INTRODUCTION TO SUPERSYMMETRY (PHYS 661)

Instructor: Philip Argyres (pronounced "are–JEER–us")
Office: Room 334 Newman Lab., 255-4118, argyres@hepth.cornell.edu.

Course Outline

I. Qualitative supersymmetry — 4 lectures, few indices
- a. Coleman-Mandula theorem
- b. Supersymmetric QM: vacua, superfields, & instantons

II. Perturbative supersymmetry — 8 lectures, the basics
- a. Chiral multiplets
- b. Nonrenormalization theorems
- c. Vector multiplets

III. Supersymmetric model building — 4 lectures, qualitative issues
- a. Supersymmetric standard model
- b. Soft supersymmetry breaking terms
- c. Messenger sectors

IV. Non-perturbative supersymmetry — 9 lectures, mostly SU(n) SQCD
- a. Higgs vacua (& instantons)
- b. Coulomb vacua (& monopoles)
- c. Chiral theories

V. Dynamical supersymmetry breaking — 2 lectures, if I get to it

There is no text covering the contents of this course. Some useful references for its various parts are:

1. Qualitative supersymmetry: E. Witten, "Dynamical breaking of supersymmetry," *Nucl. Phys.* B188 (1981) 513; S. Coleman, "The uses of instantons," in *The Whys of Subnuclear Physics* (Plenum, 1979), and in *Aspects of Symmetry* (Cambridge, 1985).

2. Perturbative supersymmetry: J. Wess and J. Bagger, "Supersymmetry and supergravity," 2nd ed., Ch. I–VIII, XXII, App. A–C. Note: I will try to follow the notation and conventions of this book in the course.

3-5. Phenomenology and non-perturbative methods: Various reviews and papers which I'll mention as we get closer to these topics.

Prerequisites are a basic knowledge of QFT (gauge theories, path integrals, 1-loop RG) on the physics side, and an aquaintance with analysis on the complex plane (holomorphy, analytic continuation) as well as rudimentary group theory (SU(3), Lorentz group, spinors) on the math side.

This is a pass/fail course. There will be no tests or final. The grade will be based on participation and doing or discussing with me problems which will be given during the course of the lectures and will also be posted on the web page

`http://www.hepth.cornell.edu/~argyres/phys661/index.html`

which can also be found by folowing the appropriate links from the high energy theory home page. These problems are meant to be simple exercises, not lengthy or difficult research projects—so please keep solutions brief, and be sure to come talk to me if you are having trouble with them. My office hours are the hour after lectures and other times by appointment.

I. Qualitative Supersymmetry

1. The Coleman-Mandula Theorem

1.1. Introduction

This course is an introduction to

4-dimensional global N=1 supersymmetric field theory,

so not, in particular, other dimensions, supergravity, or extended supersymmetry (except very briefly). I'll introduce the basics of perturbative supersymmetry and apply them to a critical survey of models of weak-scale supersymmetry in the 1st half of the semester. The first two weeks will introduce supersymmetric quantum mechanics (QM) to try to separate the features special to supersymmetry from the complications of QFT in $3+1$ dimensions. The 2nd half will be devoted to exploring the non-perturbative dynamics of supersymmetric FTs. Throughout the course I will introduce and use advanced QFT techniques (effective actions, RG flows, anomalies, instantons, ...) when needed, and I will try to emphasize qualitative explanations and symmetry-based techniques and "tricks." Many of these techniques, though used frequently in QFT, are not usually taught in the standard QFT graduate courses.

The aims of this course are two-fold. The first is to supply you with the wherewithall to evaluate the various claims/hopes for weak-scale supersymmetry. The second, and closer to my interests, is to use supersymmetric models as a window on QFT in general. From this point of view, supersymmetric FTs are just especially symmetric versions of ordinary field theories, and in many cases this extra symmetry has allowed us to solve exactly for some non-perturbative properties of these theories. This gives us another context (besides lattice gauge theory) in which to think concretely about non-perturbative QFT in $3+1$ dimensions. Indeed, the hard part of this course will be the QFT, not the supersymmetry—someone who is thoroughly familiar with QFT should find the content of this course, though perhaps unfamiliar, relatively easy to understand.

Finally, this course owes alot to Nathan Seiberg: not only is his work the main focus of the 2nd half of the course, but also much of this course is closely modeled on two series of lectures he gave at the IAS in the fall of 1994 and at Rutgers in the fall of 1995.

1.2. The Coleman-Mandula theorem, or, Why supersymmetry?

Though originally introduced in early 70's we still don't know how or if supersymmetry plays a role in nature. Supersymmetry today is like non-Abelian gauge theories before the SM: "a fascinating mathematical structure, and a reasonable extension of current ideas, but plagued with phenomenological difficulties."

Why, then, have a considerable number of people been working on this theory for the last 20 years? The answer lies in the Coleman-Mandula theorem, which singles-out supersymmetry as the "unique" extension of Poicaré invariance in 3+1 or more dimensional QFT (under some important but reasonable assumptions).

In the rest of this lecture, I follow the qualitative description of the Coleman-Mandula theorem given by E. Witten "Introduction to supersymmetry," in Proc. Intern. School of Subnuclear Physics, Erice 1981, ed. A. Zichichi (Plenum Press, 1983) p. 305.

A theory of 2 free bose fields has many conserved currents.

▸ **Exercise 1.1.** Check that in a theory of 2 free bosons

$$\mathcal{L} = \partial_\mu \phi_1 \partial^\mu \phi_1 + \partial_\mu \phi_2 \partial^\mu \phi_2, \tag{1.1}$$

the following currents are conserved:

$$\begin{aligned} J_\mu &= \phi_1 \partial_\mu \phi_2 - \phi_2 \partial_\mu \phi_1, \\ J_{\mu\rho} &= \partial_\rho \phi_1 \partial_\mu \phi_2 - \phi_2 \partial_\mu \partial_\rho \phi_1, \\ J_{\mu\rho\sigma} &= \partial_\rho \partial_\sigma \phi_1 \partial_\mu \phi_2 - \phi_2 \partial_\mu \partial_\rho \partial_\sigma \phi_1. \end{aligned} \tag{1.2}$$

There are interactions which when added to this theory still keep J_μ conserved.

▸ **Exercise 1.2.** Check that J_μ is conserved if any interaction of the form $V = f(\phi_1^2 + \phi_2^2)$ is added to the 2 boson theory.

However, there are no Lorentz-invariant interactions which can be added so that the others are conserved (nor can they be redefined by adding extra terms so that they will still be conserved). This follows from the Coleman-Mandula theorem *Phys. Rev.* D159 (1967) 1251: *In a theory with non-trivial scattering in more than 1+1 dimensions, the only possible conserved quantities that transform as tensors under the Lorentz group are the usual energy-momentum P_μ, Lorentz transformations $M_{\mu\nu}$, and scalar quantum numbers Q_i.* (This has a conformal extension for massless particles.)

The basic idea is that conservation of P_μ and $M_{\mu\nu}$ leaves only the scattering angle unknown in (say) a 2-body collision. Additional "exotic" conservation laws would determine the scattering angle, leaving only a discrete set of possible angles. Since the scattering amplitude is an analytic function of angle (*assumption # 1*) it then vanishes for all angles.

Concrete example: Suppose we have a conserved traceless symmetric tensor $Q_{\mu\nu}$. By Lorentz invariance, its matrix element in a 1-particle state of momentum p and spin zero is

$$\langle p | Q_{\mu\nu} | p \rangle \propto p_\mu p_\nu - \frac{1}{4} \eta_{\mu\nu} p^2. \tag{1.3}$$

Apply this to an elastic 2-body collision of identical particles with incoming momenta p_1, p_2, and outgoing momenta q_1, q_2, and assume that the matrix element of Q in the 2-particle state $|p_1 p_2\rangle$ is the sum of the matrix elements in the states $|p_1\rangle$ and $|p_2\rangle$. This is true if Q is "not too non-local"—say, the integral of a local current (*assumption # 2*).

▸ **Exercise 1.3.** Show that conservation of a symmetric, traceless charge $Q^{\mu\nu}$ together with energy momentum conservation implies

$$p_1^\mu p_1^\nu + p_2^\mu p_2^\nu = q_1^\mu q_1^\nu + q_2^\mu q_2^\nu, \tag{1.4}$$

for an elastic scattering of two identical scalars with incoming momenta p_1, p_2, and outgoing momenta q_1, q_2. Show that this implies the scattering angle is zero.

For the extension of this argument to non-identical particles, particles with spin, inelastic collision, see Coleman and Mandula's paper.

The Coleman-Mandula theorem does not mention spinor charges, though. So consider a free theory of a complex scalar and a free two component (Weyl) fermion

$$\mathcal{L} = \partial_\mu \overline{\phi} \partial^\mu \phi + i\overline{\psi}\overline{\sigma}^\mu \partial_\mu \psi. \tag{1.5}$$

Again, an infinite number of conserved currents exist.

▸ **Exercise 1.4.** In a theory of a free complex boson and a free 2-component fermion, show that the currents

$$\begin{aligned} S_{\mu\alpha} &= (\partial_\rho \overline{\phi} \sigma^\rho \overline{\sigma}_\mu \psi)_\alpha, \\ S_{\mu\nu\alpha} &= (\partial_\rho \overline{\phi} \sigma^\rho \overline{\sigma}_\mu \partial_\nu \psi)_\alpha, \end{aligned} \tag{1.6}$$

are conserved. (σ and $\overline{\sigma}$ obey the Dirac-like algebra $\sigma^\mu\overline{\sigma}^\nu + \sigma^\nu\overline{\sigma}^\mu \propto \eta^{\mu\nu}$. We will discuss 2-component spinors in detail in lecture 5.)

Now, there are interactions, *e.g.* $V = g(\phi\psi^\alpha\psi_\alpha + h.c.) + g^2|\phi|^4$, that can be added to this free theory such that $S_{\mu\alpha}$ (with correction proportional to g) remains conserved. (We will return to this example in much more detail in later lectures.) However, $S_{\mu\nu\alpha}$ is never conserved in the presence of interactions.

We can see this by applying the Coleman-Mandula theorem to the anticommutators of the fermionic conserved charges

$$\begin{aligned} Q_\alpha &= \int d^3x\ S_{0\alpha}, \\ Q_{\nu\alpha} &= \int d^3x\ S_{0\nu\alpha}. \end{aligned} \tag{1.7}$$

Indeed, consider the anticommutator $\{Q_{\nu\alpha}, \overline{Q}_{\mu\beta}\}$, which cannot vanish unless $Q_{\nu\alpha}$ is identically zero, since the anticommutator of any operator with its hermitian adjoint is positive definite. Since $Q_{\nu\alpha}$ has components of spin up to 3/2, the anticommutator has components of spin up to 3. Since the anticommutator is conserved if $Q_{\nu\alpha}$ is, and since the Coleman-Mandula theorem does not permit conservation of an operator of spin 3 in an interacting theory, $Q_{\nu\alpha}$ cannot be conserved in an interacting theory. (This argument is alittle too fast—after we discuss in more detail the machinery of spinor representations of the Lorentz group in lecture 5, we will be able to make this argument correctly.)

Conservation of Q_α is permitted. Since it has spin 1/2, its anticommutator has spin 1, and there is a conserved spin-1 charge: P_μ. We thus get the (N=1) supersymmetry algebra

$$\begin{aligned} \{Q_\alpha, \overline{Q}_{\dot\beta}\} &= 2\sigma^\mu_{\alpha\dot\beta} P_\mu, \\ \{Q_\alpha, Q_\beta\} &= 0, \\ [Q_\alpha, P_\mu] &= 0. \end{aligned} \tag{1.8}$$

Why don't scalar charges and Lorentz generators also appear on the right hand side of (1.8)? Haag, Sohnius, and Lopuszanski *Nucl. Phys.* B88 (1975) 257 showed that $M_{\mu\nu}$ cannot appear by associativity of the algebra, and that scalar charges can appear only in *extended* (N=2, 4) *supersymmetry*, where there are several conserved spinor charges. Though rich and beautiful, theories with extended supersymmetry are too restrictive to describe weak-scale physics, since they require all fermions to appear in real representations of gauge groups.

2. Supersymmetric QM—Vacuum properties

In this lecture we begin with a toy model of supersymmetric QFT—supersymmetric QM. Our aim is to present in the next three lectures some of the main qualitative features of supersymmetric theories and techniques without the mathematical, notational and conceptual difficulties associated with four-dimensional QFT. Much of this lecture follows E. Witten *Nucl. Phys.* B188 (1981) 513.

2.1. Algebra and representations in 0+1 dimensions

By analogy with the supersymmetry algebra in 3+1 dimensions, we take the supersymmetry algebra in QM to be

$$\{Q^+, Q\} = 2H \qquad \{Q, Q\} = 0 \qquad [Q, H] = 0. \tag{2.1}$$

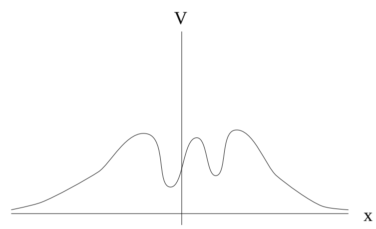
$\{Q^+, Q\}$ is positive definite. One way to see this is by taking its expectation value in any state $|\Omega\rangle$:

$$\langle\Omega|\{Q^+, Q\}|\Omega\rangle = \langle\Omega|Q^+Q|\Omega\rangle + \langle\Omega|QQ^+|\Omega\rangle = \left|Q|\Omega\rangle\right|^2 + \left|Q^+|\Omega\rangle\right|^2 \geq 0. \tag{2.2}$$

It follows from (2.1) that

$$H \geq 0. \tag{2.3}$$

This simple result lies at the core of all the "peculiar" features of supersymmetric field theories—namely, the various non-perturbative controls we have over their dynamics and the nonrenormalization theorems. Note that an inequality like this is conceptually very different from what we are used to in non-supersymmetric field theories. There there is no meaning to an overall additive constant in the energy, and the energy can be unbounded from below. With supersymmetry there is a natural zero value of the energy, which can never be unbounded from below. Of course, the energy may never attain this minimum, and there may be still be no vacuum in a supersymmetric system, since the potential $V(x)$ may slope off to infinity:



Diagonalize $H: \quad H|n\rangle = E_n|n\rangle$, so on a given eigenspace $\{Q^+, Q\} = 2E_n$. If $E_n > 0$ we can define $a^+ \equiv Q^+/\sqrt{2E_n}$, and $a \equiv Q/\sqrt{2E_n}$, so the supersymmetry algebra becomes

$$\{a^+, a\} = 1, \qquad \{a, a\} = 0, \tag{2.4}$$

a 2-dimensional Clifford algebra. Its representations should be familiar: there is one non-trivial irreducible representation which is 2-dimensional:

$$\begin{aligned} a|-\rangle &= 0 \qquad & a|+\rangle &= |-\rangle \qquad & a &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ a^+|-\rangle &= |+\rangle \qquad & a^+|+\rangle &= 0 \qquad & a^+ &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \end{aligned} \tag{2.5}$$

As with any Clifford algebra, one can define the analog of the γ_5 element which anticommutes with all the generators. In this case this is interpreted as minus one to the fermion number operator (in analogy to 3+1 dimensions):

$$(-)^F = 2a^+a - 1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.6}$$

However, when $E = 0$, the algebra becomes

$$\{Q^+, Q\} = 0. \tag{2.7}$$

There is only the trivial (one-dimensional) irrep $Q|0\rangle = Q^+|0\rangle = 0$. These states can have either fermion number: $(-)^F|0\rangle = \pm|0\rangle$. I do not know in general how $(-)^F$ is defined on these states in supersymmetric QM. In the explicit example we will introduce below, though, there will be a natural assignment. (In $2+1$ or more dimensions there is an independent definition of $(-)^F$ as the operator implementing a 2π rotation: $(-)^F = e^{2\pi i J_z}$.)

These properties are also true (qualitatively) of the supersymmetry algebra in $3+1$ dimensions. Thus the spectrum of a supersymmetric theory will have degenerate in energy (mass) and equal in number boson and fermion states at all positive energies. But, there need not be such a degeneracy among the zero energy states (supersymmetric vacua).

When there exists an $E=0$ state, we will say that "supersymmetry is unbroken", while when there is no $E=0$ state, we say that supersymmetry is (spontaneously) broken. Though this terminology is not really appropriate in QM we use it because it will describe the situation in FT. In particular, we have seen that an $E=0$ state is annihilated by the supersymmetry charges, while $E>0$ states never are.

2.2. Quantum mechanics of a particle with spin

The supersymmetry algebra can be realized in QM by a particle with two states (spin) at x described by a wavefunction

$$\Psi = \begin{pmatrix} \psi_+(x) \\ \psi_-(x) \end{pmatrix}, \tag{2.8}$$

if we define

$$\begin{aligned} Q^+ &\equiv \sigma^+ \left(P + iW'(x)\right), \qquad & \sigma^+ &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\ \sigma^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \\ Q &\equiv \sigma^- \left(P - iW'(x)\right), \qquad & P &= -i\hbar \frac{\partial}{\partial x}. \end{aligned} \tag{2.9}$$

Here $W(x)$ is a real function and we assume $\lim_{x\to\pm\infty} |W'| = \infty$ to be sure that some ground state exists. This implies

$$\begin{aligned} \{Q^+, Q\} = \underset{\text{(k.e.)}}{P^2} + \underset{\text{(p.e.)}}{(W')^2} - \underset{\text{(magn.fld.)}}{\hbar\sigma^3 W''} \equiv 2H. \end{aligned} \tag{2.10}$$

We define $(-)^F$ as simply

$$(-)^F = \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.11}$$

It is easy to see that this coincides with our previous definition of $(-)^F$ on positive energy states.

▸ **Exercise 2.1.** Show that on states of positive energy, $E>0$, $(-)^F = \frac{1}{2E}[Q^+, Q]$.

2.3. Vacua

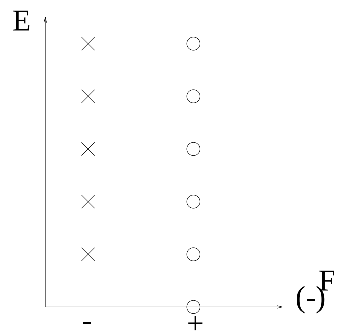
We start by looking for classical ground states. Assume there is an x_0 such that $W'(x_0) = 0$. Then

$$W'(x) = \lambda(x - x_0) + \mathcal{O}(x - x_0)^2, \tag{2.12}$$

so the potential energy is $V \simeq \frac{1}{2}\lambda^2(x - x_0)^2$ and the magnetic energy $\simeq \frac{1}{2}\lambda$.

▸ **Exercise 2.2.** Diagonalize $H = \frac{1}{2}[P^2 + (W')^2 - \hbar\sigma^3 W'']$ to find the energy levels

$$E^n_{\pm} = \hbar\lambda(n + \frac{1}{2} \mp \frac{1}{2}). \tag{2.13}$$



When we are looking for exact zero-energy states $H|\Psi\rangle = 0$, then, by the supersymmetry algebra (2.1), $Q|\Psi\rangle = Q^+|\Psi\rangle = 0$. Thus we need only look for solutions to the *first order* equations:

$$\begin{aligned} \Psi &= \begin{pmatrix} \psi_+ \\ 0 \end{pmatrix} \;\Rightarrow\; (P - iW')\psi_+ = 0 \;\Rightarrow\; \psi_+ \propto e^{-W/\hbar}, \\ \Psi &= \begin{pmatrix} 0 \\ \psi_- \end{pmatrix} \;\Rightarrow\; (P + iW')\psi_- = 0 \;\Rightarrow\; \psi_- \propto e^{+W/\hbar}. \end{aligned} \tag{2.14}$$

But for these to correspond to vacua, they must be normalizable. There are three cases:

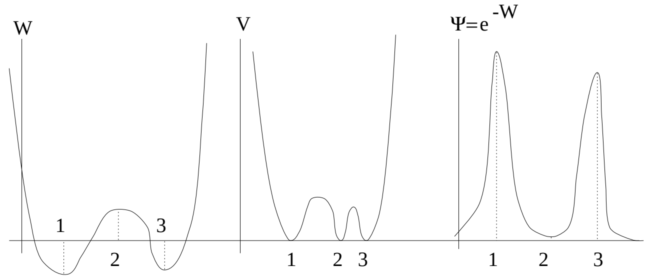
$$\begin{aligned} &(1)\ W \to +\infty \text{ as } x \to \pm\infty \;\Rightarrow\; \psi_+ \text{ normalizable, } \psi_- \text{ not;} \\ &(2)\ W \to -\infty \text{ as } x \to \pm\infty \;\Rightarrow\; \psi_- \text{ normalizable, } \psi_+ \text{ not;} \\ &(3)\ \lim_{x\to+\infty} W = -\lim_{x\to-\infty} W \;\Rightarrow\; \text{neither normalizable, and no zero-energy state.} \end{aligned} \tag{2.15}$$

With this, we have "solved" supersymmetric QM. The simplifications due to the supersymmetry algebra reducing 2nd order to first order equations is one we will see a number of times in the field theory context.

2.4. Examples

We now examine some examples illustrating these various behaviors.

(i) $W = x^4 + \dots$ (lower order) implies $V \sim x^6 + \dots$, and so will typically have three approximate ground states:

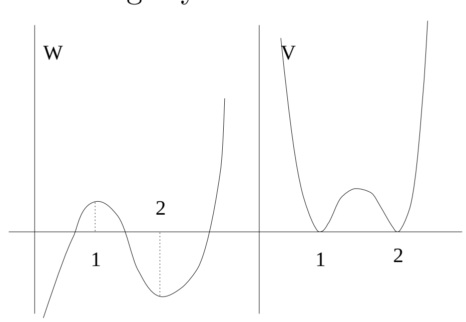


Associate the groundstate wave-functions ψ_1, ψ_2, and ψ_3 with the three classical ground states. Quantum-mechanically (exactly) we have seen that there is only a single ground state, so these states are split as

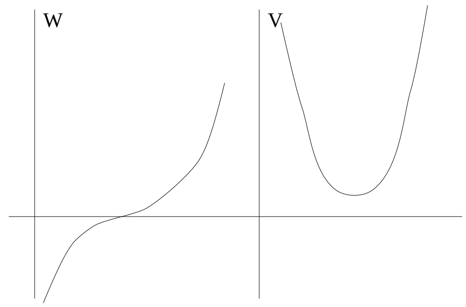
$$\begin{aligned} \Psi_{\text{exact}} = \begin{pmatrix} e^{-(x^4+\dots)/\hbar} \\ 0 \end{pmatrix} &\simeq \psi_1 + \psi_3 \\ \Psi_+ \qquad\qquad &\simeq \psi_2 \\ \Psi_- \qquad\qquad &\simeq \psi_1 - \psi_3 \end{aligned} \tag{2.16}$$

where the last two are lifted $H\psi_\pm = \epsilon\psi_\pm$ by $\epsilon \sim e^{-c/\hbar} \neq 0$. This non-perturbative effect comes from tunnelling between different (classical) ground states. We will spend lecture 3 learning how to compute this tunelling effect to leading order in $\hbar$ (*i.e.* the value of c) using semi-classical techniques.

(ii) $W = x^3 + \dots$ implying $V \sim x^4 + \dots$. In this case there are no $E = 0$ states. The two approximate vacua are mixed and slightly lifted.



(iii) W has no stationary point. Then there is no $E = 0$ state in perturbation theory, so "supersymmetry is broken classically".



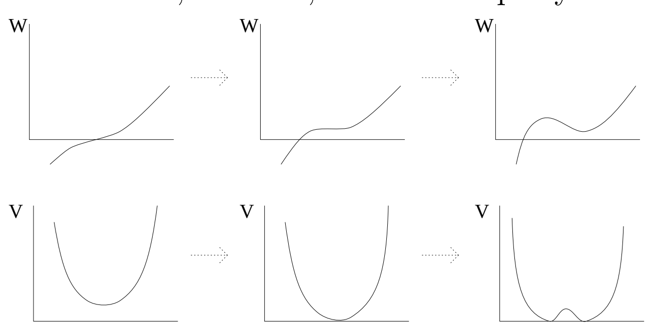
2.5. Phases: behavior of vacua as parameters vary

With this exact information in hand, let us now look at two more general arguments that give almost the same information about the ground states. Both rely on varying parameters in the QM hamiltonian.

If we vary the lower-order terms in W, for example, $W = x^3 + \lambda x^2 + \ldots$, then the various energy levels $E_i(\lambda)$ will have a *real analytic* dependence on λ simply because perturbation theory converges in λ.

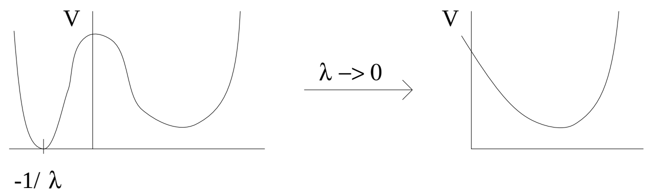
When does perturbation theory converge? Perturbation theory is a power series expansion in a parameter, say λ. Zero radius of convergence in λ (*i.e.* , non-convergence) means that there are non-analytic contributions $\mathcal{O}(e^{-1/\lambda^a})$—that is, it is an asymptotic expansion. But such contributions behave wildly as λ changes sign, so one only expects them at points where the physics does something drastic upon such a sign change. In particular, if when you change the sign (phase) of λ the Hamiltonian is no longer bounded from below, then the interval (radius) of convergence in λ is zero. (This is a famous argument of Freeman Dyson's.) Sub-leading terms in the potential energy can never have this effect, so perturbative expansions in such parameters converge.

Vary from a regime where supersymmetry is broken classically. By analyticity supersymmetry will stay broken for generic λ, since the vacuum energy is analytic in λ. (There can be special isolated values of λ, however, for which supersymmetry is not broken.)



Conversely, if supersymmetry is unbroken generically—say in some weak-coupling limit—then by analyticity, $E(\lambda) = 0$ for all λ, and so supersymmetry is always unbroken.

This is not necesarily true if you vary the leading terms in W. For example, if $W = \lambda x^4 + x^3 + \ldots$, then as $\lambda \to 0$ the zero-energy state approximately centered around $x = -1/\lambda$ will disappear by "running off to infinity".



These conclusions can be arrived at more simply by looking at the *Witten index* introduced in E. Witten, *Nucl. Phys.* B202 (1982) 253:

$$\mathrm{Tr}(-)^F e^{-\beta H}. \tag{2.17}$$

This is independent of β, since by supersymmetry only $E = 0$ states which are not paired-up can contribute—β is put in only as a regulator to make sense of the more formal $\mathrm{Tr}(-)^F$. Thus, the index really computes the number of bosonic zero-energy states minus the number of fermionic ones. If one computes $\mathrm{Tr}(-)^F \neq 0$ then supersymmetry is unbroken. If, on the other hand, it equals zero, then logically speaking it could be either broken or unbroken. But practically speaking, supersymmetry is broken since if not an arbitrarily small perturbation by a relevant operator will break it. Since the index cannot change under variations of the Hamiltonian which do not "bring states in from infinity" (*e.g.* , changing the asymptotic behavior of the potential), one can again compute it in a convenient limit of parameter space to deduce the behavior at strong coupling.

▸ **Exercise 2.3.** Deduce the generic behavior of the ground state energy in the examples given above by computing the Witten index.

Note the difference in these arguments, which gave the same information. The index argument relied only on "topological" considerations (the invariance of the index under a class of deformations), while the previous one used analyticity. In FT it will turn out that the argument based on analyticity is much more powerful. In QM, real analyticity came from the convergent nature of perturbation theory. In $3 + 1$ dimensional QFT there is a quite different source of analyticity: we will find that there are *complex* parameters that enter the Lagrangian, and a supersymmetry Ward identity shows that they can only enter holomorphically in certain terms in the effective action.

3. Supersymmetric QM—Superfields

In this lecture we rewrite the supersymmetric QM of the last lecture using anticommuting (or Grassmann) numbers. These are classical analogs of fermionic operators. This helps us do two things: (1) develop a representation theory for supersymmetric FT like that of ordinary symmetries (this lecture), and (2) compute some non-perturbative effects due to stationary points in the path integral in a semi-classical approximation (next lecture).

3.1. Fermions in QM

Let us define the QM operators ψ and ψ^+ by

$$\psi = \sqrt{\hbar}\sigma^+, \qquad \psi^+ = \sqrt{\hbar}\sigma^-, \tag{3.1}$$

so that they satisfy the algebra

$$\{\psi, \psi\} = \{\psi^+, \psi^+\} = 0, \qquad \{\psi^+, \psi\} = \hbar, \tag{3.2}$$

The point of this renaming is that now our supersymmetric QM becomes

$$\begin{aligned} Q^+ &= \psi^+(P + iW')/\sqrt{\hbar} \\ Q &= \psi(P - iW')/\sqrt{\hbar} \\ H &= \tfrac{1}{2}P^2 + \tfrac{1}{2}(W')^2 - \tfrac{1}{2}[\psi^+, \psi]\, W'', \end{aligned} \tag{3.3}$$

that is, without any explicit factors of $\hbar$ in the Hamiltonian. This allows us to identify a "classical" analog of the ψ operators, and so develop classical methods for treating fermions.

The classical limit of the algebra (3.2) of ψ's is

$$\{\psi_{cl}, \psi_{cl}\} = \{\psi^*_{cl}, \psi^*_{cl}\} = 0, \qquad \{\psi^*_{cl}, \psi_{cl}\} = 0, \tag{3.4}$$

(Here we have simply added a "cl" subscript, set $\hbar \to 0$, and changed hermitian conjugation for complex conjugation.) This is just the algebra of anticommuting (or Grassmann) numbers.

To develop the supersymmetry representation theory and use semiclassical techniques we note that the above Hamiltonian can be derived from a classical Lagrangian

$$\mathcal{L} = \frac{1}{2}\dot{x}^2 - \frac{1}{2}(W')^2 + \psi^*_{cl}\left(i\frac{d}{dt} + W''(x)\right)\psi_{cl}, \tag{3.5}$$

showing that ψ^*_{cl} is canonically conjugate to ψ_{cl}, whereas $P = \dot{x}$. (Here a dot denotes the time derivative.) We have adopted the *convention* that the complex conjugate of a product of anticommuting numbers reverses their order without introducing an extra sign.

▸ **Exercise 3.1.** Check that $\mathcal{L} = \frac{1}{2}\dot{x}^2 - \frac{1}{2}(W')^2 + i\psi^*_{cl}\dot{\psi}_{cl} + \psi^*_{cl}\psi_{cl}W''$ is real.

[Note that there is an ordering ambiguity in going from the classical $\mathcal{L}$ to the quantum H, since $\psi^+\psi$ and $\frac{1}{2}[\psi^+, \psi]$ differ by a term higher-order in $\hbar$. These higher-order "contact terms" are needed to keep supersymmetry.]

3.2. Superspace

(Notation: from now on I will denote both complex conjugation and hermitian conjugation by a bar, instead of the star and dagger used above. Also, I will drop the "*cl*" subscript and set $\hbar = 1$. It is up to the reader to figure out what is a number and what is an operator now!)

Extend space-time (in QM this is only t) to include Grassmann parameters: $t \to (t, \theta, \overline{\theta})$. Recall that Grassmann differentiation is defined by:

$$\frac{\partial}{\partial\theta}\theta = \frac{\partial}{\partial\overline{\theta}}\overline{\theta} = 1, \qquad \frac{\partial}{\partial\theta}1 = \frac{\partial}{\partial\overline{\theta}}1 = 0. \tag{3.6}$$

So with our conjugation convention,

$$\overline{\frac{\partial}{\partial\theta}} = -\frac{\partial}{\partial\overline{\theta}}. \tag{3.7}$$

Define the *covariant derivatives* by

$$\begin{aligned} D &= +\frac{\partial}{\partial\theta} - i\overline{\theta}\frac{\partial}{\partial t}, \\ \overline{D} &= -\frac{\partial}{\partial\overline{\theta}} + i\theta\frac{\partial}{\partial t}. \end{aligned} \tag{3.8}$$

Then it is easy to check that

$$\{D, \overline{D}\} = 2i\partial_t = -2H, \tag{3.9}$$

which is the supersymmetry algebra (up to a sign). Here we have made the usual identification $-i\partial_t \leftrightarrow H$.[1]

A superfield is then simply a (commuting) function on superspace:

$$X(t, \theta, \overline{\theta}) = x(t) + \theta\psi(t) - \overline{\theta\psi}(t) + \theta\overline{\theta}F(t). \tag{3.10}$$

▸ **Exercise 3.2.** Show that with our conjugation conventions, if x and F are real, then $\overline{X} = X$.

▸ **Exercise 3.3.** Calculate:

$$\begin{aligned} DX &= \psi + \overline{\theta}(F - i\dot{x}) + i\theta\overline{\theta}\dot{\psi}, \\ \overline{D}X &= \overline{\psi} + \theta(F + i\dot{x}) - i\theta\overline{\theta}\dot{\overline{\psi}}. \end{aligned} \tag{3.11}$$

[1] Tung-Mow points out that this is opposite the usual QM convention—sorry!

Our supersymmetric QM action can now be written

$$
\begin{aligned}
S &= \int dt d\theta d\overline{\theta}\left\{\tfrac{1}{2}DX\overline{D}X + W(X)\right\} \\
&= \int dt d\theta d\overline{\theta}\Big\{\tfrac{1}{2}\left(\psi+\overline{\theta}(F-i\dot{x})+i\theta\overline{\theta}\dot{\psi}\right)\left(\overline{\psi}+\theta(F+i\dot{x})-i\theta\overline{\theta}\dot{\overline{\psi}}\right) \\
&\qquad + W(x) + \left(\theta\psi - \overline{\theta\psi}+\theta\overline{\theta}F\right)W'(x) + \tfrac{1}{2}\left(\theta\psi-\overline{\theta\psi}\right)^2 W''(x)\Big\} \\
&= \int dt\left\{\tfrac{1}{2}\left(F^2+\dot{x}^2+i(\overline{\psi}\dot{\psi}-\dot{\overline{\psi}}\psi)\right) - W'F + \tfrac{1}{2}W''\cdot(\overline{\psi}\psi-\psi\overline{\psi})\right\},
\end{aligned}
\tag{3.12}
$$

where the prime denotes an X-derivative. The terms involving F are “ultralocal”—they involve no time derivatives—and so F is an auxiliary field. It’s classical equation of motion is algebraic: $F = W'$. In addition, since F appears only quadratically in the action, it gives a gaussian path integration. So we can just substitute its classical equation of motion, giving

$$
S = \int dt\left\{\tfrac{1}{2}\dot{x}^2 + i\overline{\psi}\dot{\psi} - \tfrac{1}{2}(W')^2 + \tfrac{1}{2}W''\left[\overline{\psi},\psi\right]\right\}. \tag{3.13}
$$

This is indeed our original action. [Note that I have kept track of the precise ordering of the ψ’s, and using the superfield formalism we have found the correct form to reproduce the supersymmetric action—contact terms and all.]

Let us now derive the “classical” action of the supersymmetry generators from their action on the fields (operators) in our supersymmetric QM. Recalling the basic operator algebra coming from canonical quantization

$$
[P,x] = -i, \qquad \{\overline{\psi},\psi\} = 1; \tag{3.14}
$$

the classical equations of motion

$$
\dot{\psi} = iW''\psi, \qquad F = W'; \tag{3.15}
$$

and the definition of the supersymmetry generators

$$
Q = \psi(P - iW'), \qquad \overline{Q} = \overline{\psi}(P + iW'); \tag{3.16}
$$

we compute

$$
\begin{aligned}
[Q,x] &= -i\psi \\
\{Q,\psi\} &= 0 \\
\{Q,\overline{\psi}\} &= P - iW' = \dot{x} - iF \\
[Q,F] &= [Q,W'] = -i\psi W'' = -\dot{\psi}
\end{aligned}
\tag{3.17}
$$

which can be summarized as

$$\begin{aligned}[Q,X] &= [Q,x]+[Q,\theta\psi]-[Q,\overline{\theta\psi}]+[Q,\theta\overline{\theta}F]\\ &= [Q,x]-\theta\{Q,\psi\}+\overline{\theta}\{Q,\psi\}+\theta\overline{\theta}[Q,F]\\ &= -i\psi+\overline{\theta}(\dot{x}-iF)+\theta\overline{\theta}(-\dot{\psi})\\ &= -i\left(\frac{\partial}{\partial\theta}+i\overline{\theta}\frac{\partial}{\partial t}\right)X.\end{aligned} \tag{3.18}$$

This defines Q and $\overline{Q}$ as differential operators on superspace

$$Q=+\frac{\partial}{\partial\theta}+i\overline{\theta}\frac{\partial}{\partial t}, \qquad \overline{Q}=-\frac{\partial}{\partial\overline{\theta}}-i\theta\frac{\partial}{\partial t}, \tag{3.19}$$

in analogy to ordinary symmetry transformations. It is easy to check that the supersymmetry algebra is indeed realized:

$$\{Q,\overline{Q}\}=-2i\partial_t=2H. \tag{3.20}$$

This allows us to interpret supersymmetry transformations as "translations on superspace" since acting on a superfield with a finite supersymmetry transformation gives

$$\begin{aligned}\delta_{\theta'}X &= e^{i\theta' Q}X(t,\theta,\overline{\theta})e^{-i\theta' Q}\\ &= (1+i\theta' Q)X(1-i\theta' Q)\\ &= X+i\theta'[Q,X]=X+i\theta'(\partial_\theta+i\overline{\theta}\partial_t)X\\ &= X(t{+}i\theta'\overline{\theta}\,,\,\theta{+}\theta'\,,\,\overline{\theta}),\end{aligned} \tag{3.21}$$

where θ' is a constant Grassmann parameter. Here X is any function on superspace, not necessarily just the dynamical variable appearing in our supersymmetric QM example. In components the supersymmetry transformations are of course just those found above in (3.17).

▸ **Exercise 3.4.** Check that the superspace translations

$$(t,\theta,\overline{\theta})\xrightarrow{\delta_\epsilon}(t{+}i\epsilon\overline{\theta},\theta{+}\epsilon,\overline{\theta}) \tag{3.22}$$

satisfy the supersymmetry algebra by computing the commutator of δ_ϵ with $\delta_{\overline{\epsilon}}$ and deducing the anticommutator of Q with $\overline{Q}$, *etc.* .

This can be extended to a general translation on superspace

$$\delta_{(t',\theta',\overline{\theta}')}=e^{i(t'H+\theta' Q+\overline{\theta}'\overline{Q})}, \tag{3.23}$$

so that

$$\delta_{(t',\theta',\overline{\theta}')}\delta_{(t,\theta,\overline{\theta})}=\delta_{(t+t'+i\theta'\overline{\theta}+i\overline{\theta}'\theta\,,\,\theta+\theta'\,,\,\overline{\theta}+\overline{\theta}')}. \tag{3.24}$$

Note that Q has a different relative sign compared to the covariant derivative D on superspace, and that D and $\overline{D}$ anticommute with Q and $\overline{Q}$. The existence of the covariant derivatives is due to the difference between left and right actions of the supersymmetry on superspace. In general, group associativity

$$(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3) \tag{3.25}$$

can be interpreted as saying that the action of "multiplication from the left" commutes with "right multiplication". For any left realization of supersymmetry on superspace by supersymmetry generators, there will be a right realization which (anti)commutes with it, given by the covariant derivatives on superspace. The action given by Q adn $\overline{Q}$ in (3.23) realizes (3.24) by a left action. One can equally well realize this by a right action:

$$\delta_{(t',\theta',\overline{\theta}')}\delta_{(t,\theta,\overline{\theta})} = e^{i(tH_L+\theta Q_L+\overline{\theta}\overline{Q}_L)}\delta_{(t',\theta',\overline{\theta}')}, \tag{3.26}$$

wihch must give the right hand side of (3.24). This then determines $H_L = H$, $Q_L = D$, and $\overline{Q}_L = \overline{D}$.

3.3. Supersymmetric actions and chiral superfields

Recall that Grassmann integration is the same as differentiation. The definition of integration of a single Grassmann variable θ is:

$$\int d\theta\, \theta = 1, \qquad \int d\theta\, 1 = 0. \tag{3.27}$$

It follows that any action which can be written as an integral over superspace will automatically be invariant under supersymmetry. This is because the Grassmann integration picks out the $\theta\overline{\theta}$ ("highest") component of the integrand. But, by (3.17), the supersymmetry variation of the highest component is a total space-time derivative.

Note that because Grassmann differentiation and integration are the same, we can dispense with the integration if we like. For example,

$$\int d\theta d\overline{\theta}\mathcal{L} = \int d\theta \partial_{\overline{\theta}}\mathcal{L} = -\int d\theta \overline{D}\mathcal{L}, \tag{3.28}$$

where in the last step we have added a total derivative. It is important to realize that the converse is not true: not every supersymmery-invariant term can be written as an integral over all of superspace. (Examples illustrating this are somewhat artificial in 0+1 dimensions, but they play an important role in 3+1 dimensions.)

So, consider a *"chiral" superfield*, which is a superfield satisfying the additional constraint $\overline{D}X = 0$. Noting that the cooordinate combinations θ and $\tau \equiv t - i\theta\overline{\theta}$ are themselves chiral ($\overline{D}\theta = \overline{D}\tau = 0$), it is easy to solve this constraint in general:

$$\begin{aligned} X(t,\theta,\overline{\theta}) &= X(\tau,\theta) = x(\tau) + \theta\psi(\tau) \\ &= x(t) + \theta\psi(t) - i\theta\overline{\theta}\,\dot{x}(t). \end{aligned} \tag{3.29}$$

▸ **Exercise 3.5.** Show that a product of chiral superfields is still chiral, and that $\overline{D}X$ is chiral whether X is or not.

Now supersymmetric invariants can be formed as an integral over half of superspace of a chiral field:

$$\mathcal{L} = \int dt\, d\theta X, \tag{3.30}$$

since by (3.17) the supersymmetry variation of any such term under Q vanishes, while under $\overline{Q}$ it is a total derivative. If X is chiral but not of the form $\overline{D}X$, then such a term cannot be expressed as an integral over all of superspace.

4. Supersymmetric QM—Instantons

In this lecture we perform a simple quantum-mechanical tunelling calculation, but in superspace and using a path integral formalism. This will be a warm-up for a much more complicated instanton computation in 3+1 dimensions which we will encounter about half-way through the course. The role played by bosonic and fermionic zero modes will be crucial in both contexts. A pedagogical introduction to some of this material can be found in S. Coleman, "The uses of instantons," in *The Whys of Subnuclear Physics* (Plenum, 1979), and in *Aspects of Symmetry* (Cambridge, 1985). It is strongly recommended to read at least section 2 and appendices 1-3 of Coleman's lecture, if you haven't already. First, however, we start with some technology of fermionic path integrals.

4.1. Fermionic path integrals and zero modes

In the last lecture we were essentially discussing classical supersymmetry. To see what happens quantum-mechanically, we must learn to compute expectation values (amplitudes). These can be defined by the usual path integral, but now including integrations over the Grassmann fields $\psi(t)$ and $\overline{\psi}(t)$. Since the lagrangian for the fermions is first order in time derivatives, we need to integrate over all the phase space variables in the path integral:

$$\int \mathcal{D}\overline{\psi}\mathcal{D}\psi. \tag{4.1}$$

Let us briefly review some elementary anticommuting integrals. Consider the toy action $S = \sum_i \overline{\psi}_i \lambda_i \psi_i$. Then

$$
\begin{aligned}
\int \prod_i d\overline{\psi}_i \, d\psi_i e^{-S} &= \prod_i \int d\overline{\psi}_i \, d\psi_i e^{-\overline{\psi}_i \lambda_i \psi_i} \\
&= \prod_i \int d\overline{\psi}_i \, d\psi_i (1 - \overline{\psi}_i \lambda_i \psi_i) \\
&= + \prod_i \lambda_i \int d\overline{\psi}_i \, \overline{\psi}_i \int d\psi_i \, \psi_i \\
&= \prod_i \lambda_i .
\end{aligned}
\tag{4.2}
$$

If instead $S = \sum_{ij} \overline{\psi}_i A_{ij} \psi_j$, ($A$ need not be hermitian in this example) we can reduce the path integral to the previous problem by diagonalizing A_{ij} by (different) linear transformations on ψ_j and $\overline{\psi}_i$, giving

$$
\int e^{-S} = \det A. \tag{4.3}
$$

(Note: we can always choose the diagonalizing linear transformations to have determinant one. If not, then their determinants will enter into a Jacobian for the change of variables in the grassmann integration.) Above we assumed that i and j run over the same set. Suppose instead that the number of $\overline{\psi}$'s is greater than the number of ψ's. Then by a change of bases, we can take A to

$$
A \to \begin{pmatrix} \lambda_1 & & & 0 & \\ & \ddots & & & \ddots \\ & & \lambda_N & & \end{pmatrix}, \tag{4.4}
$$

implying

$$
\begin{aligned}
\int e^{-S} &= \int \left(\prod_j^N d\overline{\psi}_j \, d\psi_j \right) \cdot \left(\prod_i{}' d\overline{\psi}_i \right) e^{-S} \\
&= \det{}' A \left(\int \prod{}' d\overline{\psi} \cdot 1 \right) = 0.
\end{aligned}
\tag{4.5}
$$

The extra $\overline{\psi}_i$'s are called the "zero modes of A". Note, though, that

$$
\int e^{-S} \prod_i{}' \overline{\psi}_i = \det{}' A \int \left(\prod{}' d\overline{\psi}_i \right) \cdot \left(\prod{}' \overline{\psi}_i \right) = \pm \det{}' A. \tag{4.6}
$$

Now let's generalize to QM—path integrals—where $\psi_i \to \psi(t)$ and $S = \int dt \overline{\psi} A \psi$, so there are an infinite number of modes. "Diagonalize" the operator A in terms of left and right eigenfunctions:

$$
\begin{aligned}
A\xi_n &= \lambda_n \eta_n, \qquad \lambda_n \in \mathbb{R}^+ \\
\overline{\eta}_n A &= \lambda_n \overline{\xi}_n \qquad (\text{or, } \overline{A} \eta_n = \lambda_n \xi_n \ )
\end{aligned}
\tag{4.7}
$$

where ξ_n and η_n are regular (commuting) functions, and can be chosen to be orthonormal[2]

$$\int dt\, \overline{\xi}_n \xi_m = \int dt\, \overline{\eta}_n \eta_m = \delta_{nm}. \tag{4.8}$$

Expand the fermionic fields in "normal modes":

$$\begin{aligned} \psi &= \sum_n \xi_n(t) a_n, \\ \overline{\psi} &= \sum_n \overline{\eta}_m(t) \overline{a}_m, \end{aligned} \tag{4.9}$$

where a_n and $\overline{a}_m$ are Grassmann numbers.

Then the path integral becomes

$$\begin{aligned} \int D\overline{\psi}\, D\psi e^{-S} &= \int \prod_n (d\overline{a}_n da_n) \exp\left\{ -\sum_{nm} \int dt\, \overline{\eta}_m(t) \overline{a}_m\, A\, \xi_n(t) a_n \right\} \\ &= \int \prod_n (d\overline{a}_n da_n) \exp\left\{ -\sum_n \lambda_n \overline{a}_n a_n \right\} \\ &= \prod_n \lambda_n \equiv \det A. \end{aligned} \tag{4.10}$$

This is true only if A has no zero modes. Zero modes of A are defined by

$$\begin{aligned} &\psi \text{ zero modes:} \qquad A\xi_0 = 0, \\ &\overline{\psi} \text{ zero modes:} \qquad \overline{\eta}_0 A = 0, \qquad (\text{or, } \overline{A}\eta_0 = 0\ ). \end{aligned} \tag{4.11}$$

Consider the case of, say, one ψ zero-mode. Then the normal mode expansion is

$$\begin{aligned} \psi &= \xi_0(t) a_0 + \sum_{n\neq 0} \xi_n(t) a_n, \\ \overline{\psi} &= \sum_{m\neq 0} \overline{\eta}_m(t) \overline{a}_m, \end{aligned} \tag{4.12}$$

so

$$\int e^{-S} = \det{}' A \left(\int da_0 \cdot 1 \right) = 0. \tag{4.13}$$

[2] To show this, note that $A\overline{A}$ is hermitean and positive. Hermiticity implies there exists an orthonormal basis of eigenvectors η_n such that $A\overline{A}\eta_n = \lambda_n^2 \eta_n$, and positivity implies we can take $\lambda_n \in \mathbb{R}^+$. Define $\xi_n = (\overline{A}\eta_n)/\lambda_n$, implying $A\xi_n = \lambda_n \eta_n$. Orthonormality of the ξ_n is easy to check.

But (do the a_0 integration first)

$$\int e^{-S}\psi(t) = \int \left(\prod_{n\neq 0} d\overline{a}_n da_n\right) da_0 \cdot \prod_{n\neq 0}(1-\overline{a}_n a_n \lambda_n)\cdot \left(a_0\,\xi_0(t) + \sum_{m\neq 0} a_m\,\xi_m(t)\right)$$
$$= \xi_0(t)\cdot \prod_n \lambda_n \equiv \xi_0(t)\det{}' A. \tag{4.14}$$

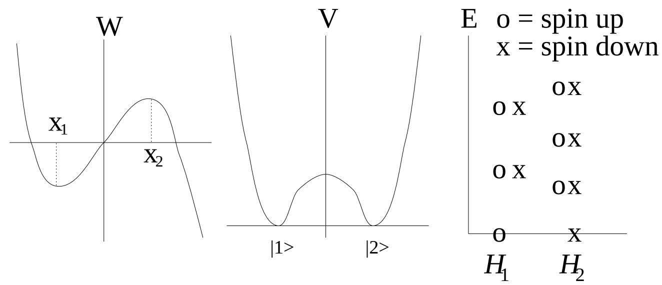
Thus a fermionic insertion "kills" a zero mode.

▶ **Exercise 4.1.** If $S=\sum_{ij}\overline{\psi}_i A_{ij}\psi_j$ and A has a single ψ zero mode, show that $\int e^{-S}\overline{\psi} = \int e^{-S}\psi(t)\overline{\psi}(t') = 0$. What is $\int e^{-S}\psi\psi\overline{\psi}$?

These simple results are very basic and are used constantly in FT.

4.2. A tunnelling problem, and its one-instanton (saddle point) approximation

To set up the problem, recall the analysis of supersymmetric QM of two lectures ago, with superpotential $W \sim x^3$. We found that supersymmetry was broken non-perturbatively—a pair of classically degenerate vacua were lifted by tunelling:



Actually the two semi-classical ground states $|1\rangle$ and $|2\rangle$ can not mix, since $\langle 1|H|2\rangle = 0$ by conservation of spin (fermion number). But $|1\rangle$ and $|2\rangle$ can still be lifted; indeed, by the arguments of lecture 2, these two states *must* be lifted (and by the same amount by supersymmetry). To leading order in $\hbar$, this lifting is due to mixing between $|1\rangle$ and the first excited state of the $|2\rangle$ sector. Let us denote the exact lowest energy states by $|0_\pm\rangle$, where the subscript refers to their spin. Then their common energy is given by

$$E_0 \simeq \langle 0_+|H|0_+\rangle = \tfrac{1}{2}\langle 0_+|\{Q,\overline{Q}\}|0_+\rangle = \tfrac{1}{2}\langle 0_+|Q\overline{Q}|0_+\rangle \simeq \tfrac{1}{2}\left|\langle 0_-|\overline{Q}|0_+\rangle\right|^2, \tag{4.15}$$

where in the last step I inserted a complete set of states, and then only kept the lowest-energy spin down state since Q commutes with the Hamiltonian and anitcommutes with $(-)^F$. Thus we are interested in computing the (semi-classical) supersymmetry-breaking order parameter

$$\epsilon \equiv \langle 0_-|\overline{Q}|0_+\rangle, \tag{4.16}$$

from which the ground state energy is given by $E_0 = \frac{1}{2}|\epsilon|^2$. The calculation of ϵ using the euclidean path integral method is done in some detail by Salomonson and van Holten *Nucl. Phys.* B196 (1982) 509.

We Wick rotate to Euclidean time, and evaluate the Feynman path integral

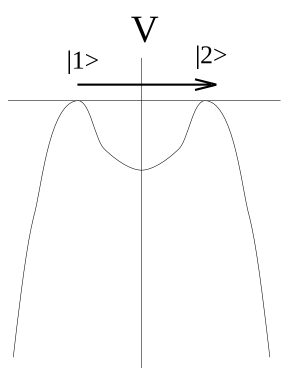
$$\begin{aligned}\int_{x(-\infty)=x_1}^{x(+\infty)=x_2} \mathcal{D}X\mathcal{D}\psi\mathcal{D}\overline{\psi}e^{-S/\hbar}\overline{Q}(t_0) &= \lim_{T\to\infty}\langle 1|e^{-H(\frac{T}{2}+t_0)}\overline{Q}e^{-H(\frac{T}{2}-t_0)}|2\rangle, \\ &= e^{-E_0T/\hbar}\langle 1|0_+\rangle\langle 0_+|\overline{Q}|0_-\rangle\langle 0_-|2\rangle, \\ &\sim \langle 0_+|\overline{Q}|0_-\rangle = \epsilon.\end{aligned} \tag{4.17}$$

Here we chose delta-function wave functions $|1\rangle = \delta(x - x_1)$ and similarly for $|2\rangle$ because they are convenient to calculate with—any function with an overlap with the ground states would do, for the $T \to \infty$ limit projects onto them, as the second line shows. In that step we inserted a complete set of energy eigenstates and kept the smallest energy exponential. In the last step we dropped the overlap factors, and the energy exponential as well, since to lowest order $E_0 = 0$. Thus, computing this path integral will allow us to read off the supersymmetry-breaking order parameter up to factors of $\mathcal{O}(1)$.

The first step in evaluating the Euclidean path integral is to find the saddle-point(s) of the Euclidean action. The Euclidean Lagrangian is

$$\mathcal{L}_E = \tfrac{1}{2}\dot{x}^2 + \tfrac{1}{2}(W')^2 - \overline{\psi}\dot{\psi} - W''\overline{\psi}\psi, \tag{4.18}$$

where each time derivative has picked up an extra factor of $-i$ due to the Wick rotation to Euclidean signature, and I have kept only the leading order terms in $\hbar$, so I have dropped the commutator in the last term. The net effect is the usual one: in Euclidean space the potential flips sign:



Solve the classical equations of motion for the Euclidean "bounce" between the two vacua. This will dominate the path integral in a saddle-point approximation. The equation of motion and boundary conditions are

$$\ddot{x} - W'W'' = 0, \qquad x(-\infty) = x_1, \quad x(+\infty) = x_2, \tag{4.19}$$

which we solve for $x_{cl}(t)$ with $\psi = \overline{\psi} = 0$ (for the moment). We can do one integral by noting that we want a zero "energy" solution: $0 = \frac{1}{2}\dot{x}^2 - \frac{1}{2}(W')^2$, implying

$$\dot{x} = \pm W'. \tag{4.20}$$

The two signs give solutions for both directions of time—we want the positive sign. Then the saddle-point action at the classical solution is

$$\begin{aligned} S &= \int dt \mathcal{L}_E(x_{cl}) = \int dt W' \, \dot{x}_{cl} = \int dt \frac{\partial W}{\partial x}\frac{\partial x}{\partial t} = \int dt \, \dot{W} \\ &= W(t{=}+\infty) - W(t{=}-\infty) = W(x_2) - W(x_1) \equiv \Delta W. \end{aligned} \tag{4.21}$$

4.3. Evaluating the path integral for the fluctuations

We now evaluate the action around x_{cl} including quadratic fluctuations:

$$\begin{aligned} S(x = x_{cl} + \delta x) = &\int dt \left[\tfrac{1}{2}\left(\dot{x}_{cl} + \delta\dot{x}\right)^2 + \tfrac{1}{2}\left(W'(x_{cl}+\delta x)\right)^2 - \delta\overline{\psi}\,\delta\dot{\psi} - W''(x_{cl}+\delta x)\delta\overline{\psi}\,\delta\psi\right] \\ &= \Delta W + \int dt\, \delta x \left(-\ddot{x}_{cl} + W_{cl}{}' W_{cl}{}''\right) \\ &\quad + \int dt \left(\tfrac{1}{2}\delta\dot{x}^2 + \tfrac{1}{2}\delta x^2[W'''W' + (W'')^2] + \delta\overline{\psi}\,\delta\dot{\psi} - W''\delta\overline{\psi}\,\delta\psi\right) \\ &= \Delta W + \tfrac{1}{2}\int dt \left(\delta x B \delta x + \delta\overline{\psi} F \delta\psi\right), \end{aligned} \tag{4.22}$$

where

$$\begin{aligned} F &\equiv \partial_t - W'', \\ B &\equiv -\partial_t^2 + W'''W' + (W'')^2 = (-\partial_t - W'')(\partial_t - W'') = \overline{F}F, \end{aligned} \tag{4.23}$$

since the second term on the second line of (4.22) vanishes by the equations of motion. We saw in the last lecture how to evaluate Gaussian (quadratic) path integrals. The fermions will give $\det F$ while the bosons give the usual $(\det B)^{-1/2}$. Actually, supersymmetry implies these cancel exactly—this is the famous cancellation between fermion and boson loop contributions. Indeed, if ξ is an eigenfunction of B with eigenvalue λ, then $\eta \equiv F\xi/\sqrt{\lambda}$ is proportional to an eigenfunction of F (in the fermionic sense) with the square root of the same eigenvalue, for

$$\begin{aligned} \overline{F}\eta &= \sqrt{\lambda}\xi, \\ F\xi &= \sqrt{\lambda}\eta, \end{aligned} \tag{4.24}$$

by (4.23). Thus, for every eigenfunction with eigenvalue λ of B there exists an eigenfunction with eigenvalue $\sqrt{\lambda}$ of F. This implies that the bosonic determinant in the denominator exactly cancels the fermion determinant in the numerator.

Thus, the path integral gives

$$\int \mathcal{D}X e^{-S}\overline{Q}(t_0) \approx e^{-\Delta W/\hbar}\left(\overline{Q}(t_0)\Big|_{cl} + \mathcal{O}(\hbar)\right). \tag{4.25}$$

Here we have simply evaluated the $\overline{Q}$ insertion at the saddle point; there will be corrections to this from correlations between the fields in $\overline{Q}$ and the fluctuations around the saddle point, giving rise to the $\mathcal{O}(\hbar)$ contribution.

4.4. Zero modes

Actually, this result is not quite right. First, we did not expect a t_0 dependence in the answer. Second, $\overline{Q}$ is fermionic, and we expected a complex number, not a Grassmann one as an answer. Finally, the cancellation of the determinants was too quick since $\det B = \det F = 0$—they both have zero modes.

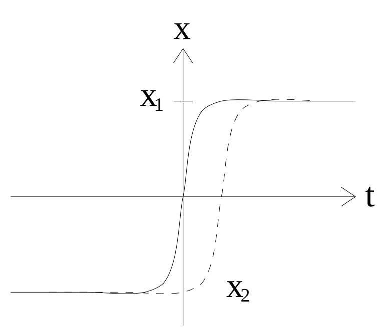
These zero modes are easily identified, since

$$F\dot{x}_{cl} = (\partial_t - W'')\dot{x}_{cl} = (\partial_t - W'')W' = W''\dot{x}_{cl} - W''W' = 0. \tag{4.26}$$

So $B = \overline{F}F$ has a bosonic zero mode $\delta x = \epsilon_0 \dot{x}_{cl}$, where ϵ_0 is any real constant, and F has a fermionic zero mode $\delta\overline{\psi} = \overline{\eta}_0 \dot{x}_{cl}$, where $\overline{\eta}_0$ is a constant Grassmann number. Note that there is no $\delta\psi$ zero mode since there is no bounded solution of $\overline{F}\delta\psi = (\partial_t + W'')\delta\psi = 0$.

What is the interpretation of these zero modes? The bosonic zero mode has a clear interpretation—time translations of x_{cl} are also solutions of the bounce equations:

$$x_{cl}(t) \to x_{cl}(t+\epsilon_0) = x_{cl}(t) + \epsilon_0 \dot{x}_{cl}(t) + \mathcal{O}(\epsilon_0^2). \tag{4.27}$$



The fermionic zero mode can be interpreted by going to a more manifestly supersymmetric formulation: our specific classical instanton solution can be combined into a superfield instanton solution $X_{cl}(t,\theta,\overline{\theta}) = x_{cl}(t) - \overline{\theta}\overline{\eta}_0 \dot{x}_{cl}(t)$. We then have

$$\begin{aligned} &[H, X_{cl}] \neq 0, \\ &[\overline{Q}, X_{cl}] \neq 0, \qquad \text{but} \\ &[Q, X_{cl}] = 0, \end{aligned} \tag{4.28}$$

and have zero modes associated with the broken symmetries: one bosonic zero mode associated with the broken time translation, and one fermionic zero mode associated with the broken $\overline{Q}$ supersymmetry. Since the Q supersymmetry generator annihilates the instanton, we say the instanton background preserves half the supersymmetries.

How does one deal with these zero modes? The bosonic zero mode is associated with a *collective coordinate* of the instanton: the "center of time" t_1 of the bounce. Clearly, physically, we should integrate over all such times for the bounce to take place. This infinite time integral is the interpretation of the infinity (the zero in $\det B$ which appears in the denominator) in the bosonic path integral. We saw in the last lecture how to deal with fermionic zero-modes: insert a fermionic operator into the path integral to absorb it. Of course, we have already done that, with the $\overline{Q}$ insertion.

So, finally, we have reduced the path integral to the zero mode integrations:

$$\begin{aligned}\epsilon &\sim \int d\overline{\eta}_0 \int_{-T/2}^{T/2} dt_1 e^{-\Delta W/\hbar} \left(\overline{Q}(t_0)\Big|_{cl} + \mathcal{O}(\hbar) \right) \\ &= e^{-\Delta W/\hbar} \int d\overline{\eta}_0 \int_{-T/2}^{T/2} dt_1 \overline{\psi}(\dot{x} - W')\Big|_{cl} \\ &= e^{-\Delta W/\hbar} \int d\overline{\eta}_0 \int_{-T/2}^{T/2} dt_1 \overline{\eta}_0 \dot{x}_{cl}(t_1; t_0)(\dot{x}_{cl} - W') \\ &= e^{-\Delta W/\hbar} \int_{x_1}^{x_2} dx (W' - W') \\ &= [0 + \mathcal{O}(\hbar)] e^{-\Delta W/\hbar}.\end{aligned} \tag{4.29}$$

Here t_0 is the time of the $\overline{Q}$ insertion, t_1 is the center of time of the instanton, and in the third step the t_1 integral is traded for an x integral using the fact that $x_{cl}(t_1; t_0) = x_{cl}(t_1 - t_0)$. It is unfortunate that to find the finite, non-zero answer expected physically, one has to go to $\mathcal{O}(\hbar)$ (one loop) in perturbation theory around the instanton background. We will not do this thankless calculation here—see the Salomonson and van Holten paper. However, we do see how including the zero modes removed all the difficulties of our previous (incorrect) answer.

Thus, putting it all together, we obtain

$$\epsilon \sim e^{-\Delta W/\hbar} \qquad \Rightarrow \qquad E_0 \sim e^{-2\Delta W/\hbar} \neq 0. \tag{4.30}$$

This reproduces the energy-lifting that can be computed more easily using the exact solution of the supersymmetric QM Schrödinger equation given in lecture 2.

This ends the pedagogical introduction to the essential physics of supersymmetry through the toy model of supersymmetric QM. The main points that turn out to generalize to 3+1 dimensions are:

- that the supersymmetry algebra determines a natural zero of the energy(-density), and that the order parameter for spontaneous supersymmetry breaking is a non-zero vacuum energy(-density);
- that low-energy (vacuum) properties of supersymmetric systems can be effectively computed by analytic continuation in appropriate parameters in the microscopic theory;
- that by extending space-time to include formal Grassmann parameters—superspace—one can find "classical" linear realizations of the supersymmetry action on superfields, providing an efficient way of constructing supersymmetric actions, and a compact notation in which the correlated interactions of a supersymmetric system are summarized in a prepotential—the "superpotential" of our QM example.

Also, we've introduced the technology for treating fermionic zero modes.

Finally, though I have been treating supersymmetric QM as a toy model, it has a mathematical interest in its own right. It has turned out to be an effective and intuitive way of proving various "index" theorems about differential operators on manifolds, and related subjects (Morse theory), and, surprisingly, this also has a generalization to higher dimensions (*via* Witten's "twisting" procedure).

II. Perturbative Supersymmetry

5. Representations of the Lorentz Group and Supersymmetry Algebra

In this lecture we properly begin our main topic—N=1 supersymmetry in four dimensions. We start, of course, with the basic field and particle representations of the supersymmetry algebra. One of the main technical difficulties of 4 dimensions compared to 1 dimension (susy QM) is simply the complication of the representation theory of the Lorentz group. In fact, this lecture will mostly be a quick review of the particle and field representations of the Poincaré and Lorentz groups without reference to supersymmetry. We follow the notation of Wess and Bagger.

