics of pions takes the form

$$S_{\rm NLSM} = \int d^4x \; g_{ij}(\pi)\, \partial_\mu \pi^i \partial^\mu \pi^j \tag{3.25}$$

Theories of this kind go by the unhelpful name of *non-linear sigma models*. The fields π^i can be thought of as coordinates on some manifold $\mathcal{M}$ that is called the *target space*. The interactions are hiding in the derivative terms and are packaged into a collection of functions $g_{ij}(\pi)$ that can be viewed as a metric on $\mathcal{M}$. The action (3.25) describes massless scalar fields, although it is always possible to add mass terms if necessary.

Actions of the type (3.25) arise in many places in physics. We first meet them in General Relativity as the action for particles (rather than fields) moving in a curved space or spacetime. But they also occur in many places in condensed matter physics and statistical physics. (The $O(N)$ models discussed in the lectures on Statistical Field Theory are an example.) You can learn more about the specific metric $g_{ij}(\pi)$ that describes pion dynamics in Section 5 of the lectures on Gauge Theory. Here, our interest is in writing down supersymmetric versions of non-linear sigma models.

We can achieve this simply by introducing more interesting D-terms. We consider n chiral superfields Φ^i with $i = 1, \ldots, n$. We'll denote the anti-chiral superfields as $\bar\Phi^{\bar\imath}$ with the $\bar\imath = 1, \ldots, n$ index a useful reminder that these label anti-chiral fields. We then consider the action

$$S = \int d^4x\, d^4\theta \; K(\Phi, \bar\Phi) \tag{3.26}$$

with $K(\Phi, \bar\Phi)$ any real function of these superfields. This function is known as the *Kähler potential*.

Previously, we took

$$K = \sum_i \Phi^{\dagger\bar{i}}\Phi^i$$

We will refer to this as the *canonical Kähler potential*. It is the form that we must take if we want our theory to renormalisable. But if we're willing to entertain low-energy effective theories then we can take a general, real function K. To compute the resulting action, we simply need to compute the D-term of $K(\Phi, \Phi^\dagger)$. This calculation is a little laborious but the result is quite beautiful. The supersymmetric non-linear sigma model takes the form

$$\begin{aligned} S = \int d^4x \ \Big[& g_{i\bar{j}} \left(\partial_\mu \phi^i \partial^\mu \bar{\phi}^{\bar{j}} + \frac{i}{2}\partial_\mu \psi^i \sigma^\mu \bar{\psi}^{\bar{j}} - \frac{i}{2}\psi^i \sigma^\mu \partial_\mu \bar{\psi}^{\bar{j}} + F^i \bar{F}^{\bar{j}} \right) \\ & + \frac{1}{2}\frac{\partial g_{i\bar{j}}}{\partial \phi^k}\left(\psi^k \psi^i \bar{F}^{\bar{j}} - i\bar{\psi}^{\bar{j}}\sigma^\mu \psi^i \, \partial_\mu \phi^k \right) + \text{h.c.} \\ & + \frac{1}{4}\frac{\partial^2 g_{i\bar{j}}}{\partial \phi^k \partial \bar{\phi}^{\bar{l}}}(\psi^i \psi^k)(\bar{\psi}^{\bar{j}} \bar{\psi}^{\bar{l}}) \Big] \end{aligned} \tag{3.27}$$

where the metric $g_{i\bar{j}}$ is related to the Kähler potential as

$$g_{i\bar{j}} = \frac{\partial^2 K}{\partial \phi^i \bar{\phi}^{\bar{j}}} \tag{3.28}$$

Note that this metric only has components with one holomorphic and one anti-holomorphic index. We can eliminate the auxiliary field F through its equation of motion

$$g_{i\bar{j}} F^i + \frac{1}{2}\frac{\partial g_{i\bar{j}}}{\partial \phi^k}\,\psi^k \psi^i = 0 \quad \text{and} \quad g_{i\bar{j}}\bar{F}^{\bar{j}} + \frac{1}{2}\frac{\partial g_{i\bar{j}}}{\partial \bar{\phi}^{\bar{l}}}\,\bar{\psi}^{\bar{l}}\bar{\psi}^{\bar{j}} = 0$$

Substituting this back into the action, we find

$$S = \int d^4x \left[g_{i\bar{j}} \left(\partial_\mu \phi^i \partial^\mu \bar{\phi}^{\bar{j}} + \frac{i}{2}\mathcal{D}_\mu \psi^i \sigma^\mu \bar{\psi}^{\bar{j}} - \frac{i}{2}\psi^i \sigma^\mu \mathcal{D}_\mu \bar{\psi}^{\bar{j}} \right) + \frac{1}{4} R_{i\bar{j}k\bar{l}}(\psi^i \psi^k)(\bar{\psi}^{\bar{j}}\bar{\psi}^{\bar{l}}) \right]$$

Rather wonderfully, all the terms now take a nice geometrical form. The kinetic term for the fermion involves a kind of covariant derivative, defined by

$$D_\mu \psi^i = \partial_\mu \psi^i + \Gamma^i_{jk}\psi^j \partial_\mu \phi^k$$

where, for a metric given by (3.28), the Christoffel symbol is given by

$$\Gamma^i_{jk} = g^{i\bar{l}}\frac{\partial g_{k\bar{l}}}{\partial \phi^j}$$

Meanwhile, the four-fermion interaction terms comes multiplying the Riemann tensor. For a metric given by (3.28), this too takes a special form

$$R_{i\bar{j}k\bar{l}} = g_{m\bar{j}} \frac{\partial \Gamma^m_{ik}}{\partial \bar{\phi}^{\bar{l}}} = \frac{\partial^2 g_{i\bar{j}}}{\partial \phi^k \partial \bar{\phi}^{\bar{l}}} - g^{m\bar{n}} \frac{\partial g_{i\bar{n}}}{\partial \phi^k} \frac{\partial g_{m\bar{j}}}{\partial \bar{\phi}^{\bar{l}}}$$

We have stumbled upon the mathematical framework of *Kähler geometry*. This is a particular form of complex geometry that can be placed on manifolds that are even dimensional and can be endowed with complex coordinates, like the ϕ^i and above. A *Kähler manifold* is a manifold that is endowed with a *Kähler two-form*

$$\Omega = 2i g_{i\bar{j}} d\phi^i \wedge d\bar{\phi}^{\bar{j}}$$

such that

$$d\Omega = 0$$

This requires that the $g_{i\bar{j}}$ satisfies

$$\frac{\partial g_{i\bar{j}}}{\partial \phi^k} = \frac{\partial g_{k\bar{j}}}{\partial \phi^i} \quad \text{and} \quad \frac{\partial g_{i\bar{j}}}{\partial \bar{\phi}^{\bar{l}}} = \frac{\partial g_{i\bar{l}}}{\partial \bar{\phi}^{\bar{j}}}$$

This condition is locally equivalent to the existence of a Kähler potential $K(\phi, \bar{\phi})$, with the metric given by (3.28).

Finally, note that the Kähler potential is not unique. The action (3.26) is invariant under any shift

$$K(\Phi, \bar{\Phi}) + \Lambda(\Phi) + \bar{\Lambda}(\bar{\Phi})$$

where $\Lambda(\Phi)$ is any holomorphic function of Φ^i. This is because $\Lambda(\Phi)$ is a chiral superfield and necessarily vanishes when integrated over all of superspace. These shifts are called *Kähler transformations*.

Supersymmetry has led us to the mathematical framework of Kähler geometry. This is just one of many close connections between supersymmetry and interesting geometric structures. Some of these connections are explored further in the lectures on Supersymmetric Quantum Mechanics.

Adding a Superpotential

The supersymmetric non-linear sigma model (3.27) describes massless fields. We can always add an additional superpotential $W(\Phi)$ to the action. We won't write down the full action, but simply comment that the scalar potential now takes the form

$$V(\phi, \bar{\phi}) = g^{i\bar{j}} \frac{\partial W}{\partial \phi^i} \frac{\partial W^\dagger}{\partial \bar{\phi}^{\bar{j}}} \tag{3.29}$$

with $g^{i\bar{j}}$ the inverse metric.

A Comment on Supergravity

Throughout these lectures we will restrict ourselves to theories with global, or rigid, supersymmetry. As we've mentioned previously, if one extends supersymmetry to a gauge symmetry, making it local, then the resulting theory necessarily includes gravity. This is supergravity. In this case, the scalar potential for a bunch of chiral multiplets again has a fixed form, depending only on the Kähler potential K and superpotential W. It is

$$V(\phi,\bar{\phi}) = e^{K/M_{\rm pl}^2}\left(g^{i\bar{j}}D_iW D_{\bar{j}}W^\dagger - 3\frac{|W|^2}{M_{\rm pl}^2}\right) \tag{3.30}$$

where

$$D_iW = \frac{\partial W}{\partial\phi^i} + \frac{1}{M_{\rm pl}^2}\frac{\partial K}{\partial\phi^i}W$$

Here $M_{\rm pl}$ is the Planck mass. In the limit that $M_{\rm pl}\to\infty$, gravity becomes arbitrarily weak and the potential (3.30) reduces to our previous potential (3.29).

Perhaps surprisingly, the supergravity potential is *not* positive definite. This is related to the fact that supersymmetric theories can exist in anti-de Sitter spacetimes with a negative cosmological constant.

3.3 Non-Renormalisation Theorems

So far our discussion of supersymmetric theories has been entirely classical. But the great advantage of supersymmetry is that it allows us to gain control over the quantum dynamics of the theory. We can start to understand this already just with chiral multiplets. In this section we will show that the superpotential does not receive quantum corrections at any order in perturbation theory. This is known as a *non-renormalisation theorem*. In contrast, all bets are off with the Kähler potential: it is no more constrained than the kinetic terms in any other quantum field theory.

The original proof of the non-renormalisation theorem used Feynman diagrams for superfields. This means that we write down a diagram in which, say, the propagators correspond to superfields. These "super-Feynman diagrams" then encode a number of normal Feynman diagrams, some with bosons running in loops and others with fermions running in loops. One can then show that the most general super-Feynman diagram doesn't contribute to the superpotential.

In these lectures, we're not going to develop the machinery of superfield Feynman diagrams. Instead, we will give a much simpler argument that uses only the symmetries of the problem.

Before we get going, an important comment. Throughout these lectures, theories of chiral superfields will typically be viewed as low-energy effective actions. More precisely, they will be viewed as *Wilsonian low-energy effective actions*. This means that they describe physics only on some suitably large length scale, or equivalently at energies less than some UV cut-off, $E \leq \Lambda_{UV}$. All short distance, or high energy, degrees of freedom have been integrated out but may, in some cases, leave an imprint on the low-energy degrees of freedom. We'll see examples of this as we proceed.

A Wilsonian effective action already takes into account any quantum effects above the cut-off Λ_{UV}. But not those below. You need to use the action to compute, for example, loop diagrams to understand the low-energy quantum dynamics. But there are no UV divergences because the action comes equipped with an explicit cut-off.

There is another, more formal kind of effective action that is common in quantum field theory. This is the *one particle irreducible*, better known as *1PI*, effective action. It arises as the Legendre transform of the (log of) the partition function. In contrast to the Wilsonian effective action, the 1PI effective action is best viewed as a classical action, with all quantum effects already taken into account. This can be problematic in the presence of massless particles since the 1PI effective action may have IR singularities. In contrast, there is no such problem with the Wilsonian effective action.

3.3.1 R-Symmetry Revisited

Given a quantum field theory, one of the first things we should do is understand its symmetries. The kind of Wess-Zumino models (or, more generally non-linear sigma models) that we've described above could have many different Abelian or non-Abelian global symmetries acting on the chiral superfields Φ^i. However, there is one that is of particular importance. This is the $U(1)$ *R-symmetry*. It is special because it does not commute with supersymmetry. Instead, as we saw in (2.25), it obeys

$$[R, Q_\alpha] = -Q_\alpha \quad \text{and} \quad [R, \bar{Q}_{\dot\alpha}] = +\bar{Q}_{\dot\alpha}$$

This means that the R-charge of the scalar ϕ and fermion ψ in a chiral superfield necessarily differ. If the scalar has charge r, then the other members of the multiplet have

$$R[\phi] = r \quad \Rightarrow \quad R[\psi] = r - 1 \quad \text{and} \quad R[F] = r - 2 \tag{3.31}$$

Another way of saying this is to return to the expansion of a chiral superfield (3.19),

$$\Phi = \phi + \sqrt{2}\theta\psi + \theta^2 F + \ldots$$

We endow the supercoordinate θ with an R-charge

$$R[\theta] = +1$$

This tallies with our expression (3.11) for the supercharge $\mathcal{Q} \sim \partial/\partial\theta + \ldots$ which tells us that $\mathcal{Q}$ and θ have opposite charges. The upshot is that if the superfield has R-charge $R[\Phi] = r$, then the other charges in (3.31) follow.

So when do theories enjoy an R-symmetry? Let's consider the simplest Wess-Zumino model (3.24) for a single chiral superfield. The D-term, which gives the kinetic terms, is clearly invariant under any R-symmetry. That leaves the superpotential. This multiplies $d^2\theta$ but Grassmann integration acts in the same way as differentiation which means that the measure has charge

$$R[d^2\theta] = -2$$

We see that the action is invariant under R-symmetry only if we can assign charges to the superfield such that the superpotential has charge

$$R[W] = +2 \tag{3.32}$$

When we have just a single superfield Φ, this is rather limiting. It holds only if the superpotential is a monomial

$$W(\Phi) = \Phi^n$$

in which case we can assign $R[\Phi] = 2/n$. For example, if we take $W(\phi) = \frac{1}{2}m\phi^2$ then the Lagrangian has an R-symmetry under which $\phi \to e^{i\alpha}\phi$ and $\psi \to \psi$. This case is a little boring because there are no interaction terms between ϕ and ψ so obviously we can rotate them independently. We could, however, take $W(\phi) = \frac{1}{3}\lambda\phi^3$ in which case we have the Yukawa term $\phi\psi\psi$ which is invariant under the R-symmetry $\phi \to e^{2i\alpha/3}$ and $\psi \to e^{-i\alpha/3}\psi$. However, if we include both mass and Yukawa terms, there is no R-symmetry. The surprise, as we will now see, is that the lack of an R-symmetry doesn't stop it being useful!

3.3.2 The Power of Holomorphy

We will now see what the R-symmetry has to do with the non-renormalisation of the superpotential. I should warn you that the argument that follows, originally due to Seiberg, is extremely slick and was developed only after a more nuts and bolts argument using Feynman diagrams had been found. But the symmetry argument is both easier and, ultimately, more powerful.

There are a number of conceptual steps that we need to take before the non-renormalisation theorem becomes clear. These are all related to the parameters that appear in the superpotential, things like the mass m and Yukawa coupling λ in (3.23). Each of these parameters is naturally complex. Moreover, like the chiral superfields themselves, the superpotential must be a *holomorphic* function of these parameters.

Of course, as written in (3.23), the superpotential is, by definition, a holomorphic function of parameters. There's an m that sits in the first term and a λ in the second and these are complex. However, the point is that any quantum corrections to the superpotential must also be holomorphic in parameters. This greatly restrains the allowed quantum corrections.

There are two ways to argue that the superpotential must be holomorphic in parameters. The first is direct, but convoluted, and invokes a kind of supersymmetric Ward identity. The second way is to say a bunch of words that hopefully makes it obvious. We're going to adopt the second way.

In any quantum field theory, we can view parameters as arising from some fixed, *background scalar fields*. This means that the parameters may come from some dynamical, but very heavy, scalar field with a potential that pins the value of the scalar to that of the parameter. If this is the case, we wouldn't notice any difference at low energies because these new fields are so heavy. We would see the fluctuations of the parameter only at high energies.

This idea is realised in our world: in the Standard Model the scale of the masses of all elementary particles is set by the expectation value of the Higgs boson. It's an idea that is extended dramatically in string theory where all dimensionless parameters of a low-energy theory also arise as the expectation value of some scalar. However, it is a way of thinking that has proven to be useful in many other arenas including, as we will now see, in supersymmetric theories. The new fields that replace the parameters are sometimes called *spurions*.

This change of perspective from parameters to spurions doesn't change the low-energy behaviour of the theory. But, remarkably, it does allow us to put constraints on what this low-energy behaviour can be. These constraints are especially strong in supersymmetric theories because the spurion must be the lowest component of a chiral superfield. And, as such, the parameters must appear holomorphically in the superpotential.

To understand what this buys us, let's return to the simple case of a single chiral superfield with superpotential

$$W_{\rm tree} = \frac{1}{2}m\Phi^2 + \frac{1}{3}\lambda\Phi^3 \tag{3.33}$$

We refer to this as the tree-level superpotential. Our goal is to understand how it is changed by quantum corrections.

As we've seen above, this theory does not have an R-symmetry. Nonetheless, thinking of the parameters as spurions suggests that we could think of enlarged symmetries under which the parameters also transform. In this larger framework, the theory has two symmetries: one R-symmetry that we call $U(1)_R$ and one global symmetry that commutes with supersymmetry that we call $U(1)_F$. The charges are

	$U(1)_R$	$U(1)_F$
Φ	1	1
m	0	-2
λ	-1	-3

All components of the superfield have the same charge under $U(1)_F$, while the charge under $U(1)_R$ tells us how the lowest scalar component of the superfield transforms, with other components given by (3.31). Relatedly, the superpotential is invariant under $U(1)_F$ but has charge $+2$ under $U(1)_R$, as in (3.32).

I stress again that neither $U(1)_R$ nor $U(1)_F$ are symmetries of our theory since a true symmetry isn't allowed to change parameters of the theory. Said another way, non-vanishing charges for m and λ are telling us that these symmetries are explicitly broken. Nonetheless, the spurions give a useful book-keeping device to characterise exactly how the symmetry is broken. Moreover, as we will now see, they also place strong constraints on the quantum corrections to theory.

Any quantum corrections to the superpotential must be consistent with the two symmetries $U(1)_R$ and $U(1)_F$. Combined with holomorphy, this becomes a very powerful constraint on what can appear. We can form a single, dimensionless combination of superfields that carries no charge at all: this is $\lambda\Phi/m$. (The superfield has the same dimension as a scalar, namely $[\Phi] = 1$. Meanwhile the mass and Yukawa coupling have dimensions $[m] = 1$ and $[\lambda] = 0$.) The only kinds of superpotentials that we can write down consistent with the symmetries are then of the form

$$W_{\rm eff} = m\Phi^2\, f\left(\frac{\lambda\Phi}{m}\right)$$

Note that holomorphy was key here. In most situations assigning a charge to a complex parameter isn't particularly restrictive since, say, $|\lambda|^2$ carries no charges and so can appear anywhere. But the fact that only holomorphic quantities can appear in the superpotential is a game changer.

We still have an arbitrary function $f(\lambda\Phi/m)$ that can appear. But this can be pinned down by studying the theory in different limits. First, for $\lambda \ll 1$, we are in the weakly coupled limit. This means that for small λ we should reproduce the tree level superpotential (3.33), perhaps with corrections at order λ^2 or higher coming from loop diagrams. In other words, the expansion of $f(x)$ about $x = 0$ must take the form

$$f(x) = \frac{1}{2} + \frac{1}{3}x + \mathcal{O}(x^2)$$

However, should also have a well defined superpotential in the limit $m \to 0$ in which we have massless particles. This tells us that we must have $f(x) = \frac{1}{2} + \frac{1}{3}x$ or, equivalently,

$$W_{\text{eff}} = \frac{1}{2}m\Phi^2 + \frac{1}{3}\lambda\Phi^3 = W_{\text{tree}}$$

This is the result we promised: the superpotential receives no quantum corrections to any order in perturbation theory in λ.

(Looking forward: in Section 6, we will study the quantum dynamics of supersymmetric gauge theories. There we will find that superpotentials are, in some circumstances, dynamically generated. But even there they will not be perturbative effects. The superpotentials will arise either by some strong coupling effect or by an instanton effect.)

While the superpotential is immune to quantum corrections, this is not true of the Kähler potential. There are now no holomorphy restrictions and nothing to prohibit corrections of order λ^2 and higher. This means that the physical masses and Yukawa couplings do, in fact, receive quantum corrections. To see this, note that typically the Kähler potential will pick up quantum correction of the form

$$K(\Phi, \Phi^\dagger) = \Phi^\dagger\Phi \to \mathcal{Z}\Phi^\dagger\Phi$$

where $\mathcal{Z} = 1 + \mathcal{O}(\lambda^2)$ is sometimes, inappropriately, called the *wavefunction renormalisation*. This renormalisation factor will have a characteristic logarithmic form

$$\mathcal{Z} = 1 + c|\lambda|^2 \log\left|\frac{\Lambda_{\text{UV}}}{m}\right|^2 + \ldots \tag{3.34}$$

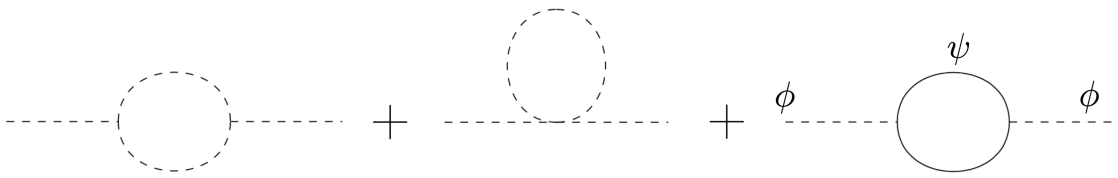


Figure 1. The three one-loop diagrams contributing to the mass of the scalar ϕ. As shown in the last diagram, the dotted line denotes the scalar ϕ and the solid line the fermion ψ.

Here c is a constant whose exact value can be calculated but isn't of interest for our purposes and ... refers to higher loop corrections. This renormalisation changes the kinetic terms for each of the fields and the action is now

$$S = \int d^4x\, d^4\theta\ \mathcal{Z}\Phi^\dagger\Phi + \int d^4x\, d^2\theta\ \left[\frac{1}{2}m\Phi^2 + \frac{1}{3}\lambda\Phi^3\right] + \text{h.c.}$$

Importantly, supersymmetry ensures that there is just a single renormalisation $\mathcal{Z}$ for the superfield, meaning that each of the component fields ϕ, ψ and F experiences the same $\mathcal{Z}$. In such a situation, we should work with the canonically normalised field $\hat{\Phi} = \mathcal{Z}^{1/2}\,\Phi$ and the action becomes

$$S = \int d^4x\, d^4\theta\ \hat{\Phi}^\dagger\hat{\Phi} + \int d^4x\, d^2\theta\ \left[\frac{1}{2}\frac{m}{\mathcal{Z}}\hat{\Phi}^2 + \frac{1}{3}\frac{\lambda}{\mathcal{Z}^{3/2}}\hat{\Phi}^3\right] + \text{h.c.}$$

In this way, the non-renormalisation of the superpotential is not enough to protect the physical mass and Yukawa coupling, which are $m_{\text{phys}} = m/\mathcal{Z}$ and $\lambda_{\text{phys}} = \lambda/\mathcal{Z}^{3/2}$ respectively.

This may seem like a disappointing end to our non-renormalisation claim: the superpotential doesn't change, but the physical parameters sitting within it do. Nonetheless, there's something important going on here. That's because supersymmetry has ensured that the mass m^2_{phys} picks up only a multiplicative renormalisation.

This contrasts strongly with the mass renormalisation expected of a scalar field in a typical quantum field theory. Typically, this mass renormalisation is additive. In particular, any one of the three diagrams shown in Figure 1 would give a contribution of the form

$$m^2_{\text{phys}} \sim m^2 + |\lambda|^2\Lambda^2_{\text{UV}}$$

This is the statement that quantum fluctuations tend to push the mass of scalar fields up to the cut-off scale. In the absence of fine tuning (or some other explanation like symmetry breaking) scalars in quantum field theory are typically heavy. Yet this doesn't

happen in supersymmetric theories: miraculously, the additive renormalisation cancels between each of the diagrams above. This occurs because, as we have seen, the same coupling λ appears in the Yukawa coupling to the fermions and in the 3-point and 4-point vertices of the scalars. The result is that, in supersymmetric theories, there is no difficulty with the masses of scalars being small. In particular, if we choose to set $m = 0$ in the superpotential so that the chiral multiplet is massless then quantum corrections do not change this.

This is the key reason that supersymmetry has attracted the interest of phenomenologists. The mass of the Higgs boson is seemingly much lighter than the cut-off scale of the Standard Model, an issue referred to as the *hierarchy problem*. (See the lectures on Particle Physics for a non-technical account of this.) The existence of supersymmetry at, say, the TeV scale would provide a natural explanation of this. Sadly, there is no evidence that this is the explanation favoured by nature.

3.3.3 Integrating Out Heavy Fields

We may sometimes find ourselves in situations in which our theory has two or more fields with different masses. In this case, we can integrate out the heavier fields, leaving ourselves with an action just for the lighter ones. This will be an important tool for us later, so we pause here to see how it works.

Consider the theory of two chiral superfields Φ and Z, both with canonical Kähler potential $K = \Phi^\dagger\Phi + Z^\dagger Z$, and with superpotential

$$W = \frac{1}{2}MZ^2 + \frac{1}{2}\lambda\Phi^2 Z \tag{3.35}$$

In this example, Z is the heavy field with mass M while Φ is massless, but interacts with Z. If we care only about physics at energies $E \ll M$, we can simply integrate out Z to leave ourselves with a theory for Φ.

Usually in quantum field theory, integrating out fields requires us to evaluate some complicated functional determinants or Feynman diagrams. But, at the level of the superpotential, things are straightforward. For a field configuration Φ, the heavy field will rapidly arrange itself to minimise its energy which it does by adjusting to

$$\frac{\partial W}{\partial Z} = 0 \quad \Rightarrow \quad Z = -\frac{\lambda}{2M}\Phi^2$$

Substituting this back into the superpotential gives our effective superpotential

$$W = -\frac{1}{8}\frac{\lambda^2}{M}\Phi^4$$

This results in a ϕ^6 interaction for the scalar, together with the Yukawa-like interaction for the fermion.

We can also reach the same conclusion by analysing the (spurious) symmetries of the theory. This time there are two global symmetries, $U(1)_\Phi$ and $U(1)_Z$ in addition to the R-symmetry. The charges of various fields and parameters are

	$U(1)_R$	$U(1)_\Phi$	$U(1)_Z$
Φ	1	1	0
Z	0	0	1
M	2	0	-2
λ	0	-2	-1

The unique superpotential consistent with these symmetries that does not involve Z is

$$W \sim \frac{\lambda^2}{M}\Phi^4 \tag{3.36}$$

This symmetry argument doesn't give the overall constant $-1/8$ but, as we've seen above, that's not difficult to get by simply solving the equation of motion.

Note that there's a different philosophy at play here from when we showed the non-renormalisation of the superpotential (3.33). In the earlier case we insisted that the superpotential was well behaved as $m \to 0$. However, in the present case the superpotential clearly diverges as $M \to 0$. But this is to be expected: the theory involving Φ alone is only supposed to make sense at energies $E \ll M$. The fact that the superpotential diverges as $M \to 0$ is telling us something physical: that we shouldn't have discarded the field Z in this limit since it wasn't heavy. This is a lesson that we will see several times as these lectures progress: our low-energy theory will break down in any limit where some field that we have ignored becomes massless.

There's also a terminological issue here. Physicists refer to the superpotential (3.36) as "holomorphic" in Φ, λ and M. Strictly speaking it's not holomorphic in M, but instead *meromorphic* because of the pole. As we explained above, the pole certainly has physical consequence, but we won't belabour the point and will continue to take about holomorphy rather than the more accurate meromorphy.

3.3.4 A Moduli Space of Vacua

We can see a twist on this same theme if we study the superpotential (3.35) in the limit $M = 0$. We have

$$W = \frac{1}{2}\lambda\Phi^2 Z \tag{3.37}$$

This theory has a feature that will become increasingly important as these lectures develop: there is not a unique ground state, or even a finite number of isolated ground states. Instead the potential energy is given by

$$V(\phi, z) = \left|\frac{\partial W}{\partial \phi}\right|^2 + \left|\frac{\partial W}{\partial z}\right|^2 = |\lambda \phi z|^2 + \frac{1}{4}\left|\lambda \phi^2\right|^2$$

We've now resorted to our earlier notation of referring to the lowest scalar component of the superfields Φ and Z by the lower case letter ϕ and z respectively. The minima of the potential are given by

$$V(\phi, z) = 0 \quad \Leftrightarrow \quad \phi = 0 \quad \text{and} \quad z = \text{anything}$$

This means that the potential has a flat direction. Provided that $\phi = 0$, there is no energy cost to turning on z. We say that there is a *moduli space of vacua*. In such a situation, the choice of ground state z is not determined dynamically. Instead, to fully specify the theory, we must also state the expectation value of the field z. Importantly, different choices of z give rise to different theories. For example, we can see immediately from the potential that the mass of ϕ is $m_\phi = |\lambda z|$. In other words, this is moduli space of inequivalent vacua.

Now the roles of z and ϕ are reversed! Provided that $z \neq 0$, the ϕ field is massive while z is massless. We can again play the kind of game that we saw above: is there a superpotential $W(Z)$ that we can write down that might arise after Φ is integrated out? It's simple to see that the answer is no. Everywhere along the moduli space, we have

$$W(Z) = 0$$

This is important. Had we found $W(Z) \neq 0$, it would have meant that there was a quantum generated potential that lifts the flat direction and that the true quantum theory has a preferred ground state. But the non-renormalisation theorem tells us that no such potential is generated. Instead we learn that the moduli space of ground states survives in the quantum theory.

The existence of a moduli space of inequivalent vacua is commonplace in supersymmetric theories but never happens in the absence of supersymmetry. In any non-supersymmetric theory, quantum corrections always generate a potential on the would-be moduli space. This is known as the *Coleman-Weinberg potential* and it picks the true ground state of the system, typically pushing the scalar either to $z = 0$ or to $z = \infty$.

We can get some intuition for the Coleman-Weinberg in a simple quantum mechanics example. Suppose that we have a quantum particle that can move in the (x, y) plane but with a potential that we take to be

$$V_{\text{toy model}} = x^2 y^2$$

The classical system has two flat directions: $x = 0$ and $y =$ anything; or $y = 0$ and $x =$ anything. Suppose that we sit at some $y \neq 0$ but classically set $x = 0$. We then look at the quantum system by supposing that y is constant and quantising the x degree of freedom. But this is just a quantum harmonic oscillator with frequency given by $\omega = y$. And the ground state energy of the quantum harmonic oscillator is $E \sim \hbar\omega = \hbar y$. In this way, the quantisation of x gives rise to an energy that pushes y back towards the origin. Indeed, this quantum mechanical system has a unique ground state, localised around the origin.

The Coleman-Weinberg potential is the analogous phenomenon in quantum field theory. It is generic but is avoided in supersymmetric theories due to a delicate cancellation between bosons and fermions, very similar to those at play in the loop diagrams in Figure 1. We'll be meeting many different vacuum moduli spaces as these lectures progress. Indeed, one of the emerging themes of these lectures is that the geometry of these moduli spaces contains important clues to the underlying physics.

For now, let's go back to our field theory (3.37) and ask: what happens to the moduli space at $z = 0$? Here the ϕ field also becomes massless and it should no longer be valid to ignore it. But how do we see this if we're focussed on the dynamics of z alone? The answer to this can be found in the Kähler potential. Classically, this takes the canonical form $K = Z^\dagger Z$, corresponding to to a flat metric

$$ds^2 = \frac{\partial^2 K}{\partial z \partial \bar{z}}\, dz\, d\bar{z} = d\bar{z} dz$$

However, as we saw above, when we integrate out the massive Φ field the Kähler potential receives a one-loop quantum correction (3.34) and becomes

$$K = Z^\dagger Z \left(1 + c|\lambda|^2 \log \left| \frac{\Lambda_{UV}}{Z} \right|^2 + \ldots \right) \tag{3.38}$$

where $|Z|$ appears in the argument of the logarithm courtesy of the role it plays as the mass of Φ. This results in a metric on the moduli space given by

$$ds^2 = \frac{\partial^2 K}{\partial z \partial \bar{z}}\, dz\, d\bar{z} = \left(-c|\lambda|^2 \log \left(\frac{\bar{z} z}{\Lambda_{UV}^2} \right) + \text{constant} + \ldots \right) d\bar{z} dz$$

We see that distances diverge as we approach $z \to 0$. The log singularity at $z = 0$ is the sign that we have attempted to integrate out a massless particle at that point.

