Physics 215C: Particles and Fields
Spring 2017

Lecturer: McGreevy
These lecture notes live here. Please email corrections to mcgreevy at physics dot ucsd dot edu.

Last updated: 2018/10/11, 17:37:18

Contents

0.1 Introductory remarks for the third quarter

Last quarter, we grappled with the Wilsonian perspective on the RG, which (among many other victories) provides an explanation of the *totalitarian principle of physics*, that anything that can happen must happen. More precisely, this means that the Hamiltonian should contain all terms consistent with symmetries, organized according to an expansion in decreasing relevance to low energy physics.

This leads directly to the idea of *effective field theory*, or, *how to do physics without a theory of everything*. (You may notice that all the physics that has been done has been done without a theory of everything.) It is a weaponized version of selective inattention.

So here are some goals, both practical and philosophical:

- We'll continue our study of coarse-graining in quantum systems with extensive degrees of freedom, aka the RG in QFT.

  I remind you that by 'extensive degrees of freedom' I mean that we are going to study models which, if we like, we can sprinkle over vast tracts of land, like *sod* (depicted in the figure at right). And also like sod, each little patch of degrees of freedom only interacts with its neighboring patches: this property of sod and of QFT is called *locality*.

  And also like sod, each little patch of degrees of freedom only interacts with its neighboring patches: this property of sod and of QFT is called *locality*. More precisely, in a quantum mechanical system, we specify the degrees of freedom by their Hilbert space; by an extensive system, I'll mean one in which the Hilbert space is of the form $\mathcal{H} = \otimes_{\text{patches of space}} \mathcal{H}_{\text{patch}}$ and the interactions are local $\mathbf{H} = \sum_{\text{patches}} \mathbf{H}(\text{nearby patches})$.

  By 'coarse-graining' I mean ignoring things we don't care about, or rather only paying attention to them to the extent that they affect the things we do care about.

  To continue the sod example in 2+1 dimensions, a person laying the sod in the picture above cares that the sod doesn't fall apart, and rolls nicely onto the ground (as long as we don't do high-energy probes like bending it violently or trying to lay it down too quickly). These *long-wavelength* properties of *rigidity* and *elasticity* are collective, emergent properties of the microscopic constituents (sod molecules) – we can describe the dynamics involved in covering the Earth with sod (never mind whether this is a good idea in a desert climate) without knowing the microscopic theory of the sod molecules ('grass'). Our job is to think about the relationship between the microscopic model (grassodynamics) and its macroscopic counterpart (in this case, suburban landscaping). In my experience, learning to do this is approximately synonymous with *understanding*.

- I would like to convince you that "non-renormalizable" does not mean "not worth your attention," and explain the incredibly useful notion of an Effective Field Theory.

- There is more to QFT than perturbation theory about free fields in a Fock vacuum. In particular, we will spend some time thinking about non-perturbative physics, effects of topology, solitons. Topology is one tool for making precise statements without perturbation theory (the basic idea: if we know something is an integer, it is easy to get many digits of precision!).

- I will try to resist making too many comments on the particle-physics-centric nature of the QFT curriculum, since the curriculum this year has been largely up to me. (In previous years when I've taught 215C, other folks taught 215A and 215B, but this time I have no one but myself to blame for misinforming you about how to think about quantum fields.) But I want to emphasize that QFT is also quite central in many aspects of condensed matter physics, and we will learn about this. From the point of view of someone interested in QFT, high energy particle physics has the severe drawback that it offers only one example! (OK, for some purposes we can think about QCD and the electroweak theory separately...)

  From the high-energy physics point of view, we could call this the study of *regulated* QFT, with a particular kind of lattice regulator. Why make a big deal about 'regulated'? Besides the fact that this is how QFT comes to us (when it does) in condensed matter physics, such a description is required if we want to know what we're talking about. For example, we need it if we want to know what we're talking about well enough to explain it to a computer. Many QFT problems are too hard for our brains. A related but less precise point is that I would like to do what I can to erase the problematic perspective on QFT which 'begins from a classical lagrangian and quantizes it' etc, and leads to a term like 'anomaly'. (We will talk about what is 'anomaly' this quarter.)