5.1. Poincaré Group—Particles

Particles (states) must transform in unitary representations of the 3+1 dimensional Poincaré group, which, since it is not compact, are all infinite dimensional. This infinite dimensionality is simply the familiar fact that particle states are labelled by the continuous parameters p_μ—their four-momenta. Such representations can be organized using a basic trick invented by Wigner, the so-called *little group*—the group of (usually compact) transformations left after fixing some of the non-compact transformations in some conventional way.

In the present case, the non-compact part of the Lorentz group is the boosts. For massive particles, we can boost to a frame in which the particle is at rest

$$P^\mu = (m, 0, 0, 0). \tag{5.1}$$

The little group in this case is just those Lorentz transformations which preserve this four-vector—that is $SO(3)$, the group of rotations. Thus massive particles are in representations of $SO(3)$, labelled by the spin $j \in \frac{1}{2}\mathbb{Z}$ of the $(2j+1)$-dimensional representation

$$|j, j_3\rangle, \qquad -j \le j_3 \le j. \tag{5.2}$$

We have derived the familiar fact that a massive particle is described by its mass and spin (as well as any internal quantum numbers).

Massless states are classified similarly. Here we can boost to

$$P^\mu = (E, E, 0, 0), \tag{5.3}$$

(for some conventional value of E) which is preserved by $SO(2)$ rotations around the z-axis.[3] Representations of $SO(2)$ are one-dimensional, labelled by a single eigenvalue, the helicity:

$$|\lambda\rangle, \tag{5.4}$$

which physically measures the component of angular momentum along the direction of motion. Algebraically λ could be any real number, but there is a topological constraint. Since the helicity is the eigenvalue of the rotation generator around the z-axis, a rotation by an angle θ around that axis produces a phase $e^{i\theta\lambda}$ on wave functions. Now, the Lorentz group is isomorphic[4] to $SL(2,\mathbb{C})/\mathbb{Z}_2$ which is topologically $\mathbb{R}_3 \times S_3/\mathbb{Z}_2$. (The $\mathbb{R}_3$ is the non-compact part corresponding to the boosts, while the doubly-connected $S_3/\mathbb{Z}_2$ corresponds to the rotations.) Thus, though a 2π rotation cannot be continuously deformed to the identity, a 4π rotation can. This implies that the phase $e^{4\pi i\lambda}$ must be one, giving the quantization of the helicity:

$$\lambda \in \tfrac{1}{2}\mathbb{Z}. \tag{5.6}$$

5.2. Lorentz Group—Fields

The Lorentz group is $SO(3,1)$ in Minkowski space. We take Minkowski space to have signature $\eta_{\mu\nu} = \mathrm{diag}(-+++)$. As a group, $SO(3,1) \simeq SL(2,\mathbb{C})/\mathbb{Z}_2$, the group of complex 2×2 matrices of determinant one (modded out by a global identification).

We will also be interested in Euclidean four-space. First, it provides a convenient classification of the Lorentz group representations, and second, later on in the course we will be performing instanton calculations in Euclidean space. In Euclidean space, the Lorentz group is $SO(4)$, which is compact and isomorphic as an algebra[5] to

$$SO(4) \underset{alg}{\simeq} SU(2)_L \times SU(2)_R \tag{5.7}$$

[3] Actually, the little group preserving P^μ is isomorphic to the non-compact group of Euclidean motions on the plane—$SO(2)$ plus two "translations". However, being a non-compact group itself, this little group's unitary representations are infinite-dimensional, except when the eigenvalues of the "translations" are zero, in which case it effectively reduces to $SO(2)$. The infinite-dimensional representations are considered unphysical because we never see particle states in nature labelled by extra continuous parameters.

[4] The reason for the $\mathbb{Z}_2$ is that $SL(2,\mathbb{C})$ is by itself a double-cover of the Lorentz group. This is easy to see:

$$M(\theta) = \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix} \tag{5.5}$$

corresponds to a Lorentz transformation producing a rotation by an angle θ about the z-axis. Hence $M = -1$ produces a rotation by 2π, which is the identity in the Lorentz group.

[5] As a group, $SU(2)_L \times SU(2)_R \underset{grp}{\simeq} Spin(4)$, where $Spin(4)$ is a double-cover of $SO(4)$ as a group (it has an extra $\mathbb{Z}_2$).

The $SO(3)$ of rotations in Minkowski space is the diagonal subgroup of $SU(2)_L \times SU(2)_R$. More concretely, if J_j are the generators of rotations, and K_j generate boosts, then $SU(2)_{L,R}$ are generated by $J_j \pm iK_j$.

Unlike particles, fields are classified by the finite-dimensional representations of the Lorentz group. These representations are conveniently labelled by $(j_L, j_R) \in SU(2)_L \times SU(2)_R$ where $j_{L,R}$ are two $SU(2)$ spins. The main examples which will interest us are:

scalar	ϕ	$(0,0)$	
left-handed spinor	ψ_α	$(\frac{1}{2},0)$	$\alpha = 1,2$ $SU(2)_L$ index
right-handed spinor	$\overline{\psi}_{\dot\alpha}$	$(0,\frac{1}{2})$	$\dot\alpha = 1,2$ $SU(2)_R$ index
vector	$A_{\alpha\dot\alpha}$	$(\frac{1}{2},\frac{1}{2})$	
self-dual 2-form	$F^+_{\alpha\beta}$	$(1,0)$	α,β symmetric
anti-s.d. 2-form	$F^-_{\dot\alpha\dot\beta}$	$(0,1)$	$\dot\alpha,\dot\beta$ symmetric

Because we have labelled the fields by representations of the Euclidean group instead of the Minkowski group (which is what we're really interested in), we have to pay attention to some differences in the way these representations behave under complex conjugation, parity, and time reversal. In Euclidean space all these representations are real (*e.g.* $(F^+)^* = F^+$) or pseudo real (*e.g.* $(\psi_\alpha)^* = \epsilon_{\alpha\beta}\psi_\beta$). But in Minkowski space, complex conjugation interchanges $SU(2)_L$ with $SU(2)_R$. So, for example,

$$(\psi_\alpha)^* = \overline{\psi}_{\dot\alpha}, \qquad (F^+)^* = F^-. \tag{5.8}$$

Parity also acts differently in Euclidean versus Minkowski signatures. In Euclidean space, parity

$$P: \qquad x_\mu \to (x_0, -x_1, -x_2, -x_3), \tag{5.9}$$

is an outer automorphism of $SO(4)$—it exchanges $SU(2)_L \leftrightarrow SU(2)_R$. When added to the rotation group, $SO(4)$ becomes $O(4)$. On the other hand, parity *plus* time reversal

$$PT: \qquad x_\mu \to -x_\mu, \tag{5.10}$$

is trivial—PT is an element of $SO(4)$. In Minkowski space-time, neither P nor PT are elements of $SO(3,1)$.

To go into more detail on the spinor representations (in Minkowski space-time), recall that $SO(3,1) \simeq SL(2,\mathbb{C})/\mathbb{Z}_2$. So a natural two-dimensional representation of the Lorentz group is in terms of 2×2 complex determinant-one matrices M. But, for a given such representation, there are actually four related representations

$$M, {}^tM^{-1}, M^*, {}^tM^{*-1}, \tag{5.11}$$

under which spinors in Minkowski space transform as

$$\begin{aligned}
{\psi_\alpha}' &= M_\alpha^\beta \psi_\beta \\
\psi^{\alpha\prime} &= (M^{-1})^\alpha_\beta \psi^\beta \\
{\overline{\psi}_{\dot\alpha}}' &= (M^*)_{\dot\alpha}^{\dot\beta} \overline{\psi}_{\dot\beta} \\
\overline{\psi}^{\dot\alpha\prime} &= (M^{*-1})^{\dot\alpha}_{\dot\beta} \overline{\psi}^{\dot\beta}.
\end{aligned} \tag{5.12}$$

The first two are $(\frac{1}{2},0)$ spinors, while the second two are $(0,\frac{1}{2})$ spinors.

The reason M and ${}^tM^{-1}$ give rise to the same representation is because there exists an $SL(2)$–invariant tensor $\epsilon_{\alpha\beta} = \epsilon_{\dot\alpha\dot\beta}$ which conjugates them. Here ϵ is the antisymmetric tensor on two indices defined by

$$\epsilon_{12} = \epsilon^{21} = -1, \qquad \epsilon_{21} = \epsilon^{12} = +1. \tag{5.13}$$

▸ **Exercise 5.1.** Show $\epsilon_{\alpha\beta}\epsilon^{\gamma\delta} = -\delta_\alpha^\gamma\delta_\beta^\delta + \delta_\alpha^\delta\delta_\beta^\gamma$. So, in particular, $\epsilon_{\alpha\beta}\epsilon^{\beta\delta} = \epsilon^{\delta\beta}\epsilon_{\beta\alpha} = \delta_\alpha^\delta$.

The basic relations are

$$\begin{aligned}
\psi^\alpha &= \epsilon^{\alpha\beta}\psi_\beta, \qquad (\text{or, } \psi_\alpha = \epsilon_{\alpha\beta}\psi^\beta\ ) \\
\overline{\psi}^{\dot\alpha} &= \epsilon^{\dot\alpha\dot\beta}\overline{\psi}_{\dot\beta}, \qquad (\text{or, } \overline{\psi}_{\dot\alpha} = \epsilon_{\dot\alpha\dot\beta}\overline{\psi}^{\dot\beta}\ ).
\end{aligned} \tag{5.14}$$

The second actually follows from the first, since $\overline{\psi}_{\dot\alpha} = (\psi_\alpha)^*$.

To reduce the number of indices we write, we introduce some important conventions for contracting spinor indices:

$$\begin{aligned}
\psi\xi \equiv \psi^\alpha\xi_\alpha &= \psi_\beta\epsilon^{\beta\alpha}\xi_\alpha = -\psi_\alpha\xi^\alpha = +\xi^\alpha\psi_\alpha = \xi\psi \\
\overline{\psi\xi} \equiv \overline{\psi}_{\dot\alpha}\overline{\xi}^{\dot\alpha} &= \qquad\qquad \cdots \qquad\qquad = \overline{\xi\psi}.
\end{aligned} \tag{5.15}$$

This gives the scalar formed from two same-handed spinors. Group-theoretically, left-handed spinors are in $(\frac{1}{2},0)$ representations, so the product of two of them is

$$(\tfrac{1}{2},0)\otimes(\tfrac{1}{2},0) = (0,0)\oplus(1,0), \tag{5.16}$$

by the usual addition of angular momentum. The scalar piece, as usual, is formed from the antisymmetric product, as shown above.

▸ **Exercise 5.2.** Show that $(\psi\xi)^* = \overline{\xi\psi}$.

The product of a left-handed with a right-handed spinor, on the other hand, gives a vector:

$$(\tfrac{1}{2},0)\otimes(0,\tfrac{1}{2}) = (\tfrac{1}{2},\tfrac{1}{2}). \tag{5.17}$$

However, we usually write vectors in a tensor notation, with space-time, not spinor, indices. To connect this 2-component spinor notation to tensor notation, we introduce the $\sigma^{\mu}_{\alpha\dot{\alpha}}$ matrices. The σ^{μ} are not matrices of a representation, they are Clebsch-Gordon coefficients—a dictionary between the spinor and vector representations:

$$\sigma^0 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{5.18}$$

Note that the complex conjugate of the σ^{μ} are just their transpose. We also define the conjugate $\overline{\sigma}$-matrices by

$$\overline{\sigma}^{\mu\dot{\alpha}\alpha} \equiv \epsilon^{\dot{\alpha}\dot{\beta}}\epsilon^{\alpha\beta}\sigma^{\mu}_{\beta\dot{\beta}}. \tag{5.19}$$

Then the combination $\psi\sigma^{\mu}\overline{\chi}$ is the $(\frac{1}{2},\frac{1}{2})$ representation in a vector notation, and $\psi\sigma^{\mu}\overline{\sigma}^{\nu}\chi$ is the $(1,0)$ piece of the $(\frac{1}{2},0)\otimes(\frac{1}{2},0)$ in tensor notation.

- ▸ **Exercise 5.3.** Show that $(\psi\sigma^{\mu}\overline{\chi})^* = \chi\sigma^{\mu}\overline{\psi}$, and $(\psi\sigma^{\mu}\overline{\sigma}^{\nu}\chi)^* = \overline{\chi}\overline{\sigma}^{\nu}\sigma^{\mu}\overline{\psi}$.
- ▸ **Exercise 5.4.** Show that $\psi\sigma^{\mu}\overline{\chi} = -\overline{\chi}\overline{\sigma}^{\mu}\psi$ and $\psi\sigma^{\mu}\overline{\sigma}^{\nu}\chi = \chi\sigma^{\nu}\overline{\sigma}^{\mu}\psi$.

The sigma matrices are used to translate between spinor and tensor notation. The basic relation is

$$P_{\alpha\dot{\alpha}} \equiv P_{\mu}\sigma^{\mu}_{\alpha\dot{\alpha}} = \begin{pmatrix} -P_0 + P_3 & P_1 - iP_2 \\ P_1 + iP_2 & -P_0 - P_3 \end{pmatrix}, \tag{5.20}$$

so that under Lorentz transformations $P'_{\alpha\dot{\alpha}} = M_{\alpha}^{\beta}P_{\beta\dot{\beta}}(M^+)^{\dot{\beta}}_{\dot{\alpha}}$. Note that this implies that $\det P = -P^{\mu}P_{\mu} = m^2$ is invariant under Lorentz transformations.

A general $SL(2,\mathbb{C})$ representation with (Euclidean) "spins" (j_L, j_R) can be written

$$X_{\alpha_1\cdots\alpha_{2j_L},\dot{\alpha}_1\cdots\dot{\alpha}_{2j_R}} \tag{5.21}$$

where the undotted and dotted indices are separately symmetrized. For example, $F_{\mu\nu} = [(\frac{1}{2},\frac{1}{2})\otimes(\frac{1}{2},\frac{1}{2})]_A = (0,1)+(1,0) = F^-_{\dot{\alpha}\dot{\beta}} + F^+_{\alpha\beta}$.

- ▸ **Exercise 5.5.** Write the metric tensor $g_{\mu\nu}$ in spinor notation.
- ▸ **Exercise 5.6.** Write $\epsilon_{\mu\nu\rho\sigma}$ in spinor notation.
- ▸ **Exercise 5.7.** Define the dual electromagnetic field strength by $({}^*F)_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$. Show $F^{\pm}_{\mu\nu} = F_{\mu\nu} \pm i({}^*F)_{\mu\nu}$. Write $F^{\pm}$ in terms of $\mathbf{E}$ and $\mathbf{B}$.
- ▸ **Exercise 5.8.** Show that the Lorentz generators in the $(\frac{1}{2},0)$ representation are given by $\sigma_{\alpha}^{[\mu\nu]\beta} = \frac{1}{4}(\sigma^{\mu}_{\alpha\dot{\alpha}}\overline{\sigma}^{\nu\dot{\alpha}\beta} - \sigma^{\nu}_{\alpha\dot{\alpha}}\overline{\sigma}^{\mu\dot{\alpha}\beta})$. Show that $\psi\sigma^{\mu\nu}\psi = 0$ (hint: use exercise 5.4).

There are a set of useful identities, called Fierz identities, for rearranging the order of spinors in products. The simplest ones are written below.

▸ **Exercise 5.9.** Using the $\epsilon_{\alpha\beta}$ completeness relation of exercise 5.1, show:

$$\begin{aligned}
0 &= (\psi_1\psi_2)(\psi_3\psi_4) + (\psi_1\psi_3)(\psi_2\psi_4) + (\psi_1\psi_4)(\psi_2\psi_3),\\
0 &= (\psi_1\psi_2)(\psi_3\sigma^\mu\overline{\psi}_4) + (\psi_1\psi_3)(\psi_2\sigma^\mu\overline{\psi}_4) + (\psi_1\sigma^\mu\overline{\psi}_4)(\psi_2\psi_3),\\
0 &= (\psi_1\sigma^\mu\overline{\psi}_2)(\psi_3\sigma^\nu\overline{\psi}_4) - (\psi_1\psi_3)(\overline{\psi}_2\overline{\sigma}^\mu\sigma^\nu\overline{\psi}_4) - (\psi_1\sigma^\nu\overline{\psi}_4)(\overline{\psi}_2\overline{\sigma}^\mu\psi_3).
\end{aligned} \tag{5.22}$$

(Hint: the second two follow with very little work from the first.)

▸ **Exercise 5.10.** Derive the completeness relations for the σ's:

$$\mathrm{Tr}\sigma^\mu\overline{\sigma}^\nu = -2\eta^{\mu\nu}, \qquad \sigma^\mu_{\alpha\dot\alpha}\overline{\sigma}^{\dot\beta\beta}_\mu = -2\delta^\beta_\alpha\delta^{\dot\beta}_{\dot\alpha}. \tag{5.23}$$

Show also the useful identities:

$$(\sigma^\mu\overline{\sigma}^\nu + \sigma^\nu\overline{\sigma}^\mu)^\beta_\alpha = -2\eta^{\mu\nu}\delta^\beta_\alpha, \qquad \sigma^\mu\overline{\sigma}^\nu = -\eta^{\mu\nu} + 2\sigma^{\mu\nu}. \tag{5.24}$$

These latter identities can be used to derive many more complicated Fierz identities.

Finally, we should mention the relation between our 2-component (Weyl) spinor notation and the 4-component notation for Dirac spinors:

$$\Psi = \begin{pmatrix} \chi_\alpha \\ \overline{\psi}^{\dot\alpha} \end{pmatrix}, \qquad \gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \overline{\sigma}^\mu & 0 \end{pmatrix}, \qquad \gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \tag{5.25}$$

This is the Weyl basis for the gamma matrices. Majorana spinors in this basis are Dirac spinors with the constraint that $\psi = \chi$. I will use the 2-spinor notation in this course. This is usually advantageous, since 2-spinors are the irreducible representations of the Lorentz group while Dirac spinors are reducible. Thus group-theoretic (symmetry) arguments are generally clearer in the 2-spinor language.

5.3. Particle representations of the supersymmetry algebra

We can now write down the supersymmetry algebra in 4 dimensions:

$$\{Q_\alpha, \overline{Q}_{\dot\alpha}\} = 2\sigma^\mu_{\alpha\dot\alpha}P_\mu = 2P_{\alpha\dot\alpha}, \qquad \{Q_\alpha, Q_\beta\} = 0. \tag{5.26}$$

This defines the normalization of the supersymmetry generators. The uniqueness of this algebra was discussed in the first lecture.

Boost massive particle states to their rest frame: $P_\mu = (-m, 0, 0, 0)$. Call this state $|\Omega\rangle$. Then, acting on this state, the supersymmetry algebra becomes

$$\{Q_\alpha, \overline{Q}_{\dot\alpha}\} = 2m\delta_{\alpha\dot\alpha}, \qquad \{Q_\alpha, Q_\beta\} = 0. \tag{5.27}$$

If we define the spinor charges

$$a_\alpha \equiv \frac{1}{\sqrt{2m}} Q_\alpha, \qquad \overline{a}_{\dot\alpha} \equiv \frac{1}{\sqrt{2m}} \overline{Q}_{\dot\alpha}, \tag{5.28}$$

then

$$\{\overline{a}_{\dot\alpha}, a_\beta\} = \delta_{\dot\alpha\beta}, \qquad \{a_\alpha, a_\beta\} = 0. \tag{5.29}$$

The representations of this two-dimensional Clifford algebra are easy to consruct, since this is just the algebra of creation and annihiliation operators. Say a_α annihilates $|\Omega\rangle$, then we find a four-dimensional representation:

$$|\Omega\rangle, \quad \overline{a}_{\dot\alpha}|\Omega\rangle, \quad \overline{a}_1\overline{a}_2|\Omega\rangle. \tag{5.30}$$

To better understand the particle content of this representation, note that if $|\Omega\rangle$ has spin j, then $\overline{a}_1\overline{a}_2|\Omega\rangle$ also has spin j, while $\overline{a}_{\dot\alpha}|\Omega\rangle$ has spins $j+\frac{1}{2}$ and $j-\frac{1}{2}$ for $j \neq 0$ (for $j = 0$ they have only $j = \frac{1}{2}$). So, explicitly, the spin content of a massive spinless supersymmetry multiplet is

$$j = 0, 0, \tfrac{1}{2}, \tag{5.31}$$

while for a massive spinning multiplet, it is

$$j{-}\tfrac{1}{2}, j, j, j{+}\tfrac{1}{2}. \tag{5.32}$$

You can check that such multiplets have equal numbers of bosonic and fermionic (propagating) degrees of freedom.

For massless particles, we boost to the frame where the four-momentum is $P_\mu = (-E, E, 0, 0)$, and denote the state by $|\Omega\rangle$. The supersymmetry algebra is then

$$\{Q_\alpha, \overline{Q}_{\dot\alpha}\} = 4E \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{5.33}$$

Defining

$$a_\alpha \equiv \frac{1}{2\sqrt{E}} Q_\alpha, \qquad \overline{a}_\alpha \equiv \frac{1}{2\sqrt{E}} \overline{Q}_{\dot\alpha}, \tag{5.34}$$

gives the algebra

$$\{a_1, \overline{a}_1\} = 1, \qquad \text{all others} = 0. \tag{5.35}$$

This implies that $a_2 = \overline{a}_2 = 0$ on all representations. Thus the massless supersymmetry multiplets are just two-dimensional:

$$(a_1|\Omega\rangle = 0): \qquad |\Omega\rangle \quad , \quad \overline{a}_1|\Omega\rangle. \tag{5.36}$$

If $|\Omega\rangle$ has helicity λ, then $\overline{a}_1|\Omega\rangle$ has helicity $\lambda+\frac{1}{2}$. (By CPT invariance, such a multiplet will always appear in a field theory with its opposite helicity multiplet $(-\lambda\,,\,-\lambda-\frac{1}{2})$.)

We will only concern ourselves with a few of these representations in this course. For massless particles, we will be interested in the "chiral multiplet" with helicities $\lambda = \{-\frac{1}{2}, 0, 0, \frac{1}{2}\}$, corresponding to the degrees of freedom associated with a complex scalar and a Weyl fermion: $\{\phi, \psi_\alpha\}$; and the "vector multiplet" with helicities $\lambda = \{-1, -\frac{1}{2}, \frac{1}{2}, 1\}$, corresponding to the degrees of freedom associated with a Weyl fermion and a vector boson: $\{\lambda_\alpha, A_\mu\}$. Other massless supersymmetry multiplets contain fields with spin 3/2 or greater. The only known consistent (classical) couplings for such fields occur in supergravity and gravity theories.

In QFT (as opposed to gravity) the chiral multiplets are the supersymmetric analog of matter fields, while the vector multiplets are the analog of the gauge fields. So, if we think of the fermions in the chiral multiplets as quarks, their scalar superpartners are given the name "squarks". Similarly, the fermionic superpartner of the gauge bosons are called "gauginos".

For massive particle multiplets, we have the massive chiral multiplet with spins $j = \{0, 0, \frac{1}{2}\}$, corresponding to a massive complex scalar and Weyl fermion field: $\{\phi, \psi_\alpha\}$; and a massive vector multiplet with $j = \{0, \frac{1}{2}, \frac{1}{2}, 1\}$ with massive field content $\{h, \psi_\alpha, \lambda_\alpha, A_\mu\}$, where h is a real scalar field. In terms of degrees of freedom, it is clear that the massive vector multiplet has the same counting as a massless chiral plus a massless vector multiplet. This is indeed the case dynamically: massive vector multiplets arise by a supersymmetric analog of the Higgs mechanism.

6. N=1 Superspace and Chiral Superfields

6.1. Superspace

Fields form representations of the supersymmetry algebra which are most conveniently handled using superspace. One should note that superspace works beautifully for N=1 supersymmetry in 4 dimensions and "lower" supersymmetry, but not for "higher supersymmetries"—extended supersymmetries in 4 dimensions or any supersymmetry in higher dimensions. I will treat superspace essentially as a convenient trick, and will feel free to use components whenever it is easier to do so.

We extend space-time by including Grassmann spinor coordinates (one for each spinor supercharge):

$$x^\mu \to (x^\mu, \theta_\alpha, \overline{\theta}_{\dot\alpha}). \tag{6.1}$$

Supertranslations are defined to be

$$(x^\mu, \theta_\alpha, \overline{\theta}_{\dot\alpha}) \to (x^\mu + i\theta\sigma^\mu\overline{\xi} - i\xi\sigma^\mu\overline{\theta}\,,\ \theta_\alpha + \xi_\alpha\,,\ \overline{\theta}_{\dot\alpha} + \overline{\xi}_{\dot\alpha}). \tag{6.2}$$

The differential operators on superspace which generate supertranslations—$\delta f = (\xi Q + \overline{\xi Q})f$—are

$$\begin{aligned} Q_\alpha &= +\frac{\partial}{\partial\theta^\alpha} - i\sigma^\mu_{\alpha\dot\alpha}\overline{\theta}^{\dot\alpha}\partial_\mu \\ \overline{Q}_{\dot\alpha} &= -\frac{\partial}{\partial\overline{\theta}^{\dot\alpha}} + i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu. \end{aligned} \tag{6.3}$$

▸ **Exercise 6.1.** Check that Q and $\overline{Q}$ as differential operators on superspace satisfy the usual supersymmetry algebra $\{Q_\alpha, Q_\beta\} = 0$ and $\{Q_\alpha, \overline{Q}_{\dot\alpha}\} = 2P_{\alpha\dot\alpha}$.

We also define the covariant derivatives

$$\begin{aligned} D_\alpha &= +\frac{\partial}{\partial\theta^\alpha} + i\sigma^\mu_{\alpha\dot\alpha}\overline{\theta}^{\dot\alpha}\partial_\mu \\ \overline{D}_{\dot\alpha} &= -\frac{\partial}{\partial\overline{\theta}^{\dot\alpha}} - i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu, \end{aligned} \tag{6.4}$$

which differ from the Q's by a relative sign, and satisfy the supersymmetry algebra with the wrong sign: $\{D_\alpha, \overline{D}_{\dot\alpha}\} = -2P_{\alpha\dot\alpha}$, and the covariant derivatives anticommute with the supersymmetry generators,

$$\{D_\alpha, Q_\beta\} = \{\overline{D}_{\dot\alpha}, Q_\beta\} = 0, \tag{6.5}$$

just as in the supersymmetric QM case.

6.2. Chiral Superfields

Superfields Φ are then just functions on superspace, which, by definition, are supersymmetry covariant. To make supersymmetry covariant objects out of them, one can then add or multiply them (since $[Q, \Phi_1\Phi_2] = [Q, \Phi_1]\Phi_2 + \Phi_1[Q, \Phi_2]$ using the Leibnitz rule for Grassmann differentiation), or act on them with space-time or covariant derivatives (since Q commutes with them). The most general superfield is

$$\begin{aligned} \Phi(x, \theta, \overline{\theta}) = \phi + \theta\psi + \overline{\theta}\overline{\chi} + \theta^2 F + \overline{\theta}^2 G + \theta\sigma^\mu\overline{\theta}A_\mu \\ + \theta^2\overline{\theta\lambda} + \overline{\theta}^2\theta\rho + \theta^2\overline{\theta}^2 D. \end{aligned} \tag{6.6}$$

This has many component fields—so many, in fact, that it gives a reducible representation of the supersymmetry algebra. This is easy to see, for if all the fields were propagating and ϕ had spin j (assuming it is massive), then there are component fields with spins j, $j \pm \frac{1}{2}$, and $j \pm 1$, which is larger than the irreducible supersymmetric particle multiplets found in the last lecture. To get an irreducible field representation we must impose a constraint on the superfield which (anti)commutes with the supersymmetry algebra. One such constraint is simply a reality condition, which turns out to lead to a vector multiplet—we will return to this representation later.

Another constraint we can impose is the so-called *chiral superfield* (χsf) constraint:

$$\overline{D}_{\dot{\alpha}}\Phi = 0. \tag{6.7}$$

This is consistent since $\overline{D}$ anticommutes with Q and $\overline{Q}$, implying in particular that if Φ is a χsf, then $Q_\alpha\Phi$ and $\overline{Q}_{\dot{\alpha}}\Phi$ are too. We can show that this gives rise to an irreducible representation of the superalgebra by simply solving the constraint.

▸ **Exercise 6.2.** Show that θ and $y^\mu \equiv x^\mu + i\theta\sigma^\mu\overline{\theta}$ are both annihilated by $\overline{D}$.

Thus the general χsf is

$$\begin{aligned}\Phi = \Phi(y,\theta) &= \phi(y) + \sqrt{2}\theta\psi(y) + \theta^2 F(y)\\ &= \phi + i\theta\sigma^\mu\overline{\theta}\partial_\mu\phi + \frac{1}{4}\theta^2\overline{\theta}^2\partial^2\phi + \sqrt{2}\theta\psi - \frac{i}{\sqrt{2}}\theta^2\partial_\mu\psi\sigma^\mu\overline{\theta} + \theta^2 F.\end{aligned} \tag{6.8}$$

Now there are only fields with spins j and $j \pm \frac{1}{2}$, consistent with an irreducible supersymmetry multiplet. (It will turn out that the F field will always be non-propagating.) Also, we will see that though this constraint implies differential relations among the component fields, they will not give rise to higher-derivative actions (non-standard kinetic terms).

Anti-chiral superfields ($\overline{\chi}$sf) can be defined in an analogous manner:

$$D_\alpha\Phi = 0. \tag{6.9}$$

This constraint can be solved in the same way, except now the anti-chiral coordinates are $\overline{\theta}$ and $\overline{y}^\mu \equiv x^\mu - i\theta\sigma^\mu\overline{\theta}$. Note that (in Minkowski space-time) $(D)^* = \overline{D}$, so if Φ is a χsf, then $\overline{\Phi}$ is an $\overline{\chi}$sf.

If Φ_i are χsf's, then $\Phi_1 + \Phi_2$ and $\Phi_1\Phi_2$ are also χsf's. Similarly for $\overline{\chi}$sf's. However, mixed objects such as $\Phi\overline{\Phi}$ are neither χsf's nor $\overline{\chi}$sf's. Note that a chiral covariant derivative such as $D\Phi$ is not a χsf—it is an $\overline{\chi}$sf!

6.3. Lagrangians—Kahler Potential

For ease of notation, we shall write

$$d^4\theta = d^2\theta d^2\overline{\theta}. \tag{6.10}$$

Then a supersymmetric-invariant Lagrangian can be written as

$$\mathcal{L} = \int d^4\theta\, \mathcal{K}(\Phi^i, \overline{\Phi}^{\bar{\imath}}, X), \tag{6.11}$$

where X stands for a collection of arbitrary (not necessarily chiral or anti-chiral) superfields. The reason $\mathcal{L}$ is automatically supersymmetry-invariant is the same as in the QM case:

▶ **Exercise 6.3.** Show that when Q acts on a superfield it always gives ∂_μ of something for the highest ($\theta^2\overline{\theta}^2$) component.

Thus, $\delta_Q\mathcal{L}$ is a total derivative.

Note, however, that a term in the Lagrangian built solely from, say, chiral superfields is automatically a total derivative:

$$\int d^4\theta\, f(\Phi) = \partial(\ldots). \tag{6.12}$$

This means that $\mathcal{K}$ is invariant under (only defined up to) the transformations

$$\mathcal{K} \sim \mathcal{K} + f(\Phi) + \overline{f}(\overline{\Phi}). \tag{6.13}$$

The simplest non-trivial example of such a Lagrangian is

$$\mathcal{K} = \overline{\Phi}\Phi, \tag{6.14}$$

which gives rise to the free Lagrangian

$$\begin{aligned}
\mathcal{L} &= \int d^4\theta\, \overline{\Phi}\Phi \\
&= \int d^4\theta (\overline{\phi} - i\theta\sigma^\mu\overline{\theta}\partial_\mu\overline{\phi} + \frac{1}{4}\theta^2\overline{\theta}^2\partial^2\overline{\phi} - \sqrt{2\theta\psi} + \frac{i}{\sqrt{2}}\overline{\theta}^2\theta\sigma^\mu\partial_\mu\overline{\psi} + \overline{\theta}^2\overline{F}) \\
&\quad \cdot (\phi + i\theta\sigma^\mu\overline{\theta}\partial_\mu\phi + \frac{1}{4}\theta^2\overline{\theta}^2\partial^2\phi + \sqrt{2}\theta\psi - \frac{i}{\sqrt{2}}\theta^2\partial_\mu\psi\sigma^\mu\overline{\theta} + \theta^2 F) \\
&= \overline{F}F + \frac{1}{4}\overline{\phi}\partial^2\phi - \frac{1}{2}\eta^{\mu\nu}\partial_\mu\overline{\phi}\partial_\nu\phi + \frac{1}{4}\partial^2\overline{\phi}\phi + \frac{i}{2}\partial_\mu\psi\sigma^\mu\overline{\psi} - \frac{i}{2}\psi\sigma^\mu\partial_\mu\overline{\psi} \\
&= \overline{F}F - \partial_\mu\overline{\phi}\partial^\mu\phi + i\partial_\mu\overline{\psi}\overline{\sigma}^\mu\psi.
\end{aligned} \tag{6.15}$$

This describes a free complex boson and a free Weyl fermion.

We can do a more general case: $\mathcal{K} = \mathcal{K}(\Phi^i, \overline{\Phi}^{\overline{\imath}})$. This is known as the supersymmetric non-linear sigma-model (nlσm), for historic reasons. It's interest lies in the fact that it is the second term in the expansion of the (non-renormalizable) low-energy effective action of a supersymmetric theory of χsf's. We will return to this point later. Define a metric on "field space"

$$g_{i\overline{\imath}} \equiv \partial_i\partial_{\overline{\imath}}\mathcal{K}(\phi, \overline{\phi}), \qquad \text{where } \partial_i \equiv \frac{\partial}{\partial\phi^i}, \ldots \tag{6.16}$$

Then, in the usual way, one defines a Christoffel symbol

$$\Gamma^i_{jk} = g^{i\overline{\imath}}g_{j\overline{\imath},k}, \qquad \Gamma^{\overline{\imath}}_{\overline{\jmath}\overline{k}} = g^{i\overline{\imath}}g_{i\overline{\jmath},\overline{k}}, \tag{6.17}$$

and a Riemann tensor,

$$R_{i\bar{\jmath}k\bar{\ell}} = g_{i\bar{\jmath},k\bar{\ell}} - \Gamma^m_{ik} g_{m\overline{m}} \Gamma^{\overline{m}}_{\bar{\jmath}\bar{\ell}}, \tag{6.18}$$

associated to this metric. In terms of these quantities one finds (this is a good, but optional, exercise)

$$\begin{aligned}\mathcal{L} = \int d^4\theta\, \mathcal{K} = g_{i\bar{\imath}} F^i \overline{F}^{\bar{\imath}} - \tfrac{1}{2} F^i g_{i\bar{\imath}} \Gamma^{\bar{\imath}}_{\bar{\jmath}\bar{k}} \overline{\psi}^{\bar{\jmath}} \overline{\psi}^{\bar{k}} - \tfrac{1}{2} \overline{F}^{\bar{\imath}} g_{i\bar{\imath}} \Gamma^i_{jk} \psi^j \psi^k \\ - g_{i\bar{\imath}} \partial_\mu \phi^i \partial^\mu \overline{\phi}^{\bar{\imath}} - i g_{i\bar{\imath}} \overline{\psi}^{\bar{\imath}} \overline{\sigma}^\mu \mathcal{D}_\mu \psi^i + \tfrac{1}{4} g_{i\bar{\jmath},k\bar{\ell}} \psi^i \psi^k \overline{\psi}^{\bar{\jmath}} \overline{\psi}^{\bar{\ell}}.\end{aligned} \tag{6.19}$$

where

$$\mathcal{D}_\mu \psi^i \equiv \partial_\mu \psi^i + \Gamma^i_{jk} \partial_\mu \phi^k \psi^j. \tag{6.20}$$

The equation of motion of $\overline{F}$ is $g_{i\bar{\imath}} F^i - \frac{1}{2} g_{i\bar{\imath}} \Gamma^i_{jk} \psi^j \psi^k = 0$. For the kinetic term to have the right sign, $g_{i\bar{\imath}}$ must be positive definite, and hence invertible, giving $F^i = \frac{1}{2} \Gamma^i_{jk} \psi^j \psi^k$. Substituting gives

$$\mathcal{L} = -g_{i\bar{\imath}} \partial_\mu \phi^i \partial^\mu \overline{\phi}^{\bar{\imath}} - i g_{i\bar{\imath}} \overline{\psi}^{\bar{\imath}} \overline{\sigma}^\mu \mathcal{D}_\mu \psi^i + \tfrac{1}{4} R_{i\bar{\jmath}k\bar{\ell}} \psi^i \psi^k \overline{\psi}^{\bar{\jmath}} \overline{\psi}^{\bar{\ell}}. \tag{6.21}$$

It should not be surprising that complex Riemannian geometry has arisen in which the complex scalar fields play the role of complex (holomorphic) coordinates on the target space (the space of chiral field vevs). Field redefinitions which preserve the chiral nature of the fields, $\phi^i \to f^i(\phi)$, are just complex coordinate transformations on the target space, implying that the target space will naturally have the structure of a manifold. The bosonic kinetic term naturally defines a postive-definite quadratic form on this manifold, thus giving it a metric structure. What is special to supersymmetry is that the target space geometry that occurs is actually *Kahler geometry*—complex geometry in which the metric is defined as above from a Kahler potential, and that the fermion fields ψ^i are naturally interpreted as vectors in the tangent space to the Kahler manifold since then $\mathcal{D}_\mu \psi^i$ is a covariant space-time derivative.

6.4. Superpotential and F-terms

So far we have written down the kinetic terms for a theory of chiral superfields. When we included some non-renormalizable terms in the nlσm, all the interaction terms included derivatives, except for the four-fermion terms. There are other non-derivative (and renormalizable) interaction terms which are supersymmetry invariant, and are found among those which cannot be written as integrals over the whole of superspace.

Recall that a general chiral superfield (or product thereof) has the component expansion $\Phi = \phi(y) + \sqrt{2}\theta\psi(y) + \theta^2 F(y)$. Consider adding a term $\mathcal{L}_\mathcal{W}$ which is an integral of a chiral superfield over *half* of superspace

$$\mathcal{L}_\mathcal{W} = \int d^2\theta\, \Phi = F(x), \tag{6.22}$$

since any terms coming from the expansion of $y^\mu = x^\mu + i\theta\sigma^\mu\overline{\theta}$ are total derivatives.

▶ **Exercise 6.4.** Show that the supersymmetry variation of the F-term of an arbitrary chiral superfield is

$$\delta_\epsilon F = -\frac{1}{\sqrt{2}}\overline{\epsilon}\overline{\sigma}^\mu\partial_\mu\psi, \tag{6.23}$$

by computing $[\epsilon Q + \overline{\epsilon}\overline{Q}, \Phi]$ and picking out the θ^2 term.

This is a total derivative, and so the F-term is indeed a supersymmetry invariant.

The general non-derivative interactions that can be written in this way are

$$\mathcal{L}_\mathcal{W} = \int d^2\theta\, \mathcal{W}(\Phi^i) + h.c. \tag{6.24}$$

▶ **Exercise 6.5.** Show that, in components,

$$\mathcal{L}_\mathcal{W} = \partial_i\mathcal{W}F^i - \tfrac{1}{2}\partial_i\partial_j\mathcal{W}\psi^i\psi^j + h.c., \tag{6.25}$$

where

$$\partial_i\mathcal{W} \equiv \frac{\partial}{\partial\phi^i}\mathcal{W}(\phi). \tag{6.26}$$

We remove F^i by solving its equations of motion from the nlσm,

$$F^i = \tfrac{1}{2}\Gamma^i_{jk}\psi^j\psi^k - g^{i\bar{\imath}}\partial_{\bar{\imath}}\overline{\mathcal{W}}, \tag{6.27}$$

giving a scalar potential

$$V(\phi,\overline{\phi}) = (\partial_i\mathcal{W})g^{i\bar{\imath}}(\partial_{\bar{\imath}}\overline{\mathcal{W}}), \tag{6.28}$$

as well as fermion masses, Yukawa terms, *etc.* —general two-fermion terms:

$$-\tfrac{1}{2}(\partial_j\partial_k\mathcal{W} - \Gamma^i_{jk}\partial_i\mathcal{W})\psi^j\psi^k + h.c. \tag{6.29}$$

Note that since the metric $g_{i\bar{\imath}}$ is positive-definite (for unitarity), the scalar potential $V \geq 0$. Thus the potential attains its minimum and supersymmetry is unbroken when

$$V = 0 \quad \Leftrightarrow \quad \partial_i\mathcal{W} = 0. \tag{6.30}$$

Note also that the F^i auxiliary fields were not only auxiliary, but also appeared only quadratically. Thus the classical step of replacing them by their equations of motion was also valid quantum-mechanically. The fact that the F^i always appear at most quadratically follows simply from the fact that they are the highest components of the chiral superfields. It has become standard terminology to refer to the terms appearing in the scalar potential V coming from the superpotential as "F-terms". (We will later see that there is another contribution to the scalar potential when vector multiplets are included—the so-called "D-terms".)

6.5. Wess-Zumino Model, Effective Actions, and the Nlσm

We can now write down the most general renormalizable Lagrangian involving chiral fields—the Wess-Zumino model. It is simply

$$\mathcal{L} = \int d^4\theta\, \overline{\Phi}_i \Phi^i + \int d^2\theta (\nu_i \Phi^i + \mu_{ij}\Phi^i\Phi^j + \lambda_{ijk}\Phi^i\Phi^j\Phi^k) + c.c. \tag{6.31}$$

Here the kinetic (Kahler) terms define the normalization of the fields (*i.e.* the Kahler metric has been diagonalized by field redefinitions), and the only free couplings are the complex parameters ν, μ and λ. (Actually, the superpotential terms linear in Φ^i can often be removed by shifting the fields. For example, if $\lambda_{ijk} = 0$ and μ_{ij} is invertible, then shifting $\Phi^i \to \Phi^i - \frac{1}{2}\mu^{ij}\nu_j$ eliminates the linear terms. In cases where they cannot be eliminated, however, they play an important role, as we will see later.) One can check by the usual power-counting argument that these are indeed all the renormalizable terms.

Let us look at this dimension-counting argument in detail, in order to clarify the importance of the WZ model *versus* the nlσm. The most general Lagrangian we could write down would look like the nlσm but with arbitrary covariant and space-time derivatives, $D^\ell \overline{D}^m \partial^n \Phi$, allowed in $\mathcal{K}$ and arbitrary space-time derivatives, $\partial^n\Phi$, allowed in $\mathcal{W}$. (Since $\overline{D}$ commutes with ∂_μ, space-time derivatives of a χsf are also χsf's.) Let us count the classical scaling dimensions of these fields.

First of all, we define x^μ to have dimension -1 (*i.e.* the dimensions of an inverse energy or mass), so by the supersymmetry algebra we read off the dimensions

object	scaling dimension
x, dx	-1
∂_x, P	$+1$
$d\theta$, ∂_θ, D, Q	$+\frac{1}{2}$
θ	$-\frac{1}{2}$
Φ	Δ

In the last line we have assigned the χsf an arbitrary scaling dimension Δ. Since the action appears exponentiated in the path integral, it should have total dimension zero, and thus the Lagrangian will have dimension 4. This implies that a dimension Δ operator (term) appearing in the Lagragian has a coefficient which scales with dimension $4 - \Delta$.

What is the correct scaling Δ of a χsf? This is typically a dynamical question, and must be computed in the quantum theory using various techniques (the renormalization group). However, at weak coupling (which we will see later is what is relevant for the WZ model), if one is interested in the physics of fluctuations around a given vacuum in which the scalar fields have zero expectation value, then one should use the free-field scaling dimensions with $\Delta = 1$, (making the coefficient of the quadratic kinetic term dimensionless). This is often called the "classical" dimension of the field, since it assigns the usual classical scale dimensions to the fields: $[\phi] = 1$, $[\psi] = 3/2$, and $[F] = 2$. This is a misnomer, though, since this scaling really comes from the behavior of quantum fluctuations around this vacuum; I will call this scaling "kinetic scaling".

In the kinetic scaling, any couplings in the action should be assigned dimensions such that the Kahler potential $\mathcal{K}$ has dimension 2 and the superpotential $\mathcal{W}$ has dimension 3, so that the whole action is dimensionless. It is then easy to see that any (Lorentz-invariant, non-total-derivative) terms in the Kahler potential and superpotential with derivatives will have coefficients with kinetic dimensions less than or equal to -1, and hence be non-renormalizable (irrelevant) by power counting. Thus the WZ theory is picked out as the most relevant terms in the kinetic scaling.

This point of view makes the nlσm we studied in the last lecture and in this lecture seem special in an arbitrary way: it contained non-renormalizable terms, but only a special subset of them (those without derivatives). Actually, however, the nlσm terms are natural in a different scaling.

There is another assignment of scaling dimensions to χsf's which should more properly be thought of as a "classical scaling". It arises when one is trying to determine from the effective action what is the vacuum picked out by the theory. In this case there is nothing *a priori* to single-out an origin in field space for the scalar fields appearing in the effective theory. In this case the scaling dimension of the scalar fields should be set to zero, so $\Delta = 0$. I will call this scaling "vacuum scaling". It is just the scalings of the classical equations of motion following from the effective action.

In the vacuum scaling, the coefficients in the superpotential have dimension $+3$ and those of the Kahler terms dimension $+2$ in the nlσm. Thus the coefficients of the Kahler terms are less relevant than those of the superpotential. This is what one expects in a vacuum scaling, since the potential, which picks out the vacuum classically, should be most important. Note that in this scaling the usual distinction between renormalizable and non-renormalizable terms is not what is important. Furthermore, terms with explicit derivatives in the $d^4\theta$ and $d^2\theta$ parts of the Lagrangian have vacuum scaling dimensions greater than or equal to 3, and thus are less relevant than either the nlσm Kahler or superpotential pieces.

Thus the nlσm includes the most relevant (in the colloquial sense!) terms in an effective action for determining the vevs of the scalar fields. Thus, with such an effective action, one can solve for the vacuum and expand about it. In this expansion, it is the kinetic scaling dimension which determines the relevant terms. Since kinetic and vacuum dimensions are different, terms which were relevant for determining the vacuum may no longer be relevant in the low energy physics by power counting. Such terms are examples of what are known in the condensed matter literature as "dangerously irrelevant" operators.

All this you actually know very well already, as a simple example will show: Consider a scalar field theory with potential $V = -\phi^2 + \phi^{100}$. Though the ϕ^{100} term is very irrelevant by power counting, it is needed to stabilize the vacuum at $\langle\phi\rangle = (1/50)^{1/98}$. Shifting to this vacuum and expanding gives a potential $V \sim 2(100\widetilde{\phi}^2/2! + 100^2\widetilde{\phi}^3/3! + 100^3\widetilde{\phi}^4/4!)$ plus irrelevant terms. I have belabored this difference between the scaling of scalar vevs and the scaling of their fluctuations because it will pay to have it clearly in mind in more complicated supersymmetric situations we will meet, where there are often intricate continuous spaces of exactly degenerate vacua of a given effective theory.

7. Selection rules and the nonrenormalization theorem

In this lecture we finally turn to the quantum mechanical aspects of the nlσm. The main result we wish to show is the celebrated "non-renormalization theorem". Though this theorem was proven laboriously to all orders in perturbation theory, it turns out to have a simple and conceptual proof due to N. Seiberg, *Phys. Lett.* 318B (1993) 469 that holds non-perturbatively. First, though, we introduce the only allowed bosonic "extension" of $N{=}1$ supersymmetry in $3+1$ dimensions, which plays an important role in proving the non-renormalization theorem.

7.1. R-Symmetries

The nlσm can, of course, have global symmetries which act on the chiral fields. By the Coleman-Mandula theorem (see lecture 1) all global symmetries must commute with the Poincaré group. However, it is not necessary that they commute with the supersymmetry algebra. In fact, associativity of the superPoincaré algebra implies[6] that there can be only at most a single (independent) hermitian $U(1)$ generator R which does not commute with the supersymmetry generators, and is conventionally normalized so that:

$$[R, Q_\alpha] = -Q_\alpha, \qquad [R, \overline{Q}_{\dot\alpha}] = +\overline{Q}_{\dot\alpha}. \tag{7.1}$$

[6] Say there were a global symmetry algebra with Hermitian generators T^a, $[T^a, T^b] = i f^{ab}_c T^c$ which did not commute with supersymmetry $[T^a, Q_\alpha] = h^a Q_\alpha$. The Jacobi identity $[T^a,[T^b,Q]] + [T^b,[Q,T^a]] + [Q,[T^a,T^b]] = 0$ implies $f^{ab}_c h^c = 0$. Now, by the Coleman-Mandula theorem, any scalar symmetry algebra is a direct sum of a semi-simple algebra $\mathcal{A}_1$ and an Abelian algebra $\mathcal{A}_2$. Since for a semi-simple Lie algebra the Killing form $g^{ab} = f^{ay}_x f^{bx}_y$ is non-degenerate (Cartan's theorem), we can go to a basis in which it is diagonal, and f^{abc} is antisymmetric (we raise and lower indices with g). Then $0 = f_{abc}h^c = f^{bad} f_{abc} h^c \propto h^d$. Thus only the components of h^c in $\mathcal{A}_2$ (the Abelian directions) can be non-zero. But then we can define the linear combination $R = \sum_a \overline{s}_a B_a / (\sum_b \overline{s}_b s_b)$ with the desired commutation relations. Note that in theories with extended supersymmetries, non-Abelian R-symmetries are allowed, *e.g.* $SU(2)\times U(1)$ in $N=2$ supersymmetry in 4 dimensions.

This single $U(1)$ under which Q has charge 1 is called the R-symmetry.

Since the R-symmetry does not commute with supersymmetry, the component fields of a χsf do not all carry the same R-charge. So we call the R-charge of the lowest component field the R-charge of the superfield, and assign R-charge $+1$ to θ_α. So, for example, if Φ has R-charge $R(\Phi) = r$, then $R(\phi) = r$, $R(\psi) = r - 1$, and $R(F) = r - 2$. Since $R(d\theta) = -1$, it follows that

$$R(\mathcal{W}) = +2. \tag{7.2}$$

In practice this last requirement provides the simplest way of finding an R-symmetry.

7.2. Holomorphy of the Superpotential

The key to the non-renormalization theorem is a supersymmetry Ward identity which implies that all coupling constants which appear in the classical (microscopic) superpotential must only appear holomorphically in quantum corrections to the superpotential. We will prove this formally as a supersymmetry Ward identity at the end of this lecture. But there is a simpler and, in the end, more powerful approach to this subject using the notion of effective actions.

Think of all the coupling constants which appear in the superpotential (*e.g.* masses, Yukawa couplings, *etc.*) as classical background chiral superfields. It then follows that these couplings can only appear in the effective superpotential holomorphically—*i.e.* if λ is a coupling, then only λ and not $\overline{\lambda}$ can appear in any quantum corrections to the superpotential, since the superpotential is a function only of χsf's, not $\overline{\chi}$sf's.

Let us examine more closely the logic of this argument. We are considering a nlσm describing the physics below some scale Λ, and we wish to understand the effective description of the model describing the physics in the IR, *i.e.* at scales below $\widetilde{\Lambda} \ll \Lambda$, that is to say, after integrating out some higher-energy degrees of freedom. (This is the essential problem of condensed matter theory, and its converse is the basic problem of particle theory!) We will henceforth describe the theory at the scale Λ as the *microscopic theory* and the effective theory at the scale $\widetilde{\Lambda}$ as the *macroscopic, or effective theory.*

We now *assume* that the macroscopic theory will also be described by a nlσm with a specified set of light chiral fields—not necessarily a simple subset of those of the microscopic theory. We have no derivation of this hypothesis—we can only test it to see if it gives consistent answers. The couplings of the effective theory will be some functions of the couplings of the microscopic theory, which we would like to solve for.

The next step of thinking of the couplings in the superpotential as background chiral superfields is just a trick—we are certainly allowed to do so if we like (since the couplings enter in the microscopic theory in the same way a backgound chiral superfield would). The point of this trick is that it makes the restrictions on possible quantum corrections allowed by supersymmetry apparent. These restrictions are just a supersymmetric version of the familiar "selection rules" of QM.

Perhaps an example from QM will make this clear: Recall the Stark effect, in which one calculates corrections to the hydrogen atom spectrum in a constant background electric field. Thus we perturb the Hamiltonian by adding a term of the form

$$\delta H = E_1 x_1 + E_2 x_2 + E_3 x_3. \tag{7.3}$$

But the resulting perturbed energy levels cannot depend on the perturbing parameters E_i arbitrarily. Indeed, one simply remarks that the electric field transforms as a vector $\mathbf{E}$ under rotational symmetries, thus giving selection rules for which terms in a perturbative expansion in the electric field strength it can contribute to. On the other hand, these selection rules are equally valid without the interpretation of the electic field as a background field transforming in a certain way under a symmetry (which it breaks). Instead, one could think of it as an abstract perturbation, and the selection rules follow simply because it is *consistent* to assign the perturbation transformation rules under the broken rotational symmetry.

The holomorphy of the superpotential is the same sort of a selection rule, but this time following from supersymmetry. The only slightly unusual feature of it is that the couplings in the superpotential do not explicitly break the supersymmetry.

We can immediately see the power of this supersymmetry selection rule. For suppose our enlarged theory (thinking of λ as a chiral superfield) has a $U(1)$ global symmetry under which λ has charge $Q(\lambda) = 1$, *i.e.* in the tree-level (classical) superpotential there is a term

$$\mathcal{W}_{\text{tree}} \supset \lambda \mathcal{O}_{-1} \tag{7.4}$$

where $\mathcal{O}_{-1}$ is some charge -1 operator. Say we are interested in the appearance of a given operator $\mathcal{O}_{-10}$ of charge $Q(\mathcal{O}_{-10}) = -10$ among the quantum corrections. Normally, one would say that this operator can appear only at tenth and higher orders in perturbation theory:

$$\delta\mathcal{W} \sim \lambda^{10}\mathcal{O}_{-10} + \lambda^{11}\overline{\lambda}\mathcal{O}_{-10} + \ldots + \lambda^{10} e^{-1/|\lambda|^2}\mathcal{O}_{-10} + \ldots, \tag{7.5}$$

(assuming that there is a regular $\lambda \to 0$ limit, so that no negative powers of λ are allowed), where I've also indicated potential non-perturbative contributions as well. However, by the above argument we learn that *only* the tenth-order term is allowed, all the higher-order pieces, including the non-perturbative ones, are disallowed since they necessarily depend on λ non-holomorphically.

Even more importantly, any operator of *positive* charge under the $U(1)$ symmetry is completely disallowed, since it would necessarily have to have inverse powers of λ as its coefficient. But since we assumed the $\lambda \to 0$ weak-coupling limit was smooth (*i.e.* the phsyics is under control there), such singular coefficients are disallowed. Note that this is again special to supersymmetry, for if non-holomorphic couplings were allowed, one could always include such operators with positive powers of $\overline{\lambda}$ instead.

This argument can be summarized prescriptively as follows: The effective (macroscopic) superpotential is constrained by

(1) holomorphy in the (microscopic) coupling constants,

(2) "ordinary" selection rules from symmetries under which the coupling constants may transform, and

(3) smoothness of the physics in various weak-coupling limits.

Much of the progress in understanding the non-perturbative dynamics of supersymmetric gauge theories of the past few years has resulted from the systematic application of the above argument (pioneered by N. Seiberg). Indeed, the second half of this course will consist of just that.

7.3. Nonrenormalization Theorem for the Nlσm

Let us now apply this argument to the nlσm. We start with a simpler special case:

$$\mathcal{W} = \tfrac{1}{2} m \Phi^2 + \tfrac{1}{3} \lambda \Phi^3. \tag{7.6}$$

By holomorphy, the effective superpotential is

$$\mathcal{W}_{\text{eff}} = f(\Phi, m, \lambda), \tag{7.7}$$

that is, a function of Φ, m, and λ and not their complex conjugates. Note that we have made the assumption that the effective theory is still described in terms of a single chiral superfield Φ. The microscopic superpotential is invariant under the global symmetries

$$\begin{array}{cccc} & U(1) & \times & U(1)_R \\ \Phi & +1 & & +1 \\ m & -2 & & 0 \\ \lambda & -3 & & -1 \end{array} \tag{7.8}$$

where we have assigned the coupling constants charges. This implies the effective superpotential must actually be

$$\mathcal{W}_{\text{eff}} = m\Phi^2 f\left(\frac{\lambda\Phi}{m}\right) = \sum_n a_n \, \lambda^n m^{1-n} \Phi^{n+2}. \tag{7.9}$$

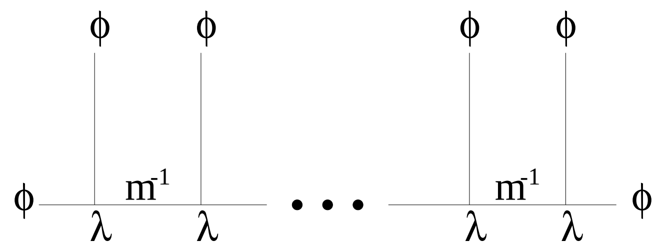
Now, expanding around the $\lambda \to 0$ limit, in which the theory is free, we see that only terms with $n \geq 0$ are allowed. Furthermore, we can also take the $m \to 0$ limit at the same time to conclude that terms with $n > 1$ are disallowed. So we learn that

$$\mathcal{W}_{\text{eff}} = \tfrac{1}{2} m \Phi^2 + \tfrac{1}{3} \lambda \Phi^3 = \mathcal{W}_{\text{micro}}. \tag{7.10}$$

The coefficients of the two allowed terms are determined by matching to perturbation theory when λ is small. Thus we see that the superpotential is (non-perturbatively) un-renormalized. Furthermore, we see that our assumption that the low-energy effective physics is described by a single chiral superfield is consistent.

The last step of taking the $m \to 0$ limit deserves a few words. Taking this limit at finite λ does not lead to weakly-coupled physics; however, by taking *both* λ and m to zero such that $m/\lambda \to 0$, we achieve the desired result. One may wonder, though, whether the $m \to 0$ limit is really smooth, for though the resulting theory is free, it also has a massless particle, and so the effective theory should have IR divergences—perhaps reflected in divergences of the superpotential? This is not the case, though, since we do not have to run the renormalization group (RG) down to $p_\mu = 0$, and so we should not see any IR divergences in our Wilsonian effective action.

Another, equivalent, way to see this point is to note that the terms with $n \geq 0$ are all generated by tree diagrams in the microscopic theory:



Since this is not a 1PI diagram for $n > 1$, it should not be included in the effective action for finite m. Equivalently, there are poles at $p_\mu = 0$ if $m = 0$ in intermediate propagators, and so should not be included in a Wilsonian effective action.[7]

▸ **Exercise 7.1.** Show that the Wess-Zumino model superpotential, $\mathcal{W} = \frac{1}{2} m_{ij} \Phi^i \Phi^j + \frac{1}{3} \lambda_{ijk} \Phi^i \Phi^j \Phi^k$, is un-renormalized.

Let us generalize this example a bit further, to

$$\mathcal{W} = \mu_1 \Phi + \mu_2 \Phi^2 + \ldots + \mu_n \Phi^n + \ldots \tag{7.11}$$

which has the global symmetries

$$\begin{array}{cccc} & U(1) & \times & U(1)_R \\ \Phi & +1 & & +1 \\ \mu_n & -n & & 2-n \end{array} \tag{7.12}$$

implying

$$\mathcal{W}_{\text{eff}} = \mu_1 \Phi \, f\left(\frac{\mu_2}{\mu_1} \Phi, \frac{\mu_3}{\mu_1} \Phi^2, \ldots, \frac{\mu_n}{\mu_1} \Phi^{n-1}, \ldots \right). \tag{7.13}$$

[7] For a recent detailed exposition of the connection between 1PI and Wilsonian effective actions in this context, see E. Poppitz and L. Randall, *Holomorphic anomalies and the nonrenormalization theorem,"* `hep-th/9608157`.

Demanding a smooth limit as all $\mu_i \to 0$ then implies

$$\mathcal{W}_{\text{eff}} = \mathcal{W}. \tag{7.14}$$

▸ **Exercise 7.2.** Prove the nonperturbative non-renormalization theorem for the nlσm.

In particular, the non-renormalization theorem shows no contradiction with our assumption that the low-energy degrees of freedom are the same as the microscopic χsf's. This is in line with our expectations from the Coleman-Gross theorem, that nlσm physics becomes more weakly coupled in the IR. We thus conclude that no "interesting" strongly-coupled IR physics occurs, such as the formation of bound states or condensates like $F_c = \langle\psi\psi\rangle$, since we need introduce no such new condensate field C inthe effective action. (The fermion bilinear is the F-component of a χsf.)

7.4. Renormalization Schemes and the Kahler Potential

Although the non-renormalization theorem implies that the superpotential gets no corrections, it says nothing about the Kahler terms. Recalling that the superpotential encodes the masses and interactions, while the Kahler potential encodes the kinetic terms, we see the non-renormalization theorem implies that there are no quadratic divergences so masses are protected from quantum corrections, but there may be wave-function renormalizations.

As an example, consider the WZ model with one χsf:

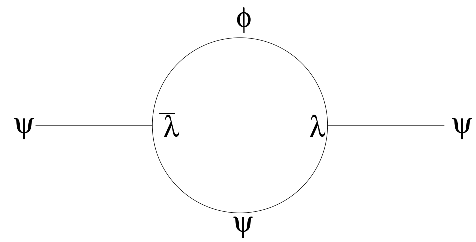
$$\mathcal{L} = \int d^4\theta\, Z^2\overline{\Phi}\Phi + \int d^2\theta \left(m\Phi^2 + \lambda\Phi^3\right) + h.c., \tag{7.15}$$

where $Z = Z(m, \lambda, \overline{m}, \overline{\lambda}, \mu)$ is not in general a holomorphic function. Here μ is the RG point. This is in a scheme in which we do *not* renormalize the fields so that $Z = 1$. If we did do this, then m and λ would indeed run; if not, then they do not. Just to see that the non-renormalization theorem is not content-less, note that if we *do* set $Z = 1$, then m^3/λ^2 is still not renormalized:

$$\hat{\mathcal{L}} = \int d^4\theta\, \hat{\overline{\Phi}}\hat{\Phi} + \int d^2\theta \left(\frac{m}{Z^2}\hat{\Phi}^2 + \frac{\lambda}{Z^3}\hat{\Phi}^3\right) + h.c., \tag{7.16}$$

where $\hat{\Phi} \equiv Z\Phi$.

Let's compute the renormalization of the Kahler potential in perturbation theory (and in components):



giving
$$Z = 1 + \lambda\overline{\lambda}\log\left|\frac{\mu}{m}\right| + \ldots \tag{7.17}$$
where the first term is the tree result, the logarithm is the usual one-loop contribution (determined by the symmetries), and the sign is correct since as $m \to 0$, $Z \to +\infty$, so that in the IR the theory becomes weakly-coupled in the $\hat{\Phi}$ variables.

7.5. Two Supersymmetry Ward Identities

In cases where supersymmetry is not broken we can derive the supersymmetric selection rule (holomorphy of the superpotential) from exact Ward identities (M. Shifman and A. Vainshtein, *Nucl. Phys.* B277 (1986) 456; *Nucl. Phys.* B359 (1991) 571; see also D. Amati, *et. al.*, *Phys. Rep.* **162**(1988)571). This approach gives a weaker result than the Wilsonian effective action approach.

Consider the correlation function of the scalar components of some chiral superfields Φ_i:
$$\mathcal{G} = \langle \phi_1(x_1)\cdots\phi_n(x_n)\rangle \tag{7.18}$$
Noting that the supersymmetry variations of the components of a chiral superfield are given by
$$\begin{aligned}
[Q_\alpha, \phi] &= \psi_\alpha, \\
[\overline{Q}_{\dot\alpha}, \phi] &= 0, \\
\{Q_\alpha, \psi_\beta\} &= \epsilon_{\alpha\beta}F, \\
\{\overline{Q}_{\dot\alpha}, \psi_\beta\} &= \sqrt{2}\sigma^\mu_{\dot\alpha\beta}\partial_\mu\phi,
\end{aligned} \tag{7.19}$$
then
$$\begin{aligned}
\sigma^\mu_{\dot\alpha\beta}\frac{\partial}{\partial x_1^\mu}\mathcal{G} &= \langle\{\overline{Q}_{\dot\alpha}, \psi_\beta\}\phi_2\cdots\phi_n\rangle \\
&= \langle 0|\overline{Q}\psi_1\phi_2\cdots\phi_n|0\rangle + \langle 0|\psi_1\phi_2\cdots\phi_n\overline{Q}|0\rangle \\
&= 0,
\end{aligned} \tag{7.20}$$
where in the last step we used the fact that the vacuum is annihilated by Q and $\overline{Q}$ if supersymmetry is unbroken. This implies that $\mathcal{G}$ is independent of the x_i. Taking the limit as all the x_i are far apart, we can thus conclude[8] that
$$\mathcal{G} \propto \langle\phi_1\rangle\cdots\langle\phi_n\rangle. \tag{7.21}$$

[8] This assumes that there is a single vacuum. If there are multiple vacua $|0_i\rangle$, then we should conclude instead that $\mathcal{G} = \sum_i a_i\langle\phi_1\rangle_i\cdots\langle\phi_n\rangle_i$ for some constants a_i.