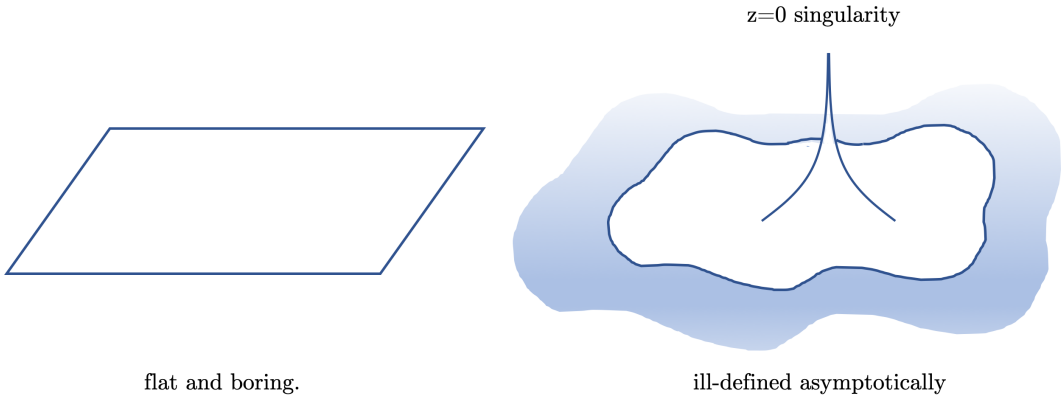


Figure 2. The classical moduli space on the left and the quantum corrected moduli space on the right, with it's singularity at $z = 0$ revealing the massless particle and its negative signature at large z showing that the quantum theory is ill-defined.

There is also some strange behaviour for large $|z|$. When $|z| \gg \Lambda_{UV}$, the first term is negative and, for large enough $|z|$, will overwhelm the constant term, giving us a negative metric. This, of course, is nonsensical. It's telling us that our scalar theory doesn't make sense at very high expectation values or, equivalently at very high energies. In other words, it is capturing the phenomenon of the Landau pole in ϕ^4 theory, but now in a novel geometric fashion. A depiction of the classical and quantum moduli spaces is shown in Figure 2.

3.4 A First Look at Supersymmetry Breaking

A symmetry is said to be *spontaneously broken* if it acts non-trivially on the ground state. This means that the Noether charge Q for the symmetry fails to annihilate the vacuum,

$$Q|0\rangle \neq 0$$

Broken symmetries have important consequences. If a discrete symmetry is spontaneously broken then it implies the existence of multiple, isolated ground states. If a continuous symmetry is spontaneously broken then it implies the existence of a massless particle called a *Goldstone boson*. These ideas underlie Landau's classification of phases of matter and were discussed in some detail in the lectures on Statistical Field Theory and the lectures on Gauge Theory. In this section, we will make a first pass at understanding when supersymmetry may be spontaneously broken and what the consequences are.

First, some basics. From the supersymmetry algebra $\{Q_\alpha, \bar{Q}_{\dot{\alpha}}\} = 2\sigma^\mu_{\alpha\dot{\alpha}} P_\mu$ we can derive an expression for the Hamiltonian

$$H = P^0 = \frac{1}{4}\{Q_1^\dagger, Q_1\} + \frac{1}{4}\{Q_2^\dagger, Q_2\}$$

We already noted in Section 2.2.2 that this implies that all states in a supersymmetric theory necessarily have energy $E \geq 0$. This means that any state with $E = 0$ must be a ground state. These states obey

$$E_{\text{ground}} = \langle 0|H|0\rangle = 0 \quad \Leftrightarrow \quad Q_\alpha|0\rangle = 0$$

In this case the supercharges annihilate the ground state which means that supersymmetry is unbroken. Conversely, supersymmetry is spontaneously broken if and only if the energy of the ground state is non-vanishing

$$E_{\text{ground}} = \langle 0|H|0\rangle > 0 \quad \Leftrightarrow \quad Q_\alpha|0\rangle \neq 0$$

In other words, the ground state energy E_{ground} is the order parameter for broken supersymmetry.

There is another way of looking at this. In theories of chiral multiplets (with a canonical Kähler potential) the potential energy is given by (3.29)

$$V(\phi, \bar{\phi}) = \sum_i |F_i|^2 = \sum_i \left|\frac{\partial W}{\partial \phi^i}\right|^2$$

The ground state energy is non-zero if and only if the F-term gets an expectation value in the vacuum

$$F_i = -\frac{\partial W^\dagger}{\partial \bar{\phi}^i} \neq 0$$

This is known as *F-term supersymmetry breaking*. (There is another option that involves vector multiplets known as *D-term supersymmetry breaking*.)

3.4.1 The Goldstino

If a normal continuous symmetry is spontaneously broken, it results in a massless particle known as a Goldstone boson. If supersymmetry is spontaneously broken, it results in a massless fermion that we call a *Goldstino*.

First, some intuition. When a normal, continuous symmetry is spontaneously broken, the symmetry sweeps out a manifold of equivalent ground states. The canonical example is the breaking of a $U(1)$ symmetry that gives rise to the $\mathbf{S}^1$ rim of the Mexican hat potential. The massless Goldstone mode then arises from fluctuations along this flat direction.

Something similar happens for supersymmetry. From the supersymmetry transformations (3.20), we see that when $F^i \neq 0$, a supersymmetry transformation acting on the vacuum turns on a linear combination of fermions

$$\delta\psi^i = \sqrt{2}\epsilon F^i$$

This is the Goldstino.

There is a simple, hands-on way to see the existence of this massless fermion within the class of theories that we're discussing here. The ground state of the system, whether supersymmetric or not, sits at

$$\frac{\partial V}{\partial \phi^i} = 0 \quad \Rightarrow \quad \sum_j \frac{\partial^2 W}{\partial\phi^i \partial\phi^j}\frac{\partial W^\dagger}{\partial \bar{\phi}^j} = -\sum_j \frac{\partial^2 W}{\partial\phi^i\phi^j} F_j = 0$$

If supersymmetry is broken then $F_j \neq 0$ for some j and the equation above then tells us that the matrix $\partial^2 W/\partial\phi^i\partial\phi^j$ necessarily has an eigenvector with vanishing eigenvalue. But $\partial^2 W/\partial\phi^i\partial\phi^j$ is the fermion mass matrix in our theory. So we learn that when supersymmetry is broken there is at least one massless fermion.

There is a more powerful, general approach to show the existence of the Goldstino that holds for the strongly coupled theories that we will discuss later. This is in close analogy to the original proof of Goldstone's theorem and we just give a bare bones sketch here. The idea is to first construct the *supercurrent* S^μ_α. This is the conserved current associated to supersymmetry transformations and, like any other conserved current, obeys $\partial_\mu S^\mu_\alpha = 0$. The supercharge Q_α arises from this current in the usual way:

$$Q_\alpha = \int d^3x \; S^0_\alpha$$

The supercurrent obeys the algebra

$$\{Q_\alpha, \bar{S}^\mu_{\dot\alpha}\} = 2\sigma^\nu_{\alpha\dot\alpha} T^\mu{}_\nu$$

with $T_{\mu\nu}$ the energy-momentum tensor. This reproduces the usual supersymmetry algebra (2.21) when integrated over space. The proof of the existence of a massless Goldstino then proceeds by computing the two-point function

$$p^\mu \langle S_{\mu\alpha}(p)\, \bar{S}_{\nu\dot{\alpha}}(-p)\rangle = -2\sigma^\mu_{\alpha\dot{\alpha}}\eta_{\mu\nu}E_0$$

with E_0 the ground state energy. This tells us that whenever $E_0 \neq 0$ there is a pole in the $\langle S\bar{S}\rangle$ 2-point function at $p = 0$. This pole corresponds to a massless fermion, the Goldstino.

These lectures are very much focussed on more formal aspects of supersymmetry rather than any possible application to our world. Nonetheless, the existence of the Goldstino raises a puzzle. Clearly we don't see supersymmetry at the energies we have explored so far, which is roughly speaking $E \lesssim 100$ GeV or so. That, in itself, is not such a big issue since it may well be that supersymmetry is broken at some higher energy scale. But, in that case the argument above suggests that we would expect to see a massless Goldstino in our world and no such particle exists. (You might wonder if perhaps the neutrino could act as a Goldstino. This isn't possible because the Goldstino is created from the vacuum and so should share its quantum numbers, while the neutrino carries electroweak charge.)

The resolution to this lies in supergravity. Recall that supergravity involves a local, or gauged, version of supersymmetry. When a normal gauge symmetry is broken, the would-be massless Goldstone boson is "eaten" by the Higgs mechanism and becomes massive. The same is true of gauged supersymmetry. In the context of supergravity, the would-be Goldstino is eaten by the gravitino and both become massive with mass of order E_0, the supersymmetry breaking scale.

3.4.2 The Witten Index

Not all theories can spontaneously break supersymmetry. There is a topological obstruction that they must overcome. This obstruction is the *Witten index*.

We met the Witten index briefly back in Section 2.3. It defined as the sum over all states

$$\mathrm{Tr}(-1)^F e^{-\beta H} \tag{3.39}$$

The trace is taken over the infinite number of states in the quantum field theory Fock space. Here F is the fermion number, so that the Witten index counts bosonic states with a $+1$ and fermionic states with a -1. In contrast to the discussion in Section

2.3, we've now included a factor of $e^{-\beta H}$, where H is the Hamiltonian. This acts as a regulator on the very high energy states. But, as we'll now show, these high energy states don't in fact contribute to the Witten index.

To make the discussion precise, we should really work on a compact space, like $\mathbf{T}^3$. This ensures that momentum is quantised and, correspondingly, the energy spectrum is discrete. There are then no subtleties in taking the trace.

The key fact about the Witten index is that any states with energy $E > 0$ necessarily come in boson-fermion pairs. This follows from the kind of representation theory that we did in Section 2.3. More precisely, if we define the combination of supercharges

$$\mathbb{Q} = Q_1 + Q_2^\dagger$$

then, from the supersymmetry algebra (2.21), it is simple to see that these obey

$$\{\mathbb{Q}, \mathbb{Q}^\dagger\} = 4H$$

Consider the action of this operator on a state with energy $H|\phi\rangle = E|\phi\rangle$ with $E \neq 0$. We can then define the fermionic creation and annihilation operators

$$a = \frac{\mathbb{Q}}{2\sqrt{E}} \quad \Rightarrow \quad \{a, a^\dagger\} = 1$$

This algebra has a two-dimensional irreducible representation $|\phi\rangle$ and $a^\dagger|\phi\rangle$, both with energy E. One of these states is bosonic and the other fermionic, ensuring that they cancel in their contribution to the Witten index.

Note that the degeneracy of $E > 0$ states is true whether or not supersymmetry is broken. If supersymmetry is unbroken, it arises because of mass degeneracy of particles in a supermultiplet. If supersymmetry is broken then the degeneracy arises simply from the addition of a zero energy Goldstino mode. (More precisely, on a compact space it arises from the quantisation of the Goldstino zero mode.) In this case, there is no need for the masses of bosonic and fermionic particles to be equal.

This argument for the degeneracy of the spectrum breaks down for states of zero energy. For such supersymmetric ground states there is no obstacle to having just a single state obeying

$$Q_\alpha|0\rangle = Q_\alpha^\dagger|0\rangle = 0$$

More generally, it may well be the case that a theory has multiple ground states. In this case, each ground state could be bosonic or fermionic. Here a "fermionic" ground state is nothing exotic: it just means that it sits in the sector of the Hilbert space with $(-1)^F|0\rangle = -|0\rangle$ rather than $(-1)^F|0\rangle = +|0\rangle$.

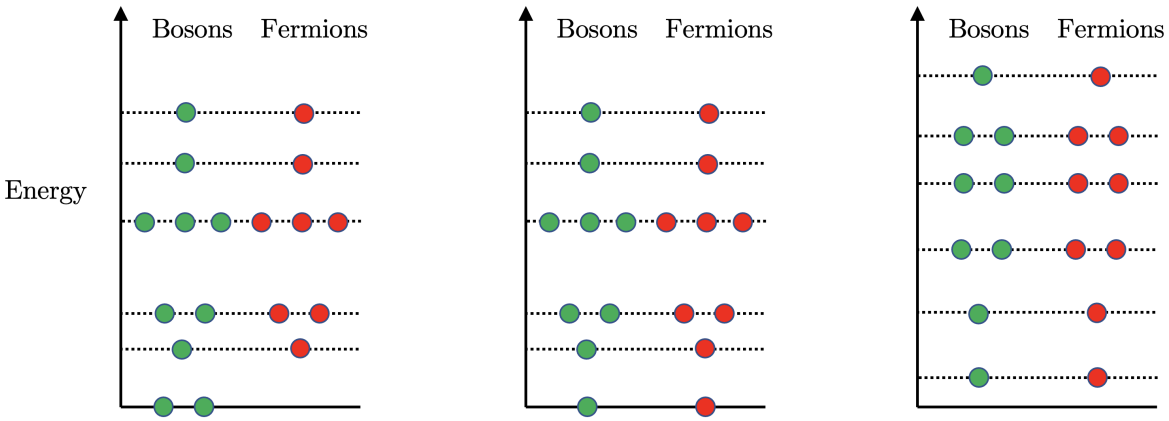


Figure 3. The spectrum on the left has $\mathrm{Tr}(-1)^F e^{-\beta H} = 2$ and cannot break supersymmetry as parameters are changed. The one in the middle has $\mathrm{Tr}(-1)^F e^{-\beta H} = 0$. It does not break supersymmetry but as parameters are varied there is nothing to protect it from turning into the spectrum on the right which does break supersymmetry.

The upshot is that the Witten index (3.39) actually counts the difference in the number of $E = 0$ ground states

$$\mathrm{Tr}(-1)^F e^{-\beta H} = n_B(E = 0) - n_F(E = 0)$$

In particular, the Witten index is independent of the value of β. Moreover, it is actually independent of any other parameter in the theory. To see this, consider a generic spectrum of a supersymmetric theory as shown in Figure 3. All $E \neq 0$ states come in pairs, while $E = 0$ states may be unpaired. As we vary parameters in the theory, some of the $E = 0$ ground states may get lifted and get non-zero energy. But they can only be lifted in pairs and the Witten index remains unchanged. In this sense, the Witten index provides a topological classification of theory.

(Actually, this last statement is only true providing that asymptotic nature of the potential does not change. We'll see an example below.)

All of this means that supersymmetry can only be spontaneously broken in theories with $\mathrm{Tr}(-1)^F = 0$. In contrast, if $\mathrm{Tr}(-1)^F \neq 0$ for some choice of parameters then the theory cannot break supersymmetry as the parameters are changed and this remains true even as the dynamics becomes strongly coupled.

An Example

All of the theories that we will explore in this section are weakly coupled and we can tell whether supersymmetry is broken simply by looking at the potential. This means that we don't really have any need for the Witten index. It starts to show its teeth only for the strongly interacting theories that we will meet in Section 6. Nonetheless, it's useful to get a feeling for how supersymmetric ground states are robust.

Consider a Wess-Zumino model with a single chiral superfield Φ with a superpotential that is a polynomial of degree $p+1$,

$$W(\phi) = a_{p+1}\phi^{p+1} + a_p\phi^p + \ldots + a_1\phi$$

A supersymmetric ground state exists if there are solutions to the equation

$$\frac{\partial W}{\partial \phi} = 0 \tag{3.40}$$

But there's always a solution to this equation because we're solving a polynomial over the complex numbers. In fact, there are always p such solutions (counted with multiplicity). As we vary the coefficients a_i the ground states move around, but they are never lifted. This reflects the fact that this theory has $\mathrm{Tr}(-1)^F e^{-\beta H} = p$. It's a little fiddly to show that all ground states contribute the same $+1$ to the Witten index, rather than with different signs. You can find the argument in the lectures on Supersymmetric Quantum Mechanics where the Witten index plays a central role throughout.

There is, however, an important caveat to the statement that the theory always has p ground states. If we set $a_{p+1} = 0$ then the superpotential becomes a polynomial of degree p and the theory has $p-1$ ground states. It's simple to see what happens here: as we take the limit $a_{p+1} \to 0$, one of the ground states starts heading off to infinity in field space $\phi \to \infty$. This provides a salutary lesson: the Witten index can change if we change how the theory behaves in the asymptotic region of field space. We will see other examples below where, as we vary parameters, a moduli space of ground states emerges then disappears again. This also provides a scenario where the Witten index can jump.

3.4.3 The O'Raifeartaigh Model

The Witten index argument, together with some basics facts about roots of polynomials, means that you have to strive to write down theories that break supersymmetry. Nonetheless, it's not too difficult to achieve. The first model was constructed in 1975

by O'Raifeartaigh. It contains three chiral superfields that we call Y, Z and Φ with the superpotential

$$W = \frac{h}{2} Y(\Phi^2 - \mu^2) + mZ\Phi \tag{3.41}$$

We take all fields to have a canonical Kähler potential so the theory is renormalisable. (We will relax this assumption below.) The parameter h is dimensionless, while $[\mu] = [m] = 1$. It's useful to note that the potential has an R-symmetry (a real one, not a spurious one) under which $R[Y] = R[Z] = 2$ and $R[\Phi] = 0$.

The fields Y and Z act like Lagrange multipliers in the superpotential, setting

$$\frac{\partial W}{\partial Y} = \frac{h}{2}\left(\Phi^2 - \mu^2\right) = 0 \quad \text{and} \quad \frac{\partial W}{\partial Z} = m\Phi = 0$$

Clearly there's no way to set both of these to zero so supersymmetry is spontaneously broken.

The potential of this model is given by

$$V(y, z, \phi) = \frac{1}{4}\left|h\phi^2 - h\mu^2\right|^2 + |m\phi|^2 + |hy\phi + mz|^2$$

Note that y and z are just names of scalar fields here; they are not to be confused with coordinates on spacetime. The minima of the potential always sits at $z = hy\phi/m$ so the final term vanishes. What happens next depends on the ratio of parameters

$$\alpha = \left|\frac{h\mu}{m}\right|$$

If $\alpha < 1$ then the minima is at $\phi = z = 0$. If $\alpha > 1$ then this minima splits into two minima at $\phi = \pm\,\text{something}$ and a saddle. Importantly, in either case y is arbitrary: it is a flat direction.

It is simple to check that the whole superfield Y is massless. The fermion is the Goldstino while the phase of y is a Goldstone boson associated to a broken R-symmetry. The surprise is that $|y|$ is also massless, with no symmetry reason to protect it. As we now explain, the classical moduli space parameterised by $|y|$ doesn't survive in the full quantum theory.

The Quantum Generated Potential

Importantly, the mass spectrum of the O'Raifeartaigh model depends on the the value of $|y|$: each point on this moduli space describes different physics. Furthermore, and in contrast to our earlier supersymmetric models, the masses of the bosons and fermions are different. This is important because it means that when we integrate out these heavy fields they will induce a Coleman-Weinberg potential on the moduli space parameterised by $|y|$. Here we give some general comments on the form of this potential.

Integrating out heavy fields in a 4d quantum field theory usually give three kinds of divergences: quartic, quadratic and logarithmic. In each case, bosons give rise to a positive potential and fermions a negative potential. In a supersymmetric theory, these exactly cancel which is the reason that moduli space of vacua are not lifted when supersymmetry is broken. As we now explain, when supersymmetry is spontaneously broken some, but not all, of this cancellation remains.

First the quartic divergences. These are given by

$$V_{\text{eff}} \sim \text{Str}\, \Lambda_{UV}^4$$

where Λ_{UV} is the UV cut-off and Str is the *supertrace* which means that we sum over all complex bosonic fields minus the sum over all fermionic fields. (Note that we're summing over the different fields of the theory here. This contrasts with the Witten index where we were performing the much larger sum over all states in the Hilbert space.) But supersymmetric theories have an equal number of bosonic and fermionic fields so all quartic divergences disappear regardless of whether supersymmetry is spontaneously broken or not.

Next up are the quadratic divergences. These take the form

$$V_{\text{eff}} \sim \Lambda_{UV}^2 \text{Str}\, \mathcal{M}^2 = \Lambda_{UV}^2 \left(\text{Tr}\, \mathcal{M}_B^2 - \text{Tr}\, \mathcal{M}_F^2 \right)$$

Here $\mathcal{M}$ is the tree-level mass matrix, including both bosons and fermions. In the second equality we've written it in terms of a sum over bosonic and fermionic fields with their appropriate mass matrices $\mathcal{M}_B$ and $\mathcal{M}_F$. Clearly this too vanishes when there is a degeneracy of masses. But a rather nice result says that it also vanishes when supersymmetry is spontaneously broken:

Claim: $\text{Str}\, \mathcal{M}^2 = 0$ for F-term supersymmetry breaking.

Proof: This holds generally in any theory with N superfields and a canonical Kähler

potential. The proof involves just a little bit of algebra. First, the $N \times N$ mass matrix for a Weyl fermion is

$$(\mathcal{M}_F)_{ij} = \frac{\partial^2 W}{\partial\phi^i \phi^j}$$

We write this in terms of the auxiliary field $\bar{F}_{\bar{i}} = -\partial W/\partial\phi^i$ as $(\mathcal{M}_F)_{ij} = -\bar{F}_{\bar{i}j}$. The mass-squared matrix that appears in the supertrace formula is the Hermitian matrix

$$(\mathcal{M}_F)^2 = (\mathcal{M}_F)_{\bar{i}j}(\mathcal{M}_F)^\dagger_{j\bar{k}} = \bar{F}_{\bar{i}j}F_{j\bar{k}}$$

Meanwhile, we have to be a little more careful with the bosons because after supersymmetry breaking the real and complex parts of the scalar will typically have different mass. (This happens, for example, in the O'Raifeartaigh Model.) This means that we should break the bosons into real and imaginary pieces and consider the $2N \times 2N$ mass matrix

$$\mathcal{M}_B^2 = \begin{pmatrix} \frac{\partial^2 V}{\partial\phi^i \bar{\phi}^j} & \frac{\partial^2 V}{\partial\phi^i \phi^l} \\ \frac{\partial^2 V}{\partial\bar{\phi}^j \bar{\phi}^k} & \frac{\partial^2 V}{\partial\bar{\phi}^j \phi^l} \end{pmatrix}$$

But $V = F_i\bar{F}_{\bar{i}}$. Plugging this expression into $\mathcal{M}_B^2$ above and taking the trace (remembering that there's a factor of $\frac{1}{2}$ because we're now working with real fields rather than complex) gives the claimed result. □

All of which means that in a theory with spontaneously broken supersymmetry, the only contribution to the effective potential comes from the logarithmic divergences. It can be shown that these too take the form a supertrace over the mass matrix

$$V_{\text{eff}} = \frac{1}{64\pi^2}\,\text{Str}\,\mathcal{M}^4 \log\left(\frac{\mathcal{M}}{\Lambda_{UV}}\right)^2$$

Again, this vanishes if supersymmetry is unbroken. But now it does not vanish if supersymmetry is spontaneously broken. This gives the quantum potential that lifts flat directions in this case.

The mass matrix $\mathcal{M}$ depends on the value of the field y, and hence V_{eff} should be viewed as a potential that lifts this flat direction. In any theory with a flat direction, quantum generated potentials typically push the field to one end or another. Computing the masses shows that here the true ground state of the system sits at $y = 0$. This is the unique ground state with spontaneously broken supersymmetry.

3.4.4 R-symmetry and the Nelson-Seiberg Argument

We could continue exploring different models (and we will below!) but it is useful to first stop and try to understand some general features of supersymmetry breaking. To this end, let's first look at a small extension of the O'Raifeartaigh model,

$$W = \frac{h}{2}Y(\Phi^2 - \mu^2) + mZ\Phi + \frac{\nu}{2}\Phi^2 + \frac{\epsilon}{2}Y^2 \tag{3.42}$$

This differs from the O'Raifeartaigh model by the addition of the last two terms. Note that these two terms break the R-symmetry and this will be important shortly. For now, we can simply study the scalar potential arising from this superpotential. It is

$$V(y, z, \phi) = \frac{1}{4}\left|h\phi^2 - h\mu^2 + 2\epsilon y\right|^2 + |m\phi|^2 + |hy\phi + mz + \nu\phi|^2$$

Now the theory does have a supersymmetric ground state, sitting at $z = \phi = 0$ and $y = h\mu^2/2\epsilon$.

If, however, we now take $\epsilon \to 0$ to remove the last term in (3.42), then the supersymmetric vacuum moves off to infinity in field space $y \to \infty$ and we once again find ourselves with a theory that breaks supersymmetry, one that appears to be very similar to the original O'Raifeartaigh model. However, in one way there is a key difference between them. To describe this difference we first need to explain what it means for theories to be "generic".

All the theories we're discussing in this section should be viewed as low-energy effective theories, coming from some unknown UV physics. But there is a mantra that can be applied to such low-energy theories: anything that is not forbidden is mandatory. This means that quantum effects will conspire to generate all possible terms in the potential provided that they are consistent with the symmetries of the theory. A low energy effective theory that includes all such terms, with no particular fine tuning of the coefficients, will be said to be "generic".

In this sense, the O'Raifeartaigh model (3.41) is generic. It has an R-symmetry and there are no further terms that one can add consistent with this symmetry.

In contrast, the extension of the O'Raifeartaigh model (3.42) is not generic. It no longer has an R-symmetry, but we have not included Z^2 terms nor Φ^3 terms nor many other terms that we could write down. Despite this, it turns out that the behaviour we have seen – namely the existence of a supersymmetric ground state – persists if we add all these extra terms. So it is sufficient for our discussion.

However, among this large class of theories that do not have an R-symmetry, we only find one that breaks supersymmetry if we set one of the coefficients to vanish: $\epsilon = 0$. This is a very particular choice of coefficient. If the theory (3.42) arose as the low-energy limit of some other theory — one which itself did not have an R-symmetry — then there would be no reason to expect that $\epsilon = 0$. For this reason, it's unlikely that the supersymmetry breaking we've found in this model is actually useful.

In fact, one can make these kind of arguments more generally. Consider a theory with N chiral superfields Φ^i and a potential $W(\phi)$. A supersymmetric ground state obeys

$$\frac{\partial W}{\partial \phi^i} = 0 \tag{3.43}$$

Supersymmetry is broken if we can cook up a superpotential for which there are no solutions to this equation. But these are N equations in N variables and for a generic W they always have a solution. That means that a supersymmetric ground state can always be found.

It is, however, appropriate to restrict W by symmetry arguments and we might wonder if that will help us find a generic W that breaks supersymmetry. For example, suppose that W is invariant under a $U(1)$ global symmetry under which the superfield Φ_i transforms with charge q_i,

$$\Phi_i \to e^{i\alpha q_i}\Phi_i$$

In this case the superpotential can always be written as a function of $W = W(X_i)$ with X_i the invariant ratios

$$X_i = \frac{\Phi_i}{\Phi_1^{q_i/q_1}} \quad i = 2, \ldots, N$$

But now the conditions for a supersymmetric ground state are just $\partial W/\partial X_i = 0$ for $i = 2, \ldots, N$ which are $N - 1$ conditions for $N - 1$ variables. Again, for a generic W there will be a solution. We see that imposing global symmetries doesn't help us in finding supersymmetry breaking potentials.

However, the story is different if there is an R-symmetry. We take the superfields to transform with charges r_i,

$$\Phi_i \to e^{i\alpha r_i}\Phi_i$$

We again form the invariant ratios

$$\tilde{X}_i = \frac{\Phi_i}{\Phi_1^{r_i/r_1}} \quad i = 2, \ldots, N$$

The key difference is that the superpotential must have R-charge +2. This means that it takes the form

$$W(\Phi_1, \tilde{X}_i) = \Phi_1^{2/r_1}\, \tilde{W}(\tilde{X}_i)$$

The conditions for a supersymmetric ground state are now $\partial\tilde{W}/\partial\tilde{X}_i = 0$. But, as long as $\Phi_1^{2/r_1} \neq 0$, we must also have $\tilde{W}(\tilde{X}) = 0$. This is now N conditions on $N-1$ variables $\tilde{X}_i$ and generically there will *not* be a solution.

This is the *Nelson-Seiberg argument*. It says that models of supersymmetry breaking with generic superpotentials should have an R-symmetry. This is indeed true of the O'Raifeartaigh model.

Our main interest in these lectures is not to construct realistic supersymmetric theories, but rather to explore the strong coupling dynamics of quantum field theories. Nonetheless, it's worth mentioning that the argument for the existence of an R-symmetry causes something of a headache if you're trying to build realistic models in which supersymmetry is spontaneously broken. In some models, like the O'Raifeartaigh model, the non-supersymmetric ground state preserves the R-symmetry (recall that, ultimately, the quantum potential pushes us to $y = 0$.). But this causes problems further down the line because, as we will see in Section 4, an R-symmetry prohibits masses for the superpartners of gauge fields, known as *gauginos*. But these must be heavy in any realistic theory.

Alternatively, we could cook up models in which both supersymmetry and the R-symmetry are spontaneously broken. But this then leads to a light Goldstone boson known as the *R-axion*. Again, we must find a way to give this a mass.

3.4.5 More Ways to (Not) Break Supersymmetry

In the remainder of this section, we briefly discuss a number of other simple models that illustrate different ways in which supersymmetry can be broken.

Runaway Potentials

Here is a model that looks like it breaks supersymmetry but, on closer inspection, does something different. It consists of two fields, Z and Φ, with superpotential

$$W = \frac{h}{2} Z\Phi^2 - \lambda\Phi$$

It has an R-symmetry with $R[\Phi] = 2$ and $R[Z] = -2$ and a scalar potential given by

$$V = \frac{1}{4}|h\phi^2|^2 + |hz\phi - \lambda|^2$$

Clearly there is no way to set both terms to zero so we seem to again have a situation in which supersymmetry is broken. However, instead something slightly different is happening and the potential slopes to zero asymptotically. To see this, look at the direction with $\phi = \lambda/hz$ for which the potential is given by

$$V(z) = \left|\frac{\lambda^2}{2hz^2}\right|^2$$

Clearly $V(z) \to 0$ as $z \to \infty$. So it is better to say that this theory has no stable ground state at all: the field is pushed to $z \to \infty$ where supersymmetry is restored. We will see behaviour like this emerging dynamically in Section 6.

Metastable Supersymmetry Breaking

Let's now consider a slightly different variant of the model (3.42) that broken R-symmetry. We take the superpotential

$$W = \frac{h}{2}Y(\Phi^2 - \mu^2) + mZ\Phi + \frac{\epsilon}{2}Z^2$$

The potential is

$$V(y, z, \phi) = \frac{1}{4}\left|h\phi^2 - h\mu^2\right|^2 + |m\phi + \epsilon z|^2 + |hy\phi + mz|^2$$

This breaks R-symmetry and so, on general grounds, we might expect it to have a supersymmetric vacuum (provided that we have taken the superpotential to be suitably generic). This is indeed the case: the supersymmetric ground state is given by $\phi^2 = \mu$ and $z = -m\phi/\epsilon$ and $y = m^2/h\epsilon$.

For ϵ very small, this ground state sits a long way from the origin of field space. Moreover, if we look close to the origin, $y = 0$, then the potential is very similar to the original O'Raifeartaigh model. In particular, when $\phi = z = 0$ there is a flat direction along y, albeit one that is not a global minimum of the the potential. When we include quantum corrections, this will be lifted and, for suitable values of the parameters, we will find a local, supersymmetry breaking vacuum at the origin. A schematic sketch of this situation is shown in Figure 4.

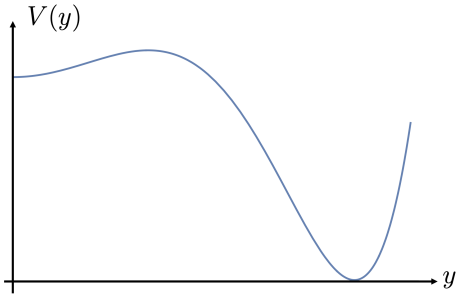


Figure 4. A schematic sketch of the metastable minima at $y = 0$ that breaks supersymmetry and the global, supersymmetric ground state at $y \sim 1/\epsilon$. (The actual potential should be plotted in higher dimensions.)

In a quantum field theory, any local minima of a potential that is not the global minimum is a metastable state, with a finite lifetime. This means that if we initially sit in the supersymmetry breaking minimum, we will eventually tunnel out into the supersymmetric ground state. Nonetheless, it is possible to use such metastable minima to build phenomenologically viable models. You just need to make sure that "eventually" $\gg$ 100 billion years (or whatever allows you to sleep easy at night).

Playing with the Kähler Potential

So far we haven't discussed the simplest theory that breaks supersymmetry. This is a single chiral multiplet with superpotential

$$W = \mu^2 \Phi$$

Clearly $\partial W/\partial \phi = \mu^2 \neq 0$. But this feels too cheap. The ground state energy may be non-zero, but the theory is just a free massless fermion (the Goldstino!) and a free complex scalar. It's hard to argue that there's any deep physics in there.

Things change however if we consider a more general Kähler potential $K = K(\phi^\dagger \phi)$. The fermion remains massless but a potential is now generated for the scalar, given by

$$V(\phi) = |\mu|^4 \left(\frac{\partial^2 K}{\partial \phi \partial \phi^\dagger} \right)^{-1}$$

The price that we pay is that the theory is no longer renormalisable. Of course, as we've stressed above, given that we view these scalar field theory as low energy effective theories, that is not necessarily a bad thing.