- There is more to QFT than the S-matrix. In a particle-physics QFT course (like this year's 215A!) you learn that the purpose in life of correlation functions or green's functions or off-shell amplitudes is that they have *poles* (at $p^\mu p_\mu - m^2 = 0$)

whose residues are the S-matrix elements, which are what you measure (or better, are the distribution you sample) when you scatter the particles which are the quanta of the fields of the QFT. I want to make two extended points about this:

1. In many physical contexts where QFT is relevant, you can actually measure the off-shell stuff. This is yet another reason why including condensed matter in our field of view will deepen our understanding of QFT.
2. This is good, because the Green's functions don't always have simple poles! There are lots of interesting field theories where the Green's functions instead have power-law singularities, like $G(p) \sim \frac{1}{p^{2\Delta}}$. If you Fourier transform this, you don't get an exponentially-localized packet. The elementary excitations created by a field whose two point function does this *are not particles*. (Any conformal field theory (CFT) is an example of this.) The theory of particles (and their dance of creation and annihilation and so on) is an important but proper subset of QFT.

- The crux of many problems in physics is the correct choice of variables with which to label the degrees of freedom. Often the best choice is very different from the obvious choice; a name for this phenomenon is 'duality'. We will study many examples of it (Kramers-Wannier, Jordan-Wigner, bosonization, Wegner, particle-vortex, perhaps others). This word is dangerous (at one point it was one of the forbidden words on my blackboard) because it is *about* ambiguities in our (physics) language. I would like to reclaim it.

  An important bias in deciding what is meant by 'correct' or 'best' in the previous paragraph is: we will be interested in low-energy and long-wavelength physics, near the groundstate. For one thing, this is the aspect of the present subject which is like 'elementary particle physics'; the high-energy physics of these systems is of a very different nature and bears little resemblance to the field often called 'high-energy physics' (for example, there is volume-law entanglement).

- We'll be interested in models with a finite number of degrees of freedom *per unit volume*. This last is important, because we are going to be interested in the thermodynamic limit. Questions about a finite amount of stuff (this is sometimes called 'mesoscopics') tend to be much harder.

- An important goal for the course is demonstrating that many fancy phenomena precious to particle physicists can emerge from very humble origins in the kinds of (completely well-defined) local quantum lattice models we will study. Here I have in mind: fermions, gauge theory, photons, anyons, strings, topological solitons, CFT, and many other sources of wonder I'm forgetting right now.

Here is a confession, related to several of the points above: The following comment in the book *Advanced Quantum Mechanics* by Sakurai had a big effect on my education in physics: *... we see a number of sophisticated, yet uneducated, theoreticians who are conversant in the LSZ formalism of the Heisenberg field operators, but do not know why an excited atom radiates, or are ignorant of the quantum-theoretic derivation of Rayleigh's law that accounts for the blueness of the sky.* I read this comment during my first year of graduate school and it could not have applied more aptly to me. I have been trying to correct the defects in my own education which this exemplifies ever since. I bet most of you know more about the color of the sky than I did when I was your age, but we will come back to this question. (If necessary, we will also come back to the radiation from excited atoms.)

So I intend that there will be two themes of this course: coarse-graining and topology. Both of these concepts are important in both hep-th and in cond-mat. Topics which I hope to discuss include:

- the uses and limitations of path integrals of various kinds
- some illustrations of effective field theory (perhaps cleverly mixed in with the other subjects)
- effects of topology in QFT (this includes anomalies, topological solitons and defects, topological terms in the action)
- more deep mysteries of gauge theory and its emergence in physical systems.
- If there is demand for it, we will discuss non-abelian gauge theory, in perturbation theory: Fadeev-Popov ghosts, and the sign of the Yang-Mills beta function. Similarly, we can talk about other topics relevant to the Standard Model of particle physics if there is demand.
- Large-N expansions?
- duality.

I welcome your suggestions regarding which subjects in QFT we should study.

0.2 Sources

The material in these notes is collected from many places, among which I should mention in particular the following:

Peskin and Schroeder, *An introduction to quantum field theory* (Wiley)

Zee, *Quantum Field Theory* (Princeton, 2d Edition)

Banks, *Modern Quantum Field Theory: A Concise Introduction* (Cambridge)

Schwartz, *Quantum field theory and the standard model* (Cambridge)

Coleman, *Aspects of Symmetry* (Cambridge)

Polyakov, *Gauge Field and Strings* (Harwood)

Wen, *Quantum field theory of many-body systems* (Oxford)

Sachdev, *Quantum Phase Transitions* (Cambridge, 2d Edition)

Many other bits of wisdom come from the Berkeley QFT courses of Prof. L. Hall and Prof. M. Halpern.