To prove the second Ward identity, suppose the (microscopic) Lagrangian has a superpotential interaction of the form

$$\mathcal{L}_{int} = \int d^2\theta\, \lambda\Phi_0 + h.c., \tag{7.22}$$

where Φ_0 is some composite χsf. Then,

$$\begin{aligned} \frac{\partial}{\partial\overline{\lambda}}\mathcal{G} &= \int d^4x_0 d^2\theta \langle\overline{\Phi}_0(x_0)\phi_1(x_1)\cdots\phi_n(x_n)\rangle \\ &= \int d^4x_0 \langle F_0\phi_1\cdots\phi_n\rangle \\ &= \int d^4x_0 \langle\{\overline{Q}_{\dot\alpha}, \overline{\psi}^{\dot\alpha}\}\phi_1\cdots\phi_n\rangle \\ &= 0 \end{aligned} \tag{7.23}$$

where in the last step I have again commuted $\overline{Q}$ to the left and right where it annihilates the vacuum. Putting the two Ward identities together, we learn that $\langle\Phi\rangle$, for an arbitrary composite χsf Φ, can depend only holomorphically on the couplings in the microscopic superpotential:

$$\frac{\partial}{\partial\overline{\lambda}}\langle\Phi\rangle = 0. \tag{7.24}$$

This is almost equivalent to the statement that the effective superpotential is holomorphic in the couplings.

8. Moduli space, "integrating out", and singularities in effective actions

In this lecture we introduce the notion of a "moduli space"—the space of vacua of a theory. For a theory to have a non-trivial moduli space necessarily means that it has more than one vacuum. In regular (non-supersymmetric) field theories the usual examples of degenerate vacua occur due to broken symmetries—*i.e.* global symmetries relate all the vacua. Any further degeneracies are considered accidental since presumably quantum corrections or small irrelevant operators will lift the non-symmetry-enforced degeneracies. In supersymmetric theories, on the other hand, moduli spaces of degenerate vacua not related by any global symmetry frequently occur, due essentially to the holomorphy of the superpotential. We will see this through some examples to follow. The existence of moduli spaces and the interpretation of their singularities turns out to be a very powerful tool for controlling supersymmetric field theories.

8.1. Moduli space examples

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

The minima are determined by $\partial\mathcal{W} = 0$, giving

$$\phi = 0, \ -\frac{m}{\lambda}. \tag{8.1}$$

If we shift $\Phi \to \Phi + m/(2\lambda)$ we get

$$\mathcal{W} = \frac{1}{3}\lambda\Phi^3 - \frac{1}{4}\frac{m^2}{\lambda}\Phi + \text{const.} \tag{8.2}$$

which has a symmetry generated by

$$\begin{aligned} \phi &\to -\phi \\ \theta &\to i\theta \end{aligned} \tag{8.3}$$

which is a $\mathbb{Z}_4$ R-symmetry relating the two ground states. Thus they have equivalent physics, though they are distinct. Note that this model is not generic—all odd powers of Φ are allowed by the $\mathbb{Z}_4$ R-symmetry.

(2) $\mathcal{W} = \frac{1}{2}\lambda L H^2$

Extrema of $\mathcal{W}$ are at

$$H = 0, \qquad L = \text{arbitrary}, \tag{8.4}$$

implying a whole moduli space, $\mathcal{M}$, of degenerate but inequivalent classical ground states. We can see that they are inequivalent because their physics is different: the spectrum at any such vacuum is one massless chiral multiplet L, and one massive chiral multiplet H with mass $|\lambda\langle L\rangle|$. This model is uniquely specified by its field content, a $U(1)$ global symmetry (with charge Q), and an R-symmetry, under which

$$\begin{aligned} Q(L) = 2, \quad Q(H) = -1, \\ R(L) = 2, \quad R(H) = 0. \end{aligned} \tag{8.5}$$

Since the Kahler potential of this model is $\mathcal{K} = L\overline{L} + H\overline{H}$, the metric induced on $\mathcal{M}$ is

$$ds^2 = dL\, d\overline{L}, \tag{8.6}$$

classically. When we include quantum effects, though $\mathcal{M}$ remains unchanged, its metric will recieve quantum corrections, for although the non-renormalization theorem implies that the superpotential gets no corrections, it says nothing about the Kahler terms.

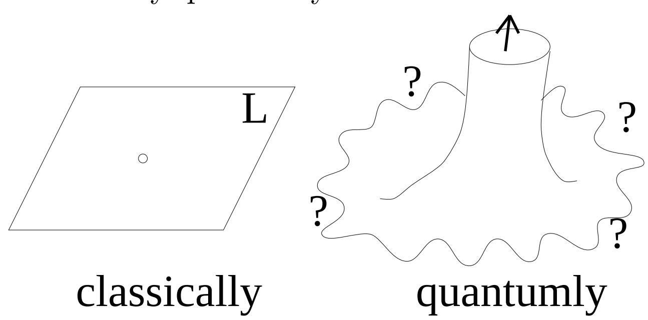
Let's look at the Kahler potential in perurbation theory:

$$\begin{aligned}\mathcal{K} &= +\overline{L}L - \#\overline{L}L|\lambda|^2 \log\left|\frac{L}{\Lambda}\right|^2 + \dots \\ &= -\#\overline{L}L|\lambda|^2 \log\left(\frac{\overline{L}L}{\Lambda^2 e^{1/|\lambda|^2}}\right) + \dots\end{aligned} \tag{8.7}$$

due to the one-loop contribution of the massive H's. This result implying that at $L = 0$ there is a Kahler metric singularity. for the one-loop perturbative result (8.7) becomes exact as $L \to 0$, since the theory is more and more weakly coupled in this limit. So the metric is given by $(ds)^2 = g_{L\overline{L}} dL d\overline{L}$ with

$$g_{L\overline{L}} = \partial_L \partial_{\overline{L}} \mathcal{K} \simeq -|\lambda|^2 \log L\overline{L} + \text{const} \to \infty \tag{8.8}$$

as $L \to 0$. (Note that for L large, $\geq \lambda e^{1/|\lambda|^2}$, the logarithm goes negative, so the metric is negative—the "Landau pole".) Thus the moduli space which classically is the complex L-plane, has a cylindrical infinity quantumly:



This singularity at $L = 0$ has a physical interpretation: it corresponds to the fact that when $L = 0$ a particle multiplet (H) is becoming massless. We will explore the reasons for this physical interpretation shortly.

(3) $\mathcal{W} = XYZ$

Extrema of $\mathcal{W}$ are at

$$XY = YZ = ZX = 0, \ \Rightarrow \ \{X = Y = 0, \ Z \text{ arbitrary; \& permutations}\}. \tag{8.9}$$

This example shows that the moduli space of vacua need not be a manifold (perhaps with singularities), but can also have intersections. This model is generic given a $U(1)_1 \times U(1)_2 \times U(1)_R$ symmetry under which

$$\begin{aligned} Q_1(X) &= +1, & Q_1(Y) &= +1, & Q_1(Z) &= -2, \\ Q_2(X) &= +1, & Q_2(Y) &= -2, & Q_2(Z) &= +1, \\ R(X) &= 0, & R(Y) &= +1, & R(Z) &= +1. \end{aligned} \tag{8.10}$$

(4) $\mathcal{W} = \mu^2 \Phi$

There are no extrema of $\mathcal{W}$, showing that supersymmetry is broken classically (spontaneously, at tree level) in this model. The potential is

$$V = |\partial\mathcal{W}|^2 = |\mu|^4, \tag{8.11}$$

showing that, in fact, there is a whole moduli space of degenerate, non-supersymmetric vacua in this model. Note, that this model is generic given a $U(1)$ R-symmetry under which

$$R(\Phi) = +2. \tag{8.12}$$

<u>(5) $\quad \mathcal{W} = \mu^2\Phi_0 + m\Phi_1\Phi_2 + g\Phi_0\Phi_1^2$</u>

This is called the *O'Raifeartaigh model*, the supersymmetric analog of spontaneous symmetry breaking. The extrema of $\mathcal{W}$ are at

$$\begin{aligned} 0 &= \partial_0\mathcal{W} = \mu^2 + g\phi_1^2, \\ 0 &= \partial_1\mathcal{W} = m\phi_2 + 2g\phi_0\phi_1, \\ 0 &= \partial_2\mathcal{W} = m\phi_1, \end{aligned} \tag{8.13}$$

which have no solution, implying supersymmetry is broken.

▸ **Exercise 8.1.** Compute the potential to find the ground state(s) and their spectra of bosons and fermions in the O'Raifeartaigh model. What happens when you take some of μ, m, $g \to 0$?

This is an important exercise to do! We will reproduce many of the qualitative features of the solution in the next lecture, when we discuss spontaneous supersymmetry breaking.

The O'Raifeartaigh potential is natural given a $\mathbb{Z}_2$ symmetry (with charge Π) and a $U(1)$ R-symmetry under which

$$\begin{aligned} \Pi(\Phi_0) = +, \quad \Pi(\Phi_1) = -, \quad \Pi(\Phi_2) = - \\ R(\Phi_0) = 2, \quad R(\Phi_1) = 0, \quad R(\Phi_2) = 2. \end{aligned} \tag{8.14}$$

8.2. Integrating-Out

So far we have not been very explicit about what scale the effective superpotential is supposed to describe. Since it includes the same chiral fields Φ_i that appeared in the microscopic superpotential, it must be valid on energy scales at least down to the mass of the heaviest Φ_i. To get the minimal effective action at a scale below the masses of some fields in the theory, those fields should be integrated out. (There is nothing wrong in principle with leaving them in, except that they are play no dynamical role in the effective theory. Let us see how this is done in a simple example.

Consider

$$\mathcal{W} = \tfrac{1}{2}mH^2 + \lambda L^2 H. \tag{8.15}$$

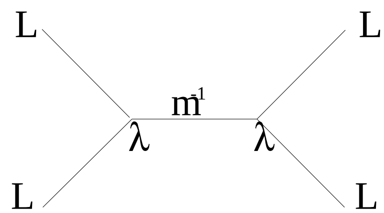
We have seen already that this superpotential gets no quantum corrections. It is easy to see that for finite m, the spectrum of this theory is one massless superfield L (for "light") and one massive superfield H (for "heavy") with mass m. Thus to get the effective superpotential for scales below m, we must integrate H out. To do this, one can simply complete the square in $\mathcal{W}$ finding

$$\mathcal{W} = \tfrac{1}{2} m \left(H + \frac{\lambda}{m} L^2 \right)^2 - \tfrac{1}{2} \frac{\lambda^2}{m} L^4. \tag{8.16}$$

Then the first term, being massive, is simply dropped, giving the effective superpotential

$$\mathcal{W}_{\text{eff}} = -\tfrac{1}{2} \frac{\lambda^2}{m} L^4. \tag{8.17}$$

It is easy to see that this is just the effective superpotential one generates in perturbation theory.



This method of integrating out H may seem special to the fact that H appeared only quadratically in $\mathcal{W}$. So let us do the integrating-out in a more general way. We know from holomorphy that $\mathcal{W}_{\text{eff}} = f(L, \lambda, m)$. The fields and couplings have charges under the global symmetries

$$\begin{array}{cccccc} & U(1)_1 & \times & U(1)_2 & \times & U(1)_R \\ L & 0 & & +1 & & +1 \\ H & +1 & & 0 & & 0 \\ m & -2 & & 0 & & +2 \\ \lambda & -1 & & -2 & & 0 \end{array} \tag{8.18}$$

Thus $\mathcal{W}_{\text{eff}} \propto \lambda^2 L^4 / m$. Matching to perturbation theory in the $\lambda \to 0$ limit recovers our earlier result.

Why didn't we take the $m \to 0$ limit, as before, to rule out all the terms? The point here is that this effective theory is only good below the scale m, and so there is no way to take the $m \to 0$ limit without meeting IR divergences. Thus, we actually expect a singularity in the effective theory at $m = 0$. The singularity indicates the breakdown of our effective theory: some new degrees of freedom besides the L field are needed—in this case the H field which becomes massless at $m = 0$.

There is another way of integrating-out H which is much more efficient in more complicated examples. At scales well below m, H is frozen at its expectation value—there is not enough energy to appreciably excite fluctuations in its field. Thus we can integrate out

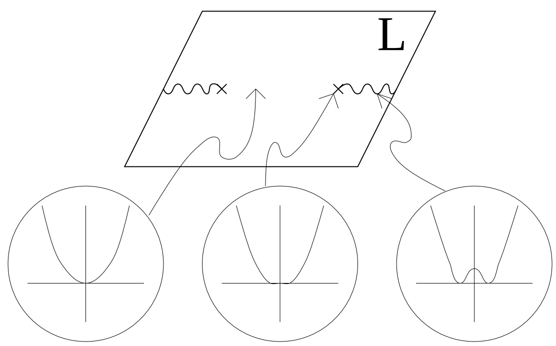
H simply by solving its (algebraic) equation of motion $\partial_H \mathcal{W} = 0$, and substituting back in $\mathcal{W}$.[9] It is not hard to see that, together with the non-renormalization theorem, this is equivalent to the above more laborious procedure using the symmetries and the classical limit.

Let us do a more complicated example:

▸ **Exercise 8.2.** Integrate H out of $\mathcal{W} = \frac{1}{2}mH^2 + \lambda L^2 H + \frac{1}{6}gH^3$ to get

$$\mathcal{W}_{\text{eff}} = \frac{1}{3}\frac{m^3}{g^2}\left[\left(1 - 3\frac{\lambda g}{m^2}L^2\right) \mp \left(1 - \frac{\lambda g}{m^2}L^2\right)\sqrt{1 - 2\frac{\lambda g}{m^2}L^2}\right]. \tag{8.19}$$

What is the interpretation of the singularities and the different signs in this effective superpotential? The L-field space has branch cuts at $L^2 = m^2/(2\lambda g)$. At these points, one can check that $\partial^2\mathcal{W}/\partial H^2 = 0$, implying that H is actually massless there, and therefore it is not OK to integrate it out there, thus the singularities there. The multivaluedness of the superpotential simply reflects the changing vacuum degeneracies in L-space:



To summarize the point of this lecture, it is a general phenomenon that singularities in the effective action correspond to "new" massless particles. In the second half of this course we will begin to form a "dictionary" of singularities and the kind of new physics they correspond to.

9. Supersymmetry breaking in the nlσm

We saw in the last lecture that the nlσm has the possibility of spontaneous supersymmetry breaking (*e.g.* the O'Raifeartaigh model). We will explore this in this lecture, concentrating on the non-perturbative things we can say about supersymmetry breaking.

---

[9] Here we are assuming that supersymmetry is not broken at or above the scale of the effective theory. Thus $\langle H\rangle = \langle h\rangle$, since $\langle\psi_h\rangle = 0$ by Lorentz-invariance and $\langle F_h\rangle = 0$ since supersymmetry is not broken.

9.1. Generalities

Just as in supersymmetric QM, supersymmetry is broken if and only if the vacuum energy density is non-zero. This follows from the supersymmetry algebra, which we can write as

$$P^\mu = -\frac{1}{4}\overline{\sigma}^{\mu\dot{\alpha}\alpha}\{Q_\alpha, \overline{Q}_{\dot{\alpha}}\}. \tag{9.1}$$

Since $\overline{\sigma}^0 = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, it follows that

$$P^0 = \frac{1}{4}\left(Q_1\overline{Q}_1 + Q_2\overline{Q}_2 + \overline{Q}_1 Q_1 + \overline{Q}_2 Q_2\right), \tag{9.2}$$

so the energy density is given by

$$E \cdot \text{volume} = \langle P^0 \rangle = \frac{1}{4}\left(\left|\overline{Q}_1|0\rangle\right| + |Q_1|0\rangle| + \left|\overline{Q}_2|0\rangle\right| + |Q_2|0\rangle|\right) \geq 0. \tag{9.3}$$

Thus

$$E = 0 \Leftrightarrow Q_\alpha|0\rangle = \overline{Q}_{\dot{\alpha}}|0\rangle = 0 \Leftrightarrow \text{supersymmetry unbroken.} \tag{9.4}$$

In the nlσm, the scalar potential was given as the sum of the squares of the F-terms

$$V = \sum_i \left|F^i\right|^2 \geq 0, \tag{9.5}$$

where, recall[10], $F^i = -g^{i\bar{\imath}}\partial_{\bar{\imath}}\overline{\mathcal{W}}$. Thus, the condition for unbroken supersymmetry is a minimum of the potential with $V = 0$, *i.e.* the existence of a solution to the equations

$$\partial_i \mathcal{W} = 0. \tag{9.6}$$

Note that if a solution exists, it is necessarily the absolute minimum since the potential is bounded below by zero.

9.2. Goldstinos

We can see this in yet another way. Supersymmetry is broken if and only if the vev of the supersymmetry variation of some field is non-zero. The field in question must be a fermion, since the supersymmetry variation of a boson is a fermion, whose vev vanishes by Lorentz invariance. Thus the condition for supersymmetry to be spontaneously broken is that there exists a fermion ψ_i such that

$$\langle \delta\psi_i \rangle = \langle \{Q_i, \psi_i\} \rangle \neq 0. \tag{9.7}$$

[10] We neglect the fermion bilinear term in F^i since $\langle \psi\psi \rangle = 0$ in the nlσm.

But in the nlσm (*i.e.* for χsf's) the supersymmetry variations of the component fields are

$$\begin{aligned} \delta\phi_i &\sim \psi_i, \qquad & \overline{\delta}\phi_i &= 0, \\ \delta\psi_i &\sim F_i, \qquad & \overline{\delta}\psi_i &\sim \partial\!\!\!/\phi_i, \end{aligned} \tag{9.8}$$

so supersymmetry breaks if and only if $\langle F_i\rangle \neq 0$. It is worth noting that $\langle\phi_i\rangle \neq 0$ does *not* break supersymmetry.

Thus, the fermion field(s) ψ_i which is not supersymmetry invariant in the ground state, is the superpartner of the non-vanishing F_i-component. This fermion, which shifts under a supersymmetry transformation, is necessarily massless, and is known as the "Goldstino" in analogy to the massless boson associated to a broken bosonic global symmetry.

To see this, note that in a supersymmetric field theory we can define a conserved spin-$\frac{3}{2}$ supersymmetry current j^μ_α, satisfying $\partial_\mu j^\mu_\alpha = 0$, and likewise a conserved symmetric energy-momentum tensor $t^{\mu\nu}$, giving the conserved charges as

$$Q_\alpha = \int d^3x\, j^0_\alpha(x), \qquad \overline{Q}_{\dot\alpha} = \int d^3x\, \overline{j}^0_{\dot\alpha}(x), \qquad P^\mu = \int d^3x\, t^{0\mu}(x). \tag{9.9}$$

The vacuum energy is then given by

$$E\eta^{\mu\nu} = \langle t^{\mu\nu}(0)\rangle = -\frac{1}{4}\overline{\sigma}^{\mu\alpha\dot\alpha}\langle\{Q_\alpha, \overline{j}^\nu_{\dot\alpha}(0)\}\rangle. \tag{9.10}$$

Thus in the case of broken supersymmetry, the supersymmetry current creates the non-invariant fermion from the vacuum. Just as in the usual proof of the Goldstone theorem, a non-zero value for this correlator implies a pole at zero momentum in the vacuum matrix element of j^0_α, thus implying the existence of a massless spinor in the spectrum, ψ_α.[11] The coupling of the supersymmetry current to the Goldstino is given by

$$\langle 0|j^\mu_\alpha|\overline{\psi}_{\dot\beta}\rangle = f\sigma^\mu_{\alpha\dot\beta}, \tag{9.12}$$

[11] The proof goes as follows:

$$\begin{aligned} E\eta^{\mu\nu} &= -\frac{1}{4}\overline{\sigma}^{\mu\alpha\dot\alpha}\langle\{Q_\alpha, \overline{j}^\nu_{\dot\alpha}(0)\}\rangle = -\frac{1}{4}\overline{\sigma}^{\mu\alpha\dot\alpha}\int d^3\mathbf{x}\langle\{j^0_\alpha(0,\mathbf{x}), \overline{j}^\nu_{\dot\alpha}(0,\mathbf{0})\}\rangle \\ &= -\frac{1}{4}\overline{\sigma}^{\mu\alpha\dot\alpha}\int d^4x\langle j^0_\alpha(t,\mathbf{x})\,\overline{j}^\nu_{\dot\alpha}(0,\mathbf{0})\,\delta(t) + \overline{j}^\nu_{\dot\alpha}(0,\mathbf{0})\,j^0_\alpha(t,\mathbf{x})\,\delta(t)\rangle \\ &= -\frac{1}{4}\overline{\sigma}^{\mu\alpha\dot\alpha}\int d^4x\frac{\partial}{\partial x^\rho}\langle j^\rho_\alpha(x)\,\overline{j}^\nu_{\dot\alpha}(0)\,\theta(t) - \overline{j}^\nu_{\dot\alpha}(0)\,j^\rho_\alpha(x)\,\theta(-t)\rangle \\ &= -\frac{1}{4}\overline{\sigma}^{\mu\alpha\dot\alpha}\int d^4x\frac{\partial}{\partial x^\rho}\langle\mathcal{T}j^\rho_\alpha(x)\overline{j}^\nu_{\dot\alpha}(0)\rangle = -\frac{1}{4}\overline{\sigma}^{\mu\alpha\dot\alpha}\int d^3\Omega\, R^3\hat{n}_\rho\langle\mathcal{T}j^\rho_\alpha(x)\overline{j}^\nu_{\dot\alpha}(0)\rangle\Big|_{|x|=R\to\infty} \\ &= -\frac{1}{4}\overline{\sigma}^{\mu\alpha\dot\alpha}\int d^3\Omega\, R^3\hat{n}_\rho\langle 0|j^\rho_\alpha|\overline{\psi}_{\dot\beta}\rangle\frac{\overline{\sigma}^{\sigma\dot\beta\beta}\hat{n}_\sigma}{R^3}\langle\psi_\beta|\overline{j}^\nu_{\dot\alpha}|0\rangle \propto |f|^2\overline{\sigma}^{\mu\alpha\dot\alpha}\eta_{\rho\sigma}\sigma^\rho_{\alpha\dot\beta}\overline{\sigma}^{\sigma\dot\beta\beta}\sigma^\nu_{\beta\dot\alpha} \propto |f|^2\eta^{\mu\nu}. \end{aligned} \tag{9.11}$$

which is the only coupling allowed by Lorentz invariance. The vacuum energy density is

$$E \propto |f|^2, \tag{9.13}$$

the square of the Goldstino decay constant, f, which is therefore proportional to F^i. (There is another coupling of a massless fermion to the supercurrent that we can write:

$$\langle 0|j^\mu_\alpha|\chi_\beta\rangle = gp^\mu \epsilon_{\alpha\beta}, \tag{9.14}$$

but this does not lead to a vanishing vacuum energy density.)

9.3. IR Subtleties

There are some subtle aspects of supersymmetry breaking which are worth mentioning, since they differ from the way usual (bosonic) global symmetries are broken. For ordinary symmetry-breaking, we need to go to the infinite-volume limit, otherwise QM tunnelling will mix all the degenerate vacua. In the infinite-volume limit, though, these vacua are "infinitely far apart" in the Hilbert space, and so do not mix. This is signalled by the fact that in the infinite-volume limit, in a vacuum breaking the symmetry, the charge operator does not exist as an operator on the Hilbert space—this is related to the divergence used to show the existence of the Goldstone boson. On the other hand, in the finite-volume limit (*e.g.* supersymmetric QM) we found that supersymmetry breaking can occur—there can be vacua with non-zero energy.

This fact gives a very different flavor to the question of whether supersymmetry is broken or not non-perturbatively. For ordinary symmetries, one often regularizes the theory by putting it in a box (finite volume) where the symmetry is never broken, then turning on a small breaking perturbation and seeing what happens as the volume is taken large and the perturbation taken to zero (in that order!). In this way it is fairly easy to tell if the symmetry *is* broken, though it can be quite subtle to be sure that it is *not* broken. In the supersymmetric case, if supersymmetry is not broken at finite volume, then it is not broken in the infinite-volume limit (the limit of zero energy in the infinite-volume limit is still zero); but if susy is broken at finite volume, one still needs to see whether or not it is restored in the infinite-volume limit. The net effect of this is to make it easier to tell if supersymmetry is *not* broken than it is to tell if supersymmetry *is* broken.

Put another way, if a theory has no massless fermions say at leading order in perturbation theory, then we can be sure that for some range of parameters no higher-order (or non-perturbative) effects will break supersymmetry, since there is no massless fermion available to be the Goldstino. The physically important case in which this argument does not work, however, is in chiral gauge theories, where mass terms for the chiral fermions are forbidden, and so are always "available" to play the role of the Goldstino. In this case an arbitrarily small perturbation or correction may conceivably break supersymmetry.

A second subtlety, related to the above, concerns the existence of the supersymmetry generator Q_α in theories with spontaneously broken supersymmetry. Recall that by the supersymmetry algebra, all finite (non-zero) energy states come in boson-fermion pairs (since on each energy eigenspace the supersymmetry algebra is a Clifford algebra whose only representation has an equal number of bosons and fermions). If this were true in QFTs with spontaneously broken supersymmetry, then we could rule out supersymmetry right now! Well, in the finite volume case, the boson-fermion pair is simply the ground state $|\Omega\rangle$ and the degenerate state $|\Omega + \{p_\mu{=}0 \text{ Goldstino}\}\rangle$; thus the Hilbert space is indeed supersymmetry-covariant. But in the infinite-volume limit, the zero-momentum Goldstino state is not defined (it is not normalizable), and so one of these states does not exist in the Hilbert space. This implies that the Q_α charge which related these two states must cease to exist as an operator in the infinite-volume limit, and the Hilbert space is not supersymmetry-covariant.

[This last point is important in understanding the possiblity of spontaneous *partial* breaking of supersymmetry. In extended supersymmetry, an argument goes that for one of the supersymmetries to be broken, the vacuum energy for those generators must be non-zero, and therefore all the supersymmetries will be broken. This is certainly true at finite volume, however the last paragraph shows a loop-hole in the infinite-volume limit: the broken supersymmetry generators cease to exist, and the extended supersymmetry algebra need no longer be satisfied. In fact, the IR divergence in Q_α is just right to allow a shift in the definition of the energy that the broken supersymmetry generators close on.]

9.4. Goldstinos again, mass sum rules, and mass splittings

In the nlσm, $\mathcal{L} = \int d^4\theta\mathcal{K} + \int d^2\theta\mathcal{W} + h.c.$, the scalar potential is given by

$$V(\phi^i, \overline{\phi}^{\bar{\imath}}) = \mathcal{D}_i\mathcal{W} g^{i\bar{\imath}} \overline{\mathcal{D}}_{\bar{\imath}} \overline{\mathcal{W}} \tag{9.15}$$

while the two-fermion terms are

$$\mathcal{L}_{\psi\psi} = -\tfrac{1}{2}\mathcal{D}_i\mathcal{D}_j\mathcal{W}\psi^i\psi^j + h.c. \tag{9.16}$$

Here $\mathcal{D}_i$ is the covariant derivative with respect to the Kahler metric on the target space. Thus, for example, $\mathcal{D}_i\mathcal{W} = \partial_i\mathcal{W}$, and $\mathcal{D}_i\mathcal{D}_j\mathcal{W} = \partial_i\partial_j\mathcal{W} - \Gamma^k_{ij}\partial_k\mathcal{W}$.

Evaluated at the vacuum, the last formula just gives the fermion mass terms

$$\mathcal{L}_{\psi\psi} \equiv -\tfrac{1}{2}m_{Fij}\psi^i\psi^j + h.c., \qquad m_{Fij} = \mathcal{D}_i\mathcal{D}_j\mathcal{W}, \tag{9.17}$$

in terms of which the fermion masses-squared are computed as the eigenvalues of the matrix

$$\mathcal{M}_F^2 = \begin{pmatrix} \overline{m}_{F\bar{\imath}\bar{k}} g^{k\bar{k}} m_{Fkj} & 0 \\ 0 & m_{Fik} g^{k\bar{k}} \overline{m}_{F\bar{k}\bar{\jmath}} \end{pmatrix}, \tag{9.18}$$

in a notation where each degree of freedom is treated separately. Thus, each eigenvalue appears twice, corresponding to the masses of the two physical degrees of freedom of a Weyl spinor.

The vacuum is at a minimum of the potential V: $\partial_i V = 0$.

▶ **Exercise 9.1.** Show that

$$\partial_i V = \mathcal{D}_i \mathcal{D}_j \mathcal{W} g^{j\bar{\jmath}} \overline{\mathcal{D}}_{\bar{\jmath}} \overline{\mathcal{W}}. \tag{9.19}$$

(Hint: show that $\partial_i g^{j\bar{\jmath}} = -g^{j\bar{k}} g^{k\bar{\jmath}} \partial_i g_{k\bar{k}}$.)

Thus, if supersymmetry is broken, so $V \neq 0$ at its minimum, then $\overline{\mathcal{D}}_{\bar{\jmath}} \overline{\mathcal{W}} \neq 0$, and we learn that $\mathcal{D}_i \mathcal{D}_j \mathcal{W} = m_{Fij}$ has a zero eigenvalue, and therefore that $\mathcal{M}_F^2$ has a pair of zero eigenvalues. This is just the Goldstino.

The boson mass terms are found by evaluating the second derivative of the potential at the vacuum:

$$\begin{aligned} \mathcal{L}_{\phi\phi} &= -\tfrac{1}{2} \partial_i \partial_j V \phi^i \phi^j - \partial_i \overline{\partial}_{\bar{\jmath}} V \phi^i \overline{\phi}^{\bar{\jmath}} - \tfrac{1}{2} \overline{\partial}_{\bar{\imath}} \overline{\partial}_{\bar{\jmath}} V \overline{\phi}^{\bar{\imath}} \overline{\phi}^{\bar{\jmath}}, \\ &\equiv -\tfrac{1}{2} \begin{pmatrix} \overline{\phi}^{\bar{\imath}} & \phi^i \end{pmatrix} \begin{pmatrix} m_{B\bar{\imath}j} & m_{B\bar{\imath}\bar{\jmath}} \\ m_{Bij} & m_{Bi\bar{\jmath}} \end{pmatrix} \begin{pmatrix} \phi^j \\ \overline{\phi}^{\bar{\jmath}} \end{pmatrix}. \end{aligned} \tag{9.20}$$

It is straightforward to compute:

$$\begin{aligned} m_{Bij} &= \mathcal{D}_i \mathcal{D}_j \mathcal{D}_k \mathcal{W}\, g^{k\bar{k}}\, \overline{\mathcal{D}}_{\bar{k}} \overline{\mathcal{W}}, \\ m_{Bi\bar{\jmath}} &= \mathcal{D}_i \mathcal{D}_k \mathcal{W}\, g^{k\bar{k}}\, \overline{\mathcal{D}}_{\bar{\jmath}} \overline{\mathcal{D}}_{\bar{k}} \overline{\mathcal{W}} = m_{Fik} g^{k\bar{k}} \overline{m}_{F\bar{k}\bar{\jmath}}, \end{aligned} \tag{9.21}$$

in terms of which the boson masses-squared are given by the eigenvalues of the matrix

$$\mathcal{M}_B^2 = \begin{pmatrix} m_{B\bar{\imath}j} & m_{B\bar{\imath}\bar{\jmath}} \\ m_{Bij} & m_{Bi\bar{\jmath}} \end{pmatrix}. \tag{9.22}$$

When supersymmetry is not broken, then $m_{Bij} = 0$ since it is proportional to $\overline{\mathcal{D}}_{\bar{k}} \overline{\mathcal{W}}$, and the boson mass-squared eigenvalues are equal in pairs to the eigenvalues of $m_{Bi\bar{\jmath}}$, which are precisely the eigenvalues of $\mathcal{M}_F^2$, the squares of the fermion masses.

If, on the other hand, supersymmetry is broken, then $m_{Bij} \propto \overline{\mathcal{D}}_{\bar{k}} \overline{\mathcal{W}} \neq 0$, so there will appear off-diagonal terms in the boson mass-squared matrix. This implies that the boson masses will be split from each other and from the fermions by (schematically)

$$\delta m_B^2 = \pm |m_{Bij}| \propto |\mathcal{D}^3 \mathcal{W}| \cdot |\mathcal{D}\mathcal{W}| \sim \lambda_{ijk} F^k, \tag{9.23}$$

where we have recognized $\mathcal{DW}$ as an F-component, and $\mathcal{D}^3\mathcal{W}$ as the dimensionless (marginal) coupling λ in the superpotential. This makes sense dimensionally, and physically, since F is the order parameter for supersymmetry breaking. We see that the effects of supersymmetry breaking are "transmitted" by fields with $\langle F^i\rangle \neq 0$ to the other fields.

Note also that these mass splittings satisfy the sum rule (S. Ferrara, L. Girardello, and F. Palumbo, *Phys. Rev.* D20 (1979) 403)

$$\mathrm{Tr}\mathcal{M}_B^2 - \mathrm{Tr}\mathcal{M}_F^2 = 0, \tag{9.24}$$

which is often written

$$\sum_i (-)^{2j_i}(2j_i+1)m_i^2 = 0, \tag{9.25}$$

where the sum is over all the (real) particles, and j_i is the spin and m_i the mass of the ith particle. The above sum rule is trivially satisfied when supersymmetry is unbroken, since the number of bosonic and fermionic degrees of freedom are equal, and their massess the same, in each multiplet. This sum rule is modified when vector multiplets (gauge fields) are included.

9.5. Effective actions of broken supersymetry

When we try to write an effective action at scales Λ for a theory in which supersymmetry is spontaneously broken at a scale $\widetilde{\Lambda} > \Lambda$, we will lose manifest supersymmetry, since the mass splittings within the multiplets will typically mean that some of the fields within a multiplet should be integrated-out while others remain dynamical. The question arises whether the resulting low-energy theory looks like a completely arbitrary non-supersymmetric effective theory, or whether some sign of the high-energy supersymmetry persists.[12]

We can address this question by simply not integrating-out the massive components of the supermultiplets, thus keeping a manifestly supersymmetric formalism. In particular, if supersymmetry is broken, then the F-components of some fields are getting vevs of order $\widetilde{\Lambda}^2$. So, for an effective action on scales smaller than $\widetilde{\Lambda}$, these F-components will not be dynamical. We can encode the effects of supersymmetry breaking by including these fields anyway in the effective action. Such fields are called *spurions*, and are simply chiral fields which are assumed to have a given $\langle F\rangle \neq 0$. In particular the dynamics of the spurion field which determines its vevs is not included in our effective action, thus we do not vary with respect to the F-components of the spurion fields to derive equations of motion.

[12] The Goldstino persists, but is "eaten" when one includes gravity. We will discuss this in more detail when we address supersymmetric phenomenology.

As a simple example, consider a theory of a single χsf. We can add in effective supersymmetry breaking "by hand" by coupling to a spurion χsf U, with $\langle U\rangle = \langle u\rangle + \theta^2\langle F_u\rangle$. Here we will also consider the scalar component's vev to be fixed by dynamics which we have integrated-out. So, for example, if we add to the superpotential the term

$$\int d^2\theta\, U\Phi^2 \supset 2\langle u\rangle(\phi F_\phi + \psi\psi) + \langle F_u\rangle\phi^2, \tag{9.26}$$

we see that we split the Φ chiral multiplet by giving different masses to ϕ and ψ.

In general, if we couple spurions to the other fields in the effective theory in a generic way, we will get generic non-supersymmetric low-energy physics. One has to make other physical assumptions about how the spurion fields couple to the low-energy physics to get constraints. We will discuss such assumptions in later lectures on supersymmetric phenomenology.

The low-energy theory with spurions may not satisfy the mass sum rules of the last section. This may be due to two things. First, some of the massive fields which have been integrated-out should be kept in the sum rule. This can typically account for errors in the sum rule $\mathcal{O}(\widetilde{\Lambda}^2)$, the scale the supersymmetry breaking. Larger errors are typically due to the second source, which is simply that there is no consistent high-energy nlσm physics that can give rise to the spurion vevs as assigned!)

9.6. Fermion condensates

The component expansion of a product chiral superfield $\Phi = \Phi_1\Phi_2$ is (schematically)

$$\Phi = \phi_1\phi_2 + \theta\,(\psi_1\phi_2 + \psi_2\phi_1) + \theta^2\,(\phi_1 F_2 + \phi_2 F_1 - \psi_1\psi_2)\,. \tag{9.27}$$

Thus, its F-term can acquire a non-zero expectation value, and supersymmetry spontaneously break, if the fermions were subject to a strong force, causing a condensate like $\langle\psi_1\psi_2\rangle$ to form. We have seen, however, that in the nlσm such condenstates do not form. In gauge theories, however, we might expect such condensates to form in analogy to chiral symmetry breaking in QCD, giving rise to *dynamical supersymmetry breaking* (= spontaneous supersymmetry breaking through a non-perturbative mechanism). We will hopefully start to address this possibility near the end of the course.

9.7. R-symmetries and genericity

There are some general statements that can be made about *when* supersymmetry can and cannot be broken in our theories. Suppose we have an effective description of our theory at some scale as a supersymmetric nlσm,[13] so supersymmetry is unbroken if and

[13] The following arguments won't change when gauge fields are added.

only if there exists a solution to $\partial_i \mathcal{W} = 0$. But these are n (complex) analytic equations in n (complex) unknowns (the vevs of the chiral fields ϕ^i), and so there will *generically* exist a solution. So we learn that for a generic superpotential, supersymmetry is unbroken.

It is a general working hypothesis that unless there are some symmetries to restrict the model, there will be generated by quantum corrections all possible terms in the effective action, and thus that the superpotential will be generic.

So, what if the superpotential is constrained by an ordinary (not R) global symmetry? Say it is a $U(1)$ symmetry with charges $Q(\Phi_i) = q_i$. The ground state may or may not spontaneously break this symmetry. If it does not, then the vevs of all the charged fields are zero. We can then reduce the question to whether there is a solution to $\partial_i \mathcal{W} = 0$ restricted to the submanifold where all charged bosons vanish. This just takes us back to the previous situation, and supersymmetry is not broken, generically. If, on the other hand, the $U(1)$ symmetry is spontaneously broken, then at least one of the charged fields will have a non-zero vev. Without loss of generality, we can take it to be Φ_1 and choose $q_1 = 1$. The superpotential must then have the form

$$\mathcal{W}(\Phi_1, \ldots, \Phi_n) = f(U_2, \ldots, U_n), \text{ where } U_i \equiv \Phi_i / \Phi_1^{q_i}. \tag{9.28}$$

(This should be taken as a local statement in the target space of the nlσm.) Then $\partial_i \mathcal{W} = 0$ if and only if $(\partial/\partial U_i)\mathcal{W} = 0$, giving $n-1$ equations for $n-1$ variables, and so for a generic superpotential again supersymmetry is unbroken. (Because $\langle \phi_1 \rangle \neq 0$, this change of variables is non-singular.)

▸ **Exercise 9.2.** Show that in a nlσm with non-Abelian global symmetries, and without an R-symmetry, supersymmetry is generically not broken.

Finally, suppose that there is an R-symmetry, with charges $R(\Phi_i) = r_i$. Again, if it is not spontaneously broken, then generically supersymmetry is not either. If it is, we can choose $r_1 = 1$ and $\langle \phi_1 \rangle \neq 0$, so that $\mathcal{W}$ can be written

$$\mathcal{W} = \Phi_1^2 f(U_2, \ldots, U_n), \text{ where } U_i \equiv \Phi_i / \Phi_1^{r_i}. \tag{9.29}$$

Then $\partial_i \mathcal{W} = 0$ is equivalent to the set of equations $\partial/\partial U_i f = 0$ as well as $f = 0$. These are now n equations for $n-1$ unknowns, and so typically have no solution. Thus generically supersymmetry *is* broken in this situation.

Our net result is that *if the superpotential is a generic function (constrained only by global symmetries) then supersymmetry is spontaneously broken if and only if there is a spontaneously broken R-symmetry.*

▸ **Exercise 9.3.** Show that a discrete symmetry does not help break supersymmetry generically.

▸ **Exercise 9.4.** Find examples with *non-generic* $\mathcal{W}$ such that:
(i) $U(1)_R$ is broken and supersymmetry is unbroken,

(ii) $U(1)_R$ is unbroken and supersymmetry is broken, and
(iii) there is no $U(1)_R$ but supersymmetry is broken.

This would seem to be bad news for supersymmetry phenomenology. For this result implies that along with supersymmetry breaking goes a Goldstone boson for the spontanously broken $U(1)_R$. Furthermore, we expect that global continuous symmetries (broken or not) do not exist with gravity.

This depends on the effective theory being generic. There is a set of measure zero in "theory space" which evades this problem. This way out, however, gives rise to a "naturalness" problem: it seems unnatural for the effective theory to have exactly the required special couplings with no symmetry reason to enforce them.

However, we learned in the last lecture that holomorphy of the superpotential (following from supersymmetry) puts extra constraints which imply the effective superpotential is *not* generic in this way. Indeed, the nlσm superpotential is unrenormalized, so if the microscopic theory is a special one evading the above result, then the effective theory will automatically be one also. This is often called "technical naturalness", which just means that the naturalness problem has just been shifted to a smaller scale (the microscopic theory) where we may not be able to address it. These issues of naturalness and supersymmetry breaking will be a theme of lectures to come.

10. Vector superfields and superQED

The non-perturbative results concerning the nlσm of the previous lectures may have seemed disappointing since, in the end, the IR phsyics of the nlσm was free. Luckily, supersymmetric field theory is richer than just chiral superfields. In particular, gauge fields appear in *vector superfields*. This lecture will focus on classical vector superfields and the effective actions describing their couplings to χsf's.

10.1. Abelian vector superfield (vsf)

A vsf V is a general scalar superfield satisfying a reality condition:

$$\begin{aligned} V &= \overline{V} \\ &= B + \theta\chi + \overline{\theta}\overline{\chi} + \theta^2 C + \overline{\theta}^2 \overline{C} - \theta\sigma^\mu\overline{\theta} A_\mu \\ &\quad + i\theta^2\overline{\theta}\left(\overline{\lambda} + \tfrac{1}{2}\overline{\sigma}^\mu\partial_\mu\chi\right) - i\overline{\theta}^2\theta\left(\lambda - \tfrac{1}{2}\sigma^\mu\partial_\mu\overline{\chi}\right) + \tfrac{1}{2}\theta^2\overline{\theta}^2\left(D + \partial^2 B\right), \end{aligned} \tag{10.1}$$

where B, D, and A_μ are real, and C complex. Since one component is a vector field, A_μ, we expect the interactions of this superfield to have a gauge invariance. The only supersymmetry covariant generalization of the usual $U(1)$ gauge invariance is

$$V \to V + i(\hat{\Lambda} - \overline{\hat{\Lambda}}), \tag{10.2}$$

where $\hat{\Lambda}$ is an arbitrary χsf with the usual component expansion

$$\hat{\Lambda} = \Lambda + \sqrt{2}\theta\psi_\Lambda + \theta^2 F_\Lambda + i\theta\sigma^\mu\overline{\theta}\partial_\mu\Lambda + \tfrac{i}{\sqrt{2}}\theta^2\overline{\theta}\overline{\sigma}^\mu\partial_\mu\psi_\Lambda + \tfrac{1}{4}\theta^2\overline{\theta}^2\partial^2\Lambda. \tag{10.3}$$

In components, the effect of the above gauge transformation is

$$\begin{aligned} \delta B &= i(\Lambda - \overline{\Lambda}) \\ \delta\chi &= i\sqrt{2}\psi_\Lambda \\ \delta C &= iF_\Lambda \\ \delta A_\mu &= \partial_\mu(\Lambda + \overline{\Lambda}) \\ \delta\lambda &= 0 \\ \delta D &= 0. \end{aligned} \tag{10.4}$$

We see that this correctly transforms A_μ with gauge parameter $\mathrm{Re}(\Lambda)$. Note however that other components transform with gauge parameters that depend also on $\mathrm{Im}(\Lambda)$. This means that the gauge invariance of vsf's is larger than just that of ordinary gauge fields. In this example, it is $U(1)_\mathbb{C}$ instead of the usual $U(1)_\mathbb{R}$. This is the general pattern: vsf's are invariant under the *complexification* $G_\mathbb{C}$ of the gauge group G. B, χ, and C are all gauge artifacts, and λ and D are gauge invariant. A_μ is the photon, the Weyl fermion λ is called the "photino", and D is an auxiliary field.

10.2. Wess-Zumino (WZ) gauge

Indeed, we can fix to the *Wess-Zumino gauge* where this is apparent, though at the cost of breaking manifest supersymmetry:

$$B = \chi = C = 0. \tag{10.5}$$

In this gauge we easily see that

$$\begin{aligned} V &= -\theta\sigma^\mu\overline{\theta}A_\mu + i\theta^2\overline{\theta\lambda} - i\overline{\theta}^2\theta\lambda + \tfrac{1}{2}\theta^2\overline{\theta}^2 D, \\ V^2 &= -\tfrac{1}{2}\theta^2\overline{\theta}^2 A_\mu A^\mu, \\ V^3 &= 0, \end{aligned} \tag{10.6}$$

so

$$e^V = 1 + V - \tfrac{1}{4}\theta^2\overline{\theta}^2 A_\mu A^\mu \tag{10.7}$$

in WZ gauge.

Note that WZ gauge does not completely fix the gauge—indeed, the usual $U(1)$ gauge invariance is still left. Thus WZ gauge fixes the supersymmetric $U(1)_\mathbb{C}$ gauge invariance to $U(1)_\mathbb{R}$.

10.3. Field-strength χsf

The field strength appears in the superfield

$$W_\alpha = -\tfrac{1}{4}\overline{D}^2 D_\alpha V, \qquad \overline{W}_{\dot\alpha} = -\tfrac{1}{4}D^2\overline{D}_{\dot\alpha}V. \tag{10.8}$$

One can easily check that W_α is gauge invariant:

$$\begin{aligned}\delta W_\alpha &\propto \overline{D}^2 D_\alpha(\hat\Lambda - \overline{\hat\Lambda}) = \overline{D}^2 D_\alpha \hat\Lambda = \overline{D}_{\dot\alpha}\{\overline{D}^{\dot\alpha}, D_\alpha\}\hat\Lambda \\ &\propto \overline{D}_{\dot\alpha}\partial^{\dot\alpha\alpha}\hat\Lambda = \partial^{\dot\alpha\alpha}\overline{D}_{\dot\alpha}\hat\Lambda = 0.\end{aligned} \tag{10.9}$$

Note also that W_α is a χsf:

$$\overline{D}_{\dot\alpha}W_\alpha = D_\alpha \overline{W}_{\dot\alpha} = 0. \tag{10.10}$$

In WZ gauge,

$$W_\alpha = -i\lambda_\alpha(y) + \left[\delta_\alpha^\beta D - \tfrac{i}{2}(\sigma^\mu\overline{\sigma}^\nu)_\alpha^\beta F_{\mu\nu}\right]\theta_\beta + (\sigma^\mu\partial_\mu\overline{\lambda})_\alpha\theta^2, \tag{10.11}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Note that $\sigma^\mu\overline{\sigma}^\nu F_{\mu\nu}$ projects onto the self-dual $(1,0)$ part F^+ of F. Since $\overline{F^+} = F^-$, $\overline{W}_{\dot\alpha}$ contains F^-. Recall that $F^\pm = F \pm i\widetilde{F}$ where $\widetilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$, so $\widetilde{\widetilde{F}} = -F$.

Finally, the superspace version of the Bianchi identity is

$$D^\alpha W_\alpha - \overline{D}_{\dot\alpha}\overline{W}^{\dot\alpha} = 0. \tag{10.12}$$

10.4. Pure $U(1)$ gauge theory

The superspace analog of $\mathcal{L} = \frac{1}{4g^2}F^{\mu\nu}F_{\mu\nu}$ is

$$\mathcal{L} = \int d^2\theta\, \frac{-i\tau}{16\pi} W^\alpha W_\alpha + h.c., \tag{10.13}$$

where

$$\tau \equiv \frac{\vartheta}{2\pi} + i\frac{4\pi}{g^2}, \tag{10.14}$$

and the integral is over only half of superspace since W^2 is chiral.[14] Here ϑ is the "theta-angle", which will play an important role quantumly.

[14] From the definition of W_α in terms of V, one can express $W^2 = \overline{D}^2(WDV)$, implying that the gauge kinetic term can also be written as an integral over all of superspace: $\int d^4\theta(WDV + \overline{W}\overline{D}V)$. However, the integrand in this expression is gauge-variant, so it is inconvenient to use in computing effective actions.

▸ **Exercise 10.1.** Show that in components in WZ gauge,

$$\int d^2\theta \, \frac{-i\tau}{16\pi} W^\alpha W_\alpha + h.c., = -\frac{1}{4g^2} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2g^2} D^2 - \frac{i}{g^2} \lambda \partial\!\!\!/ \overline{\lambda} + \frac{\vartheta}{32\pi^2} F_{\mu\nu} \widetilde{F}^{\mu\nu}. \tag{10.15}$$

There is one other gauge-invariant term which can be formed from a vsf, the *Fayet-Iliopoulos* term:

$$\delta\mathcal{L} = \int d^4\theta (2\kappa V) = \kappa D, \tag{10.16}$$

in WZ gauge. This is gauge-invariant only for Abelian gauge theories, and when κ is a real constant. This term can be written in a manifestly gauge-invariant way as an integral over a*quarter* of superspace:

$$\int d^4\theta V \sim \int d\overline{\theta}^2 d\theta^\alpha D_\alpha V \sim \int d\theta^\alpha \overline{D}^2 D_\alpha V \sim \int d\theta^\alpha W_\alpha. \tag{10.17}$$

Thus the FI term can be written as

$$\mathcal{L}_{FI} = \tfrac{1}{2} \int d\theta^\alpha \kappa W_\alpha + h.c. \tag{10.18}$$

Here we can take κ to be a complex number; only its real part enters by virtue of the Bianchi identity, which implies that $\int d\theta^\alpha W_\alpha = \int d\overline{\theta}_{\dot\alpha} \overline{W}^{\dot\alpha}$.

10.5. Coupling to χsf's—supersymmetric QED

Under a gauge transformation vsf and χsf's transform as

$$\begin{aligned} V &\to V + i(\hat\Lambda - \overline{\hat\Lambda}) \\ \Phi^i &\to e^{-iq_i\hat\Lambda}\Phi^i, \end{aligned} \tag{10.19}$$

where q_i is the charge of the Φ^i χsf. Note that since $\hat\Lambda$ is complex, the bigger gauge freedom, namely $U(1)_{\mathbb{C}}$ rather than $U(1)_{\mathbb{R}}$, is manifest. The Lagrangian for SQED is then, in addition to the gauge kinetic and FI terms,

$$\mathcal{L} = \int d^2\theta \, \mathcal{W}(\Phi^i) + h.c. + \sum_i \int d^4\theta \, \overline{\Phi}_i e^{q_i V} \Phi^i, \tag{10.20}$$

where the superpotential is restricted to contain only gauge-invariant interactions.

There is another constraint on the possible Lagrangians coming from *anomalies.* As we will discuss in a few lectures, this theory is inconsistent unless the charges, q_i, of the χsf's satisfy

$$\sum_i q_i = \sum_i q_i^3 = 0. \tag{10.21}$$

The first constraint is from the mixed gauge-gravitational anomaly, and the second is from the pure gauge anomaly. These constaints can always be satisfied by pairs of χsf's with opposite charges. Non-trivial solutions exist for five or more charged χsf's (in fact there is a continuum of solutions). Non-trivial solutions with commensurate charges are harder to find. One such has fifteen chiral fields, corresponding to the hypercharge assignments of one generation of the standard model. (I think there is probably a smaller non-trivial commensurate solution with just eight charges.)

The expansion of the kinetic terms (in WZ gauge) gives the usual minimal coupling:

$$\int d^4\theta \overline{\Phi} e^{qV} \Phi = \overline{F}F - \left|(\partial_\mu + \tfrac{i}{2}qA_\mu)\phi\right|^2 - i\overline{\psi}\overline{\sigma}(\partial_\mu + \tfrac{i}{2}qA_\mu)\psi - \tfrac{i}{\sqrt{2}}q(\phi\overline{\lambda\psi} - \overline{\phi}\lambda\psi) + \tfrac{1}{2}qD\overline{\phi}\phi. \tag{10.22}$$

Thus the terms involving the auxiliary fields F^i and D are

$$\mathcal{L} \supset \frac{1}{2g^2}D^2 + \kappa D + \tfrac{1}{2}D\sum_i q_i\overline{\phi}_i\phi^i + \sum_i \overline{F}_i F^i + \sum_i F_i \frac{\partial \mathcal{W}}{\partial \phi^i} + h.c. \tag{10.23}$$

Integrating-out D and F^i gives the scalar potential

$$-\mathcal{L} \supset V(\overline{\phi}_i, \phi^i) = \sum_i \left|\frac{\partial \mathcal{W}}{\partial \phi^i}\right|^2 + \frac{g^2}{4}\left(2\kappa + \sum_i q_i \left|\phi^i\right|^2\right)^2, \tag{10.24}$$

again in WZ gauge.

This immediately implies that supersymmetry is unbroken if and only if the F and D-terms vanish:

$$\begin{aligned} 0 = \overline{F}_i &= -\frac{\partial \mathcal{W}}{\partial \phi^i}, \qquad \forall i, \\ 0 = D &= -\frac{g^2}{2}\left(2\kappa + \sum_i q_i|\phi^i|^2\right). \end{aligned} \tag{10.25}$$

I will refer to these equations as the *vacuum equations*. It is worth emphasizing that these vacuum equations are only valid in WZ gauge. It is not too hard to derive the scalar potential without gauge-fixing:

▸ **Exercise 10.2.** Show, by integrating-out D and F^i without gauge fixing, that the scalar potential in SQED is

$$V = \sum_i e^{-q_i B}\left|\partial_i \mathcal{W}\right|^2 + \tfrac{g^2}{4}\left(2\kappa + \textstyle\sum_i q_i e^{q_i B}\left|\phi^i\right|^2\right)^2 + C\left(\textstyle\sum_i q_i \phi^i \partial_i \mathcal{W}\right) + h.c. \tag{10.26}$$

(*Hint:* keep only the scalar fields with no derivatives and show first that then $e^{qV} = e^{qB}[1 + q\theta^2 C + q\overline{\theta}^2\overline{C} + \tfrac{1}{2}\theta^2\overline{\theta}^2(qD + q^2 C\overline{C})]$.)

Note that the last terms vanish, since by gauge invariance $\sum_i q_i\phi^i\partial_i\mathcal{W} = 0$. This more general expression will be useful when we come to solving the D-term constraints. It is also useful when the gauge invariance is spontaneously broken, in which case in unitary gauge B is a physical (propagating) field.

10.6. General Abelian gauged nlσm

We can write the most general Abelian gauged nlσm:

$$\begin{aligned}\mathcal{L} = &\int d^4\theta \left[\mathcal{K}\left(\overline{\Phi}^{\bar{\imath}} e^{-q_{ai}V^a}, \Phi^i\right) + 2\kappa_a V^a\right] \\ &+ \int d^2\theta \left[\mathcal{W}(\Phi^i) + \frac{1}{16\pi i}\tau_{ab}(\Phi^i) W^a W^b\right] + h.c.\end{aligned} \tag{10.27}$$

where $\mathcal{K}$ is a general gauge invariant Kahler potential, κ_a are real numbers, $\mathcal{W}$ is a general gauge invariant superpotential, and τ_{ab} are generalized gauge couplings and theta-angles,

$$\tau_{ab} = \frac{\vartheta_{ab}}{2\pi} + i\frac{4\pi}{(g^2)^{ab}}, \tag{10.28}$$

which can depend only on χsf's. The i and $\bar{\imath}$ indices run over the different chiral multiplets, while the a, b indices run over the different $U(1)$ vector multiplets. The anomaly cancellation conditions require the charges to satisfy

$$\sum_i q_{ai} = \sum_i q_{ai} q_{bi} q_{ci} = 0, \qquad \text{for all } a, b, c. \tag{10.29}$$

The component expansion of this SQED lagrangian gives rise to the scalar potential

$$\begin{aligned}V = &\sum_i e^{-q_{ia}B^a} \left|\partial_i \mathcal{W}\right|^2 \\ &+ \tfrac{1}{4}(g^2)^{ab} \left(2\kappa_a + \textstyle\sum_i q_{ia} e^{q_{ic}B^c} |\phi^i|^2\right) \left(2\kappa_b + \textstyle\sum_i q_{ib} e^{q_{ic}B^c} |\phi^i|^2\right).\end{aligned} \tag{10.30}$$

Unitarity requires the symmetric coupling matrix $(g^2)^{ab}$ to be positive definite, implying the vacuum energy vanishes and supersymmetry is unbroken if and only if the vacuum equations[15]

$$\begin{aligned}0 &= \partial_i \mathcal{W}, && \forall i \quad \text{“F-terms”} \\ 0 &= 2\kappa_a + \sum_i q_{ia} e^{q_{ic}B^c} |\phi^i|^2, && \forall a \quad \text{“D-terms”}\end{aligned} \tag{10.31}$$

are satisfied. Here B^a is the lowest component of the V^a vsf. It is gauge-variant; in WZ gauge we set $B^a = 0$, giving rise to the “usual” vacuum equations.

[15] More properly, the χsf indices i should be raised and lowered with the Kahler metric $g^{i\bar{\imath}}$.

10.7. Non-Abelian gauge theories

Suppose we have a unitary non-Abelian gauge group G with hermitian generators T^a satisfying the algebra

$$[T^a, T^b] = if_c^{ab} T^c. \tag{10.32}$$

Linear representations of this algebra realize the generators as matrices acting on some representation space. If necessary, we will denote the representation by a subscript, *e.g.* T_R^a for the R representation.

We define the gauge parameter

$$\Lambda = T^a \Lambda_a, \tag{10.33}$$

By analogy with the Abelian case, we promote this to a χsf, $\hat{\Lambda}$. The lowest component of $\hat{\Lambda}$ is a complex field, so we have again enlarged the gauge symmetry from G to $G_{\mathbb{C}}$. The gauge transformation rule of a χsf Φ_r in the representation r of G is

$$\Phi_r \to e^{-i\hat{\Lambda}_r} \Phi_r, \tag{10.34}$$

where $\hat{\Lambda}$, of course, is also in the r representation.

A non-Abelian vsf is defined in the same way as

$$V = T^a V_a, \tag{10.35}$$

where the V_a are hermitian. Guessing that the Kahler terms should remain of the same form as in the Abelian case, $\mathcal{K} = \overline{\Phi}_r e^{V_r} \Phi_r$, we get the gauge transformation rule for the vsf as

$$e^V \to e^{-i\overline{\hat{\Lambda}}} e^V e^{i\hat{\Lambda}}. \tag{10.36}$$

Expanding this out to leading order gives

$$V \to V + i(\hat{\Lambda} - \overline{\hat{\Lambda}}) + \ldots, \tag{10.37}$$

implying that an analog of WZ gauge exists for non-Abelian vsf's,

$$\begin{aligned} V &= -\theta\sigma^\mu\overline{\theta} A_\mu + i\theta^2\overline{\theta\lambda} - i\overline{\theta}^2\theta\lambda + \tfrac{1}{2}\theta^2\overline{\theta}^2 D, \\ V^2 &= -\tfrac{1}{2}\theta^2\overline{\theta}^2 A_\mu A^\mu, \\ V^3 &= 0, \end{aligned} \tag{10.38}$$

where all the components are matrix-valued fields in some representation of G, and the gauge parameter is determined up to a single hermitian scalar part

$$\hat{\Lambda} = \mathrm{Re}\Lambda + i\theta\sigma^\mu\overline{\theta}\partial_\mu(\mathrm{Re}\Lambda) + \tfrac{1}{4}\theta^2\overline{\theta}^2\partial^2(\mathrm{Re}\Lambda). \tag{10.39}$$

▸ **Exercise 10.3.** Show in WZ gauge that the non-Abelian gauge transformation $e^V \to e^{-i\overline{\hat{\Lambda}}} e^V e^{i\hat{\Lambda}}$ gives the usual gauge transformation rule for A_a^μ, and gauge-covariant transformation rules for the other components λ_a and D_a.

The field-strength χsf is defined as

$$W_\alpha = -\frac{1}{4}\overline{D}^2 e^{-V} D_\alpha e^V . \tag{10.40}$$

It transforms inthe adjoint representation of the gauge group.

▸ **Exercise 10.4.** Check that W_α transforms covariantly, $W_\alpha \to e^{-i\overline{\hat{\Lambda}}} W_\alpha e^{i\hat{\Lambda}}$, under gauge transformations.

▸ **Exercise 10.5.** Show that $\mathcal{D}_\alpha \equiv e^{-V} D_\alpha e^V$ is a gauge-covariant super-covariant derivative.

The Bianchi identity is just as in the Abelian case, except it involves the gauge-covariant derivatives.

We can write the general gauged nlσm (with simple gauge group) as

$$\mathcal{L} = \int d^4\theta\, \mathcal{K}\left(\overline{\Phi}^{\bar{\imath}} e^{V_{R_i}}, \Phi^i\right) + \int d^2\theta \left[\mathcal{W}(\Phi^i) + \frac{1}{32\pi i}\tau(\Phi^i)\,\mathrm{tr}_f W_\alpha^2\right] + h.c. \tag{10.41}$$

where all the symbols have the same meaning as in the Abelian gauged nlσm. (V_{R_i} denotes V in the R_i representation of the gauge group, and tr_f denotes a trace in the fundamental representation; we'll discuss this normalization in detail in lecture 13.) Note that the FI term does not appear, since it is not allowed by gauge invariance.

The scalar potential from expanding this action in components is once again a sum of squares of F and D-terms. Since the F terms do not involve the vsf's, they are just the same as in the nlσm. It is easy to check that the D-terms are (in WZ gauge)

$$D^a = \sum_i \overline{\phi}_i T^a_{R_i} \phi^i, \tag{10.42}$$

where I have assumed canonical Kahler terms (otherwise there would be a factor of the Kahler metric lowering the i index on $\overline{\phi}$).

11. Supersymmetry and gauge symmetry breaking

In this lecture we will explore in a qualitative way the classical low-energy physics of the classical gauged nlσm. This analysis is only expected to be accurate in the case the gauge dynamics are IR-free, which occurs for Abelian gauge groups and for non-Abelian gauge groups with "enough" matter.

11.1. Supersymmetry breaking generalities

We saw in the last lecture that the scalar potential was the sum of squares of the D as well as the F terms. We thus have an extra condition to satisfy for there to be a supersymmetric vacuum compared to the nlσm case. If the F term conditions cannot be satisfied by themselves, then, just as in the un-gauged nlσm, supersymmetry will be broken. This kind of breaking is called "F-term" or sometimes "O'Raifeartaigh" breaking, and its systematics are just as we discussed for the nlσm. It will turn out that if the F term conditions have a solution, then the D term conditions will always also have a solution *if there are no FI terms.* Thus the FI terms play a special role in the discussion of supersymmetry breaking. Breaking due to them is called "D-term" or sometimes "Fayet-Iliopoulos" breaking. Recall that FI terms are only allowed for Abelian gauge groups.

In D-term breaking, the non-zero vev of a D component is the order parameter for supersymmetry breaking. Thus, in particular, the scale of supersymmetry breaking, or equivalently, the scale of the mass-splittings within multiplets, is given by

$$\delta m^2 \sim g^2 D. \tag{11.1}$$

The factor of the gauge coupling enters since the D term is coupled to the other fields in the theory as part of the gauge multiplet. Furthermore, since the D-term breaking only occurs when there is an FI term, one expects the D vev to be proportional to κ, the FI constant.

The mass sum rule that we worked out for the nlσm also holds in the gauged case. When there is no gauge symmetry breaking and we have canonical (quadratic) Kahler terms, this is straight-forward to see. The only unusual term comes from the $D^a q_{ai}\phi^i\overline{\phi}_i$ term that appears in the component expansion of the Kahler terms. Upon solving for the auxiliary D fields, and when there is an FI term, this gives rise to an "extra" mass term $\sim (g^2)^{ab}\kappa_a q_{ib}|\phi^i|^2$ for the χsf bosons. In the mass-squared trace formula this term vanishes, however, since it is proportional to $\sum_i q_{ia} = 0$ by the mixed gauge-gravitational anomaly cancellation. Thus we have

$$\sum_j m_j^2 (-)^{2J_j}(2J_j + 1) = 0, \tag{11.2}$$

where the sum is over all physical particles j with mass m_j and spin J_j. In the case of spontaneous gauge symmetry breaking, it is a little more involved to derive this formula, since one then has to take into account the vector boson mass and the mass terms mixing the gaugino with the χsf fermions, but the idea is straight-forward; see S. Ferrara, L. Girardello, and F. Palumbo, *Phys. Rev.* D20 (1979) 403.