For example, suppose that, when expanded around the origin, the Kähler potential takes the form

$$K(\phi, \phi^\dagger) = |\phi|^2 - \frac{1}{M^2}|\phi|^4 + \ldots$$

This kind of behaviour can arise from integrating out heavy particles of mass M. (We found a log correction to the Kähler potential from integrating out particles in (3.38), but other interactions can give the power-law above.) We should view M as the UV cut-off of the theory. Other energy scales in the game should necessarily be much smaller than the cut-off which, for us, means $\mu \ll M$.

With such a Kähler potential, the actual potential energy reads

$$V(\phi, \phi^\dagger) = |\mu|^4 \left(1 + \frac{4}{M^2}|\phi|^2 + \ldots\right)$$

This now has a minima at $\phi = 0$. The net result is that the scalar ϕ has a mass $m_\phi = 2\mu^2/M^2$.

A comment on the scales here. As we've mentioned repeatedly, all the theories in this section should be viewed as low-energy effective theories arising from some high energy completion. In the present case, our theory is valid at energy scales $\sim \mu$. We have integrated out stuff at the much higher scale $M \gg \mu$ and this is what gives rise to the correction to the Kähler potential. It's necessary that there is a separation of scales here. Although the scalar ϕ is not massless, it is light in the sense that $2\mu^2/M \ll \mu$.

Different Kähler potentials can give the different kinds of behaviour that we saw above, including runaway potentials and metastable vacua.

4 Supersymmetric Gauge Theories

Finally, we turn to the main subject of these lectures: supersymmetric gauge theory. In this section we will describe the classical structure of supersymmetric gauge theories. In Section 6 we turn to their quantum dynamics.

4.1 Abelian Gauge Theories

A gauge field A_μ sits inside a real superfield satisfying $V(x,\theta,\bar\theta) = V^\dagger(x,\theta,\bar\theta)$. Expanding out such a superfield in components, we have

$$\begin{aligned} V(x,\theta,\bar\theta) = C(x) + \theta\chi(x) + \bar\theta\bar\chi(x) + i\theta^2 M(x) - i\bar\theta^2 M^\dagger(x) + \theta\sigma^\mu\bar\theta\, A_\mu(x) \\ + \theta^2\bar\theta\left(\bar\lambda(x) + \frac{i}{2}\bar\sigma^\mu\partial_\mu\chi(x)\right) + \bar\theta^2\theta\left(\lambda(x) + \frac{i}{2}\sigma^\mu\partial_\mu\bar\chi(x)\right) \\ + \frac{1}{2}\theta^2\bar\theta^2\left(D(x) - \frac{1}{2}\Box C(x)\right) \end{aligned} \tag{4.1}$$

The real superfield contains two real scalars, C and D, and a complex scalar M, together with two Weyl fermions χ_α and λ_α. Importantly, it also contains a real vector field A_μ. This will play the role of the gauge field in what follows. We've defined some of the components to include derivatives of others. This should simply be thought of as a redefinition of $D(x)$ and $\lambda(x)$, admittedly one that you wouldn't write down unless you had an inkling of what was coming.

If A_μ is to be a gauge field, then it must enjoy a gauge transformation. These too sit in a superfield. We start by taking a chiral superfield Ω

$$\Omega = \omega + \sqrt{2}\theta\rho + \theta^2 G + i\theta\sigma^\mu\bar\theta\partial_\mu\omega - \frac{i}{\sqrt{2}}\theta^2\partial_\mu\rho\,\sigma^\mu\bar\theta - \frac{1}{4}\theta^2\bar\theta^2\Box\omega$$

Then $i(\Omega - \Omega^\dagger)$ is a real superfield. Consider the generalised gauge transformation

$$V \to V + i(\Omega - \Omega^\dagger) \tag{4.2}$$

The vector component of the real superfield shifts as

$$A_\mu \to A_\mu - 2\partial_\mu(\mathrm{Re}\,\omega) := A_\mu + \partial_\mu\alpha \tag{4.3}$$

But this is precisely the form of a gauge transformation. But under this generalised gauge transformation, it's not just A_μ that shifts. The other fields in $V(x,\theta,\bar\theta)$ also transform as

$$\begin{aligned} C &\to C - 2\,\mathrm{Im}\,\omega \\ \chi &\to \chi + \sqrt{2}i\rho \\ M &\to M + G \end{aligned} \tag{4.4}$$

Importantly, however, $\lambda \to \lambda$ and $D \to D$ remain unchanged. This can be traced to the extra derivative terms that we included in the superfield expansion (4.1) which were designed to soak up the shift by a chiral superfield.

We can now use this gauge transformation to simply set $C = \chi = M = 0$. This is known as *Wess-Zumino gauge*. Note that it's not a gauge choice that has done anything to fix A_μ. It's more a "super gauge choice" to fix the extraneous components in the superfield. In Wess-Zumino gauge, the superfield takes the simpler form

$$V_{WZ} = \theta\sigma^\mu\bar{\theta}\, A_\mu + \theta^2\,\bar{\theta}\bar{\lambda} + \bar{\theta}^2\theta\lambda + \frac{1}{2}\theta^2\bar{\theta}^2 D \tag{4.5}$$

It contains a gauge field A_μ, a Weyl fermion λ_α and an extra real scalar D that, as the top component of a superfield, will prove to be auxiliary. If we quantise A_μ and λ then we find the single-particle excitations of the gauge multiplet that we anticipated in Section 2.3.2.

If you act with a supersymmetry transformation on V_{WZ}, then it will take you out of Wess-Zumino gauge. This isn't a big headache; it just means that you have to do a compensating transformation to put yourself back in Wess-Zumino gauge afterwards. The supersymmetry transformations then act on the fields A_μ, λ and D as

$$\begin{aligned} \delta A_\mu &= \epsilon\sigma_\mu\bar{\lambda} + \lambda\sigma_\mu\bar{\epsilon} \\ \delta\lambda &= \epsilon D + (\sigma^{\mu\nu}\epsilon)F_{\mu\nu} \\ \delta D &= i\epsilon\sigma^\mu\partial_\mu\bar{\lambda} - i\partial_\mu\lambda\bar{\sigma}^\mu\bar{\epsilon} \end{aligned} \tag{4.6}$$

Note that the supersymmetry transformations (3.15) alone give us a term proportional to $\partial_\mu A_\nu$ in $\delta\lambda$. The compensating gauge transformation to take us back into Wess-Zumino gauge adds another term so this becomes the gauge invariant field strength

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

Finally, note that

$$V_{WZ}^2 = \frac{1}{2}\theta^2\bar{\theta}^2 A_\mu A^\mu \quad \text{and} \quad V_{WZ}^3 = 0 \tag{4.7}$$

This will be useful when constructing supersymmetric actions shortly.

4.1.1 The Field Strength and Action

We will build the action out of a field strength superfield, constructed from V by

$$W_\alpha = -\frac{1}{4}\bar{\mathcal{D}}^2\mathcal{D}_\alpha V$$

This has some nice properties. First, it is a chiral superfield, obeying $\bar{\mathcal{D}}_{\dot{\alpha}} W_\alpha = 0$. This follows from the fact that $\bar{\mathcal{D}}^3 = 0$. Second, it is invariant under the superfield gauge symmetry (4.2): the $\Omega^\dagger$ term is killed immediately by $\mathcal{D}_\alpha \Omega^\dagger = 0$, while the two $\bar{\mathcal{D}}$'s contrive to kill the Ω term. (You need one $\bar{\mathcal{D}}$ to get past the $\mathcal{D}_\alpha$ and the other $\bar{\mathcal{D}}$ to kill Ω.) The upshot is that any action formed from W_α will be automatically gauge invariant.

Next, we compute the components of W_α. This is a straightforward calculation but the number of terms involved gets rather large. Happily, things are easier if we appreciate that W_α is a chiral superfield since this means we only need to worry about the θ terms, with the $\bar{\theta}$ terms following automatically from the expansion (3.19). In components, the field strength superfield reads

$$W_\alpha(x,\theta) = \lambda_\alpha(x) + \theta_\alpha D(x) + (\sigma^{\mu\nu}\theta_\alpha) F_{\mu\nu}(x) - i\theta^2 \sigma^\mu_{\alpha\dot{\alpha}} \partial_\mu \bar{\lambda}^{\dot{\alpha}}(x) + \ldots$$

The first component of the chiral superfield W_α is a spinor, rather than a scalar, reflecting the fact that W_α is itself a spinor chiral superfield. Importantly, W_α contains the field strength $F_{\mu\nu}$.

Since W_α is chiral, we can integrate it over half of superspace to get a supersymmetric action. We have

$$\int d^2\theta \; W^\alpha W_\alpha = -\frac{1}{2} F_{\mu\nu} F^{\mu\nu} + \frac{i}{2} F_{\mu\nu} {}^\star F^{\mu\nu} - 2i\lambda\sigma^\mu \partial_\mu \bar{\lambda} + D^2$$

where the second term involves the *dual field strength*

$${}^\star F^{\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma}$$

This is like $F_{\mu\nu}$ but with the electric and magnetic fields swapped (one of them with a minus sign).

The term $iF_{\mu\nu} {}^\star F^{\mu\nu}$ is imaginary and so, at first glance, it looks like it will cancel when we add the hermitian conjugate $\int d^2\bar{\theta} \; W^\dagger_{\dot{\alpha}} W^{\dagger\,\dot{\alpha}}$. However, it turns out that this term plays an important role (at least this is true in the non-Abelian theories that we will discuss shortly) and we wish to keep it. This is achieved by introducing the gauge coupling constant e^2. Because this coupling constant sits in an F-term it is necessarily complex. We define

$$\tau = \frac{\vartheta}{2\pi} + \frac{4\pi i}{e^2}$$

And then write the Lagrangian

$$S_{\text{Maxwell}} = -\int d^4x \left[\int d^2\theta \frac{i\tau}{16\pi} W^\alpha W_\alpha + \text{h.c.}\right]$$
$$= \int d^4x \left[-\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} + \frac{\vartheta}{32\pi^2}F_{\mu\nu}{}^\star F^{\mu\nu} - \frac{i}{e^2}\lambda\sigma^\mu\partial_\mu\bar{\lambda} + \frac{1}{2e^2}D^2\right] \quad (4.8)$$

This is the supersymmetric Maxwell action. The propagating degrees of freedom are the $U(1)$ gauge field and a fermion λ that, in this context, is called the *gaugino* or, more specifically, the *photino*. There is also a real, auxiliary field D.

The parameter e^2 is the coupling constant. It doesn't do anything in Maxwell theory, which is free, but will come into play when we add matter. Note that we're working in a convention where there is a factor of $1/e^2$ that sits in front of the Maxwell action. As we'll see, the gauge coupling doesn't then sit anywhere else. This differs from the convention that we first met in Quantum Field Theory where the Maxwell term was canonically normalised but there was a gauge coupling inside the covariant derivatives. The two conventions are related by a rescaling $A_\mu \to eA_\mu$. Note that the photino λ similarly has an unconventionally normalised kinetic term, with a $1/e^2$.

Finally, there is the parameter ϑ. This is known as the *theta angle*. (We've used calligraphic script ϑ to distinguish it from the superspace coordinate θ.) Classically, the theta angle doesn't do anything. This is because it multiplies a total derivative

$$^\star F_{\mu\nu}F^{\mu\nu} = 2\partial_\mu(\epsilon^{\mu\nu\rho\sigma}A_\nu\partial_\rho A_\sigma)$$

However, things are more interesting in the quantum theory and the addition of such topological terms in the path integral can affect the dynamics. This is rather subtle for Maxwell theory, but underlies the story of 3d topological insulators. The effect is more pronounced in Yang-Mills theory and we'll discuss it further in Section 6. You can read (a lot) more about the theta angle in the lectures on Gauge Theory.

4.1.2 Supersymmetric QED

Next we add matter. This comes in the form of chiral multiplets Φ_i, where $i = 1, \ldots, N$. We want these to be charged under the $U(1)$ gauge field so that under a gauge transformation

$$A_\mu \to A_\mu + \partial_\mu\alpha$$

The components of the chiral multiplet transform with charges $q_i \in \mathbb{Z}$. This means that the lowest components transform as

$$\phi_i \to e^{i\alpha q_i}\phi_i \quad (4.9)$$

By necessity, the fermions ψ_i and auxiliary fields F_i in the chiral multiplet Φ_i must have the same charge,

$$\psi_i \to e^{i\alpha q_i}\psi_i \quad \text{and} \quad F_i \to e^{i\alpha q_i}F_i \tag{4.10}$$

From (4.3), this gauge transformation sits within a larger superfield transformation, under which

$$\Phi_i \to \exp\left(-2iq_i\Omega\right)\Phi_i$$

Note that this actually includes a larger symmetry than (4.9) and (4.10), because we can have, for example $\phi_i \to e^{i\alpha q_i}\phi_i$ with $\alpha \in \mathbb{C}$, rather than $\mathbb{R}$. This shifts the modulus of ϕ_i, as well as the phase. In mathematical language, the gauge group has been extended from $U(1)$ to the complexification $U(1)_{\mathbb{C}}$. However, the extra piece also shifts the additional fields in the vector multiplet, in particular the field C that transforms as $C \to C + \mathrm{Im}\,\alpha$, as shown in (4.4). This means that if we use the extra transformation to go to Wess-Zumino gauge, with $C = 0$, then we don't get to use it again on ϕ_i, and we're back to the more familiar $U(1)$ gauge transformations with $\alpha \in \mathbb{R}$.

The canonical Kähler potential that we've used so far is not gauge invariant:

$$\sum_{i=1}^{N}\Phi_i^\dagger\Phi_i \to \sum_{i=1}^{N}\exp\left(-2iq_i(\Omega-\Omega^\dagger)\right)\Phi_i\Phi_i^\dagger$$

However, it's simple to fix up. We simply need to use the new Kähler potential

$$K(\Phi_i,\Phi_i^\dagger,V) = \sum_{i=1}^{N}\Phi_i^\dagger e^{2q_iV}\Phi_i$$

with the transformation of V given in (4.2) rendering the whole expression gauge invariant. In Wess-Zumino gauge, the formulae (4.7) truncates at $e^{2qV} = 1+2qV+q^2V^2$. Integrating over superspace then gives

$$\int d^4\theta\ \Phi^\dagger e^{2qV}\Phi = \int d^4x\ \left[|\mathcal{D}_\mu\phi|^2 - i\bar\psi\bar\sigma^\mu\mathcal{D}_\mu\psi + |F|^2 - \sqrt{2}q\left(\phi\bar\lambda\bar\psi + \phi^\dagger\lambda\psi\right) + qD|\phi|^2\right]$$

Here the covariant derivatives are given by

$$\mathcal{D}_\mu\phi = \partial_\mu\phi - iqA_\mu\phi \quad \text{and} \quad \mathcal{D}_\mu\psi = \partial_\mu\psi - iqA_\mu\psi$$

The full action for an Abelian gauge theory then comes from combining the Maxwell action (4.8) with the matter fields. It is

$$\begin{aligned} S &= S_{\rm Maxwell} + \sum_{i=1}^{N} \int d^4x\ d^4\theta\ \Phi_i^\dagger e^{2q_i V} \Phi_i \\ &= \int d^4x \ \Big[-\frac{1}{4e^2} F_{\mu\nu}F^{\mu\nu} + \frac{\vartheta}{32\pi^2} F_{\mu\nu}{}^\star F^{\mu\nu} - \frac{i}{e^2}\lambda\sigma^\mu\partial_\mu\bar\lambda + \sum_{i=1}^{N} \left(|\mathcal{D}_\mu\phi_i|^2 - i\bar\psi\bar\sigma^\mu\mathcal{D}_\mu\psi_i \right) \\ &\qquad + \frac{1}{2e^2}D^2 + \sum_{i=1}^{N} \Big(|F_i|^2 - \sqrt{2}q_i \left(\phi_i\bar\lambda\bar\psi_i + \phi_i^\dagger\lambda\psi_i \right) + q_i D|\phi_i|^2 \Big) \Big] \end{aligned} \tag{4.11}$$

The first line contains the kinetic terms, the second the interactions. Note that there is a Yukawa coupling between the gaugino λ and the chiral multiplet fields, with $\phi^\dagger$ partnering ψ so that the Yukawa term is gauge invariant. In addition, there is a scalar potential that arises when we integrate out the auxiliary fields. The F terms don't do anything unless we also add a superpotential, while integrating out the D term results in the potential

$$V(\phi) = \frac{1}{2e^2}D^2 \quad \text{with} \quad D = e^2\left(\sum_{i=1}^{N} q_i|\phi_i|^2 \right) \tag{4.12}$$

Provided that there are both positive and negative charges q_i (and there must be as we explain below) then the potential has flat directions in which

$$\sum_{i=1}^{N} q_i|\phi_i|^2 = 0 \tag{4.13}$$

The existence of a moduli space of vacua is an important feature of supersymmetric gauge theories. We will study it more closely in Section 4.3

A First Look at the Anomaly

There's nothing wrong with (4.11) as a classical theory. But, as a quantum theory, it has a problem. It turns out that for most choices of the charges q_i, the quantum theory is sick. It has an inconsistency that goes by the name of a *gauge anomaly*.

We will have a lot to say about anomalies, gauge and otherwise, later in these lectures. For now we simply mention that the quantum theory only makes sense if the charges satisfy the following two conditions

$$\sum_{i=1}^{N} q_i = \sum_{i=1}^{N} q_i^3 = 0 \tag{4.14}$$

These conditions are not special to supersymmetric theories. They hold for any theory that has Weyl fermions coupled to a $U(1)$ gauge group. We'll say more about where these conditions come from in Section 5.2. For now, note that they require us to have fields with both positive and negative charges which, in turn, ensures that there are solutions to (4.13) with $\phi_i \neq 0$.

There are non-trivial solutions to the consistency conditions (4.14) but, for the most part, we will work with trivial solutions in which chiral multiplets come in pairs so that for each Φ with charge q there is a second chiral multiplet that we call $\tilde{\Phi}$ with charge $-q$. The conditions (4.14) are then automatically satisfied. Each pair Φ and $\tilde{\Phi}$ is sometimes referred to as a *flavour*. If a flavour is said to have charge q, it means that Φ has charge q and $\tilde{\Phi}$ charge $-q$.

The simplest example comprises of a $U(1)$ gauge field interacting with N flavours (which means $2N$ chiral multiplets) of charge $+1$. This theory is known as *supersymmetric QED*, or SQED for short. The action is

$$\begin{aligned} S_{\rm SQED} &= S_{\rm Maxwell} + \sum_{i=1}^{N} \int d^4x\, d^4\theta\, \left(\Phi_i^\dagger e^{2_i V} \Phi_i + \tilde{\Phi}_i^\dagger e^{-2_i V} \tilde{\Phi}_i \right) \\ &= \int d^4x \, \Big[-\frac{1}{4e^2} F_{\mu\nu}F^{\mu\nu} + \frac{\vartheta}{32\pi^2} F_{\mu\nu}{}^\star F^{\mu\nu} - \frac{i}{e^2}\lambda\sigma^\mu\partial_\mu\bar{\lambda} \\ &\quad + \sum_{i=1}^{N} \left(|\mathcal{D}_\mu\phi_i|^2 + |\mathcal{D}_\mu\tilde{\phi}_i|^2 - i\bar{\psi}\bar{\sigma}^\mu\mathcal{D}_\mu\psi_i - i\bar{\tilde{\psi}}_i\bar{\sigma}^\mu\mathcal{D}_\mu\tilde{\psi}_i \right) \\ &\quad - \sqrt{2}\sum_{i=1}^{N} \left(\phi_i^\dagger\lambda\psi_i - \tilde{\phi}_i^\dagger\lambda\tilde{\psi}_i + \text{h.c.} \right) - \frac{e^2}{2}\left(\sum_{i=1}^{N} |\phi_i|^2 - |\tilde{\phi}_i|^2 \right)^2 \Big] \end{aligned} \tag{4.15}$$

where we've integrated out both D-term and F-terms so the scalar potential takes the form (4.12).

When we first met QED in the lectures on Quantum Field Theory, we coupled a Dirac fermion to a $U(1)$ gauge field. This Dirac fermion contains two chiral fermions, one left-handed ψ and one right-handed $\bar{\chi}$, both with the same charge. If we conjugate the right-handed fermion then it becomes a left-handed fermion χ. We now have two left-handed fermions with equal and opposite charges. That's precisely the fermionic matter content in each flavour in (4.15).

Adding Further Terms

There are further terms that we can add to the action (4.15) (or, indeed, to the more general action (4.11)). We can add any superpotential $W(\Phi)$ provided that it is gauge

invariant. For example, we can always add to (4.15) the superpotential

$$W(\Phi, \tilde{\Phi}) = \sum_{i=1}^{N} m_i \tilde{\Phi}_i \Phi_i$$

This gives a mass $|m_i|$ to each chiral multiplet. In particular, the fermions get a Dirac mass. Note that such mass terms are only possible if there are pairs of chiral superfields with opposite charges.

There is one further, slightly curious term that we can add. This is known as the *Fayet-Iliopoulos term*,

$$\mathcal{L}_{\rm FI} = \int d^4\theta \; 2\zeta V = \zeta D \tag{4.16}$$

It is gauge invariant because D doesn't shift under the generalised gauge symmetry (4.2). Here $\zeta \in \mathbb{R}$ is the Fayet-Ilipoulos, or FI, parameter. Since this multiplies the D-term, it changes only the scalar potential (4.12) which becomes

$$V(\phi) = \frac{e^2}{2}\left(\sum_{i=1}^{N} q_i |\phi_i|^2 - \zeta\right)^2$$

In particular, supersymmetric vacua with $V(\phi) = 0$ now require some scalar field to get a non-vanishing expectation value which, in turn, breaks the $U(1)$ gauge symmetry.

4.2 Non-Abelian Gauge Theories

We can repeat everything above for non-Abelian gauge fields. We work with a gauge group G with Lie algebra

$$[T^A, T^B] = i f^{ABC} T^C$$

The factor of i in the commutation relations ensures that the generators are Hermitian, so $(T^A)^\dagger = T^A$. We normalise the generators in the fundamental (i.e. minimal) representation as

$$\mathrm{Tr}\, T^A T^B = \frac{1}{2}\delta^{AB} \tag{4.17}$$

In what follows, generators T^A will always be taken to be in the fundamental representation. If we need generators in other representations R then we will denote them as T^A_R. In these lectures we will mostly work with

$$G = SU(N_c)$$

with the subscript on N_c short for the number of "colours". We'll also mention results for other gauge groups as we go and, for now, keep things general.

4.2.1 Super Yang-Mills

Constructing supersymmetric Yang-Mills theory is slightly more fiddly version of what we did for Maxwell theory. We introduce a real superfield V in the adjoint of the gauge group. As usual, we can view an object in the adjoint representation as living in the Lie algebra by writing

$$V = V^A T^A \quad A = 1, \ldots, \dim G$$

For $G = SU(N_c)$, if we take T^A to be in the fundamental representation then this means that V is an $N_c \times N_c$ matrix. In terms of the components, we have a gauge field, but this is now accompanied by a fermion λ and auxiliary field D, both of which must also sit in the adjoint representation. Equivalently, all of them naturally live in the Lie algebra

$$A_\mu = A_\mu^A T^A \quad , \quad \lambda_\alpha = \lambda_\alpha^A T^a \quad , \quad D = D^A T^A$$

Again, for $SU(N_c)$ this means that each of these should be thought of as an $N_c \times N_c$ matrix (in addition to any vector or spinor index they carry). The fermion is again called a *gaugino* or sometimes a *gluino*.

We again want to generalise the usual non-Abelian gauge symmetry to something that can act on a superfield. We do this by taking an adjoint valued chiral superfield

$$\Omega = \Omega^A T^A$$

Since Ω is in the Lie algebra, $e^{i\Omega} \in G$ and this acts on the real superfield as

$$e^{2V} \to e^{-2i\Omega^\dagger} e^{2V} e^{2i\Omega}$$

From the Baker-Cambell-Hausdorff formula, $e^X e^Y = e^{X+Y+\frac{1}{2}[X,Y]+\cdots}$, we get the transformation law for the superfield itself

$$V \to V + i(\Omega - \Omega^\dagger) - i[V, \Omega + \Omega^\dagger] + \ldots$$

We can use the shift that appears in the first term to once again go to Wess-Zumino gauge where the real superfield takes the form (4.5), now with all fields in the adjoint of G. You can check that the remaining gauge symmetry acts on A_μ in the usual way,

$$A_\mu \to U A_\mu U^{-1} + iU\partial_\mu U^{-1}$$

with $U \in G$. The field strength lives in a chiral multiplet, defined as

$$W_\alpha = -\frac{1}{8}\bar{\mathcal{D}}^2 \left(e^{-2V} \mathcal{D}_\alpha e^{2V} \right)$$

Evaluated in Wess-Zumino gauge, we use the fact that $V^3 = 0$, as in (4.7), to expand $e^{2V} = 1 + 2V + 2V^2$. A short calculation then shows that

$$\begin{aligned} W_\alpha(y,\theta) &= -\frac{1}{4}\bar{\mathcal{D}}^2\left(\mathcal{D}_\alpha V - [V, \mathcal{D}_\alpha V]\right) \\ &= \lambda_\alpha(y) + \theta_\alpha D(y) + (\sigma^{\mu\nu}\theta)_\alpha F_{\mu\nu}(y) - i\theta^2\sigma^\mu_{\alpha\dot\beta}\mathcal{D}_\mu\bar\lambda^{\dot\beta}(y) \end{aligned}$$

with the non-Abelian field strength and covariant derivative defined by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu] \quad \text{and} \quad \mathcal{D}_\mu\lambda = \partial_\mu\lambda - i[A_\mu, \lambda]$$

To construct the action, we again define the complexified gauge coupling

$$\tau = \frac{\vartheta}{2\pi} + \frac{4\pi i}{g^2} \tag{4.18}$$

The action is then given by

$$\begin{aligned} S_{\text{SYM}} &= -\int d^4x\ \text{Tr}\left[\int d^2\theta\ \frac{i\tau}{8\pi}W^\alpha W_\alpha + \text{h.c.}\right] \\ &= \int d^4x\ \text{Tr}\Big[-\frac{1}{2g^2}F_{\mu\nu}F^{\mu\nu} + \frac{\vartheta}{16\pi^2}F_{\mu\nu}{}^\star F^{\mu\nu} - \frac{2i}{g^2}\lambda\sigma^\mu\mathcal{D}_\mu\bar\lambda + \frac{1}{g^2}D^2\Big] \end{aligned} \tag{4.19}$$

This is *super Yang-Mills*. After all that work, it's actually a very simple theory: just Yang-Mills coupled to a single, adjoint Weyl fermion. The factor of 2 differences compared to the Maxwell action (4.8) can be traced to the normalisation convention (4.17).

4.2.2 Supersymmetric QCD

We can add matter transforming in any representation R of the gauge group. The matter sits, as always, in a chiral superfield Φ that now transforms as

$$\Phi \to \exp\left(-2i\Omega^A T_R^A\right)\Phi \tag{4.20}$$

We construct a gauge invariant, supersymmetric action with the superfield expression

$$\begin{aligned} \int d^4x\, d^4\theta\ \Phi^\dagger e^{2V}\Phi &= \mathcal{D}_\mu\phi^\dagger\mathcal{D}^\mu\phi - i\bar\psi\bar\sigma^\mu\mathcal{D}_\mu\psi + F^\dagger F \\ &\quad - \sqrt{2}(\bar\psi\lambda^A T_R^A\phi + \phi^\dagger\lambda^A T_R^A\psi) + \phi^\dagger D^A T_R^A\phi \end{aligned}$$

Here the covariant derivatives include the gauge field transforming in the appropriate representation R.

Again, various anomaly cancellation conditions must be satisfied when coupling Weyl fermions to non-Abelian gauge groups in complex representations. The simplest way forward is to work instead with Dirac fermions. This means that we take pairs of chiral superfields, Φ transforming in some representation R and $\tilde{\Phi}$ in the conjugate representation $\bar{R}$. (In much of the literature, these superfields are denoted Q and $\tilde{Q}$ but we'll stick with Φ and $\tilde{\Phi}$ to avoid any unnecessary confusion with the supercharges.)

The most common is to take R to be the fundamental representation. We could, for example, consider $G = SU(N_c)$ gauge group with N_f flavours of fermions, each in the fundamental representation. The action is then

$$\begin{aligned} S_{\rm SQCD} = \int d^4x \ {\rm Tr}\Big[-\frac{1}{2g^2}F_{\mu\nu}F^{\mu\nu} + \frac{\vartheta}{16\pi^2}F_{\mu\nu}{}^\star F^{\mu\nu} - \frac{2i}{g^2}\lambda\sigma^\mu\mathcal{D}_\mu\bar{\lambda}\Big] \\ + \sum_{i=1}^{N_f}\Big[|\mathcal{D}_\mu\phi_i|^2 + |\mathcal{D}_\mu\tilde{\phi}|^2 - i\bar{\psi}_i\bar{\sigma}^\mu\mathcal{D}_\mu\psi_i - i\tilde{\psi}_i\sigma^\mu\mathcal{D}_\mu\bar{\tilde{\psi}}_i\Big] \\ - \sqrt{2}\sum_{i=1}^{N_f}\Big[\phi_i^\dagger\lambda\psi_i - \tilde{\phi}_i\bar{\lambda}\bar{\tilde{\psi}}_i + {\rm h.c.}\Big] - V(\phi,\tilde{\phi}) \end{aligned} \tag{4.21}$$

Here the covariant derivatives are

$$\mathcal{D}_\mu\phi = \partial_\mu\phi - iA_\mu\phi \quad \text{and} \quad \mathcal{D}_\mu\psi = \partial_\mu\psi - iA_\mu\psi$$

for the fields in the fundamental representation, and

$$\mathcal{D}_\mu\tilde{\phi} = \partial_\mu\tilde{\phi} + i\tilde{\phi}A_\mu \quad \text{and} \quad \mathcal{D}_\mu\tilde{\psi} = \partial_\mu\tilde{\psi} + i\tilde{\psi}A_\mu$$

for those in the anti-fundamental representation. Finally, the scalar potential is again given by the D-terms

$$V(\phi,\tilde{\phi}) = \frac{1}{2g^2}D^AD^A \quad \text{with} \quad D^A = -g^2\sum_{i=1}^{N_f}\Big(\phi_i^\dagger T^A\phi_i - \tilde{\phi}_iT^A\tilde{\phi}_i^\dagger\Big) \tag{4.22}$$

with T^A the $N_c \times N_c$ generators in the fundamental representation. This is the action of *supersymmetric QCD*, or SQCD for short. In a nod to the real world, we refer to the fermions ψ and $\tilde{\psi}$ as *quarks*. Their supersymmetric scalar partners ϕ and $\tilde{\phi}$ are called *squarks*.

Once again, we can also add masses for the quark multiplets by including the gauge invariant superpotential

$$\mathcal{W}(\Phi,\tilde{\Phi}) = \sum_{i=1}^{N_f} m_i\tilde{\Phi}_i\Phi_i$$

This gives an extra term to the scalar potential

$$\delta\mathcal{L}_{\text{mass}} = -\sum_{i=1}^{N_f} |m_i|^2 \left(|\phi_i|^2 + |\tilde{\phi}_i|^2\right)$$

as well as Dirac masses for ψ_i and $\tilde{\psi}_i$.

There is no FI parameter that we can add for non-Abelian theories. The non-Abelian analog of (4.16) would involve Tr D but the trace of the generators of any non-Abelian Lie algebra always vanishes. Fayet-Iliopoulos terms can only be introduced for $U(1)$ gauge theories.

4.3 The Moduli Space of Vacua

In the absence of a superpotential, supersymmetric gauge theories do not have a unique ground state. Instead, the D-term potential has a flat direction with $V(\phi) = 0$. This is the moduli space of vacua. It will turn out that this moduli space holds the key to understanding the quantum dynamics of supersymmetric gauge theories. For this reason, we will spend some time studying its structure.

Consider, for example, $U(1)$ SQED with a single flavour. If we don't turn on a FI parameter then the D-term is (4.15)

$$D = -g^2(|\phi|^2 - |\tilde{\phi}|^2)$$

Clearly any solution with

$$|\phi|^2 = |\tilde{\phi}|^2 = v^2$$

has zero energy. To fully specify the classical theory, we must decide where on this moduli space we want to sit.

At all points on the moduli space, there are always massless particles. Indeed, the low-energy physics is dominated by the fluctuations along the moduli space, which always correspond to massless particles, together with their fermionic superpartners.

Meanwhile, the masses of heavy particles typically depend on where you sit on the moduli space which, in the current example, means that value of v^2. Because ϕ is charged under the $U(1)$ gauge field, when it gets an expectation value, the Higgs mechanism kicks in and the photon gets a mass of order

$$m_\gamma^2 \sim e^2 v^2$$

But the Yukawa terms in (4.15) mean that a particular combination of fermions also gets a mass, given by

$$m_{\text{fermion}} \sim ev$$

The fact that this is the same as m_γ is, of course, no coincidence: the photon, massive fermion and an additional massive scalar in the spectrum form a massive vector multiplet of the kind discussed in Section 2.3. The origin of the moduli space, at $\phi = \tilde{\phi} = 0$, is special because here the vector multiplet becomes massless.