0.3 Conventions

Following most QFT books, I am going to use the $+ - --$ signature convention for the Minkowski metric. I am used to the other convention, where time is the weird one, so I'll need your help checking my signs. More explicitly, denoting a small spacetime displacement as $dx^\mu \equiv (dt, d\vec{x})^\mu$, the Lorentz-invariant distance is:

$$ds^2 = +dt^2 - d\vec{x} \cdot d\vec{x} = \eta_{\mu\nu} dx^\mu dx^\nu \quad \text{with} \quad \eta^{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}_{\mu\nu} .$$

(spacelike is negative). We will also write $\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(\partial_t, \vec{\nabla}_x\right)^\mu$, and $\partial^\mu \equiv \eta^{\mu\nu}\partial_\nu$. I'll use $\mu, \nu ...$ for Lorentz indices, and $i, k, ...$ for spatial indices.

The convention that repeated indices are summed is always in effect unless otherwise indicated.

A consequence of the fact that english and math are written from left to right is that time goes to the left.

A useful generalization of the shorthand $\hbar \equiv \frac{h}{2\pi}$ is

$$\text{đ}k \equiv \frac{\mathrm{d}k}{2\pi}.$$

I will also write $\not{\delta}^d(q) \equiv (2\pi)^d \delta^{(d)}(q)$. I will try to be consistent about writing Fourier transforms as

$$\int \frac{\mathrm{d}^d k}{(2\pi)^d} e^{ikx} \tilde{f}(k) \equiv \int \text{đ}^d k \; e^{ikx} \tilde{f}(k) \equiv f(x).$$

IFF $\equiv$ if and only if.
RHS $\equiv$ right-hand side. LHS $\equiv$ left-hand side. BHS $\equiv$ both-hand side.
IBP $\equiv$ integration by parts. WLOG $\equiv$ without loss of generality.
$+\mathcal{O}(x^n) \equiv$ plus terms which go like x^n (and higher powers) when x is small.
$+h.c. \equiv$ plus hermitian conjugate.

We work in units where $\hbar$ and the speed of light, c, are equal to one unless otherwise noted. When I say 'Peskin' I usually mean 'Peskin & Schroeder'.

Please tell me if you find typos or errors or violations of the rules above.

11 Resolving the identity

The following is an advertisement: When studying a quantum mechanical system, isn't it annoying to have to worry about the order in which you write the symbols? What if they don't commute?! If you have this problem, too, the path integral is for you. In the path integral, the symbols are just integration variables – just ordinary numbers, and you can write them in whatever order you want. You can write them upside down if you want. You can even change variables in the integral (Jacobian not included).

(What order do the operators end up in? As we showed last quarter, in the kinds of path integrals we're thinking about, they end up in *time-order*. If you want a different order, you will want to use the *Schwinger-Keldysh* extension package, sold separately.)

This section is about how to go back and forth from Hilbert space to path integral representations, aka Hamiltonian and Lagrangian descriptions of QFT. You make a path integral representation of some physical quantity by sticking lots of $\mathbb{1}$s in there, and then resolving each of the identity operators in some basis that you like. Different bases, different integrals. Some are useful, mostly because we have intuition for the behavior of integrals.

11.1 Quantum-classical correspondence

[Kogut, Sachdev chapter 5, Goldenfeld §3.2]

Let me say a few introductory words about quantum spin systems, the flagship family of examples of well-regulated QFTs. Such a thing is a collection of two-state systems (aka qbits) $\mathcal{H}_j = \text{span}\{|\uparrow_j\rangle, |\downarrow_j\rangle\}$ distributed over space and coupled somehow:

$$\mathcal{H} = \bigotimes_j \mathcal{H}_j \ , \quad \dim(\mathcal{H}) = 2^N$$

where N is the number of sites.
One qbit: To begin, consider just one two-state system. There are four independent hermitian operators acting on this Hilbert space. Besides the identity, there are the three Paulis, which I will denote by $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ instead of $\boldsymbol{\sigma}^x, \boldsymbol{\sigma}^y, \boldsymbol{\sigma}^z$:

$$\mathbf{X} \equiv \boldsymbol{\sigma}^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{Y} \equiv \boldsymbol{\sigma}^y = \begin{pmatrix} 0 & -\mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix}, \quad \mathbf{Z} \equiv \boldsymbol{\sigma}^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

This notation (which comes to us from the quantum information community) makes the important information larger and is therefore better, especially for those of us with limited eyesight.