Note that since the supersymmetry variation of the gaugino is

$$\delta_\alpha \lambda_\beta = \{Q_\alpha, \lambda_\beta\} \sim \epsilon_{\alpha\beta} D + \sigma^{\mu\nu}_{\alpha\beta} F^+_{\mu\nu}, \tag{11.3}$$

(which follows simply if you recall that λ is the lowest component of the field-strength χsf) then in the case of D-term breaking the gaugino is shifted, and so is identified with the Goldstino.

Finally, it is worth mentioning that, although condensates of χsf fermions can generate (F-term) supersymmetry-breaking, gaugino condensates do *not* lead to supersymmetry breaking since they are the lowest components of χsf's.[16]

11.2. Generic solutions of the vacuum equations

Note that if the superpotential were generic, then the F-terms in (10.25) give n complex analytic equations for n complex unknowns, and so would typically have a solution. However, $\mathcal{W}$ is subject to one constraint—gauge invariance—which may reduce the number of independent equations by one. If the gauge symmetry is not broken (so no charged χsf's get vevs), then the D-term equation will be satisfied if and only if $\kappa = 0$. Thus the FI term *generically* leads to broken supersymmetry when the F-terms do not break the gauge invariance by themselves.

When the F-term solution breaks the gauge symmetry, then, as we saw in lecture 9, the F-terms are equivalent to $n-1$ equations for the $n-1$ unknowns $u_i = \phi_i/\phi_1^{q_i}$, and thus will typically have a one-complex-dimensional space of solutions. However one real dimension of this space is a gauge artifact: the phases of the ϕ_i are unobservable by gauge invariance. There remains one real dimension, plus the single real D-term equation. But since this last equation is a *real* equation, one can not predict the generic existence of solutions. We will show in the next lecture the general result that *in the absence of FI terms and when the F-term equations have a solution, then there always exists a simultaneous solution to the D-term vacuum equation.* Thus, in the presence of gauge symmetry breaking, a generic superpotential with no FI term will lead to a unique supersymmetric vacuum.

A simple example is a theory with four charged χsf's $\Phi_{\pm 1}$ and $\Phi_{\pm 2}$ where their charges are given by their subscripts. Say the F-terms break gauge invariance with $\phi_{-1} \neq 0$ and the solution $u_1 \equiv \phi_1\phi_{-1} = 0$, $u_2 \equiv \phi_2\phi_{-1}^2 = 0$, and $u_3 = \phi_{-2}\phi_{-1}^{-2} = 1$. Then solving the D-term gives $4|\langle\phi_{-1}\rangle|^2 = -1 \pm \sqrt{1+8\kappa}$. Clearly, for the whole range $\kappa < 0$ there is no solution.[17]

Finally, if there is a spontaneously broken $U(1)_R$ symmetry, then, just as in the last lecture, supersymmetry will be generically spontaneously broken—the D term just adds an additional constraint to the F-term equations.

[16] In supergravity theory this is no longer true. We will discuss the differences between supergravity and supersymmetry in the lectures on phenomenology.

[17] There is no solution in this case for $\kappa = 0$ either, since that would imply $\langle\phi_{-1}\rangle = 0$, which is not consistent with the F-term solution. This is not a generic situation, however. For example,

11.3. Higgsing and unitary gauge

When a charged χsf gets a non-zero vev, the gauge symmetry is spontaneously broken. As in the usual Higgs mechanism, gauge bosons become massive, “eating” neutral scalars. We can see this simply in a model with two charged χsf’s, $\Phi_\pm$ with charges ± 1 (they must have opposite charges by anomaly cancellation), no FI term, and the superpotential

$$\mathcal{W} = -m\Phi_+\Phi_- + \frac{1}{2m}\Phi_+^2\Phi_-^2, \qquad m \in \mathbb{R}^+. \tag{11.4}$$

Then in WZ gauge the F-term conditions for a supersymmetric minimum are satisfied by either $\phi_+ = \phi_- = 0$ or $\phi_+\phi_- = m^2$. The first solution does not interest us since it does not break the gauge symmetry. The D-term condition, $|\phi_+|^2 - |\phi_-|^2 = 0$ implies that $\phi_+ = e^{i\alpha}\phi_-$ for some angle α, thus giving the supersymmetric but non-gauge invariant vacua

$$\phi_+ = me^{+i\alpha/2}, \qquad \phi_- = me^{-i\alpha/2}. \tag{11.5}$$

These vacua are related by the gauge invariance:

$$\phi_+ \simeq \phi_+ e^{+i\beta}, \qquad \phi_- \simeq \phi_- e^{-i\beta}, \tag{11.6}$$

so we can choose $\beta = -\alpha/2$ to find only one vacuum, say, $\phi_+ = \phi_- = m$.

However, in WZ gauge it is hard to see the physical field content. So instead, let us go to *unitary gauge*, in which we fix the whole $U(1)_{\mathbb{C}}$ gauge invariance by rotating the *whole* χ*sf*

$$\Phi_+ \to m. \tag{11.7}$$

We are free to do this so long as $\langle\Phi_+\rangle \neq 0$. We have chosen m as a convenient value—we could just as well have chosen any non-zero complex number.

In this gauge

$$\begin{aligned}\mathcal{L} &= \int d^4\theta\, \left(e^V m^2 + \overline{\Phi}_- e^{-V}\Phi_-\right) + \int d^2\theta\, \left(\frac{\tau}{16\pi i}W^2 - m^2\Phi_- + \frac{m}{2}\Phi_-^2\right) + h.c. \\ &= \int d^4\theta \left(\overline{\widetilde{\Phi}} e^{-V}\widetilde{\Phi} + \overline{\widetilde{\Phi}} e^{-V} m + m e^{-V}\widetilde{\Phi}\right) + \int d^2\theta \frac{\tau}{16\pi i}W^2 + h.c. \\ &\quad + \int d^4\theta\, m^2\left(e^V + e^{-V}\right) + \int d^2\theta\, \frac{m}{2}\widetilde{\Phi}^2 + h.c.\end{aligned} \tag{11.8}$$

---

if we turned on an arbitrarily small vev for one of the other fields, say $u_1 = \epsilon$, then it is easy to that there are no solutions only for $\kappa < |\epsilon|$. In general, it would seem that such non-generic supersymmetry breaking arises when only fields with like-sign charges get vevs, which in turn, by gauge invariance, depends on fields appearing with negative exponents in the superpotential. We will see examples of how such terms can arise when we study strongly-coupled gauge theories later in the course.

where we have shifted $\widetilde{\Phi} \equiv \Phi_- - m$. The first line in the last expression contains the kinetic terms for V and $\widetilde{\Phi}$. The second line contains the terms

$$\int d^4\theta \, m^2 \left(e^V + e^{-V}\right) \supset \int d^4\theta \, m^2 V^2$$
$$\supset -m^2 \left(\tfrac{1}{2}A_\mu A^\mu + \tfrac{1}{2}\partial_\mu B \partial^\mu B - BD - \tfrac{1}{2}C\overline{C} + i\overline{\chi}\partial\!\!\!/\chi + i\chi\lambda - i\overline{\chi}\overline{\lambda}\right). \tag{11.9}$$

giving the vector boson a mass and making B and χ dynamical. Thus $\widetilde{\Phi}$ plays the role of the Higgs boson in this example.

▸ **Exercise 11.1.** Solve for the vacua and spectrum of the *Fayet-Iliopoulis model*

$$\mathcal{L} = \int d^4\theta \left(\overline{\Phi}_- e^{-V}\Phi_- + \overline{\Phi}_+ e^{+V}\Phi_+ + \kappa V\right) + \int d^2\theta \left(\frac{1}{4g^2}W^2 + m\Phi_+\Phi_-\right) + h.c. \tag{11.10}$$

as a function of its parameters g, κ, and m.

12. D-terms and Higgs branches

In what follows we will set to zero the FI terms and the superpotential, thus $\kappa_a = \mathcal{W} = 0$. We will see that the resulting D-term constraints always have flat directions—whole moduli spaces of solutions. We start by looking at a simple example.

12.1. Example 1

Consider a $U(1)$ theory with two χsf's $\Phi_\pm$ of charges ± 1. Then the D-term constraint is in WZ gauge

$$0 = |\phi_+|^2 - |\phi_-|^2, \tag{12.1}$$

implying

$$\phi_+ = e^{i\alpha}\phi_- \tag{12.2}$$

for some angle α. The resulting 3-real-dimensional space of vacua $\{\phi_+, \alpha\}$ must be divided by the $U(1)_{\mathbb{R}}$ gauge equivalence which remains in WZ gauge:

$$\phi_+ \simeq e^{+i\beta}\phi_+, \qquad \phi_- \simeq e^{-i\beta}\phi_-, \tag{12.3}$$

for β any real angle. We can use this gauge freedom to fix the angle α, by choosing $\beta = \alpha/2$ so that $\phi_+ = \phi_-$. Actually, this choice does not completely fix the gauge freedom, since $\beta = \pi + \alpha/2$ would have sone just as well. Thus the moduli space can be described as

$$\mathcal{M} = \{\phi_+\}/\{\phi_+ \to -\phi_+\}, \tag{12.4}$$

which means the space of all ϕ_+ quotient the (residual gauge) identification of ϕ_+ with $-\phi_+$.

This space can be conveniently parametrized in terms of the gauge-invariant variable

$$M \equiv \phi_+\phi_-. \tag{12.5}$$

M is a good coordinate on $\mathcal{M}$ since every ϕ_+ gives rise to a unique M (since $\phi_+ = \phi_-$), while every value of M dtermines a ϕ_+ up to a sign. Thus

$$\mathcal{M} = \{M\}. \tag{12.6}$$

So, topologically, the moduli space $\mathcal{M} \simeq \mathbb{C}$.

Metrically, however, it has a singularity:

$$\mathcal{K} = \overline{\phi}_+\phi_+ + \overline{\phi}_-\phi_- = 2\overline{\phi}_+\phi_+ = 2\sqrt{M\overline{M}}, \tag{12.7}$$

so the metric is

$$ds^2 = \frac{1}{2}\frac{dM\, d\overline{M}}{\sqrt{M\overline{M}}}. \tag{12.8}$$

Thus there is a metric singularity at $M = 0$, which corresponds to $\phi_\pm = 0$, where the $U(1)$ gauge symmetry is not spontaneously broken, and so the vsf is massless there. This is another example of singularities in moduli space corresponding to new massless physics. The metric $(ds)^2$ is flat everywhere except at the origin, where it has a $(\mathbb{Z}_2)$ conical singularity. Thus metrically the moduli space is

$$\mathcal{M} = \mathbb{C}/\mathbb{Z}_2. \tag{12.9}$$

Note that this is only a classical equivalence. Quantumly, the Kahler potential gets corrections, and changes the metric structure of the moduli space. We will return to the issue of quantum corrections later.

12.2. Solving the D-term constraints

More generally, the moduli space of a theory with no superpotential is given by the space of all scalar vevs satisfying the D-term constraints, modulo gauge equivalences:

$$\mathcal{M} = \{\phi^i | D^a = 0\}/G. \tag{12.10}$$

I claim this space is equivalent to

$$\mathcal{M} = \{\phi^i\}/G_{\mathbb{C}}, \tag{12.11}$$

the space of all scalar vevs of the χsf's modulo *complexified* gauge transformations—*i.e.* the D-terms are just a reflection of the larger $G_{\mathbb{C}}$ gauge invariance that we have seen necessarily appears in supersymmetric gauge theory with gauge group G.

We can see this explicitly in the Abelian gauged nlσm as follows. Recall that without gauge fixing, in addition to the scalar fields from the χsf's the real scalar components B^a of the vsf's also appears in the potential, giving rise to the D-term conditions

$$0 = \sum_i q_{ia} e^{q_{ib} B^b} |\phi^i|^2, \qquad \forall a \tag{12.12}$$

(recall that we are assuming that there are no FI terms). We will now show that these equations can always be satisfied by adjusting only the B^a's, for arbitrary ϕ^i's. Consider first a single Abelian gauge field, so that there is only one D-term equation and one B field. Then the right-hand side of (12.12) is positive for $B \to +\infty$, since (by anomaly cancellation) the greatest q_i, call it q_{max}, is postive and so as $B \to +\infty$ the right-hand side is dominated by $q_{max} e^{q_{max} B} |\phi|^2$. Similarly, for $B \to -\infty$ the right-hand side is negative. Therefore, there exists some value of B for which $D = 0$. Furthermore, there is a *unique* such value, which follows simply by taking the derivative of the D term with respect to B, and noting that it is positive definite. In the case of many gauge fields, the same argument works simply by applying it one equation at a time.

Thus we see that the moduli space is the set of all ϕ^i's, since we did not have to constrain them in any way to satisfy the D-term equation. However, we did not fix the $U(1)_{\mathbb{C}}$ gauge invariances either, so we must divide-out this space by all such complexified gauge transformations, giving the result (12.11). Note that this description makes the complex structure of $\mathcal{M}$ manifest. This result is also valid when G is non-Abelian; a closely-related argument for the non-Abelian case appears at the end of chapter 8 of Wess and Bagger.

Note that if we turn on a superpotential in this theory, our analysis of the D-term equations is not changed. If there is a moduli space $\mathcal{M}'$ of solutions to the F-term constraints coming from the superpotential, then this whole space will also be solutions of the D-terms, and so the whole moduli space is

$$\mathcal{M} = \mathcal{M}'/G_{\mathbb{C}} = \{\phi^i | F^j = 0\}/G_{\mathbb{C}}. \tag{12.13}$$

This establishes the result we used in our qualitative discussion of the low-energy physics of the gauged nlσm in the last lecture.

12.3. Dividing by $G_{\mathbb{C}}$

The usefulness of the above result resides not only in showing that solutions to the D-terms always exist, but also in providing a relatively simple description of the resulting moduli space. This is because of the following theorem describing the quotient by the complexified group $G_{\mathbb{C}}$:

$$\{\phi^i\}/G_{\mathbb{C}} = \{\mathcal{P}_G(\phi^i)\}/\{\text{algebraic relations}\}, \tag{12.14}$$

where

$$\mathcal{P}_G(\phi^i) = G\text{-invariant holomorphic monomials of the } \phi^i, \tag{12.15}$$

and the {algebraic relations} refers to any algebraic relations among such monomials. G-invariant holomorphic monomials are monomials in the ϕ^i's only (so no $\overline{\phi}^i$'s and no inverse powers of ϕ^i's) which are gauge-singlets.

We will see in a series of examples that this gives an effective description of the moduli space of supersymmetric gauge theories (without FI terms). But first, let us derive this theorem in the case that $G = U(1)$. Then a $U(1)_{\mathbb{C}}$ transformation rotates the fields by

$$U(1)_{\mathbb{C}}: \quad \phi_i \to e^{q_i\Lambda}\phi_i, \qquad \Lambda \in \mathbb{C}. \tag{12.16}$$

To parametrize the effects of dividing the space of all ϕ_i's (and their complex conjugates) by $U(1)_{\mathbb{C}}$, let us try to find a set of coordinates which span this quotient space. Say $f(\phi_i, \overline{\phi}_i)$ is such a coordinate function. Then f must be $U(1)_{\mathbb{C}}$-invariant, since if not, a value of f will not specify a submanifold of $\mathcal{M}$ since a $U(1)_{\mathbb{C}}$ transformation changes it. Without loss of generality we can expand f as a sum of terms $t_{\{n,\overline{n}\}}$ each of which is of the form

$$t_{\{n,\overline{n}\}} = \prod_i \phi_i^{n_i}\overline{\phi}_i^{\overline{n}_i} \tag{12.17}$$

for some set of exponents $\{n_i, \overline{n}_i\}$. In order for f to be $U(1)_b C$-invariant, each such term must be separately invariant. Thus the set of all such terms can be taken as a possible basis of coordinate functions on $\mathcal{M}$—f is clearly not an independent coordinate.

In order for $t_{\{n,\overline{n}\}}$ to be well-defined on the space of ϕ_i's, their exponents must be integers. For $t_{\{n,\overline{n}\}}$ to be $U(1)_{\mathbb{C}}$-invariant we must have

$$\sum_i q_i n_i = \sum_i q_i \overline{n}_i = 0. \tag{12.18}$$

This separate cancellation of the n_i and $\overline{n}_i$ powers is because Λ is complex in (12.16). Thus each term is a product of two $U(1)_{\mathbb{C}}$-invariant terms—one made only from ϕ_i's and the other from only $\overline{\phi}_i$'s. So, again, we can take the purely holomorphic terms $t_{\{n\}} = t_{\{n,\overline{n}=0\}}$ as a basis of complex coordinate functions. (The purely anti-holomorphic terms are their complex conjugates.) If there were a negative exponent, say $n_i < 0$ in $t_{\{n\}}$, then t would not be a good coordinate near points where $\phi_i = 0$, which are certainly points in $\mathcal{M}$. So, finally, we have a basis of good coordinates on $\mathcal{M}$:

$$\{t_{\{n\}} | \sum_i q_i n_i = 0, \text{ and } n_i \geq 0\}. \tag{12.19}$$

However, these coordinates need not all be independent, and will in general satisfy a set of algebraic relations. For example, given two terms $t_{\{n\}}$ and $t_{\{m\}}$, then $t_{\{n+m\}} = t_{\{n\}} \cdot t_{\{m\}}$, and thus $t_{\{n+m\}}$ should not be considered as an independent coordinate on $\mathcal{M}$.

We have thus developed a description of $\mathcal{M}$ as the space of all t's modulo algebraic relations among them, which is our desired result. This mathematical result is apparently a difficult theorem in the case of non-Abelian G;[18] however, as a physical statement it should seem very plausible.

12.4. Example 2

Consider a $U(1)^n$ theory with $2n+2$ fields with charges

$$\begin{array}{ccccccccc} & U(1)_1 & \times & U(1)_2 & \times & \cdots & \times & U(1)_n \\ \Phi_1^\pm & \pm 1 & & 0 & & \cdots & & 0 \\ \Phi_2^\pm & 0 & & \pm 1 & & \cdots & & 0 \\ \vdots & \vdots & & & & \ddots & & \vdots \\ \Phi_n^\pm & 0 & & 0 & & \cdots & & \pm 1 \\ \Phi_0^\pm & \pm 1 & & \pm 1 & & \cdots & & \pm 1 \end{array} \tag{12.20}$$

A basis of gauge-invariant holomorphic monomials is

$$\begin{aligned} M_i &= \phi_i^+ \phi_i^-, \qquad i = 1, \ldots, n, \\ M_0 &= \phi_0^+ \phi_0^-, \\ B &= \phi_0^- \prod_{i=1}^n \phi_i^+, \\ \widetilde{B} &= \phi_0^+ \prod_{i=1}^n \phi_i^-, \end{aligned} \tag{12.21}$$

which are subject to the single constraint

$$B\widetilde{B} = M_0 \prod_{i=1}^n M_i. \tag{12.22}$$

---

[18] It is a result of a branch of mathematics known as "geometric-invariant theory"; I don't know of a good reference.

Thus

$$\mathcal{M} = \{M_0, M_i, B, \widetilde{B}\}/\{B\widetilde{B} = M_0 \prod_i M_i\}. \tag{12.23}$$

Counting invariants minus relations, we see that the complex dimension of $\mathcal{M}$ is

$$\dim_{\mathbb{C}}(\mathcal{M}) = n + 2. \tag{12.24}$$

This matches our physical expectations: there are $2(n+1)$ complex fields and n complex gauge invariances which, generically, are all broken, leaving us with $2(n+1) - n = n + 2$ complex flat directions.

Let us now examine the singularities of $\mathcal{M}$. Define

$$y = B\widetilde{B} - M_0 \prod_i M_i. \tag{12.25}$$

A singularity occurs when both

$$y = 0, \text{ and } dy = 0. \tag{12.26}$$

The first is simply the condition that we satisfy the constraint that determines $\mathcal{M}$. The second implies there is a singualrity in the tangent space to $\mathcal{M}$, so there are no good local coordinates. The dy constraint is

$$0 = dy = Bd\widetilde{B} + \widetilde{B}dB - \sum_{a=0}^{n} \left(\prod_{b \neq a} M_b \right) dM_a, \tag{12.27}$$

implying that singularities occur whenever

$$B = \widetilde{B} = 0, \text{ and at least } \textit{two } M_a = 0. \tag{12.28}$$

Associated to these singularities are points of enhanced gauge symmetry. For example, when

$$M_1 = M_2 = B = \widetilde{B} = 0 \quad \Rightarrow \quad \phi_1^{\pm} = \phi_2^{\pm} = 0, \tag{12.29}$$

so the diagonal $U(1) \subset U(1)_1 \times U(1)_2$ is unbroken (since $\phi_0^{\pm} \neq 0$). (To deduce that ϕ_1^+ as well as ϕ_1^- vanish, we have to use the D-term equations.) As another example, at the singularity

$$M_0 = M_1 = B = \widetilde{B} = 0 \quad \Rightarrow \quad \phi_0^{\pm} = \phi_1^{\pm} = 0, \tag{12.30}$$

implying that $U(1)_1$ is unbroken.

▸ **Exercise 12.1.** What happens to the singularities in this example as one of the gauge couplings, say g_1 of the $U(1)_1$, goes to zero?

12.5. Example 3

One $U(1)$ and two χsf's with charges

$$
\begin{array}{cc}
 & U(1) \\
\Phi_1^{\pm} & \pm 1 \\
\Phi_2^{\pm} & \pm 2
\end{array}
\tag{12.31}
$$

We thus expect $\dim_{\mathbb{C}}\mathcal{M} = 4-1 = 3$, and indeed we find it, with the basis of four invariants

$$
\begin{array}{ll}
M_1 = \phi_1^+\phi_1^- & B = \phi_2^+\phi_1^-\phi_1^- \\
M_2 = \phi_2^+\phi_2^- & \widetilde{B} = \phi_2^-\phi_1^+\phi_1^+
\end{array}
\tag{12.32}
$$

and the one relation

$$
B\widetilde{B} = M_2 M_1^2. \tag{12.33}
$$

From

$$
\begin{aligned}
0 = y &= B\widetilde{B} - M_2M_1^2, \\
0 = dy &= Bd\widetilde{B} + \widetilde{B}dB - M_1^2 dM_2 - 2M_1M_2 dM_1,
\end{aligned}
\tag{12.34}
$$

we find singularities at

$$
B = \widetilde{B} = M_1 = 0 \ \ \forall M_2, \quad \Rightarrow \quad \phi_1^{\pm} = 0 \text{ and } \phi_2^{\pm} = \text{ arbitrary.} \tag{12.35}
$$

This implies a $\mathbb{Z}_2$ gauge symmetry in the low-energy theory involving the light fields $\Phi_1^{\pm}$ and M_2. So, we see that the enhanced gauge symmetry need not be continuous.

It is worth noting that in this example the Kahler potential in the low-energy theory is given by

$$
\mathcal{K} = \overline{\Phi}_1^+\Phi_1^+ + \overline{\Phi}_1^-\Phi_1^- + \sqrt{\overline{M}_2 M_2}. \tag{12.36}
$$

Thus even though our *coordinate* description of $\mathcal{M}$ was singular for $M_2 \neq 0$, the *metric* on $\mathcal{M}$ is not singular. As we have emphasized before, metric singularities in $\mathcal{M}$ are always associated with massless particles. Here we gain new massless particle when $M_2 \to 0$, since then a $U(1)$ is restored and its associated photon becomes massless.

In all these examples, we see that the (classical) physics we are describing is the Higgs mechanism. Thus these moduli spaces which appear as solutions to the D terms are called Higgs branches. They are often also called D-flat directions.

12.6. Non-Abelian examples

We will save examples of solutions to the D-flatness conditions in non-Abelian gauge theories for later in the course.

13. Quantum gauge theories

We would now like to turn to the quantum-mechanical properties of supersymmetric gauge theories. First, however, we will review the main quantum-mechanical features of ordinary gauge theories. These can be collected into three topics: (1) β-functions, (2) ϑ-angles, and (3) anomalies, which we'll get to next lecture. First, however, a note on the normalization of the gauge kinetic term.

13.1. Normalizations and gauge groups

When one writes the kinetic terms in a non-Abelian gauge theory:

$$\mathcal{L} = -\frac{1}{4g^2}\sum_a F^a_{\mu\nu}F^{a\mu\nu} + \frac{\vartheta}{64\pi^2}\sum_a F^a_{\mu\nu}F^a_{\rho\sigma}\epsilon^{\mu\nu\rho\sigma}, \tag{13.1}$$

one is implicitly using a conventional normalization of the gauge group generators. This follows because the non-Abelian field strength is defined as

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu, \tag{13.2}$$

and the structure constants of the gauge algebra are defined by

$$[T^a, T^b] = i f^{abc} T^c. \tag{13.3}$$

Clearly, if one should rescale the group generators, one must rescale the gauge potentials as well to keep the definition of F the same. One finds the scalings:

$$T' = \alpha T, \quad f' = \alpha f, \quad A' = \frac{1}{\alpha}A, \quad F' = \frac{1}{\alpha}F. \tag{13.4}$$

Thus, to keep the gauge coupling constants in (13.1) invariant under these rescalings, we need to insert a compensating factor.

The natural factor is the quadratic invariant $C(R)$ (related to the quadratic Casimir) of some representation R of the gauge group G, defined by

$$\mathrm{tr}_R(T^aT^b) = C(R)\delta^{ab}, \tag{13.5}$$

where tr_R denotes a trace in the R representation. Under the above rescaling,

$$C'(R) = \alpha^2 C(R), \tag{13.6}$$

so it is a suitable factor to multiply the gauge kinetic action (13.1) by to make the coupling constants normalization-independent.

There remains, however, the arbitrary choice of representation R to use. This is a matter of convention. For the classical gauge groups (Sp, SU, and SO) it is standard to use the fundamental (or defining) representation, and to write (13.1) as

$$\mathcal{L} = -\frac{1}{2g^2}\mathrm{tr}_f(F^2) + \frac{\vartheta}{16\pi^2}\mathrm{tr}_f(F\widetilde{F}), \tag{13.7}$$

where tr_f denotes the trace in the fundamental representation. Recall that

$$\widetilde{F}^{\mu\nu} \equiv \tfrac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}, \tag{13.8}$$

so that $\widetilde{\widetilde{F}} = F$. One often sees the gauge kinetic term written with the gauge fields in the adjoint representation of G. In this case the invariant formula is (13.7) with tr_f replaced by Tr (using the conventional notation that a capitalized trace refers to a trace in the adjoint representation) multiplied by an overall factor of C_f/C_{adj}, where C_{adj} is the quadratic invariant of the adjoint representation (which, for the adjoint is the same as its quadratic Casimir, sometimes denoted C_2).

The forms for gauge kinetic terms one sees in the literature often implicitly use the following conventional normalizations of the gauge group generators (for the classical groups) given by the quadratic invriants for their defining representations. We include some additional useful group theory information:

G	rank(G)	$d(G)$	C	C_f	$d(fund)$	type fun.rep.
$Sp(2N)$	N	$N(2N{+}1)$	$N{+}1$	1/2	$2N$	pseudoreal
$SU(N)$	$N{-}1$	$N^2{-}1$	N	1/2	N	complex
$SO(N)$	$[N/2]$	$N(N{-}1)/2$	$N{-}2$	1	N	real

Here $d(G)$ is the dimension of the group and $d(fund)$ is the dimension of the fundamental (defining) representation.

It should be clear however, that the real invariant quantities are the *ratios* of the quadratic invariants. From these ratios can be defined the *index* of a representation $T(R)$. Thus the indices will enter in physical quantities, and not the quadratic invariants. Inthe case of the classical groups the index is simply

$$T(R) \equiv C(R)/C_f. \tag{13.9}$$

(Mathematically, there is a more general definition, applicable to all simple Lie algebras.) It is a theorem that the index of any representation is an integer. We see that, by definition, the index of the fundamental representation is 1, and, from the above table, that the indices of the adjoint representations are $2N{+}2$, $2N$, and $N{-}2$ for $Sp(2N)$, $SU(N)$, and $SO(N)$, respectively.

A useful formula for computing the indices of other representations is

$$C(R_1)d(R_2) + d(R_1)C(R_2) = \sum_i C(R_i), \tag{13.10}$$

where $d(R)$ is the dimension of the representation, and here $R_1 \times R_2 = \sum_i R_i$. Also, one should be aware of the following equivalences among Lie algebras:

$$\begin{aligned} SO(3) &\simeq SU(2) \simeq Sp(2) \\ SO(4) &\simeq SU(2) \times SU(2) \simeq Sp(2) \times Sp(2) \\ SO(5) &\simeq Sp(4) \\ SO(6) &\simeq SU(4). \end{aligned} \tag{13.11}$$

In computing the index one should use the right-most groups in these equivalences.

13.2. RG flow of gauge coupling constants

So, a gauge theory lagrangian contains the gauge kinetic term

$$\mathcal{L} \supset \frac{1}{2g_0^2} \mathrm{tr}_f(F^2). \tag{13.12}$$

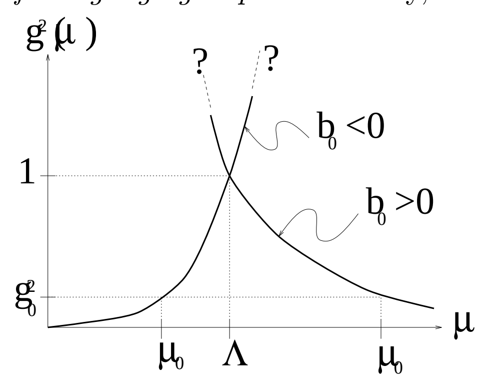
Here we are thinking of $\mathcal{L}$ as an effective action at a scale μ_0, and g_0 is the coupling at that scale. The one-loop RG implies the coupling "runs" as a function of the scale μ according to

$$\mu \frac{dg}{d\mu} = -\frac{b_0}{16\pi^2} g^3 + \mathcal{O}(g^5) \quad \Rightarrow \quad \frac{1}{g^2(\mu)} \simeq -\frac{b_0}{8\pi^2} \log\left(\frac{\Lambda}{\mu}\right), \tag{13.13}$$

where we have defined

$$\Lambda \equiv \mu_0 e^{-8\pi^2/b_0 g_0^2}, \tag{13.14}$$

the (strong-coupling) *scale of the gauge group.* Pictorially, this RG running looks like:



Here we have shown the running of two coulings, one with $b_0 < 0$ and one with $b_0 > 0$. The first case is weakly-coupled in the IR, and so their gauge groups are refered to as IR-free (IRF) gauge groups. The second case is weakly-coupled in the UV, and are refered to as asymptotically-free (AF) gauge groups. We see that the scale of the gauge group is just the scale where its coupling becomes strong (of order one). Beyond this scale perturbation results (such as the one-loop running) are no longer reliable. The trading of the information of a gauge coupling at a given scale for the strong-coupling scale of the gauge group

$$\{g_0, \mu_0\} \leftrightarrow \Lambda, \tag{13.15}$$

is known as "dimensional transmutation". In a theory with many gauge groups, $G_1 \times G_2 \times \cdots \times G_n$, there will be correspondingly many gauge group scales $\{\Lambda_1, \cdots, \Lambda_n\}$.

The basic "phenomenology" of the running of classically marginal couplings in four-dimensional QFT is that the couplings

$$\lambda\phi^4 + g\phi\psi^2 + \frac{1}{e^2}F^2, \qquad (F \text{ Abelian}), \tag{13.16}$$

are all IRF, while non-Abelian gauge couplings may be AF if the gauge group is coupled to "not too much" matter. More precisely

$$b_0 = \frac{11}{6}T(\text{adj}) - \frac{1}{3}\sum_i T(R_i) - \frac{1}{6}\sum_a T(R_a), \tag{13.17}$$

where the sum on i is over Weyl fermions with R_i the representation of the ith fermion, and the sum on a is over complex bosons in representations R_a.

13.3. ϑ angles and instantons

Gauge theories can also contain another term built solely out of the gauge fields:

$$\mathcal{L} \supset \frac{\vartheta}{16\pi^2}\text{tr}_f(F\widetilde{F}) = \frac{\vartheta}{8\pi^2}\epsilon^{\mu\nu\rho\sigma}\partial_\mu \text{tr}_f(A_\nu\partial_\rho A_\sigma + \frac{2}{3}A_\nu A_\rho A_\sigma). \tag{13.18}$$

Since this term is a total derivative, classically it has no effect; however, quantumly it can have an effect due to fluctuations of the gauge fields. The typical gauge configuration appearing in the path integral is not smooth, and the ϑ term can take an arbitrary value for these gauge fields.

In some cases, however, semi-classical configurations called *instantons*—saddle-point contributions to the path integral coming from expanding about finite-action solutions to the Euclidean equations of motion—can give finite contributions to the path integral. It turns out that for sufficiently regular gauge configurations (we will derive this later)

$$\frac{1}{16\pi^2}\int d^4x\, \text{tr}_R(F\widetilde{F}) = n \quad \in \mathbb{Z}. \tag{13.19}$$

In this expression, R is an arbitrary non-trivial representation of the gauge group.[19] The ϑ-term is said to be a "topological invariant" since continuous deformations of F don't change n. The integer n is called the *instanton number*. We will later construct an explicit example of an instanton solution (in a supersymmetric context). In an n-instanton background, the ϑ-term contributes to the action

$$S \supset \int d^4x \, \frac{\vartheta}{16\pi^2} \mathrm{tr}_f(F\widetilde{F}) = n\vartheta. \tag{13.20}$$

The path integral computes $\int \mathcal{D}\phi \ldots e^{iS}$, so

$$\vartheta \to \vartheta + 2\pi \tag{13.21}$$

is a (semi-classical) symmetry of the theory.

This symmetry which we derived semi-classically is actually an exact symmetry of any gauge theory.[20] This follows from the existence of non-trivial homotopy classes ("winding numbers") of gauge transformations $\mathcal{G}_n$ labelled by an integer n. Such large gauge transformations ($n \neq 0$) are not connected to the identity, so it is possible for them to be realized projectively on physical states: $\mathcal{G}_n|\Psi\rangle = e^{in\vartheta}|\Psi\rangle$. The angular parameter ϑ is a new parameter in gauge theory that only appears quantumly. If we demand a description of the theory in which physical states do *not* transform by a phase under large gauge transformations—corresponding to dividing out by the full classical gauge symmetry, suitable for a path-integral description—then the ϑ-angle is translated into the coeffcient of the $F\widetilde{F}$ term in the action. This is because the $F\widetilde{F}$ term is the unique gauge-invariant, total derivative term one can write down whose contribution on a constant-time slice computes the winding number of the gauge configuration.

We can put a lower bound on the (Euclidean) action of an instanton:

$$\begin{aligned} 0 \leq \int d^4x \mathrm{tr}_f(F \pm \widetilde{F})^2 &= \int d^4x [2\mathrm{tr}_f(F^2) \pm 2\mathrm{tr}_f(F\widetilde{F})] \\ \Rightarrow \int d^4x \mathrm{tr}_f F^2 &\geq \left| \int d^4x \mathrm{tr}_f F\widetilde{F} \right| = 16\pi^2|n|. \end{aligned} \tag{13.22}$$

[19] Since the rest of the expression is independent of R, we can take the one with the smallest index—the defining representations for the classical groups—so the quantization of the $F\widetilde{F}$ term is $\int d^4x \sum_a F^a \widetilde{F}^a = 16\pi^2 n/C_f$, invariantly.

[20] The large gauge transformations that follow only exist for non-Abelian gauge theories on Minkowski space; however, as we will see near the end of the course, even in an Abelian gauge theory there can be such large gauge transformations in the presence of monoples.

So instanton contributions will be down by factors of (at least)

$$e^{-S_{inst}} = e^{-8\pi^2/g^2} = \left(\frac{\Lambda}{\mu_0}\right)^{b_0}, \tag{13.23}$$

and so are non-perturbative effects going as a power of the gauge group scale. In this formula, the RG scale μ_0 will turn out be interpreted as the size of the instanton, and will need to be integrated over when we perform a more detailed instanton calculation later.

We should emphasize that instantons represent only a semi-classical approximation to the non-perturbative physics of gauge theories. We will see examples later in this course of non-perturbative effects which cannot be interpreted as coming from instanton effects.

14. Anomalies

In the Abelian nlσm we had restrictions on the allowed charges of the χsf's from anomaly cancellation. In the non-Abelian context this places restrictions on the representations of the χsf's. We will now discuss the origin and systematics of anomalies. A good reference for the physics of anomalies which is mostly complementary to the approach I'll take here are the introductory sections of L. Alvarez-Gaumé and E. Witten, *Nucl. Phys.* B234 (1983) 269.

Anomalies refer to classical symmetries which are broken by quantum effects. This means that in the full quantum theory there is no (gauge-invariant or covariant) conserved current for an anomalous symmetry. This is important in the case of classical global symmetries, implying as it does that the classical Ward identities are violated, but it does not affect the consistency of the theory. A familiar and important example of an anomalous symmetry is scale-invariance: as we saw above, quantum effects in a classically scale-invariant Yang-Mills theory make the gauge coupling run with scale. Another kind of anomaly occurs in the conservation currents for chiral rotations, and I'll call them chiral anomalies. It will turn out (in the next lecture) that in supersymmetric theories scale and chiral anomalies are related.

If anomalous chiral rotations are gauged, then the resulting theory is *inconsistent*, since we only know how to couple spin-1 fields (gauge fields) in a unitary way to conserved currents.

Chiral anomalies (local or global) arise in four dimensional QFT only in theories where fermions are coupled to gauge fields. They can be computed in perturbation theory and only occur at one loop level. This is a reflection of the fact that they can also be thought of as IR effects. From this perspective, the existence of anomalies depends only on the field content and charges of the light fields in the theory, and not on details of the interactions. Later we will try to bring out this feature of anomalies later in this lecture, but for the moment, we will summarize the perturbative approach.

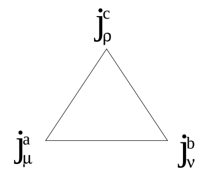
In free field theory we can write conserved currents for fermions transforming in a representation R of a symmetry group G as

$$j^a_\mu = \overline{\psi}^{\dot\alpha}_i \overline{\sigma}^\mu_{\dot\alpha\alpha} \psi^\alpha_j (T^a_R)^{ij} \tag{14.1}$$

where T^a_R are hermitian generators in the representation R of the Lie algebra of G:

$$[T^a_R, T^b_R] = i f^{ab}_c T^c_R. \tag{14.2}$$

We can compute the 3-point function of currents at one loop

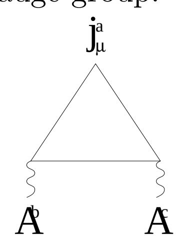


$$\langle j^a_\mu(x_1) j^b_\nu(x_2) j^c_\rho(x_3)\rangle = \mathrm{tr}_R(T^a T^b T^c) f_{\mu\nu\rho}(x_i). \tag{14.3}$$

Now couple these currents to gauge fields in the usual way,

$$\mathcal{L} = \mathcal{L}_{free} + \sum A^\mu_a j^a_\mu, \tag{14.4}$$

and compute $\langle jAA\rangle$, taking care to regulate, impose Bose symmetry on the gauge fields, and covariantize with respect to the gauge group.[21]



Differentiating the result, one finds the "Abelian anomaly"[22]

$$\partial^\mu j^a_\mu \propto \mathrm{tr}_R(T^a\{T^b, T^c\}) F^{\mu\nu}_b F^{\rho\sigma}_c \epsilon_{\mu\nu\rho\sigma} = \partial^\mu K^a_\mu. \tag{14.6}$$

[21] Note that in gauge-covariantizing the result for non-Abelian gauge groups means adding terms with more powers of the external gauge potential. These correspond to one-loop, higher-point anomalous diagrams in perturbation theory. The "Wess-Zumino consistency conditions" imply, though, that all these higher-point amplitudes can be derived from the 3-point amplitude. See, for an explanation with few details, sections 13.3 and 13.4 of Green, Schwarz and Witten, *Superstring Theory*, vol 2.

[22] There is also a "non-Abelian anomaly" for the gauge-covariant derivative:

$$\mathcal{D}^\mu j^a_\mu = \frac{\epsilon^{\mu\nu\rho\sigma}}{24\pi^2} \partial_\mu \mathrm{tr}_f[T^a(A_\nu \partial_\rho A_\sigma + \tfrac{1}{2} A_\nu A_\rho A_\sigma)], \tag{14.5}$$

which is also non-zero, but not the same as the Abelian anomaly.

This implies the j^a_μ current is not conserved. (Though the combination $j^a_\mu - K^a_\mu$ is conserved, it is not gauge-covariant.)

To summarize, we learn that if $\mathrm{tr}_R(T^a\{T^b, T^c\}) \neq 0$ for

- 3 global currents $\Rightarrow$ harmless (though important),
- 1 global and 2 local currents $\Rightarrow$ global current not conserved,
- 3 local currents $\Rightarrow$ inconsistent.

As an example, consider massless QED:

$$\begin{array}{ccc} & U(1)_{\text{gauge}} \quad \times & U(1)_{\text{global}} \\ \underset{\sim}{\psi}_\alpha & -1 & +1 \\ \psi_\alpha & +1 & +1 \end{array} \tag{14.7}$$

with currents

$$j^\mu_L \equiv \overline{\psi}\overline{\sigma}^\mu\psi, \qquad\qquad j^\mu_R \equiv \overline{\widetilde{\psi}}\overline{\sigma}^\mu\widetilde{\psi}. \tag{14.8}$$

Then the gauge current is $j_V = j_R - j_L$ and the global current is $j_A = j_R + j_L$. The anomalies are then

$$\begin{aligned} \langle j_A j_V j_V\rangle &\propto \{1\cdot(-1)^2 + 1\cdot(+1)^2\} \neq 0 \quad \Rightarrow \quad \partial_\mu j^\mu_A \neq 0, \\ \langle j_V j_V j_V\rangle &\propto \{(-1)^3 + (+1)^3\} = 0 \quad \Rightarrow \quad \partial_\mu j^\mu_A = 0, \end{aligned} \tag{14.9}$$

so the global symmetry is anomalous and the gauge symmetry is non-anomalous.

14.1. Gauge anomalies

As a more complicated example, we can calculate the anomaly conditions for a $U(1)^n$ gauge theory with fermion ψ_i with charges q_{ia} under $U(1)_a$. The gauge currents are

$$j^\mu_a = \sum_i \overline{\psi}_i\overline{\sigma}^\mu\psi_i q_{ia}, \tag{14.10}$$

so the gauge anomaly is

$$\langle j_a j_b j_c\rangle \propto \sum_i q_{ia}q_{ib}q_{ic}, \tag{14.11}$$

which must vanish for consistency of the theory. Also, one can insert two $t^{\mu\nu}$ (energy-momentum) tensors in a triangle diagram and couple them to gravity in the usual way $(\int d^4x\sqrt{g}g_{\mu\nu}t^{\mu\nu})$ giving the mixed gauge-gravitational anomaly

$$\langle j_a tt\rangle \propto \sum_i q_{ia}, \tag{14.12}$$

which must also vanish for consistency. This reproduces the anomaly conditions introduced in an earlier lecture.

The generalization to non-Abelian gauge anomalies is straight-forward. Consider a theory with fermion ψ_i in representations R_i of a gauge group G. Then the gauge anomalies cancel if

$$\sum_i \mathrm{tr}_{R_i}(T^a\{T^b,T^c\}) = 0 \qquad \forall a,b,c. \tag{14.13}$$

This is actually a much less restrictive condition for non-Abelian groups than it might seem. First of all, if the anomaly is non-zero, it implies that there is a symmetric G-invariant tensor d^{abc}. Furthermore, real representations give no contribution to the anomaly since for a real representation, $T^a = -(T^a)^T$ (up to a unitary similarity transformation), implying

$$\begin{aligned} \mathrm{tr}_R(T^a\{T^b,T^c\}) &= \mathrm{tr}_R\left[(-T^a)^T\left\{(-T^b)^T,(-T^c)^T\right\}\right] = -\mathrm{tr}_R(\{T^c,T^b\}T^a) \\ &= -\mathrm{tr}_R(T^a\{T^b,T^c\}). \end{aligned} \tag{14.14}$$

The only groups which have both complex representations and a symmetric three-index invariant tensor are the $SU(n)$ groups for $n \geq 3$. (Note that $SO(6) \simeq SU(4)$).

[There is one other anomaly which requires, in our conventions, that the total index of fermions transforming in pseudoreal representations be an even integer. This is relevant only for the $Sp(2n)$ groups. (Note that $Sp(2) \simeq SU(2)$.) This is an anomaly under large gauge transformations, and so is not seen in the perturbative approach; see E. Witten *Phys. Lett.* 117B (1982) 324.]

Finally, there are the mixed gauge-gravitational anomalies, which imply that $\sum_i \mathrm{tr}_{R_i}(T^a) = 0$ for consistency. But for semi-simple groups the generators are automatically traceless. Thus the mixed gauge-gravitational anomalies only constrain the coupling to $U(1)$ gauge factors.

14.2. Chiral anomalies (anomalies in global symmetries)

Having satisfied the consistency conditions from the gauge anomalies, we now turn to the physics of anomalous global symmetries. Suppose we have a gauge group G with generators T^a, a global symmetry group $\widetilde{G}$ generated by $\widetilde{T}^a$, and Weyl fermions ψ_i transforming in the $R_i \times \widetilde{R}_i$ representation of $G \times \widetilde{G}$. Then from the triangle diagram with one global current j^a_μ insertion and two gauge insertions we find the anomaly is proportional to

$$\partial^\mu j^a_\mu \propto \sum_i \mathrm{tr}(\widetilde{T}^a_{\widetilde{R}_i} T^b_{R_i} T^c_{R_i}) = \sum_i \mathrm{tr}_{\widetilde{R}_i}(\widetilde{T}^a)\mathrm{tr}_{R_i}(T^b T^c). \tag{14.15}$$

Again, if $\widetilde{G}$ is semi-simple, $\mathrm{tr}(\widetilde{T}^a) \equiv 0$, so there is no anomaly. Thus there are only anomalies in global $U(1)$ symmetries.

So, let us restrict ourselves to the case where the global symmetry is $U(1)$ and the fermions ψ_i have global charge q_i and transform as above in the R_i representation of some gauge group G. Then the anomaly is (this time including all the factors)

$$\partial^\mu j_\mu = \frac{1}{16\pi^2}\sum_i q_i \mathrm{tr}_{R_i}(F\widetilde{F}) = \frac{\sum_i q_i T(R_i)}{16\pi^2 C}\mathrm{tr}_f(F\widetilde{F}). \tag{14.16}$$

This implies that the symmetry is anomalous if $\sum_i q_i T(R_i) \neq 0$. We will normally normalize our $U(1)$ generator so that the charges q_i are integers (so the parameter α of such a rotation is an angle $\alpha \simeq \alpha + 2\pi$).

14.3. Fujikawa's derivation of the anomaly

The formula (14.16) is the key result of our discussion of anomalies. We will now derive it following the approach of K. Fujikawa, *Phys. Rev. Lett.* 42 (1979) 1195, which most clearly reveals the topological and geometrical nature of anomalies. In this approach, the anomaly is seen as the non-invariance of the fermionic path-integral measure under chiral rotations.

Consider the effective action for a Weyl fermion transforming in a representation R in a background non-Abelian gauge field:

$$e^{iS_{eff}[A]} \equiv \int \mathcal{D}\psi\mathcal{D}\overline{\psi} e^{iS[\psi,\overline{\psi},A]}, \tag{14.17}$$

where we will take the action to be free (for convenience)

$$S = \mathrm{tr}_R \int d^4x\, \overline{\psi} \not{D} \psi \qquad \left(= \mathrm{tr}_R \int d^4x\, \psi \overline{\not{D}}\,\overline{\psi} \right). \tag{14.18}$$

Here

$$\begin{aligned} \not{D}^{\dot\alpha\alpha} &\equiv i\overline{\sigma}^{\mu\dot\alpha\alpha}(\partial_\mu + iA_\mu), \\ \overline{\not{D}}_{\alpha\dot\alpha} &\equiv i\sigma^\mu_{\alpha\dot\alpha}(\partial_\mu - iA_\mu), \end{aligned} \tag{14.19}$$

is the Dirac operator which depends on the background gauge configuration. This action obviously is invariant under the global chiral (or "axial") symmetry rotations

$$\psi \to e^{i\alpha}\psi \qquad \left(\Rightarrow \quad \overline{\psi} \to e^{-i\alpha}\overline{\psi}\right). \tag{14.20}$$

Indeed, we can derive the associated conservation equation by the Noether method, in which we look at the response of the path integral to a *local* chiral rotation:

$$\psi \to e^{i\alpha(x)}\psi \equiv \psi'. \tag{14.21}$$

Under this transformation the action changes to

$$\begin{aligned} S \to S' &= \mathrm{tr}_R \int d^4x\, \overline{\psi} e^{-i\alpha} \not{D} e^{i\alpha}\psi \\ &= \mathrm{tr}_R \int d^4x\, \left(\overline{\psi}\not{D}\psi + \overline{\psi} i(\not{\partial}\alpha)\psi\right) \\ &= S - i\int d^4x\, \alpha(x)\partial^\mu \mathrm{tr}_R(\overline{\psi}\overline{\sigma}_\mu\psi). \end{aligned} \tag{14.22}$$

Since we can consider this transformation on the fermion fields just as a change of variable in the path integral, we must have that $\partial^\mu j_\mu = 0$, where

$$j_\mu \equiv \mathrm{tr}_R \overline{\psi}\overline{\sigma}_\mu\psi. \tag{14.23}$$

Of course, this was not correct: we did not include the Jacobian for the change of variables in the measure of the path integral. In order to evaluate this, let us diagonalize (formally) our action. Noting that

$$\begin{aligned} (D^2)^\alpha_\beta &\equiv \overline{\not{D}}_{\beta\dot\alpha}\not{D}^{\dot\alpha\alpha}, \\ (\overline{D}^2)^{\dot\alpha}_{\dot\beta} &\equiv \not{D}^{\dot\alpha\alpha}\overline{\not{D}}_{\alpha\dot\beta} \end{aligned} \tag{14.24}$$

are negative-semi-definite hermitian operators, we can diagonalize them by

$$\begin{aligned} (D^2)^\alpha_\beta \xi_{\alpha n} &= -\lambda_n^2 \xi_{\beta n}, \\ (\overline{D}^2)^{\dot\alpha}_{\dot\beta}\overline{\eta}^{\dot\beta}_n &= -\lambda_n^2 \overline{\eta}^{\dot\alpha}_n, \end{aligned} \tag{14.25}$$

for non-negative real λ_n, where ξ_n and $\overline{\eta}_n$ are a basis of *commuting* (not Grassmann) spinor functions normalized so that

$$\begin{aligned} \mathrm{tr}_R \int d^4x\, \xi_n^* \xi_m &= \delta_{nm} \\ \mathrm{tr}_R \int d^4x\, \overline{\eta}_n^* \overline{\eta}_m &= \delta_{nm}. \end{aligned} \tag{14.26}$$

Here the asterisk denotes complex-conjugation of the spinor functions, whose spinor indices are contracted in the usual way (raising and lowering them with $\epsilon_{\alpha\beta}$), and the bar on the $\overline{\eta}$ spinors is just to remind us that it is a right-handed as opposed to a left-handed spinor (*i.e.* the bars should not be interpreted as complex conjugation of the spinor functions). The eigenvalues of D^2 and $\overline{D}^2$ are the same, as can be seen by defining

$$\begin{aligned} \not{D}\xi_n &= \lambda_n \overline{\eta}_n \\ \overline{\not{D}}\overline{\eta}_n &= -\lambda_n \xi_n, \end{aligned} \tag{14.27}$$

and checking that the previous two equations are satisfied. This is just the 4-dimensional version of the 1-dimensional diagonalization of a fermionic action that we did when we discussed fermionic path integrals in supersymmetric QM in lecture 4. Just as in that case, there can also be zero modes ξ_0 and $\overline{\eta}_0$ of $\not{D}$ and $\overline{\not{D}}$:

$$\not{D}\xi_0 = 0, \qquad \overline{\not{D}}\overline{\eta}_0 = 0. \tag{14.28}$$

They do not have to be equal in number.

Now, expand the spinor ψ in this basis:

$$\begin{aligned} \psi(x) &= \sum_n a_n \xi_n(x) \\ \overline{\psi}(x) &= \sum_n b_n \overline{\eta}_n^*(x), \end{aligned} \tag{14.29}$$

where a_n and b_n are Grassmann numbers. The path integral measure is then (defined to be)

$$\int \mathcal{D}\psi\mathcal{D}\overline{\psi} \equiv \int \prod_n da_n db_n. \tag{14.30}$$

Under a change of variables $\psi \to \psi'$ by a local chiral rotation (14.21), the Grassmann coefficients change as

$$\begin{aligned} a_n \to a_n{}' &= C_{nm} a_m, \qquad & C_{nm} &\equiv \int d^4x\, e^{i\alpha(x)} \mathrm{tr}_R(\xi_n^* \xi_m) \\ b_n \to b_n{}' &= \overline{C}_{nm} b_m, \qquad & \overline{C}_{nm} &\equiv \int d^4x\, e^{-i\alpha(x)} \mathrm{tr}_R(\overline{\eta}_m^* \overline{\eta}_n) \end{aligned} \tag{14.31}$$

where we have used the orthonomormality relations to compute

$$a_n{}' = \mathrm{tr}_R \int d^4x\, \xi_n^* \psi', \qquad b_n{}' = \mathrm{tr}_R \int d^4x\, \overline{\psi}' \overline{\eta}_n. \tag{14.32}$$

(Note that since ξ and $\overline{\eta}$ are not Grassmann, the order of factors in the above expressions is important!) Under this change of variables, the fermionic measure transforms as

$$\prod_n da_n db_n \to \prod_n da_n{}' db_n{}' = (\det C)^{-1} (\det \overline{C})^{-1} \prod_n da_n db_n. \tag{14.33}$$

The reason for the inverse determinant is due to the nature of Grassmann integration. For example, since $1 = \int d\theta\theta = \int d(c\cdot\theta)(c\cdot\theta)$, therefore $d(c\cdot\theta) = \frac{1}{c}d\theta$. We can evaluate this determinant by expanding in powers of α: $C_{nm} = \delta_{nm} + \delta C_{nm} + \mathcal{O}(\alpha^2)$, where

$$\begin{aligned} \delta C_{nm} &= i \int d^4x\, \alpha \mathrm{tr}_R(\xi_n^* \xi_m) \\ \delta \overline{C}_{nm} &= -i \int d^4x\, \alpha \mathrm{tr}_R(\overline{\eta}_m^* \overline{\eta}_n). \end{aligned} \tag{14.34}$$

Using the identities

$$\det{}^{-1} C = e^{-\mathrm{tr}\log C} \simeq e^{-\mathrm{tr}\log(1+\delta C)} \simeq e^{-\mathrm{tr}\delta C}, \tag{14.35}$$

we find to leading order in α

$$(\det C\overline{C})^{-1} = \exp\left\{-i\int d^4x\,\alpha(x)\mathbf{A}(x)\right\}, \qquad \mathbf{A} \equiv \sum_n \mathrm{tr}_R\left(\xi_n^*\xi_n - \overline{\eta}_n^*\overline{\eta}_n\right). \tag{14.36}$$

Putting all the pieces together, the total variation of the path integral under this change of variables (the local chiral rotation) is

$$\int \mathcal{D}\psi\mathcal{D}\overline{\psi}e^{iS} \to \int \mathcal{D}\psi\mathcal{D}\overline{\psi}e^{iS+\int d^4x\alpha(x)(\partial^\mu j_\mu(x)+\mathbf{A}(x))} \tag{14.37}$$

implying the anomalous conservation equation

$$\partial^\mu j_\mu = \mathbf{A}. \tag{14.38}$$

We would now like to evaluate the anomaly $\mathbf{A}$. In the limit of a global chiral rotation, $\alpha(x) \to$constant, it is clear that the anomaly is ill-defined, so we need to regulate $\int d^4x\,\mathbf{A}(x)$. We can do this using the "heat kernel" method which adds a gaussian suppression to UV modes above a regulator scale M:

$$\begin{aligned}\left[\int d^4x\,\mathbf{A}(x)\right]_M &\equiv \mathrm{tr}_R\int d^4x\sum_n\left(\xi_n^* e^{D^2/M^2}\xi_n - \overline{\eta}_n^* e^{\overline{D}^2/M^2}\overline{\eta}_n\right)\\ &= \sum_n e^{-\lambda_n^2/M^2}\mathrm{tr}_R\int d^4x\left(\xi_n^*\xi_n - \overline{\eta}_n^*\overline{\eta}_n\right)\\ &= \nu - \overline{\nu} \equiv \mathrm{index}(\not{D}[A]),\end{aligned} \tag{14.39}$$

where in the last step the contribution from *all* modes with non-zero eigenvalues cancelled, leaving only the ν ξ zero-modes, and the $\overline{\nu}$ $\overline{\eta}$ zero-modes. The difference in number of left- and right-handed zero-modes is called the index of the Dirac operator.

To evaluate the integrand of the above expression in this regulator we note that the above integral just computes a trace which we can evaluate in a plane-wave basis:

$$\left[\int d^4x\,\mathbf{A}(x)\right]_M = \mathrm{tr}(e^{D^2/M^2}) - \mathrm{tr}(e^{\overline{D}^2/M^2}) = \mathrm{tr}_R\int\frac{d^4k}{(2\pi)^4}\langle k|e^{D^2/M^2} - e^{\overline{D}^2/M^2}|k\rangle. \tag{14.40}$$

Assuming the gauge background is varying only on scales $k < \Lambda$, (so that we are computing the anomaly for an effective theory for energies less than Λ), the integrand is approximately

$$\mathbf{A}(x) = \left[\lim_{x\to y}\int_\Lambda\frac{d^4k}{(2\pi)^4}e^{ik\cdot(x-y)}e^{-k^2/M^2}\right]\,\mathrm{tr}_R\left(e^{-i\not{F}^+(x)/M^2} - e^{+i\not{F}^-(x)/M^2}\right) \tag{14.41}$$

where we have used the fact that

$$
\begin{aligned}
D^2 &= \partial^2 \delta^\alpha_\beta - i\ \not{F}^{+\alpha}{}_\beta + \mathcal{O}(\partial A) \\
\overline{D}^2 &= \partial^2 \delta^{\dot\alpha}_{\dot\beta} + i\ \not{F}^{-\dot\alpha}{}_{\dot\beta} + \mathcal{O}(\partial A).
\end{aligned}
\tag{14.42}
$$

Recall that $\not{F}^+ = \sigma^{\mu\nu}F_{\mu\nu}$ and $\not{F}^- = \overline{\sigma}^{\mu\nu}F_{\mu\nu}$, are the $(1,0)$ and $(0,1)$ parts of the non-Abelian background field strength. Rescaling the momenta as $k \to kM$, we find in the limit $M \to \infty$ that

$$
\begin{aligned}
\mathbf{A}(x) &= \frac{i}{16\pi^2}\mathrm{tr}_R\left(iM^2\left[\not{F}^- - \not{F}^+\right] + \tfrac{1}{2}\left[(\not{F}^-)^2 - (\not{F}^+)^2\right] + \mathcal{O}(M^{-2})\right) \\
&= C(R)\frac{1}{16\pi^2}\mathrm{tr}_R(F\widetilde{F})
\end{aligned}
\tag{14.43}
$$

which is the anomaly. Note that in the process of deriving the anomaly, we have shown that its integral is an integer.

14.4. Some anomalous physics

Let us now extract the physics of the chiral anomaly.

Firstly, the anomalous breaking can be expressed in an effective action by assigning the ϑ-angle transformation properties under the anomalous symmetry:

$$
\begin{aligned}
\psi^i &\to e^{iq_i\alpha}\psi^i \\
\vartheta &\to \vartheta + \alpha\left[\sum_i q_i T(R_i)\right].
\end{aligned}
\tag{14.44}
$$

This follows by the Noether procedure, where the classical conservation equation is derived by varying the action with respect to a local symmetry transformation. Since the right-hand side of the anomalous conservation equation (14.16) is proportional to $F\widetilde{F}$, it will be generated by a shift in ϑ. In this way we understand the anomalous breaking of the $U(1)$ symmetry as occuring due to an *explicit* breaking: a term $(\vartheta F\widetilde{F})$ in the action is not invariant.

Secondly, since the anomaly appears only through the ϑ term, it follows that at most one global $U(1)$ symmetry per gauge factor can be anomalous—by making appropriate linear combinations of their generators, one can choose all others to be non-anomalous.

Thirdly, by an anomalous chiral rotation (say of a single massless Weyl fermion coupled to the gauge group) one can rotate the ϑ-angle away. If there is no massless fermion charged under the gauge group, then any such rotation will at the same time give a CP-violating phase to some fermion mass, and in this case the ϑ-angle has observable consequences. (This is the origin of the strong CP problem in the standard model due to a possible ϑ-angle in $SU(3)$ QCD.)

The last point also applies if there are Yukawa couplings; however, in that case a scalar field will generally also transform under the anomalous chiral $U(1)$. If it is gauge neutral (or at scales below the scale of spontaneous breaking of any gauge groups it is charged under), then the anomalous conservation law can be considered to arise from an effective coupling of the argument of ϕ to $F\widetilde{F}$:

$$\mathcal{L}_{eff} \supset \frac{\vartheta + a(x)}{16\pi^2}\mathrm{tr}_f(F\widetilde{F}), \tag{14.45}$$

where

$$\phi(x) = \rho(x)e^{ia(x)}, \qquad \rho \in \mathbb{R}^+, \quad a \in [0, 2\pi). \tag{14.46}$$

For a term such as (14.45) to be well defined, the modulus of ϕ must be positive: $\langle\rho\rangle \neq 0$. In this case of a spontaneously broken anomalous $U(1)$ symmetry (called a Peccei-Quinn symmetry), we are free to absorb the ϑ-angle in a redefinition of a, the argument of ϕ. The field a is called an axion.

Finally, the global charge violation in, say, a scattering process due to the anomaly is:

$$\begin{aligned}\Delta Q &= \int_{-\infty}^{+\infty} dt\, \partial_0 Q = \int dt d^3x\, \partial_0 j_0 = \int d^4x \left(\partial^i j_i + \frac{\sum_i q_i T(R_i)}{16\pi^2}\mathrm{tr}_f(F\widetilde{F})\right) \\ &= \left[\sum_i q_i T(R_i)\right] n\end{aligned} \tag{14.47}$$

where we dropped a total derivative of the current, and n in the last line is the (change in the) instanton number. We learned above that processes changing the instanton number are non-perturbative, so we see that though j_μ is not conserved, its charge is conserved *in perturbation theory.*

14.5. 't Hooft anomaly matching conditions

There is one other property of anomalies that will be important to us. It concerns the triangle diagrams with three global currents which we saw before lead to no anomalous symmetry-breaking. Nevertheless, the following beautiful argument of 't Hooft shows that they compute scale-independent information about the theory.

Consider a theory described by a Lagrangian $\mathcal{L}$ at some scale μ, with global (non-anomalous) symmetries generated by currents j_a^μ. Gauge these symmetries by adding in new gauge fields A_μ^a, which I'll call "spectator" gauge fields, thus giving the new theory

$$\mathcal{L}' = \mathcal{L} + \int d^4x \left[\frac{1}{g^2}\mathrm{tr}F^2 + j_a \cdot A^a\right]. \tag{14.48}$$

This may not be a consistent theory, however, due to non-vanishing triangle diagrams for the newly-gauged currents j_a. In that case, add in a set of new (spectator) free fermion fields ψ^S in representations to exactly cancel the anomalies and couple them only to the spectator gauge fields. Denoting the currents of the spectator fermions by j_a^S, we then have the enlarged and anomaly-free theory

$$\mathcal{L}'' = \mathcal{L} + \int d^4x \left[\frac{1}{g^2}\mathrm{tr}F^2 + + \overline{\psi}^S \partial\!\!\!/ \psi^S + (j_a^S + j_a) \cdot A^a \right]. \tag{14.49}$$

Since the spectator theory can be made arbitrarily weakly-coupled by taking $g \to 0$, the IR dynamics of the enlarged theory are just the IR dynamics of the original theory plus the arbitrarily weakly-coupled spectator theory. Thus the anomalies in the spectator theory are just the same as in the UV, and since the whole theory is anomaly-free, the anomaly from the IR currents j_a must also still be the same as in the UV. We can now throw away the spectator theory (take $g = 0$), to learn that the coefficient of the triangle diagram $\mathrm{tr}_R(T^a\{T^b, T^c\})$ for the *global* currents must be the same in the IR as in the UV.

The importance of this result is that the original theory might have been strongly-coupled in the IR in terms of its UV degrees of freedom, so the IR effective action may *a priori* be described by a completely different set of fermionic fields transforming under the global symmetries than appeared in the microscopic description. But 't Hooft's argument gives constraints on the possible IR fermion content, by demanding that their "anomalies" be the same as those of the UV fermions.

15. Non-renormalization in supersymmetric gauge theories

We now turn to the quantum-mechanical properties of supersymmetric gauge theories. Our aim is to prove non-renormalization theorems for supersymmetric gauge theories, along the lines of the non-renormalization theorem we proved for the nlσm. The results we find will be considerably weaker than in the nlσm case, essentially because gauge theories can be AF (and thus strongly-coupled in the IR).