The Geometry of Moduli Space

We denote the moduli space of vacua as $\mathcal{M}$. As we now explain, this manifold naturally comes with a number of interesting geometric structures.

First $\mathcal{M}$ is defined by the requirement that $V(\phi) = 0$. In the absence of a superpotential, this is equivalent to $D(\phi) = 0$. (Note that here ϕ denotes all chiral multiplet scalars and, for SQED and SQCD, this means both ϕ and $\tilde{\phi}$.). However, we should also remember that the gauge group G acts on these scalars. The gauge symmetry is not really a symmetry of the theory, but rather a redundancy in our description. This means that any two values of ϕ related by a gauge transformation should be viewed as physically equivalent. The upshot is that the vacuum moduli space $\mathcal{M}$ is defined as the quotient

$$\mathcal{M} = \{\phi \,|\, D(\phi) = 0\}/G \tag{4.23}$$

We have stumbled upon a construction known to mathematicians as the *symplectic reduction*. It's particularly natural because, as we've seen above, the D-term constraint $D(\phi) = 0$ is fully specified by the action of the group G. In this way, the group G gets to act twice: once as a constraint, and again as a quotient. Mathematicians call the constraint $D(\phi) = 0$ the *moment map*. If G includes an Abelian factor, the associated FI parameter is known as the *level*.

There are two, further ways to describe the moduli space $\mathcal{M}$. We will now describe these, but won't prove the equivalence with (4.23). Instead, we will content ourselves with some heuristic justification, followed by some examples[2].

[2]A full proof can be found in the paper by Marcus Luty and Wati Taylor, Varieties of vacua in classical supersymmetric gauge theories.

The fact that the group G "acts twice", is even more apparent if the second way of writing the moduli space: it is the holomorphic quotient

$$\mathcal{M} = \{\,\phi\,\}/G_{\mathbb{C}} \tag{4.24}$$

with $G_{\mathbb{C}}$ the *complexified gauge group*. This means that we take the real parameters α that usually specify a gauge transformation – that is $\phi \to e^{iq\alpha}\phi$ for Abelian G or $\phi \to e^{i\alpha^a T^a_R}\phi$ for non-Abelian – and quotient by transformations with $\alpha \in \mathbb{C}$. You should think of the D-term constraint in (4.23) as like a gauge-fixing condition for the non-Hermitian part of the $G_{\mathbb{C}}$ transformations.

In fact, looking back at our construction of supersymmetric gauge theories, the gauge transformations started life in a chiral superfield Ω where everything was complex. They became real only after moving to Wess-Zumino gauge. From the perspective of supersymmetric gauge theory, the equivalence of (4.23) and (4.24) is best seen by looking at the more general gauge transformations before imposing Wess-Zumino gauge.

The final description of the moduli space will, in some circumstances, turn out to be the most useful. The manifold $\mathcal{M}$ can alternatively be viewed as

$$\mathcal{M} = \{\text{Gauge invariant, holomorphic monomials}\} \,/\, \{\text{Algebraic relations}\} \tag{4.25}$$

This is a description of $\mathcal{M}$ in terms of what mathematicians call an *algebraic variety*. This definition is best elucidated by examples that we will turn to below, but here we give the basic gist.

There are three key ideas that we need to explain in this definition: gauge invariant, holomorphic, and the algebraic relations. We cover each in turn:

- Because gauge symmetry is merely a redundancy in our choice of description, it should be possible to describe the dynamics of massless particles in terms of some gauge invariant fields. This is the basic idea underlying the characterisation (4.25)

- It's always possible to build such gauge invariant fields by taking combinations like $\phi^\dagger\phi$. These are invariant under G, but not invariant under the larger $G_{\mathbb{C}}$ that defines the moduli space according to (4.24). The need to impose invariance under $G_{\mathbb{C}}$, or equivalently the need to impose the D-term constraint $D = 0$, means that we should work with holomorphic gauge invariant combinations, meaning monomials that involve ϕ alone and not $\phi^\dagger$. Alternatively, and more physically, supersymmetry means that we should be able to describe the fields in terms of chiral multiplets, and these are necessarily holomorphic.

- Finally, it will turn out that, for some examples, not all of the gauge invariant combinations are independent. This is why there is the need to quotient by certain relations between them. This is best illustrated when we turn to examples below.

Mathematically, the equivalence between the quotient constructions (4.23) and (4.24) and the algebraic description (4.25) goes by the name of *geometric invariant theory*.

4.3.1 The Moduli Space of SQED

We'll start by looking at the simpler case of SQED. This is a $U(1)$ gauge theory coupled to N flavours. If we set the FI parameter to zero for now, then the D-term condition is (4.15)

$$\sum_{i=1}^{N} |\phi^i|^2 - |\tilde{\phi}_i|^2 = 0 \tag{4.26}$$

In addition, we should quotient by the $U(1)$ gauge action

$$\phi^i \to e^{i\beta}\phi^i \quad \text{and} \quad \tilde{\phi}_i \to e^{-i\beta}\tilde{\phi}_i \tag{4.27}$$

We started with $2N$ fields ϕ and $\tilde{\phi}$. There is one real constraint (4.26) which, together with the quotient (4.27) reduces the *complex* dimension of the vacuum moduli space by one. We then have

$$\dim \mathcal{M} = 2N - 1 \tag{4.28}$$

Let's see how to reproduce this counting when thinking of $\mathcal{M}$ as an algebraic variety defined by (4.25). The gauge invariant monomial are the bilinears

$$M_j{}^i = \tilde{\phi}_j \phi^i \tag{4.29}$$

We will refer to these, not entirely accurately, as "mesons". There are N^2 such fields and, at first glance, it looks like we have way too many. However, they are not all independent and this is where the algebraic relations in (4.25) come into play.

The meson matrix M is built from vectors ϕ and $\tilde{\phi}$ and so has, at most, rank 1. This means that there are $N-1$ eigenvalues that are guaranteed to vanish. In general, the determinant of an $N \times N$ matrix A can be written as

$$\epsilon_{i_1 \ldots i_N} A^{i_1}_{j_1} \ldots A^{i_N}_{j_N} = \det A \, \epsilon_{j_1 \ldots j_N}$$

The rank 1 matrix M therefore obeys

$$\epsilon_{i_1 \ldots i_N} (M_{j_1}{}^{i_1} - \lambda \delta_{j_1}{}^{i_1}) \ldots (M_{j_N}{}^{i_N} - \lambda \delta_{j_N}{}^{i_N}) = \det(M - \lambda)\, \epsilon_{j_1 \ldots j_N} = \lambda^{N-1}(\lambda - \lambda_0)\epsilon_{j_1 \ldots j_N}$$

This tells us that if we expand out the left-hand side, all terms of order λ^{N-2} and lower must vanish for a rank 1 matrix. In other words, we have the constraints

$$\epsilon_{i_1 \ldots i_N} M_{j_1}^{\ i_1} M_{j_2}^{\ i_2} = 0 \tag{4.30}$$

with all other constraints following by contracting with further $M_j^{\ i}$. Our next task is to count how many independent constraints we have here. The $i_3, \ldots, i_N$ indices are left hanging so by picking these we can restrict i_1 and i_2 to run over any pair. But the resulting constraints aren't all independent. For example, there is a constraint that arises from $(i_1, i_2) = (1, 2)$ and another that arises from $(i_1, i_2) = (1, 3)$. But dividing the first constraint by the second, and rearranging, gives the constraint that arise from $(i_1, i_2) = (2, 3)$. In fact, it's not hard to convince yourself that the constraints that come from $(i_1, i_2) = (1, \text{anything but } 1)$ are independent and sufficient to give all others. Clearly there are $N - 1$ of these.

For each of these constraints, we still have the (j_1, j_2) indices hanging. These too are anti-symmetrised and the same argument that we gave above for (i_1, i_2) also holds for (j_1, j_2). This means that the total number of constraints from (4.30) is $(N - 1)^2$. The algebraic variety $\mathcal{M}$, defined by all mesons (4.29) subject to the constraints (4.30) then has complex dimension

$$\dim \mathcal{M} = N^2 - (N - 1)^2 = 2N - 1$$

in agreement with our earlier counting (4.28).

The Metric on the Vacuum Moduli Space

The vacuum moduli space inherits a natural metric. Indeed, if we restrict to very low energies the dynamics is that of the massless fields, corresponding to fluctuations along the moduli space. This is the realm of the non-linear sigma model that we discussed in Section 3.2.4. On general grounds, we know that not only is there a metric on $\mathcal{M}$ but this metric must be Kähler.

It is straightforward to compute this metric. Here we do it in two different ways for the simplest case of $N = 1$ flavour. The easiest way to proceed is to start with the Kähler potential

$$K = \phi^\dagger \phi + \tilde{\phi}^\dagger \tilde{\phi}$$

Note that the Kähler potential for a gauge theory involves terms like e^{2qV}, with V the real superfield, to ensure gauge invariance. We simply set the gauge fields to zero in the

following calculation, so the Kähler potential is the canonical one above. Restricting to the moduli space (4.26), we have $|\phi|^2 = |\tilde{\phi}|^2$. Furthermore, if we work with the meson field $M = \tilde{\phi}\phi$, the Kähler potential becomes

$$K = 2|\phi|^2 = 2\sqrt{M^\dagger M} \tag{4.31}$$

The associated metric is just

$$ds^2 = \frac{|dM|^2}{2|M|} \tag{4.32}$$

We see immediately that the metric is singular at the origin $M = 0$. This singularity is telling us something important: when $\phi = \tilde{\phi} = 0$, there are new massless degrees of freedom. This is simply the photon and its superpartner which become massless at the origin because the Higgs mechanism turns off.

This is a lesson that we've seen before. When we integrated out heavy fields in Section 3.3, we found that the low-energy effective theory had singularities at points where the heavy fields became light. This is a general feature of low-energy effective theories, and one that will be important in Section 6 when we come to discuss the quantum dynamics of these theories. For now, the lesson is worth repeating one more time: singularities in the low-energy effective action signal the emergence of new, massless degrees of freedom.

There is a more prosaic way to do this same calculation that highlights our original quotient description of the vacuum moduli space (4.23). The general solution to the constraint (4.26) is

$$\phi = ve^{i\alpha}e^{i\beta} \quad \text{and} \quad \tilde{\phi} = ve^{i\alpha}e^{-i\beta}$$

with $v > 0$. The $e^{\pm i\beta}$ has been taken to coincide with the gauge action (4.27), so that v and α provide the coordinates on the moduli space $\mathcal{M}$.

At this point, there's an important factor of 2 that we have to take care of. The parameter β corresponding to the $U(1)$ gauge transformation has range $\beta \in [0, 2\pi)$. In contrast, we have $\alpha \in [0, \pi)$. This follows because we can always implement a gauge transformation with $\beta = \pi$ which flips the sign of ϕ and $\tilde{\phi}$ or, equivalently, takes $\alpha \to \alpha + \pi$.

The metric on $\mathcal{M}$ is inherited from the kinetic terms for the scalar fields. To this end, we promote v, α and β to fields that vary slowly over spacetime. The covariant derivatives are

$$\begin{aligned}\mathcal{D}_\mu\phi &= (\partial_\mu v + iv(\partial_\mu\alpha + \partial_\mu\beta - A_\mu))\, e^{i(\alpha+\beta)} \\ \mathcal{D}_\mu\tilde{\phi} &= (\partial_\mu v + iv(\partial_\mu\alpha - \partial_\mu\beta + A_\mu))\, e^{i(\alpha-\beta)}\end{aligned}$$

We now choose $A_\mu = \partial_\mu\beta$ to absorb the variation of β. This how the quotient in (4.23) manifests itself in this calculation. The kinetic terms for the scalar fields, restricted to the vacuum moduli space, then become

$$\mathcal{L}_{\text{eff}} = |\mathcal{D}\phi|^2 + |\mathcal{D}\tilde{\phi}|^2 = 2\Big[\partial v^2 + v^2\,\partial\alpha^2\Big] \tag{4.33}$$

which we interpret as a metric like the non-linear sigma models (3.25) we discussed earlier. It's straightforward to check that this coincides with the metric (4.32) written in terms of the meson field.

At first glance, (4.33) looks like a flat metric. And, indeed, it is. But it's not the flat metric on $\mathbb{C}$ because the angular coordinate α doesn't have periodicity 2π. Instead, it's the flat metric on $\mathbb{C}/\mathbb{Z}_2$ and has a conical singularity at the origin $v = 0$. This how we see the emergence of the massless photon at this point.

Turning on the FI Parameter

A small variation on this calculation provides yet another perspective on the importance of singularities in the low-energy effective action. We again consider SQED with $N = 1$ flavour, but this time turn on a FI parameter. The D-term constraint now reads

$$|\phi|^2 - |\tilde{\phi}|^2 = \zeta \tag{4.34}$$

We assume that $\zeta \geq 0$. In the ground state, we necessarily have $|\phi|^2 \neq 0$ meaning that the photon now gets a mass on all points of the moduli space.

We can see how this manifests itself in the moduli space metric. The condition (4.34) is solved by

$$\phi = \sqrt{v^2 + \zeta}\, e^{i\alpha} e^{i\beta} \quad \text{and} \quad \tilde{\phi} = v e^{i\alpha} e^{-i\beta}$$

Our previous calculation to compute the metric on $\mathcal{M}$ is now a little more involved. The subtlety lies in figuring out what expression we should take for the gauge field A_μ. The answer can be found in its equation of motion. Or, more precisely, the equation of

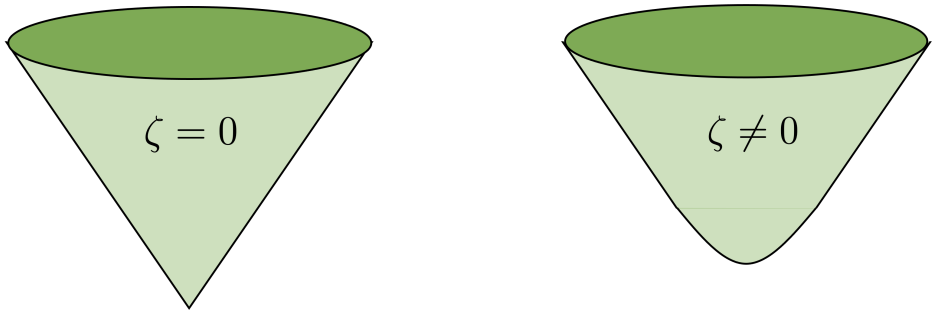


Figure 5. The moduli space of SQED. When $\zeta = 0$, the moduli space is the singular cone $\mathbb{C}/\mathbb{Z}_2$ shown on the left. The singularity at the origin reflects the existence of the massless photon. When $\zeta \neq 0$ the singularity is resolved and the moduli space is the smooth cone shown on the right. Now the photon is Higgsed everywhere on the moduli space.

motion in the limit $e^2 \to \infty$ where we neglect the Maxwell term. This is the appropriate limit when the gauge field responds immediately to fluctuations in the scalar and gives

$$A_\mu = \frac{\zeta}{2v^2 + \zeta}\,\partial_\mu \alpha + \partial_\mu \beta$$

It reduces to our previous, pure gauge, choice when $\zeta = 0$. Inserting this expression into the kinetic terms for ϕ and $\tilde{\phi}$, we compute the metric on the vacuum moduli space

$$\mathcal{L}_{\text{eff}} = |\mathcal{D}\phi|^2 + |\mathcal{D}\tilde{\phi}|^2 = \frac{2v^2 + \zeta}{v^2 + \zeta}\left[\partial v^2 + \frac{4v^2(v^2+\zeta)^2}{(2v^2+\zeta)^2}\partial\alpha^2\right] \tag{4.35}$$

Importantly, as we approach the origin, $v^2 \to 0$, the metric is well approximated by

$$ds^2 \approx dv^2 + 4v^2 d\alpha^2 = dv^2 + v^2 d(2\alpha)^2$$

That extra factor of 2 makes all the difference! We now get the flat metric with the angular coordinate $2\alpha \in [0, 2\pi)$ which means that close to $v = 0$ the metric really does look like flat space. The resulting moduli space is sketched in Figure 5.

4.3.2 The Moduli Space of SQCD

We now play the same game for SQCD. We will take gauge group

$$G = SU(N_c)$$

coupled to N_f fundamental flavours, ϕ_a^i in the fundamental representation and $\tilde{\phi}_i^a$ in the anti-fundamental. Here $a = 1, \ldots, N_c$ labels is the gauge group index while $i = 1, \ldots N_f$ is the flavour index.

The generators $(T^A)^a{}_b$ in the fundamental representation are the set of Hermitian, traceless, complex $N_c \times N_c$ matrices. Meanwhile, the generators in the anti-fundamental representation are simply $\bar{T}^A = -T^A$. The $N_c^2 - 1$ D-term conditions (4.22) are then

$$\phi_i^\dagger T^A \phi^i - \tilde{\phi}_i T^A \tilde{\phi}^{\dagger\, i} = 0 \qquad A = 1, \ldots N_c^2 - 1$$

where there is an implicit sum over $i = 1, \ldots, N_f$. To get a better sense of these constraints, let us first relax the requirement that T^A is traceless. (This is what we would get if the gauge group was $U(N_c)$ rather than $SU(N_c)$.) In this case, the T^A provide a basis for all Hermitian matrices and the D-term condition is N_c^2 constraints

$$\phi_i^{\dagger\, a} \phi_b^i - \tilde{\phi}_i^a \tilde{\phi}_b^{\dagger\, i} = 0 \qquad a, b = 1, \ldots N_c \;\text{ for } U(N_c)$$

But the fact that we're working with $SU(N_c)$ rather than $U(N_c)$ means that there's no reason to set the trace to zero. So our true D-term constraint is

$$\phi_i^{\dagger\, a} \phi_b^i - \tilde{\phi}_i^a \tilde{\phi}_b^{\dagger\, i} = \frac{1}{N_c} \left(\phi_i^{\dagger\, c} \phi_c^i - \tilde{\phi}_i^c \tilde{\phi}_c^{\dagger\, i} \right) \delta^a{}_b \tag{4.36}$$

At first glance, this looks like it's still N_c^2 conditions. But if you take the trace then you find that both sides are trivially equal. This means that, in fact, it's only $N_c^2 - 1$ conditions, with no condition on the trace. This is what we wanted.

To understand the vacuum moduli space, we must first solve the equations (4.36). As we will now see, the nature of the solutions is different for $N_f < N_c$ and $N_f \geq N_c$. We deal with each in turn.

$\underline{N_f < N_c}$

We'd like to count the dimension of the moduli space $\mathcal{M}$, defined by (4.36) modulo gauge transformations. It's tempting to think that there are just $N_c^2 - 1$ constraints in (4.36) but how do we know that they are all independent? In fact, it's simple to see that these constraints cannot all be independent when $N_f < N_c$ because then we would have more constraints than degrees of freedom. Yet solutions to (4.36) certainly exist! To proceed, we use the fact that the D-terms and gauge symmetry are closely entwined. The D-terms only bite when the gauge symmetry does.

When $N_f < N_c$, we can always use an $SU(N_c)$ gauge transformations and $SU(N_f)$ flavour rotations to put the matrix ϕ in the block-diagonal form

$$\phi_a^i = \left(\begin{array}{ccc} v_1 & \ldots & 0 \\ & \ddots & \\ 0 & \ldots & v_{N_f} \\ \hdashline 0 & \ldots & 0 \end{array} \right) \tag{4.37}$$

Here the columns have length N_c and the rows length N_f. We can then use the other $SU(N_f)$ to rotate $\tilde{\phi}$ to be in upper-diagonal form. (We can't make it fully diagonal because we've already used up the $SU(N_c)$ to diagonalise ϕ). However, now we invoke the D-term conditions (4.36). The only solutions to these conditions require that the off-diagonal terms in $\tilde{\phi}$ vanish. (You could check this for a simple case, say $N_c = 3$ and $N_f = 2$ to get a feel for why this is the case.) We're left with

$$\tilde{\phi}_a^{\dagger\, i} = \phi_a^i$$

As before, points on the moduli space related by a gauge transformation are to be physically identified. On a generic point on the moduli space (with $v_i \neq v_j \neq 0$ when $i \neq j$) the gauge group is broken to

$$SU(N_c) \to SU(N_c - N_f)$$

The number of broken gauge generators is then

$$\#\text{ broken generators} = (N_c^2 - 1) - ((N_c - N_f)^2 - 1)$$

Each of these is eaten by one of the original $2N_cN_f$ bosons ϕ and $\tilde{\phi}$. This means that the resulting vacuum moduli space has complex dimension

$$\dim \mathcal{M} = 2N_cN_f - [\#\text{ broken generators}] = N_f^2$$

Note that we only divide out by the points on the moduli space related by the $SU(N_c)$ gauge symmetry. There will still be points on the moduli space related by the flavour symmetry $SU(N_f)$ but these are physically distinct vacua.

We can also view the moduli space as an algebraic variety. Once again, the holomorphic monomials are the meson fields

$$M_j^{\ i} = \tilde{\phi}_j^a \phi_a^i \tag{4.38}$$

This time the name "meson" is more appropriate: we have contracted the gauge indices of ϕ and $\tilde{\phi}$ to form a gauge invariant composite. The mesons form N_f^2 fields but, in contrast to SQED, there is no constraint on M. The contracted gauge indices in (4.38) run over $a = 1, \ldots, N_c > N_f$ so there is no obstacle to M being maximal rank. We see immediately that $\dim \mathcal{M} = N_f^2$, in agreement with our result above.

We can compute the metric on $\mathcal{M}$ along the same lines as we saw for SQED. The Kähler potential is

$$K = \phi_i^{\dagger\, a}\phi_a^i + \tilde{\phi}_i^a\tilde{\phi}_a^{\dagger\, i}$$

We want to write this in terms of the meson field (4.38). To do this, first note that for $N_f < N_c$ the trace term on the right-hand side of the D-term (4.36) vanishes when restricted to the moduli space and we have

$$\phi_i^{\dagger\, a}\phi_b^i = \tilde{\phi}_i^a\tilde{\phi}_b^{\dagger\, i} \tag{4.39}$$

From this, we have

$$(M^\dagger M)_j{}^i = \tilde{\phi}_a^{\dagger\, i}\phi_k^{\dagger\, a}\phi_b^k\tilde{\phi}_j^b = (\tilde{\phi}_a^{\dagger\, i}\tilde{\phi}_k^a)(\tilde{\phi}_b^{\dagger\, k}\tilde{\phi}_j^b)$$

where, in the last equality, we've used (4.39). Taking the square root of this matrix equation tells us that $(\tilde{\phi}^\dagger\tilde{\phi})^i_{\ j} = (\sqrt{M^\dagger M})^i_{\ j}$, and so the Kähler potential is

$$K = 2\,\mathrm{Tr}\sqrt{M^\dagger M} \tag{4.40}$$

Just like the Kähler potential for SQED (4.31), the resulting metric will have singularities whenever M^{-1} ceases to exist. Again, these singularities correspond to new degrees of freedom becoming massless. At a generic point on the moduli space, there will be massless gauge bosons associated to the unbroken $SU(N_c - N_f)$ gauge symmetry. But along the loci on which M is not invertible we have an enhancement of the gauge group and new massless gauge bosons.

$\underline{N_f \geq N_c}$

For $N_f \geq N_c$, the story is different. First, we can now use $SU(N_c)$ and $SU(N_f)$ transformations to find solutions to the D-term equations (4.36), again in block-diagonal form

$$\phi_a^i = \left(\begin{array}{ccc|c} v_1 & \dots & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \dots & v_{N_c} & 0 \end{array}\right) \quad \text{and} \quad \tilde{\phi}_a^{\dagger\, i} = \left(\begin{array}{ccc|c} \tilde{v}_1 & \dots & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \dots & \tilde{v}_{N_c} & 0 \end{array}\right)$$

with

$$|v_a|^2 = |\tilde{v}_a|^2 + \rho \qquad a = 1, \dots, N_c$$

where ρ must be independent of a. This reflects the fact that the trace term on the right-hand side of (4.36) can now be non-zero.

At a generic point on $\mathcal{M}$, the $SU(N_c)$ gauge symmetry is completely broken. The complex dimension of the moduli space is therefore

$$\dim \mathcal{M} = 2N_cN_f - (N_c^2 - 1) \tag{4.41}$$

How can we describe this moduli space as an algebraic variety? The meson fields (4.38) provide N_f^2 degrees of freedom, but now there are constraints of the kind we met for SQED since M is at most rank N_c. In addition, there are also new gauge invariant fields. These are *baryons*, built from the totally anti-symmetric invariant tensor of $SU(N_c)$,

$$\begin{aligned} B^{i_1\ldots i_{N_c}} &= \phi^{i_1}_{a_1}\ldots\phi^{i_{N_c}}_{a_{N_c}}\epsilon^{a_1\ldots a_{N_c}} \\ \tilde{B}_{i_1\ldots i_{N_c}} &= \tilde{\phi}^{a_1}_{i_1}\ldots\tilde{\phi}^{a_{N_c}}_{i_{N_c}}\epsilon_{a_1\ldots a_{N_c}} \end{aligned}$$

Each of these is anti-symmetric in the N_c different flavour indices $i_1,\ldots,i_{N_c}$. There are then a bunch of further constraints between these baryons and mesons. Rather than doing this in full generality, we'll instead just describe how this works for the two cases that will prove most interesting in Section 6.

- $N_f = N_c$: In this case, anti-symmetry properties mean that there is just a single baryon of each type

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

  The meson M can have rank N_f, so there are no constraints there. But there is a single relation between the mesons and baryons, given by

  $$\tilde{B}B = \det M \tag{4.42}$$

  This means that there are N_f^2+2 degrees of freedom in M, B and $\tilde{B}$ and a single relation, giving a moduli space of dimension $\dim\mathcal{M} = N_f^2+1$ in agreement with (4.41). The relation (4.42) will play a starring role when we come to consider the quantum theory in Section 6.3.

- $N_f = N_c + 1$: Now there are N_f baryons of each type,

  $$B_j = \epsilon_{ji_1\ldots i_{N_c}}B^{i_1\ldots i_{N_c}} \quad \text{and} \quad \tilde{B}^j = \epsilon^{ji_1\ldots i_{N_c}}\tilde{B}_{i_1\ldots i_{N_c}}$$

  This time the constraints are less obvious, but they turn out to be

  $$\mathrm{Adj}(M)_i{}^j = B_i\tilde{B}^j \quad \text{and} \quad M_j{}^iB_i = M_j{}^i\tilde{B}^j = 0 \tag{4.43}$$

  where $\mathrm{Adj}(M)$ is the adjugate matrix, which is the transpose of the matrix of cofactors. The adjugate matrix is most familiar when M is invertible, in which case $\mathrm{Adj}(M) = (\det M)M^{-1}$. However, the conditions $BM = M\tilde{B} = 0$ tell us that M has a zero eigenvalue and so is not invertible.

At this point, things start to get a little messy! It turns out that not all the relations (4.43) are independent, but there's no way to write them as a smaller set. Mathematicians say that the resulting variety is not a *complete intersection*. We'll simply duck the issue which, it turns out, will not hinder us from understanding the physics.

There is one sense in which the use of the words "mesons" and "baryons" might be misleading. In QCD, mesons and baryons are bound states of quarks, stuck together because of confinement. But confinement is a surprising and poorly understood property of the quantum theory. Here we are not invoking anything so dramatic. Indeed, we haven't yet discussed any quantum effects and what we've call SQCD might better be called SCCD for our current purposes. Instead, we're using meson and baryon fields simply because they are gauge invariant and so free of any gauge redundancy. We'll turn on the Q in SQCD in Section 6 where we'll see how this tallies with ideas of confinement.

4.3.3 Briefly, Gauged Linear Sigma Models in 2d

We've learned that we can construct interesting geometric spaces as the moduli spaces of vacua of supersymmetric gauge theories. This kind of construction goes by the name of *gauged linear sigma models*. It turns out that it's a particularly useful method when wielded in quantum field theories in $d = 1 + 1$ dimensions.

To see why, first consider the action for a non-linear sigma model in general d-dimensional spacetime

$$S = \int d^d x \; g_{ij}(\pi) \, \partial_\mu \pi^i \partial^\mu \pi^j \tag{4.44}$$

Here π^i are coordinates on a manifold $\mathcal{M}$ with metric g_{ij}.

When $d = 0 + 1$, we're dealing with the quantum mechanics of particle moving on $\mathcal{M}$. But we know what happens in this case: the wavefunction will spread over $\mathcal{M}$ and there will typically be a unique ground state.

This is conceptually very different from what happens in $d = 3 + 1$ dimensions. There, each point on $\mathcal{M}$ defines a different ground state of the system. There is no spread of the wavefunction.

The reason for this different behaviour can be traced to the long-distance property of the propagator. The propagator grows in $d = 0 + 1$ and $d = 1 + 1$ dimensions (logarithmically in the latter case) while it decays in $d = 2 + 1$ and higher. This fact

is closely related to the *Mermin-Wagner theorem* which says that global symmetries cannot be spontaneously broken in $d = 0 + 1$ and $d = 1 + 1$ dimensions. (We met this theorem in the lectures on Statistical Field Theory and Gauge Theory.)

In the context of non-linear sigma models of the type (4.44), this long-distance behaviour of the propagator is telling us that $d = 0+1$ and $d = 1+1$ dimensions are special because the wavefunction spreads over the manifold $\mathcal{M}$. This means that the ground state of the system has a chance of knowing something about the global structure of the manifold $\mathcal{M}$, like its topology. Indeed, studying the dynamics of low-dimensional quantum systems on $\mathcal{M}$ has been a very fruitful source of developments in mathematics. This beginnings of this story are told in the lectures on Supersymmetric Quantum Mechanics.

The story is particularly rich for theories in $d = 1 + 1$ dimensions where, in addition to the wavefunction spreading over $\mathcal{M}$, the UV divergences of the quantum field theory mean that the metric on $\mathcal{M}$ is renormalised. At one-loop, the running is captured by the beautifully geometric RG equation

$$\mu \frac{\partial g_{ij}}{\partial \mu} = R_{ij} \tag{4.45}$$

where μ is the RG scale and R_{ij} the Ricci tensor. This formula is known as *Ricci flow*. It plays an important role in String Theory and has a number of applications in pure mathematics. Note that the flow stops only if the metric becomes Ricci flat, with $R_{ij} = 0$. At this point we have a 2d conformal field theory. However, not all manifolds admit such a Ricci flat metric.

Things become even more interesting when we throw supersymmetry into the mix. This is what we called $\mathcal{N} = (2, 2)$ supersymmetry in Section 2.4.3. It not only gives us an important level of control over the dynamics but, as we've seen already in these lectures, dovetails nicely with some interesting mathematical structures. It turns out that the gauge theory approach to realising non-linear sigma models as the vacuum moduli space is particularly powerful in this context. Here we just give a hint of how this works

First, the anomaly cancellation conditions (4.14) are for 4d quantum field theories and are not needed in two dimensions. (A 4d Weyl fermion reduces to a 2d Dirac fermion and so the theories we construct are not chiral in 2d.) This means that there is nothing to stop us considering $U(1)$ coupled to N chiral multiplets of charge $+1$ in

$d = 1 + 1$ dimensions. The D-term condition is

$$\sum_{i=1}^{N} |\phi_i|^2 = \zeta$$

where we turn on a FI parameter $\zeta > 0$. Taken on its own, this condition defines a sphere $\mathbf{S}^{2N-1}$. But we still have to quotient by the $U(1)$ action to get the vacuum moduli space and this gives

$$\mathcal{M} = \mathbf{S}^{2N-1}/U(1) = \mathbb{CP}^{N-1}$$

Here $\mathbb{CP}^{N-1}$ is complex projective space, defined as the space of complex lines in $\mathbb{C}^N$. This can also be seen in the definition (4.24) of the moduli space.

Things get more interesting if we add, in addition, a chiral superfield P with charge $-q$. The D-term condition is now

$$D = \sum_{i=1}^{N} |\phi_i|^2 - q|p|^2 - \zeta = 0$$

After quotienting by the $U(1)$ action, the vacuum moduli space is a non-compact manifold. But we now have the option of introducing a gauge invariant superpotential

$$W(P, \Phi) = PG(\Phi_1, \ldots, \Phi_N)$$

with G a homogeneous polynomial of degree q. The potential energy now also includes contributions from the F-terms

$$V_F = |p|^2 \sum_{i=1}^{N} \left| \frac{\partial G}{\partial \phi_i} \right|^2 + |G|^2$$

If we choose G to be *transverse*, meaning

$$\frac{\partial G}{\partial \phi_i} = 0 \quad \forall i \quad \Leftrightarrow \quad \phi_i = 0$$

then $V_F = 0$ only if $p = 0$ which means that we're back onto the $\mathbb{CP}^{N-1}$ vacuum manifold. But now, in addition, we must satisfy $G(\phi) = 0$. The resulting vacuum moduli space is now a compact manifold given by a degree q hypersurface, $\mathcal{M} \subset \mathbb{CP}^{N-1}$.