15.1. Supersymmetric selection rules

We start by examining the analog of the holomorphy of the superpotential for an AF supersymmetric gauge theory. In the action at a scale μ_0, the terms which can be written only as integrals over half of superspace, and therefore must have holomorphic dependence on their fields couplings, are the gauge kinetic and superptential terms, which we'll assemble into a "generalized superpotential"

$$\widetilde{\mathcal{W}}_0 = \frac{\tau_0}{32\pi i}\mathrm{tr}(W_\alpha^2) + \mathcal{W}_0(\Phi^i, \lambda_r). \tag{15.1}$$

Here λ_r are the couplings appearing in the tree-level superpotential

$$\mathcal{W}_0 = \sum_r \lambda_r X_r, \tag{15.2}$$

where X_r are gauge-invariant composite operators of the Φ^i's. The factors of μ_0 are inserted for convenience to absorb the clssical scaling dimensions.

Recall that the gauge coupling is

$$\tau_0 \equiv \frac{\vartheta}{2\pi} + i\frac{4\pi}{g_0^2} = \frac{1}{2\pi i}\log\left[\left(\frac{|\Lambda|}{\mu_0}\right)^{b_0} e^{i\vartheta}\right], \tag{15.3}$$

where we have used the *definition* of the strong coupling scale $|\Lambda|$ in the last step. (The absolute value is to remind us that it is a postive real number.) It is thus natural to define a *complex* "scale" in supersymmetric gauge theories by

$$\Lambda = |\Lambda| e^{i\vartheta/b_0} \qquad \Rightarrow \qquad \tau_0 = \frac{b_0}{2\pi i}\log\left(\frac{\Lambda}{\mu_0}\right). \tag{15.4}$$

Recall that b_0 is the coefficient of the one-loop beta function, given by

$$b_0 = \frac{11}{6}T(adj) - \frac{1}{3}\sum_i T(R_i) - \frac{1}{6}\sum_a T(R_a). \tag{15.5}$$

where the indices i run over Weyl fermions and a run over complex bosons. In a supersymmetric gauge theory, the vector multiplet always includes a Weyl fermion in the adjoint representation (the gaugino), while each chiral multiplet Φ^i has one Weyl fermion and one complex boson, transforming in the same representation of the gauge group R_i. Thus, for supersymmetric gauge theories, b_0 simplifies to

$$b_0 = \tfrac{3}{2}T(adj) - \tfrac{1}{2}\textstyle\sum_{i\in\chi\mathrm{sf}} T(R_i). \tag{15.6}$$

We assume we are dealing with an AF theory, so if we take the scale $\mu_0 \gg |\Lambda|$, then the theory is weakly coupled (we might also have to take some of the superpotential couplings to be small). Let us consider how this effective theory will change is we run it down in scale a little to $\mu < \mu_0$. As long as the ratio μ/μ_0 is not too small, the theory should remain weakly-coupled, and we expect that the effective theory should be describable in terms of the same degrees of freedom. The effective generalized superpotential will then be

$$\frac{\tau(\Lambda, \Phi^i, \lambda_r; \mu)}{32\pi i}\mathrm{tr}(W_\alpha^2) + \mathcal{W}(\Phi^i, \lambda_r, \Lambda; \mu) + \text{irrelevant operators.} \tag{15.7}$$

Here we have written the effective coupling τ and the superpotential as general holomorphic functions of the χsf's and the bare couplings, as befits terms that appear only as integrals over half of superspace. The irrelevant operators include terms with higher powers of $\mathrm{tr}(W_\alpha^2)$, since W_α has scaling dimension 1 even in vacuum scaling.

However, we have forgotten to take into account the angular nature of the ϑ-angle:

$$\vartheta \simeq \vartheta + 2\pi \qquad \Rightarrow \qquad \tau \simeq \tau + 1. \tag{15.8}$$

Thus $\tau(\Lambda, \Phi, \lambda)$ is *not* a general holomorphic function, rather it is a "section of a line bundle" over target space—which means only that as we rotate the phase of Λ^{b_0},

$$\Lambda^{b_0} \to e^{2\pi i}\Lambda^{b_0}, \tag{15.9}$$

we must have $\tau \to \tau + 1$. This constrains the functional form of τ to be

$$\tau(\Phi, \Lambda, \lambda; \mu) = \frac{b_0}{2\pi i}\log\left(\frac{\Lambda}{\mu}\right) + f(\Lambda^{b_0}, \Phi^i, \lambda_r; \mu), \tag{15.10}$$

where f is now an arbitrary holomorphic *function* of its arguments.

Since we are dealing with an AF theory, the $\Lambda \to 0$ limit corresponds to the weak-coupling limit, in which the effective couplings should not diverge. Thus we have

$$\tau = \frac{b_0}{2\pi i}\log\left(\frac{\Lambda}{\mu}\right) + \sum_{n=1}^{\infty}\Lambda^{b_0 n} a_n(\Phi^i, \lambda_r; \mu), \tag{15.11}$$

(*i.e.* inverse powers of Λ^{b_0} so not appear). By comparing this expression to the perturbative expansion, where $\log\Lambda \sim 1/g^2$ and $\Lambda^{b_0 n} \sim n$-instanton action, we see that *the gauge coupling τ in the Wilsonian effective action only gets one-loop corrections in perturbation theory, though non-perturbative corrections are allowed.* This is a "not-much renormalization" theorem.

The superpotential satisfies a similar constraint. If we turned-off the gauge coupling ($\Lambda \to 0$) then we would have the old (nlσm) non-renormalization theorem for the superpotential in the Wilsonian effective action which says it does not get renormalized at all: $\mathcal{W}(\Phi, \lambda; \mu) = \mathcal{W}_0(\Phi, \lambda)$. Turning on the gauge group can then only add new terms holomorphic in Λ^{b_0} and vanishing as $\Lambda \to 0$:

$$\mathcal{W} = \mathcal{W}_0 + \sum_{n=1}^{\infty}\Lambda^{b_0 n} b_n(\Phi^i, \lambda_r; \mu), \tag{15.12}$$

implying no perturbation-theory corrections, but possible non-perturbative corrections.

We should emphasize the limitations of these "non-renormalization" theorems: they are only derived for weakly-coupled theories where the description in terms of the microscopic degrees of freedom is good. As we run the RG down to the IR, the theory will become strongly-coupled, and our description in terms of the Φ_i and W_α fields will break down. More technically, as we run down in scale at some point we can no longer be sure that the "irrelevant operators" in (15.7)—as well as other irrelevant operators appearing elsewhere in the effective action besides the generlized superpotential—are really irrelevant. In particular, these non-renormalization theorems in no way solve the essential strong-coupling problem of AF gauge theories by themselves.

15.2. Symmetries and selection rules

The above non-renormalization theorems can be sharpened in an important way by using the selection rules of other global symmetries in the theory. An important new element is the treatment of the selection rules stemming from anomalous symmetries.

Consider the global symmetry, $U(1)_j$,which rotates only one χsf, Φ^j. Thus

$$U(1)_j: \ \Phi^k \to e^{i\alpha\delta_{jk}}\Phi^k. \tag{15.13}$$

As we saw in the last lecture, this symmetry is anomalous, and can be considered as having the effect of rotating the ϑ-angle

$$U(1)_j: \ \vartheta \to \vartheta + \alpha T(R_j), \tag{15.14}$$

where R_j is the gauge-group representation of Φ^j. So for supersymmetric theories we can express this by giving Λ an effective charge under $U(1)_j$:

$$U(1)_j: \ \Lambda^{b_0} \to e^{i\alpha T(R_i)}\Lambda^{b_0}. \tag{15.15}$$

This gives a selection rule for possible terms appearing in τ and $\mathcal{W}$ due to this anomalous symmetry.

Another symmetry is the R-symmetry of our theory. Recalling the R-charge of the superspace Grassmann variable: $R(\theta) = 1$, it follows that for $\int d^2\theta \mathrm{tr} W^2$ to be R-invariant we must have

$$R(W_\alpha) = 1 \qquad \Rightarrow \qquad R(\lambda_\alpha) = 1, \ R(F_{\mu\nu}) = R(A_\mu) = 0. \tag{15.16}$$

This implies that $R(V) = 0$, so that the Kahler terms are automatically R-invariant. If we define the R-charges of the χsf's to be zero then

$$R(\Phi^i) = 0 \qquad \Rightarrow \qquad R(\phi^i) = 0, \ R(\psi^i) = -1. \tag{15.17}$$

Looking at the charges of the fermions, we see that the anomaly for this symmetry can be compensated by assigning R-charge to Λ of

$$R(\Lambda^{b_0}) = T(adj) - \sum_i T(R_i). \tag{15.18}$$

Let us apply these symmetries to the holomorphic parts of our effective theory—our generalized superpotential $\widetilde{\mathcal{W}}$—which contains the gauge kinetic terms as well as the superpotential. The microscopic generalized superpotential for our theory was

$$\widetilde{\mathcal{W}}_0 = \frac{1}{64\pi^2}\log(\Lambda/\mu_0)^{b_0} S + \sum_r \lambda_r X_r, \tag{15.19}$$

where I have defined the composite chiral superfields

$$S \equiv -\mathrm{tr}(W^\alpha W_\alpha), \tag{15.20}$$

and the gauge-invariant composite X_r's built out of the Φ^i's appearing in the superpotential built out of the Φ^i's. Then our non-renormalization theorem says that the effective generalized superpotential is

$$\widetilde{\mathcal{W}} = \frac{1}{64\pi^2}\log(\Lambda/\mu)^{b_0} S + \sum_r \lambda_r X_r + f(\Lambda^{b_0}, \lambda_r, S, \Phi^i; \mu), \tag{15.21}$$

for some some holomorphic function f which must vanish in the weak-coupling limits $\Lambda, \lambda_r \to 0$.

We now would like to apply the selection rules following from the $\prod_i U(1)_i \times U(1)_R$ global symmetries to constrain the form of f. The analysis of these constraints is made simpler by ignoring the constraints that come from gauge-invariance—we can put back in this extra constraint at the end of our analysis. So, we analyze instead the selection rule constraints on a generalized superpotential where we introduce a tree-level coupling for each χsf Φ^i:

$$\widetilde{\mathcal{W}} = \frac{1}{64\pi^2}\log(\Lambda/\mu)^{b_0} S + \sum_i \lambda_i \Phi^i + f(\Lambda^{b_0}, \lambda_r, S, \Phi^i; \mu). \tag{15.22}$$

The charges of all the fields and couplings can be summarized as:

$$\begin{array}{cccc} & U(1)_i & \times & U(1)_R \\ S & 0 & & 2 \\ \Phi^j & \delta_{ji} & & 0 \\ \lambda_j & -\delta_{ji} & & 2 \\ \Lambda^{b_0} & T(R_i) & & T(adj) - \sum_i T(R_i) \end{array} \tag{15.23}$$

We emphasize that the λ_i couplings are not gauge invariant; however, transformation properties of any physical coupling λ_r of a gauge-invariant composite operator $X_r = \prod \Phi^j$ in the microscopic superpotential under the global $U(1)_R \times \prod_i U(1)_i$ symmetries will be the same as those of the corresponding product of λ_i's $\prod_j \lambda_j$.

The general term in f is of the form

$$\Lambda^{b_0 n} S^m \prod_i \left(\lambda_i^{n_i} (\Phi^i)^{m_i}\right). \tag{15.24}$$

The selection rules then imply the conditions

$$\begin{aligned} U(1)_i: &\qquad 0 = nT(R_i) - n_i + m_i \\ U(1)_R: &\qquad 2 = n\left(T(adj) - \sum_i T(R_i)\right) + 2m + 2\sum_i n_i. \end{aligned} \tag{15.25}$$

Summing the $U(1)_i$ conditions and adding them to the $U(1)_R$ condition implies

$$2 = nT(adj) + 2m + \sum_i (n_i + m_i). \tag{15.26}$$

Since by our previous arguments $n \geq 1$ and $n_i \geq 0$, the only solutions to this constraint have some $m_i < 0$ (or $m < 0$). Thus all the possible non-perturbative corrections involve inverse powers of the χsf's. In particular, we find that there are *no non-perturbative corrections to the tree-level couplings.* In other words, the function f contains non-perturbative corrections to the generalized superpotential in the form of new operators that may be generated, but none of them are proportional to the tree-level operators S and X_r. This implies that the one-loop β-function for the gauge coupling is exact, and that the λ_r couplings are unrenormalized. A caveat on these seemingly very strong results is that they refer only to the renormalization of the couplings in the Wilsonian effective action, and not to the physical couplings; we will return to this point in the next lecture.

15.3. IRF gauge theories and FI terms

So far we have been discussing only AF (and therefore non-Abelian) gauge theories. Clearly similar arguments can be applied to IRF gauge theories, as long as we take the scale of our theory *low* enough—$\mu_0 \ll \Lambda$—so that the theory is weakly coupled. Then the RG running to the IR will just make the theory more weakly coupled, so the effective theory should be described by the same degrees of freedom. We once again find that the gauge coupling is only renormalized at one loop in perturbation theory, and that all non-perturbative corrections must be proportional to inverse powers of Λ^{b_0}. This changes the sign of n in the above selection rules, and so allows non-perturbative corrections to β-functions. Thus, even though these theories are IRF, it may not be trivial to compute their low-energy effective couplings! (Since IRF theories are UV strongly-coupled, the question of their IR effective couplings is largely moot, unless they are realized as effective theories of some microscopic physics with different degrees of freedom, *e.g.* an AF gauge theory whose gauge group is spontaneously broken down to IRF groups.)

In the case when the IRF gauge groups are Abelian, then there exist techniques relying on the *electric-magnetic duality* of the low-energy effective actions which have proved to be strong-enough to exactly determine the non-perturbative corrections to the low-energy couplings. We will briefly introduce touch on this topic at the end of the course.

We have seen that Abelian gauge theories admit one extra kind of term, the FI D-term,

$$\mathcal{L}_{FI} = \tfrac{1}{2} \int d\theta^\alpha \kappa_0 W_\alpha + h.c. \tag{15.27}$$

Here κ_0 is a complex number, since for this term to be a supersymmetry invariant, κ_0 must satisfy

$$D^\alpha \kappa_0 = \overline{D}_{\dot\alpha}\kappa_0 = 0, \tag{15.28}$$

which implies it must be a constant. Only the real part of κ_0 enters the action by virtue of the Bianchi identity

$$D^\alpha W_\alpha = \overline{D}_{\dot\alpha}\overline{W}^{\dot\alpha}. \tag{15.29}$$

The supersymmetric selection rule for the FI term is thus that there can be arbitrary κ_0-dependent corrections to the superpotential, gauge, and Kahler terms, but that the superpotential and gauge couplings and the (vevs of) any χsf's can *not* enter into quantum corrections of the FI term. The only other parameters in the theory are the bare scale μ_0 at which the theory is defined, the RG scale μ, and the gauge charges of the χsf's q_i. Furthermore, by gauge-invariance, the charges and κ_0 itself can only enter in physical amplitudes in the combinations $g\kappa_0$ and gq_i, where g is the $U(1)$ gauge coupling. Thus the effective FI-term coefficient, by $\kappa(\mu)$ must satisfy

$$g\kappa = \mu^2 f\left(\frac{g\kappa_0}{\mu_0^2}, gq_i; \frac{\mu}{\mu_0}\right), \tag{15.30}$$

where I have used dimensional analysis to fix some of the μ-dependence. However, by our supersymmetric selection rule, κ must be g-independent, implying $\kappa = \kappa_0\, f_0(\mu/\mu_0) + q_i f_i(\mu/\mu_0)$ for some functions f_a. Now, in the limit $q_i \to 0$, the $U(1)$ field is completely decoupled, and so the FI-coefficient should have its free value, implying $f_0 = (\mu/\mu_0)^2$ from the tree-level (classical) rescaling. The term proportional to q_i comes by definition from the first order in pertrubation theory, namely the tadpole graph with the vector superfield attached to a loop of χsf's. The exact result for the effective FI constant is thus

$$\kappa = \kappa_0\left(\frac{\mu}{\mu_0}\right)^2 + \left(\sum_i q_i\right)\log\left(\frac{\mu}{\mu_0}\right). \tag{15.31}$$

Finally, in any realistic theory, the requirement of absence of mixed gauge-gravitational anomalies requires the trace of the $U(1)$ charges to vanish. Thus we learn that the FI term is exactly unrenormalized.

16. Exact β-functions, supersymmetric anomaly, and finiteness

The exact one-loop β-function that we derived in the last lecture has to be interpreted with some care, for, as was pointed out by V. Novikov, M. Shifman, A. Vainshtein, and V. Zakharov, *Phys. Lett.* 166B (1986) 329, and M. Shifman, A. Vainshtein, and V. Zakharov, *Phys. Lett.* 166B (1986) 334, and M. Shifman and A. Vainshtein, *Nucl. Phys.* B277 (1986) 456, the β-function we have determined is not what we mean by the physical β-function. Nevertheless, there is a definite relation between the two which will allow us to derive an exact relation between the physical β-function and the anomalous dimensions of the χsf's.

16.1. Wave-function renormalization

We write the action for an AF theory at a large enough scale μ_0 so that it is weakly coupled as

$$\mathcal{L}_{\mu_0} = \int d^4\theta\, \overline{\Phi}_i e^V \Phi^i + \int d^2\theta \left[-\frac{b_0}{64\pi^2} \log(\Lambda/\mu_0)\, \mathrm{tr}(W_\alpha^2) + \sum_r \lambda_r X_r \right] + h.c. \tag{16.1}$$

Here λ_r are the couplings appearing in the tree-level superpotential asociated with the gauge-invariant composite operators

$$X_r = \prod_i (\Phi^i)^{r_i} \tag{16.2}$$

for some integers r_i.

We have seen that the Wilsonian effective coupling, which we'll denote $\tau_W(\mu)$ runs with scale as

$$2\pi i \tau_W(\mu) = b_0 \log\left(\frac{\Lambda}{\mu}\right), \tag{16.3}$$

where recall that

$$b_0 = \tfrac{3}{2} T(adj) - \tfrac{1}{2} \textstyle\sum_{i \in \chi\mathrm{sf}} T(R_i). \tag{16.4}$$

Then we have seen that the effective theory at a scale $\mu < \mu_0$, as long as the ratio μ/μ_0 is not too small, is

$$\begin{aligned}\mathcal{L}_\mu = &\int d^4\theta\, Z_i(\mu) \overline{\Phi}_i e^V \Phi^i + \int d^2\theta\, \frac{\tau_W(\mu)}{32\pi i}\, \mathrm{tr}(W_\alpha^2) \\ &+ \int d^2\theta \left[\sum_r \left(\frac{\mu}{\mu_0}\right)^{3-d_r} \lambda_r X_r + \text{non-perturbative operators} \right] + h.c. + \text{irrelevant operators.}\end{aligned} \tag{16.5}$$

This is the Wilsonian effective action at the scale μ. Here we have included the wave-function renormalizations of the Kahler terms,[23] and the classical scalings of the superpotential couplings: d_r is the classical (kinetic) scaling dimension of the X_r composite operator, given by

$$d_r = \sum_i r_i. \tag{16.6}$$

[23] One may wonder why I haven't written a possible multiplicative renomalization for the vector superfield V: $Z_i \overline{\Phi}_i e^{Z_{1F} V} \Phi^i$? This can certainly arise; however, the relation between the field-strength superfield W_α and V can also get renormalized: $W_\alpha \sim \overline{D}^2 (e^{-Z_3 V} D_\alpha e^{Z_3 V})$. Gauge invariance requires that $Z_{1F} = Z_3$, so that these redefinitions can be trivially scaled out of the effective action.

(Back in lectures 8 and 9 when we did wave-function renormalization on the nlσm I left out these factors out of laziness and because we weren't being very specific about the scales at which we were defining our effective actions.)

In order to compare the couplings of this effective action to physical couplings that would be measured in, say, a scattering experiment with energy transfer of order μ, the first thing we must do is normalize the kinetic terms to their canonical form—*i.e.* perform wave-function renormalization. This is done simply by rescaling the χsf's by

$$\Phi^i \to \frac{\mu}{\mu_0}\sqrt{Z_i(\mu)}\,\Phi^i. \tag{16.7}$$

Then the renormalized Lagrangian has the same form as the Wilsonian one, but with the bare superpotential couplings replaced by running ones:

$$\lambda_r(\mu) \equiv \left(\frac{\mu}{\mu_0}\right)^{3-d_r}\left(\prod_i Z_i^{-r_i/2}\right)\lambda_r. \tag{16.8}$$

This immediately implies the exact RG equation for the physical superpotential couplings

$$\frac{d\lambda_r(\mu)}{d\log\mu} = \lambda_r(\mu)\left(3 - d_r - \tfrac{1}{2}\sum_i r_i\gamma_i(\mu)\right), \tag{16.9}$$

where we have defined the anomalous dimension of the Φ^i χsf as

$$\gamma_i(\mu) \equiv \frac{d\log Z_i(\mu)}{d\log\mu}. \tag{16.10}$$

Of course, we have no "exact" method of computing these anomalous dimensions.

It will be useful to note that the RG equation for the effective mass $m_i(\mu)$, defined as the coefficient of the term $m_i\Phi^i\Phi^i$ in the wave-function renormalized superpotential (assuming such a term is allowed by gauge invariance), is $dm_i/d\log\mu = m_i(1-\gamma_i)$. So γ_i is the anomalous dimension of the mass.

16.2. Relation between Wilsonian and physical β-functions.

Wave-function renormalization is not the whole story in relating Wilsonian to physical couplings. One way this can be understood is by noting that physical couplings are those measured in, say, scattering processes, whose amplitudes are expressed in terms of vacuum correlation functions. So, when expanded in perturbation theory, the effective coupling at a scale μ will be computed by diagrams with external momenta at $p^2 \sim \mu^2$. At tree-level, this will just involve the Wilsonian coupling, $\lambda(\mu)$, as we have computed above. However, at one loop (and higher), which we have seen contributes to the gauge coupling, the relevant amplitude will involve an integral over internal states running in the loop.

When we rescale the fields as in (16.7) we will not only rescale the external legs, but also the internal loop states. Thus the resulting physical couplings in general will not depend simply on the wave-function renormalizations Z_i at one loop and higher.

A physical way of computing this non-linear effect in the case of the gauge couplings is to suppose that all the χsf's have (gauge-invariant) bare masses m_i such that at the scale μ their physical masses $m_i(\mu) > \mu$. Also, add an extra χsf Σ in the adjoint representation of the gauge group G, along with some superpotential which gives it a tree-level mass m_Σ and vev $\langle\Sigma\rangle$ Higgsing the gauge group down to its maximal Abelian subgroup: $G \to U(1)^{\mathrm{rank}G}$. (We will decouple this field at the end by sending its mass $m_\Sigma \to \infty$.) The charged vector bosons $W^\pm$ get masses $\sim \langle\Sigma\rangle g$ by the Higgs mechanism, which in supersymmetric notation gives the physical masses

$$m_W(\mu) = \frac{\mu Z_\Sigma^{1/2}}{\mu_0}\langle\Sigma\rangle\sqrt{\frac{4\pi i}{\tau_P}}, \tag{16.11}$$

where τ_P is the *physical coupling*, and the factor of μ/Z_Σ comes from the wave-function renormalization of the Σ superpotential. Likewise, the χsf's physical masses are given by

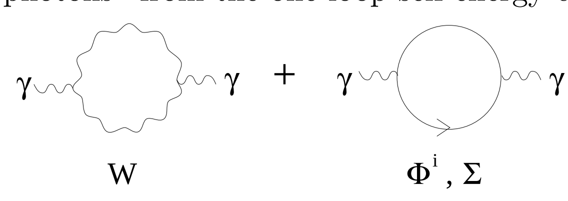
$$\begin{aligned} m_\Sigma(\mu) &= \frac{\mu}{\mu_0 Z_\Sigma} m_\Sigma, \\ m_i(\mu) &= \frac{\mu}{\mu_0 Z_i} m_i. \end{aligned} \tag{16.12}$$

Since we will be taking the bare mass of Σ to infinity, we can ceratinly take its mass to be greater than the microscopic scale μ_0 at which we are defining the theory. In this case the dynamics of the Σ field are free, and so its wave-function renormalization is just

$$Z_\Sigma(\mu) = \frac{\mu}{\mu_0}. \tag{16.13}$$

(Another way of seeing this is that since it is free, its physical mass must be the same as its bare mass.)

Since there is no one-loop running of the gauge coupling due to the contributions of these charged fields at scales below their masses, we can compute the Wilsonian effective coupling for the $U(1)^r$ "photons" from the one-loop self-energy diagrams



which are cut off by their physical masses, giving:[24]

$$2\pi i\tau_W(\mu) = T(adj)\log\left(\frac{\Lambda}{m_W(\mu)}\right) - \frac{1}{2}T(adj)\log\left(\frac{\Lambda}{m_\Sigma(\mu)}\right) - \frac{1}{2}\sum_i T(R_i)\log\left(\frac{\Lambda}{m_i(\mu)}\right). \tag{16.14}$$

Since this Wilsonian β-function is computed at a scale where all the matter has been integrated-out, it is the same as the physical coupling. Furthermore, adjusting the bare masses so that the physical masses of the W-bosons and the χsf's are all just below the scale μ, we see that the physical $U(1)^r$ coupling we have computed is the same as the physical gauge coupling of our original theory at the scale μ. Now, substituting in the μ-dependence of the physical masses worked-out above, we thus compute the physical β-function to be given by the exact expression[25]

$$\begin{aligned}\beta(\mu) \equiv -2\pi i\frac{d\tau_W}{d\log\mu} &= \tfrac{3}{2}T(adj) + T(adj)\left(\tfrac{3}{2} - \tfrac{\beta(\mu)}{4\pi i\tau_P(\mu)}\right) - \tfrac{1}{2}\textstyle\sum_i T(R_i)\left(1-\gamma_i(\mu)\right),\\ &= \frac{b_0 + \frac{1}{2}\sum_i T(R_i)\gamma_i}{1+\frac{T(adj)}{4\pi i\tau_P}}.\end{aligned} \tag{16.15}$$

This is the one-loop Wilsonian β-function corrected by the anomalous mass dimensions (which involve higher-loop effects) as well as by an overall factor which depends on the exact coupling. Conceptually, the anomalous dimensions of the χsf's entered because they were circulating in the loops; in the first diagram (which occurs only for non-Abelian gauge theories) the vsf circulating in the loop also gives some contribution due to its wave-function renormalization. This result was first derived by V. Novikov, M. Shifman, A. Vainshtein, and V. Zakharov, *Nucl. Phys.* B229 (1983) 381, by different methods.

16.3. The supersymmetric form of the anomaly

Before discussing some of the physical implications of the above exact results, we pause to give an alternate, indirect argument, involving the "supermultiplet of anomalies". The basic idea is that the non-trivial dependence of the physical couplings on the anomalous

[24] I am being alittle sloppy here: what I am really calculating is the Wilsonian coupling of the $U(1)^r$ photons, and so instead of indices of representations of the group G, I should have sums of squares of components of the weight vectors of these representations (*i.e.* their charges under the $U(1)^r$'s). It is a Lie algebra fact that these sums are indeed proportional to the indices of the representations; I have simply picked the factors of proportionality so that I recover the one-loop β-function in the end.

[25] I have fudged the sign in the denominator! The answer written below is the correct one, thought I can't seem to get the sign right...

dimensions that we have just found is just an example of the familiar fact that quantum effects break the classical invariance under rescalings, such as are involved in performing the wave-function renormalizations (16.7)—*i.e.* scale invariance is an anomalous symmetry in QFT. To compare to physical processes we not only need to rescale the χsf's but also the vsf:

$$\begin{aligned} \Phi^j &\to \sqrt{Z_j}\,\Phi^j \\ W_\alpha &\to \sqrt{\frac{\tau_P}{32\pi i}}\,W_\alpha . \end{aligned} \tag{16.16}$$

Note that the physical coupling τ_P appears in the rescaling of the vsf. We can think of these as complexified chiral rotations of the superfields:

$$\begin{aligned} \Phi^j &\to e^{i\alpha_j}\Phi^j, \qquad &\text{with } i\alpha_j = \tfrac{1}{2}\log(Z_j), \\ W_\alpha &\to e^{i\alpha_0}W_\alpha, \qquad &\text{with } i\alpha_0 = \tfrac{1}{2}\log(\tau_P/32\pi i). \end{aligned} \tag{16.17}$$

Such chiral rotations are anomalous, implying that the ϑ-angle is shifted, which in complexified language translates to

$$2\pi i\tau_W \to 2\pi i\tau_P = 2\pi i\tau_W + \tfrac{1}{2}T(adj)\log(\tau_P/32\pi i) + \tfrac{1}{2}\sum_i T(R_i)\log(Z_i), \tag{16.18}$$

giving rise to precisely the same β-function as above (again with the wrong sign, though!).

More formally, this supersymmetrization of the anomaly can be understood in terms of supermultiplets of currents. It turns out that the (classical) chiral current j_μ rotating all the Φ^i χsf's is part of a supermultiplet of currents:

$$\{j_\mu, s^{\dot\alpha}_\mu, t_{\mu\nu}\} \tag{16.19}$$

which is written as a real superfield $V_{\alpha\dot\alpha} = \sigma^\mu_{\alpha\dot\alpha} j_\mu + \ldots$. Here $s^{\dot\alpha}_\mu$ and $t_{\mu\nu}$ are the supersymmetry current and energy-momentum tensor. It follows that the divergences of these currents also fall into a supermultiplet:

$$\overline{D}^{\dot\alpha} V_{\alpha\dot\alpha} \sim \{\not{s}_\alpha, \partial^\mu j_\mu, t^\mu_\mu\}. \tag{16.20}$$

The classical conservation of the supercurrent $\overline{D}^{\dot\alpha} V_{\alpha\dot\alpha} = 0$ thus implies the vanishing of the trace of the energy-momentum tensor. A non-zero trace of the energy-momentum tensor just reflects broken scale-invariance since $t^\mu_\mu = \partial^\mu j^D_\mu$, where $j^D_\mu = x^\nu t_{\mu\nu}$ is the dilatation current. So, conservation of the supercurrent is broken classically by explicitly non-scale-invariant superpotential interactions and quantum-mechanically by scaling modifications from gauge β-functions and matter anomalous dimensions:

$$\overline{D}^{\dot\alpha} V_{\alpha\dot\alpha} = D_\alpha \left(3\mathcal{W} - \sum_i \Phi^i \frac{\delta\mathcal{W}}{\delta\Phi^i} - \frac{b_0}{64\pi^2}\mathrm{tr}(W^2_\alpha) + \frac{1}{8}\overline{D}^2 \sum_i \gamma_i Z_i \overline{\Phi}_i e^V \Phi^i \right). \tag{16.21}$$

These considerations apply equally to the (anomalous) chiral rotation of a single χsf Φ^j, whose current will also be part of a multiplet of currents. We can find the superfield form of the "divergence multiplet" of this current by noting that the divergence of the chiral current is computed in the Noether method by varying the action under $\Phi^i \to e^{i\alpha}\Phi^i$ and pulling out the piece proportional to α. Rewriting the Kahler term as $\int d^2\theta\,\overline{D}^2(\overline{\Phi}e^V\Phi)$, we thus find that the divergence multiplet is proportional to $\overline{D}^2(\overline{\Phi}e^V\Phi)$. Noting that the chiral anomaly involves $\mathrm{tr}(F\widetilde{F})$ which appears as a component of the superfield $\mathrm{tr}(W^\alpha W_\alpha)$, it is not hard to write down the *supersymmetric anomaly* (also called the Konishi anomaly)

$$\frac{1}{4}\overline{D}^2\left(Z_i\overline{\Phi}_i e^V\Phi^i\right) = \frac{1}{64\pi^2}T(R_i)\mathrm{tr}(W_\alpha^2) + \Phi_i\frac{\delta\mathcal{W}}{\delta\Phi^i} \tag{16.22}$$

valid for all i. Here we have gone to the Wilsonian effective action and so included the Kahler-term renormalization, as well as an effective superpotential $\mathcal{W}$.

▸ **Exercise 16.1.** Expand the supersymmetric anomaly in components (in WZ gauge), and verify that it includes the chiral anomaly.

Thus, the chiral anomaly is a component of a *multiplet of anomalies*.

Plugging the anomaly-corrected equations of motion (16.22) into (16.21) then implies our previous exact conditions for (quantum) scale invariance:

$$\begin{aligned} 0 = \beta_g &\propto b_0 + \tfrac{1}{2}\sum_i T(R_i)\gamma_i \\ 0 = \beta_{\lambda_r} &\propto 3 - d_r - \tfrac{1}{2}\sum_i r_i\gamma_i. \end{aligned} \tag{16.23}$$

16.4. Scale-invariance and finiteness

Theories with vanishing β-functions are said to be "fixed points" of the RG. Thus (16.23) give exact conditions for quantum scale-invariant theories. Though we do not know the exact functional form of the anomalous dimensions $\gamma_i(g,\lambda_r)$, there are still cases where we can use these conditions to find non-trivial results.

For example, consider an $SU(N)$ gauge theory with $2N$ χsf's Q^i in the $\mathbf{N}$ of $SU(N)$, and another $2N$ χsf's $\widetilde{Q}_i$ in the $\overline{\mathbf{N}}$. (This is "superQCD" with $2N$ flavors.) If the theory also has a χsf Φ in the adjoint of the gauge group, we can consider adding the operator

$$\mathcal{W} = \lambda\sum_i \mathrm{tr}\widetilde{Q}_i\Phi Q^i. \tag{16.24}$$

It is easy to check that with this matter content, the theory is one-loop scale-invariant: $b_0 = 0$. Since the Q's and $\widetilde{Q}$'s all enter symmetrically (there is an $SU(2N)$ flavor symmetry), they will all have the same anomalous dimension γ_Q. We then find

$$\beta_g \propto \beta_\lambda \propto \gamma_\Phi + 2\gamma_Q, \tag{16.25}$$

so there is only one condition on the λ–g parameter plane for scale invariance. Thus there will be a line of fixed-points; furthermore, at weak-coupling $(g, \lambda \ll 1)$ the anomalous dimensions vanish, so this line of fixed points goes through the origin. (In fact, it turns out that the exact curve is $\lambda = g$ and the scale-invariant theories along have an N=2-extended supersymmetry.)

As another example, consider a theory with three adjoint χsf's Φ^i $(i = 1, 2, 3)$ and the superpotential $\mathcal{W} = \lambda \mathrm{tr} \Phi^1 \Phi^2 \Phi^3$. Then their anomalous dimensions are all equal $\gamma_i = \gamma$ and $\beta_g \propto \beta_\lambda \propto \gamma$. Again there is a fixed line passing through weak-coupling. (In fact, the exact line is $g = \lambda$ and this time the fixed-point theory has N=4-extended supersymmetry.) Here the fixed-point theories are not only scale-invariant, but also *finite*: all the anomalous dimensions as well as the β-functions vanish!

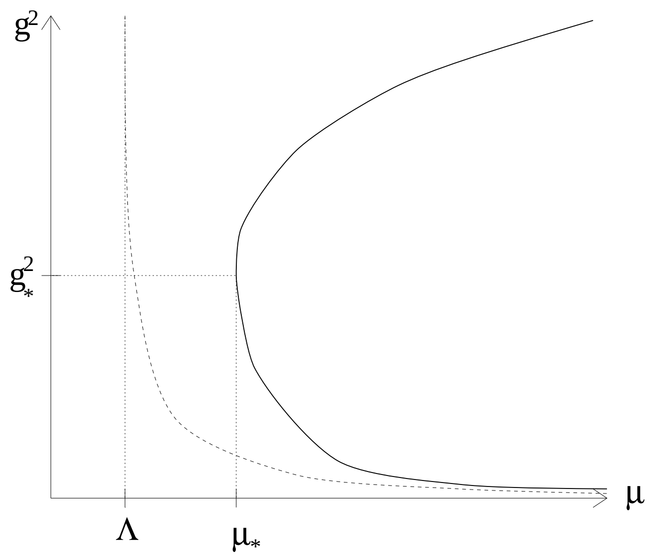
There are many more examples one can construct along these lines (including ones for which there is no accidental enhancement to extended supersymmetry); see R. Leigh and M. Strassler, *Nucl. Phys.* B447 (1995) 95.

16.5. Breakdown of the exact results

The exact β function (16.15) for, say, a pure non-Abelian gauge theory can be written

$$\beta(\mu) = \frac{\frac{3}{2} T(adj)}{1 + \frac{T(adj)}{4\pi i \tau}}, \tag{16.26}$$

which can be integrated to give the following RG flow in the τ-plane (for $\vartheta = 0$):



where $g_*^2 \sim 1/T_{adj}$ and $\mu_* \sim \Lambda / T_{adj}^{1/3}$. (The dashed line shows the unmodified one-loop running of the coupling for comparison.) What is the meaning of the pole in the β-function at $g^2 = g_*^2$ and the associated strong-coupling "branch" of the coupling? What happens at scales less than μ_*?

The point is that the exact β-function has become invalid by the time we reach these interesting-looking points: new relevant operators (which were irrelevant at weak coupling) were probably generated upon flowing to strong couling in the IR. Thus we see that the supersymmetric non-renormalization theorems and related exact results do not solve by themselves the strong-coupling problem of AF gauge theory.

One interesting piece of information about strong-coupling which we do extract is that (for, say, $SU(N)$ gauge group) strong coupling occurs at couplings $\mathcal{O}(1/N)$, while the strong-coupling scale of the theory is $\mathcal{O}(\Lambda/N^{1/3})$, results which may be familiar from the large-N expansion.

It will be the purpose of the last third of the course to show how exact information on strongly-coupled supersymmetric gauge theories *can* be obtained. First, however, we will turn to an important weakly-coupled application of supersymmetric theories...

III. Supersymmetric Models of Particle Physics

17. Naturalness, supergravity, and the MSSM

In the next two lectures I'd like to give an overview of the application of supersymetry to building models of particle physics beyond the standard model (SM). I will try to emphasize the qualitative problems and successes that have been encountered in this endeavor. Since no single compelling model has emerged, though, I will not dwell on the details.

17.1. Physics beyond the SM and naturalness

The SM is not the final theory, but at best an effective theory "low-energy" theory: it does not include gravity, some of its couplings are not AF (it is strongly-coupled in the UV), and it has too many parameters. We can write the SM Lagrangian schematically as

$$\mathcal{L} = M^4 + M^2\phi^2 + M^0\left[(\mathcal{D}\phi)^2 + \overline{\psi}\not{\mathcal{D}}\psi + g^{-2}F_{\mu\nu}^2 + y\phi\overline{\psi}\psi + \lambda\phi^4\right] + M^{-2}\overline{\psi\psi}\psi\psi + \dots \quad (17.1)$$

where ψ represent the three generations of quarks and leptons, $F_{\mu\nu}$ the $SU(3)\times SU(2)\times U(1)$ gauge bosons, and ϕ the Higgs; and g are the gauge couplings, y the Yukawas, and λ the Higgs self-coupling.

This is the form we would expect for an effective theory good below the scale M. The philosophy of *naturalness* would lead us to expect that in such a theory all terms allowed by the symmetries should appear with coefficients of order 1, the idea being that these are the "generic" values the couplings will take at the scale M where new physics comes into play, and the M-dependence of the coefficients are then determined by the RG running. One should emphasize that naturalness is an untested principle; one can imagine other principles which may govern the matching to high energy physics, for example dynamical mechanisms which set the high-energy couplings at non-generic-looking fixed-point values, or a historical principle (anthropic principle) that picks out non-generic but "interesting" couplings. One bit of evidence in favor of naturalness is the fact that, with the possible exception of ϑ_{QCD} and the cosmological constant, all terms in the SM allowed by the gauge symmetries do indeed occur.

In any case, accepting the philosophy of naturalness, the natural question to ask is: what is the scale M of the SM? We can address this order by order in M.

- $\underline{M^{-2}\text{ terms:}}$ The absence of proton decay, FCNC's, and various non-standard processes meditated by irrelevant (dimension 6) operators implies their coefficients must be small. Assuming naturalness, this gives

$$M > 10^{15}\text{ GeV}. \quad (17.2)$$

- $\underline{M^0 \text{ terms:}}$ The classically marginal (dimension 4) operators all run logarithmically in perturbation theory:
  - $\underline{g}$: Only the $U(1)$ hypercharge is not AF, but it gets strong only at scales much greater than M_{pl}, the Planck mass where gravity becomes important, and so is not significant. All three couplings run to about the same value near $M \sim 10^{14-15}$ GeV, but do not unify.
  - $\underline{y}$: The Yukawas run as $dy/d\log\mu \sim y^3 - g^2 y$, and are AF since the gauge couplings dominate in all cases.
  - $\underline{\lambda}$: The Higgs self-coupling runs as $d\lambda/d\log\mu \sim \lambda^2 + y^2\lambda - y^4$ plus small gauge pieces since the Higgs is neutral under $SU(3)$. Now, λ has not been measured, but it is typically *not* AF, unless λ is very small. We can get a sense of the scales on which λ will become strong by looking at the next term.
- $\underline{M^2 \text{ terms:}}$ The Higgs potential, $-M^2\phi^2 + \lambda\phi^4$ implies $m_\phi^2 \sim M\sqrt{\lambda}$ and $\langle\phi\rangle \sim M^2/\sqrt{\lambda}$. Since by the Higgs mechanism $m_W \sim \langle\phi\rangle g$, we learn that $m_H^2 \sim 10\lambda m_W^2$, and from failed searches for the Higgs, we learn that $\lambda > .1$. So, from its RG runnning its associated strong-coupling scale should be

$$M \equiv M_{EW} \sim 10^{3-4} \text{ GeV}. \tag{17.3}$$

- $\underline{M^4 \text{ terms:}}$ This term is a constant—the cosmological constant—and so has no effect unless we include gravity, in which case it has drastic effects on the large-scale structure of the universe. From cosmological observations we learn that

$$M \equiv M_{cc} < 10^{-11} \text{ GeV}. \tag{17.4}$$

  Note that when we include gravity we add another M^2 term to our effective Lagrangian, $M^2\sqrt{g}R$. Matching to Newton's constant gives the Planck scale

$$M \equiv M_{pl} \sim 10^{19} \text{ GeV}. \tag{17.5}$$

It is clear that naturalness is in big trouble for physics beyond the SM, for it implies three very different scales for that physics:

$$\begin{array}{ccccc} \text{Cosmological const.} & \ll & \text{Electroweak scale} & \ll & \text{Planck scale} \\ 10^{-11} & \ll & 10^3 & \ll & 10^{19}\ \text{GeV} \end{array} \tag{17.6}$$

This naturalness problem is called the gauge hierarchy problem, and is often described as the unnatural fine-tuning a GUT or gravity theory would require to keep the Higgs mass so much smaller than the Planck mass. The resulting tension between needing new physics at the TeV scale *versus* not wanting to add large irrelevant operators is what keeps the whole enterprise of physics beyong the SM going despite the lack of experimental results.

There are two main proposals for the new TeV physics:

- Strongly-coupled physics (technicolor, top quark bound states, ...). Such schemes reportedly have serious problems getting a realistic spectrum and at the same time not giving rise to large irrelevant operators. An important part of the interest in supersymmetric models derives from the failure of this idea, and so it is important to understand its problems. Unfortunatley, I do not have the time (nor am I qualified) to give a survey of this topic.
- Supersymmetry. The idea is that the presence of this symmetry makes the small Higgs mass natural, since it is related to the mass of the Higgsino (its fermionic partner under supersymmetry) which in turn is protected by gauge invariance. Then, if supersymmetry is (effectively) broken at a scale $M_S \sim M_{EW}$, then the Higgs will naturally get a mass of the right magnitude.

  This supersymmetry scenario immmediately raises some questions:

(1) Why is $M_S \sim M_{EW}$? What causes electroweak breaking (gives the Higgs a vev)? This is answered by the hope that whatever mechanism breaks supersymmetry also happens to break the electroweak symmetry. We will see an example of this later.

(2) Why is $M_S \ll M_{pl}$? Isn't this just another naturalness problem?

There are two answers. First, recalling the non-renormalization theorem, one might say that this hierarchy is *technically natural*, since once one does the necessary fine-tuning of the tree-level couplings at the Planck scale, the couplings do not get radiative corrections. This seems unsatisfactory, for it only shifts the problem to the Planck scale where we have much less control over the physics. The second answer is to say that some symmetry of the Planck scale physics (say from string theory) sets the tree-level couplings to zero (*i.e.* a natural value). Then, by the non-renormalization theorems, only non-perturbative corrections can be generated, thus naturally generating the hierarchy of scales. Thus in this scenario, called *dynamical symmetry breaking*, the supersymmetry breaking occurs at the strong-coupling scale of some gauge group. One should note that this would seem to just amount to a supersymmetric version of technicolor models, and might be expected to bring along all their problems. We will see how this can be dealt with in the next lecture.

(3) Since the vacuum energy is the order parameter for supersymmetry breaking, doesn't $M_S \sim M_{EW}$ imply a large cosmological constant $M_{cc} \sim M_{EW}$?

This observation is a potentially damning one for supersymmetry, for it would seem to lead to a huge contradiction with observation. However—the first nice surprise of supersymmetry—there is a way out:

17.2. Supergravity

Supergravity is the effective field theory of gravity and supersymmetry below the Planck scale (where gravity becomes strong). I will just summarize the main features of supergravity—deriving them is a course in itself. See the second half of Wess and Bagger for more details.

The first point about supergravity is that in the presence of gravity, supersymmetry must be gauged. (The quickest way to see this is by noting that in the presence of non-trivial curvature, the covariantly-constant spinor needed to defined global supercharges does not exist.) Thus, the graviton will get a spin-$\frac{3}{2}$ gravitino as a superpartner. The gravitino is the "gauge fermion" of supergravity.

Making globally supersymmetric FT actions generally covariant and supergravity co-variant (in "curved superspace") gives rise to new terms in the theory. In particular, the scalar potential becomes

$$V(\phi) = e^{\overline{\phi}_i \phi^i / M_{pl}^2} \left\{ \left| \partial_i \mathcal{W} + \frac{\overline{\phi}_i}{M_{pl}} \mathcal{W} \right|^2 - \frac{3}{M_{pl}^2} \left| \mathcal{W} \right|^2 \right\} + \frac{1}{2} g^2 D_a^2, \tag{17.7}$$

where ϕ^i are the complex scalars of the χsf's, $\mathcal{W}$ is the superpotential, D_a are the usual D-terms, and I have assumed a canonical form for the Kahler terms. Note that as $M_{pl} \to 0$, we recover the global supersymmetric result.

The important property of this formula is that *the vacuum energy is no longer positive-definite in supergravity.* This answers question (3) above: supergravity-breaking can still occur with zero cosmological constant. If this hadn't been true, supersymmetry would have been a dead issue.

The order parameter for supergravity breaking is still the non-vanishing of an F or D auxiliary field vev:

$$F_i = \partial_i \mathcal{W} + (\overline{\phi}_i / M_{pl}) \mathcal{W} \neq 0 \qquad \text{or} \qquad D_a \neq 0, \tag{17.8}$$

which is no longer the vacuum energy. In particular, from the scalar potential, we see that though $E_{vac} > 0$ (de Sitter space) always implies broken supergravity, $E_{vac} \leq 0$ (anti de Sitter or Minkowski space) can be consistent with unbroken supergravity. On the other hand, there is no explanation of why $M_{cc} = 0$—an unnatural fine-tuning is still required:

$$\sum_i F_i \overline{F}^i = \frac{3}{M_{pl}^2} \left| \mathcal{W} \right|^2, \tag{17.9}$$

which we will assume from now on. (In other words: we give up on the cosmological constant problem.)

The resulting scale of supergravity breaking is $M_S^2 = \langle F \rangle e^{\phi \cdot \overline{\phi}/M_{pl}^2}$ or $\langle D \rangle / g^2$. For $M_S \neq 0$, the *super-Higgs effect* takes place: the gravitino eats the Goldstino and gets a mass

$$m_{3/2} = \frac{|\mathcal{W}|}{M_{pl}^2} e^{\phi \overline{\phi}/M_{pl}^2} = \frac{M_S^2}{\sqrt{3} M_{pl}}, \tag{17.10}$$

where in the last step we assumed $M_{cc} = 0$. Though the gravitino couples weakly (with gravitational strength), its longitudinal (Goldstino) components will couple with strength $F \sim y/M_S$ where y is a dimensionless coupling. Thus the mass splittings of supermultiplets will be

$$\delta m \sim F M_S^2 \sim y M_S. \tag{17.11}$$

Another feature unique to supergravity is that gaugino condensation can break supergravity. Recall that it does not break global supersymmetry. This has the following interesting consequence. Assume that the gaugino condensate is $\langle \lambda^\alpha \lambda_\alpha \rangle \sim \Lambda^3$, where Λ is the strong-coupling scale of some gauge group. Then the scale of supersymmetry breaking is $M_S \sim \Lambda^{3/2} M_{pl}^{-1/2}$, since it must vanish in the limit $M_{pl} \to \infty$. The associated mass of the gaugino is then $m_\lambda \sim \Lambda^3 M_{pl}^{-2}$. This mechanism gives a way of dynamically breaking supergravity, since if the gauge strong-coupling scale $\Lambda \ll M_{pl}$, then plausibly gravitational physics will play no role in the formation of the condensate. We will see later that we can reliably compute when gaugino condensation occurs in globally supersymmetric theories. Coupling to gravity then implies dynamical supersymmetry breaking.

Another low-energy modification coming from gravity is in the mass sum-rule derived in earlier lectures, which read $\mathrm{STr} M^2 = 0$. In supergravity this is modified to $\mathrm{STr} M^2 = 2(N-1) m_{3/2}^2$, where N is the number of χsf's.

In what follows, since we are interested in constructing an effective supersymmetric theory around the electro-weak scale, we will decouple supergravity. We will have to be careful to keep the low-energy effects of supergravity mentioned above, though. This amounts to taking the formal limit $M_{pl} \to \infty$ while $m_{3/2}$ is held fixed.

17.3. The minimal supersymmetric SM (MSSM)

Let us now build the smallest supersymmetric model which contains the SM, and see if we can answer our previous questions about what breaks the electroweak symmetry and why $M_S \ll M_{pl}$.

First the particle content (which for simplicity we write for one generation only):

superfield	particles of spin			gauge quantum numbers		
name	1	1/2	0	SU(3)	SU(2)	U(1)
V^a	g^a	$\widetilde{g}^a$		**8**	**1**	0
V^i	W^i	$\widetilde{W}^i$		**1**	**3**	0
V	B	$\widetilde{B}$		**1**	**1**	0

Q	u, d	$\widetilde{u}, \widetilde{d}$	$\mathbf{3}$	$\mathbf{2}$	$+1/6$
U	u^c	$\widetilde{u}^c$	$\overline{\mathbf{3}}$	$\mathbf{1}$	$-2/3$
D	d^c	$\widetilde{d}^c$	$\overline{\mathbf{3}}$	$\mathbf{1}$	$+1/3$
L	ν, e	$\widetilde{\nu}, \widetilde{e}$	$\mathbf{1}$	$\mathbf{2}$	$-1/2$
E	e^c	$\widetilde{e}^c$	$\mathbf{1}$	$\mathbf{1}$	$+1$
H_1	$\widetilde{h}_1$	h_1	$\mathbf{1}$	$\mathbf{2}$	$-1/2$
H_2	$\widetilde{h}_2$	h_2	$\mathbf{1}$	$\mathbf{2}$	$+1/2$

Since superfields have common gauge quantum numbers, we are forced to add a superpartner (denoted by the fields with tildes) to each field in the SM. The only exception is the SM Higgs, h_1, which could be the superpartner of the lepton doublet. We will see shortly that this doesn't work. Thus we make the Higgs into its own supermultiplet. This adds a new fermion to the model, the Higgsino $\widetilde{h}_1$, which wrecks anomaly cancellation of $SU(2) \times U(1)$ gauge groups. Thus we are forced to add at least one other Higgs multiplet, H_2, in the complex conjugate representation.

The general action for this model is simple to write down. There will be the canonical Kahler and gauge-kinetic terms with gauge couplings g_1, g_2,and g_3 for the $U(1)$, $SU(2)$,and $SU(3)$ gauge groups respectively, and there will be arbitrary gauge-invariant superpotential terms:

$$\begin{aligned} \mathcal{W} =& y_U QUH_2 + y_D QDH_1 + y_E LEH_1 + \mu H_1 H_2 \\ &+ (QDL + UDD + ELL + EH_1H_1 + LH_2) . \end{aligned} \tag{17.12}$$

Here we have written only the renormalizable terms. The Yukawas in the first line are just like those appearing in the SM: they give masses to the up and down quarks and electron if the Higgs fields $h_{1,2}$ get vevs. We can now see why putting the SM Higgs in the L χsf would not have worked: this would amount to deleting the $H_{1,2}$ fields above, and we would lose the Yukawa giving the up quark a mass.

The terms in the second line above all mediate baryon and lepton-number violating interactions. Since they are renormalizable operators, they will naturally give rise to instantaneous proton decay, for instance. So, we would like to set these terms to zero by some symmetry. Indeed, if we set these terms to zero, the theory has an additional global $U(1)_R$ symmetry, under which the χsf's have charges

$$R(Q, U, D, L, E) = 1, \quad \text{and} \quad R(H_1, H_2) = 0. \tag{17.13}$$

In components this implies that $R(\text{SM particles}) = 0$ and $R(\text{super-partners}) = \pm 1$. (This applies also to the vector multiplets, since the R-charge of a vector superfield must be 1.) Recalling that the superpotential must have R-charge 2, it follows that all the unwanted terms are disallowed, as well as the $\mu H_1 H_2$ term, in a natural way.

This symmetry is problematical, however. First, it disallows a Majorana mass for the gauginos which, in the case of the gluino, turns out to create phenomenological problems.

Second, the absence of the μ-term also is problematical for electroweak breaking. One could avoid these problems by spontaneously breaking this symmetry, but that gives rise to an axion (since the symmetry is actually anomalous). For the axion to be invisible, the scale of the breaking has to be on the order of 10^{10-11} GeV, which creates another hierarchy problem. Finally, there are arguments that global symmetries (spontaneously broken or not) are always explicitly broken by non-perturbative gravitational effects.

These problems can be addressed by assuming that such an R-symmetry is explicitly broken at the Planck scale, but a discrete subgroup is left unbroken. There are a number of possibilities for this discrete subgroup, but the simplest is *R-parity*: a $\mathbb{Z}_2 \subset U(1)_R$ which is simply $(-)^R$ on all states. Invariance under this symmetry implies that all terms in the action must have even R-parity. This re-allows gaugino masses and the μ-term. The issue of this symmetry being explicitly broken by gravitational effects can be circumvented by assuming it is actually a discrete gauge symmetry. In this case this $\mathbb{Z}_2$ represents a redundancy in our description of the theory, and not a symmetry at all. There is a discrete analog of anomaly cancellation (Ibanez and Ross *Phys. Lett.* 260B (1991) 291, and Banks and Dine *Phys. Rev.* D45 (1992) 1424), which is satisfied by R-parity.

This, then, is the MSSM. Let us see if it answers our naturalness problems.

First, consider possible irrelevant operators we can add to the MSSM. The first ones are of the form $QQQL$ and $UDDE$, which give rise to dimension-5 operators mediating proton decay. This would seem to present a problem: in the SM the fist irrelevant operators we could add were dimension-6, and so naturally suppressed by a mass-squared. The lack of observed proton decay then implies that that mass scale should be greater than 10^{15} GeV. With dimension-5 operators, however, there is only one power of mass suppression, and the corresponding limit would be naively greater than 10^{27} GeV, which would present another naturalness problem compared to the Planck scale! However, the above naive estimate is wrong: the contribution to proton decay from dimension-5 operators is loop suppressed, giving two extra factors of Yukawas, plus the required color antisymmetrization implies that one of the Yukawas must connect between generations, giving a Cabbibo angle suppression. These all combine to imply that the natural scale of the dimension-5 operators be greater than only 10^{15} GeV, just as in the SM, and consistent with the Planck mass. The channel for proton decay is different, though, going preferentially to kaons rather than pions, a potential signal of supersymmetry if proton decay is ever observed.

Moving on to the dimensionless couplings, another rare process that might be wrecked by the MSSM are FCNC's. Since the gauge and generation structure of the MSSM is the same as that of the SM, the tree-level absence of FCNC's still holds. However, the one-loop suppression of the SM may be wrecked since the superpartners can now contribute in loops. It turns out that one needs the matrices which diagonalize the quark masses be nearly equal to those diagonalizing the squarks. This is a fine-tuning problem, and can be quite severe for some supergravity models which we will consider in the next lecture.

A more severe problem arises because of the possible relevant operator $\mu H_1 H_2$ in the superpotential. This is a supersymmetric mass term for the Higgs fields. It shows that our original motivation for supersymmetry is not realized: a bare mass term for the Higgs is *not* disallowed by supersymmetry! Thus the basic hierarchy problem remains, now called the "μ problem": why is $\mu \ll M_{pl}$? Note that we cannot appeal to some gravitational physics which sets $\mu = 0$ since for an acceptable electroweak breaking phenomenology, we need $\mu \sim M_{EW}$. One hope for solving this problem is that $\mu = 0$ at the Planck scale, but that the non-perturbative effects which give rise to supersymmetry breaking also generate the μ term. We will discuss a mechanism whereby this may happen in the next lecture.

Finally, we may ask whether the MSSM breaks either or both of supersymmetry and electroweak symmetry? The answer is no, as is easily seen by writing out all the F-terms:

$$\begin{aligned}
F_Q = 0 =& y_U U H_2 + y_D D H_1 \\
F_U = 0 =& Q H_2 \\
F_D = 0 =& Q H_1 \\
F_L = 0 =& E H_1 \\
F_E = 0 =& L H_1 \\
F_{H_1} = 0 =& \mu H_2 + y_D Q D + y_E L E \\
F_{H_2} = 0 =& \mu H_1 + y_U Q U.
\end{aligned} \tag{17.14}$$

There are many D-flat directions. Assuming $SU(3) \times U(1)_{EM}$ is not broken, we set $L = E = Q = U = D = 0$. Then the F-terms imply that $H_1 = H_2 = 0$, so that $SU(2) \times U(1)$ is not broken either, and we have found a supersymmetric ground state.

The upshot of this whole discussion is that all the questions about the viability of the MSSM have been shifted to the question of how supersymmetry is broken.

18. Supersymmetry breaking, hidden sectors, and soft terms

We would like to understand what constraints, if any, experiment and naturalness impose on how supersymmetry can be broken. We will learn that it needs to be done in a round-about way, through a *hidden sector* of the theory which is only weakly coupled to the *visible sector*, the sector containing the fields of the MSSM. I will follow somewhat the presentation in the review by Bagger, `hep-ph/9604232`.

18.1. The need for a hidden sector

Suppose we add extra fields to the MSSM whose dynamics gives rise to supersymmetry breaking at a scale M_S. If these fields have tree-level renormalizable couplings to the fields of the MSSM, then we naturally must have $M_S \sim M_{EW}$ in order to stabilize the hierarchy. From supergravity we see that the gravitino will have a mass $m_{3/2} \sim M_S^2/M_{pl} \sim 10^{-16} M_{EW}$, so we can ignore supergravity effects. Typical mass-splittings of supermultiplets will be $\delta m \sim M_S \sim M_{EW}$, so the superpartners should have masses about 10^{3-4} GeV.

Is this spectrum consistent with the tree-level mass sum rule $\mathrm{STr} M^2 = 0$ that we derived earlier? (Since the gravitino mass is so small, we can ignore its contribution.) Recall from the derivation of this rule, that the tree-level masses of fermions were unchanged by supersymmetry breaking, but the masses of their scalar partners were split above and below the fermion mass. This would seem inconsistent with having all the squark masses ~ 1 TeV. One might imagine adding a sufficient number of massive multiplets to the MSSM with strong-enough tree-level couplings to result in a spectrum such that the sum rule is satisfied and all superpartners are massive. However, Dimopoulos and Georgi, *Nucl. Phys.* B193 (1981) 150, showed that this can not work. They argued that since the squarks are color triplets and we don't want to spontaneously break $SU(3)$ color, their net coupling at tree-level to the massive fields breaking supersymmetry must vanish, which then implies that some squarks must be lighter than the quarks—a phenomenological disaster.

So, we learn that tree-level renormalizable couplings between the supersymmetry-breaking sector of the theory and the MSSM sector of the theory are undesirable. Assuming that no such couplings occur, we are led to models with hidden sectors.

18.2. Soft-breaking terms

We can now parametrize our ignorance about how supersymmetry-breaking occurs in the hidden sector, by encoding its effects in the visible sector in a *spurion field* U, which is just a field whose F-component gets a vev (thus breaking supersymmetry) and enters in the visible sector only through non-renormalizable operators. Such a field is sometimes called a "messenger field". There is no reason for there to be only one such field; if there are many, one speaks of a "messenger sector".

If the messenger sector consists of fields which are simultaneously charged under both visible (*e.g.* SM) gauge groups *and* hidden sector gauge groups (but have no tree-level couplings to the visible sector), we then say we have a "gauge-mediated" model of supersymmetry breaking. Such models have recieved growing attention lately; see M. Dine and A. Nelson, *Phys. Rev.* D48 (1993) 1277, M. Dine, A. Nelson, and Y. Shirman, *Phys. Rev.* D51 (1995) 1362, and M. Dine, A. Nelson, Y. Nir, and Y. Shirman, *Phys. Rev.* D53 (1996) 2658. Another possibility is if there are no such fields, so only gravity couples the visible and hidden sectors. This class of models is called "gravity-mediated" or "supergravity hidden sector" models.

To be more precise, we are assuming that the visible sector is an effective supersymmetric theory containing the MSSM with unbroken $SU(3) \times SU(2) \times U(1) \times \mathbb{Z}_2$ gauge group, valid below some scale M. The hidden sector is some other supersymmetric effective theory valid below the same scale M, but which spontaneously breaks supersymmetry at a scale $M_S < M$. In gravity-mediated models, we will take $M \sim M_{pl}$, typically. In gauge-mediated models it will turn out that we need $M \sim 10^7$ GeV.

We take the spurion U to be a constant χsf whose F-term is of order M_S^2:

$$U \sim \theta^2 M_S^2. \tag{18.1}$$

Since it can couple only through non-renormalzable interactions, by naturalness they must be of the form:

$$\begin{aligned} \delta\mathcal{K} &= \frac{1}{M}\overline{\Phi}\Phi(U+\overline{U}) + \frac{1}{M^2}\overline{\Phi}\Phi\overline{U}U + \ldots \\ \delta\mathcal{W} &= \frac{1}{M}\left(\mathrm{tr}W_\alpha^2 + \mu\Phi^2 + y\Phi^3\right)U + \ldots \end{aligned} \tag{18.2}$$

Here Φ stands for all the χsf's in the visible sector. Expanding in components, such terms give rise to *soft terms*—classically relevant operators in the visible sector fields which explicitly break supersymmetry. Other, less relevant, operators can be added to (18.2), but it is easy to see that they result in operators supressed by factors of M_S/M. For instance, the marginal operators are all down by a factor of $(M_S/M)^2$.

Assuming that the visible sector is just the MSSM, they are terms of the form

$$m_0^2\overline{\phi}\phi, \quad m_{1/2}\widetilde{g}\widetilde{g}, \quad \mu B h_1 h_2, \quad yA\phi^3, \tag{18.3}$$

where ϕ stands for squarks, sleptons, and Higgs, and $\widetilde{g}$ for gauginos. Terms of this type are the only ones allowed by gauge symmetry. Putting in the three generations, this gives rise to 60 or so parameters. From the form of the spurion couplings (*i.e.* naturalness) we find

$$m_0 \sim m_{1/2} \sim B \sim A \sim \frac{M_S^2}{M}. \tag{18.4}$$

Thus, though the hidden sector breaks supersymmetry at the scale M_S, because of its weak (or gravitational) coupling, it only breaks it at the lower scale M_S^2/M in the visible sector. As we discussed before, we want this scale to be of order the electroweak scale, so

$$M_{EW} \sim \frac{M_S^2}{M}. \tag{18.5}$$

Thus, in the gravity-mediated models with $M = M_{pl}$, we find that $M_S \sim 10^{11}$ GeV. In the gauge-mediated models with $M_S \sim 10^5$ GeV, we have $M \sim 10^7$ GeV.

There are a couple of important comments we should make. First, in writing the spurion couplings, we assumed that they were gauge singlets; this is justified in the case of the gravity-mediated models, since if they weren't, they would have broken, say, the electroweak symmetry up near the scale M_S (perhaps suppressed by a couple of factors of the gauge coupling from loops). In the gauge-mediated models, on the other hand, by definition we want the messenger sector to carry gauge charges, and that is why in these models we need $M_S \sim 10^{2-3} M_{EW}$ since they can only differ by loop suppression factors. Since the messenger sector carries gauge charges, the non-renormalizable terms through which it can couple to the visible sector are more restricted, giving rise to many fewer soft-coupling parameters.

The second comment concerns the μ term and the μB term. Recall that naturally $\mu \sim M$, so likewise the soft-term μB is naturally $\sim M_S^2$, and is not suppressed. The μ-problem remains with us, and is just "multiplied" into two μ-problems in hidden-sector models.