To give a sense of why the gauge theory description is useful in understanding the geometric properties of the vacuum manifold, here's a short anecdote. It turns out that the gauge theory flows to a conformal field theory only when $q = N$. (Only then does the FI parameter not run.) In this case, the vacuum moduli space X is a degree N hypersurface $\mathbb{CP}^{N-1}$. But it is known that such spaces defines what mathematicians call a *Calabi-Yau manifold*. One of the key properties of these spaces (conjectured by Calab and proven by Yau) is that they admit a Ricci flat metric. This ties in nicely with the gauge theory expectation because, as we have seen in (4.45), such a Ricci flat metric is necessary for conformal symmetry.

There are many more geometrical properties that can be extracted from a study of gauge theories in 2d dimensions, including mirror symmetry of Calabi-Yau manifolds[3].

4.4 Extended Supersymmetry

We discussed the representations of extended supersymmetry algebras in Section 2.4. For theories with $\mathcal{N} = 2$ supersymmetry (or eight supercharges) there are two different multiplets:

$$\begin{aligned}\mathcal{N} = 2 \text{ vector multiplet} &= \mathcal{N} = 1 \text{ vector multiplet } (A_\mu, \lambda_\alpha, D) \\ &+ \mathcal{N} = 1 \text{ chiral multiplet } (\phi, \chi_\alpha, F)\end{aligned}$$

Here the chiral multiplet necessarily sits in the adjoint representation of the gauge group. There is also the $\mathcal{N} = 2$ matter multiplet

$$\begin{aligned}\mathcal{N} = 2 \text{ hypermultiplet} &= \mathcal{N} = 1 \text{ chiral multiplet } (q, \psi_\alpha, F) \\ &+ \mathcal{N} = 1 \text{ chiral multiplet } (\tilde{q}, \tilde{\phi}_\alpha, \tilde{F})\end{aligned}$$

If the first of these transforms in the representation R of the gauge group then the second transforms in the conjugate representation $\bar{R}$. We can tune the matter content and interactions of $\mathcal{N} = 1$ theories to give theories with extended supersymmetry.

With $\mathcal{N} = 4$ there is just a single multiplet (at least restricting to non-gravitational theories) with content

$$\begin{aligned}\mathcal{N} = 4 \text{ vector multiplet} &= \mathcal{N} = 1 \text{ vector multiplet } (A_\mu, \lambda^1_\alpha, D) \\ &+ 3 \times \mathcal{N} = 1 \text{ chiral multiplets } (\phi^i, \lambda^{i+1}_\alpha, F^i) \qquad i =, 1, 2, 3\end{aligned}$$

In addition to the gauge field, we have three complex scalars and four Weyl fermions, all sitting in the adjoint representation of the gauge group.

[3]The use of gauge theories as a method to understand geometry was pioneered by Edward Witten in the paper Phases of $N = 2$ Theories. You can read more in Kentaro Hori's lecture notes which comprise Part 2 and Part 3 of the book Mirror Symmetry.

To construct theories with $\mathcal{N} = 2$ and $\mathcal{N} = 4$ supersymmetry, we could try to build an extended superspace. It turns out that there is a superspace for $\mathcal{N} = 2$ theories, known as harmonic superspace, but it's rather cumbersome to work with. In contrast, there is no superspace for $\mathcal{N} = 4$ theories. Instead, we will build Lagrangians for both by tuning the interactions of $\mathcal{N} = 1$ theories. The key is to get Lagrangians that exhibit larger R-symmetries.

4.4.1 $\mathcal{N} = 2$ Theories

$\mathcal{N} = 2$ super Yang-Mills comprises of a vector multiplet V and an adjoint chiral multiplet Φ. The $\mathcal{N} = 2$ Lagrangian is constructed by simply turning off any superpotential for Φ. It is

$$\begin{aligned}\mathcal{L} &= -\mathrm{Tr}\left[\int d^2\theta\ \frac{i\tau}{8\pi}W^\alpha W_\alpha + \mathrm{h.c.}\right] + \frac{1}{g^2}\int d^4\theta\ \Phi^\dagger e^{2V}\Phi \\ &= \frac{2}{g^2}\mathrm{Tr}\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - i\lambda\sigma^\mu\mathcal{D}_\mu\bar{\lambda} - i\chi\sigma^\mu\mathcal{D}_\mu\chi + \mathcal{D}_\mu\phi^\dagger\mathcal{D}^\mu\phi\right] + \frac{\vartheta}{16\pi^2}\mathrm{Tr}\,F_{\mu\nu}{}^\star F^{\mu\nu} \\ &\quad + \frac{2}{g^2}\,\mathrm{Tr}\Big[\sqrt{2}i\lambda[\phi^\dagger,\chi] + \sqrt{2}i\bar{\lambda}[\phi,\bar{\chi}] - \frac{1}{2}[\phi^\dagger,\phi]^2\Big] \end{aligned} \tag{4.46}$$

The potential term comes from integrating out the D-term from the $\mathcal{N} = 1$ vector multiplet: we'll look more closely at the moduli space of vacua below.

Of more immediate importance are the fermion terms: the two Weyl fermions λ and χ sit on the same footing in the final Lagrangian, despite their origins in different $\mathcal{N} = 1$ multiplets. This means that there is an $SU(2)$ symmetry that rotates them, under which they sit in a doublet $\mathbf{2}$. The bosonic field ϕ does not transform under this symmetry, which tells us that this must be an $SU(2)_R$ R-symmetry. This is the smoking gun for $\mathcal{N} = 2$ supersymmetry. There is also a $U(1)_R$ symmetry, under which $R[\phi] = 2$ and $R[\lambda] = R[\chi] = 1$.

There is another way to derive the $\mathcal{N} = 2$ Lagrangian. You can write down a minimal super Yang-Mills theory in $d = 5 + 1$ dimensions, consisting of a gauge field coupled to a Weyl fermion. Upon dimensional reduction, this gives the Lagrangian (4.46).

We can couple matter to (4.46) in the form of hypermultiplets. These comprise of two chiral multiplet, Q and $\tilde{Q}$. (Note: until now the letter Q has always meant a supercharge, but it's not unusual to also use it to denote a chiral multiplet, with Q standing for "quark".) As we mentioned above, if Q sits in the representation R then $\tilde{Q}$ necessarily sits in the conjugate representation $\bar{R}$. This suffices to determine the

interaction with the vector multiplet V,

$$\mathcal{L}_{\rm vector} = \int d^4\theta \ \left[Q^\dagger e^{2V} Q + \tilde{Q}^\dagger e^{-2V} \tilde{Q} \right]$$

But in addition we should couple Q and $\tilde{Q}$ to the $\mathcal{N} = 2$ vector multiplet field Φ in such a way that the $SU(2)_R$ symmetry between λ and χ remains. This is achieved by the superpotential term

$$\mathcal{L}_{\rm chiral} = \sqrt{2} \int d^2\theta \ \tilde{Q} \Phi Q + \text{h.c.}$$

The interactions between $\tilde{Q}$ and Q themselves are greatly limited by the extended supersymmetry: we can add only mass terms

$$W = \sqrt{2} m \tilde{Q} Q$$

A general $\mathcal{N} = 2$ theory is specified by the gauge group G and the representations R_i of any matter multiplets, together with their masses. (If G contains Abelian factors, we can also add FI terms. We will not include these in the following.) The scalar potential comes, as always, from integrating out D and F-terms. After some rearranging, the potential can be expressed as the sum of positive definite terms. For $SU(N_c)$, it is

$$\begin{aligned} V(\phi, q, \tilde{q}) = \frac{1}{g^2} \, \text{Tr}[\phi^\dagger, \phi]^2 + \frac{g^2}{2} \sum_{A=1}^{\dim G} \left(\sum_i q_i^\dagger T_R^A q_i - \tilde{q}_i T_R^A \tilde{q}_i^\dagger \right)^2 + g^2 \sum_{A=1}^{\dim G} \left| \sum_i \tilde{q}_i T_R^A q_i \right|^2 \\ + \sum_i q_i^\dagger \{ \phi^\dagger - m_i^\dagger, \phi - m_i \} q_i + \tilde{q}_i \{ \phi^\dagger - m_i^\dagger, \phi - m_i \} \tilde{q}_i^\dagger \end{aligned} \tag{4.47}$$

(Initially, the D-term contains both ϕ and the q's and $\tilde{q}$'s. The first two terms on the first line both arise from this D-term, but the cross-term has sneaked into the third line, where it turns $\phi^\dagger \phi$ into the anti-commutator $\{\phi^\dagger, \phi\}$.)

The hypermultiplet scalars q and $\tilde{q}^\dagger$ transform as a doublet **2** under the $SU(2)_R$ symmetry. Conversely, their fermionic superpartners ψ and $\tilde{\psi}$ are singlets under $SU(2)_R$. The second and third terms in the potential (4.47) can be rewritten in way that makes the $SU(2)_R$ symmetry manifest. We introduce the doublet

$$\omega_i = \begin{pmatrix} q_i \\ \tilde{q}_i^\dagger \end{pmatrix}$$

The second term in (4.47) is a real D-term while the third is a complex F-term. But, with $\mathcal{N} = 2$ supersymmetry they are better viewed as a potential $V = \frac{1}{g^2}\vec{D}^2$ arising from triplet of D-terms

$$\vec{D}^A = g^2 \sum_i \omega_i^\dagger T_R^A \vec{\sigma} \omega_i$$

where $\vec{\sigma}$ are the Pauli matrices. The triplet $\vec{D}$ transforms in the **3** of $SU(2)_R$.

The potential (4.47) has some interesting properties. Let's take the masses to vanish: $m_i = 0$. In this case, the second line takes the schematic form $|\phi|^2(|q|^2 + |\tilde{q}|^2)$. That means that if we're looking for vacuum states with $V(\phi, q, \tilde{q}) = 0$ then there are two possibilities: either $\phi = 0$ and the hypermultiplet scalars $q, \tilde{q}$ are turned on; or $\tilde{q} = q = 0$ and the vector multiplet scalar ϕ is turned on. Geometrically, this means that the vacuum moduli space factorises as

$$\mathcal{M} = \mathcal{M}_C \times \mathcal{M}_H$$

There are defined as follows:

- $\mathcal{M}_C$ is called the *Coulomb branch*. It is defined as the space $\tilde{q} = q = 0$ with ϕ restricted to obey

  $$[\phi^\dagger, \phi] = 0$$

  This is solved by ϕ sitting in the Cartan sub-algebra. For $G = SU(N_c)$. this means that $\phi = \text{diag}(\phi_1, \ldots, \phi_{N_c})$ with $\sum_a \phi_a = 0$. At a typical point, the gauge group is broken to the Cartan subalgebra with a bunch of surviving, massless photons. For example, for $G = SU(N_c)$, this means $G \to U(1)^{N_c - 1}$. At some special points, the surviving gauge group will be enhanced further.

  When the gauge group is broken to $U(1)$'s, all charged matter experiences a Coulomb force, hence the name of this branch of vacua.

- $\mathcal{M}_H$ is called the *Higgs branch*. It is defined as the space $\phi = 0$ with $\tilde{q}$ and q constrained to obey the conditions

  $$\vec{D}^A = 0$$

  In addition, we should quotient by the action of G. At a general point, the gauge group is completely Higgsed, hence the name of this branch of vacua.

The Higgs branch has real dimension that is a multiple of four and is a special case of a Kähler manifold, known as a *hyperKähler manifold*. (For what it's worth, a hyperKähler manifold has three independent complex structures while a Kähler manifold has just one.) The definition of the Higgs branch is an extension of the idea of symplectic reduction that gives a hyperKähler metric and is known as the *hyperKähler quotient construction*.

4.4.2 $\mathcal{N} = 4$ Theories

The more supersymmetry we have, the more restrictive the theory.

With $\mathcal{N} = 1$ supersymmetry, we are free to specify the gauge group and (chiral) matter content. In addition to the gauge coupling and masses, both suitably complexified, we can also introduce any superpotential interactions that we wish.

With $\mathcal{N} = 2$ supersymmetry, we are again free to specify the gauge group and (now non-chiral) matter content. But we have no freedom in the choice of interactions: the only arbitrary parameters are the gauge coupling and masses.

With $\mathcal{N} = 4$ supersymmetry, we get to specify only the gauge group and gauge coupling. All other terms in the Lagrangian are then dictated by supersymmetry.

There are a number of different ways to construct $\mathcal{N} = 4$ super Yang-Mills. It can be viewed as minimal super Yang-Mills in $d = 9 + 1$ dimensions, dimensionally reduced to $d = 3 + 1$. Alternatively, it can be viewed as an $\mathcal{N} = 2$ theory with a single adjoint hypermultiplet. The theory contains four adjoint Weyl fermions, transforming in the **4** of $SU(4)_R$ R-symmetry and six real scalars φ^i with $i = 1, \ldots, 6$, transforming in the **6**. The scalar potential is

$$V(\varphi) = -g^2 \sum_{i<j} [\varphi^i, \varphi^j]^2$$

There is now just a Coulomb branch, with G broken to the Cartan subalgebra at a generic point.

5 Boot Camp: Quantum Gauge Dynamics

Our ultimate aim in these lectures is to understand the quantum dynamics of supersymmetric gauge theories. But before we can appreciate this, we really need to understand something about the quantum dynamics of ordinary gauge theories. The purpose of this section is to provide the necessary background.

I should warn you that, in contrast to the rest of these lecture notes, we won't attempt to prove any of the statements made in this section. Indeed, some of them – like the phenomenon of confinement – can't currently be proven, although we do have overwhelming evidence that it takes place, both from numerics and from toy models, not least supersymmetric theories. (Not to mention experimental results like the fact that you are literally stuck together by confinement.) Other phenomena – like the one-loop beta function and the anomaly – have some technical calculations underlying them. Here we omit the technicalities and just state the relevant facts, meaning that you can relax and enjoy this section as something akin to the middle eight in a song. If you want to see the gory details that underlie these results then they can all be found in the lectures on Gauge Theory.

5.1 Strong Coupling

Our interest throughout this section will be on non-Abelian gauge theories. We start with Yang-Mills. The Lagrangian is

$$\mathcal{L}_{YM} = \int d^4x \ -\frac{1}{2g^2}\,\mathrm{Tr}\, F_{\mu\nu}F^{\mu\nu} \tag{5.1}$$

Here the field strength is given by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]$. As you can see, we work with the convention in which the coupling constant sits in front of the kinetic term.

5.1.1 The Beta Function

The key feature of Yang-Mills which makes it both subtle and hard is that the coupling g^2 runs under RG. At a scale μ the coupling is given by

$$\frac{1}{g^2(\mu)} = \frac{1}{g_0^2} - \frac{b_0}{(4\pi)^2}\log\frac{\Lambda_{UV}^2}{\mu^2} \tag{5.2}$$

where g_0^2 is the coupling constant evaluated at the cut-off scale Λ_{UV}. Here b_0 is the coefficient of the 1-loop beta function and, for pure Yang-Mills, is given by

$$b_0 = \frac{11}{6}I(\mathrm{adj})$$

G	$SU(N)$	$Sp(N)$	$SO(N)$	E_6	E_7	E_8	F_4	G_2
$I(\text{adj})$	$2N$	$2(N+1)$	$N-2$	4	3	1	3	4

Table 1. The quadratic Casimir $I(\text{adj})$ for all compact Lie groups.

This depends on a group theoretic factor $I(\text{adj})$, known as the *quadratic Casimir*. It has another avatar as the Dynkin index in the adjoint representation. (Note that we've defined $I(R)$ with a factor of 2 difference from the Gauge Theory lecture notes.) The quadratic Casimirs for the various compact Lie groups are shown in Table 1. In these lectures, we will focus almost exclusively on gauge group $G = SU(N)$.

The running of the coupling constant is often summarised in terms of the one-loop beta function

$$\beta(g) \equiv \mu \frac{dg}{d\mu} = -\frac{b_0}{(4\pi)^2} g^3 \tag{5.3}$$

whose solution gives the logarithmic behaviour (5.2).

The all-important feature of the beta function is the overall minus sign. This means that the theory is weakly coupled at high energies, a phenomenon known as *asymptotic freedom*. Conversely, it means that the theory is strongly coupled at low energies. It is this low-energy physics that we would like to understand.

What do we mean by low and high energy here? Where's the dividing line? The answer to this can be found within the formula (5.2). This is because we can construct a *strong coupling scale*

$$\Lambda = \mu \exp\left(-\frac{8\pi^2}{b_0 g^2(\mu)}\right) \tag{5.4}$$

This has the property that $d\Lambda/d\mu = 0$. In other words, it is an RG invariant. This is the scale at which the Yang-Mills theory becomes strong.

There's already something remarkable about the existence of the scale Λ. Classically, the Yang-Mills theory (5.1) has no dimensionful parameter. That means that there is nothing to set a scale. Instead, there is just a dimensionless coupling constant g^2. But the logarithmic running succeeds in turning this into a dimensionful parameter Λ! One way to see this is to note that to define the quantum theory, we necessarily had a

dimensionful parameter lurking all along. This is the UV cut-off of the theory, Λ_{UV}. The strong coupling scale (5.4) is related to the UV cut-off by

$$\Lambda = \Lambda_{UV}\, e^{-8\pi^2/b_0 g_0^2}$$

This means that if the bare coupling is small, $g_0 \ll 1$, as it should be then the physical scale Λ is exponentially suppressed relative to the UV cut-off: $\Lambda \ll \Lambda_{UV}$.

5.1.2 Confinement and the Mass Gap

When the coupling is small, quantum field theories look similar to their classical counterparts. For example, classical Maxwell theory provides a decent guide to what you might expect from QED. In contrast, when the coupling is large, all bets are off. The quantum theory and classical theory may be completely different. Yang-Mills provides the archetypal example.

If you solve the classical Yang-Mills equations, you will find waves that propagate at the speed of light. This suggests that the quantum theory will give rise to a massless particle called a gluon, similar to the photon. Indeed, if you stare at the action there is no A_μ^2 term that might suggest a mass.

Nonetheless, we now know that quantum Yang-Mills contains no massless particles. We say that the theory is *gapped* which means that the first excited state has a finite energy above the ground state. This additional energy is, of course, just $E = mc^2$ where m is the mass of the lightest particle in the theory. The gap is of order the strong coupling scale, $m \sim \Lambda$.

We don't currently have the technology to prove the Yang-Mills mass gap. Indeed, it is generally considered one of the most important and challenging open problems in mathematical physics. We do, however, have very compelling numerical evidence that this occurs, together with some intuition built from various toy models and heuristic explanations for why it occurs. You can read about some of these in the lectures on Gauge Theory. We'll meet others later in these lectures.

In our world, the strong force is governed by an $SU(3)$ gauge theory known as QCD. The associated strong coupling scale is $\Lambda \approx 300$ MeV and is usually referred to as $\Lambda_{\rm QCD}$. No massless gluons are seen in Nature, but there is good evidence for states known as glueballs with masses around the scale Λ.

The existence of a mass gap goes hand in hand with another phenomenon: this is *confinement*. To explain this, consider placing two charged test particles in the Yang-Mills field. To be specific, we'll consider $G = SU(N)$ and take a quark in the fundamental representation $\mathbf{N}$ and an anti-quark in $\bar{\mathbf{N}}$. We simply ask: what force do they feel?

It's best to compute the potential energy between the two particles. You can first do this in the classical theory. There's a little bit of group theoretic fiddliness but the final result is very intuitive: the potential energy scales with the separation r between particles as

$$V(r) \sim \frac{g^2}{r} \tag{5.5}$$

This, of course, is the same scaling that we see in the Coulomb force of electromagnetism.

What about the quantum theory? If the separation between particles is small, meaning $r \ll 1/\Lambda$, you don't notice much difference. At these short distances the theory is weakly coupled and we again see the Coulomb-like potential (5.5) between test particles. We should replace the coupling constant in (5.5) with $g^2(\mu) = g^2(1/r)$ so it's more accurate to say that the potential scales as $V(r) \sim \log r/r$ but this is a mild correction to the physics.

In contrast, at large separation things are radically different. For distances $r \gg 1/\Lambda$, the potential between test particles takes the form

$$V(r) \sim \sigma r \tag{5.6}$$

The coefficient σ necessarily has dimension $[\sigma] = 2$ and this scale, like everything else in Yang-Mills, is set by $\sigma \sim \Lambda^2$. For reasons that we will explain shortly, σ is called the *string tension*. The force law (5.6) is, to put it mildly, a dramatic departure from what we're used to. The potential energy now *increases* with separation. Indeed, it costs an infinite amount of energy to pull the quark anti-quark pair to infinity. This kind of potential energy is said to be *confining*.

The phenomenon of confinement is, like the mass gap, something that we can't prove from first principles. Once again, however, there is clear numerical evidence together with a plethora of heuristic explanations.

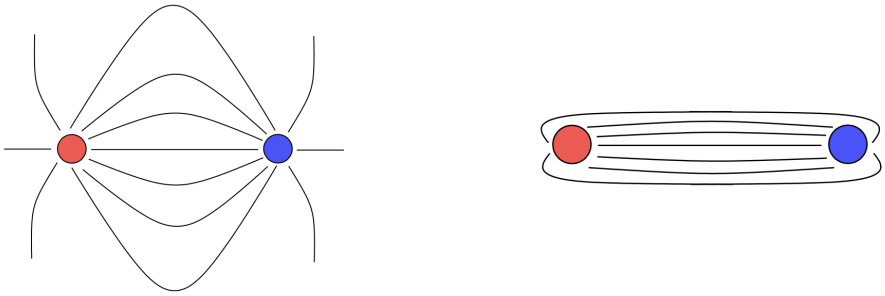


Figure 6. A rough sketch of the non-Abelian field lines in the Coulomb phase, on the left, and in the confining phase, on the right.

To get some very rough intuition for what's going on, we can repeat Faraday's old experiment (now in thought only!) and try to understand what the field lines look like. At short separation, in the Coulomb-like phase (5.5), the field lines form the familiar pattern, first spreading out radially before they bend over to combine with those emitted by the anti-particle. This is shown on the left-hand side of Figure 6. However, as the particles are separated to larger distances, the fact that the gauge field is massive makes itself known. The field lines no longer spread out, but instead lie closely together to form a collimated flux tube. This flux tube acts very much like a string, connecting the two quarks. If its tension, or energy per unit length, is σ then it gives rise to a confining force law like (5.6).

The above description of confinement should be taken with something of a pinch of salt. After all, we are in a strongly interacting quantum field theory and there is no single field configuration that governs the physics. Instead, there are many fields configurations that we should sum over that contribute to the path integral. The discussion above should be understood to mean that those field configurations that resemble the flux tube dominate.

The story above was told in terms of test particles. When we introduce dynamical matter fields into the theory, one would naively expect the associated particles to bind together like the test particles above. And, roughly speaking, this is indeed what happens, at least if the number of light species is small enough. (We'll flesh out this statement shortly.) For example, in QCD the quarks bind together into mesons and baryons. Mesons contain a quark anti-quark pair while baryons contain three quarks and are a colour singlet by dint of the ϵ^{abc} invariant tensor. For $G = SU(N)$ we would get mesons which again contain a quark anti-quark pair and baryons containing N quarks.

There is much more to say about confinement. In particular, the correct, mathematical description of the confining phase lies involves a non-local operator known as the Wilson loop

$$W[C] = \mathrm{Tr}\,\mathcal{P}\exp\left(i\oint_C A\right)$$

Here C is a closed curve in spacetime, while $\mathcal{P}$ stands for "path ordering". In a Coulomb-like phase, the expectation value scales as $\langle W[C]\rangle \sim \exp(-L[C])$ where $L[C]$ is the length of the perimeter of C. Meanwhile, in the confining phase the expectation value scales as $\langle W[C]\rangle \sim \exp(-A[C])$ where $A[C]$ is the area spanned by the curve C. An explanation of why this is the right diagnostic, together with its significance, can be found in the lectures on Gauge Theory.

5.1.3 Adding Matter

Until now, we've considered pure Yang-Mills and its response to test particles. Now we wish to add dynamical matter. The first thing that this does is change the beta function.

Suppose that we have a bunch of Weyl fermions transforming in some representations R_f and a bunch of scalars transforming in some representation R_s. Then the one-loop beta function (5.2) becomes

$$b_0 = \frac{11}{6}I(\mathrm{adj}) - \frac{2}{6}\sum_{\mathrm{fermions}} I(R_f) - \frac{1}{6}\sum_{\mathrm{scalars}} I(R_s) \tag{5.7}$$

Here the group theoretical factors are Dynkin indices. For the representation R, the Dynkin index $I(R)$ is defined by the normalisation of the trace

$$\mathrm{Tr}\; T_R^A\, T_R^B = \frac{1}{2}I(R)\,\delta^{AB} \tag{5.8}$$

Our previous normalisation (4.17) means that we're taking the fundamental representation to have $I(\mathrm{fund}) = 1$. Some examples of $I(R)$ for $SU(N)$ representations are collected in Table 2.

Strictly speaking, the beta function takes the form (5.7) only if the matter is massless. If the matter has some mass m, then the beta function runs like (5.7) for energies $\mu > m$, but as we drop below the mass scale m the matter decouples and its contribution to the one-loop beta function is removed.

Irrep	$\square$	adj	$\square\square$	$\begin{smallmatrix}\square\\\square\end{smallmatrix}$
dim	N	N^2-1	$\frac{1}{2}N(N+1)$	$\frac{1}{2}N(N-1)$
$I(R)$	1	$2N$	$N+2$	$N-2$
$A(R)$	1	0	$N+4$	$N-4$

Table 2. Some group theoretic properties of $SU(N)$ representations. Here $\square\square$ is the symmetric representation and $\begin{smallmatrix}\square\\\square\end{smallmatrix}$ the anti-symmetric. Conjugate representations have $I(\bar{R}) = I(R)$ and $A(\bar{R}) = -A(R)$.

Again, the first thing to notice is the signs. Both fermions and scalars give a contribution to the beta function that has the opposite sign to the gauge bosons. This means that if we have too much matter then we will have $b_0 < 0$ and, correspondingly, $\beta(g) > 0$ and the theory will be weakly coupled in the infra-red. In this case, the quantum theory looks very much like classical Yang-Mills at low energies, with massless gauge bosons. Here we would like to understand what happens when $b_0 > 0$ and the theory is strongly coupled.

To illustrate this, we will consider a specific set of matter particles. We take

$$G = SU(N_c)$$

with N_f flavours of quarks in the fundamental representation. This means that we have a collection of left-handed Weyl spinors $\psi^a_{\alpha i}$ and $\tilde{\psi}^i_{\alpha a}$. Here $a = 1, \ldots, N_c$ is the gauge index and $i = 1, \ldots, N_f$ the flavour index. We take ψ to transform in the fundamental $\mathbf{N}_c$ representation and $\tilde{\psi}$ in anti-fundamental representation $\bar{\mathbf{N}}_c$ representation. (If we take the complex conjugate of $\tilde{\psi}$, we get a Dirac spinor in the $\mathbf{N}_c$ representation.) The action is

$$\mathcal{L}_{QCD} = -\frac{1}{2g^2}\,\mathrm{Tr}\, F_{\mu\nu}F^{\mu\nu} - \sum_{i=1}^{N_f}\left[i\bar{\psi}_i\bar{\sigma}^\mu\mathcal{D}_\mu\psi_i + i\bar{\tilde{\psi}}^i\bar{\sigma}^\mu\mathcal{D}_\mu\tilde{\psi}^i\right] \tag{5.9}$$

with

$$\mathcal{D}_\mu\psi = \partial_\mu\psi - iA_\mu\psi \quad \text{and} \quad \mathcal{D}_\mu\tilde{\psi} = \partial_\mu\tilde{\psi} + i\tilde{\psi}A_\mu$$

We could add a mass for the quarks by introducing terms like

$$\mathcal{L}_{\text{mass}} = \sum_{i=1}^{N_f} m_i\tilde{\psi}^i\psi_i + \text{h.c.}$$

However, our interest will be on the case with massless quarks, with $m_i = 0$.

You might wonder why this is interesting. After all, the quarks in our world aren't massless. But they are *almost* massless! The up and down quarks have masses of a few MeV, much less than the relevant scale $\Lambda_{\rm QCD} \approx 300$ MeV. Meanwhile, the strange quark has a mass $m_{\rm strange} \approx 95$ MeV, still smaller than $\Lambda_{\rm QCD}$ although not by much. This means that understanding the behaviour of massless QCD is not a bad starting point for understanding the full theory.

5.1.4 Chiral Symmetry Breaking

The important observation is that massless QCD (5.9) has an extra symmetry that the massive theory doesn't have, under which the ψ and $\tilde{\psi}$ fermions rotate independently. The global symmetry includes

$$G_F = SU(N_f)_L \times SU(N_f)_R \tag{5.10}$$

Here $SU(N_f)_L$ acts on the ψ while $SU(N_f)_R$ acts on the $\tilde{\psi}$,

$$\psi_i \to (L^\star)_i{}^j \psi_j \quad \text{and} \quad \tilde{\psi}^i \to R^i{}_j \tilde{\psi}^j \tag{5.11}$$

with $L \in SU(N_f)_L$ and $R \in SU(N_f)_R$. (In fact, the full symmetry of the classical theory is $U(N_f)_L \times U(N_f)_R$; we'll discuss these additional $U(1)$ factors in Section 5.2.) The group G_F is known as the *chiral symmetry*, chiral because it acts on Weyl spinors rather than Dirac spinors. This kind of symmetry only exists when the masses $m_i = 0$.

The question that we want to ask is: what becomes of this chiral symmetry? The answer to this depends on the number of flavours N_f in a way that is not fully understand. However, for suitably small N_f the theory develops a vacuum expectation value

$$\langle \tilde{\psi}^i \psi_j \rangle \sim \Lambda^3 \delta^i{}_j$$

The formation of this condensate is a strong coupling effect and, like confinement, poorly understood. In contrast, the consequence of the condensate is both well understood and dramatic. First, note that the condensate does not preserve the chiral symmetry (5.11). Indeed, it transforms as

$$\langle \tilde{\psi}^i \psi_j \rangle \to \Lambda^3 \, R^i{}_k (L^\dagger)^k{}_j \tag{5.12}$$

This is the phenomenon of *chiral symmetry breaking*, sometimes shortened to χSB. The surviving subgroup requires us to set $L = R$ in (5.11), meaning

$$SU(N_f)_L \times SU(N_f)_R \to SU(N_f)_{\rm diag}$$

The spontaneous breaking of chiral symmetry means that massless QCD actually has a moduli space of vacua, since each choice of $L \neq R$ in (5.12) gives a different, equally valid, ground state, albeit one that is entirely equivalent to the original because they are related by a global symmetry. The vacuum moduli space is the coset

$$\mathcal{M} = [SU(N_f)_L \times SU(N_f)_R] / SU(N_f)_{\text{diag}} \tag{5.13}$$

with dimension

$$\dim \mathcal{M} = N_f^2 - 1$$

There is an important difference between this vacuum moduli space and those that arise in supersymmetric theories. All points on $\mathcal{M}$ in QCD are equivalent because any point is related to any other by the action of a symmetry. This is not the case for the supersymmetric moduli space.

Nonetheless, there is one important feature that is common whenever we have flat directions and this is the importance of massless particles, corresponding to fluctuations along $\mathcal{M}$. When the flat directions arise from broken symmetries, as in the present case, these massless particles are *Goldstone bosons*.

We learn something interesting. Yang-Mills theory has a mass gap. But massless QCD, at least for $N_f > 1$, does not. Even if the theory confines, giving massive baryons and glueballs, chiral symmetry breaking means that there are massless Goldstone bosons. These can be identified with certain meson states called *pions*.

Of course, in our world the pions are not massless. But this is because the constituent quarks are not exactly massless so the chiral symmetry is not exact. Nonetheless, the chiral symmetry is an approximate symmetry which, in turn, means that the would-be Goldstone bosons are light, but not exactly massless. Indeed, the pions are notably lighter than all other hadrons in QCD.

5.1.5 Phases of Massless QCD

We're now in a position to describe the different phases of massless QCD as we vary N_c and N_f. There is much that we don't yet understand (here "we" means everyone, not just those following these lectures!) and there are a few subtleties that I will sweep under the carpet. But, with broad brush, we can sketch the different phases of the theory.

We start with low N_f:

- When $N_f = 0$, we have pure Yang-Mills. The theory sits in the confining phase, with a mass gap.
- When $N_f = 1$, there is no chiral symmetry group (5.10) and so no chiral symmetry breaking. The theory is again thought to have a mass gap, with quarks bound in mesons and baryons.
- When $2 \leq N_f \leq N^\star$ the theory confines and exhibits chiral symmetry breaking. This means that the low energy theory consists of freely interacting Goldstone bosons, parameterising the moduli space (5.13).

  The big question here is: what is the maximum value $N^\star$ for which chiral symmetry breaking occurs? We don't know the answer to this. Various approaches, including numerics, suggest that it is somewhere around

  $$N^\star \approx 4N_c$$

  Our lack of knowledge of this simple question highlights just how poorly we understand strongly interacting field theories.

Now let's jump to high values of N_f and we'll then try to fill in the details in the middle.

- When $N_f \geq \frac{11}{2}N_c$, the beta function is positive. You can see this from the general expression (5.7) which, for massless QCD, becomes

  $$b_0 = \frac{11}{3}N_c - \frac{2}{3}N_f \tag{5.14}$$

  This means that theory is weakly coupled in the infra-red: the low-energy physics consists of massless gluons, weakly interacting with massless quarks. As we go to smaller and smaller energies, the interactions become weaker and weaker. Strictly speaking, in the far IR, the physics is free.

  On the flip side, these become arbitrarily strongly coupled in the UV, with the gauge coupling diverging at some very high scale. This doesn't mean that we should discard them, but they don't make sense at arbitrarily high energies scales. Said another way, we can't take the UV cut-off Λ_{UV} to infinity while keeping any low-energy interactions. Nonetheless, it's quite possible that these theories may arise as the low-energy limit of some other theory. We will see examples in Section 6 when we discuss supersymmetric extensions of QCD.

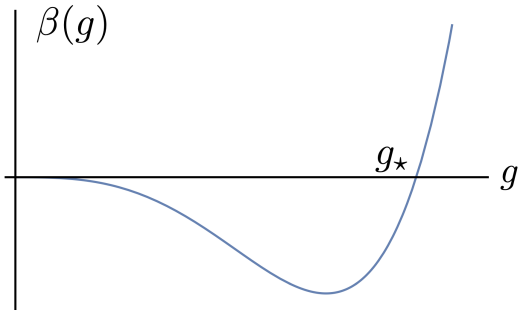


Figure 7. The beta function for N_f slightly below the asymptotic freedom bound has a zero which indicates the existence of an interacting conformal field theory.

That leaves us with the physics in the middle region. We'll keep working down from the asymptotic freedom bound $11N_c/2$.

- When $N^{\star\star} < N_f < \frac{11}{2}N_c$, things are more interesting. To see what happens, we need the two-loop beta function

$$\beta(g) = -\frac{b_0}{(4\pi)^2}g^3 - \frac{b_1}{(4\pi)^4}g^5 + \ldots$$

with the one-loop coefficient b_0 given in (5.14) and the two-loop coefficient

$$b_1 = \frac{34N_c^2}{3} - \frac{N_f(N_c^2-1)}{N_c} - \frac{10N_fN_c}{3}$$

In the window of interest, $b_0 > 0$ and $b_1 < 0$, so we can play the one-loop contribution against the two-loop contribution to find a zero of the beta function

$$g_\star^2 = -(4\pi)^2\frac{b_0}{b_1}$$

with $\beta(g_\star) = 0$. The beta function is shown in Figure 7. The existence of such a fixed point is telling us that we have an interacting conformal field theory: there are massless modes, but they are no longer free in the infra-red. This is known as the *Banks-Zaks fixed point*.

Importantly, when N_f lies just below the asymptotic freedom bound, so $\frac{N_f}{N_c} = \frac{11}{2} - \epsilon$, this fixed point lies at $g_\star \ll 1$ which means that we can trust the analysis without having to worry about higher order corrections. Moreover, because $g_\star$ is small we can use perturbation theory to calculate anything that we want.

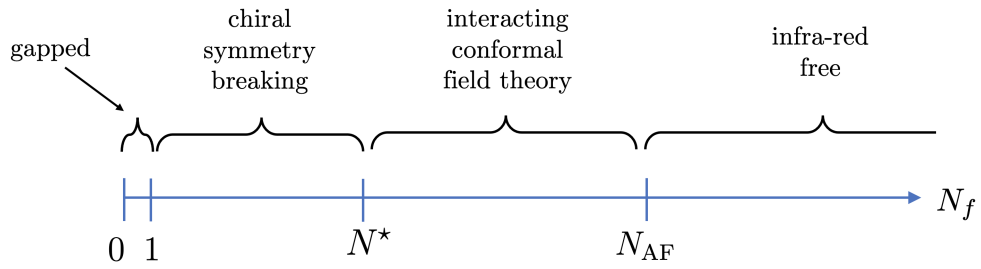


Figure 8. The expected phases of massless QCD. The asymptotic freedom bound is $N_f = \frac{11}{2}N_c$. The lower edge of the conformal window is not known but is expected to be somewhere around $N_f \approx 4N_c$.

However, as N_f decreases, the value of the fixed point $g_\star$ increases until we can no longer trust the analysis above. The expectation is that we get a conformal field theory only for some range of N_f, lying within $N^{\star\star} < N_f < \frac{11}{2}N_c$. This is known as the *conformal window*. We don't currently know the value of $N^{\star\star}$.

That leaves us with understanding what happens in the middle when $N^\star < N_f \leq N^{\star\star}$. Our best guess is that there is no such regime, and the upper edge of the chiral symmetry breaking phase coincides with the lower edge of the conformal window,

$$N^{\star\star} = N^\star$$

This guess is motivated partly by numerics and partly by a lack of any compelling alternative. For us, the lesson to take away is that strongly interacting quantum field theories are hard and even the most basic questions are beyond our current abilities. A summary of the expected behaviour of massless QCD is shown in Figure 8.

5.2 Anomalies

The next topic that we need to cover is *anomalies*. This is a beautiful subject and, in many ways, the place in which quantum field theory intersects most cleanly with topics in mathematics. Here we won't describe any of these mathematical underpinnings, but instead just cover the minimum material necessary for our later applications.

The main idea is to understand how certain symmetries manifest themselves in quantum field theory. To this, end consider a single left-handed Weyl fermion in $d = 3 + 1$ dimensions. The action is

$$S = \int d^4x \; i\bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi$$

This action is clearly invariant under the $U(1)$ global symmetry $\psi \to e^{i\alpha\psi}$, with the corresponding current $j^\mu = \bar{\psi}\sigma^\mu\psi$. To illustrate the anomaly, we will couple this current to a gauge field A_μ with charge $q \in \mathbb{Z}$. The action is now

$$S = \int d^4x \; i\bar{\psi}\bar{\sigma}^\mu \mathcal{D}_\mu \psi$$

where the covariant derivative contains the new coupling $\mathcal{D}_\mu\psi = \partial_\mu\psi - iqA_\mu\psi$. This is action is now invariant under the gauge symmetry

$$\psi \to e^{iq\alpha(x)}\psi \quad \text{and} \quad A_\mu \to A_\mu + \partial_\mu\alpha \tag{5.15}$$

Before we proceed, I should mention that there are two distinct ways to think about the gauge field A_μ and this distinction will be important when we come to look at the various implications of anomalies. They are:

- A_μ could be a *dynamical gauge field*. In the classical theory, this means that we treat it as a dynamical variable, with its own equation of motion, typically after adding a Maxwell term to the action. In the quantum theory, it means that we integrate over A_μ in the path integral.

- A_μ could be a *background gauge field*. This means that it is something fixed, under our control, and should be viewed as a parameter of the theory. Turning it on typically breaks Lorentz symmetry, but could be useful to explore how our system responds to the presence of an electric or magnetic field. In the quantum theory, A_μ appears as a source on which the partition function depends.

We will consider gauge fields of both types in what follows. However, for now, we will consider A_μ to be a background gauge field, something that is under our control.

While the classical theory is clearly invariant under the gauge transformation (5.15), the question that we really want to ask is: what about the quantum theory? For this, we should turn to the path integral, with the partition function in Euclidean space defined as

$$Z[A] = \int D\psi D\bar{\psi} \; \exp\left(-\int d^4x \; i\bar{\psi}\bar{\sigma}^\mu\mathcal{D}_\mu\psi\right)$$

Clearly the action in the exponent remains invariant under gauge transformations. But now we must also worry about the measure in the path integral, and this takes some care to define. The statement of the anomaly is that the measure is *not* invariant under

gauge transformations. Instead, it turns out that the measure, and hence the partition function, changes by a phase

$$Z[A] \to \exp\left(\frac{iq^3}{32\pi^2}\int d^4x\ \alpha F_{\mu\nu}{}^\star F^{\mu\nu}\right)\ Z[A] \tag{5.16}$$

with ${}^\star F^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$.

This subtlety only happens for fermions. If we have scalar fields charged under a symmetry, then the measure is perfectly invariant. At heart, this is related to the fact that there is no difficulty in giving masses to scalar fields while preserving symmetries, but giving masses for fermions necessarily breaks certain symmetries.

The purpose of this section is to understand the implication of this calculation and a number of variants. As we now explain, there are three different avatars of the anomaly. We deal with them each in turn.

5.2.1 Gauge Anomalies

The first implication of the anomaly (5.16) is that it is an obstruction to gauging. Although the action is invariant under the gauge symmetry, the measure is not and neither is the partition function. That means that we cannot promote the gauge field A_μ to a dynamical field, where we integrate over it in the path integral. If we attempted to do this, we would get a sick theory.

There are a number of ways to see why the theory is sick but here is a simple one. Recall that when we first attempted to quantise the gauge field A_μ in the lectures on Quantum Field Theory we had some work to do to decouple the negative norm states that arise from quantising A_0. That work ultimately boiled down to using the gauge invariance to remove these states. But in an anomalous theory, we no longer have that gauge invariance at our disposal and the Hilbert space will involve negative norm states. That's bad.

The upshot is that a $U(1)$ gauge theory, coupled to a single Weyl fermion, is not consistent. To proceed, we must have multiple, left-handed Weyl fermions ψ_i, each with some charge q_i. (If we have right-handed fermions, simply conjugate them to make them left-handed.) The phase in (5.16) is then proportional to the sum of q_i^3. The gauge theory is consistent only if

$$\sum_i q_i^3 = 0 \tag{5.17}$$

This was one of the conditions that we met previously in (4.14). This condition is sometimes written in a different way. One, very simple way to solve this constraint is to take pairs of Weyl fermions with charges $\pm q$. If we conjugate one of them to become a right-handed Weyl fermion, we then have a single Dirac fermion with charge q. These are called vector-like theories and QED is the most familiar example.

There are, however, more interesting solutions to (5.17) that do involve $\pm$ pairs. These are known as *chiral gauge theories*.

The discussion above holds for an Abelian gauge symmetry. There is a similar story for a non-Abelian gauge symmetry G. For a single Weyl fermion, transforming in the representation R of G, the anomaly is proportional to the group theoretic factor $A(R)$. For the fundamental representation, $A(R) = 1$. For other representations, it is given by

$$\mathrm{Tr}\, T_R^A \{T_R^B, T_R^C\} = A(R)\, \mathrm{Tr}\, T^A \{T^B, T^C\}$$

Some examples of $A(R)$ for $SU(N)$ representations are collected in Table 2. To be consistent, a non-Abelian gauge theory coupled to a bunch of left-handed Weyl fermions must obey

$$\sum_i A(R_i) = 0 \tag{5.18}$$

which is the non-Abelian version of (5.17). If R is a complex representation, then it's simple to show that $A(\bar{R}) = -A(R)$. This means that we can again always satisfy (5.18) by taking Dirac fermions, rather than Weyl fermions, since these have a left-handed fermion in a representation R and another in $\bar{R}$.

One consequence of the relation $A(\bar{R}) = -A(R)$ is that $A(R) = 0$ for any real representation. This means that there is no obstacle to coupling a single Weyl fermion in a real representation to a non-Abelian gauge group. Indeed, we've seen this already in these lectures: pure super-Yang-Mills has a single adjoint Weyl fermion, but the adjoint representation is real so there is no problem.

Relatedly, here's a comment that will prove useful shortly: only massless fermions contribute to the anomaly. If you have a Weyl fermion ψ in a complex representation R of a group G, then to give it a mass preserving G you need a second Weyl fermion $\tilde{\psi}$ in representation $\bar{R}$. You can then write down a Dirac mass term $m\tilde{\psi}\psi$. But the two Weyl fermions $\tilde{\psi}$ and ψ cancel in their contribution to the anomaly. Alternatively, you can write down as Majorana mass $m\psi\psi$ for any fermion in a real representation of G

but, as we have seen, there is no contribution to the anomaly from fermions in a real representation. This means that only fermions that cannot get a mass preserving G contribute to the anomaly for G.

When we previously discussed the requirements of anomaly cancellation in (4.14), we gave a further condition on $U(1)$ gauge theories. We asked that they also satisfy

$$\sum_i q_i = 0 \tag{5.19}$$

This, it turns out, is a little more subtle and it follows from the requirement that the theory can be consistently coupled to gravity. There is no corresponding requirement for non-Abelian gauge theories (essentially because $\mathrm{Tr}\, T^A = 0$ for any generator of a simply connected Lie algebra).

The upshot is that if you want to have a theory with a dynamical gauge field, then you better make sure that the anomaly (5.17) or (5.18) cancels. Furthermore, if you want your theory to be compatible with gravity, then you have one further hoop (5.19) to jump through.

5.2.2 Chiral (or ABJ) Anomalies

Here is a slight variant on the same calculation that leads to a physically very different conclusion. Again, consider a single Weyl fermion, now coupled to a background non-Abelian gauge field A in some representation R of the global symmetry G. It's useful to think of $G = SU(N)$, and R either the fundamental or adjoint representation. We can construct the partition function

$$Z[A] = \int D\psi D\bar{\psi}\ \exp\left(-\int d^4x\ i\bar{\psi}\bar{\sigma}^\mu \mathcal{D}_\mu \psi\right)$$

now with $\mathcal{D}_\mu\psi = \partial_\mu\psi - iA^A_\mu T^A_R\psi$. We know that the partition function isn't invariant under gauge transformations of G. But here we instead ask a different question: is it invariant under $U(1)$ rotations of the fermion?

$$\psi \to e^{iq\alpha}\psi \tag{5.20}$$

The answer is again no, with the partition function transforming as

$$Z[A] \to \exp\left(\frac{iqI(R)}{16\pi^2}\int d^4x\ \alpha \mathrm{Tr}\, F_{\mu\nu}{}^\star F^{\mu\nu}\right)\ Z[A] \tag{5.21}$$

with $I(R)$ the Dynkin index defined previously in (5.8). This looks very similar to our previous result, but it should now be thought of a mixed anomaly between the $U(1)$ symmetry (5.20) and the non-Abelian symmetry G. This can be seen in the coefficient $qI(R)$ which is still cubic but now a mix of Abelian and non-Abelian generators.

An interesting consequence of this is that, in the presence of background gauge fields for G, the $U(1)$ symmetry is no longer conserved. If we repeat Noether's theorem, including the anomaly (5.21), we find that the $U(1)$ current associated to the symmetry (5.20) now obeys

$$\partial_\mu j^\mu = \frac{qI(R)}{32\pi^2} \operatorname{Tr} F_{\mu\nu}{}^\star F^{\mu\nu} \tag{5.22}$$

When the right-hand side is non-zero, the current is no longer conserved.

An important example of this occurs in the theory of massless QCD that we introduced in the last section. The gauge group is $G = SU(N_c)$ and the Lagrangian is (5.9),

$$\mathcal{L}_{QCD} = -\frac{1}{2g^2} \operatorname{Tr} F_{\mu\nu} F^{\mu\nu} - \sum_{i=1}^{N_f} \left[i\bar{\psi}_i \bar{\sigma}^\mu \mathcal{D}_\mu \psi_i + i\bar{\tilde{\psi}}^i \bar{\sigma}^\mu \mathcal{D}_\mu \tilde{\psi}^i \right] \tag{5.23}$$

We have added extra fermions to cancel the gauge anomaly in G, as we should. But, as we will see, a mixed anomaly of the type (5.21) remains.

Classically, the theory (5.23) has a $U(N_f)_L \times U(N_f)_R$ global symmetry, with each factor rotating ψ and $\tilde{\psi}$ independently. We studied the $SU(N_f)_L \times SU(N_f)_R$ subgroup in some detail in the previous section, but didn't mention the two $U(1)$ factors. These are usually written as

$$\begin{aligned} U(1)_B: \quad & \psi_i \to e^{i\beta}\psi_i \quad \text{and} \quad \tilde{\psi}^i \to e^{-i\beta}\tilde{\psi}^i \\ U(1)_A: \quad & \psi_i \to e^{i\alpha}\psi_i \quad \text{and} \quad \tilde{\psi}^i \to e^{i\alpha}\tilde{\psi}^i \end{aligned} \tag{5.24}$$

The subscript B stands for "baryon" since this is the vector-like symmetry under which baryons are charged. Since ψ and $\tilde{\psi}$ have opposite charges under $U(1)_B$, there is no obstacle to gauging it should we wish. Moreover, the $\pm$ charges also cancel on the right-hand side of (5.22), and the $U(1)_B$ current is conserved in the quantum theory.

In contrast, the *axial symmetry* $U(1)_A$ has the same charges for ψ and $\tilde{\psi}$. This means that the associated current is, following (5.22), no longer conserved. Instead, it obeys

$$\partial_\mu j^\mu_A = \frac{N_f}{16\pi^2} \operatorname{Tr} F_{\mu\nu}{}^\star F^{\mu\nu} \tag{5.25}$$

Note that the gauge fields on the right-hand side are now dynamical $SU(N_c)$ gauge fields that fluctuate. There is now no way to set them to zero. There is no axial $U(1)_A$ symmetry in the quantum theory.

This also explains why we didn't include $U(1)_A$ when discussing chiral symmetry breaking in the previous section. Since it is not a symmetry, there is no corresponding Goldstone boson. (In the real world, the meson associated to $U(1)_A$ is called the η' and is significantly heavier than the pion Goldstone bosons.)

This, then, is the second avatar of the anomaly. It manifests itself as a symmetry of the classical theory that does not survive the quantisation procedure. In fact, this is how the anomaly was first discovered. In this context, it usually goes by the name of the *chiral anomaly*, or the *ABJ anomaly* after Adler, Bell and Jackiw who first uncovered this subtle effect of quantum field theory. (Yes, *that* Bell.)

There is one further way to think about the chiral anomaly. Non-Abelian gauge theories have an additional, topological term

$$S_\vartheta = \frac{\vartheta}{16\pi^2}\int d^4x \ \mathrm{Tr}\, F_{\mu\nu}{}^\star F^{\mu\nu}$$

This is the *theta term*. We already met it when constructing super Yang-Mills theory in (4.19). Comparing with the form of the mixed anomaly (5.21), we see that axial transformation (5.24) can be thought of as shifting the theta angle

$$U(1)_A : \ \vartheta \to \vartheta + 2\alpha \tag{5.26}$$

We've met this kind of idea previously in Section 3.3, where we found it useful to think of parameters – supurions – transforming under symmetries (which, of course, means that the symmetries aren't actually symmetries). In Section 6, we'll learn how we can combine the shift of the ϑ angle with holomorphy in supersymmetric theories.

5.2.3 't Hooft Anomalies

So far we have discussed two manifestations of the anomaly:

- For a gauge symmetry, the anomaly better cancel. Or else.
- A mixed anomaly between a global symmetry and gauge symmetry means that the global symmetry isn't.

But what if we have an anomaly just for a global symmetry? What are the consequences? From what we've discussed above, we know that the symmetry isn't conserved if we couple it to background gauge fields. But nothing compels us to do so. So what else can we learn from this?

The answer is both subtle and powerful. An anomaly for a purely global symmetry puts strong constraints on the low-energy dynamics of the theory. The anomaly should be thought of as a robust way of characterising the theory, and this characterisation cannot change under RG flow, nor under any other deformation of the theory, providing that the symmetry remains unchanged. Such anomalies in global symmetries are referred to as *'t Hooft anomaly*.

We will first explain the basic idea and then give a concrete example. Suppose that we have some quantum field theory – typically a non-Abelian gauge theory – that is weakly coupled in the UV, but flows to strong coupling in the IR. We will abstractly call the UV theory $\mathcal{T}_{UV}$. We assume that it has some global symmetry G_F. This should be a true symmetry of the quantum theory, meaning that it has no mixed anomalies with the gauge symmetry.

This UV theory may have an anomaly for G_F. If G_F is Abelian, anomaly is simply $\sum q^3$ as in (5.17); if it is non-Abelian the anomaly is $\sum A(R)$ as in (5.18). Either way, we will denote this anomaly as $\mathcal{A}_{UV}$ and assume $\mathcal{A}_{UV} \neq 0$,

The theory now flows under RG to a theory $\mathcal{T}_{IR}$ in the IR which, as we've seen, will typically be very different. We have the following result:

Claim: Either the symmetry G_F is spontaneously broken, or the anomalies match meaning

$$\mathcal{A}_{UV} = \mathcal{A}_{IR}$$

This is a wonderfully powerful result. If G_F is spontaneously broken then we necessarily have massless Goldstone bosons. But if G_F is unbroken then we must have massless fermions that reproduce the anomaly. This is known as *'t Hooft anomaly matching*.

Proof: The argument for 't Hooft anomaly matching is very slick. Suppose that $\mathcal{A}_{UV} \neq 0$ then we know from the discussion above that we're not allowed to couple G_F to dynamical gauge fields. That would lead to a sick theory.

To proceed, we introduce a bunch of extra massless Weyl fermions transforming under G_F. We call these *spectator fermions*. These won't interact directly with our original fields in $\mathcal{T}_{UV}$, but they are designed so that the total anomaly of the original fields and these new fermions vanishes:

$$\mathcal{A}_{UV} + \mathcal{A}_{\text{spectator}} = 0$$

Now there's nothing to stop us introducing dynamical gauge fields for G_F. We do so, but with a very very small coupling constant. We'll see the importance of this shortly.

Now let's go back to our original theory $\mathcal{T}_{UV}$. It will flow to strong coupling at some scale Λ and we'd like to understand the physics $\mathcal{T}_{IR}$ below this scale. If the gauge coupling for G_F is small enough, then this RG flow takes place entirely unaffected by the presence of the G_F gauge fields. This means that one of two things could have happened. It may be that the strong coupling dynamics of $\mathcal{T}_{UV}$ spontaneously breaks the symmetry G_F. (For example, as we've seen, this is expected to happen if we take G_F to be the chiral symmetry of QCD.) This was the first possibility of our claim. Alternatively, G_F may be unbroken at low-energies. In this case, we're left with $\mathcal{T}_{IR}$, together with the spectator fermions, all coupled to the G_F gauge fields. But this can only be consistent if

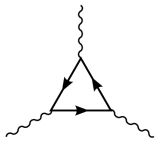
$$\mathcal{A}_{IR} + \mathcal{A}_{\text{spectator}} = 0$$

Clearly, this is only consistent if $\mathcal{A}_{IR} = \mathcal{A}_{UV}$. □

Triangle Diagrams

Until now, we've explained the anomaly as a transformation of the fermion measure in the path integral. However, the anomalies also show up in perturbation theory when computing corrections to Ward identities like (5.25). In this way of looking at things, one has to compute so called *triangle diagrams*. Schematically, these take the form

$$\text{Anomaly} = \sum_{\text{fermions}}$$



where you sum over all Weyl fermions running in loops. The outer legs are currents, either gauge or global. The fact that there are three legs reflects the fact that the anomalies are always proportional to the cube of generators. Our three kinds of anomalies are related to the different types of currents on the legs

- Gauge^3: This is a gauge anomaly.
- Global $\times$ Gauge^2: This is the chiral anomaly.
- Global^3: This is the 't Hooft anomaly.

An Application: Confinement Implies Chiral Symmetry Breaking

We saw in the last section that massless QCD exhibits two, distinct strong coupling phenomena: confinement and chiral symmetry breaking. We will now show that they're not quite as unrelated as they first appear.

As we've seen, the $U(1)_A$ symmetry of massless QCD is anomalous. The true symmetry group is therefore

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

Let's first compute the 't Hooft anomalies in the ultra-violet, where the quarks contribute. There is no 't Hooft anomaly for $U(1)_B^3$ because this is a vector-like symmetry. In contrast, there is a 't Hooft anomaly associated to the chiral, $SU(N_f)$ factors. In fact, there are two. The first is the purely non-Abelian anomaly

$$[SU(N_f)_L]^3 : \quad \mathcal{A} = \sum A(\overline{\square}) = -N_c$$

Here the anomaly $\mathcal{A}$ arises because each quark ψ carries a colour index $a = 1, \ldots, N_c$. The ψ fermions transform in the $\overline{\square}$ of $SU(N_f)_L$ and $A(\overline{\square}) = -1$. But there are N_c such fermions. Hence the result $N_c A(\overline{\square}) = -1$. There is a similar anomaly for $SU(N_f)_R$. In addition, there is a mixed 't Hooft anomaly between $U(1)_B$ and $SU(N_f)$. This is

$$[SU(N_f)_L]^2 \times U(1)_B : \quad \mathcal{A}' = \sum qI(\overline{\square}) = N_c$$

which again simply counts the number of quarks.

Now the question is: what happens in the infra-red? For suitably low N_f, we've already explained the chiral symmetry G_F is expected to be broken down to $U(1)_B \times SU(N_f)_{\text{diag}}$, but we didn't give any justification for this. The idea of 't Hooft anomaly matching goes some way to help.

Here is the idea. We will assume that the theory confines and, moreover, that in the infra-red, the physics is described by weakly interacting mesons and baryons. (This is in contrast to the conformal field theories that we see at larger N_f.) In such a situation, 't Hooft anomaly matching shows that the chiral symmetry *must* be broken.

Here is the argument. Suppose that G_F is unbroken in the infra-red. Then they must be massless fermions around that can reproduce the anomalies $\mathcal{A}$ and $\mathcal{A}'$. Moreover, by assumption, these massless fermions must be bound states of quarks, either mesons or baryons.

Mesons certainly can't do the job because these are bosons. Baryons, meanwhile, contain N_c quarks so these too are bosons when N_c is even. This is telling us that when N_c is even, a confining theory contains no fermions at low-energies and so certainly can't reproduce the anomalies. We learn that chiral symmetry breaking must occur when N_c is even.

What about N_c odd? Now baryons are fermions. Is it possible that some of these baryons could be massless and reproduce the 't Hooft anomalies? This time we have something of a calculation to do. First, you have to figure out what representations of G_F the baryons sit in. Then you have to figure out what combination of massless baryons could match the anomalies $\mathcal{A}$ and $\mathcal{A}'$. It takes some work, but the answer is that the baryons can never reproduce the anomalies. (You can find the calculation in Section 5.6 of the lectures on Gauge Theory.) This means that if QCD confines into weakly interacting colour singlets, then chiral symmetry is necessarily broken.

5.3 Instantons

One of the new ingredients in these lectures is the Yang-Mills theta angle

$$S_\vartheta = \frac{\vartheta}{16\pi^2}\int d^4x\ \mathrm{Tr}\, F_{\mu\nu}{}^\star F^{\mu\nu}$$

This deserves some explanation.

First, the theta term is a total derivative,

$$S_\vartheta = \frac{\theta}{8\pi^2}\int d^4x\ \partial_\mu K^\mu \quad \text{with} \quad K^\mu = \epsilon^{\mu\nu\rho\sigma}\mathrm{Tr}\left(A_\nu\partial_\rho A_\sigma - \frac{2i}{3}A_\nu A_\rho A_\sigma\right)$$

This means that it does not affect the classical equations of motion. Nonetheless, it can affect the quantum dynamics of gauge theories. This arises because the path integral receives contributions from field configurations that have something interesting going on at infinity so that the boundary term S_ϑ is non-vanishing. This something interesting can be found in the topology of the gauge group.

To explain this, we first Wick rotate so that we work in Euclidean spacetime $\mathbb{R}^4$. Configurations that have a finite action from the Yang-Mills term must asymptote to pure gauge,

$$A_\mu \to i\Omega\partial_\mu\Omega^{-1} \quad \text{as } x\to\infty \tag{5.27}$$

with $\Omega \in G$. This means that finite action, Euclidean field configurations involve a map

$$\Omega(x) : \mathbf{S}^3_\infty \ \mapsto G$$

with $\mathbf{S}^3_\infty = \partial\mathbb{R}^4$. Maps of this kind fall into disjoint classes. This arises because the gauge transformations can "wind" around the spatial $\mathbf{S}^3$ in such a way that one gauge transformation cannot be continuously transformed into another. Such winding is characterised by *homotopy theory*. In the present case, the maps are labelled by an element of the homotopy group which is

$$\Pi_3(G) = \mathbb{Z}$$

for all simple, compact Lie groups G. In words, this means that the winding of gauge transformations (5.27) at infinity is classified by an integer n.

This statement is most intuitive for $G = SU(2)$ since $SU(2) \cong \mathbf{S}^3$ and the homotopy group counts the winding from one $\mathbf{S}^3$ to another. For higher dimensional G, it turns out that it's sufficient to pick an $SU(2)$ subgroup of G and consider maps which wind within that. You then need to check that these maps cannot be unwound within the larger G.

It can be shown that, in general, the winding $n \in \mathbb{Z}$ is computed by

$$n(\Omega) = \frac{1}{24\pi^2}\int_{\mathbf{S}^3_\infty} d^3S\ \epsilon^{ijk}\mathrm{Tr}\,(\Omega\partial_{\rm i}\Omega^{-1})(\Omega\partial_{\rm j}\Omega^{-1})(\Omega\partial_{\rm k}\Omega^{-1}) \tag{5.28}$$

Evaluated on any configuration, the theta term becomes (5.27)

$$S_\vartheta = \vartheta n \tag{5.29}$$

It is the contribution from configurations with $n \neq 0$ in the path integral that means that observables in quantum gauge theories can depend on ϑ.

We can say more if we work in a regime in which the theory is weakly coupled. Here the path integral is dominated by the saddle points, which are solutions to the classical equations of motion. This means that any ϑ dependence should come from field equations that wind at infinity, so $n \neq 0$, and solve the classical equations of motion,

$$\mathcal{D}_\mu F^{\mu\nu} = 0 \tag{5.30}$$

There is a cute way of finding solutions to this equation. The Yang-Mills action is

$$S_{YM} = \frac{1}{2g^2}\int d^4x\ \mathrm{tr}\, F_{\mu\nu}F^{\mu\nu}$$

Note that in Euclidean space, the action comes with a + sign. This is to be contrasted with the Minkowski space action (5.1) which comes with a minus sign. We can write this as

$$S_{YM} = \frac{1}{4g^2}\int d^4x\ \mathrm{tr}\,(F_{\mu\nu} \mp {}^\star F_{\mu\nu})^2 \pm \frac{1}{2g^2}\int d^4x\ \mathrm{tr}\, F_{\mu\nu}{}^\star F^{\mu\nu} \geq \frac{8\pi^2}{g^2}|n|$$

where, in the last line, we've used the result (5.29). We learn that in the sector with winding n, the Yang-Mills action is bounded by $8\pi^2 n/g^2$. The action is minimised when the bound is saturated. This occurs when

$$F_{\mu\nu} = \pm {}^\star F_{\mu\nu} \tag{5.31}$$

These are the (anti) self-dual Yang-Mills equations. The argument above shows that solutions to these first order equations necessarily minimise the action is a given topological sector and so must solve the equations of motion (5.30). In fact, it's straightforward to see that this is the case since it follows immediately from the Bianchi identity $\mathcal{D}_\mu {}^\star F^{\mu\nu} = 0$.

Solutions to the (anti) self-dual Yang-Mills equations (5.31) have finite action, which means that any deviation from the vacuum must occur localised in Euclidean spacetime. In other words, they are point-like objects in $\mathbb{R}^4$. Because they occur for just an "instant of time" they are known as *instantons*.

There is much to say about instantons. You can read about the role they play in quantum Yang-Mills in the lectures on Gauge Theory and more about the structure of the solutions to (5.31) in the lectures on Solitons. For our purposes, it will suffice to point out that the contributions of instantons to any quantity comes with the characteristic factor

$$e^{-S_{\mathrm{instanton}}} = e^{-8\pi^2|n|/g^2} e^{i\vartheta n} \tag{5.32}$$

Famously, the function $e^{-8\pi^2/g^2}$ has vanishing Taylor expansion about the origin $g^2 = 0$. This is telling us that effects due to instantons are smaller than any perturbative contribution, which takes the form g^{2n}. Nonetheless, that doesn't mean that instantons are useless since they can contribute to quantities that apparently vanish in perturbation theory.

The theta dependence $e^{i\vartheta n}$ associated to an instanton is also interesting. It is a complex phase. The fact that it is complex can be traced to the $\epsilon^{\mu\nu\rho\sigma}$ tensor in S_ϑ. This means that S_θ contains a single time derivative and so, upon Wick rotation, still

sits in the path integral with a factor of i. The fact that $n \in \mathbb{Z}$ means that ϑ is a periodic variable, with

$$\vartheta \in [0, 2\pi)$$

Instantons are usually referred to as *non-perturbative* effects. This is a little bit of a misnomer. The use of instantons requires weak coupling $g^2 \ll 1$, so in this sense they are just as perturbative as usual perturbation theory. The name *non-perturbative* really means "not perturbative around the vacuum". Instead, the perturbation theory occurs around the instanton solution.