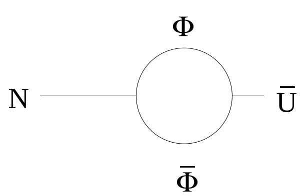
18.3. Stabilizing the hierarchy

Aside from the μ problem, one may ask whether the hierarchy $M_{EW} < M_S$ is stable against quantum effects, given that supersymmetry is broken so we do not have the non-renormalization theorems to protect us. The fact that the two sectors only communicate through non-renormalizable interactions goes a long way towards stabilizing the hierarchy—it turns out (see Bagger's review for more details) that there are only a few possible sources of destabilizing contributions.

The first are tadpoles of gauge-singlet fields in the visible sector. Suppose there is such a gauge-singlet field N. It will generically couple non-renormalizably in the Kahler potential along with the spurion as

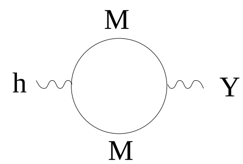
$$\delta\mathcal{K} \sim \frac{1}{M}\overline{\Phi}\Phi(N + \overline{N} + U + \overline{U}). \tag{18.6}$$

The vev of U gives rise to a tadpole for the F-component of N of order M_S^2 from diagrams like:



since U gives M_S^2, the two vertices give $1/M$ each, and the loop is quadratically divergent, so gives M^2.

Recently Dienes, Kolda, and March-Russell, `hep-ph/9610479`, pointed-out another potentially destabilizing situation: if the model has a $U(1)_h$ gauge factor in the hidden sector, and a messenger field M charged under both $U(1)_h$ and $U(1)_Y$ (SM hypercharge), then the mixing:



can destabilize the hierarchy if the D-term of the $U(1)_h$ vector superfield is of order M_S^2.

One has to avoid these situations, giving constraints on the models. There are, of course, simple ways to avoid them. For example, if the visible sector is only the MSSM, then there are no gauge singlets N. If the hidden sector has no $U(1)$ factors, then the gauge-mixing instability cannot arise. In any case, it is believed that once these conditions are met, the hierarchy is stable against radiative corrections. (This still ignores the μ problem, though.)

18.4. Gravity-mediated models

We will now look at the typical properties of gravity-mediated hidden sector models in a little more detail. At the end I will briefly compare these properties to those expected in the gauge-mediated models.

It has been noted that although the gauge couplings do not unify in the SM, in the MSSM with superpartner masses around 1 TeV, they do unify at a scale $M_{GUT} \sim 10^{16}$ GeV. If one assigns fundamental significance to this meeting, then one would like the effective scale M of the visible sector to be greater than or equal to M_{GUT}. This is easily accomodated in the gravity-mediated models, where we can take M as large as the Planck scale.

As we have noted above, μ and μB are not naturally of the electroweak scale. Since they control the Higgs potential, though, they have a strong effect on the scale of electroweak breaking, so they must be of order M_{EW}. In the context of gravity-mediated models, a mechanism has been proposed which solves this μ problem, due to Giudice and Masiero, *Phys. Lett.* 206B (1988) 480. It assumes that $\mu = 0$ in the limit of exact supersymmetry (due, say, to some unknown gravity-related physics or string symmetry) and also that the Kahler terms coupling the spurion to the Higgses is non-minimal:

$$\delta\mathcal{K} \sim \frac{1}{M}\left(H_i\overline{H}^i U + \overline{H}_1\overline{H}_2 U + h.c.\right) + \frac{1}{M^2}H_i\overline{H}^i U\overline{U} + \ldots . \tag{18.7}$$

Plugging in $U \sim \theta^2 M_S^2$ gives rise to the terms

$$\frac{M_S^2}{M}(F_i\overline{h}^i + F_2h_1 + F_1h_2 + h.c.) + \frac{M_S^4}{M^2}h_i\overline{h}^i \sim \left(\frac{M_S^2}{M}\right)^2 (|h_1|^2 + |h_2|^2 + h_1h_2 + h.c.) \tag{18.8}$$

after eliminating the auxiliary F-fields, implying that the effective μ and μB terms are indeed naturally of order the electroweak scale.

To extract more detailed predictions from models, one often assumes relations among the soft-breaking parameters to reduce the size of parameter space. One such assumption is that all the soft-breaking parameters unify at a high scale, say M_{GUT}, so there are only five parameters describing the low-energy visible sector:

$$m_0, m_{1/2}, A, B, \mu \sim M_{EW}. \tag{18.9}$$

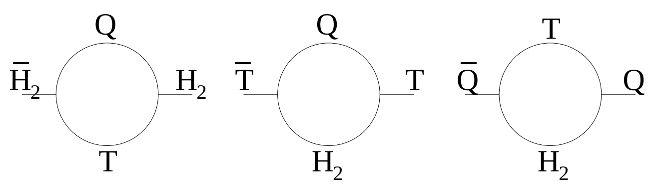
This is often called *universality*, despite which name it has no real theoretical justification. Although these parameters are all the same at M_{GUT}, their RG running will be different, giving rise to potentially widely different values at M_{EW}.

As we noted last lecture, some kind of universality is needed if we are to avoid FCNC's. In gravity-mediated models with $M \sim M_{GUT}$ or M_{pl}, the universality between generations required is very precise, giving rise to a serious naturalness problem.

We still have not explained electroweak breaking. With fairly loose universality assumptions it automatically follows from the RG flow of the couplings. We can see the essential idea by just looking at the third (top) generation which has the biggest Yukawas. The quark doublet superfield Q (including t and b), the singlet χsf T (including t^c), and H_2 are coupled through their Yukawa y_T. The masses of their scalar components approximately satisfy the RG equation

$$\frac{d}{d\log\mu}\begin{pmatrix} m_H^2 \\ m_Q^2 \\ m_T^2 \end{pmatrix} = \frac{y_T^2}{8\pi^2}\begin{pmatrix} 3 & 3 & 3 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} m_H^2 \\ m_Q^2 \\ m_T^2 \end{pmatrix} \tag{18.10}$$

from the loop diagrams



Universality implies that at the scale M the masses are given by

$$\begin{pmatrix} m_H^2 \\ m_Q^2 \\ m_T^2 \end{pmatrix} = m_0^2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{m_0^2}{2}\left[\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}\right], \tag{18.11}$$

where we have decomposed it into eigenvectors of the matrix appearing in the RG equation. The first eigenvector has eigenvalue zero, so hardly runs at all, while the second eigenvector has eigenvalue 6 and so is exponentially suppressed by the RG flow from M to M_{EW}. Thus we find at the scale M_{EW} that $m_H^2 \sim -\frac{1}{2}m_0^2$. This sign for the mass of h_2 is just what we need to drive electroweak breaking. Furthermore, putting in the known value of the W-boson mass, we predict the top quark mass to be $m_t \sim 175$ GeV—the top quark mass is at a fixed point of the RG flow. This encouraging property is a general feature of hidden-sector models.

After electroweak breaking, 3 of the Higgs bosons are eaten by $W^{\pm}$ and Z, and 3 neutral (real) scalars and one charged scalar remains of the Higgs sector. The spin-1/2 Higgsinos mix with the Winos and Binos to give the *neutralinos* (4 neutral Weyl spinors) and the *charginos* (2 charged Dirac spinors). Besides the "direct" masses $\sim |h_i|^2$ propotional to the m_0 soft-parameter, the Higgs sector potential was also determined by the μ and $B\mu$ parameters. It is common practice to trade μ and $B\mu$ for the parameter

$$\tan\beta \equiv \frac{\langle h_2 \rangle}{\langle h_1 \rangle} \tag{18.12}$$

after fitting to the mass of the Z, the Fermi coupling and the fine structure constant. One is typically left with the four parameters m_0 (governing the squark and Higgs masses), $m_{1/2}$ (governing the gaugino masses), $\tan\beta$ (governing the Higgs vevs), and A (governing squark trilinear couplings).

An important effect of R-parity is to make the lightest super-partner (LSP) stable. Cosmological constraints require the LSP to be neutral. In gravity-mediated models the gravitino gets a mass $M_S^2/M_{pl} \sim 10^3$ GeV. Matching to experiment, it is typically found in gravity-mediated models that a neutralino is the LSP over much of parameter space.

18.5. Comparison to gauge-mediated models

Since the effective scale M of the gauge mediated models is around 100 TeV, they can not accomodate the noted super-GUT unification of couplings.

There seems to be no analog of the Guidice-Masiero mechanism to solve the μ-problem in these models (*cf.* Dvali, Guidice, and Pomarol, `hep-ph/960328`). Of course, since the difference between M_S and M_{EW} in these models is only a factor of 100 or so, one might be willing to put up with this degree of unnaturalness.

One the other hand, since the ways supersymmetry breaking can be transmitted to the visible sector is constrained by gauge invariance in these models, it easy easy and natural to get the kind of generation universality that is needed to solve the FCNC problem. Also, the automatic occurrence of electroweak breaking and heavy top also works for the gauge-mediated models. It is claimed that various cosmological problems associated with the light gravitino and axions in gauge-mediated models are avoided if $M_S < 100$ TeV.

The experimental signatures of gauge-mediated models are distinct from those of the gravity-mediated models mainly because the neutralino is no longer the LSP—it is the next-lightest super-partner (NLSP). It typically decays to the gravitino with a long-enough lifetime to lead to observable vertex separation. The gravitino, which is the LSP, has a mass of only a keV or so, and is very weakly coupled.

To sum up, the gravity-mediated models can accomodate GUT unification, are sensitive to UV physics up to a very high scale, have a severe naturalness problem to avoid FCNC's, may avoid the μ problem, and have hidden sectors which are truly hidden. The Gauge-mediated models do not accomodate GUT unification, are sensitive to UV physics only up to 100 TeV or so, have no FCNC problem, have a mild μ naturalness problem, and have hidden sectors which are fairly observable.

For those interested in delving further into the phenomenology of electroweak-scale supersymmetry, I recommend the reviews by H. Nilles, *Phys. Rep.* 110 (1984) 1; H. Haber and G. Kane, *Phys. Rep.* 117 (1985) 75; X. Tata, `hep-ph/9510287`; and M. Peskin, `hep-ph/9604339`.

IV. Nonperturbative Supersymmetry

19. Supersymmetric Yang-Mills Theory

Supersymmetric Yang-Mills theory is a pure gauge theory with (simple) gauge group G. It's Lagrangian is simply

$$\mathcal{L} = \int d^2\theta \, \frac{\tau}{32\pi i} \mathrm{tr} W_\alpha^2 + h.c. \tag{19.1}$$

where, as usual, $\tau = (\vartheta/2\pi) + i(4\pi/g^2)$ is the coupling.

It's classical global symmetry is a $U(1)_R$ with $R(W_\alpha) = 1$, so that the gaugino has R-charge 1:

$$U(1)_R : \ \lambda_\alpha \to e^{i\alpha}\lambda_\alpha. \tag{19.2}$$

This symmetry is anomalous, therfore shifting

$$\vartheta \to \vartheta + \alpha T(adj). \tag{19.3}$$

The anomalous $U(1)_R$ rotations can be used to shift $\vartheta = 0$. Thus the ϑ-angle can have no observable effect in this theory. Since the ϑ-angle is an angle, we see that a $\mathbb{Z}_{T(adj)}$ discrete subgroup of the $U(1)_R$ is unbroken. These models thus have a discrete chiral symmetry.

The theory is asymptotically free, so the gauge coupling runs with scale as

$$\frac{8\pi^2}{g^2(\mu)} = \frac{3}{2} T(adj) \log\left(\frac{\mu}{\Lambda}\right), \tag{19.4}$$

at least at weak coupling (small scales). Here Λ is the strong-coupling scale of the theory. Aside from the discrete choice of the gauge group, G, this model has no free parameters.

What happens in this theory at strong coupling, *i.e.* at scales below Λ? Is supersymmetry broken? Witten (*Nucl. Phys.* B185 (1982) 253) showed that in these theories

$$\mathrm{Tr}(-)^F = \mathrm{rank}(G) + 1. \tag{19.5}$$

Since this is non-zero, this implies that supersymmetry is not broken. Furthermore, there are at least $\mathrm{rank}(G) + 1$ degenerate vacua.

Since

$$\tfrac{1}{2} T(adj) = \mathrm{rank}(G) + 1 \qquad \text{for } Sp(N_c) \text{ and } SU(N_c) \tag{19.6}$$

gauge groups, it is a natural guess that the discrete chiral $\mathbb{Z}_{T(adj)}$ symmetry is spontaneously broken:

$$\mathbb{Z}_{T(adj)} \to \mathbb{Z}_2 \tag{19.7}$$

by gaugino condensation:

$$\langle\lambda\lambda\rangle \sim \Lambda^3 e^{4\pi i n/T(adj)}, \qquad n = 1, \ldots, \tfrac{1}{2}T(adj). \tag{19.8}$$

Novikov, Shifman, Vainshtein, and Zakharov, *Nucl. Phys.* B229 (1983) 407, showed that the one-instanton contribution to the correlator

$$\mathcal{G} = \langle\lambda\lambda(x_1)\,\lambda\lambda(x_2)\cdots\lambda\lambda(x_{T(adj)/2})\rangle \tag{19.9}$$

is non-zero. Recalling that $\lambda\lambda(x)$ is the lowest component of a χsf, and by a supersymmetry Ward identity and cluster decomposition, a correlator of such fields is just the product of their vevs, it follows that $0 \neq \mathcal{G} = \langle\lambda\lambda\rangle^{T(adj)/2}$, confirming the chiral-symmetry-breaking picture.

It is worth making a few comments on the issues encountered in such instanton calculations. One can see these already in the original (non-supersymmetric) instanton solution of A. Belavin, A. Polyakov, A. Schwarz, and Y. Tyupkin, *Phys. Lett.* 59B (1975) 85, and G. 't Hooft, *Phys. Rev.* D14 (1976) 3432 (here for gauge group $SU(2)$):

$$A^a_\mu = \frac{2\overline{\eta}^a_{\mu\nu}(x-x_0)^\nu}{(x-x_0)^2+\rho^2} \tag{19.10}$$

in $\partial^\mu A^a_\mu = 0$ gauge. Here $\overline{\eta}^a_{\mu\nu}$ are parameters determining how the gauge and space-time indices are tied together, x_0^μ are four parameters labelling the position of the "center" of the instanton, and ρ is a real parameter labelling the size of the instanton. To these parameters correspond bosonic zero modes (corresponding to the gauge-invariance, translational-invariance, and scale-invariance, broken by the instanton solution). As discussed in lecture 4, one is supposed to integrate over the bosonic zero modes, giving rise to a term like

$$\int d^4x_0 \int \frac{d\rho}{\rho^5} e^{-8\pi^2/g^2(\rho)}, \tag{19.11}$$

where the ρ^{-5} is to get the dimensions right, and the $e^{-8\pi^2/g^2}$ is the usual one-instanton contribution. The running coupling constant is naturally evaluated at the scale of the instanton. It is natural to interpret $\int d^4x_0$ (after exponentiating the one-instanton contribution in a "dilute instanton gas approximation") as the space-time integration of a term in the effective Lagrangian induced by the instanton. However, recalling that

$$e^{-8\pi^2/g^2(\rho)} \sim (\rho\Lambda)^{3T(adj)/2}, \tag{19.12}$$

we see that the ρ-integration is IR divergent (*i.e.* as $\rho \to \infty$) since $T(adj) \geq 4$ for all gauge groups. Of course, the theory becomes strongly-coupled in the IR below the scale Λ, and one might expect a semi-classical approximation based on the microscopic (UV) description of the theory to break down. Nevertheless, this prevents us from reliably calculating the instanton contribution to the effective action.

Novikov *et. al.* bypassed this difficulty in the supersymmetric case by proposing to calculate the instanton contribution to the correlation function $\mathcal{G}$, instead of the effective Lagrangian, which one might expect to be ill-defined in a strongly-coupled theory anyway. By looking at a correlator of the lowest components of χsf's, supersymmetry ensured that IR divergences would cancel, since by the Ward identity $\mathcal{G}$ has no x-dependence.

The following "accident" shows why one might expect a one-instanton contribution to this correlator. Besides the bosonic zero modes, supersymmetric instantons also have fermionic zero modes, and it is a fact that an instanton in this theory has $T(adj)$ fermionic zero modes. From our earlier discussion of fermionic functional integration, recall that to get a non-zero result in the presence of backgrounds with fermionic zero modes, one must insert that many fermionic operators into the correlation function. Thus $\langle(\lambda\lambda)^{T(adj)/2}\rangle$ can potentially get a non-zero one-instanton contribution. By dimensional analysis (since the dimension of λ is 3/2), this correlator must be proportional to $\Lambda^{3T(adj)/2}$. On the other hand, we have seen that a one-instanton contribution depends on the gauge coupling as $e^{-8\pi/g^2}$, which, by the RG running from the one-loop β-function is proportional to $\Lambda^{3T(adj)/2}$.

Unfortunately, we will not have time in the rest of the course to go into more detail on one-instanton calculations like this one, though we will point out an important case where an instanton contribution to the effective action can be reliably computed.

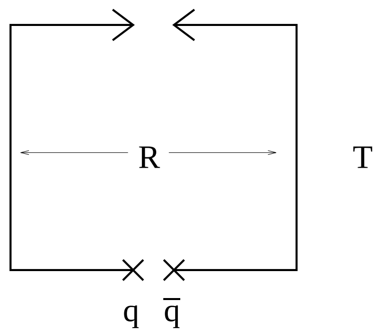
A piece of "lore" about the super-Yang-Mills theories is that they are thought, based on our experience with QCD and with lattice simulations, to have a mass gap and confinement. Confinement is defined to occur when the energy required to separate two test charges grows linearly with their separation:

$$E(R) \sim \sigma R. \tag{19.13}$$

This is often described by the behavior of the expectation value of the *Wilson loop* operator in the limit of large space-time loops:

$$\langle \mathrm{Tr}\mathcal{P}e^{i\oint A}\rangle \sim e^{-\sigma\cdot\mathrm{Area}}. \tag{19.14}$$

The Wilson loop operator tests the response of the theory to the presence of an external source distributed along the loop. If one chooses the loop to be a rectangle of width R and length T (in the time direction), one can interpret the Wilson loop as measuring the action of a process which creates a pair of heavy charges separated by R and holds them there for a time T before annihilating them. The Wilson loop then computes $e^{-TE(R)}$ which for a confining potential then gives the above area law:



On the other hand, a Higgs mechanism would be expected to give a perimeter-type law for large Wilson loops, since the energy for separated charges is expected to fall off exponentially due to screening by the Higgs vev.

It seems hard to gain any new insight into super-Yang-Mills using the non-renormalization theorems, essentially since the theory has no free parameters. It turns out, however, that progress can be made by looking at a more complicated theory: super-QCD, which is just super-Yang-Mills with non-chiral matter. The idea is that super-QCD has adjustable parameters (masses and vevs of the matter χsf's) and so is more amenable to analysis. The super-Yang-Mills theory can be realized as a limit of super-QCD in which all the masses are taken to infinity.

19.1. Digression on the meaning of confinement

Before we turn to super-QCD, I would like to take a moment to explain some subtleties concerning the distinction between confinement and the Higgs mechanism. These subtleties will be important in interpreting our solutions for the vacuum structure of super-QCD. The following discussion follows 't Hooft's 1981 Cargese lectures.

Gauge symmetry is not a symmetry, it is a redundancy in our description of the physics. Evidence of this fact is that we divide our space of states by gauge transformations, considering two states differing by a gauge transformation as physically equivalent. This is different from what we do in the case of global symmetries: two states connected by a global symmetry transformation are inequivalent states, though they have identical physics.

All physical states in a gauge theory are gauge-invariant, by definition. Confinement is sometimes described by saying that only color-singlet (*i.e.* gauge-invariant) combinations of quarks and gluons are observable as asymptotic states. So isn't confinement trivially a consequence of gauge invariance? Furthermore, if the vacuum is always gauge invariant, there can be no such thing as "spontaneous gauge symmmetry breaking"! Is the Higgs mechanism, in which a field gets a gauge-non-invariant vev, in contradiction with gauge invariance?

The answer to both these questions is no. We can see what is wrong with the above naive descriptions of Higgs and confining behavior in gauge theories through a simple example.

Consider the $SU(2)$ gauge theory with a doublet scalar ϕ (the Higgs), a doublet Weyl spinor ψ (the left-handed electron and neutrino), and a singlet Weyl spinor χ (the right-handed electron):

$$\mathcal{L} = \frac{1}{g^2}F_{\mu\nu}^2 + D_\mu\overline{\phi}D^\mu\phi + V(|\phi|) + \overline{\psi}\not{D}\psi + \overline{\chi}\not{\partial}\chi + y\chi(\psi\phi) + h.c. \tag{19.15}$$

If the minimum of V is at $|\phi| = v$, we usually describe the resulting Higgs mechanism by choosing $\langle\phi\rangle = \binom{v}{0}$ and expand about that vacuum as

$$\phi = \begin{pmatrix} v + h_1 \\ h_2 \end{pmatrix}, \qquad \psi = \begin{pmatrix} e_L \\ \nu \end{pmatrix}, \qquad \chi = e_R, \qquad A_\mu^a = (Z_\mu, W_\mu, \overline{W}_\mu). \tag{19.16}$$

This description seems to break the gauge symmetry. However, dividing by gauge transformations, it is indistinguishable from the following "confined" description where all physical particles are gauge singlets:

$$e_L \sim \frac{1}{v}\psi\overline{\phi}, \qquad \nu \sim \frac{1}{v}\psi\phi, \qquad \mathrm{Re}(h_1) \sim \frac{1}{v}\phi\overline{\phi} - v, \qquad Z_\mu \sim \frac{1}{v^2}\overline{\phi}D_\mu\phi, \qquad W_\mu \sim \frac{1}{v^2}\phi D_\mu\phi. \tag{19.17}$$

The reason that we describe the Higgs mechanism in terms of fictitious global quantum numbers (like ν *versus* e_L, *etc.*) is because of our familiarity with global symmetry breaking and the fact that in the weak-coupling limit ($g \to 0$) a theory with a local symmetry looks globally symmetric.

In the proper gauge-invariant description, all the physical states are "mesons" or "baryons" of the scalars bound to other fields. Is there, then, no fundamental distinction between Higgs and confinement?

The intuitive distinction between the two is clear: in a confining phase, the static potential between two charges grows linearly with separation, $V \sim \Lambda^2 r$, so that as long as we are observing the system on energy scales much smaller than Λ, the charges appear confined. In the Higgs phase, on the other hand, the static potential falls off exponentially with separation, $V \sim \Lambda_1 e^{-\Lambda_2 r}$, due to the screening by the scalar condensate. Similarly, in QED (the "Coulomb phase") charges feel a Coulomb potential, $V \sim 1/r$, so charges can be infinitely separated. Thus we can talk about a single electron state $\psi(x)$ even though it is not gauge-invariant: it can be thought of as an electron-positron state with a Wilson line running between them, with the positron sent off to infinity. Thus the gauge-invariant description of an electron is actually $\psi(x)\exp\{i\int_x^\infty A\cdot dx\}$; the Wilson line has observable effects when the topology of space-time is not simply-connected, *e.g.* the Aharanov-Bohm effect.

The question of whether a gauge theory shows one of these behaviors is a dynamical one. A kinematical question that can be addressed is whether these various long distance behaviors correspond to separate phases, or whether one can smoothly deform, say, Higgs to confining behavior. A Higgs vacuum can arise if there is a scalar condensate which can screen massive charges in the gauge group. One can always imagine the possibility in any theory (whether it has fundamental scalars or not) that the strong-coupling dynamics might form such a massless scalar composite, and therefore that a Higgs vacuum might arise.

As long as the scalar is in a faithful representation[26] of the gauge group, it can screen all charges, and there can be no invariant distinction between the Higgs vacuum and a confining vacuum since all the asymptotic states are gauge-singlets. The only way a (non-trivial) representation of a simple group can fail to be faithful is if it does not transform under the center of the gauge group.[27] If the microscopic field content of a theory is such that no (composite) scalars in faithful representations can be formed, then there is an invariant distinction between Higgs and confining phases: in the Higgs phase, all but the discrete central charges are screened,[28] whereas in the confining phase all the asymptotic states will be invariant under the center of the gauge group. QCD is such a theory, where the $\mathbb{Z}_3$ center is tied to the electric charges of the fields, so the distinction between confinement and Higgs phases in this case is whether there are charge-1/3 asymptotic states or not.

20. Classical supersymmetric QCD

In the next five lectures we will study the IR physics of super-QCD; see the review of K. Intriligator and N. Seiberg, `hep-th/9509066`, for another presentation of this material. For the purposes of the rest of this course, super-QCD is defined as an $SU(n_c)$ gauge theory with n_f "quark" χsf's Q^i_a transforming in the $\mathbf{n_c}$ representation of the gauge group, and n_f "anti-quarks" $\widetilde{Q}^a_i$. Here $a, b, c = 1, \dots n_c$ are color indices, and $i, j, k = 1, \dots, n_f$ are flavor indices. The most general renormalizable Lagrangian for super-QCD, and the definition of our theories is

$$\mathcal{L} = \int d^4\theta \left(\overline{Q}_i e^V Q^i + \overline{\widetilde{Q}}^i e^{-V} \widetilde{Q}_i \right) + \int d^2\theta \left(\frac{\tau}{32\pi i} \mathrm{tr} W^2 + \mathcal{W}(Q, \widetilde{Q}) \right) + h.c. \qquad (20.1)$$

[26] A faithful representation is one for which for every $g \in G$ with $g \neq 1$, then $R(g) \neq 1$, where $R(g)$ is the representation matrix.

[27] The center of a group is the subgroup consisting of the set of all elements which commute with all elements of the group. For example, the center of $SU(n_c)$ is $\mathbb{Z}_{n_c}$ realized as overall phase rotations by the n_c-th roots of unity.

[28] Since the center is a discrete subgroup of the gauge group, and discrete gauge groups have no long-range fields, there can exist asymptotic states charged under the center.

Here the notation in the Kahler terms is meant to imply that e^V is in the fundamental representation, while e^{-V} is in the anti-fundamental. We will start our analysis by setting the superpotential to zero:

$$\mathcal{W} = 0. \tag{20.2}$$

Later we will examine the effect of superpotential terms.

Our aim is to extract the IR physics of this theory. By IR physics, I will mean the physics at arbitrarily low energy scales—*i.e.* the vacuum structure and the massless particles. The reason for this is that this is all that is captured by the Kahler and superpotential terms that we have been keeping in our effective actions. The non-renormalization theorems that we have proved only apply to these terms. Any finite-energy effects will presumably also get contributions from higher-derivative terms in the effective action. We start by analyzing the classical vacuum structure of the theory.

20.1. Symmetries and vacuum equations

With zero superpotential, the theory has a non-Abelian global symmetry $U(n_f)$ rotating the quarks, and similarly for the antiquarks; in addition there is a $U(1)_R$ symmetry. We can choose a basis of the $U(1)$ factors to have the following action on the scalar components of the χsf's (we will denote the scalar components by the same symbol as the whole superfield):

	$SU(n_c)$		$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_A$	$U(1)_{R'}$
Q^i_a	$\mathbf{n_c}$	|	$\mathbf{n_f}$	$\mathbf{1}$	1	1	1
$\widetilde{Q}^a_i$	$\overline{\mathbf{n_c}}$	|	$\mathbf{1}$	$\overline{\mathbf{n_f}}$	-1	1	1

Here $U(1)_B$ is the "baryon number", and the axial $U(1)$ and the $U(1)_{R'}$ as defined are anomalous. Later we will find the non-anomalous R-symmetry which is a linear combination of these two $U(1)$'s.

Since there is no superpotential, the scalar potential is just the D-terms $V(Q, \widetilde{Q}) \sim \mathrm{tr}D^2$. The condition for a supersymmetric vacuum is then

$$\begin{aligned} 0 = D &= \sum_A \left(\overline{Q}^a_i (T^A_{n_c})^b_a Q^i_b + \overline{\widetilde{Q}}^i_a (T^A_{\overline{n_c}})^a_b \widetilde{Q}^b_i \right) \\ &= \sum_A (T^A)^b_a \left(\overline{Q}^a_i Q^i_b - \widetilde{Q}^a_i \overline{\widetilde{Q}}^i_b \right), \end{aligned} \tag{20.3}$$

where in the first line $T^A_{n_c}$ are the generators in the fundamental and $T^A_{\overline{n_c}}$ are those of the anti-fundamental, and in the second line we have used that fact that the two are related by $(T^A_{\overline{n_c}})^a_b = -(T^A_{n_c})^a_b \equiv (T^A)^a_b$. Since the T^A are a basis of hermitian traceless matrices, the D-term conditions can be written as

$$\overline{Q}^a_i Q^i_b - \widetilde{Q}^a_i \overline{\widetilde{Q}}^i_b = \frac{1}{n_c} (\overline{Q}^c_i Q^i_c - \widetilde{Q}^c_i \overline{\widetilde{Q}}^i_c) \delta^a_b. \tag{20.4}$$

These conditions are not hard to solve, using the fact that by appropriate color and flavor rotations we can put an arbitrary Q^i_a into the diagonal form

$$Q = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_{n_f} \\ & & \\ & & \end{pmatrix}, \qquad a_i \in \mathbb{R}^+, \qquad (n_f < n_c). \tag{20.5}$$

Here the columns are labelled by the flavor index and the rows by the color index, and we have shown the result for $n_f < n_c$. Now $\widetilde{Q}$ can simultaneously be put in upper-diagonal form by the remaining $SU(n_f)$ symmetry rotations. Plugging into the D-term equations then gives

$$\widetilde{Q}^T = Q. \tag{20.6}$$

Of course, we could have solved the D-term conditions in a gauge-invariant way by using our result from lecture 12 that the D-flat directions are parametrized by the algebraically independent set of holomorphic gauge-invariant monomials in the fields. Such a basis (for $n_f < n_c$) is

$$M^i_j = \widetilde{Q}^a_j Q^i_a, \tag{20.7}$$

giving ${n_f}^2$ massless χsf's whose vev's parametrize the moduli space of vacua. (We will discuss shortly how we know these are a basis of gauge-invariant states.)

As a check on these results, we can count that the two answers imply the same dimension of our moduli space. From the solution (20.5) and (20.6) we see that at a generic point in the moduli space the gauge symmetry is spontaneously broken from $SU(n_c)$ to $SU(n_c - n_f)$, implying that $({n_c}^2 - 1) - [(n_c - n_f)^2 - 1] = 2n_f n_c - {n_f}^2$ gauge bosons get a mass. But, by the Higgs mechanism, each such massive gauge boson eats a χsf, implying that of the original $2n_f n_c$ massless χsf's, only $2n_f n_c - (2n_f n_c - {n_f}^2) = {n_f}^2$ survive, matching the sounting we found from the gauge-invariant solution.

Thus we see that the basic physics occuring here classically is just the Higgs mechanism: the squark vevs generically break $SU(n_c) \to SU(n_c - n_f)$. Of course, for non-generic values of the squark vevs, the unbroken gauge symmetry can be enhanced, corresponding to points where $\mathrm{rank}(M) < n_f$, or, equivalently, where $\det(M) = 0$.

We can also compute the classical Kahler metric on the moduli space. The Kahler form is $\mathcal{K} = \overline{Q}^a_i Q^i_a + \overline{\widetilde{Q}}{}^i_a \widetilde{Q}^a_i$. The D-term equations imply

$$\overline{Q}^a_i Q^i_b = \widetilde{Q}^a_i \overline{\widetilde{Q}}{}^i_b \qquad (n_f < n_c) \tag{20.8}$$

since the trace terms automatically vanish for $n_f < n_c$. Squaring this equation gives $(M\overline{M})^i_j = (\overline{\widetilde{Q}}\widetilde{Q})^i_k(\overline{\widetilde{Q}}\widetilde{Q})^k_j$ which implies that $\overline{\widetilde{Q}}\widetilde{Q} = (M\overline{M})^{1/2}$, and so the Kahler potential is

$$\mathcal{K} = 2\mathrm{Tr}(\overline{M}M)^{1/2}. \tag{20.9}$$

This implies the Kahler metric is singular whenever M is not invertible, corresponding to points of enhanced gauge symmetry.

One case that bears special mention is when $n_f = n_c - 1$. Then generically the gauge symmetry is completely broken (there is no $SU(1)$ group). In this case we might expect the IR physics to be under control even quantumly, since there are no AF gauge groups left.

20.3. Vacua for $n_f \geq n_c$.

In this case, diagonalizing Q and $\widetilde{Q}$ subject to the D-term equations gives the solutions (up to gauge and flavor rotations)

$$\begin{aligned} Q &= \begin{pmatrix} a_1 & & & \\ & \ddots & & \\ & & a_{n_c} & \end{pmatrix}, \qquad a_i \in \mathbb{R}^+, \\ \widetilde{Q} &= \begin{pmatrix} \widetilde{a}_1 & & & \\ & \ddots & & \\ & & \widetilde{a}_{n_c} & \end{pmatrix}, \qquad (n_f \geq n_c), \end{aligned} \tag{20.10}$$

where

$$|\widetilde{a}_i|^2 = a_i^2 + \rho, \qquad \rho \in \mathbb{R}, \tag{20.11}$$

for some constant ρ independent of i. Thus, generically, the gauge symmetry is completely broken on the moduli space.

The gauge-invariant description is in terms of the following set of holomorphic invariants:

$$\begin{aligned} M^i_j &= Q^i\widetilde{Q}_j && \text{“mesons”} \\ B^{i_1\ldots i_{n_c}} &= Q^{i_1}_{a_1}\cdots Q^{i_{n_c}}_{a_{n_c}}\epsilon^{a_1\ldots a_{n_c}} && \text{“baryons”} \\ \widetilde{B}_{i_1\ldots i_{n_c}} &= \widetilde{Q}^{a_1}_{i_1}\cdots \widetilde{Q}^{a_{n_c}}_{i_{n_c}}\epsilon_{a_1\ldots a_{n_c}} && \text{“anti-baryons”}. \end{aligned} \tag{20.12}$$

The baryons and anti-baryons vanished identically for $n_f < n_c$ because of the antisymmetrization of the squarks. It is clear that M, B, and $\widetilde{B}$ form a basis of gauge-invariants, since they are all that can be made from the $SU(n_c)$ invariant tensors δ^a_b and $\epsilon^{a_1\cdots a_{n_c}}$.

However, they are an overcomplete basis. One way of seeing this is to note that there are $2\binom{n_f}{n_c} + {n_f}^2$ meson and baryon fields, but by the Higgs mechanism there are only

$2n_f n_c - ({n_c}^2{-}1)$ massless χsf's. Thus there must be relations among the baryons and mesons.

These constraints are easy to find. Since the product of two color epsilon tensors is the antisymmetrized sum of Kronecker deltas, it follows that

$$B^{i_1\ldots i_{n_c}}\widetilde{B}_{j_1\ldots j_{n_c}} = M^{[i_1}_{j_1}\cdots M^{i_{n_c}]}_{j_{n_c}}, \tag{20.13}$$

where the square brackets denote antisymmetrization. Also, since any expression antisymmetrized on $n_c{+}1$ color indices must vanish, it follows that any product of M's, B's, and $\widetilde{B}$'s antisymmetrized on $n_c{+}1$ upper or lower flavor indices must vanish.

A convenient notation for writing these constraints is to denote the contraction of an upper with a lower flavor index by a "$\cdot$", and the contraction of all flavor indices with the totally antisymmetric tensor on n_f indices by a "$*$". For example

$$(*B)_{i_{n_c+1}\ldots i_{n_f}} = \epsilon_{i_1\ldots i_{n_f}} B^{i_1\ldots i_{n_c}}. \tag{20.14}$$

Then (20.13) can be rewritten in this notation as

$$(*B)\widetilde{B} = *(M^{n_c}). \tag{20.15}$$

The constraint coming from antisymmetrizing the $n_c{+}1$ flavor indices in the product of one M with a baryon is written

$$M\cdot *B = M\cdot *\widetilde{B} = 0. \tag{20.16}$$

As long as both B and $\widetilde{B}$ are non-zero, an induction argument shows that the above two constraints imply all the other D-term constraints: A constraint with, say, k B's and an arbitrary number of M's antisymmetrized on $n_c{+}1$ upper indices can be replaced by a constraint with $k{-}1$ B fields by (20.15). Repeating this process reduces all constraints to (20.15) plus the single constraint with no B fields $*(M^{n_c+1}) = 0$. But this latter constraint is implied by (20.15) and (20.16): $0 = \widetilde{B}(M\cdot *B) = M\cdot *(M^{n_c}) = *(M^{n_c+1})$. When only one of B or $\widetilde{B}$ vanishes, the above arguments fail, and extra constraints would seem to be needed beyond (20.15) and (20.16). I do not know of a simple set of constraints in this case.

It will be useful to write out the first simplest cases explicitly. The first case is when $n_f = n_c$. Then we expect ${n_f}^2{+}1$ massless χsf's, but we have ${n_f}^2{+}2$ invariants. The single constraint is just (20.15), which can be written more simply in terms of $*B$ and $*\widetilde{B}$ which have no flavor indices, as

$$y \equiv \det M - (*B)(*\widetilde{B}) = 0. \tag{20.17}$$

Since

$$dy = (\det M)(M^{-1})^i_j dM^j_i - (*B)d(*\widetilde{B}) - (*\widetilde{B})d(*B), \tag{20.18}$$

singularities of the moduli space $y = dy = 0$ occur at

$$B = \widetilde{B} = *(M^{n_c-1}) = 0. \tag{20.19}$$

This last constraint implies $\mathrm{rank}(M) < n_c - 1$, so (by referring back to our explicit solutions for Q and $\widetilde{Q}$) we see that there will be at least an unbroken $SU(2)$ gauge group.

In the case $n_f = n_c+1$, there are ${n_f}^2$ massless χsf's, and ${n_f}^2+2n_f$ invariants. (20.15) and (20.16) give ${n_f}^2+2n_f$ constraints. Therefore, the constraints are not independent in this case. Nevertheless, there is not a smaller set of holomorphic constraints.

21. Quantum super-QCD: $n_f < n_c$

First, we look at the RG running of the gauge coupling. By our previous formulas, and recalling that for $SU(n_c)$

$$b_0 = \frac{3}{2}T(adj) - \frac{1}{2}2n_f T(n_c) = 3n_c - n_f, \tag{21.1}$$

the one-loop running is

$$\frac{8\pi^2}{g^2(\mu)} = (3n_c - n_f)\log\left(\frac{\mu}{\Lambda}\right). \tag{21.2}$$

Thus, the theory is AF for $n_f < 3n_c$ and IR-free for $n_f \geq 3n_c$. (For $n_f = 3n_c$, though there is no one-loop running, typically the two-loop running due to the anomalous dimensions of the quarks makes the theory IR-free.)

Next, we look for possible anomalies in the classical global symmetries. As we discussed before, an anomalous symmetry can be thought of as one under which the gauge strong-coupling scale Λ^{b_0} transforms. We therefore have:

	$SU(n_c)$	|	$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_A$	$U(1)_{R'}$	$U(1)_R$
Q^i_a	$\mathbf{n_c}$	|	$\mathbf{n_f}$	$\mathbf{1}$	1	1	1	$1-\frac{n_c}{n_f}$
$\widetilde{Q}^a_i$	$\overline{\mathbf{n_c}}$	|	$\mathbf{1}$	$\overline{\mathbf{n_f}}$	-1	1	1	$1-\frac{n_c}{n_f}$
$\Lambda^{3n_c-n_f}$	$\mathbf{1}$	|	$\mathbf{1}$	$\mathbf{1}$	0	$2n_f$	$2n_c$	0

where in the last column we have defined a new non-anomalous R-charge as

$$R = R' - \frac{n_c}{n_f}A. \tag{21.3}$$

(We need a coefficient of 1 in front of R' to keep the vector superfield having R-charge 1.) We can immediately write down how the mesons and baryons transform under these symmetries:

	$SU(n_c)$	|	$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_A$	$U(1)_{R'}$	$U(1)_R$
M	$\mathbf{1}$	|	$\mathbf{n_f}$	$\overline{\mathbf{n_f}}$	0	2	2	$2-2\frac{n_c}{n_f}$
B	$\mathbf{1}$	|	$\binom{n_f}{n_c}$	$\mathbf{1}$	n_c	n_c	n_c	$n_c-\frac{n_c{}^2}{n_f}$
$\widetilde{B}$	$\mathbf{1}$	|	$\mathbf{1}$	$\binom{n_f}{n_c}$	$-n_c$	n_c	n_c	$n_c-\frac{n_c{}^2}{n_f}$

Now, as we go far out along the classical flat direction, generically the gauge group will be broken as $SU(n_c) \to SU(n_c - n_f)$ for $n_f < n_c - 1$, and completely broken otherwise. In the cases where the gauge group is completely broken, we would then expect the IR physics to just be a nlσm for the light fields M, B and $\widetilde{B}$ (subject to constraints). In the cases where there is an unbroken gauge group, because the massless meson degrees of freedom along the flat directions are gauge neutral, all their couplings to the gauge fields are non-renormalizable. Thus in the IR we expect the theory to decouple into a nlσm for the light M fields and an $SU(n_c - n_f)$ super-YM theory. We expect the latter factor to have a gap, though, leaving only the nlσm.

Assuming, then, that the meson and baryon χsf's are the correct IR degrees of freedom (at least at generic points on moduli space and with "large enough" vevs), the next question to ask is: are the classical flat directions lifted quantumly? We can write all the possible terms that could appear in the superpotential consistent with the symmetries and the selection rule for the anomalous $U(1)_A$ symmetry:

$$\mathcal{W} \sim \left[\frac{\Lambda^{3n_c - n_f}}{\det M}\right]^{\frac{1}{n_c - n_f}} . \tag{21.4}$$

This follows since $\det M$ or one of its powers is the only $SU(n_f) \times SU(n_f)$ invariant, and the above powers are fixed by the $U(1)_A$ and $U(1)_R$ symmetries. In the weak-coupling limit, $\Lambda \to 0$, we expect this contribution to vanish, and so we see it can only appear for $n_f < n_c$. We thus learn that for $n_f \geq n_c$, no superpotential can be generated, and therefore that the classical flat directions are not lifted. We will explore the $n_f \geq n_c$ theories in later lectures.

If the superpotential term (21.4) *were* generated dynamically for $n_f < n_c$, what would its implications be? Since, qualitatively, $\det M \sim M^{n_f}$, the scalar potential derived from the superpotential goes as

$$V = |\mathcal{W}'|^2 \sim |M|^{-2n_c/(n_c - n_f)}. \tag{21.5}$$

This potential has not minimuum and slopes to zero as $M \to \infty$. This implies that the resulting theory has no ground state!

There are a number of ways this conclusion might be avoided. First, might quantum effects make the $M = \infty$ "point" in moduli space actually be at finite distance in field space? No, because for large M, we expect perturbation theory to be good, so the classical

Kahler potential $\mathcal{K} \sim |M|$ should be valid, implying that $M \to \infty$ really is infinitely far away in field space. Could quantum corrections to the Kahler potential for small M lead to new minima of the scalar potential? The reason this is a possibility is that the scalar potential is actually $V \sim g^{i\bar{i}}\partial_i \mathcal{W} \partial_{\bar{i}} \overline{\mathcal{W}}$, where $g^{i\bar{i}}$ is the inverse Kahler metric. As long as this metric is not degenerate, then the only zeroes of the potential are when $\partial_i \mathcal{W} = 0$, which we saw occurs only at $M = \infty$. Modifications to the Kahler metric could create *local* minima in the scalar potential at finite values of M, but these would always be metastable, since there would be lower-energy states for large enough M. A final possibility is that quantum corrections actually make the Kahler metric singular at finite M. This is a breakdown of unitarity, implying that other massless fields would need to be added to our IR description. I do not see how to rule out such a possibility, however it seems unlikely that new massless degrees of freedom would enter and yet there would be no sign of them in the superpotential (recall that when we integrated-out massless fields, we typically found singularities in the superpotential).

21.1. $n_f = n_c - 1$

How can we tell whether this superpotential really is generated dynamically? In the case $n_f = n_c - 1$, the gauge symmetry is completely broken, and instanton techniques can then reliably compute terms in the effective action. (This is because the IR divergence we mentioned before in the the instanton calculation in the super-YM case is cutoff by the scalar meson vev M.) Furthermore, in this case the superpotential goes as Λ^{b_0}, which is just what we expect from a one-instanton effect. Such a one-instanton calculation has indeed been carried out in the $n_c = 2$, $n_f = 1$ case, finding a non-zero result; for a summary and the latest status of this computation, see D. Finnell and P. Pouliot, `hep-th/9503115`. This result not only implies that the superpotential term is generated in the $SU(2)$ theory with one flavor, but also in all super-QCD theories with $n_f = n_c - 1$.

To see this, assume the $SU(n_c)$ with $n_c - 1$ flavors generates the superpotential

$$\mathcal{W} = c \cdot \frac{\Lambda^{2n_c+1}}{\det M}. \tag{21.6}$$

We will determine c by looking at a convenient flat direction, namely

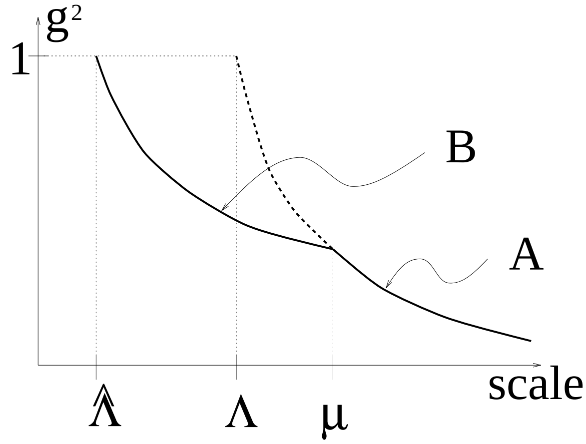
$$\langle Q \rangle = \langle \widetilde{Q}^T \rangle = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_{n_f} \\ 0 & \cdots & 0 \end{pmatrix}, \qquad \text{with} \quad a_1, \ldots, a_{n_f-1} \sim \mu \gg a_{n_f} \gg \Lambda. \tag{21.7}$$

Along this direction, the large vevs break $SU(n_c)$ with $n_f = n_c - 1$ massless flavors down to $SU(2)$ with one massless flavor. In that case the superpotential is

$$\mathcal{W}_{eff} = \hat{\Lambda}^5 / a_{n_f}^2 \tag{21.8}$$

where $\hat{\Lambda}$ is the strong-coupling scale of the $SU(2)$ and the coefficient $c = 1$ by the instanton calculation. By RG matching, the $SU(2)$ scale is given by[29]

$$\left(\hat{\Lambda}/\mu\right)^5 = (\Lambda/\mu)^{2n_c+1}\,. \tag{21.9}$$



Here the "A" curve is the running of the coupling for the $SU(n_c)$ theory with $n_f = n_c{-}1$, and the "B" curve is for the $SU(2)$ theory with $n_f = 1$. This implies that $\hat{\Lambda}^5 = \Lambda^{2n_c+1}\mu^{4-2n_c} = \Lambda^{2n_c+1}\mu^{2-2n_f} = \Lambda^{2n_c+1}/(a_1^2 \cdots a_{n_f-1}^2)$. Thus, comparing to (21.6), we see that $c = 1$.

21.2. $n_f \leq n_c - 1$: effects of tree-level masses

We can extend this result to other numbers of flavors by considering the effect of a tree-level superpotential giving masses to the squarks. In the end, by matching to the Witten-index result in the pure super-YM theory in the infinite-mass limit, we will give a separate argument for the appearance of the dynamical superpotential.

Consider adding to our $n_f = n_c{-}1$ theory a tree-level superpotential term

$$\mathcal{W}_t = m^i_j M^j_i \quad (= \mathrm{Tr}(mM)\ ). \tag{21.10}$$

We can consistently assign charges to the mass-matrix m as follows:

	$SU(n_c)$	$\vert$	$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_A$	$U(1)_{R'}$	$U(1)_R$
m	$\mathbf{1}$	$\vert$	$\overline{\mathbf{n_f}}$	$\mathbf{n_f}$	0	-2	0	$\frac{2n_c}{n_c-1}$
M	$\mathbf{1}$	$\vert$	$\mathbf{n_f}$	$\overline{\mathbf{n_f}}$	0	2	2	$\frac{-2n_c}{n_c-1}$

[29] Actually, this is just a one-loop matching. There can be a threshold factor, due to two-loop effects which enters as a possible multiplicative factor between the two sides. This factor is scheme-dependent (since it can be absorbed in a redefinition of what we mean by the strong-coupling scales). There exists a scheme, however, in which the threshold factors are always 1: the "$\overline{\mathrm{DR}}$-scheme".

Λ^{2n_c+1}	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{1}$	0	$2n_c{-}2$	$2n_c$	0

These then give selection rules implying

$$\mathcal{W}_{eff} = \mathrm{Tr}(mM) \cdot f\left(\frac{Mm}{\mathrm{Tr}(mM)}, \frac{\Lambda^{2n_c+1}}{(\det M)\mathrm{Tr}(mM)}\right). \tag{21.11}$$

When $m = 0$ and $\Lambda = 0$ we should recover the results

$$\begin{aligned} \mathcal{W}_{eff}(m{=}0) &= \frac{\Lambda^{2n_c+1}}{\det M} \\ \mathcal{W}_{eff}(\Lambda{=}0) &= \mathrm{Tr}(mM). \end{aligned} \tag{21.12}$$

Taking the limits $m \to 0$ and $\Lambda \to 0$ in (21.11) in various ways, you can show by holomorphy that

$$\mathcal{W}_{eff} = \mathrm{Tr}(mM) + \frac{\Lambda^{2n_c+1}}{\det M} \qquad (n_f = n_c{-}1). \tag{21.13}$$

The F-term equation for M is then

$$0 = \frac{\partial \mathcal{W}_{eff}}{\partial M^i_j} = m^j_i - (M^{-1})^j_i \frac{\Lambda^{2n_c+1}}{\det M}. \tag{21.14}$$

This implies that

$$\det M = \Lambda^{(2n_c+1)(n_c-1)/n_c} (\det m)^{-1/n_c}, \tag{21.15}$$

and plugging back into (21.14) gives a supersymmetric vacuum at

$$\langle M^i_j \rangle = (m^{-1})^i_j (\det m)^{1/n_c} \Lambda^{(2n_c+1)/n_c}. \tag{21.16}$$

So, as long as we turn on *any* non-degenerate masses for the quarks, we find n_c discrete supersymmetric vacua because of the n_c-th root in (21.16). This is precisely what we expected physically, since after giving masses to the quarks the low energy theory should be pure super-YM, which has n_c vacua according to the Witten index.

Instead of turning on a non-degenerate mass-matrix, we can turn on a degenerate one so as to integrate-out only some of the flavors:

$$m = \begin{pmatrix} 0 & 0 \\ 0 & \widehat{m} \end{pmatrix} \tag{21.17}$$

where the upper left-hand block is $n_f \times n_f$ for some $n_f < n_c{-}1$. Then, in the same block decomposition,

$$M = \begin{pmatrix} \widehat{M} & X \\ Y & Z \end{pmatrix} \tag{21.18}$$

where the fields in the X, Y, and Z blocks all get masses, and so can be integrated-out. Letting $i,j=1,\ldots,n_f$, and $I,J=1,\ldots,n_c{-}1$, the equations of motion for the X and Y blocks are $0=\partial\mathcal{W}_{eff}/\partial M^I_j=\partial\mathcal{W}_{eff}/\partial M^i_J$, giving $0=(M^{-1})^I_j=(M^{-1})^i_J$, which imply

$$M^{-1}=\begin{pmatrix}\widehat{M}^{-1} & 0\\ 0 & Z^{-1}\end{pmatrix}\quad\Rightarrow\quad M=\begin{pmatrix}\widehat{M} & 0\\ 0 & Z\end{pmatrix} \tag{21.19}$$

so that the X and Y blocks vanish. The equation of motion for the Z block is

$$0=\frac{\partial\mathcal{W}_{eff}}{\partial M^I_J}=\widehat{m}^J_I-(Z^{-1})^J_I\frac{\Lambda^{2n_c+1}}{\det\widehat{M}\det Z}, \tag{21.20}$$

implying that

$$\begin{aligned}\det Z&=\frac{\Lambda^{2n_c+1}}{\det\widehat{M}}\left(\frac{\det\widehat{M}}{\Lambda^{2n_c+1}\det\widehat{m}}\right)^{1/(n_c-n_f)},\\ \langle Z^I_J\rangle&=(\widehat{m}^{-1})^I_J\left(\frac{\Lambda^{2n_c+1}\det\widehat{m}}{\det\widehat{M}}\right)^{1/(n_c-n_f)},\end{aligned} \tag{21.21}$$

by a similar calculation as in (21.15) and (21.16). Plugging into $\mathcal{W}_{eff}$ (21.13) gives

$$\begin{aligned}\mathcal{W}_{eff}&=\mathrm{Tr}(\widehat{m}Z)+\frac{\Lambda^{2n_c+1}}{\det\widehat{M}\det Z}\\ &=(n_c-n_f)\left(\frac{\widehat{\Lambda}^{3n_c-n_f}}{\det\widehat{M}}\right)^{1/(n_c-n_f)},\end{aligned} \tag{21.22}$$

where we have defined

$$\widehat{\Lambda}^{3n_c-n_f}\equiv(\det\widehat{m})\Lambda^{2n_c+1}. \tag{21.23}$$

By RG matching, we recognize this as the strong-coupling scale of the $SU(n_c)$ theory with n_f flavors. Dropping the hats, (21.22) implies that the superpotential term is dynamically generated for the theories with $n_f<n_c{-}1$ with coefficient $n_c{-}n_f$.

There are a few interesting points to note about this superpotential. A one-instanton contribution goes as $\Lambda^{b_0}=\Lambda^{3n_c-n_f}$, but (21.22) goes as $\Lambda^{b_0/(n_c-n_f)}$. The interpretation of this is not clear: does it mean that there are semi-classical field configurations with fractional instanton number which compute this effect? In any case, the usual instanton contribution to the effective action is not well-defined in this case due to the IR divergences from the unbroken $SU(n_c{-}n_f)$ gauge group.

A second interesting point is that these fractional powers imply the superpotential is multivalued as a function of $\langle M\rangle$. We can understand the meaning of this by considering the limit where $\det M\gg\Lambda^{n_f}$, so the theory is broken at a large scale down to $SU(n_c{-}n_f)$, classically. Since this occurs at weak coupling, we take as light degrees of freedom the M^i_j

meson χsf's and the $SU(n_c{-}n_f)$ vsf W_α. From our non-renormalization theorem of lecture 15, the effective action for the W_α fields will be

$$\mathcal{L}_{SU(n_c-n_f)} = \int d^2\theta \frac{-\hat{b}_0}{64\pi^2} \log\left(\frac{\widehat{\Lambda}}{\mu}\right) \mathrm{tr} W_\alpha^2 + h.c. \tag{21.24}$$

where $\widehat{\Lambda}$ is the scale of the $SU(n_c{-}n_f)$ super-YM theory and $\hat{b}_0 = 3(n_c{-}n_f)$ is it beta-function. By the usual RG matching

$$\widehat{\Lambda}^{3(n_c-n_f)} \det M = \Lambda^{3n_c-n_f} \tag{21.25}$$

where Λ is the scale of the original $SU(n_c)$ theory with n_f flavors. We argued earlier that the two sectors corresponding to the $SU(n_c{-}n_f)$ super-YM theory and the M nlσm decouple in the IR. However, the two sectors are coupled by irrelevant terms through the above dependence of $\widehat{\Lambda}$ on M. In particular, from (21.24) and (21.25) we have

$$\mathcal{L}_{SU(n_c-n_f)} = \int d^2\theta \frac{-1}{64\pi^2} \log\left(\frac{\Lambda^{3n_c-n_f}}{\det M}\right) \mathrm{tr} W_\alpha^2 + h.c. \supset \frac{1}{64\pi^2} \mathrm{Tr}(F_M M^{-1}) \lambda_\alpha \lambda^\alpha + h.c., \tag{21.26}$$

where in the second line we have expanded in components: M stands for the lowest component of the M χsf, and F_M is its F-component. On the other hand, the dynamically-generated superpotential $\int d^2\theta\, \mathcal{W}_{eff}$, (21.22), from the M-sector of the theory gives rise to the terms

$$\mathcal{L}_M \supset -\mathrm{Tr}(F_M M^{-1}) \left(\frac{\Lambda^{3n_c-n_f}}{\det M}\right)^{1/(n_c-n_f)} + h.c. \tag{21.27}$$

Solving the F_M-term equations of motion from $\mathcal{L}_M + \mathcal{L}_{SU(n_c-n_f)}$ then gives

$$\langle \lambda\lambda \rangle = 64\pi^2 \left(\frac{\Lambda^{3n_c-n_f}}{\det M}\right)^{1/(n_c-n_f)} = 64\pi^2 \hat{\Lambda}^3, \tag{21.28}$$

confirming the expected gaugino condensation in the pure super-YM theory. Thus the $n_c{-}n_f$ branches in the superpotential (21.22) correspond to the $n_c{-}n_f$ vacua of the $SU(n_c{-}n_f)$ super-YM theory.

Just as we did for the $n_f = n_c{-}1$ theory, we can add in tree-level masses m^i_j. The usual argument using holomorphy, symmetry, and weak-coupling limits implies

$$\begin{aligned} \mathcal{W}_{eff} &= \mathrm{Tr}(mM) + (n_c - n_f) \left(\frac{\Lambda^{3n_c-n_f}}{\det M}\right)^{1/(n_c-n_f)} \\ \langle M^i_j \rangle &= (m^{-1})^i_j \left(\Lambda^{3n_c-n_f} \det m\right)^{1/n_c}. \end{aligned} \tag{21.29}$$

This result was first obtained by A. Davis, M. Dine, and N. Seiberg, *Phys. Lett.* 125B (1983) 487.

21.3. Integrating out and in

The technique of adding masses and integrating-out massive degrees of freedom can be generalized, and in many cases is a useful tool for determining exact superpotentials;[30]

[30] K. Intriligator, R. Leigh, and N.Seiberg, `hep-th/9403198`; K. Intriligator, `hep-th/9407106`.

I will present the basic idea in a somewhat simplified form explained in more detail in K. Intriligator and N. Seiberg, `hep-th/9509066`, section 2.3.

Consider a gauge theory with scale Λ whose D-flat directions are parametrized by a set of gauge-invariant composite χsf's $\mathcal{O}^i$. Then, the dynamics may generate an effective superpotential

$$\mathcal{W}_{dyn} = f(\mathcal{O}^i, \Lambda^{b_0}). \tag{21.30}$$

We could probe this theory by adding tree-level couplings

$$\mathcal{W}_{tree} = \sum_i J_i \mathcal{O}^i \tag{21.31}$$

to the theory, and then use holomorphy, symmetries and weak-coupling limits to constrain the resulting effective superpotential, as we have done above. However, there are many cases in which this can be done more simply:

Think of the couplings J_i as sources for each light degree of freedom. *Assuming the dynamics is trivial (gaussian) in the IR* (so there are no IR divergences to keep the 1PI effective action from existing), we can compute the resulting effective superpotential as a 1PI effective superpotential, $\langle\mathcal{W}\rangle$, by the usual Legendre transform[31]

$$\langle\mathcal{W}\rangle(J_i, \Lambda^{b_0}) \equiv \mathcal{W}_{dyn}(\mathcal{O}^i, \Lambda^{b_0}) + \sum_i J_i \mathcal{O}^i, \tag{21.32}$$

where we replace $\mathcal{O}^i$ on the right-hand side by inverting $J_i = -\partial\mathcal{W}_{dyn}/\partial\mathcal{O}^i$, so that $\langle\mathcal{O}^i\rangle = \partial\langle\mathcal{W}\rangle/\partial J_i$. In this case the effective superpotential is automatically linear in the sources:

$$\mathcal{W}_{eff} = \mathcal{W}_{dyn}(\mathcal{O}^i, \Lambda^{b_0}) + \sum_i J_i \mathcal{O}^i \tag{21.33}$$

and the Legendre transform just corresponds to integrating-out the χsf's coupled to the sources.

This can be extended to the gaugino condensate χsf

$$S \equiv -\frac{1}{64\pi^2}\mathrm{tr}(W^\alpha W_\alpha) \tag{21.34}$$

as well, by treating $\log\Lambda^{b_0}$ as its source:

$$\mathcal{W}_{eff} = \mathcal{W}_{dyn} + \sum_i J_i \mathcal{O}^i + \log\Lambda^{b_0}\, S. \tag{21.35}$$

[31] One may wonder why we can apply the 1PI effective *action* technology to the *superpotential*. This follows simply from the fact that we add sources to χsf's in the action as $\mathcal{L} = \cdots + \int d^2\theta\, J_i \mathcal{O}^i$. So, to compute $\langle\mathcal{O}^i\rangle$ we must differentiate with respect to the F-component of the source χsf: $\langle\mathcal{O}^i\rangle = (\partial/\partial F_{J_i})\int d^2\theta\,\langle\mathcal{W}\rangle = (\partial/\partial F_{J_i})[\sum_j(\partial\langle\mathcal{W}\rangle/\partial J_j)F_{J_j}] = \partial\langle\mathcal{W}\rangle/\partial J_j$.

Since the Legendre transform is invertible, we can reverse this procedure and "integrate-in" fields as well. As an example, consider the pure $SU(n_c)$ super-YM theory, where we have

$$\langle S\rangle = \left(\Lambda^{3n_c}\right)^{1/n_c} = \frac{\partial\langle\mathcal{W}\rangle(\Lambda)}{\partial(\log\Lambda^{3n_c})}. \tag{21.36}$$

Solving for $\langle\mathcal{W}\rangle$ gives $\langle\mathcal{W}\rangle = n_c\Lambda^3$, and taking the (inverse) Legendre transform with respect to the source $\log\Lambda^{3n_c}$ then gives

$$\mathcal{W}_{dyn}(S) = \langle\mathcal{W}\rangle - \log\Lambda^{3n_c}\cdot S = n_c S(1-\log S). \tag{21.37}$$

Thus

$$\mathcal{W}_{eff} = \mathcal{W}_{dyn} + \log\Lambda^{3n_c}\cdot S = S\left[\log\left(\frac{\Lambda^{3n_c}}{S^{n_c}}\right) + n_c\right], \tag{21.38}$$

a result first obtained by G. Veneziano and S. Yankielowicz, *Phys. Lett.* 113B (1982) 321, and T. Taylor, G. Veneziano, and S. Yankielowicz, *Nucl. Phys.* B218 (1993) 493. However, the meaning of this effective action is not clear, since it implies the χsf S is always massive.

22. Quantum super-QCD: $n_f \geq n_c$

We now move up in the number of flavors to the $n_f = n_c$ and $n_f = n_c+1$ cases. These were first solved by N. Seiberg in `hep-th/9402044`.

In the last lecture we found for $n_f < n_c$ that

$$\langle M^i_j\rangle = (m^{-1})^i_j(\Lambda^{3n_c-n_f}\det m)^{1/n_c}, \tag{22.1}$$

for an arbitrary mass matrix m. It is not hard to see from symmetries and holomorphy that this expression is the only one allowed, even for $n_f \geq n_c$, though this does not fix its coefficient. But if we consider a theory with $n_f > n_c$ and take masses such that

$$m_1,\ldots,m_{n_c-1},\Lambda \ll m_{n_c},\ldots,m_{n_f}, \tag{22.2}$$

and integrate-out the heavy masses, we arrive at an effective $SU(n_c)$ theory with n_c-1 light flavors with a strong-coupling scale

$$\hat\Lambda^{3n_c-(n_c-1)} = m_{n_c}\cdots m_{n_c}\Lambda^{3n_c-n_f}, \tag{22.3}$$

by the usual RG matching. Plugging into (22.1) then implies that (22.1) must also hold for all n_c and n_f.

Now, consider taking the limit in (22.1) as $m\to 0$. For $n_f < n_c$ this limit always implied $M\to\infty$, and so there was no vacuum. But for $n_f \geq n_c$, we can take the limit in such a way that M remains finite. By taking $m\to 0$ in different ways, we can "map out" the space of vacua of the $n_f \geq n_c$ theories. The fact that flat directions survive in these theories accords with the fact that no superpotential is dynamically generated.

22.1. $n_f = n_c$

In this case, recall that the classical moduli space was parameterized by the "meson" and "baryon" composite χsf's

$$\begin{aligned} M^i_j &= Q^i \widetilde{Q}_j, \\ *B &= Q^{i_1}_{a_1} \cdots Q^{i_{n_c}}_{a_{n_c}} \epsilon^{a_1 \ldots a_{n_c}} \epsilon_{i_1 \ldots i_{n_c}} \\ *\widetilde{B} &= \widetilde{Q}^{a_1}_{i_1} \cdots \widetilde{Q}^{a_{n_c}}_{i_{n_c}} \epsilon_{a_1 \ldots a_{n_c}} \epsilon^{i_1 \ldots i_{n_c}}, \end{aligned} \tag{22.4}$$

which satisfy the constraint

$$\det M - (*B)(*\widetilde{B}) = 0. \tag{22.5}$$

Now, turning on meson masses, by (22.1) gives

$$M = m^{-1} (\det m)^{1/n_c} \Lambda^2, \qquad \Rightarrow \qquad \det M = \Lambda^{2n_c}. \tag{22.6}$$

Since this last formula is independent of m, it will be true in the $m \to 0$ limit. Also, note that when $\det m \neq 0$, that the baryon expectation values must vanish:

$$*B = *\widetilde{B} = 0 \quad \text{if } \det m \neq 0, \tag{22.7}$$

since the vacuum must transform trivially under $U(1)_B$ because all the fields carrying baryon number are integrated-out if $\det m \neq 0$. Taking the limit $m \to 0$, we conclude that $*B = *\widetilde{B} = 0$.

These conclusions, (22.6) and (22.7), are not consistent with the classical constraint (22.5). Therefore, the *classical constraints are modified quantumly*, even though no superpotential is dynamically generated.

To see what the quantum-modified constraints are, we need to do a little more work, since so far we have only probed the vacua by adding a source for M. Symmetries, holomorphy, the fact that $\det M = \Lambda^{2n_c}$ when $*B = *\widetilde{B} = 0$, and from demanding that in the weak-coupling limit $\Lambda \to 0$ the quantum constraint reduce to the classical one, the general form of the quantum constraint must be

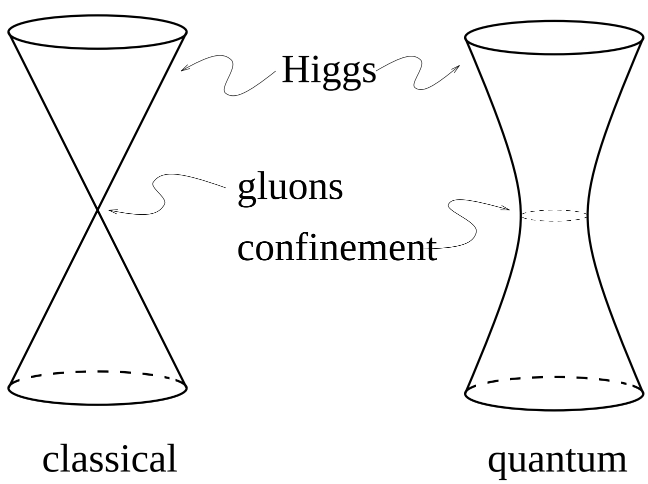
$$\det M - (*B)(*\widetilde{B}) = \Lambda^{2n_c} \left(1 + \sum_{\alpha,\beta>0} c_{\alpha,\beta} \frac{(\Lambda^{2n_c})^\alpha (*B*\widetilde{B})^\beta}{(\det M)^{\alpha+\beta}} \right). \tag{22.8}$$

Classically we have vacua with arbitrary values of $\langle *B \rangle$ and $\langle *\widetilde{B} \rangle$, so by going far-enough out on the classical moduli space where the physics is the Higgs mechanism taking place at arbitrarily weak coupling, we are assured that there will be vacua of the full quantum theory with non-zero baryon vevs with associated meson vevs which statisfy the classical constraint arbitrarily well. But, fixing $*B*\widetilde{B}$ at some large (compared to Λ^{2n_c}) constant value, we see that in addition to the asymptotically classical solution with $\det M \sim *B*\widetilde{B}$, any non-zero $c_{\alpha,\beta}$ gives rise to additional solutions going as $\det M \sim (*B*\widetilde{B})^{(\beta-1)/(\beta+\alpha)}$. Such solutions extend out to the perturbative regime of large meson and baryon vevs; since no vacua like this are seen in perturbation theory, we must have all the $c_{\alpha,\beta} = 0$, giving the quantum constraint:

$$0 = y = \det M - (*B)(*\widetilde{B}) - \Lambda^{2n_c}. \tag{22.9}$$

What is the physics of these vacua? First, note that $dy = 0$ at $*B = *\widetilde{B} = \det M = 0$, which does not lie on the constraint surface $y = 0$. Thus we do not expect any enhanced gauge symmetries on this moduli space. The $*B$, $*\widetilde{B}$ and M vevs at typical points on the moduli space spontaneously break all the global symmetries. There are special submanifolds where the global symmetry can be enhanced. For example, at the point $M^i_j = \Lambda^2\delta^i_j$, $*B = *\widetilde{B} = 0$, the global $SU(n_f) \times SU(n_f) \times U(1)_B \times U(1)_R$ symmetry is only broken to $SU(n_f)_{diag} \times U(1)_B \times U(1)_R$, and the light degrees of freedom are the $*B$ and $*\widetilde{B}$ baryons, as well as the Golstone bosons of the diagonal breaking of the flavor symmetry. This, then, is a supersymmetric version of a vacuum with chiral symmetry-breaking, and massless pions and baryons. Another enhanced symmetry point is $M = 0$ and $*B = *\widetilde{B} = i\Lambda^{n_c}$, where only the $U(1)_B$ of the global symmetry is broken. There is no chiral symmetry-breaking, and the light fields are the mesons M, as well as a $B\widetilde{B}$ composite (the Goldstone boson of the baryon number).

The difference between the classical and quantum moduli spaces can be summarized by the following cartoon:



Classically, the physics is the Higgs mechanism, and at the singularity at the origin, the gauge symmetry is unbroken so there are massless quarks and gluons. In the quantum theory, on the other hand, there is no vacuum with massless gluons, it being replaced by the circle of theories at the neck of the hyperboloid which have chiral symmetry breaking. This is the expected physics of a confining vacuum. We see that in this theory there is no phase transition separating a "Higgs phase" from a "confining phase". This is in accord with the fact that we have squarks in the fundamental representation which can screen any sources in a Wilson loop.

One may wonder how one can generate a superpotential by integrating-out some quarks from this theory, when it doesn't have a superpotential to start with. The point, however, is that the *fluctuations* (and not just the vevs) of the meson and baryon fields are constrained by (22.9). One way of seeing this is to note that even after turning on meson masses—which should enable us to probe possible vacua off the constraint surface if they exist—the meson vev still satisfies the constraint (22.6). One can not just naively integrate-out the meson fields without taking into account the constraint which couples the meson and baryon fluctuations. To impose this constraint in the action, we add a Lagrange-multiplier χsf, A, to enforce the constraint. The Lagrange-multiplier can be thought of as a χsf with no kinetic (Kahler) terms, and therefore no fluctuations. The superpotential (with mass term for the squarks) becomes

$$\mathcal{W} = \mathrm{Tr}(mM) + A\left[\det M - (*B)(*\widetilde{B}) - \Lambda^{2n_c}\right]. \tag{22.10}$$

Taking the mass matrix to be

$$m = \begin{pmatrix} 0 & 0 \\ 0 & \hat{m} \end{pmatrix}, \qquad M = \begin{pmatrix} \hat{M} & X \\ Y & Z \end{pmatrix} \tag{22.11}$$

where the upper left-hand block is $n_f \times n_f$, one can then use the F-term equations of motion for the M, $*B$, $*\widetilde{B}$, and A fields to show that $*B = *\widetilde{B} = X = Y = 0$, and solve for the others (just as in the last lecture) giving

$$\mathcal{W}_{eff} = (n_c - n_f)\left(\frac{\hat{\Lambda}^{3n_c-n_f}}{\det \hat{M}}\right)^{1/(n_c-n_f)} \qquad \text{with} \quad \hat{\Lambda}^{3n_c-n_f} \equiv \Lambda^{2n_c}\det \hat{m}. \tag{22.12}$$

22.2. $n_f = n_c{+}1$

Recall from the discussion in lecture 20 that a basis of composite χsf's in this case is M^i_j, $*B_i$, and $*\widetilde{B}^i$, satisfying the classical constraints:

$$\begin{aligned} 0 &= (M^{-1})^i_j \det M - (*B)_j(*\widetilde{B})^i, \\ 0 &= (*B)_i M^i_j = M^i_j(*\widetilde{B})^j. \end{aligned} \tag{22.13}$$

We probe the quantum moduli space by turning on quark masses m^i_j. As before, when $\det m \neq 0$, $*B_i = *\widetilde{B}^i = 0$ and

$$M^i_j = (m^{-1})^i_j(\Lambda^{2n_c-1}\det m)^{1/n_c}, \tag{22.14}$$

which imply, in particular, that

$$(*B \cdot M)_i = (M \cdot *\widetilde{B})^i = 0, \quad \text{and} \quad (M^{-1})^i_j \det M = \Lambda^{2n_c-1} m^i_j. \tag{22.15}$$

Unlike the $n_f = n_c$ case where turning on masses kept $\langle M \rangle$ on the constraint surface ($\det M = \Lambda^{2n_c}$), turning on masses in this case allows M to take any value off the constraint surface. This implies we will not be able to implement the quantum constraints with Lagrange mulitpliers in the superpotential—they will have to arise as equations of motion. Also unlike the $n_f = n_c$ case, in the limit $m \to 0$, $\langle M \rangle$ is on the *classical* constraint surface. The possible corrections to the classical constraints consistent with this data involve positive powers of $*B \cdot M \cdot *\widetilde{B}/\det M$ by the symmetries; assuming that turning on baryon sources can probe vacua with arbitrary baryon vevs, all these terms must vanish by taking appropriate $M \to 0$ limits. In this way we see that the classical constraints (22.13) remain valid in the full quantum theory.

To see how these can arise as equations motion, we write down the most general dynamical superpotential (consistent with the symmetries):

$$\mathcal{W}_{eff} = \frac{1}{\Lambda^{2n_c-1}} \left[\alpha(*B \cdot M \cdot *\widetilde{B}) + \beta \det M + \det M \, f\left(\det M / *B{\cdot}M{\cdot}*\widetilde{B}\right)\right]. \tag{22.16}$$

We normally would not allow such a term since it does not vanish in the weak-coupling limit $\Lambda \to 0$; however, in this case we will see that it reproduces the classical constraints, so it can be kept. The arbitrary function f must vanish in order to have a smooth $M \to 0$ limit; alternatively, only $f = 0$ will reproduce the classical constraints. The F-term equations of motion are:

$$\begin{aligned} \frac{\partial \mathcal{W}_{eff}}{\partial M} \Rightarrow &\quad 0 = \alpha(*B)_j(*\widetilde{B})^i + \beta(M^{-1})^i_j \det M = 0, \\ \frac{\partial \mathcal{W}_{eff}}{\partial(*B, *\widetilde{B})} \Rightarrow &\quad 0 = (*B \cdot M)_i = (M \cdot *\widetilde{B})^i. \end{aligned} \tag{22.17}$$

These are the classical constraints if $\alpha = -\beta$.

Adding in a single mass to integrate-out one flavor matches to the $n_f = n_c$ case when $\alpha = 1$. The algebra is as follows. In the superpotential

$$\mathcal{W} = \frac{1}{\Lambda^{2n_c-1}} \left(*B \cdot M \cdot *\widetilde{B} - \det M\right) + \mathrm{Tr}(mM), \tag{22.18}$$

let

$$m = \begin{pmatrix} 0 & 0 \\ 0 & \widehat{m} \end{pmatrix}, \quad M = \begin{pmatrix} \widehat{M} & X \\ Y & Z \end{pmatrix}, \quad *B = \begin{pmatrix} W \\ \widehat{*B} \end{pmatrix}, \quad *\widetilde{B} = \begin{pmatrix} \widetilde{W} \\ \widehat{*\widetilde{B}} \end{pmatrix}, \tag{22.19}$$

where the upper left-hand blocks are $n_c \times n_c$. We integrate-out X, Y, W, and $\widetilde{W}$ using the equations of motion, leaving us with the equations $\det M = Z \det \widehat{M}$, $*B \cdot M \cdot *\widetilde{B} = Z \widehat{*B} \widehat{*\widetilde{B}}$, and $\mathrm{Tr} mM = Z\Lambda^{2n_c-1}\widehat{m}$, which, when plugged back into $\mathcal{W}$ give

$$\mathcal{W} = \frac{Z}{\Lambda^{2n_c-1}} \left(\widehat{*B} \widehat{*\widetilde{B}} - \det \widehat{M} + \widehat{\Lambda}^{2n_c} \right), \quad \text{where } \widehat{\Lambda}^{2n_c} \equiv \widehat{m}\Lambda^{2n_c-1}. \tag{22.20}$$

Dropping the hats, and identifying Z/Λ^{2n_c-1} with the Lagrange multiplier field A, we indeed recover the $n_f = n_c$ case. Note also, that with $\widehat{\Lambda}$ fixed, $\widehat{m} \to \infty$ implies $\Lambda \to 0$. Thus the kinetic terms for A go as $(\partial Z)^2 \sim \Lambda^{4n_c-2}(\partial A)^2 \to 0$, showing that A is indeed a Lagrange multiplier, and not a fluctuating field.

We have just shown that the classical and quantum moduli spaces of the $n_f = n_c+1$ theories are the same. In particular, unlike the $n_f = n_c$ case, the singular point at $M = *B = *\widetilde{B} = 0$ remains in the moduli space. Classically this was the point with unbroken $SU(n_c)$ gauge group and massless quarks and gluons. Quantumly, it seems to be a point with massless meson and baryon composites, confinement (no gluons), and no chiral symmetry breaking. On the other hand, we have seen before that singularities in the holomorphic coordinate description of the moduli space are often (though not necessarily) associated with new light degrees of freedom that were not included in our original effective action. How can we tell if that is what actually occurs in this case?

While there is no proof that there cannot be new light degrees of freedom at the origin, the following argument suggests that there are not. We can test the consistency of assuming that only the composite meson and baryon fields are the light degrees of freedom at the origin through the 't Hooft anomaly matching conditions. At the origin, the full global symmetry group is unbroken, under which the microscopic quark χsf's and the macroscopic meson and baryon χsf's have charges:

	$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_R$
Q	$\mathbf{n_f}$	1	$+1$	$\frac{1}{n_f}$
$\widetilde{Q}$	1	$\overline{\mathbf{n_f}}$	-1	$\frac{1}{n_f}$
M	$\mathbf{n_f}$	$\overline{\mathbf{n_f}}$	0	$\frac{2}{n_f}$
B	$\overline{\mathbf{n_f}}$	$\mathbf{1}$	n_f-1	$1-\frac{1}{n_f}$
$\widetilde{B}$	$\mathbf{1}$	$\mathbf{n_f}$	$1-n_f$	$1-\frac{1}{n_f}$

In terms of the microscopic and macroscopic fermion fields this gives

	$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_R$
λ	$\mathbf{n_c{}^2-1}$	1	0	1
ψ_Q	$\mathbf{n_f}$	1	$+1$	$\frac{1}{n_f}-1$
$\psi_{\widetilde{Q}}$	1	$\overline{\mathbf{n_f}}$	-1	$\frac{1}{n_f}-1$
ψ_M	$\mathbf{n_f}$	$\overline{\mathbf{n_f}}$	0	$\frac{2}{n_f}-1$
ψ_B	$\overline{\mathbf{n_f}}$	$\mathbf{1}$	n_f-1	$-\frac{1}{n_f}$

$\psi_{\widetilde{B}}$	$\mathbf{1}$	$\mathbf{n_f}$	$1-n_f$	$-\frac{1}{n_f}$

One can then check that all the anomalies match. For example:

$$\begin{aligned}\mathrm{Tr}R &= 2n_f n_c \cdot \left(\frac{1}{n_f}-1\right) + \left({n_c}^2-1\right)\cdot 1 = -{n_f}^2 + 2n_f - 2 \qquad \text{(micro)} \\ &= {n_f}^2 \cdot \left(\frac{2}{n_f}-1\right) + 2n_f \cdot \left(-\frac{1}{n_f}\right) = -{n_f}^2 + 2n_f - 2 \qquad \text{(macro)},\end{aligned} \tag{22.21}$$

and

$$\begin{aligned}\mathrm{Tr}R^3 &= 2n_f n_c \cdot \left(\frac{1}{n_f}-1\right)^3 + \left({n_c}^2-1\right)\cdot 1^3 = -{n_f}^2 + 6n_f - 12 + \frac{8}{n_f} - \frac{2}{{n_f}^2} \qquad \text{(micro)} \\ &= {n_f}^2 \cdot \left(\frac{2}{n_f}-1\right)^3 + 2n_f \cdot \left(-\frac{1}{n_f}\right)^3 = -{n_f}^2 + 6n_f - 12 + \frac{8}{n_f} - \frac{2}{{n_f}^2} \qquad \text{(macro)},\end{aligned} \tag{22.22}$$

etc. . Because we compute the anomaly by counting states only if their kinetic terms are non-singular, the matching of the anomalies can be taken as evidence for the Kahler potential being smooth at the origin.

In summary, for the $n_f = n_c{+}1$ theories, we have seen that the quantum and classical moduli spaces are the same. The classical moduli space was described by constraints which arose "trivially" from the definition of the composite χsf's in terms of the microscopic quark χsf's; while those same constraints in the quantum theory arose as equations of motion.

22.3. $n_f \geq n_c{+}2$

Just as in the $n_f = n_c{+}1$ case, we can probe the quantum moduli space by turning on masses m^i_j and using

$$\langle M^i_j \rangle = (m^{-1})^i_j (\Lambda^{3n_c - n_f} \det m)^{1/n_c}. \tag{22.23}$$

For $m \neq 0$ all values of M can be obtained, and by taking $m \to 0$ in various ways we again find that we can arrive at any point on the classical moduli space with vanishing baryon vevs. This then implies, using the symmetries and weak-coupling limits, that the quantum moduli space must coincide with the classical one.

This immmediately raises the question of the interpretation of the singularity at $M = B = \widetilde{B} = 0$. Unlike the $n_f = n_c$ case, the superpotential (which gives rise to the constraints as equations of motion, and goes as $(\det M)^{1/(n_f - n_c)}$) in this case is singular at the origin, which is a sign that there are extra light degrees of freedom there. Also, the 't Hooft anomaly-matching conditions are not satisfied if one assumes that only M, B, and $\widetilde{B}$, are light there.

We will spend the next two lectures understanding what happens at these points.

23. Phases of N=1 gauge theories, and superconformal invariance

In asking the question of what is the IR physics at the origin of moduli space in the $n_f \geq n_c$+1 super-QCD theories, we finally have to face up to the question of what is the range of possible IR behaviors of a QFT. We will refer to these different kinds of IR behavior as possible "phases" of the theory.

By definition, the IR behavior of a theory is its behavior at arbitrarily low energy scales, and so, in particular, below any mass or strong-coupling scales in the theory. Therefore the IR behavior must be described by a scale-invariant theory. What are some examples of scale-invariant QFT's?

The simplest is the trivial, or Gaussian, theory. This QFT has no propagating fields. It is the IR behavior of a theory which has only massive degrees of freedom, since on scales below all masses, all the modes of these fields can be integrated-out. Other examples are the completely Higgsed and confining phases of gauge theories.

The next simplest are theories of free, massless fields. We can write down such theories for arbitrary numbers of fields of arbitrary spin. Examples are free theories of massless scalars, spin-$\frac{1}{2}$ particles, and pure $U(1)$ gauge theory. One can describe the behavior of such theories in terms of their correlation functions: the only connected ones are the two-point functions, whose spatial dependence is given by the canonical (free) scaling dimension of the fields. For example,

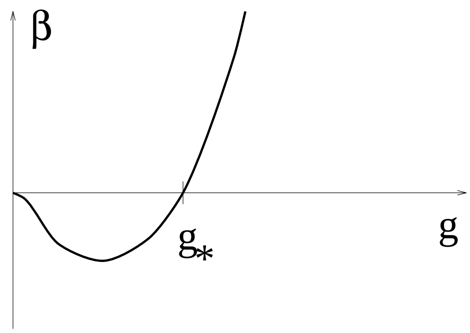
$$\begin{aligned}\langle\phi(x)\phi(y)\rangle &= \frac{1}{|x-y|^2},\\ \langle\psi^\alpha(x)\psi^\beta(y)\rangle &= \frac{\epsilon^{\alpha\beta}}{|x-y|^3},\\ \langle F^{\mu\nu}(x)F^{\rho\sigma}(y)\rangle &= \frac{g^{\mu\rho}g^{\nu\sigma}-g^{\mu\sigma}g^{\nu\rho}}{|x-y|^4}.\end{aligned} \tag{23.1}$$

Finally, there are theories of massless fields with scale-invariant interactions. Actually, examples of such theories in four dimensions are few and far between. For example, adding Yukawa and $\lambda\phi^4$ couplings to theories of fermions and scalars leads only to free IR theories, since those classically marginal couplings turn out to be irrelevant (IR-free) in perturbation theory. Another example is massless QED, where a massless electron is coupled to a $U(1)$ gauge field: this again is an IR-free theory with the electron charge ($U(1)$ coupling) running to zero at long distances. Non-Abelian gauge theories can be AF, meaning that they get strongly-coupled in the IR, but this does not tell us what their IR behavior is, just that it is difficult to deduce it from their microscopic description. If one adds enough matter, say $n_f \geq 3n_c$ in our $SU(n_c)$ super-QCD examples, then the β-function changes sign and the theory is IR-free again.

One might think then that one could tune the field content of a non-Abelian gauge theory to make the β-function exactly zero, thus acheiving exact scale-invariance. Taking $n_f = 3n_c$ in super-QCD does not work, though, since that only cancels the one-loop β-function; the two-loop contributions make the theory IR free. However, this observation suggests how to show that interacting scale-invariant theories do exist. (The following observation was made by T. Banks and A. Zaks, *Nucl. Phys.* B196 (1982) 189.) For $n_f < 3n_c$ a two-loop computation (see lecture 16) gives the β-function

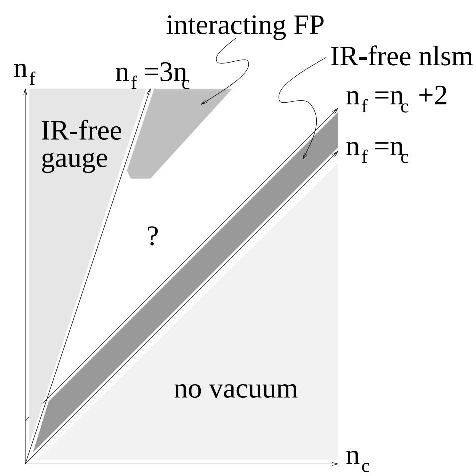
$$\beta(g) = -\frac{g^3}{16\pi^2}(3n_c - n_f) + \frac{g^5}{128\pi^4}\left(2n_c n_f - 3{n_c}^2 - \frac{n_f}{n_c}\right) + \mathcal{O}(g^7), \tag{23.2}$$

which gives an IR fixed point ($\beta = 0$) at a coupling $g_* \sim \sqrt{3n_c - n_f}$.



We can trust the existence of this fixed point as long as the coupling g_* is small, so that the higher-order terms can be safely neglected. Define $n_f = n_c(3-\epsilon)$, and take the limit $n_c \to \infty$ (so that $\epsilon \sim 1/n_c$). Then $n_c g_*^2 \sim 4\pi^2\epsilon/3$. Recalling that $n_c g_*^2$ is the expansion parameter for large n_c, we see that there does exist a limit in which the fixed point is a weak coupling.

This shows the existence of non-trivial scale-invariant four-dimensional QFTs. (Note that this argument had nothing to do with supersymmetry—it could have just as well been done in non-supersymmetric QCD). It's implications for super-QCD can be summarized in a kind of "phase diagram":



In this and the next lecture we will answer the question posed by the question mark.

How can we tell the difference between a free scale-invariant theory and an interacting one? At the level of 2-point functions, we would expect the scaling dimensions of interacting fields to be different from their canonical values. With an extra assumption, we can prove this; we can also derive restrictions on the allowed ranges of scaling dimensions following from unitarity.

Before we do that, however, I should mention another common way of probing the IR behavior of QFT's which is different from the above method of looking at the correlators of fields in the low-energy effective theory. Another way to probe these theories is by their response to classical sources—massive particles interacting with the massless fields. This probe can give different information than the correlator method. For example, massive QED and massless QED are both IR-free field theories. But the static potential between sources of charge e separated by a distance r in massive QED is

$$V(r) = \frac{e^2}{r}, \tag{23.3}$$

while that of massless QED is

$$V(r) = \frac{e^2(r)}{r} \sim \frac{1}{r\log(\Lambda r)}, \tag{23.4}$$

due to the running of the coupling constant. As another example, the Abelian Higgs model, and pure $SU(3)$ Yang-Mills are both thought to have mass gaps, and are therefore trivial in the IR. But in hte Abelian Higgs model

$$V(r) \sim \Lambda e^{-\Lambda r} \tag{23.5}$$

due to screening, while in the YM theory, the potential is thought to go as

$$V(r) \sim \Lambda^2 r \tag{23.6}$$

due to confinement. In all of these potentials, Λ is some scale that appears in the full theory. This shows that even though the static potential probes a long-distance aspect of the behavior of these theories, it does not just probe the scale-invariant (arbitrarily low energy) properties of the theory.

23.1. Conformal invariance

A scale-invariant theory (one with $\beta = 0$) is one which is invariant not only under the Poincaré algebra, but also under dilatations (scale transformations), which we take to be generated by an operator D. The only non-zero commutator of dilatations with the generators of the Poincaré algebra is

$$[D, P^\mu] = P^\mu, \tag{23.7}$$

since energy momentum has scaling dimension 1, while the generators $M^{\mu\nu}$ of Lorentz rotations have dimension 0.

This algebra has a unique extension to the larger algebra of *conformal transformations*. It is thought to be only a mild assumption that scale-invariant QFT's are actually conformally invariant. In particular, there is no known example (I think) of a scale-invariant but not conformally invariant 4-dimensional QFT; see J. Polchinksi, *Nucl. Phys.* B303 (1988) 226, for a detailed discussion of this issue.

The conformal algebra has in addition to the Poincaré and dilatation generators, a vector of conformal generators K^μ. I will write out the full algebra in spinor notation (since that will be convenient for the supersymmetric generalization). Furthermore, I will write the algebra in Euclidean space. Then, D, $P^{\alpha\dot\alpha}$, $K^{\alpha\dot\alpha}$, $M^{\alpha\beta}$, and $\widetilde{M}^{\dot\alpha\dot\beta}$, respectively generate dilatations, translations, special conformal transformations, and the $SU(2) \times \widetilde{SU}(2) \in SO(4)$ Lorentz rotations.

The Lorentz algebra and charges are

$$\begin{aligned}
[M^{\alpha\beta}, M^{\gamma\delta}] &= i(M^{\alpha\delta}\varepsilon^{\beta\gamma} + M^{\alpha\gamma}\varepsilon^{\beta\delta} + M^{\beta\delta}\varepsilon^{\alpha\gamma} + M^{\beta\gamma}\varepsilon^{\alpha\delta}) \\
[M^{\alpha\beta}, X^\gamma] &= i(X^\alpha\varepsilon^{\beta\gamma} + X^\beta\varepsilon^{\alpha\gamma}),
\end{aligned} \tag{23.8}$$

where X is any generator with a single undotted index. The same formulas hold for the other $SU(2)$ (*i.e.*, with dotted indices). Here $\varepsilon^{\alpha\beta}$ and $\widetilde{\varepsilon}^{\dot\alpha\dot\beta}$ are antisymmetric 2-index tensors with $\varepsilon^{12} = \widetilde{\varepsilon}^{\dot1\dot2} = +1$. Defining

$$J^3 = \frac{i}{2}M^{12}, \qquad J^+ = \frac{1}{2}M^{11}, \qquad J^- = \frac{1}{2}M^{22}, \tag{23.9}$$

puts the algebra into familiar form

$$\begin{aligned}
[J^3, J^\pm] &= \pm J^\pm, & [J^+, J^-] &= 2J^3, \\
[J^3, X^1] &= \frac{1}{2}X^1, & [J^3, X^2] &= -\frac{1}{2}X^2.
\end{aligned} \tag{23.10}$$

The quadratic casimir $J^3(J^3+1)+J^-J^+ = j(j+1)$ measures the spin j of a representation. The casimir can be written in terms of the $M^{\alpha\beta}$ as

$$j(j+1) = \frac{1}{8}M^{\alpha\beta}M^{\gamma\delta}\varepsilon^{\alpha\gamma}\varepsilon^{\beta\delta}, \tag{23.11}$$

where summation over repeated indices is implied. An analogous definition exists for the other spin $\tilde{\jmath}$.

We choose D to be hermitian. Hermitian conjugation conjugates the $SU(2)$'s as well:

$$\begin{aligned} &D^{+} = D, \\ &(M^{\alpha\beta})^{+} = +\varepsilon^{\alpha\gamma}\varepsilon^{\beta\delta}M^{\gamma\delta}, \qquad (\widetilde{M}^{\dot\alpha\dot\beta})^{+} = +\widetilde{\varepsilon}^{\dot\alpha\dot\gamma}\widetilde{\varepsilon}^{\dot\beta\dot\delta}\widetilde{M}^{\dot\gamma\dot\delta}, \\ &(P^{\alpha\dot\alpha})^{+} = -\varepsilon^{\alpha\beta}\widetilde{\varepsilon}^{\dot\alpha\dot\beta}K^{\beta\dot\beta}, \qquad (K^{\alpha\dot\alpha})^{+} = -\varepsilon^{\alpha\beta}\widetilde{\varepsilon}^{\dot\alpha\dot\beta}P^{\beta\dot\beta}. \end{aligned} \tag{23.12}$$

Summation on repeated $SU(2)$ indices is implied. (I do not raise and lower indices in the usual way only to avoid confusing minus signs.)

The non-zero dimensions of the generators are given by

$$[D, P^{\alpha\dot\alpha}] = +P^{\alpha\dot\alpha}, \qquad [D, K^{\alpha\dot\alpha}] = -K^{\alpha\dot\alpha}. \tag{23.13}$$

The special conformal generators and their superpartners satisfy

$$[P^{\alpha\dot\alpha}, K^{\beta\dot\beta}] = \frac{i}{2}(M^{\alpha\beta}\widetilde{\varepsilon}^{\dot\alpha\dot\beta} + \widetilde{M}^{\dot\alpha\dot\beta}\varepsilon^{\alpha\beta}) + D\varepsilon^{\alpha\beta}\widetilde{\varepsilon}^{\dot\alpha\dot\beta}. \tag{23.14}$$

23.2. Representations of the conformal algebra

We will construct finite-dimensional representations of this algebra by "radial quantization". This means that we pick an S^3 centered around a point x_0 in 4 dimensions as our "space", and parametrize the radius of the sphere as $r = e^t$, with t our "time" variable. This gives us a one-to-one map between states and fields (local operators) in such a QFT by

$$\begin{aligned} |0\rangle &\leftrightarrow 1, \\ |\mathcal{O}\rangle &\leftrightarrow \mathcal{O}(x_0). \end{aligned} \tag{23.15}$$

Thus our "time" translation operator is D, so when we diagonalize this "Hamiltonian" in radial quantization, what we are really doing is studying the representation theory of the maximal compact subalgebra, $SO(4)$, of the conformal algebra with fixed dimension D. From the hermiticity of D and (23.13), we see that in radial quantization P is adjoint to K.

We can then find the finite-dimensional representations of the conformal algebra by looking at the *primary* or *highest-weight* states which are those annihilated by the K's

$$K^{\alpha\dot\alpha}|j, \tilde{\jmath}, d\rangle = 0. \tag{23.16}$$

Here we have labelled the highest-weight state by its eigenvalues under the $SU(2) \times \widetilde{SU(2)}$ rotations and the dilatations. The rest of the states in the representation of which $|j, \tilde{\jmath}, d\rangle$ is the highest-weight state are formed by acting on it with the "lowering" operator $P^{\alpha\dot\alpha}$:

$$\prod P|j, \tilde{\jmath}, d\rangle, \tag{23.17}$$

and are called *descendant* states. In terms of our state-operator correspondence, we have

$$\begin{aligned} |j,\tilde{\jmath},d\rangle &\leftrightarrow \mathcal{O}_{j,\tilde{\jmath},d}(x_0), \\ \prod P_\mu |j,\tilde{\jmath},d\rangle &\leftrightarrow (\prod \partial_\mu)\mathcal{O}_{j,\tilde{\jmath},d}(x_0). \end{aligned} \tag{23.18}$$

All the unitary irreducible representations of the conformal algebra can be classified as follows (Mack, *Comm. Math. Phys.* 55 (1977) 1):

$$\begin{array}{llll} j{=}\tilde{\jmath}=0 & d=0 & \text{identity} & \\ j\tilde{\jmath}=0 & d=j+\tilde{\jmath}+1 & \text{free, massless} & h=j-\tilde{\jmath} \\ j\tilde{\jmath}=0 & d>j+\tilde{\jmath}+1 & \text{(anti)chiral} & s=j+\tilde{\jmath} \\ j\tilde{\jmath}\neq 0 & d=j+\tilde{\jmath}+2 & \text{free} & s=j+\tilde{\jmath} \\ j\tilde{\jmath}\neq 0 & d>j+\tilde{\jmath}+2 & \text{general} & s=|j-\tilde{\jmath}|,...,j+\tilde{\jmath} \end{array} \tag{23.19}$$

The mass and spin/helicity (s/h) refer to the Poincaré content of the representations; all reps are massive unless noted otherwise. The spins in the last case take all integer steps between the limits.

It is not hard to derive these constraints from the conformal algebra. (The hard part is showing that they are sufficient.) Fields corresponding to representations with Lorentz spins j, $\tilde{\jmath}$ are denoted

$$\phi^{\alpha_1\ldots\alpha_{2j}\dot{\alpha}_1\ldots\dot{\alpha}_{2\tilde{\jmath}}}, \tag{23.20}$$

where the chiral Lorentz indices are separately symmetrized. (From now on the various $SU(2)$ indices of a single field will always be understood to be symmetrized.) All descendants are generated by applying $P^{\alpha\dot{\alpha}}$ to ϕ. The dimension of ϕ is $D(\phi)=d$.

In the scalar case, $j{=}\tilde{\jmath}{=}0$,

$$||P^{\alpha\dot{\alpha}}\phi||^2 = \langle 0,0,d|[K^\mu,P_\mu]|0,0,d\rangle = 4\langle 0,0,d|D|0,0,d\rangle = 4d, \tag{23.21}$$

implying $d>0$, and a null state when $d=0$. At the next level, $||P^2\phi||^2 = 8d(d-1)$, implying $d\geq 1$ and a null state at $d=1$ which is just the free massless wave equation.

In the chiral case, $j{\neq}0$, $\tilde{\jmath}{=}0$, $||\varepsilon^{\alpha\beta_1}P^{\alpha\dot{\alpha}}\phi^{\beta_1\ldots\beta_{2j}}||^2 = 2(d{-}j{-}1)$, implying $d\geq j{+}1$, and a null state when $d=j{+}1$. This null state gives the free massless wave equation since $P^{\beta_1}_{\dot{\alpha}}P^{\dot{\alpha}}_{\alpha}\phi^{\alpha\beta_2\ldots\beta_{2j}} = -\frac{1}{2}P^2\phi^{\beta_1\ldots\beta_{2j}}$.

Finally, for $j\tilde{\jmath}\neq 0$, $||\varepsilon^{\alpha\beta_1}\tilde{\varepsilon}^{\dot{\alpha}\dot{\beta}_1}P^{\alpha\dot{\alpha}}\phi^{\beta_1\ldots\beta_{2j}\,\dot{\beta}_1\ldots\dot{\beta}_{2\tilde{\jmath}}}||^2 = d{-}j{-}\tilde{\jmath}{-}2$, implying $d\geq j{+}\tilde{\jmath}{+}2$, with a null state when the inequality is saturated.

An interesting consequence of this classification was pointed out in P. Argyres, R. Plesser, N. Seiberg, and E. Witten, *Nucl. Phys.* B461 (1996) 71. Note that an Abelian field-strength field $F^{\mu\nu}$ is decomposed into $(1,0)$ and $(0,1)$ representations $F^{\pm}$. Then, if $D(F^{\pm})=2$, $dF^{\pm}=0$, which implies the free Maxwell equations and the Bianchi identities

$dF = d*F = 0$. On the other hand, if the field-strength is interacting, then $D(F^{\pm}) > 2$, implying $J^{\pm} \equiv dF^{\pm} \neq 0$. Since F^+ and F^- are independent representations of the conformal algebra, we learn from the equations of motion

$$dF = J^+ - J^- \equiv J_e \neq 0, \quad \text{and} \quad d*F = J^+ + J^- \equiv J_m \neq 0, \tag{23.22}$$

that the electric and magnetic currents J_e and J_m cannot vanich as quantum fields in this theory. We have shown that a $U(1)$ CFT in four dimensions is interacting if and only if it has both electrically *and* magnetically charged conformal fields in its spectrum.

A related, and perhaps somewhat suprising-sounding point, is that all Abelian gauge *charges* will vanish in a fixed-point theory (though they may still couple to massive degrees of freedom). In the case of the interacting $U(1)$ field strength F, though we have seen that its conserved electric and magnetic currents do not vanish, there is no charge at infinity associated with them, because of the rapid decay of correlation functions of F due to its anomalous dimension. This is true even if we include massive or background sources, since the long-distance behavior of the fields is governed by the CFT. If, on the other hand, F were free, then we have seen that its associated conserved currents, and thus the charges, vanish. Note, however, massive sources can have long-range fields in this case since F has its canonical dimension. (We do not reach a contradiction by taking the mass of a chrged source to zero since its $U(1)$ couplings flow to zero in the IR.) Non-Abelian gauge charges need not vanish in the CFT since the above arguments only apply to gauge-invariant fields or states.

23.3. N=1 superconformal algebra and representations

When we extend the conformal algebra by including the supersymmetry generators Q^α, $\widetilde{Q}^{\dot\alpha}$, we are forced by associativity to include three additional generators: the fermionic *superconformal generators* S^α and $\widetilde{S}^{\dot\alpha}$, and a scalar bosonic R generating $U(1)_R$ rotations.

In radial quatization, we choose R to be hermitian, and the others to satisfy

$$\begin{aligned} &R^+ = R, \\ &(Q^\alpha)^+ = +\varepsilon^{\alpha\beta} S^\beta, \qquad (S^\alpha)^+ = -\varepsilon^{\alpha\beta} Q^\beta, \\ &(\widetilde{Q}^{\dot\alpha})^+ = +\widetilde{\varepsilon}^{\dot\alpha\dot\beta} \widetilde{S}^{\dot\beta}, \qquad (\widetilde{S}^{\dot\alpha})^+ = -\widetilde{\varepsilon}^{\dot\alpha\dot\beta} \widetilde{Q}^{\dot\beta}. \end{aligned} \tag{23.23}$$

The non-zero dimensions of the generators are given by

$$\begin{aligned} &[D, Q^\alpha] = +\frac{1}{2} Q^\alpha, \qquad [D, S^\alpha] = -\frac{1}{2} S^\alpha, \\ &[D, \widetilde{Q}^{\dot\alpha}] = +\frac{1}{2} \widetilde{Q}^{\dot\alpha}, \qquad [D, \widetilde{S}^{\dot\alpha}] = -\frac{1}{2} \widetilde{S}^{\dot\alpha}, \end{aligned} \tag{23.24}$$

and likewise for the $U(1)_R$ charges

$$
\begin{aligned}
&[R, Q^\alpha] = +Q^\alpha, \qquad [R, S^\alpha] = -S^\alpha, \\
&[R, \widetilde{Q}^{\dot\alpha}] = -\widetilde{Q}^{\dot\alpha}, \qquad [R, \widetilde{S}^{\dot\alpha}] = +\widetilde{S}^{\dot\alpha}.
\end{aligned}
\tag{23.25}
$$

The conformal generators and their superpartners satisfy

$$
\begin{aligned}
&[K^{\alpha\dot\alpha}, Q^\beta] = i\widetilde{S}^{\dot\alpha}\varepsilon^{\alpha\beta}, \qquad [P^{\alpha\dot\alpha}, S^\beta] = i\widetilde{Q}^{\dot\alpha}\varepsilon^{\alpha\beta}, \\
&[K^{\alpha\dot\alpha}, \widetilde{Q}^{\dot\beta}] = iS^\alpha\widetilde{\varepsilon}^{\dot\alpha\dot\beta}, \qquad [P^{\alpha\dot\alpha}, \widetilde{S}^{\dot\beta}] = iQ^\alpha\widetilde{\varepsilon}^{\dot\alpha\dot\beta},
\end{aligned}
\tag{23.26}
$$

while the supersymmetry algebra and its conformal extension are given by:

$$
\begin{aligned}
&\{Q^\alpha, \widetilde{Q}^{\dot\alpha}\} = 2P^{\alpha\dot\alpha}, \qquad \{S^\alpha, \widetilde{S}^{\dot\alpha}\} = 2K^{\alpha\dot\alpha}, \\
&\{Q^\alpha, S^\beta\} = M^{\alpha\beta} - i(D - \frac{3}{2}R)\varepsilon^{\alpha\beta}, \\
&\{\widetilde{Q}^{\dot\alpha}, \widetilde{S}^{\dot\beta}\} = \widetilde{M}^{\dot\alpha\dot\beta} - i(D + \frac{3}{2}R)\widetilde{\varepsilon}^{\dot\alpha\dot\beta}.
\end{aligned}
\tag{23.27}
$$

In radial quantization, there is again a one-to-one map between states and local operators at the origin. Primary states, $|j, \tilde{\jmath}, d, r\rangle$, are in a representation of $SO(4) \times U(1)_R$ with fixed D, and are annihilated by the "raising operators" K^μ, S_i^α and $\widetilde{S}_i^{\dot\alpha}$. Descendants are formed from the primary states by acting on them with the Q and $\widetilde{Q}$ operators, (since P can be expressed as an anticommutator of Q and $\widetilde{Q}$). The classification of unitary irreducible representations is then (Dobrev and Petkova *Phys. Lett.* 162B (1985) 127):

$$
\begin{array}{lllll}
j{=}\tilde{\jmath} = 0 & d = 0 & r = 0 & \text{identity} & \\
\tilde{\jmath} = 0 & d = +\frac{3}{2}r & +\frac{3}{2}r \geq j + 1 & \text{chiral} & +\frac{3}{2}r = j + 1 \Rightarrow \text{free, massless} \\
j = 0 & d = -\frac{3}{2}r & -\frac{3}{2}r \geq \tilde{\jmath} + 1 & \text{antichiral} & -\frac{3}{2}r = \tilde{\jmath} + 1 \Rightarrow \text{free, massless} \\
 & \multicolumn{2}{l}{d \geq |\frac{3}{2}r - j + \tilde{\jmath}| + j + \tilde{\jmath} + 2} & \text{general} & \left\{ \begin{array}{c} j\tilde{\jmath}{\neq}0,\ \frac{3}{2}r{=}j{-}\tilde{\jmath} \\ \&\ d = j{+}\tilde{\jmath}{+}2 \end{array} \right\} \Rightarrow \text{free}
\end{array}
\tag{23.28}
$$

Thus, in general, $d \geq |\frac{3}{2}r|$, with equality only for the chiral or anti-chiral fields. In the above classification, the chiral fields are defined as those with $\tilde{\jmath} = 0$. It is easy to see from the superconformal algebra that this implies the usual condition for χsf's: $\widetilde{Q}^{\dot\alpha}\phi = 0$; similarly for the anti-chiral fields.

Consider the operator product expansion:

$$
\mathcal{O}_{d_1 = \frac{3}{2}r_1}(x_1)\, \mathcal{O}_{d_2 = \frac{3}{2}r_2}(x_2) = \sum_n (x_1 - x_2)^{d_n - d_1 - d_2} \mathcal{O}_{d_n, r_n = r_1 + r_2}(x_1).
\tag{23.29}
$$

Here $\mathcal{O}_1$ and $\mathcal{O}_2$ are chiral primary fields, while the fields on the right-hand side need not be; however, since the R-charges of the two sides must be the same, all fields on the right-hand side will have R-charge $r = r_1 + r_2$. Since for all fields $d_n \geq \frac{3}{2} r_n$, the exponents $d_n - d_1 - d_2$ on the right-hand side will all be zero or positive. This implies there are no singularities in the OPE of chiral fields. This implies that we can define a *chiral ring* of operators, by the products of chiral primary operators at the same point:

$$\mathcal{O}_{r_1}(x)\, \mathcal{O}_{r_2}(x) = \mathcal{O}_{r_1+r_2}(x). \tag{23.30}$$

24. N=1 duality

Let us apply this representation theory of the superconformal algebra to the singularity at the origin of the $n_f \geq n_c+2$ moduli space. Recall that the global symmetry group and charges of the super-QCD theory for $n_f \leq 3n_c$ is

	$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_R$
Q	$\mathbf{n_f}$	$\mathbf{1}$	1	$\frac{n_f - n_c}{n_f}$
$\widetilde{Q}$	$\mathbf{1}$	$\overline{\mathbf{n_f}}$	-1	$\frac{n_f - n_c}{n_f}$
M	$\mathbf{n_f}$	$\overline{\mathbf{n_f}}$	0	$2\frac{n_f - n_c}{n_f}$
B	$\binom{n_f}{n_c}$	$\mathbf{1}$	n_c	$n_c\frac{n_f - n_c}{n_f}$
$\widetilde{B}$	$\mathbf{1}$	$\overline{\binom{n_f}{n_c}}$	$-n_c$	$n_c\frac{n_f - n_c}{n_f}$

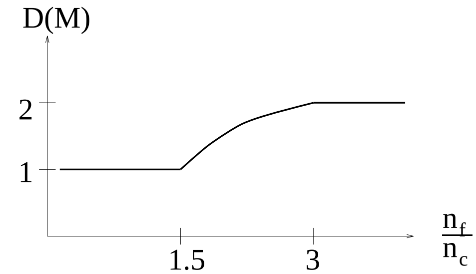
For sufficiently large n_c and n_f close to (but less than) $3n_c$, then we have seen that the fixed point is close to zero-coupling. The zero-coupling theory is free so is conformally invariant, and the $U(1)_R$ symmetry in the superconformal algebra is just the microscopic $U(1)_R$ shown above. So for the fixed-point at small value of the coupling, it is reasonable to assume that the $U(1)_R$ symmetry in its superconformal algebra is the same, since there is not enough "time" for relevant operators at the zero-coupling point to flow to irrelevant operators at the fixed point, and so make a new, "accidental", $U(1)_R$ symmetry in the IR.

Actually, since there is also the $U(1)_B$ symmetry, the $U(1)_R$ symmetry appearing in the superconformal algebra at the fixed point could be a combination of the $U(1)_R$ and $U(1)_B$ defined above. Notice, however, that this will not affect the R-charge of the meson field, since its baryon number vanishes. We thus read off the scaling dimension of M:

$$D(M) = \frac{3}{2} R(M) = 3\frac{n_f - n_c}{n_f}. \tag{24.1}$$

So, for $n_f \leq 3n_c$, $D(M) \leq 2$. For $n_f > 3n_c$ this formula implies $D(M) > 2$; however, we know that in this range the IR theory is free, so the quark χsf's have their canonical dimension of 1, and thus the meson χsf must have dimension 2. The reason the above formula fails in this case is that the IR free theory (being free) has an unbroken $U(1)_A$ in the IR which can mix with the $U(1)_R$ defined above.

The relation (24.1) also implies that $D(M) \leq 1$ for $n_f \leq \frac{3}{2}n_c$. Since dimensions less than 1 are not allowed by unitarity, it must be that a new accidental R-symmetry arises in this range. It is suggestive that right at $n_f = \frac{3}{2}n_c$, $D(M) = 1$, implying that M is free. This led N. Seiberg, `hep-th/9411149`, to guess that $D(M) = 1$ for $n_f \leq \frac{3}{2}n_c$, and so should be treated as an elementary field in an IR-free theory in this range.



Since the global symmetry must be the same as in the microscopic theory, one wants this IR free theory to have n_f fundamental flavors in an $SU(\widetilde{n_c})$ gauge theory. In order to be IR-free we need $\widetilde{n_c} < \frac{1}{3}n_f$ when $n_f \leq \frac{3}{2}n_c$. A simple choice that works is

$$\widetilde{n_c} \equiv n_f - n_c. \tag{24.2}$$

We will refer to this theory as the "dual theory", while we will call the original $SU(n_c)$ theory the "direct theory". (It is sometimes also referred to as the "magnetic theory", while the direct theory is called the "electric theory"; the reasons for these names will only become clear a few lectures from now.)

We assign the quantum numbers to the fundamental fields in the dual theory as

	$SU(\widetilde{n_c})$	$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_R$
M	$\mathbf{1}$	$\mathbf{n_f}$	$\overline{\mathbf{n_f}}$	0	$2\frac{n_f-n_c}{n_f}$
q	$\widetilde{\mathbf{n_c}}$	$\overline{\mathbf{n_f}}$	$\mathbf{1}$	$\frac{n_c}{n_f-n_c}$	$\frac{n_c}{n_f}$
$\widetilde{q}$	$\overline{\widetilde{\mathbf{n_c}}}$	$\mathbf{1}$	$\overline{\mathbf{n_f}}$	$\frac{-n_c}{n_f-n_c}$	$\frac{n_c}{n_f}$

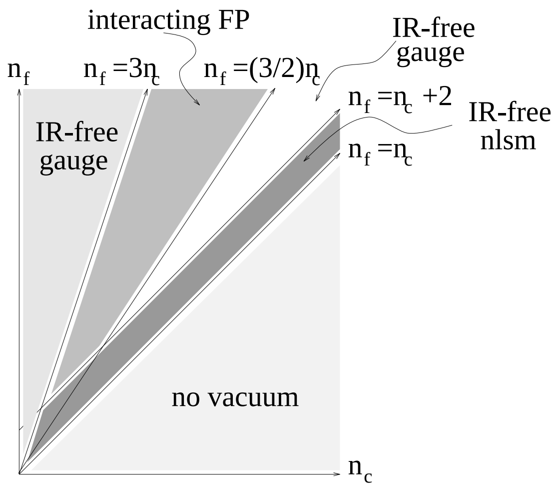
Here $SU(\widetilde{n_c})$ column are the gauge charges, while the rest are the (non-anomalous) global symmetries. The R-charges of the dual quark χsf's are fixed by anomaly cancellation. The normalization of their baryon number is chosen so that the dual baryons, $b \equiv q^{\widetilde{n_c}}$ and $\widetilde{b} \equiv \widetilde{q}^{\widetilde{n_c}}$, will have the same baryon number as the direct baryon fields B and $\widetilde{B}$. Indeed, with these assignments, we find the global charges of the gauge-invariant composite χsf's in the dual model to be

	$SU(n_f)$	$SU(n_f)$	$U(1)_B$	$U(1)_R$
$\widetilde{m}$	$\overline{\mathbf{n_f}}$	$\mathbf{n_f}$	0	$2\frac{n_c}{n_f}$
b	$\overline{\binom{n_f}{n_f-n_c}}$	$\mathbf{1}$	n_c	$(n_f-n_c)\frac{n_c}{n_f}$

$\widetilde{b}$	$\mathbf{1}$	$\binom{n_f}{n_f-n_c}$	$-n_c$	$(n_f-n_c)\frac{n_c}{n_f}$

where we have defined the dual meson to be $\widetilde{m} \equiv q\widetilde{q}$. Comparing with the global charges of the baryons in the direct theory, we see that they are the same, since as flavor representations $\overline{\binom{n_f}{n_c}} = \binom{n_f}{n_f-n_c}$.

This educated guess for an alternative (IR-equivalent) description of the vacuum physics at the origin of moduli space implies the following "phase diagram"



answering the question posed in the last lecture (compare the phase diagram of that lecture). The nature of this proposed solution is quite surprising: the AF direct gauge theory, at least for some range of n_f, is IR-equivalent to an IR-free gauge theory! This naturally raises the question of what is the relation between the IR-free gauge bosons and the direct (microscopic) gauge fields? No precise answer to this question is known.

24.1. Checks

Is there any way of checking this proposal?

The first thing to note is that the global symmetries of the direct and dual theories are the same. One can check that the 't Hooft anomaly-matching conditions all work.

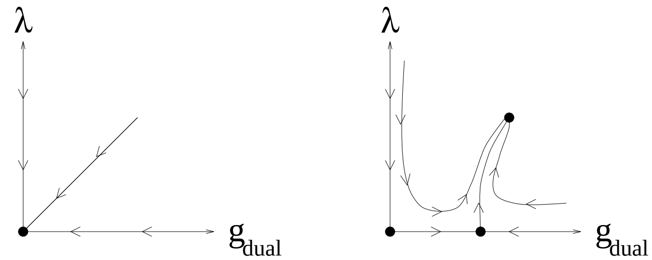
The next thing to check is whether these two theories have the same moduli space of vacua: do they have the same light gauge-singlet χsf's? In the direct theory away from the origin, we have M, B, and $\widetilde{B}$. In the dual theory, the elementary M, and the composite b and $\widetilde{b}$ fields have the same symmetry properties, and so can plausibly be identified. However, the dual theory also has the composite dual meson $\widetilde{m}$. To remove this operator from the dual theory, we must add some superpotential interaction. There is only one term allowed by the symmetries:

$$\mathcal{W}_{dual} = \lambda M q \widetilde{q}, \tag{24.3}$$

where λ is a dimensionless coupling. Such a coupling is just what we need to remove $\widetilde{m}$ as an independent degree of freedom in the IR, since the F-term equation for M implies that $q\widetilde{q} \equiv \widetilde{m} = 0$. Thus, for non-zero λ we at least have the right counting of light degrees of freedom away from the origin of moduli space.

The superpotential in the dual theory raises a new question, however: what is the correct value of λ? Actually, this is the wrong question, since the superpotential coupling is not exactly marginal.

For example, at the fixed point (the vacuum at the origin of moduli space) when $n_c+2 < n_f < \frac{3}{2}n_c$, the dual theory is IR-free, so the gauge-coupling, g_{dual} flows to zero. In a free theory, a Yukawa coupling like (24.3) is irrelevant, so λ also flows to zero. Thus the origin of the λ–g_{dual} plane is the fixed point. In the regime when $\frac{3}{2}n_c < n_f < 3n_c$, the λ and g_{dual} couplings are still irrelevant for large couplings, but $g_{dual} = 0$ is an UV fixed-point, since there is supposed to be an IR fixed point at $g_{dual} = g_* > 0$ when $\lambda = 0$. (Recall the form of the 2-loop β-function found at the beginning of last lecture.) However, at this IR fixed point $D(M) = 1$ since it is free (it has no couplings), and $D(q) = D(\widetilde{q}) = 3n_c/(2n_f)$ from their R-charges, implying $D(Mq\widetilde{q}) < 3$, and so is a relevant operator. Thus the superpotential will cause the theory to flow to a fixed point at non-zero $\lambda = \lambda_*$. These RG flows can be illustrated as:



$n_c+2<n_f<(3/2)n_c$ $\qquad$ $(3/2)n_c<n_f<3n_c$

Thus we expect the superpotential term to be irrelevant everywhere in the vicinity of the fixed point, except at the fixed point itself. (This situation is often described by saying that the operator in the superpotential is marginal but not exactly marginal.)

We can therefore trade λ for a scale in the dual theory, and so (as long as it is not zero) its value can have no effect on the scale-invariant far-IR physics. In the case where both the direct theory and the dual theory are AF, each has a gauge strong-coupling scale, Λ and Λ_{dual} respectively. However, the dual theory also has a second scale, which we can define as $\mu \sim \lambda\Lambda_{dual}$. The statement that these two theories are "dual" just means that they flow to the same theory at mass scales well below the smallest of Λ, Λ_{dual} and μ. We can trade λ for μ in the superpotential by noting that in the microscopic theory, M is a composite operator of canonical dimension 2 (in the UV), while in the dual theory it is a fundamental field of dimension 1 (in the UV). Then, if we define a new meson field by

$$M = M_{direct} \equiv \mu M_{dual}, \tag{24.4}$$

the dual superpotential becomes

$$\mathcal{W}_{dual} = \frac{1}{\mu} M q\widetilde{q}. \tag{24.5}$$

By the symmetries and holomorphy, the relation between the direct and dual strong-coupling scales must be

$$\Lambda_{direct}^{3n_c - n_f} \Lambda_{dual}^{3(n_f-n_c)-n_f} = (-)^{n_f - n_c} \mu^{n_f}. \tag{24.6}$$

The factor of $(-)^{n_f-n_c}$ can be determined by considering the dual of the dual theory. In this case we expect to regain the original theory with gauge group $SU(n_c)$ and quarks Q (since $n_f - \widetilde{n_c} = n_c$):

$$\begin{array}{ccccc} Q & \stackrel{\text{dual}}{\longrightarrow} & q\,,\, M & \stackrel{\text{dual}}{\longrightarrow} & Q, N, M \\ & & & & \\ \mathcal{W}=0 & & \mathcal{W} = \frac{1}{\mu} M q\widetilde{q} & & \mathcal{W} = \frac{1}{\mu} MN + \frac{1}{\widetilde{\mu}} NQ\widetilde{Q} \end{array} \tag{24.7}$$

where in the original theory M is the composite meson $M = Q\widetilde{Q}$, and similarly in the first dual $N = q\widetilde{q}$. From the superpotential of the (dual)^2 theory, we see that the (now fundamental) N mesons are massive and can be integrated-out, giving the required $M = Q\widetilde{Q}$ only if $\widetilde{\mu} = -\mu$. Then (24.6) implies that $\Lambda_{(dual)^2} = \Lambda_{direct}$ with the factor of $(-)^{n_f-n_c}$.

24.2. Matching flat directions

We will now analyze the moduli space of deformations of the two theories and show they are the same. We will do somewhat less than this, mainly because (as mentioned in lecture 20) we do not have a convenient description of this moduli space for general n_f and n_c. So we will outline what happens when we turn on vevs for the meson field in the two theories. The equivalence of the baryonic directions in moduli space are, as far as I know, less well understood.

Recall that the moduli space of the direct theory is the same as its classical moduli space, and that in the classical moduli space there are flat directions with arbitrary meson vevs with $\text{rank}(M) < n_c$; see eq. (20.10). Suppose we turn on a vev with $\text{rank}(M) = 1$:

$$\langle M \rangle = \begin{pmatrix} a^2 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}, \tag{24.8}$$

corresponding to giving only one component of the squarks a vev. The effect of this on the direct theory for large a is to Higgs the theory from $SU(n_c)$ with n_f flavors down to

$SU(n_c{-}1)$ with $n_f{-}1$ flavors. On the other hand, turning on this vev in the dual theory gives rise through the dual superpotential,

$$\mathcal{W}_{dual} = \frac{1}{\mu}\mathrm{Tr}Mq\widetilde{q} = \frac{a^2}{\mu}q_1 \cdot \widetilde{q}^1, \tag{24.9}$$

to a mass term for the q_1 and $\widetilde{q}^1$ dual quarks (here the 1 is a flavor index). Again for large a, integrating-out the massive quarks takes the dual theory from $SU(n_f{-}n_c)$ with n_f flavors to the $SU(n_f{-}n_c)$ theory with $n_f{-}1$ flavors. The equivalence of the direct and dual theories implies that the resulting theories after Higgsing or integrating-out should again be realted by our dual map, which indeed they are:

$$\begin{array}{ccc} \underline{\text{direct}} & & \underline{\text{dual}} \\ & & \\ SU(n_c),\, n_f & \longleftrightarrow & SU(n_f{-}n_c),\, n_f \\ \downarrow & & \downarrow \\ \text{Higgs} & & \text{mass} \\ \downarrow & & \downarrow \\ SU(n_c{-}1),\, n_f{-}1 & \longleftrightarrow & SU(n_f{-}n_c),\, n_f{-}1 \end{array} \tag{24.10}$$

Doing the more general case of higher-rank M is equivalent to simply repreating this procedure. Nothing new happens until we take $\mathrm{rank}(M) = n_c$, in which case it can be shown that the resulting direct and dual theories coincide, giving identical non-singular moduli spaces of meson and baryon vevs.

Alternatively to turning on vevs in the direct theory, we can turn on masses to the fundamental quarks. The corresponding deformation of the dual theory should again give rise to an equivalent theory. Suppose we turn on a mass for just the Q_1 and $\widetilde{Q}^1$ quarks:

$$\mathcal{W}_{direct} = mQ_1 \cdot \widetilde{Q}^1. \tag{24.11}$$

For large m, integrating-out this quark then takes the $SU(n_c)$ theory with n_f flavors to an $SU(n_c)$ theory with $n_f{-}1$ flavors. In the dual theory, on the other hand, turning on this mass corresponds to the superpotential

$$\mathcal{W}_{dual} = \frac{1}{\mu}\mathrm{Tr}(Mq\widetilde{q}) + mM_1^1, \tag{24.12}$$

which, upon integrating-out the M_1^1 component (by its F-term equation), gives rise to

$$\langle q_1 \cdot \widetilde{q}^1 \rangle = -m\mu. \tag{24.13}$$

For large m this is just Higgses the dual theory from $SU(n_c{-}n_f)$ with n_f flavors down to $SU(n_c{-}n_f{-}1)$ with $n_f{-}1$ flavors. This is again dual to the corresponding direct theory:

$$\begin{array}{ccc} \underline{\text{direct}} & & \underline{\text{dual}} \\ \\ SU(n_c),\, n_f & \longleftrightarrow & SU(n_f{-}n_c),\, n_f \\ \downarrow & & \downarrow \\ \text{mass} & & \text{Higgs} \\ \downarrow & & \downarrow \\ SU(n_c),\, n_f{-}1 & \longleftrightarrow & SU(n_f{-}n_c{-}1),\, n_f{-}1 \end{array} \tag{24.14}$$

Again, one can extend this to turning on mass matrices of arbtrary rank by repeating this procedure. This procedure ends with turning on a mass matrix with rank$(m) = n_f{-}n_c$, where again the resulting direct and dual theories can be shown to be the same.

24.3. The meaning of $N{=}1$ duality

So far we have presented strong evidence for Seiberg's duality conjecture. This conjecture posits the IR equivalence of two quite different-looking gauge theories—essentially it says the two theories are in the same universality class. There are a number of questions raised by this result: Can the conjecture be proven? Are there other classes of IR-equivalent gauge theories? Do we learn any general lessons about the structure and dynamics of gauge theories from these universality classes?

The first question has a partial positive answer. Using techniques that go beyond the scope of these lectures, it has been shown in P. Argyres, R. Plesser, and N. Seiberg, `hep-th/9603042` that the direct and dual gauge theories are in the same universality class. The idea is to find a larger theory (in this case the $N{=}2$ supersymmetric $SU(n_c)$ theory with n_f flavors) and show that by tuning a relevant parameter (an $N{=}2$ breaking mass term) one can flow arbitrarily close to either the direct or dual UV fixed point theories. One then argues that there can be no phase transition as one adjusts a relevant parameter in a supersymmetric theory, implying that the direct and dual theories must indeed be in the same universality class. The absence of phase transitions can be seen by putting the theory in a finite volume; then, as in our discussion of supersymmetric quantum mechanics, the vacuum energy will be an analytic function of the parameter. In these non-chiral theories, a Witten-index argument implies that supersymmetry is unbroken at finite volume, so the vacuum energy is always zero, and there can be no vacuum level-crossing (zero temparture phase transition) as the parameter is varied. This argument shows the IR equivalence of the non-chiral (Kahler) as well as the chiral parts of the the theory. This is in contrast to the above consistency arguments which only probed the chiral ring of the two theories.

The second question of generalizations of $N{=}1$ duality has given rise to a fairly rich "phenomenology" of dual sets of theories for other gauge groups and matter content.

Perhaps especially interesting among these dualities are chiral–non-chiral dual pairs. Only the simplest of these dualities have found derivation as described in the last paragraph. No simple constructive rules for predicting other dual sets has been given. Also, the question of IR equivalences among non-simple (product) gauge theories has not been systematically explored.

Finally, the question of what general lessons can be derived from the existence and systematics of these gauge universality classes has not been answered. There are many suggestions that these dualities are related to a different kind of duality among QFT's called *S-duality*. S-duality is the *exact* quantum equivalence of theories with an exactly marginal operator at different values of the coefficient of this operator. (In a few lectures we will discuss the simplest example of such an S-duality: electric-magnetic duality in Abelian gauge theories.) However, there is as yet no clear statement of the relation between $N{=}1$ and S- dualities.

25. Dynamical Supersymmetry Breaking

In this lecture I will present three examples of models where dynamical supersymmetry breaking (DSB) can be shown to occur. They represent, in some sense, three different "mechanisms" for supersymmetry breaking. Showing that DSB occurs involves minimizing the scalar potential which involves the inverse Kahler metric. Since we do not have control over the Kahler metric, we will be forced to make assumptions or go to weak coupling.

25.1. Supersymmetry breaking by dynamically generated superpotentials

By the Witten index argument, supersymmetry breaking can not take place in non-chiral theories where one can turn on bare masses (consistent with gauge symmetry) for all the χsf's in the theory, since under this deformation one flows to pure super-YM theory at low energies which doesn't break supersymmetry. This suggests we look to chiral theories for supersymmetry breaking.