This also means that the theta dependence (5.32) is only expected at weak coupling $g^2 \ll 1$. As we've seen, in the far infra-red non-Abelian gauge theories are typically strongly coupled and the theta dependence of quantities can take a different form. We'll see examples in what follows.

An Example: An Instanton in $SU(2)$

It is fairly straightforward to write down the instanton solutions with winding $n = 1$. For $SU(2)$, such a configuration is given by

$$A_\mu = \frac{1}{x^2 + \rho^2}\, \eta^a_{\mu\nu} x^\nu \sigma^a \tag{5.33}$$

Here ρ is a parameter whose role we will describe shortly. The $\eta^a_{\mu\nu}$ are usually referred to as 't Hooft matrices. They are three 4×4 matrices which provide an irreducible representation of the $su(2)$ Lie algebra. They are given by

$$\eta^1_{\mu\nu} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad , \quad \eta^2_{\mu\nu} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad , \quad \eta^3_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

These matrices are self-dual: they obey $\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\eta^i_{\rho\sigma} = \eta^i_{\mu\nu}$. (Note that we're not being careful about indices up vs down as we are in Euclidean space with no troublesome minus signs.) In the solution (5.33), the 't Hooft matrices intertwine the $su(2)$ group index $a = 1, 2, 3$ with the spacetime index μ and this implements the asymptotic winding of the gauge fields.

The associated field strength is given by

$$F_{\mu\nu} = -\frac{2\rho^2}{(x^2 + \rho^2)^2}\eta^a_{\mu\nu}\sigma^a$$

This inherits its self-duality from the 't Hooft matrices: $F_{\mu\nu} = {}^\star F_{\mu\nu}$ and therefore solves the Yang-Mills equations of motion, $\mathcal{D}_\mu F_{\mu\nu} = 0$.

We can get some sense of the form of this solution. First, the non-zero field strength is localised around the origin $x = 0$. (By translational invariance, we can shift $x^\mu \to x^\mu - X^\mu$ to construct a solution localised at any other point X^μ.) The solution depends on a parameter ρ which can be thought of as the size of the instanton lump. The fact that the instanton has an arbitrary size follows from the classical conformal invariance of the Yang-Mills action.

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 exact, 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,

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

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

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

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,

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

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

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

where the surviving $\mathbb{Z}_2$ acts as fermion parity $\lambda \to -\lambda$. 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

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

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

$$U_i = \mathrm{diag}(e^{i\theta_1^i}, \ldots, e^{i\theta_{N_c}^i})$$

The zero energy modes are the coordinates θ_a^i, 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

$$\sum_{a=1}^{N_c} \theta_i^a = 0 \mod 2\pi \tag{6.10}$$

We should quantise each of these periodic rotors θ_i^a, 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	$SU(N)$	$Sp(N)$	$\text{Spin}(2N+1)$	$\text{Spin}(4N)$	$\text{Spin}(4N+2)$	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

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

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

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

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

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

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 $G = \text{Spin}(N)$ 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 $\text{Spin}(N)$ 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

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

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)

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

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

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

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_{eff} from an effective superpotential

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

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

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

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

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

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_{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_{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 $(\square, \overline{\square})$ 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 $W_{\text{eff}} \sim M^{-N_f/(N_c - N_f)}$. If we ignore the Kähler potential, then the scalar potential takes the form

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

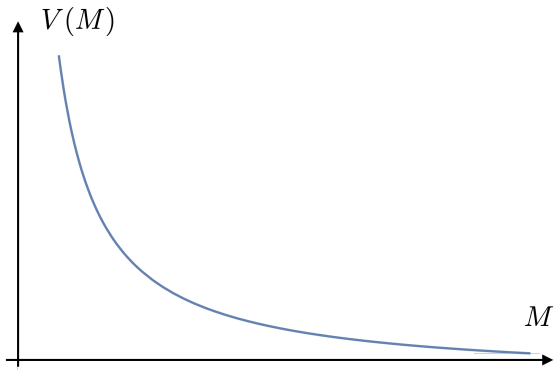


Figure 9. The runaway potential on the moduli space for $N_f < N_c$ massless flavours.

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,

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

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	$\frac{2(N_f-N_c)}{N_f}$
$\Lambda^{3N_c-N_f}$	**1**	**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

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

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

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

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),

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

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 just a single complex meson $M = \tilde{\Phi}\Phi$. The superpotential now has a critical point,

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

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

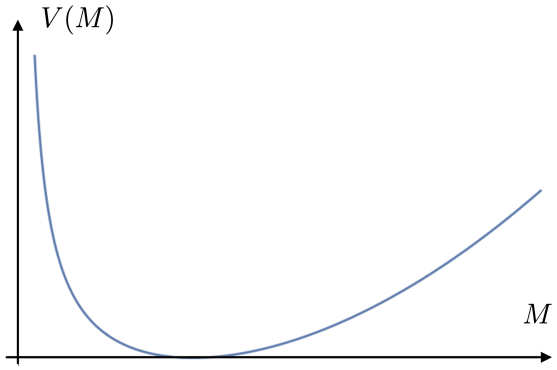


Figure 10. The rescued runaway, with a supersymmetric minimum when mass is added.

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

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

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

Comparing the two results above, we have the matching condition

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

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

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

In this case, $\Lambda_{\rm new} > \Lambda_{\rm old}$. 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 $\Lambda_{\rm new}$. 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

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

This, of course, we've seen before. It is precisely the superpotential (6.12) for super Yang-Mills, now with the strong coupling scale $\Lambda_{\rm new}$. 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

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

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

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

with $\text{Adj}(M)$ the adjugate matrix. If M is invertible then this coincides with the more familiar $\delta(\det M) = (\det M)\,\text{tr}(M^{-1}\delta M)$. 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^{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),

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

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 $\partial W/\partial M_j^i = 0$ 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

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

Taking traces and determinants gives

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

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

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

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

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

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

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

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)

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

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 $N_c < N_f < 3N_c$ 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

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

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".

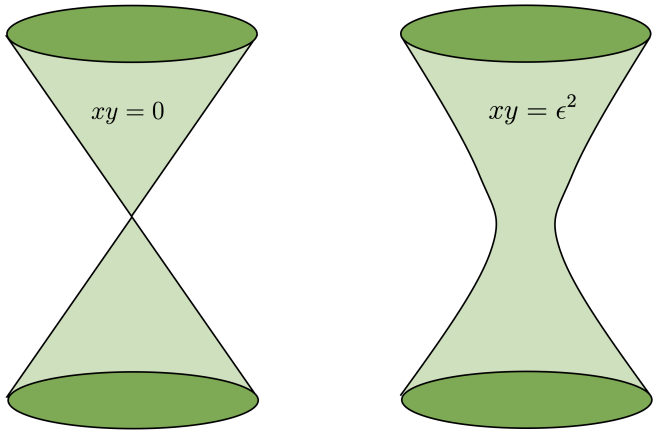


Figure 11. The singular space $xy = 0$ on the left and the smooth space $xy = \epsilon^2$ on the right. This is a cartoon for the moduli space of SQCD when $N_f = N_c$. On the left the classical, singular modular space; on the right, the smooth quantum moduli space.

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 at 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 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.

$SU(N_f)^3_{\text{diag}}$: 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 $\square \otimes \overline{\square} = \text{adj} \oplus \mathbf{1}$.

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

$\underline{SU(N_f)^2_{\text{diag}} \cdot U(1)_B}$: 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 $\mathcal{A}_{IR} = 0$.

$\underline{SU(N_f)^2_{\text{diag}} \cdot U(1)_R}$: 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

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

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 $SU(N_f)_{\text{diag}}$ 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

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

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

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

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

$$B_j = \epsilon_{ji_1\ldots i_{N_c}} B^{i_1\ldots i_{N_c}} \quad \text{and} \quad \tilde{B}^j = \epsilon^{ji_1\ldots i_{N_c}} \tilde{B}_{i_1\ldots i_{N_c}}$$

This satisfy the constraints (4.43)

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

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 $\overline{\square}$. 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

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

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

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

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

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

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 $S_{\rm CFT}$. (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

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

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

$\underline{SU(N_f)_L^3}$: 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 $\overline{\square}$ of $SU(N_f)_L$ while the quarks sit in the $\square$. This was also need to ensure that the meson fields M have the same quantum numbers.

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

$\underline{SU(N_f)_L^2 \cdot U(1)_R}$: 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 '$U(1)_B^3$ 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

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

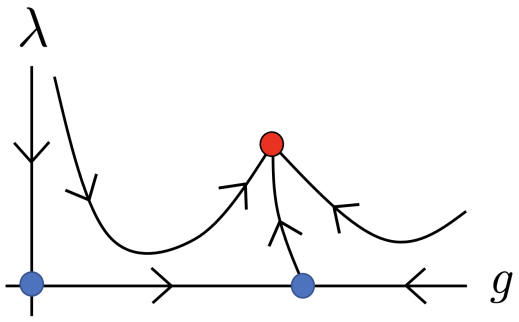


Figure 12. The RG flow in mSQCD The free fixed point and the fixed point in the conformal window are shown as black points. The superpotential induces a further flow to the red point. This is conjectured to coincide with the fixed point of SQCD.

$R[M] = 2/3$. 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.

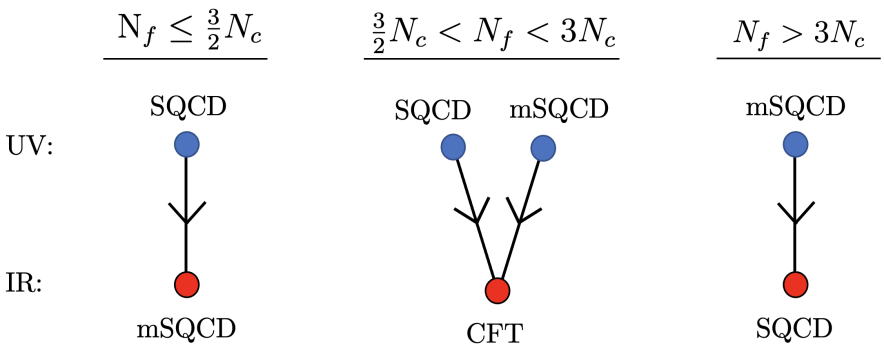


Figure 13. Seiberg duality is a statement about RG flows, although the precise statement changes as we vary N_f/N_c.

As we increase $N_f \geq 3N_c$, 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,

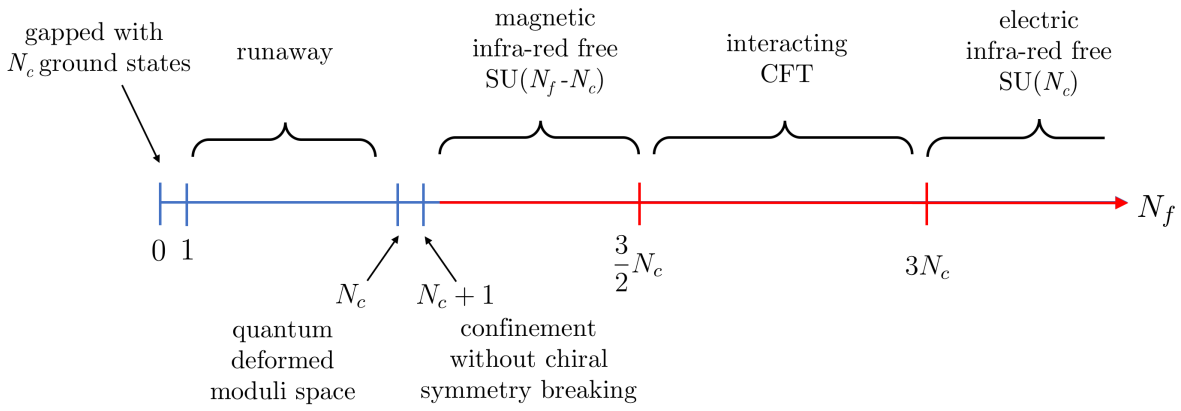


Figure 14. The phases of massless SQCD. For the values of N_f shown in red, we have a dual description in terms of mSQCD. This dual description is weakly coupled from $N_f = N_c + 2$ to $N_f = 3N_c/2 + \epsilon$.

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

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

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

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

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

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

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

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

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 $F^{\mu\nu} \to {}^\star F^{\mu\nu}$. In terms of the underlying electric and magnetic fields, this means

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

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

$$\partial_\mu F^{\mu\nu} = J_e^\mu \quad \text{and} \quad \partial_\mu {}^\star F^{\mu\nu} = J_m^\mu$$

with J_e^μ and J_m^μ 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.

7 More Supersymmetric Gauge Dynamics

There are many more interesting properties of $\mathcal{N} = 1$ gauge theories. In this section, we describe a few of them.

7.1 Other Gauge Groups

One obvious generalisation of the previous results comes from looking at other gauge groups. There is a similar story for both $Sp(N)$ and $SO(N)$ gauge groups, with a runaway potential for a small number of flavours and a dual description available in the conformal window. It tuns out that $SO(N)$ is significantly more complicated, with a number of twists and turns along the way[13]. Here we give the details only for the much simpler case $Sp(N)$.

The classical Lie group $Sp(N)$ is subgroup of $SU(2N)$ that leaves invariant the anti-symmetric tensor

$$J = \mathbb{1}_N \otimes i\sigma^2$$

The group $Sp(N)$ has dimension $N(2N+1)$, rank N and the fundamental representation has dimension $2N$. For the lowest rank we have

$$Sp(1) = SU(2)$$

Be warned: you will find different naming conventions for this group in the literature. Some authors prefer $USp(2N)$ to $Sp(N)$, where the argument now describes the dimension of the smallest representation rather than the rank. More confusingly, other authors write $Sp(2N)$ for $Sp(N)$!

7.1.1 $Sp(N)$ Quantum Dynamics

In this section, we consider $Sp(N_c)$ gauge theory coupled to $2N_f$ chiral multiplets Q_i in the fundamental representation[14].

The representations of $Sp(N_c)$ are pseudoreal which means that there's no sense in which the matter comes in conjugate pairs. Nonetheless, there's a subtle effect in $Sp(N_c)$ gauge theories called the *Witten anomaly* that means that $Sp(N_c)$ gauge theories only make sense when coupled to an even number of fundamental Weyl fermions. Hence the $2N_f$ above.

[13] A question on the examples sheet covers the key duality. You can find the full details in the original paper by Ken Intriligator and Nati Seiberg.

[14] This theory was first discussed by Ken Intriligator and Philippe Pouliot.

To understand this theory, we can largely follow the path laid down in the previous section. The 1-loop beta function is given by

$$b_0 = 3(N_c + 1) - N_f$$

Next, the symmetries. In the case of $N_f = 0$, the $U(1)_R$ symmetry is anomalous with a surviving $\mathbb{Z}_{2(N_c+1)}$. This, in turn, is spontaneously broken to $\mathbb{Z}_2$ by a gluino condensate $\langle \operatorname{Tr} \lambda\lambda \rangle \neq 0$, giving $N_c + 1$ ground states. Indeed, this coincides with the Witten index

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

When $N_f > 0$, there is a surviving R-symmetry. Taking into account the anomaly, the symmetries of the theory are

	$Sp(N_c)$	$SU(2N_f)$	$U(1)_A$	$U(1)_R$
Q	$\square$	$\square$	1	$1 - \frac{N_c+1}{N_f}$
Λ^{b_0}	**1**	**1**	$2N_f$	0

This is largely sufficient for us to understand what becomes of the quantum dynamics of this theory.

First, we should understand the classical dynamics. For $Sp(N_c)$ gauge theories there are no baryons and the classical moduli space is parameterised solely by mesons,

$$M_{ij} = Q_{ia} Q_{jb} J^{ab} \tag{7.1}$$

with $a, b = 1, \ldots, 2N_c$ the group index and $i, j = 1, \ldots 2N_f$ the flavour index. Importantly, these mesons are anti-symmetric in the flavour indices: $M_{ij} = -M_{ji}$.

When $N_f \leq N_c$, there are no further constraints on these mesons. The classical moduli space has dimension $\dim \mathcal{M} = N_f(2N_f - 1)$. At a generic point, the gauge group is broken from $Sp(N_c)$ to $Sp(N_f - N_c)$.

For $N_f > N_c$, there is a constraint arising from the fact that the mesons M have $\operatorname{rank}(M) \leq 2N_c$. This classical constraint can be written as

$$\epsilon^{i_1 \ldots i_{2N_f}} M_{i_1 i_2} M_{i_3 i_4} \ldots M_{i_{2N_c+1} i_{2N_c+2}} = 0 \tag{7.2}$$

At a generic point, the $Sp(N_c)$ gauge group is broken completely. As with the $SU(N_c)$ theories, this moduli space has singularities whenever the rank drops below the maximal. These signify the emergence of massless, unbroken gauge bosons.

So much for the classical theory. What about the quantum? Given our earlier results about SQCD, we might expect that a superpotential is generated, lifting the moduli space for some low N_f. We can use the symmetries above to determine what superpotential is possible. First, we need to form an object that is invariant under the $SU(2N_f)$ flavour symmetry. For $SU(N_c)$ SQCD, this was the determinant of the meson matrix. But for $Sp(N_c)$, we have something a little different. This is because the meson (7.1) is necessarily anti-symmetric in the i, j flavour indices which means that it's natural to consider the *Pfaffian*, defined by

$$(\text{Pf}\, M)^2 = \det M$$

This has U(1) charges $R[\text{Pf}\, M] = 2(N_f - N_c - 1)$ and $A[\text{Pf}\, M] = 2N_f$.

Runaway for $N_f \leq N_c$

The symmetries allow a unique dynamically generated superpotential

$$W = C\left(\frac{\Lambda^{3(N_c+1)-N_f}}{\text{Pf}\, M}\right)^{1/N_c+1-N_f} \tag{7.3}$$

for some coefficient C. This superpotential only makes sense for $N_f \leq N_c$ where it gives rise to a runaway potential, lifting all ground states. For the case $N_f = N_c$, the gauge group is completely broken and here the superpotential arises from an instanton with the characteristic signature Λ^{b_0}. An explicit weak coupling calculations shows that $C \neq 0$ and the superpotential is indeed generated.

As for SQCD, giving the flavours a mass stabilises the vacua at a finite distance and reveals the $N_c + 1$ ground states expected by the Witten index. If we crank up the mass and integrate out the massive flavours, we can derive the runaway superpotential, together with the coefficient C, for all smaller values of N_f.

Deformed Moduli Space for $N_f = N_c + 1$

For $N_f = N_c + 1$, the classical constraint (7.2) reads

$$\text{Pf}\, M = 0$$

For this choice of N_f, we have $R[M] = 0$ and there is an opportunity for the classical constraint to pick up a quantum deformation to

$$\text{Pf}\, M \sim \Lambda^{2(N_c+1)} \tag{7.4}$$

The classical moduli space had singularities arising from massless gauge bosons. These are removed in the quantum moduli space, signalling confinement.

To see this the quantum deformation does indeed occur, we can repeat the analysis of SQCD and integrate out the last flavour. The only real difference comes from the fact that M_{ij} is now anti-symmetric. We start with a superpotential imposing the constraint, together with a mass term for the final flavour which we call Z

$$W = X(\text{Pf}\, M - \Lambda_{\text{old}}^{2(N_c+1)}) + mZ \quad \text{with} \quad Z = M_{2N_c+1,2N_c+2} \tag{7.5}$$

where we're not being too careful about the overall coefficient in front of the quantum deformation. (There are some annoying factors of 2 that appear in the $Sp(N_c)$ analysis that aren't there for $SU(N_c)$.) We write the meson matrix as

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

The equation of motion for Z and X give

$$X = -\frac{m}{\text{Pf}\,\tilde{M}} \quad \text{and} \quad Z = \frac{\Lambda_{\text{old}}^{2(N_c+1)}}{\text{Pf}\,\tilde{M}}$$

Substituting this back into the constrained superpotential (7.5) reproduces the expected runaway behaviour (7.3) with the matched RG scales $\Lambda_{\text{new}}^{2N_c+1} = \Lambda_{\text{old}}^{2(N_c+1)} m$.

We can also do some 't Hooft anomaly matching. When M satisfies the quantum modified constraint (7.4), the global symmetry is broken to

$$SU(2N_f) \times U(1)_R \to Sp(N_f) \times U(1)_R$$

There is no need to match the $Sp(N_f)$ anomalies because the relevant group theoretic cubic invariant simply vanishes for $Sp(N_c)$. But we still have others

$\underline{Sp(N_f)^2 \cdot U(1)_R}$: In the UV we have just the quarks with $R[\psi] = -1$. The 't Hooft anomaly is

$$\mathcal{A}_{\text{UV}} = -2N_c$$

In the IR, we have only mesons. The chiral superfields have R-charge $R[M] = 0$, so the fermions have charge -1. They transform in the anti-symmetric representation of $Sp(N_f)$. This has dimension $\dim(\square\!\square) = N_f(2N_f - 1) - 1$ and Dynkin index $I(\square\!\square) = 2N_f - 2$. The 't Hooft anomaly is then

$$\mathcal{A}_{IR} = -(2N_f - 2) = -2N_c$$

$\underline{U(1)_R^3}$: In the UV we have both gluinos and quarks, contributing

$$\mathcal{A}_{\rm UV} = N_c(2N_c+1)\times(+1)^3 + 4N_cN_f\times(-1)^3 = -N_c(2N_c+3)$$

In the IR, we have just the mesons, giving

$$\mathcal{A}_{\rm IR} = -N_f(2N_f-1) - 1$$

which agrees with $\mathcal{A}_{UV}$. A similar counting also shows that the mixed $U(1)_R$-gravitational anomaly matches.

Confinement Without χSB for $N_f = N_c + 2$

Now there can be neither a superpotential generated on the moduli space, nor a quantum deformation of the constraints. We are left with the classical moduli space, subject to the classical constraint (7.2). This space has a singularity at the origin.

As with SQCD, the constraints are not imposed by a Lagrange multiplier, but instead arise as the equations of motion from the superpotential

$$W = \frac{\text{Pf}\, M}{\Lambda^{2N_c+1}}$$

Once again, we propose that the quantum interpretation of this singularity is different from the classical interpretation. The gauge gauge bosons, which are classically massless, are thought to confine with the singularity at $M = 0$ arising because all $\frac{1}{2}\times(2N_f)\times(2N_f-1)$ elements of the anti-symmetric meson matrix M are massless.

Once again, this proposal must pass the stringent tests of 't Hooft anomaly matching. We have

$\underline{SU(2N_f)^3}$: In the UV, the quarks give $\mathcal{A}_{UV} = 2N_c$. In the infra-red, the mesons sit in the anti-symmetric representation and $\mathcal{A}_{IR} = \mathcal{A}(\square\!\square)$. This is given by $\mathcal{A}(\square\!\square) = 2N_f - 4 = \mathcal{A}_{UV}$.

$\underline{SU(2N_f)^2\cdot U(1)_R}$: The quarks now have R-charge $R[\psi] = -(N_c+1)/(N_c+2)$ and so contribute to the UV 't Hooft anomaly as $\mathcal{A}_{UV} = -2N_c(N_c+1)/(N_c+2)$. In the IR, the mesons have R-charge $R[M] = 2/N_f$ and, of course, the fermions in this chiral multiplet have R-charge $R[M]-1$. For $SU(2N_f)$, the Dynkin index of the anti-symmetric representation is $I(\square\!\square) = 2N_f - 2$, so we have $\mathcal{A}_{IR} = 2(N_f-1)\times(2/N_f-1) = \mathcal{A}_{UV}$.

$\underline{U(1)_R^3}$: The gluinos and quarks give

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

Meanwhile, the mesons give

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

$\underline{U(1)_R}$: This time the mixed $U(1)_R$-gravitational anomaly gives a different counting. We have

$$\mathcal{A}_{UV} = N_c(2N_c+1)\times(+1) + 4N_cN_f \times \left(\frac{1}{N_f}-1\right) = -2N_f^2+5N_f-2$$

Meanwhile, the mesons give

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

Again, we see that all 't Hooft anomalies match as they should.

7.1.2 Seiberg Duality

For $N_f \geq N_c+3$, we turn to a dual description. The claim is that $Sp(N_c)$ with $2N_f$ chiral multiplets is dual to

$$Sp(\tilde{N}_c) \text{ with } 2N_f \text{ chiral multiplets } q \text{ in the fundamental and singlets } M_{ij}$$

Here M_{ij} sits in the anti-symmetric representation of the $SU(2N_f)$ flavour symmetry and is coupled to the other fields through the superpotential

$$W = M_{ij}\, q_a^i q_b^j J^{ab}$$

with $a,b = 1,\ldots,\tilde{N}_c$ and $i,j = 1,\ldots,N_f$. The rank of the dual gauge group should be taken to be

$$\tilde{N}_c = N_f - N_c - 2$$

One can perform all the same tests of Seiberg duality that we saw for $SU(N_c)$ SQCD. The proposal passes them all.

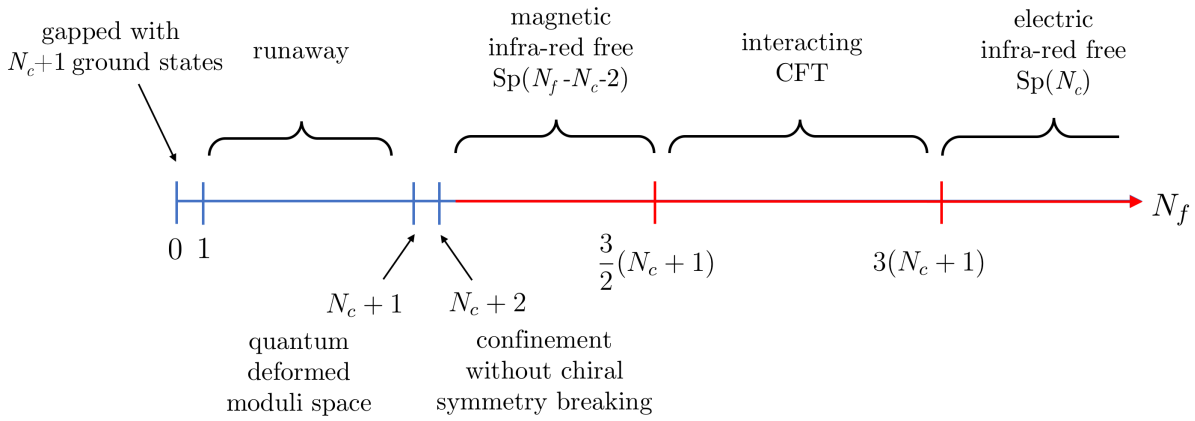


Figure 15. The phases of $Sp(N_c)$ gauge theory with $2N_f$ massless, fundamental chiral multiplets.

For now, we can use the duality to put together the phase diagram for $Sp(N_c)$ with $2N_f$ fundamental chiral multiplets. It looks very similar to the $SU(N_c)$ case, with just the numbers changing.

Jumping first to large N_f, the original electric theory is infra-red free when $N_f \geq 3(N_c+1)$. For $N_c+3 \leq N_f \leq \frac{3}{2}(N_c+1)$, the magnetic theory is infra-red free. For $\frac{3}{2}(N_c+1) < N_f < 3(N_c+1)$, both theories flow to the same conformal fixed point. The upshot is that the phase diagram for $Sp(N_c)$ theories looks very similar to that of $SU(N_c)$ SQCD. It is shown in Figure 15.

7.1.3 $SU(2)$ Gauge Theory Revisited

As we mentioned at the beginning of this section, $Sp(1) = SU(2)$. That means that we now have two different stories for $SU(2)$ gauge theory, one presented here and the other in Section 6. We should check to make sure that they are consistent.

Things start out looking fine. For $N_f = 0$, the Witten index tells us that there are two ground states. For $N_f = 1$, there is just a single meson field M and in both descriptions we have the superpotential

$$W = \frac{\Lambda^5}{M}$$

For $N_f = 2$, our two descriptions are the same, but with slightly different names for various objects. In the $SU(N_c)$ language, we introduced four mesons M_{ij}, with $i,j = 1,2$ and two baryons B and $\tilde{B}$, making 6 in total. In the $Sp(1)$ language, we only

have mesons that, to avoid confusion, we'll call $\hat{M}_{ij}$. These have $i, j = 1, \ldots, 4$ with the requirement that $\hat{M}_{ij} = -\hat{M}_{ji}$ again making 6 in total. One can show that

$$\det M - \tilde{B}B = \text{Pf}\, \hat{M}$$

This means that both the classical constraint, and the quantum deformed constraint, coincide in the two descriptions.

There is a similar story when $N_f = 3$. Now in the $SU(2)$ description there are 9 mesons M and 6 baryons B and $\tilde{B}$, while in the $Sp(1)$ description there are $\frac{1}{2} \times 6 \times 5$ mesons $\hat{M}$.

Things start to get more interesting when we move into the realm $N_f \geq 4$ where the dual description is available to us. The gauge invariant operators M, B and $\tilde{B}$ still match the mesons $\hat{M}$. But the dual descriptions are very different.

To see this, let's look at $SU(2)$ with $N_f = 4$ flavours. The two dual descriptions are based on $SU(N_f - N_c)$ and $Sp(N_f - N_c - 2)$ gauge theories respectively, which happily coincide for $N_c = 2$ and $N_f = 4$. But the singlet fields which couple through a superpotential are different. The $SU(N_f - N_c)$ dual gives

$$SU(2) \text{ with } N_f = 4 \text{ flavours and } W = \textstyle\sum_{i,j=1}^{4} \tilde{q}_i M_{ij} q_i$$

The global symmetry of this theory is $SU(4)^2 \times U(1)$, acting on the $\tilde{q}$ and q individually. Meanwhile the $Sp(N_f - N_c - 2)$ dual gives

$$SU(2) \text{ with } N_f = 8 \text{ chiral multiplets and } W = \textstyle\sum_{i,j=1}^{8} q_i \hat{M}_{ij} q_i$$

Now we haven't split the matter into two sets, q and $\tilde{q}$. Correspondingly, the theory has a much larger $SU(8)$ global symmetry. From our discussion above, both of these theories must flow to the same IR fixed point. This means that the first theory must develop the full $SU(8)$ flavour symmetry in the infra-red. In fact, it turns out that there are a number of other ways to split the matter multiplets, giving different duals. You can read more about this in the lectures by Yuji Tachikawa.

For $N_f \geq 5$, things start to look even more different. For example, when $N_f = 5$ one dual is an $SU(3)$ gauge theory while the other is an $Sp(2) = \text{Spin}(5)$ gauge theory. We see that dual theories can come in different forms: there is nothing that tells us that there is a unique dual (or, indeed, any dual) for a given gauge theory.

7.2 A Chiral Gauge Theory

A chiral gauge theory is defined to be one in which left and right handed fermions transform differently under the gauge group. In the supersymmetric context, this means that chiral multiplets do not come in conjugate pairs.

It's not completely straightforward to write down consistent chiral gauge theories because we have to make sure that there are no gauge anomalies. Furthermore, in the absence of supersymmetry, chiral theories are those that we understand least, in large part because the Nielsen-Ninomiya theorem provides an obstacle to simulating these theories on a computer. Notably, the Standard Model is an example of a chiral gauge theory, albeit one where the chiral interactions are weakly coupled and so we can use perturbation theory to understand what's going on.

The purpose of this section is to describe the dynamics of some simple supersymmetric chiral theories.

7.2.1 $SU(N)$ with a Symmetric

Consider a $G = SU(N)$ gauge theory, with a single chiral multiplet S in the symmetric representation and $N+4$ chiral multiplets $\tilde{Q}_i$ in the anti-fundamental. This is a consistent chiral theory because the the symmetric representation $\square\square$ contributes $\mathcal{A}(\square\square) = N+4$ to the $SU(N)$ anomaly, which is subsequently cancelled by the $\tilde{Q}^i$ with $i = 1, \ldots, N+4$, each of which contributes $\mathcal{A}(\overline{\square}) = -1$.

The symmetry structure of the theory is

	$SU(N)$	$SU(N+4)$	$U(1)_F$	$U(1)_R$
S	$\square\square$	$\mathbf{1}$	$N+4$	$-\frac{N-2}{N+2}$
$\tilde{Q}$	$\overline{\square}$	$\square$	$-(N+2)$	1

There is a large classical moduli space, parameterised as always by gauge invariant, holomorphic monomials of the matter fields. These are:

$$\begin{aligned} \text{mesons}: &\quad M^{ij} = \tilde{Q}^i S \tilde{Q}^j \\ \text{flavour singlet}: &\quad U = \det S \\ \text{baryons}: &\quad B = \tilde{Q}^N \\ \text{more baryons}: &\quad B' = (\tilde{Q}S)^N \end{aligned} \tag{7.6}$$

where the baryons are contracted with an $SU(N)$ epsilon symbol; there are $\binom{N+4}{N}$ of them. As always, there are some constraints among these operators, including $M^N = UB^2$ and $B' = UB$.