As an example (first analyzed in Affleck, Dine, and Seiberg *Phys. Lett.* 137B (1984) 187) consider an $SU(3)\times SU(2)$ gauge theory with the following field content:

	$SU(3)$	$SU(2)$	|	$U(1)_Q$	$U(1)_u$	$U(1)_d$	$U(1)_L$	$U(1)_R$
Q	$\mathbf{3}$	$\mathbf{2}$	|	1	0	0	0	1
u	$\overline{\mathbf{3}}$	$\mathbf{1}$	|	0	1	0	0	1
d	$\overline{\mathbf{3}}$	$\mathbf{1}$	|	0	0	1	0	1
L	$\mathbf{1}$	$\mathbf{2}$	|	0	0	0	1	1

The first two columns are the gauge symmetries, and the last five columns are (classsical) global symetries. Note that actually $U(1)_u\times U(1)_d \subset U(2)$, but we will not need to use the larger symmetry. Thus the field content is like that of one generation of the standard model, but without the $U(1)$-hypercharge gauge group.

There are 11 gauge bosons and 14 χsf's, implying at most 3 flat directions. They can be parametrized by the gauge-singlet composite χsf's

$$X_1 \equiv QuL, \quad X_2 \equiv QdL, \quad Y \equiv Q^2 ud. \tag{25.1}$$

Their global charges, and the charges of the strong-coupling scales Λ_2 and Λ_3 of the gauge groups (reflecting anomalies in the various classical global symetries) are

	$U(1)_Q$	$U(1)_u$	$U(1)_d$	$U(1)_L$	$U(1)_R$
Λ_3^7	2	1	1	0	6
Λ_2^4	3	0	0	1	4
X_1	1	1	0	1	3
X_2	1	0	1	1	3
Y	2	1	1	0	4

Thus, out of the five classical $U(1)$'s, three linear combinations will be non-anomalous, including one R-symmetry. We can take as convenient non-anomalous combinations

$$R' = R - 2Q - u - d + 2L, \quad S = Q - 4u + 2d - 3L, \quad T = u - d, \tag{25.2}$$

which refer to the generators of the various $U(1)$'s. In this basis the charges of the singlet χsf's are

	$U(1)_T$	$U(1)_S$	$U(1)_{R'}$
X_1	1	-6	2
X_2	-1	0	2
Y	0	0	-2

Can the classical flat directions be lifted quantumly? Using the selection rules from all the symmetries, the only superpotential that can be generated dynamically is

$$\mathcal{W} = c \cdot \left(\frac{\Lambda_3^7}{Y} \right). \tag{25.3}$$

The coefficient c can be determined by "turning off" the $SU(2)$ gauge couplings (*i.e.* taking $\Lambda_2 \to 0$), in which case the theory becomes $SU(3)$ super-QCD with 2 flavors which we saw previously has $c = 1$. Thus the flat directions are lifted, and the theory has no vacuum.

Let us now modify the theory by turning on a tree-level superpotential

$$\mathcal{W}_{tree} = \lambda X_2 \tag{25.4}$$

We assign λ charges

	$U(1)_Q$	$U(1)_u$	$U(1)_d$	$U(1)_L$	$U(1)_R$
λ	-1	0	-1	-1	-1

This explicitly breaks the global $U(1)_T$ symmetry, leaving $U(1)_S$ and $U(1)_{R'}$ as unbroken symmetries. Using the selection rules, holomorphy, and taking the weak-coupling limits $\lambda \to 0$ and $\Lambda_3 \to 0$, we find that the exact effective superpotential is

$$\mathcal{W} = \frac{\Lambda_3^7}{Y} + \lambda X_2. \tag{25.5}$$

There are still no solutions to the F-term equations from this superpotential, so either there is no vacuum, or supersymmetry is broken. Since the λ term is rising at infinity in field space, it is plausible that the potential will have a minimum and that supersymmetry will be broken.

To calculate whether this breaking actually occurs requires minimizing the scalar potential $V = g^{i\bar{i}} \partial_i \mathcal{W} \partial_{\bar{i}} \overline{\mathcal{W}}$, which involves the Kahler potential. Assuming the minimum occurs at large vevs for λ small, then the Kahler potential will be close to its classical value,

$$\mathcal{K} = \overline{Q}Q + \overline{u}u + \overline{d}d + \overline{L}L. \tag{25.6}$$

since the physics is weakly-coupled in this regime. Writing the superpotential in terms of the microscopic fields,

$$\mathcal{W} = \frac{\Lambda_3^7}{Q^2 du} + \lambda QdL, \tag{25.7}$$

we find the minimum of the scalar potential at

$$\begin{aligned} \langle X_1 \rangle &= 0 \\ \langle X_2 \rangle &\sim \Lambda_3^3 \lambda^{-3/7} \\ \langle Y \rangle &\sim \Lambda_3^4 \lambda^{-4/7} \\ E &\sim \Lambda_3^4 \lambda^{10/7} \end{aligned} \tag{25.8}$$

where E is the vacuum energy. This solution justifies our assumption that the vacuum was at large vevs for small λ.

This model of DSB has been generalized to many different models with product gauge groups; notably ones with gauge group $SU(n) \times Sp(2m)$, M. Dine, A. Nelson, Y. Nir, and Y. Shirman, `hep-ph/9507378`.

25.2. Explicitly breaking the R-symmetry

That the model we just constructed has DSB is not too surprising, since the model has a spontaneously broken $U(1)_R$ symmetry. Recall, by a general counting argument that the F-term conditions in such a situation generically do not have a solution. Since the model also has no classical flat directions, generically the vacuum will be stablized (not run off to infinity) and supersymmetry will be broken. The spontaneously broken supersymmetry gives rise to a massless Goldstino, while the spontaneously broken $U(1)_{R'}$ gives rise to a massless Goldstone boson. The former is lifted by gravitational effects, while the latter is phenomenologically unacceptable. In particular, in more realistic models, the Goldstone boson gains a small mass from the QCD anomaly, becoming an "R-axion"; astrophysical bounds are quite restrictive for such a particle, implying that the scale of spontaneous breaking of the $U(1)_{R'}$ symmetry can only be in a narrow range ($10^{10} - 10^{12}$ GeV).[32]

Is there any way to avoid having an exact $U(1)_{R'}$ symmetry in a model with DSB?

There are general arguments based on quantum gravitational effects[33] and on string theory[34] which lead us to expect that any apparent global symmetries in an effective Lagrangian are just accidental IR symmetries, and therefore approximate. We might therefore expect there to be non-renormalizable operators in the superpotential which explicitly break $U(1)_{R'}$. For some unspecified (generic) physics which gives rise to such non-renormalizable terms in an effective theory below a scale m, we generically expect supersymmetry to be restored by our previous argument. This would seem to imply that we need to fine tune the $U(1)_R$–breaking physics at scale m in order to have DSB.

This conclusion is wrong, however, because the low-energy effective superpotential obtained by integrating-out the m-scale physics is not generic due to holomorphy. For example (A. Nelson and N. Seiberg `hep-th/9309299`), consider adding to the model of the last section two $SU(2)$ singlet quark χsf's:

	$SU(3)$	$SU(2)$	
S	$\mathbf{3}$	$\mathbf{1}$	
$\widetilde{S}$	$\overline{\mathbf{3}}$	$\mathbf{1}$	

with a large mass m and the most general renormalizable couplings:

$$\mathcal{W}_{tree} = \lambda QdL + mS\widetilde{S} + \lambda_1 Q\widetilde{S}L + \lambda_2 Q^2 S + \lambda_3 ud\widetilde{S}. \tag{25.9}$$

It is easy to check that this theory has no $U(1)_R$ symmetry. Is supersymmetry restored? First integrate-out S and $\widetilde{S}$ at tree-level, giving

$$\mathcal{W}_{tree} = \lambda X_2 + \frac{\lambda_1\lambda_2}{m} Q^3 L + \frac{\lambda_2\lambda_3}{m} Y. \tag{25.10}$$

The second term vanishes classically, by Bose statistics, but quantumly this constraint is replaced by $Q^3L \sim \Lambda_2^4$, as can be seen by turning off Λ_3, leaving an $SU(2)$ super-QCD with

[32] See, *e.g.* , J. Kim *Phys. Rep.* 149 (1987) 1.

[33] S. Hawking, *Comm. Math. Phys.* 43 (1975) 199, *Phys. Lett.* 195B (1987) 337; G. Lavrelashvili, V. Rubakov, and P. Tinyakov, *JETP Lett.* 46 (1987) 167; S. Giddings and A. Strominger, *Nucl. Phys.* B307 (1988) 854; S. Coleman, *Nucl. Phys.* B310 (1988) 643; T. Banks, *Physicalia* 12 (1990) 19.

[34] T. Banks and L. Dixon *Nucl. Phys.* B307 (1988) 93.

2 flavors. To find the exact effective superpotential, use the usual symmetries, holomorphy, and limits argument, to find

$$\mathcal{W} = \lambda X_2 + \frac{\lambda_2\lambda_3}{m}Y + \frac{m\Lambda_3^6}{Y} + \frac{\lambda_1\lambda_2\Lambda_2^4}{m}. \tag{25.11}$$

Here the last term is a constant, which has no effect in global supersymmetry; its coefficient was determined by matching to the constraint for Q^3L in the $\Lambda_3 \to 0$ limit. This exact superpotential again breaks supersymmetry, but has no R-symmetry.

J. Bagger, E. Poppitz, and L. Randall, `hep-th/9405345`, have pointed out in the context of supergravity models that the vanishing of the cosmological constant requires the addition of a (finely-tuned) constant to the effective low-energy (field-theory) superpotential. Such a constant, though it has no effect in global supersymmetry, does explicitly break any R-symmetry. This then gives rise to an explicit mass term for the R-axion due to a cross-term in the scalar potential (as calculated in supergravity, see lecture 17, section 2) between the constant and non-constant parts of the superpotential. Because the R-axion mass, m_a, arises as a gravitational effect, it is typically varies inversely with M_P, the Planck mass. In the hidden-sector models described in lecture 18, it is easy to see that it typically goes as $m_a \sim M_S^3/M_P$, giving $m_a \sim 10^7$ GeV for the gravity-mediated models, and $m_a \sim 10$ MeV for the gauge-mediated models. The large mass in the gravity-mdeiated models makes the R-axion harmless; its mass in the gauge-mediated models is just at the limit of viability: any lower and it would cool supernovas too quickly. This suggests that it may be phenomenologically allowed to look to field theory models with spontaneously-broken R-symmetries for DSB.

25.3. DSB by confinement in a smooth quantum moduli space

To summarize so far, the mechanism for DSB of the last two sections relied on the existence of a dynamically generated superpotential which lifts the classical moduli space and drives scalar fields to large vevs. Additional tree-level interactions gave a potential which rises at large vevs, stabilizing the ground state. Under special circumstances supersymmetry is broken in this vacuum. This mechanism had the desriable feature that the DSB could take place for large vevs, where the Kahler potential was under control.

We will now turn to other, less well-understood, mechanisms for DSB. They are less well-understood mainly because the DSB takes place at strong coupling, where the Kahler potential is not reliably calculable in perturbation theory. Special arguments (basically 't Hooft anomaly matching at points of unbroken global symmetries) will be needed to gain some control over the Kahler potential.

We have seen many examples in super-QCD of theories which have no dynamically-generated superpotential, and the classical moduli space remains unmodified quantumly (*e.g.* the $n_f \geq n_c+1$ $SU(n_c)$ theories). In some circumstances it might be possible to have

DSB in such a theory. The following example was found by K. Intriligator, N. Seiberg, and S. Shenker, `hep-th/9410203`.

Consider an $SU(2)$ gauge theory with a single χsf, Q, in the spin-3/2 representation. This theory classically has a $U(1)\times U(1)_R$ global symmetry, of which only the R-symmetry is non-anomalous. The only gauge-invariant composite χsf is $U = Q^4$, a totally symmetric combination. The charges of these fields and the strong-coupling scale Λ are

	$SU(2)$		$U(1)$	$U(1)_R$
Q	**4**		1	3/5
Λ	**1**		10	0
U	**1**		4	12/5

The classical moduli space is the U-plane. The classical Kahler potential is $\mathcal{K} \sim \overline{Q}Q \sim (\overline{U}U)^{1/4}$, implying a $\mathbb{Z}_4$ singularity at $U = 0$, where classically the gauge bosons become massless. Since it is at strong coupling, this classical prediction of a singularity is not reliable. Quantumly, by the usual arguements, no superpotential can be generated, and so the flat directions are not lifted.

What happens at $U = 0$? The singularity in the Kahler potential could remain, implying that new massless degrees of freedom arise there. In fact, if this is what happens, then the vacuum at the origin would be an interaction conformal fixed point, since from the R charge we see that the light field has an anomalous dimension $D(U) = 18/5$.

A simpler possibility is that at $U = 0$ the only massless quanta are just the U field. If this were the case, it would imply that the singularity in the Kahler potential was smoothed-out by quantum effects, and was regualar at the origin. This can be tested by checking the 't Hooft anomaly matching conditions:

$$\begin{aligned}\mathrm{Tr}R &= 3\cdot 1 + 4\cdot(-2/5) \qquad \text{(micro)}\\ &= (7/5) \qquad \text{(macro)},\end{aligned} \tag{25.12}$$

and

$$\begin{aligned}\mathrm{Tr}R^3 &= 3\cdot 1^3 + 4\cdot(-2/5)^3 \qquad \text{(micro)}\\ &= (7/5)^3 \qquad \text{(macro)}.\end{aligned} \tag{25.13}$$

We take this admittedly rather thin evidence as confirmation of the quantum smoothing of the $U = 0$ singularity.

We now turn on a tree-level superpotential

$$\mathcal{W}_{tree} = \lambda U, \tag{25.14}$$

and by the usual arguments, the exact effective superpotential is the same: $\mathcal{W} = WW_{tree}$. Being linear in U, it lifts the flat directions and breaks supersymmetry. This conclusion could only be modified by a singularity in the Kahler metric, which we have argued does not occur. Using a Kahler potential which goes like $\mathcal{K} \sim \overline{U}U$ for small U, it is easy to show that the vacuum energy goes as $V \sim |\lambda^2\Lambda^6|$, the factors of Λ entering by dimensional considerations.

One should note that the tree-level superpotential is non-renormalizable. We should therefore take the coupling λ to go as the inverse of some cut-off scale m:

$$\lambda \sim \frac{1}{m}. \tag{25.15}$$

Thus any vacuum we find using this superpotential will only be reliable if the U vev satisfies

$$|\langle U\rangle| \ll m^4, \tag{25.16}$$

otherwise we potentially more irrelevant operators could be important, implying that we need to include the exact superpotential arising from integrating-out the physics above the cut-off scale. Our DSB solution can satisfy this condition as long as we take $m \gg \Lambda$, since then the Kahler potential reliably goes as

$$\mathcal{K} \sim (\overline{U}U)^{1/4} \quad \text{for} \quad \Lambda^4 \ll |U| \ll m^4, \tag{25.17}$$

which means the minimum of the scalar potential will occur for $|U| < \Lambda^4$.

Note also that the superpotential explicitly broke the R-symmetry. To summarize, this model had DSB due to the Kahler potential being (with some evidence) modified quantumly, without any dynamically generated superpotential.

25.4. DSB by quantum deformation of classical moduli space

In the $n_f = n_c$ theories of $SU(n_c)$ super-QCD we saw that though no superpotential was dynamically generated, the classical moduli space was deformed quantumly, "smoothing-out" a classical singularity by changing the constraints among the light χsf's. K. Intriligator and S. Thomas, `hep-th/9603158`, have pointed out that DSB can occur in theories with this behavior as well.

Consider the $SU(2)$ super-QCD with $n_f = 2$. In previous lectures we described this as two quarks Q_i in the $\mathbf{2}$ of $SU(2)$ and two anti-quarks $\widetilde{Q}^i$ in the $\overline{\mathbf{2}}$. But $SU(2)$ (unlike the higher-rank unitary groups) has only real representations; in particular the $\mathbf{2}$ and $\overline{\mathbf{2}}$ representations are equivalent. So, this theory can be described more symmetrically as $SU(2)$ with four doublet quark χsf's Q_i, $i = 1, \ldots, 4$. Thus, this model actually has a $U(4)$ rather than a $U(2) \times U(2)$ flavor symmetry. The classical moduli space is parametrized by the vevs of the singlets

$$M_{ij} = -M_{ji} \equiv Q_iQ_j, \tag{25.18}$$

which are antisymmetric to make an $SU(2)$ color singlet. (In our old division into quarks and anti-quarks, this "meson" field combines both the old mesons and baryons.) The classical moduli space is subject to the constraint

$$\mathrm{Pf}M \equiv \epsilon^{ijkl} M_{ij} M_{kl} = 0 \qquad \text{(classical)}; \tag{25.19}$$

Quantumly, the constraint is modified to

$$\mathrm{Pf}M = \Lambda^4 \qquad \text{(quantum)}. \tag{25.20}$$

While the singular point $M_{ij} = 0$ lies on the classical moduli space, it is not part of the quantum moduli space.

Now consider adding to this $SU(2)$ theory six singlet fields $S^{ij} = -S^{ji}$ with couplings

$$\mathcal{W}_{tree} = \lambda S^{ij} Q_i Q_j = \lambda S^{ij} M_{ij}. \tag{25.21}$$

This leaves unbroken the $SU(4)$ flavor symmetry; there is also an anomaly-free $U(1)_R$ symmetry under which $R(Q) = 0$ and $R(S) = 2$. The usual symmetry and holomorphy argument implies the exact effective superpotential is the same as the tree-level potential: $\mathcal{W} = \mathcal{W}_{tree}$. Classically there is a moduli space of supersymmetric vacua with $M_{ij} = 0$ and S^{ij} arbitrary. Quantumly, the S^{ij} equations of motion, $\lambda M_{ij} = 0$, are incompatible with the quantum constraint (25.20). The classical moduli space is therefore completely lifted for $\lambda \neq 0$, and supersymmetry is broken.

Recall that the quantum modification of the moduli space can be realized in the superpotential with a Lagrange multiplier field A enforcing the constraint:

$$\mathcal{W} = \lambda S \cdot M + A(\mathrm{Pf}M - \Lambda^4). \tag{25.22}$$

For $\lambda \ll 1$ the supersymmetry breaking vacuum is close to the $SU(2)$ quantum moduli space, so we can analyze the model by pertubing around this space. The quantum constraint generically breaks the $SU(4)$ flavor symmetry to $SU(2) \times SU(2)$, but there is a point with enhanced symmetry $Sp(4) = SO(5)$ when

$$M_0 = \pm\Lambda^2 \begin{pmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & 0 & 1 \\ & & -1 & 0 \end{pmatrix}. \tag{25.23}$$

The fluctuations of M_{ij} away from M_0 subject to the quantum constraint are fields M_5 in the **5** of the $SO(5)$ flavor symmetry. (Enhanced symmetry points are always extrema of the scalar potential, so are good candidates for minima.)

As long as $\lambda S^{ij} \ll \Lambda$, the M_{ij} fields are light, since the λSM superpotential term looks like a small mass term for the quarks. So

$$\hat{M}_5 = M_5/\Lambda \tag{25.24}$$

comprises the massless spectrum near the vacuum at M_0, and the Kahler potential will have its canonical form at this point $\mathcal{K} \sim \overline{M}_5 M_5/|\Lambda|^2$ up to small corrections. Under the $SO(5)$ the S^{ij} break up as $S_5 \oplus S_0$ transforming as **5** and **1**. The λSM superpotential becomes

$$\mathcal{W} = \lambda M_5 S_5 \pm 2\lambda\Lambda^2 S_0 \tag{25.25}$$

at M_0. The first term lifts the M_5 and S_5 moduli which are present for $\lambda = 0$, giving them masses $\sim \lambda\Lambda$. The second term is linear in S_0, therefore breaking supersymmetry and giving two degenerate vacua of energy $V \sim |\lambda^2\Lambda^4|$.

When $\lambda S^{ij} \gg \Lambda$, the quarks are heavy, and we integrate them out by enforcing the equations of motion for M_{ij}, and the superpotential becomes

$$\mathcal{W} = \pm 2\lambda\Lambda^2\sqrt{\mathrm{Pf}S}, \tag{25.26}$$

in the limit of small λ. The F-term equations for the S^{ij} have no solution, so supersymmetry is again broken. The F-component of S_0 again breaks the supersymmetry, giving a pair of degenerate vacua of energy $V \sim |\lambda^2\Lambda^4|$. The S_5 components get masses $\sim \lambda\Lambda^2/S_0$. Note that this model seems to have a flat direction corresponding to the S_0 direction. This would be an exactly flat direction if the Kehler potential were precisely canonical; quantum corrections to the Kahler potential can lift this degeneracy. Unfortunately these quantum corrections are not calculable, so it is not known whether the minimum lies at $S_0 \sim 0$, $\mathcal{O}(\Lambda)$, or ∞. If the minimum is at infinity, then the model has no ground state, though DSB still occurs since the potential approaches a non-zero constant there.

These latter vacua can also be described in terms of gaugino condensation: after integrating out the quarks, the low energy theory is just the S^{ij} singlets plus pure $SU(2)$ super-YM with a strong-coupling scale $\hat{\Lambda}^6 = \lambda^2(\mathrm{Pf}S)\Lambda^4$ by the usual matching. Gaugino condensation in the pure $SU(2)$ gives a superpotential $\mathcal{W} = \pm 2\hat{\Lambda}^3$, which is (25.26).

This mechanism for DSB can be easily generalized to other theories with quantum-deformed moduli spaces, like the $SU(n_c)$ super-QCD with $n_f = n_c$, or the $Sp(2n_c)$ with $n_f = n_c{+}1$ fundamental flavors.

The mechanism for DSB differs from the first mechanism presented in this lecture in two surprising ways: it does not need a chiral theory, and the classical superpotential has flat directions. Being a non-chiral theory would seem to contradict the Witten index argument which said that the Witten index does not change under deformations of a theory by relevant parameters like masses for vector-like matter; so giving all the matter in a vector-like theory large masses, the low-energy theory is just a pure super-YM theory for which the Witten index does not vanish. The loophole is just that these theories have classically flat directions, along which vacua can "run off to infinity" as masses are added, changing the value of the index. (In other words, even though the masses are relevant operators, they still change the large-vev behavior of the superpotential.)

The existence of the classical flat directions may seem to contradict the conventional wisdom that models of DSB should have no flat directions, otherwise the vacuum will run off to infinity. This lore does not apply in this case since, in contrast to flat directions along which an AF gauge group is Higgsed and becomes weaker for larger vevs, here the matter fields become more massive and the theory becomes more strongly coupled, leading to a vacuum energy which does not vanish even infinitely far along the classical flat direction.

26. Theories with low-energy photons: monopoles and electric-magnetic duality

In these last two lectures of the course I will give a brief introduction to the description of low-energy effective actions in the "Coulomb phase". The Coulomb phase is the part of the moduli space where there are unbroken (or unconfined) $U(1)$ Abelian gauge factors in the IR—massless photons. The simplest example of a Coulomb branch occurs in the $SU(2)$ super-YM theory with and adjoint χsf Φ. This example with a special superpotential interaction is actually $N{=}2$ supersymmetric, and was solved by Seiberg and Witten, `hep-th/9407087`. The $N{=}1$ case was solved by Intriligator and Seiberg, `hep-th/9408155`, which we will follow.

So, let us consider $SU(2)$ with an adjoint Φ^a_b. The adjoint representation can be thought of as the set of hermitian traceless 2×2 matrices, acted on by gauge transformations as $\Phi \to g\Phi g^{-1}$ where $g \in SU(2)$. Since the scalar component of Φ is complex, it takes values in the set of complex traceless 2×2 matrices (no hermiticity condition). The classical moduli space is parametrized by the singlet composite χsf

$$U = \mathrm{Tr}\Phi^2. \tag{26.1}$$

Higher powers of Φ in the color trace are just polynomials in U. Thus the classical moduli space is the complex U-plane. The classical Kahler potential is $\mathcal{K} \sim (\overline{U}U)^{1/2}$, so there is a $\mathbb{Z}_2$ conical singularity at the origin, corresponding to the vacuum where the full $SU(2)$ symmetry is restored.

This moduli space is actually in a Coulomb phase. One way of seeing this is to note that Φ has χsf degrees of freedom, U has one, so only two were given mass. Thus only two of the three $SU(2)$ are Higgsed, so it must be that $SU(2) \to U(1)$ on the U-plane. This can be seen more directly by noting that the D-term equations imply

$$[\overline{\Phi}, \Phi] = 0, \tag{26.2}$$

which implies that Φ can be diagonalized by color rotations:

$$\Phi = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}, \tag{26.3}$$

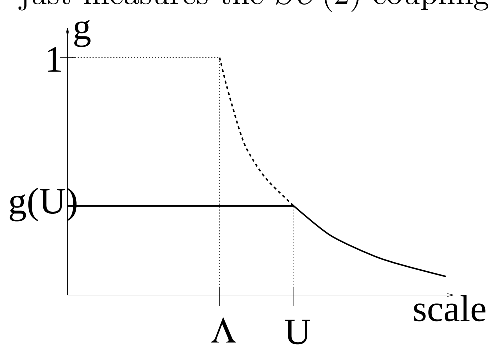
and there is a discrete gauge identification $a \simeq -a$. It is easy to see that (26.3) leaves the diagonal $U(1) \subset SU(2)$ unbroken, and the light field U is neutral under this $U(1)$.

The light degrees of freedom are thus a $U(1)$ vsf W_α and the χsf U. This theory has an anomaly-free $U(1)_R$ symmetry under which $R(W_\alpha) = 1$ and $R(U) = 0$. There is also an anomalous $U(1)_A$ under which W_α is neutral, $U \to e^{i\alpha}U$, and $\vartheta \to \vartheta + 2\alpha$. By the usual arguments, there is no dynamically generated superpotential for U, so the classical flat directions are not lifted.

This is not the whole story, though, since there is also the kinetic term for the vsf:

$$\mathcal{L}_{eff} = \text{Kahler} + \int d^2\theta\, \tau(U, \Lambda^4) \text{tr}(W_\alpha^2) + h.c. \tag{26.4}$$

where $\tau(U) = \frac{\vartheta(U)}{2\pi} + i\frac{4\pi}{g^2(U)}$ is the low-energy coupling, which can depend on U. Our goal will be to detemine the coupling function $\tau(U)$. What is the meaning of this $U(1)$ coupling in an IR effective theory? Classically, the AF $SU(2)$ theory is being Higgsed at the scale $U^{1/2}$ down to $U(1)$; since the fields charged under the $U(1)$ (*e.g.* the $W^\pm$ bosons) get masses of order $U^{1/2}$, they decouple at smaller scales, and the $U(1)$ coupling does not run. Thus the IR coupling τ just measures the $SU(2)$ coupling at the scale $U^{1/2}$.



By asymptotic freedom, for $\langle U \rangle \gg \Lambda^2$ this one-loop description of the physics should be accurate. The question is what happens for $\langle U \rangle < \Lambda^2$.

Recall that under the anomalous $U(1)_A$ rotations the ϑ-angle and therefore τ shifts. By the angular nature of the ϑ-angle, the shift $\tau \to \tau{+}1$ is a symmetry. This, plus holomorphy and matching to the one-loop β-function at weak coupling (large U) implies

$$\tau(U) = \frac{1}{2\pi i} \log\left(\frac{\Lambda^4}{U^2}\right) + \sum_{n=0}^{\infty} c_n \left(\frac{\Lambda^4}{U^2}\right)^n. \tag{26.5}$$

The first term is just the one-loop $SU(2)$ β-function. The non-perturbative term with coefficient c_n corresponds to an n-instanton contribution. (Since the model is Higgsed, the instantons have an effective IR cutoff at the scale U, so these instanton effects are calculable; the first two coefficients have been calculated.)

As we make a large circle in the U-plane, the effective coupling shifts, $\tau \to \tau-2$, corresponding to an unobservable ϑ-angle shift $\vartheta \to \vartheta-4\pi$. Note that there is a global discrete symmetry of this model which acts on the U-plane as

$$\mathbb{Z}_2 : \qquad U \to -U, \tag{26.6}$$

and so takes $\tau \to \tau-1$ (a 2π shift in the ϑ-angle). This $\mathbb{Z}_2$ is part of the anomaly-free $\mathbb{Z}_4$ subgroup of the anomalous $U(1)_A$.

So far we have been doing the "standard" analysis of the low-energy effective action for this theory. But there are two puzzles which indicate that we are missing some basic physics:

(1.) The effective coupling $\tau(U)$ is holomorphic, implying that $\mathrm{Re}\tau$ and $\mathrm{Im}\tau$ are harmonic functions on the U-plane. Since they are not constant functions, they therefore must be unbounded both above and below. In particular this implies that $\mathrm{Im}\tau = \frac{1}{g^2}$ will be negative for some U, and the effective theory will be non-unitary!

(2.) If we add a tree-level mass $\mathcal{W}_{tree} = m\mathrm{tr}\Phi^2 = mU$, then, for $m \gg \Lambda$, Φ can be integrated out leaving a low-energy pure $SU(2)$ super-YM theory with scale $\widehat{\Lambda}^6 = m^2\Lambda^4$. This theory has a gap, confinement, and two vacua with gaugino condensates $\langle\lambda\lambda\rangle = \pm m\Lambda^2$. But, in our low-energy theory on the U-plane, there are no light charged degrees of freedom to Higgs the photon.

The remainder of this lecture presents the physical ingredients which resolve these puzzles. In the next lecture we return to this $SU(2)$ theory and solve for $\tau(U)$.

26.1. Monopoles

The first ingredient we need to be aware of is monopoles. A recent review is J. Harvey, `hep-th/9603086`.

First, following Dirac, we ask whether it is possible to add megnetic charges without disturbing the quantum consistency of the electromagnetic coupling. A static magnetic field

$$\vec{B} = \frac{Q_m \hat{r}}{4\pi r^2} \tag{26.7}$$

corresponds to a magnetic charge $\int_{S^2_\infty} \vec{B}\cdot d\vec{S} = Q_m$ at $r=0$. To couple a charged particle to a background field quantumly we need the vector potential A_μ. There is no solution for A_μ which is regular away from $r=0$; however we can write the solution as one which is regular in two "patches". Divide a two-sphere S^2 of fixed radius r into a northern half N

with $0 \leq \theta \leq \pi/2$, a southern half S with $\pi/2 \leq \theta \leq \pi$ and the overlap region which is the equator at $\theta = \pi/2$. The vector potential on the two halves is then taken to be

$$\vec{A}_N = \frac{Q_m}{4\pi r}\frac{(1-\cos\theta)}{\sin\theta}\hat{e}_\phi, \qquad \vec{A}_S = -\frac{Q_m}{4\pi r}\frac{(1+\cos\theta)}{\sin\theta}\hat{e}_\phi. \tag{26.8}$$

Note that on the two halves of the two-sphere the magnetic field $\vec{B} = \vec{\nabla} \times \vec{A}$ agrees with (26.7). This construction only makes sense if the difference between A_N and A_S on the overlap region is a gauge transformation. At $\theta = \pi/2$

$$\vec{A}_N - \vec{A}_S = \vec{\nabla}\chi, \qquad \chi = \frac{Q_m}{2\pi}\phi, \tag{26.9}$$

so that the difference is a gauge transformation; however, the gauge function χ is not continuous. In quantum mechanics, a gauge transformation acts on wave functions carrying (electric) charge Q_e as $\psi \to e^{-iQ_e\chi}\psi$ so physical quantities will be continuous as long as $e^{-iQ_e\chi}$ is continuous. This then gives us the condition $e^{-iQ_eQ_m} = 1$ or

$$Q_eQ_m = 2\pi n, \qquad n \in \mathbb{Z} \tag{26.10}$$

which is the famous Dirac quantization condition.

Monopoles can be constructed as finite-energy classical solutions of non-Abelian gauge theories spontaneously broken down to Abelian factors (G. 't Hooft, *Nucl. Phys.* B79 (1974) 276; A. Polyakov, *JETP Lett.* 20 (1974) 194). We illustrate this in an $SU(2)$ theory broken down to $U(1)$ by an adjoint Higgs:

$$\mathcal{L} = -\frac{1}{4g^2}F^a_{\mu\nu}F^{a\mu\nu} + \frac{1}{2}D^\mu\Phi^a D_\mu\Phi^a - V(\Phi) \tag{26.11}$$

where V has a minimum on the sphere in field space $\sum_a \Phi^a\Phi^a = v^2$. Different directions on this sphere are gauge-equivalent. In the vacuum $\langle\Phi^a\rangle$ lies on this sphere, Higgsing $SU(2) \to U(1)$ and giving a mass $m_W = gv$ to the $W^\pm$ gauge bosons. The unbroken $U(1)$ has coupling g, so satisfies Gauss's law $\vec{D}\cdot\vec{E} = g^2 j^0_e$, where j^μ_e is the electric current density. Thus the electric charge is computed as $Q_e = \frac{1}{g^2}\int_{S^2_\infty} \varepsilon E \cdot d\vec{S}$. In the vacuum, the unbroken $U(1)$ is picked out by the direction of the Higgs vev, so $\vec{E} = \frac{1}{v}\Phi^a\vec{E}^a$. With this normalization of the electic charge, we find that the $W^\pm$ bosons have $Q_e = \pm 1$.

Static, finite-energy configurations must approach the vacuum at spatial infinity. Thus for a finite energy configuration the Higgs field Φ^a, evaluated as $r \to \infty$, provides a map from the S^2 at spatial infinity into the S^2 of the Higgs vacuum. Such maps are characterized by an integer, n_m, which measures the winding of one S^2 around the other. Mathematically, the second homotopy group of S^2 is the integers, $\pi_2(S^2) = \mathbb{Z}$. The winding, n_m, is the magnetic charge of the field configuration. To see this, the total energy from the Higgs field configuration:

$$\text{Energy} = \int d^3x \tfrac{1}{2} D_\mu \Phi^a D^\mu \Phi^a + V(\Phi) \geq \int d^3x \tfrac{1}{2} D_\mu \Phi^a D^\mu \Phi^a. \tag{26.12}$$

To have finite energy configurations we must therefore ensure that the covariant derivative of Φ^a falls off faster than $1/r$ at infinity. The general solution for the gauge field consistent with this behavior is

$$A^a_\mu \sim -\frac{1}{v^2}\epsilon^{abc}\Phi^b\partial_\mu\Phi^c + \frac{1}{v}\Phi^a A_\mu \tag{26.13}$$

with A_μ arbitrary. The leading-order behavior of the field strength is then

$$F^{a\mu\nu} = \frac{1}{v}\Phi^a F^{\mu\nu} \tag{26.14}$$

with

$$F^{\mu\nu} = -\frac{1}{v^3}\epsilon^{abc}\Phi^a\partial^\mu\Phi^b\partial^\nu\Phi^c + \partial^\mu A^\nu - \partial^\nu A^\mu \tag{26.15}$$

and the equations of motion imply $\partial_\mu F^{\mu\nu} = \partial_\mu * F^{\mu\nu} = 0$. Thus we learn that outside the core of the monopole the non-Abelian gauge field is purely in the direction of Φ^a, that is the direction of the unbroken $U(1)$. The magnetic charge of this field configuration is then computed to be

$$Q_m = \int_{S^2_\infty} \vec{B}\cdot d\vec{S} = \frac{1}{2v^3}\int_{S^2_\infty} \epsilon^{ijk}\epsilon^{abc}\Phi^a\partial^j\Phi^b\partial^k\Phi^c = 4\pi n_m \tag{26.16}$$

where n_m is the winding number of the Higgs field configuration, recovering the Dirac quantization condition.[35]

Note that for such non-singular field configurations, the electric and magnetic charges can be rewritten as

$$\begin{aligned} Q_e &= \frac{1}{g^2}\int_{S^2_\infty}\vec{E}\cdot d\vec{S} = \frac{1}{g^2 v}\int_{S^2_\infty}\Phi^a\vec{E}^a\cdot d\vec{S} = \frac{1}{g^2 v}\int d^3x\vec{E}^a\cdot(\vec{D}\Phi)^a \\ Q_m &= \int_{S^2_\infty}\vec{B}\cdot d\vec{S} = \frac{1}{v}\int_{S^2_\infty}\Phi^a\vec{B}^a\cdot d\vec{S} = \frac{1}{v}\int d^3x\vec{B}^a\cdot(\vec{D}\Phi)^a \end{aligned} \tag{26.17}$$

using the vacuum equation of motion and the Bianchi identity $\vec{D}\cdot\vec{E}^a = \vec{D}\cdot\vec{B}^a 0$ and integration by parts.

---

[35] The reason this is the Dirac quantization condition (26.10) only for even values of n is that in this theory we could add fields in the fundamental **2** representation of $SU(2)$, which would carry electric charge $Q_e = \pm 1/2$.

If we consider a static configuration with vanishing electric field the energy (mass) of the configuration is given by

$$
\begin{aligned}
m_M &= \int d^3x \left(\frac{1}{2g^2} \vec{B}^a \cdot \vec{B}^a + \frac{1}{2} \vec{D}\Phi^a \cdot \vec{D}\Phi^a + V(\Phi) \right) \geq \int d^3x \left(\frac{1}{2g^2} \vec{B}^a \cdot \vec{B}^a + \frac{1}{2} \vec{D}\Phi^a \cdot \vec{D}\Phi^a \right) \\
&= \frac{1}{2} \int d^3x \left(\frac{1}{g} \vec{B}^a - \vec{D}\Phi^a \right) \cdot \left(\frac{1}{g} \vec{B}^a - \vec{D}\Phi^a \right) + \frac{vQ_m}{g},
\end{aligned}
\tag{26.18}
$$

giving the BPS bound

$$
m_M \geq \left| \frac{vQ_m}{g} \right| . \tag{26.19}
$$

This semi-classical bound can be extended to *dyons* (solitonic states carrying both electric and magnetic charges):

$$
m_D \geq gv \left| Q_e + i \frac{Q_m}{g^2} \right| . \tag{26.20}
$$

E. Witten, *Phys. Lett.* 86B (1979) 283, pointed out that the ϑ angle has a non-trivial effect in the presence of magnetic monopoles: it shifts the allowed values of electric charge in the monopole sector of the theory. To see this, consider gauge transformations, constant at infinity, which are rotations in the $U(1)$ subgroup of $SU(2)$ picked out by the Higgs vev. That is, rotations in $SU(2)$ about the axis $\hat{\Phi}^a = \Phi^a / |\Phi^a|$. The action of such an infinitesimal gauge transformation on the field is

$$
\delta A^a_\mu = \frac{1}{v} (D_\mu \Phi)^a \tag{26.21}
$$

with Φ the background monopole Higgs field. Let $\mathcal{N}$ denote the generator of this gauge transformation. Then if we rotate by 2π about the $\hat{\Phi}$ axis we must get the identity

$$
e^{2\pi i \mathcal{N}} = 1. \tag{26.22}
$$

Including the ϑ term, it is straightforward to compute $\mathcal{N}$ using the Noether method,

$$
\mathcal{N} = \frac{\partial \mathcal{L}}{\partial \partial_0 A^a_\mu} \delta A^a_\mu = Q_e - \frac{\vartheta Q_m}{8\pi^2}, \tag{26.23}
$$

where we have used the definitions (26.17) of the electric and magnetic charge operators. This result implies

$$
Q_e = n_e + n_m \frac{\vartheta}{2\pi} \tag{26.24}
$$

where n_e is an arbitrary integer and $n_m = Q_m / 4\pi$ determines the magnetic charge of the monopole. We will henceforth label dyons by the integers (n_e, n_m). Note that the BPS bound becomes

$$
M_D \geq gv \left| \left(n_e + n_m \frac{\vartheta}{2\pi} \right) + i n_m \frac{4\pi}{g^2} \right| = gv |n_e + \tau n_m| . \tag{26.25}
$$

In theories with extended supersymmetry, the (quantum-corrected) BPS bound can be computed exactly, and states saturating the bound can be identified; see Seiberg and Witten, `hep-th/9407087`.

Maxwell's vacuum equations are invariant under the duality transformation

$$\vec{E} \to \vec{B}, \qquad \vec{B} \to -\vec{E}. \tag{26.26}$$

If we write Maxwell's equations in covariant form

$$\partial_\mu F^{\mu\nu} = j_e^\nu, \qquad \partial_\mu {*F}^{\mu\nu} = 0, \tag{26.27}$$

the duality transformation takes the form $F_{\mu\nu} \to *F_{\mu\nu}$. The duality symmetry of the free Maxwell equations is broken by the presence of electric source terms. For this reason it is of no practical interest in everyday applications of electromagnetism. However, if we include magnetic source terms so that $\partial_\mu * F^{\mu\nu} = j_m^\nu$ with j_m^ν the magnetic four-current, we make Maxwell's equations symmetric under the duality transformation and simultaneous interchange of electric and magnetic currents.

This duality of the classical equations of motion can be shown to hold quantumly as well. To see this, consider a $U(1)$ theory with coupling

$$\tau = \frac{\vartheta}{2\pi} + i\frac{4\pi}{g^2}. \tag{26.28}$$

Then the action can be written

$$S = \int d^4x \, \frac{\tau}{32\pi i}(F + i{*F})^2 + h.c. \tag{26.29}$$

In the quantum theory, we compute physical quantities in this theory by taking a path integral over all gauge potential configurations $\int \mathcal{D}A_\mu e^{iS}$. This can be rewritten as a path integral over field strength configurations as long as we insert the Bianchi identity as a constraint: $\int \mathcal{D}F_{\mu\nu}\mathcal{D}V_\rho e^{iS'}$, where

$$S' = S + \frac{1}{8\pi}\int d^4x \, V_\mu \epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma}. \tag{26.30}$$

Here V_μ is a Lagrange multiplier enforcing the Bianchi identity. In the presence of a monopole, $\epsilon^{0\mu\nu\rho}\partial_\mu F_{\nu\rho} = 8\pi\delta^{(3)}(x)$; the normalization of V_μ in (26.30) is chosen so that V_μ couples to a monopole with charge one. This Lagrange multiplier term can be rewritten

$$\frac{1}{8\pi}\int V_\mu \epsilon^{\mu\nu\rho\sigma}\partial_\mu F_{\rho\sigma} = \frac{1}{8\pi}\int {*\widetilde{F}}F = \frac{1}{16\pi}\int *(\widetilde{F} + i{*\widetilde{F}})(F + i{*F}) + h.c., \tag{26.31}$$

where $\widetilde{F}_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$ is the field strength of V. We can now perform the Gaussian functional integral over F and find an equivalent action, $\widetilde{S}$, for V:

$$\widetilde{S} = \int d^4x \, \frac{1}{32\pi i}\left(\frac{-1}{\tau}\right)(\widetilde{F} + i{*\widetilde{F}})^2. \tag{26.32}$$

We can just as easily perform these steps in the supersymmetric theory. Treating the field strength χsf W_α in $\int d^2\theta(\tau/32\pi i)\cdot(W^2)$ as an independent field in the path integral, the Bianchi identity, $D^\alpha W_\alpha = \overline{D}_{\dot\alpha}\overline{W}^{\dot\alpha}$, can be implemented by a real vsf V Lagrange multiplier. We add to the action

$$\frac{i}{16\pi}\int d^4x\, d^4\theta\, V D W = \frac{-i}{16\pi}\int d^4x\, d^2\theta\, DVW = \frac{1}{16\pi}\int d^4x\, d^2\theta\, \widetilde{W}W, \tag{26.33}$$

plus its hermitian conjugate. Performing the Gaussian integral over W gives an equivalent action

$$\widetilde{S} = \int d^4x\, d^2\theta\, \frac{1}{32\pi i}\left(\frac{-1}{\tau}\right)\widetilde{W}^2 + h.c. \tag{26.34}$$

Because we normalized the dual $U(1)$ potential so that a monopole couples to it with charge one, we see that we have iinterchanged what we mean by electric and magnetic charges. In general, under this transformation, a massive (n_e, n_m) dyonic source in the original description will couple to the dual $U(1)$ with charges $(n_m, -n_e)$. (The minus sign arises because the square of a duality transformation (26.26) is charge conjugation.) We have thus learned that the free $U(1)$ gauge theory with coupling τ is quantum equivalent to another such theory with coupling $-1/\tau$, and electric and magnetic charges of any massive sources interchanged. This is the electric-magnetic duality "symmetry". It is not really a symmetry since it acts on the coupling—it is an equivalence between two descriptions of the physics. The electric-magnetic duality transformation (S) and the invariance of the physics under 2π shifts of the ϑ-angle (T),

$$\begin{aligned} S: &\quad \tau \to \frac{-1}{\tau}, \qquad (n_e, n_m) \to (n_m, -n_e), \\ T: &\quad \tau \to \tau+1, \qquad (n_e, n_m) \to (n_e+n_m, n_m), \end{aligned} \tag{26.35}$$

generate the group $SL(2,\mathbb{Z})$ of duality transformations:

$$\tau \to \frac{a\tau+b}{c\tau+d}, \quad \begin{pmatrix} n_e \\ n_m \end{pmatrix} \to \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} n_e \\ n_m \end{pmatrix}, \qquad a,b,c,d \in \mathbb{Z}, \quad ad-bc=1. \tag{26.36}$$

The T transformation on the charges follows form the Witten effect (26.24).

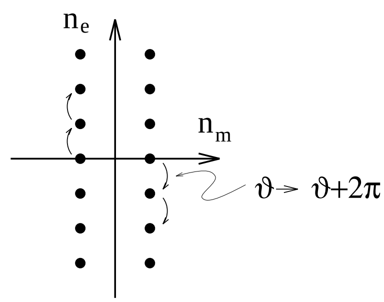
27. Theories with low-energy photons: dual Higgs mechanism and confinement

We now return to the supersymmetric $SU(2)$ gauge theory with an adjoint χsf Φ. Recall that the classical moduli space was the complex U-plane, where $U = \mathrm{tr}\Phi^2$, on which there is an unbroken $U(1)$ gauge group with field-strength χsf W_α. There is a global discrete $\mathbb{Z}_2$ symmetry $U \to -U$ on the moduli space. The low-energy effective coupling for the $U(1)$ has an expansion

$$\tau(U) = \frac{1}{2\pi i} \log\left(\frac{\Lambda^4}{U^2}\right) + \sum_{n=0}^{\infty} c_n \left(\frac{\Lambda^4}{U^2}\right)^n . \tag{27.1}$$

One of the puzzles we had with this theory was that if we turned on a large bare mass for Φ, the resulting theory should have a gap, and thus the low-energy $U(1)$ should be Higgsed. But there are no light charged χsf's in the spectrum, at least out at weak coupling $U \gg \Lambda$, to do the job.

We also learned last lecture that this theory can have magnetic monopoles. Indeed, one can show that there are semi-classically stable solitons with charges $(n_e, n_m) = (0, \pm 1)$ in this theory, and they turn out to lie in chiral multiplets of the supersymmetry algebra. Furthermore, from (27.1) we see that changing the phase of U shifts the effective ϑ-angle. In particular under the global $\mathbb{Z}_2$: $U \to e^{i\pi} U$, $\tau \to \tau - 1$. From the associated duality transformation on the charges of any massive states (*i.e.* the Witten effect), we see that there will be $(\mp 1, \pm 1)$ dyons in the spectrum. Repeating this procedure, we find there must be a whole tower of semi-classically stable dyons of charges $(n, \pm 1)$ for arbitrary integers n.



The existence of these dyon states suggests a possible resolution to one of our puzzles: perhaps at some strong coupling point on the moduli space, *e.g.* $U = U_0$ with $U_0 \sim \Lambda^2$, one of these dyons becomes massless, thereby providing the light charged χsf needed to Higgs the $U(1)$.

27.1. Physics near U_0

Making this assumption, let us check that it gives rise to the desired physics. We add to our low-energy effective theory two χsf's M and $\widetilde{M}$ oppositely charged under the $U(1)$. (We need two for anomaly-cancellation in the $U(1)$.) Since we are supposing that these states become massless at $U = U_0$, we can expand the effective potential around this point as

$$\mathcal{W} = (U - U_0) M \widetilde{M} + \mathcal{O}((U - U_0)^2). \tag{27.2}$$

The D-term equations from the coupling to the $U(1)$ gauge field imply

$$|M| = |\widetilde{M}|, \tag{27.3}$$

while the F-term equations from (27.2) are

$$\begin{aligned} 0 &= \frac{\partial \mathcal{W}}{\partial U} = M\widetilde{M}, \\ 0 &= \frac{\partial \mathcal{W}}{\partial M} = (U - U_0)\widetilde{M}. \end{aligned} \tag{27.4}$$

The solutions are $M = \widetilde{M} = 0$ with U arbitrary, which is just the U-plane Coulomb branch.

Now we add a bare mass term for the adjoint Φ, and see if we lift the flat directions and obtain a discrete set of gapped vacua. The bare mass term is $\mathcal{W}_{tree} = m\mathrm{tr}\Phi^2 = mU$. By the selection rule for the anomalous $U(1)_A$ under which m has charge -1 and the $U(1)_R$ under which m is assigned charge 2, and the usual non-renormalization argument, the low-energy effective superpotential must be of the form

$$\mathcal{W} = (U - U_0)M\widetilde{M} + mU + \mathcal{O}((U - U_0)^2). \tag{27.5}$$

The D- and F-term equations become

$$\begin{aligned} 0 &= |M| - |\widetilde{M}|, \\ 0 &= M\widetilde{M} + m, \\ 0 &= (U - U_0)\widetilde{M}, \end{aligned} \tag{27.6}$$

whose solutions are $|M| = |\widetilde{M}| = m^{1/2}$ and $U = U_0$. Thus the Coulomb branch is indeed lifted, and there is only a single vacuum. This vacuum has a gap, since the charged χsf's M and $\widetilde{M}$ get non-zero vevs, thereby Higgsing the $U(1)$.

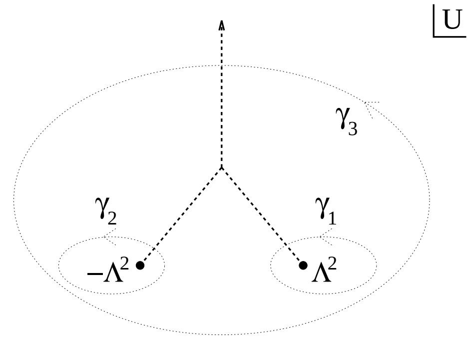
In this analysis, we have implicitly assumed (in writing down the D terms) that M and $\widetilde{M}$ were electrically charged with respect to the $U(1)$ field. But, by electric-magnetic duality, our analysis is valid for any dyonic charges. This is because M and $\widetilde{M}$ are the only light charged fields in the theory near U_0, so we can by an electric-magnetic duality transformation rotate any (n_e, n_m) to a description in which they are proportional to $(1, 0)$.[36] Then in this description the above analysis is valid.

Now, for $m \gg \Lambda$ we expect to recover the two gapped vacua of the pure $SU(2)$ super-YM theory. Recalling the $\mathbb{Z}_2$ symmetry of the theory, it is natural to assume that there are two points on the U-plane where charged χsf's become massless in the $m = 0$ theory, and they are at $U = \pm U_0$. Since Λ is the only scale in the theory, we take $U_0 = \Lambda^2$. (We can take this as the definition of Λ, if we like.)

[36] More precisely, there is an $SL(2, \mathbb{Z})$ transformation which takes them to $(q, 0)$ where q is the greatest common divisor of n_e and n_m.

Can this assumption be checked? Yes, by examining the behavior of τ as a function of U. Recall the other puzzle we had about the physics on the Coulomb branch: since $\tau(U)$ is holomorphic, $1/g^2 \sim \mathrm{Im}\tau$ is harmonic and therefore unbounded from below, violating unitarity.

This puzzle is resolved by noting that τ is not, in fact, a holomorphic function of U. In particular, by electric-magnetic duality, as we traverse closed loops in the U-plane, τ need not come back to the same value, only one related to it by an $SL(2,\mathbb{Z})$ transformation. Mathematically, this is described by saying that τ is a section of a flat $SL(2,\mathbb{Z})$ bundle. This multi-valuedness of τ can be described by saying that τ is a holomorphic function on a cut U-plane, with cuts emanating from some singularities, and with the jump in τ across the cuts being an element of $SL(2,\mathbb{Z})$. The two points $U = \pm\Lambda^2$ at which we are assuming there are massless charged χsf's are the natural candidates for the branch points:



Here I have placed the cuts in an arbitrary manner connecting the two possible strong-coupling singularities, and a possible singularity at weak coupling ($U = \infty$). The presence of these cuts allows us to avoid the conclusion that $\mathrm{Im}\tau$ is unbounded.

Upon traversing the various loops γ_i in the above figure, τ will change by the action of an $SL(2,\mathbb{Z})$ element. These elements are called the *monodromies* of τ, and will be denoted $\mathcal{M}_i$.

We first calculate $\mathcal{M}_3$, the monodromy around the weak-coupling singularity at infinity. By taking γ_3 of large enough radius, τ will be accurately given by its one-loop value, the first term in (27.1). Taking $U \to e^{2\pi i}U$ in this formula gives $\tau \to \tau - 2$, giving for the monodromy at infinity[37]

$$\mathcal{M}_3 = \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix}. \tag{27.7}$$

[37] This actually only determines the monodromy up to an overall sign. The sign is determined by noting that $U \to e^{2\pi i}U$ has the effect of $\Phi \to -\Phi$ on the elementary Higgs field, so it reverses the sign of the low-energy electromagnetic field which in terms of $SU(2)$ variables is proportional to $\mathrm{tr}(\Phi F)$. Thus it reverses the sign of electric and magnetic charges, giving an "extra" factor of $-\mathbb{1} \in SL(2,\mathbb{Z})$.

In order to calculate the $\mathcal{M}_{1,2}$ monodromies, let us first calculate the monodromy we would expect if the χsf becoming massless at the associated singularity had charge (n_e, n_m). By a duality transformation we can change to a basis where this charge is purely electric: $(\widetilde{n_e}, 0)$. In this basis the physics near the $U = U_0$ singularity is just that of QED with the electron becoming massless. This theory is IR free, so the behavior of the low-energy effective coupling will be dominated by its one-loop expression, at least sufficiently near U_0 where the mass of the charged χsf $\sim U - U_0$ is arbitrarily small:

$$\widetilde{\tau} = \frac{\widetilde{n_e}^2}{\pi i} \log(U - U_0) + \mathcal{O}(U - U_0)^0. \tag{27.8}$$

By traversing a small loop around U_0, $(U - U_0) \to e^{2\pi i}(U - U_0)$, we find the monodromy

$$\widetilde{\tau} \to \widetilde{\tau} + 2\widetilde{n_e}^2 \qquad \Longrightarrow \qquad \widetilde{\mathcal{M}} = \begin{pmatrix} 1 & 2\widetilde{n_e}^2 \\ 0 & 1 \end{pmatrix}. \tag{27.9}$$

Now let us duality-transform this answer back to the basis where the charges are (n_e, n_m). The required $SL(2, \mathbb{Z})$ element will be denoted $\mathcal{N} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and satisfies

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} n_e \\ n_m \end{pmatrix} = \begin{pmatrix} \widetilde{n_e} \\ 0 \end{pmatrix}, \quad \text{and} \quad ad - bc = 1 \quad \text{with} \quad a, b, c, d \in \mathbb{Z}. \tag{27.10}$$

The transformed monodromy is then

$$\mathcal{M} = \mathcal{N} \widetilde{\mathcal{M}} \mathcal{N}^{-1} = \begin{pmatrix} 1 + 2n_e n_m & 2n_e^2 \\ -2n_m^2 & 1 - 2n_e n_m \end{pmatrix}. \tag{27.11}$$

Now, by deforming the γ_i contours in the U-plane, we find that the three monodromies must be related by

$$\mathcal{M}_3 = \mathcal{M}_1 \mathcal{M}_2. \tag{27.12}$$

Assuming that a χsf with charges (n_{e1}, n_{m1}) becomes massless at $U = \Lambda^2$, while one with charges (n_{e2}, n_{m2}) does so at $U = -\Lambda^2$, and substituting into (27.12) using (27.7) and (27.11) gives as solutions

$$(n_{e1}, n_{m1}) = \pm(n, 1), \qquad (n_{e2}, n_{m2}) = \pm(n-1, 1), \qquad \text{for all} \quad n \in \mathbb{Z}. \tag{27.13}$$

This set of charges actually represents a single physical solution. This is because taking $U \to e^{i\pi} U$ takes us to an equivalent theory by the $\mathbb{Z}_2$ symmetry; but this corresponds to shifting the low-energy ϑ-angle by 2π which in turn shifts all dyon electric charges by their magnetic charges. Repeated applications of this shift can take any of the above solutions to the solution

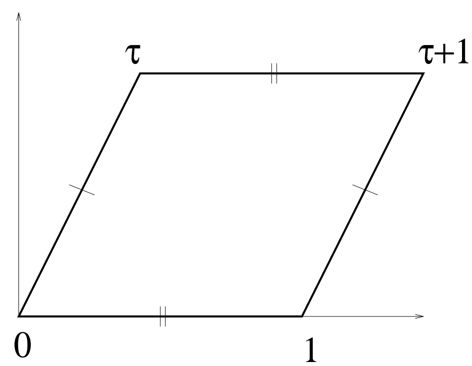
$$(n_{e1}, n_{m1}) = \pm(0, 1), \qquad (n_{e2}, n_{m2}) = \pm(-1, 1). \tag{27.14}$$

The plus and minus sign solutions must both be there by anomaly cancellation in the low-energy $U(1)$. We thus learn that there is a consistent solution with a monopole becoming massless at $U = \Lambda^2$ and a charge $(-1, 1)$ dyon becoming massless at $U = -\Lambda^2$. Some progress has been made in weakening the initial assumption that there are just two strong-coupling singularities, see R. Flume *et. al.*, `hep-th/9611123`.

27.3. $\tau(U)$

Now that we have the monodromies around the singularities, we now turn to finding the low-energy coupling τ on the U-plane. The basic idea is that τ is determined by holomorphy and demanding that it match onto the behavior we have deterined above at $U = \infty$ and $U = \pm\Lambda^2$. Seeing how to solve this "analytic continuation" problem analytically is not obvious, however. Seiberg and Witten did it by introducing an auxiliary mathematical object: a family of tori varying over the Coulomb branch.

This is a useful construction because the low-energy effective coupling τ has the same properties as the complex structure of a 2-torus. In particular, the complex structure of a torus can be described by its *modulus*, a complex number τ, with $\mathrm{Im}\tau > 0$. In this description, the torus can be thought of as a parallelogram in the complex plane with oposite sides identified:



Furthermore, the modulus τ of such a torus gives equivalent complex structures modulo $SL(2,\mathbb{Z})$ transformations acting on τ. Therefore, if we associate to each point in the U-plane a holomorphically-varying torus, its modulus will automatically be a holomorphic section of an $SL(2,\mathbb{Z})$ bundle with positive imaginary part, which are just the properties we want for the effective coupling τ.

At $U = \pm\Lambda^2$, magnetically charged states become massless, implying that the effective coupling $\mathrm{Im}\tau \to 0$. (Recall that by $U(1)$ IR freedom, when an electrically charged state becomes massless, the coupling $g \to 0$, implying $\tau \to +i\infty$. Doing the duality transform $\tau \to -1/\tau$ gives the above result for a magnetic charge becoming massless.) From the parallelogram, we see this implies that the torus is degenerating: one of its cycles is vanishing.

Now, a general torus can be described analytically as the Riemann surface which is the solution $y(x)$ to the complex cubic equation

$$y^2 = (x - e_1)(x - e_2)(x - e_3). \tag{27.15}$$

We can think of this as a double-sheeted cover of the x-plane, branched over the three points e_i and the point at infinity. We let this torus vary over the U-plane by letting the

e_i vary: $e_i = e_i(U, \Lambda)$. By choosing the cuts to run between pairs of these branch points, and "gluing" the two sheets together along these cuts, one sees that the Riemann surface is indeed topologically a torus. Furthermore, the condition for a nontrivial cycle on this torus to vanish is that two of the branch points collide. Since we want this to happen at the two points $U = \pm\Lambda^2$, it is natural to choose $e_1 = \Lambda^2$, $e_2 = -\Lambda^2$, and $e_3 = U$:

$$y^2 = (x - \Lambda^2)(x + \Lambda^2)(x - U). \tag{27.16}$$

Furthermore, note that this choice has a manifest $U \to -U$ symmetry, under which $x \to -x$ and $y \to \pm iy$.

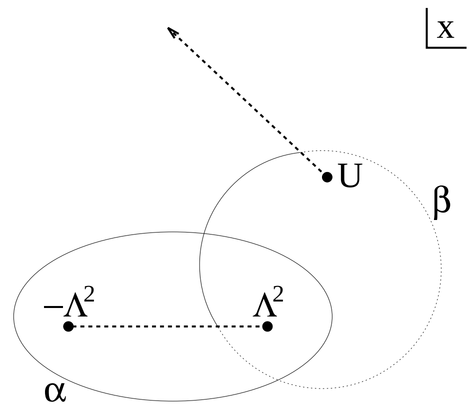
Given this family of tori, one can compute their moduli as a ratio of line integrals:

$$\tau(U) = \frac{\oint_\beta \omega}{\oint_\alpha \omega}, \tag{27.17}$$

where ω is the (unique) holomorphic one-form on the Riemann surface,

$$\omega = \frac{dx}{y} = \frac{dx}{\sqrt{(x^2 - \Lambda^4)(x - U)}}, \tag{27.18}$$

and α and β are any two non-trivial cycles on the torus which intersect once. For example, we might take α to be a cycle on the x-plane which loops around the branch points at $\pm\Lambda^2$, while β is the one which loops around the branch points at Λ^2 and U. If we chose the cuts on the x-plane to run between $\pm\Lambda^2$ and between U and ∞, then the α cycle would lie all on one sheet, while the β cycle would go onto the second sheet as it passes through the cut:

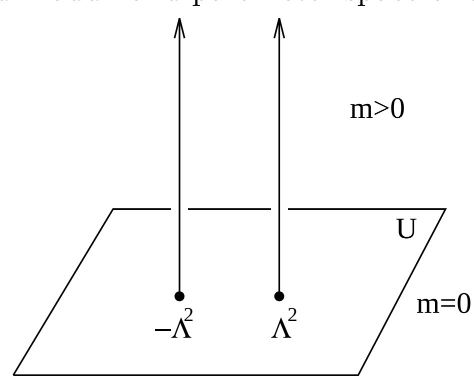


Since the integrand in (27.18) is a closed one form ($d\omega = 0$), the value of τ does not depend on the exact locations of α and β, but only on how they loop around the branch points.

We can now check that our family of tori (27.16) indeed give rise to the correct low-energy τ. By taking $U \to \infty$, it is not hard to explicitly evaluate (27.18) to find agreement with the first term in the weak-coupling expansion (27.1).[38] Also, without having to explicitly evaluate the integrals in (27.18), one can check that it reproduces the correct monodromies as U goes around the singularities at $\pm\Lambda^2$ by tracking how the α and β cycles are deformed as U varies. Finally, it turns out that the family of tori (27.16) is the unique one with these properties.

27.4. Dual Higgs mechanism and confinement

In summary, we have found the solution for $SU(2)$ with a massive adjoint in which, at zero mass, there is a complex U-plane of degenerate vacua in a Coulomb phase. The vacua at $U = \pm\Lambda^2$ are special since a monopole and dyon, respectively, becomes massless there. When we turn on a non-zero mass for the adjoint, all the vacua on the U-plane are lifted, except for the two massless points. At those points, the scalar monopole or dyon fields condense, Higgsing the (appropriate electric-magnetic dual) $U(1)$. This is illustrated in a picture of the combined moduli and parameter space of the model:



One puzzle that may remain concerning this solution is that for an adjoint mass $m \gg \Lambda$ we expected to find two confining vacua of the low-energy pure super-YM theory, yet we seem to have found instead two Higgs vacua. This is not quite right, though, since the Higgs mechanism taking place is not the usual condensation of an electrically charged scalar field, but of magnetically (and dyonically) charged scalars.

To see what this means, let us recall the basic physics of the Higgs mechanism. When an electric charge condenses, it screens any background electromagnetic fields, damping them exponentially—this is a consequence of the photon aquiring a non-zero mass. This means that electric sources in the theory are essentially free, for their electric fields can be "absorbed" by the electric condensate, and their interaction energy will drop off exponentially. Magnetic charges, on the other hand, behave very differently, because the magnetic field lines have no condensate source to end on. The result is that magnetic field lines tend to be excluded from the vacuum; this is called the Meissner effect in superconductors. The

[38] Though perhaps only up to an $SL(2,\mathbb{Z})$ transformation if I made the wrong choice for my α and β cycles.

minimum energy configuration is for the magnetic field to be confined to a thin flux tube connecting opposite magnetic charges, leading to confining forces between them. Thus, in the Higgs mechanism, electric charges are screened and magnetic charges are confined.

To see what happens when magnetic charges, instead of electric charges, condense, we simply do an electric-magnetic duality transformation. Thus in the dual Higgs effect, magnetic charges are screened, and electric charges are confined. So we have indeed found confinement in our $SU(2)$ solution at the monopole point. This is a concrete realization of a picture of confinement in non-Abelian gauge theories proposed in the '70's by S. Mandelstam and by G. 't Hooft.

Finally, at the dyonic point, by another duality transformation, it is not hard to see that both electric and magnetic charges are confined, though any dyonic charges proportional to $(-1, 1)$ will just be screened. This is a realization of an "oblique confinement" phase of non-Abelian gauge theories proposed by 't Hooft.