There is no superpotential that we can write down consistent with the symmetries, so this moduli space survives in the quantum theory. (The flavour singlet U has charge under $U(1)_B$, while other flavour singlets that you might think you could construct, such as $\det M$ or M^4B^2 vanish identically.)

We can move out along the moduli space in various directions, breaking the gauge and global symmetries in some manner. The physics far out along the moduli space can be understood using weakly coupled analysis (possibly with some strong coupling physics of the unbroken part of the gauge group still to deal with). Here we would like to understand what happens at the origin of the moduli space.

First note that there's no issue with asymptotic freedom in these theories. As the number of flavours increases, so too does the number of colours and the theories are asymptotically free for all N. However, there is an issue with the unitarity bound (6.40). This tells us that any chiral operator in an interacting superconformal theory must have R-charge

$$R_{IR}[\mathcal{O}] > \frac{2}{3} \tag{7.7}$$

where, crucially, R_{IR} is the R-charge at the superconformal point. In general, this may not coincide with the R-symmetry that we identify in the UV. Indeed, there's an ambiguity in our choice of R-symmetry in the table above: we made a specific choice, but we could equally as well have chosen a new R-symmetry which involved the old one, together with a mix of $U(1)_F$. In general, the IR R-symmetry could be a mix

$$R_{IR} = R + \alpha F \tag{7.8}$$

for some $\alpha \in \mathbb{R}$. We don't yet have any way to determine which combination should be identified with the R-symmetry of the conformal field theory.

We will, in fact, explain how we can identify R_{IR} in Section 7.2.4. But for now, let's take the most general case (7.8) and look at the R-charges of two of our chiral operators, M and U. They are

$$R_{IR}[M] = \frac{N+6}{N+2} - \alpha N \quad \text{and} \quad R_{IR}[U] = N\left[-\frac{N-2}{N+2} + \alpha(N+4)\right]$$

You can see immediately that, for large N, there is going to be a problem satisfying the unitarity bound (7.7). The first term for $R_{IR}[U]$ is negative, so we must take $\alpha > 0$. But then, for large enough N, we will necessarily have $R_{IR}[M] < 0$. A short calculation shows that there is no choice of R-symmetry for which $R_{IR}[M] > \frac{2}{3}$ and $R_{IR}[U] > \frac{2}{3}$ whenever $N \geq 13$.

This suggests that the chiral theory flows to a free infra-red theory when $N \geq 13$ and to an interacting SCFT when $N < 13$. In fact, for the intermediate case of $N = 13$, there is a choice for which $R_{IR}[M] = R_{IR}[U] = \frac{2}{3}$, suggesting again that these fields may be free.

7.2.2 A Chiral Duality

To better understand the infra-red physics, we can try to find a dual description. It turns out that the chiral gauge theory described above has a rather startling dual[15]. It has gauge group

$$\tilde{G} = \text{Spin}(8)$$

This group, which is the double cover of $SO(8)$, is rather special as it has three, inequivalent represents all of dimension 8. These are the vector $\mathbf{8}_v$, the spinor $\mathbf{8}_s$ and the conjugate spinor $\mathbf{8}_c$. The dual theory has a single chiral multiplet p in the spinor representation and $N + 4$ chiral multiplets q_i in the vector representation. In addition, there are Spin(8) singlet fields M^{ij} and U and a superpotential

$$W = M^{ij} q_i q_j + U p p \tag{7.9}$$

The symmetry structure of the theory is

	Spin(8)	$SU(N+4)$	$U(1)'_F$	$U(1)'_R$
q	$\mathbf{8}_v$	$\overline{\square}$	-1	1
p	$\mathbf{8}_s$	$\mathbf{1}$	$N+4$	-5
M	1	$\square\square$	2	0
U	$\mathbf{1}$	$\mathbf{1}$	$-2(N+4)$	12

Let's first see why these two theories might be dual to each other. First, each have the same global symmetry $SU(N + 4) \times U(1)^2$. Note, however, that we haven't yet made any attempt to match the two Abelian symmetries across the duality. We'll do this shortly.

[15] This was first found by Philippe Pouliot and Matt Strassler. A closely related duality was previously found by Pouliot, and his name is sometimes attached to these dualities. They are also referred to simply as *Seiberg Dualities* as a catch-all for this kind of behaviour.

In addition, the gauge invariant chiral superfields match. For our Spin(8) theory, the obvious qq and pp mesons are killed by the equations of motion of the superpotential. (Indeed, this is largely the purpose of the superpotential.) We do, however, have the singlets M^{ij} and U whose names already suggest how they might map to the original theory,

$$\begin{aligned} \tilde{Q}^i S \tilde{Q}^j &\quad\longleftrightarrow\quad M^{ij} \\ \det S &\quad\longleftrightarrow\quad U \end{aligned}$$

Moreover, we can use these to understand how the Abelian symmetries map across both sides of the duality. The symmetries match if we rescale the global symmetry a

$$2F = -NF'$$

We can't rescale the R-symmetry because it's fixed by the requirement that $R[\text{gluino}] = 1$. However, the two R-symmetries on either side of the duality can differ by a flavour symmetry. You can check that the R-symmetries match if we take

$$R' = R + \frac{1}{N}\frac{N+6}{N+2}F$$

With these redefinitions, our group of symmetries read

	Spin(8)	$SU(N+4)$	$U(1)_F$	$U(1)_R$
q	$\mathbf{8}_v$	$\overline{\square}$	$\frac{1}{2}N$	$\frac{N-2}{2(N+2)}$
p	$\mathbf{8}_s$	$\mathbf{1}$	$-\frac{1}{2}N(N+4)$	$\frac{N^2+4}{2(N+2)}$
M	1	$\square\!\square$	$-N$	$\frac{N+6}{N+2}$
U	$\mathbf{1}$	$\mathbf{1}$	$N(N+4)$	$-\frac{N(N-2)}{N+2}$

These most likely aren't the R-symmetries that you would have chosen. But they're the R-symmetries we've got!

We haven't yet discussed the baryons of either theory. It turns out that these too agree, as do the moduli spaces, but there's a subtlety awaiting us so we will postpone that discussion to Section 7.2.3. Instead, with the symmetries in hand we can turn to the next check: 't Hooft anomaly matching. For example, those involving the non-Abelian global symmetry are

$\underline{SU(N+4)^3}$: In the electric theory, we have $\mathcal{A}_{\text{el}} = N$. In the magnetic theory, the q contribute $\mathcal{A}_{\text{mag}} = -8$ while the mesons M contribute $\mathcal{A}_{\text{mag}} = (N+4)+4$, so

$\mathcal{A}_{\rm el} = \mathcal{A}_{\rm mag}$ as it should.

$\underline{SU(N+4)^2 \cdot U(1)_F}$: In the electric theory, we have $\mathcal{A}_{\rm el} = -N \times (N+2)$. In the magnetic theory we have $\mathcal{A}_{\rm mag} = 8 \times (\frac{1}{2}N) + (N+4+2) \times (-N) = \mathcal{A}_{\rm el}$.

$\underline{SU(N+4)^2 \cdot U(1)_R}$: Since $R[\tilde{Q}] = 1$ the corresponding fermions are uncharged and we have $\mathcal{A}_{\rm el} = 0$. In the magnetic theory, $\mathcal{A}_{\rm mag} = 8 \times (\frac{N-2}{2(N+2)} - 1) + (N+4+2) \times (\frac{N+6}{N+2} - 1) = 0$.

We won't check all of the others, but here are a couple to give you a sense. For the mixed $U(1)_R$-gravitational anomaly we have

$\underline{U(1)_R}$: This has $\mathcal{A}_{\rm el} = (N^2-1) + \frac{1}{2}N(N+1) \times (-\frac{N-2}{N+2} - 1) = \frac{(N-2)(N+1)}{N+2}$ where the contributions are from the gluino and the S field respectively. In the Spin(8) magnetic theory, we have

$$\begin{aligned}\mathcal{A}_{\rm mag} = 28 &+ 8(N+4)\left(\frac{N-2}{2(N+2)} - 1\right) + 8\left(\frac{N^2+4}{2(N+2)} - 1\right) \\ &+ \frac{1}{2}(N+4)(N+5)\left(\frac{N+6}{N+2} - 1\right) - \left(\frac{N(N-2)}{N+2} - 1\right) = \frac{(N-2)(N+1)}{N+2}\end{aligned}$$

while for the $U(1)_R^3$ anomaly we have

$\underline{U(1)_R^3}$: This has $\mathcal{A}_{\rm el} = -\frac{1}{2}N(N+1)\left(\frac{N-2}{N+2} + 1\right)^3$. Meanwhile,

$$\begin{aligned}\mathcal{A}_{\rm mag} = 28 &+ 8(N+4)\left(\frac{N-2}{2(N+2)} - 1\right)^3 + 8\left(\frac{N^2+4}{2(N+2)} - 1\right)^3 \\ &+ \frac{1}{2}(N+4)(N+5)\left(\frac{N+6}{N+2} - 1\right)^3 - \left(\frac{N(N-2)}{N+2} - 1\right)^3\end{aligned}$$

A little algebra (or Mathematica) shows you that $\mathcal{A}_{\rm el} = \mathcal{A}_{\rm mag}$. Needless to say, the other 't Hooft anomalies involving $U(1)_F$ and mixed $U(1)_R$, $U(1)_F$ also coincide. As alway, the agreement of these fairly complicated algebraic expressions gives some confidence that the two theories are indeed related in some way.

Consequences for the Infra-Red Dynamics

Let's now run with the conjecture that these two theories are dual. The magnetic Spin(8) theory has the one-loop beta function given by

$$b_0 = \frac{3}{2} \times (8-2) - \frac{1}{2}(N+5) = \frac{1}{2}(13-N)$$

We see that the theory is asymptotically free only when $N < 13$. But this agrees perfectly with our previous analysis of the conformal window of the electric theory! The duality tells us that the chiral theory is indeed infra-red free when $N \geq 13$, but the free theory is a Spin(8) gauge theory, with the matter described above. Needless to say, it's unlikely that we would have guessed this starting the $SU(N)$ gauge theory.

Meanwhile, for $2 \leq N \leq 12$, both theories are expected to flow to an interacting SCFT. The statement of Pouliot duality here is that, once we include the superpotential (7.9), the two theories flow to the same SCFT.

A Deformation of the Duality

As always, given a duality we can deform it in different ways to derive new (or perhaps old) dualities. Indeed, understanding how connections in the web of different dualities is an important consistency check on any new proposal.

There are many ways to deform our chiral duality. Here we just mention two particularly straightforward ones. First, suppose that we add

$$W = \det S \tag{7.10}$$

to the electric side. We have the same gauge theory, just with this additional superpotential.

It's obvious what happens on the magnetic side: the superpotential (7.9) becomes

$$W = Mqq + U(pp + 1)$$

where we're not being careful about including coefficients, dimensionful or otherwise, for these various terms. The equation of motion for U now means that $p \neq 0$ in the ground state. This induces a Higgs mechanism and breaks the magnetic gauge symmetry Spin(8) $\rightarrow$ Spin(7) in such a way that the other chiral superfields q, that previously transformed in $\mathbf{8}_v$, now transform in the $\mathbf{8}$ spinor representation of Spin(7).

This gives us a new duality: the electric chiral theory with superpotential (7.10) is dual to Spin(7) gauge theory with $N+4$ chiral multiplets in the spinor representation $\mathbf{8}$, coupled to singlets through $W = Mqq$. (This is actually the original "Pouliot duality".) The magnetic theory is now infra-red free for any $N \geq 11$.

This version of Pouliot duality has a surprising feature. Our original $SU(N)$ theory was a chiral gauge theory. But its Spin(7) dual is non-chiral! In particular, for $N \geq 11$, the chiral $SU(N)$ theory flows in the infra-red to the non-chiral Spin(7) theory. There is a lesson in this: the question of whether or not a theory is chiral depends on the energy scale at which you look. It is not a property that is preserved under RG.

Another Deformation

Alternatively, we could give an expectation value to $U = \det S$. On the electric side, this gives a mass to the spinor p, allowing us to integrate it out. We're left just with $SO(8)$ gauge theory coupled to $N+4$ chiral multiplets in the $\mathbf{8}_v$, still, of course, coupled to the superpotenial $W = Mqq$. (I'm ignoring global issues of the gauge group here.)

What happens on the original electric side? We give an expectation value to the symmetric $S \neq 0$. This breaks $SU(N) \to SO(N)$, so we're left with an $SO(N)$ gauge theory coupled to $N+4$ fundamental chiral multiplets. The claim is that this is dual to the $SO(8)$ theory above.

In fact, this is part of the $SO(N)$ Seiberg dualities which, in general, relate an $SO(N_c)$ theory to an $SO(N_f - N_c + 4)$ theory.

7.2.3 Briefly, the Konishi Anomaly

There's a loose thread hanging from our discussion of Pouliot duality. The electric theory includes two baryon operators

$$B = \tilde{Q}^N \quad \text{and} \quad B' = (\tilde{Q}S)^N$$

We haven't yet seen what they map to on the magnetic side. Happily, the Spin(8) theory also contains two baryon operator which, schematically, take the form

$$b = q^4 p^2 \quad \text{and} \quad b'' = q^8$$

Here the q^8 in b'' are contracted with an epsilon tensor. We need a little group theory to explain how b is put together. The vectors q^4 combine in an anti-symmetric fashion into $\mathbf{35}_s + \mathbf{35}_c$ and the latter is contracted with the two spinors which combine symmetrically into $\mathbf{35}_s$ so that the whole thing is a singlet of Spin(8).

It seems reasonable to think that these operators might map into each other under duality. To see this, we can check the flavour and R-symmetry charges. We have

$$\begin{aligned} F[B] = -N(N+2) \quad &\text{and} \quad R[B] = N \\ F[B'] = 2N \quad &\text{and} \quad R[B'] = \frac{4N}{N+2} \end{aligned}$$

and

$$\begin{aligned} F[b] = -N(N+2) \quad &\text{and} \quad R[b] = N \\ F[b''] = 4N \quad &\text{and} \quad R[b''] = \frac{4(N-2)}{N+2} \end{aligned}$$

It's close but, sadly, no cigar! First, it's clear that under the duality we should match

$$b \quad \longleftrightarrow \quad B$$

But while the flavour charge of B' and b'' agree, their R-charge does not! What's going on?

In fact, there is a subtlety in this duality that didn't rear its head in our previous examples. To fully understand the structure of chiral superfields, we should include one further field from each theory, each of which involves the chiral superfield that houses the field strength. We call this W_α for the electric theory and $\tilde{W}_\alpha$ for the magnetic theory. Then consider

$$B'' = (\tilde{Q}^{N-4}S^{N-2})W_\alpha W^\alpha \quad \text{and} \quad b' = q^4\,\tilde{W}_\alpha\tilde{W}^\alpha$$

If we use the fact that $R[W^2] = R[\tilde{W}^2] = 2$, we find $F[B''] = F[b'']$ and $R[B''] = R[b'']$ and $F[b'] = F[B']$ and $R[b'] = R[B']$. So this solves our matching problem: the baryons on one side are paired chiral fields that include the field strength of the other

$$\begin{aligned} b' \quad &\longleftrightarrow \quad B' \\ b'' \quad &\longleftrightarrow \quad B'' \end{aligned}$$

But this also opens up a whole can of worms! Why are we suddenly including the field strength in the story? Or, said differently, why didn't we include the field strength in Section 6 when discussing $SU(N)$ SQCD?

The answer to this is a little subtle. Here I don't give all the details, but sketch the basic idea. It turns out that one can derive an equation in SQCD that, for each chiral multiplet, reads

$$\bar{D}^2(Q^\dagger Q) = Q\frac{\partial W}{\partial Q} + \frac{1}{8\pi^2}\mathrm{Tr}\, W_\alpha W^\alpha$$

This equation is known as the *Konishi anomaly* and is the supersymmetric version of the chiral anomaly which says that a rotation $Q \to e^{i\alpha Q}$ results in a shift of the theta angle. It tells us that, at least as far as the chiral ring is concerned, the operator $\mathrm{Tr}\, W_\alpha W_\alpha$ can be replaced by $Q\partial W/\partial Q$, so we're not missing anything if we neglect it.

However, in other theories there are a number of these additional chiral multiplets, dressed with W_α, that you need to include. This first rears its head in the duality for $SO(N_c)$ theories (which we didn't describe in these lectures notes, in part to duck this particular issue). For the chiral duality that we've described above, it turns out that you need to include the extra B'' and b' (and, in fact, one further operator from each theory that depends linearly on W_α or $\tilde{W}_\alpha$ respectively).

7.2.4 Briefly, a-Maximisation

We've seen a few times in these lectures that many theories don't have a unique R-symmetry. Instead, we can always add any linear combination of other Abelian flavour symmetries and this also provides a good candidate R-symmetry. This becomes an issue only when we flow to an interacting SCFT, where the R-symmetry dictates the dimension of chiral operators

$$\Delta[\mathcal{O}] = \frac{3}{2} R_{IR}[\mathcal{O}]$$

But for this to be useful, we need to know exactly what R-symmetry we're dealing with in the infra-red.

Happily, there is a simple prescription to determine this. This prescription, known as *a-maximisation*[16], is straightforward to state but somewhat harder to prove. Here we just give the statement, dressed with a little context.

First, in any conformal field theory the trace of the stress tensor necessarily vanishes: $\langle T^{\mu}{}_{\mu} \rangle = 0$. At least, this is true in flat space. But if the theory is placed on a curved manifold, there is a so-called *trace anomaly* and we get

$$\langle T^{\mu}{}_{\mu} \rangle = \frac{c}{16\pi^2} C_{\mu\nu\rho\sigma} C^{\mu\nu\rho\sigma} - \frac{a}{16\pi^2} {}^*R_{\mu\nu\rho\sigma} {}^\star R^{\mu\nu\rho\sigma}$$

where $C_{\mu\nu\rho\sigma}$ is the Weyl tensor and ${}^*R$ is the dual of the Riemann tensor. (We proved the analogous statement for 2d CFTs in the lectures on String Theory.) The two coefficients a and c are known as *central charges* and provide a way to characterise the CFT.

Of the two, a is the more interesting. First, it can be proven that a always decreases under RG flow. Second, in superconformal field theories it turns out that a is determined by the R-charge

$$a = \frac{3}{32} \sum_{\text{fermions}} 3R_{IR}^3 - R_{IR}$$

where the sum should be taken over left-handed Weyl fermions.

[16] It was first derived by Ken Intriligator and Brian Wecht. After landing a faculty job in London, Brian moved on to an infinitely more interesting career as a ninja, fighting dinosaurs with lasers and, ultimately, finding a much larger audience for his derivation of a-maximisation.

Once again, it's important that we use the right R-symmetry R_{IR} when computing the central charge a. However, the beauty of this calculation is that it gives us a way to figure out what the right central charge is. Suppose that we have a collection of candidate central charges in the UV, parameterised by some coefficients α as in (7.8). For each of these we can compute the would-be central charge

$$a(\alpha) = \frac{3}{32} \sum_{\text{fermions}} 3R(\alpha)^3 - R(\alpha)$$

The R-symmetry that appears in the superconformal algebra turns out to the one that maximises the value of a. This gives a simple way to compute R_{IR} and, therefore, the dimensions of chiral operators in the SCFT.

7.3 Dynamical Supersymmetry Breaking

All the gauge theories that we've discussed so far have supersymmetric vacua with vanishing energy. In some cases these vacua are pushed off to infinity by a runway potential, but we can always rescue them by giving masses to the matter multiplets, bringing them in to finite distance. One might wonder: do all supersymmetric gauge theories have supersymmetric ground states? Or is it possible that some gauge theories spontaneously break supersymmetry, with a ground state that has energy $E > 0$?

We already met some models that break supersymmetry back in Section 3.4. There, we worked only with chiral multiplets and the game was to cook up a superpotential which for which no critical points exist. In searching for gauge theories that break supersymmetry, the game is similar. The difference is that now there is the option for the superpotential to be generated by quantum effects. Such theories are said to break supersymmetry dynamically.

Where should we look for dynamical supersymmetry breaking? An obvious obstacle is the Witten index. This is non-vanishing for super Yang-Mills theory with any gauge group. (It is given by the dual Coxeter number and is listed for all gauge groups in Table 3.) If we add matter in any vector-like representation, we can always give it a mass and reduce to super Yang-Mills with its non-vanishing Witten index. This suggests two places to look for supersymmetry breaking.

- We could consider chiral gauge theories in which it's not possible to give the matter mass.
- Alternatively, we could consider gauge theories with a quantum moduli space of vacua for which the Witten index is ill-defined. It may then be possible to deform these theories in some other way that doesn't involve giving masses.

In this section, we give two examples of dynamical supersymmetry breaking, one of each kind.

7.3.1 The $SU(3) \times SU(2)$ Model

One of the simplest chiral gauge theories we can write down is based on the gauge group

$$G = SU(3) \times SU(2)$$

We introduce a collection of four chiral multiplets, with quantum numbers given by

	$SU(3)$	$SU(2)$	$U(1)_Y$	$U(1)_R$	$U(1)_A$	$U(1)'_A$
Q	$\mathbf{3}$	$\mathbf{2}$	1	-1	1	1
$\tilde{U}$	$\overline{\mathbf{3}}$	$\mathbf{1}$	-4	0	1	0
$\tilde{D}$	$\overline{\mathbf{3}}$	$\mathbf{1}$	2	0	1	0
L	$\mathbf{1}$	$\mathbf{2}$	-3	3	0	1
Λ_3^7	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{1}$	0	-4	0
Λ_2^4	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{1}$	0	0	-4

We've also included both non-anomalous and anomalous $U(1)$ symmetries in this table. Classically there is a $U(1)^4$ symmetry, but quantum mechanically only a $U(1)^2$ survives. The anomalous $U(1)$ symmetries are $U(1)_A$ and $U(1)'_A$, as shown by the transformation of the strong coupling scales. The exponents in these strong coupling scales can be traced to the one-loop beta functions, which are

$$SU(3): \ b_0 = 9 - 2 = 7 \quad \text{and} \quad SU(2): \ b_0 = 6 - 2 = 4$$

If you know the smallest amount of particle physics, these quantum numbers should look very familiar! They are the representations of the quarks and leptons of the Standard Model. (The right-handed electron is missing.) The symmetry $U(1)_Y$ coincides (up to a normalisation) with the hypercharge symmetry of the Standard Model, here a global rather than gauge symmetry.

It's curious that, as we shall see, this theory dynamically breaks supersymmetry although it doesn't seem particularly useful for real-world purposes: the MSSM must include the Higgs fields (which, of course, also sit in chiral multiplets). Various phenomenological constraints means that supersymmetry breaking is thought to take place in an entirely different sector before being communicated to the Standard Model by so-called "messenger" fields. Here we study the theory simply to get a feeling for what chiral gauge theories do.

First, the classical moduli space. As we've seen, this is parameterised by the gauge invariant holomorphic monomials. For our current theory, there are three:

$$Y_1 = \tilde{U}QL \quad , \quad Y_2 = \tilde{D}QL \quad , \quad Z = \tilde{U}Q\,\tilde{D}Q$$

where the $SU(2)$ gauge indices are contracted with an ϵ^{ab} symbol in each. These have R-charge $R[Y_1] = R[Y_2] = 2$ and $R[Z] = -2$. This means that we can add a tree level superpotential that preserves the R-symmetry,

$$W_{\text{tree}} = \lambda \tilde{D}QL = \lambda Y_2$$

with λ a (classically) dimensionless constant. This superpotential is renormalisable and also preserves $U(1)_Y$.

The superpotential W_{tree} lifts the vacuum moduli space. To see this, note that the critical point requires

$$\frac{\partial W_{\text{tree}}}{\partial L} = 0 \quad \Rightarrow \quad \tilde{D}Q = 0 \quad \Rightarrow \quad Y_2 = Z = 0$$

and

$$\frac{\partial W_{\text{tree}}}{\partial \tilde{D}} = 0 \quad \Rightarrow \quad Q\tilde{L} = 0 \quad \Rightarrow \quad Y_1 = X_2 = 0$$

This means that if there is supersymmetric ground state then it necessarily sits at the origin of moduli space where the theory is strongly coupled.

Now let's turn to the quantum dynamics. For λ suitably small, we can ignore the tree-level superpotential and import our results from Section 6. Things are easiest if we assume that $|\Lambda_3| \gg |\Lambda_2|$ so that the $SU(3)$ dynamics becomes strong first. In this case we have $SU(3)$ with $N_f = 2$ flavours which, we know, is the situation where a non-perturbative superpotential is generated by instantons. Adding this to our tree-level superpotential gives

$$W = \lambda Y_1 + \frac{\Lambda_3^7}{Z} \tag{7.11}$$

The quantum generated superpotential gives a runaway that pushes the ground state towards infinity. Meanwhile, we've already seen that the tree level superpotential pushes the ground state towards the origin. The net result is shown in Figure 16, with a ground state that sits at energy $E > 0$ and hence breaks supersymmetry.

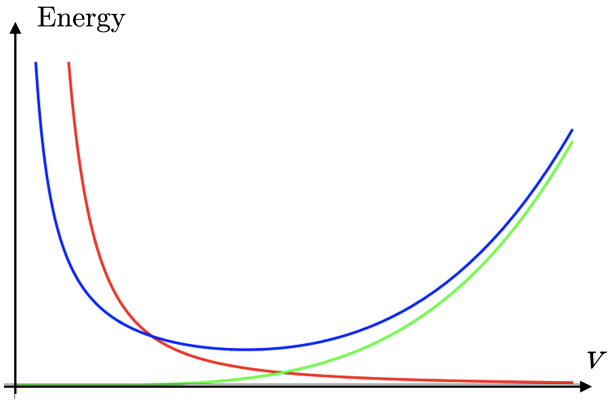


Figure 16. The tree level superpotetnial, shown in green competes with the dynamically induced superpotential, shown in red. The sum of the two, shown in blue, has a minimum at $E > 0$ and so breaks supersymmetry.

The above analysis was very quick. You might wonder if perhaps one can play off the two contributions to find a minimum at zero energy after all. In fact there's a cute argument that say this can't happen. Here's why. First note that each of Y_1, Y_2 and Z carry non-zero R-charge. Wherever the minimum of (7.11) sits, one of these must get an expectation value and so R-symmetry is broken with a corresponding Goldstone mode called an *R-axion*. This is a compact scalar. If supersymmetry is unbroken, then there must be another non-compact, massless scalar that joins with the R-axion to form the lowest component of a chiral multiplet. Usually such non-compact scalars take us out along the moduli space. But we've seen that the moduli space is lifted by the tree-level superpotential, so no such massless scalar exists and supersymmetry is necessarily broken.

We could be more precise, finding the minima of the potential in terms of the fundamental fields but this is a little fiddly. However, there's one feature that is important. From the form of the superpotential (7.11), we would expect the expectation value v of the fundamental fields to scale as

$$v \sim \frac{\Lambda_3}{\lambda^{1/7}}$$

This means that for $\lambda \ll 1$, we have $v \gg |\Lambda_3| \gg |\Lambda_2|$. As long as the expectation values break the gauge group completely the theory is weakly coupled and we can compute everything reliably. In particular, we are free to use the canonical Kähler potential in this regime.

7.3.2 The Quantum Moduli Space Revisited

As a second example of supersymmetry breaking, we take a theory that has a moduli space of vacua, and hence an ill defined Witten index. We then deform it in such a way that supersymmetry is broken.

To this end, consider $SU(2)$ gauge theory coupled to four chiral multiplets Φ^i, $i = 1, \ldots, , 4$, each in the fundamental representation. The gauge invariant operators consist of six mesons

$$M^{ij} = \Phi^i_a \Phi^j_b \epsilon^{ab}$$

(This is the $Sp(1)$ language of Section 7.1. In the $SU(2)$ language of Section 6, both mesons and baryons are housed in the 4×4 matrix $M^{ij} = -M^{ji}$.)

Classically, the mesons obey the constraint $\mathrm{Pf}\, M = 0$ where the Pfaffian is defined by

$$\mathrm{Pf}\, M = \epsilon_{ijkl} M^{ij} M^{kl}$$

We now add six singlet fields $S_{ij} = -S_{ji}$ to our original theory. These couple to the original fields through the tree-level superpotential

$$W_{\mathrm{tree}} = \lambda S_{ij} \Phi^i \Phi^j$$

This lifts the moduli space parameterised by M which must take value $M = 0$, but the theory retains a classical moduli space, parameterised by the expectation values of S_{ij}.

Now we turn to the quantum theory. We know from our discussion in Section 6.3 (or from Section 7.1) that, before adding the singlets, the quantum moduli space is deformed in the quantum theory and becomes $\mathrm{Pf}\, M = \Lambda^4$. The superpotential of our theory with the singlets is now

$$W = \lambda S_{ij} M^{ij} + X \left(\mathrm{Pf}\, M - \Lambda^4 \right)$$

with X a Lagrange multiplier field. But it's clear that the equations of motion of X and of S cannot be simultaneously satisfied: therefore this simple model breaks supersymmetry[17].

[17] This model was first proposed by Izawa and Yanagida and Intriligator and Thomas.

In fact, we should be a little more careful. This theory has a flat direction, albeit one with energy $E > 0$. To see suppose that we place ourselves far out along the classical direction $S \neq 0$. This gives the original quarks Φ a large mass and so they can be integrated out. The low-energy superpotential is

$$W_{\text{eff}} \sim \left(\lambda^2 \Lambda^4 S_{ij} S^{ij}\right)^{1/2} \tag{7.12}$$

The behaviour on S follows on symmetry grounds, including the fact that $R[S] = 2$. The behaviour on the couplings can be deduced from matching scales after integrating out the quarks, with $\Lambda^6_{\text{new}} = \Lambda^4_{\text{old}} m^2 = \Lambda^4_{\text{old}} \lambda^2 S^2$ and the superpotential is simply $W_{\text{eff}} = \Lambda^3_{\text{new}}$ as in (6.12).

If we assume a canonical Kähler potential for S, then the superpotential (7.12) results in the potential

$$V \sim |\lambda \Lambda^2|^2 \, \frac{S_{ij} S^{\dagger\, ij}}{|S_{ij} S^{ij}|}$$

As we vary the phases of different S_{ij} components, this potential diverges in some directions, but also has flat directions in which $V \sim |\lambda \Lambda^2|^2$.

Because we've broken supersymmetry, these flat directions will surely be lifted by quantum effects. (They are sometimes called pseudo-flat directions for this reason). The concern is that these quantum effects might lead to a runaway behaviour, so that rather than breaking supersymmetry we instead have a theory with no good ground state. Integrating out the quarks gives a logarithmic correction to the Kähler potential for S, along the lines of (3.38). You need to be careful about the signs, but it turns out that this causes the potential to grow as we move out along the flat directions. The ground state is pushed towards smaller values of S and breaks supersymmetry.

Because this model is vector like, we could add masses for the quark fields. What then happens? To see this, it's actually useful to add to mass terms: one for the quarks and another for S. After the quantum modification of the moduli space, the superpotential becomes

$$W_{\text{eff}} = \lambda S_{ij} M^{ij} + m_{ij} M^{ij} + \tilde{m} \text{Pf}\, S + X(\text{Pf}\, M - \Lambda^4)$$

Now there are supersymmetric ground states! They sit at

$$M^{ij} \sim \epsilon^{ijkl} m_{jk} \left(\frac{\Lambda^4}{\text{Pf}\, m}\right)^{1/2} \quad \text{and} \quad S_{ij} \sim \frac{m_{ij}}{\tilde{m}} \left(\frac{\Lambda^4}{\text{Pf}\, m}\right)^{1/2}$$

The square roots allow for two different signs, and these are the two expected supersymmetric ground states since $\mathrm{Tr}(-1)^F e^{-\beta H} = 2$ for $SU(2)$ super Yang-Mills. But we can also see what happens as the masses are removed. As $m_{ij} \to 0$, we get a smooth limit for M^{ij} (because $\mathrm{Pf}\, m \sim m^2$). But as $\tilde{m} \to 0$, the supersymmetric ground state decouples as $S \to \infty$. Naively, one might think that this leads to runaway behaviour (as it does, for example, for $SU(N_c)$ with $N_f < N_c$ flavours). The novelty in the current case is that there is an infinite barrier between the supersymmetric ground state at $S \to \infty$ and the supersymmetry breaking ground state at finite S. If you like, the maximum of this barrier must also have moved off to infinity as $\tilde{m} \to 0$.

It is straightforward to construct generalisations of this model using other theories that exhibit a quantum deformed moduli space, including $SU(N_c)$ with $N_f = N_c$ and $Sp(N_c)$ with $N_f = N_c + 1$.