They satisfy

$$\mathbf{XY} = \mathbf{iZ}, \quad \mathbf{XZ} = -\mathbf{ZX}, \quad \mathbf{X}^2 = \mathbb{1},$$

and all cyclic permutations $\mathbf{X} \to \mathbf{Y} \to \mathbf{Z} \to \mathbf{X}$ of these statements.
Multiple qbits: If we have more than one site, the paulis on different sites *commute*:

$$[\boldsymbol{\sigma}_j, \boldsymbol{\sigma}_l] = 0, \quad j \neq l \qquad i.e. \quad \mathbf{X}_j \mathbf{Z}_l = (-1)^{\delta_{jl}} \mathbf{Z}_l \mathbf{X}_j,$$

where $\boldsymbol{\sigma}_j$ is any of the three paulis acting on $\mathcal{H}_j$.

In this section we're going to study the 'path integral' associated with the $\mathbf{Z}$-basis resolution, $\mathbb{1} = |+\rangle\langle+| + |-\rangle\langle-|$. The *labels* on the states are classical spins ± 1 (or equivalently, classical bits). I put 'path integral' in quotes because it is instead a 'path sum', since the integration variables are discrete. This discussion will allow us to further harness our knowledge of stat mech for QFT purposes. An important conclusion at which we will arrive is the (inverse) relationship between the correlation length and the energy gap above the groundstate.

One qbit from classical Ising chain. Let's begin with the classical ising model in a (longitudinal) magnetic field:

$$Z = \sum_{\{s_j\}} e^{-K \sum_{\langle jl \rangle} s_j s_l - h \sum_j s_j}. \tag{11.1}$$

Here I am imagining we have classical spins $s_j = \pm 1$ at each site of some graph, and $\langle jl \rangle$ denotes pairs of sites which share a link in the graph. You might be tempted to call K the inverse temperature, which is how we would interpret if we were doing classical stat mech; resist the temptation.

First, let's think about the case when the graph in (11.1) is just a chain:

$$Z_1 = \sum_{\{s_l = \pm 1\}} e^{-S}, \qquad S = -K \sum_{l=1}^{M_\tau} s_l s_{l+1} - h \sum_{l=1}^{M_\tau} s_l \tag{11.2}$$

These ss are now just M_τ numbers, each ± 1 – there are 2^{M_τ} terms in this sum. (Notice that the field h breaks the $s \to -s$ symmetry of the summand.) The parameter $K > 0$ is the 'inverse temperature' in the Boltzmann distribution; I put these words in quotes because I want you to think of it as merely a parameter in the classical hamiltonian.

For definiteness let's suppose the chain loops back on itself,

$$s_{l+M_\tau} = s_l \quad \text{(periodic boundary conditions)}.$$

Using the identity $e^{\sum_l (\ldots)_l} = \prod_l e^{(\ldots)_l}$,

$$Z_1 = \sum_{\{s_l\}} \prod_{l=1}^{M_\tau} T_1(s_l, s_{l+1}) T_2(s_l)$$

where

$$T_1(s_1, s_2) \equiv e^{K s_1 s_2}, \quad T_2(s) \equiv e^{hs} \ .$$

What are these objects? The conceptual leap is to think of $T_1(s_1, s_2)$ as a 2×2 matrix:

$$T_1(s_1, s_2) = \begin{pmatrix} e^K & e^{-K} \\ e^{-K} & e^K \end{pmatrix}_{s_1 s_2} = \langle s_1 | \, \mathbf{T}_1 \, | s_2 \rangle \,,$$

which we can then regard as matrix elements of an operator $\mathbf{T}_1$ acting on a 2-state *quantum* system (hence the boldface). And we have to think of $T_2(s)$ as the *diagonal* elements of the same kind of matrix:

$$\delta_{s_1, s_2} T_2(s_1) = \begin{pmatrix} e^h & 0 \\ 0 & e^{-h} \end{pmatrix}_{s_1 s_2} = \langle s_1 | \, \mathbf{T}_2 \, | s_2 \rangle \,.$$

So we have

$$Z_1 = \mathrm{tr} \left(\underbrace{(\mathbf{T}_1 \mathbf{T}_2)(\mathbf{T}_1 \mathbf{T}_2) \cdots (\mathbf{T}_1 \mathbf{T}_2)}_{M_\tau \ \text{times}} \right) = \mathrm{tr} \mathbf{T}^{M_\tau} \tag{11.3}$$

where I've written

$$\mathbf{T} \equiv \mathbf{T}_2^{\frac{1}{2}} \mathbf{T}_1 \mathbf{T}_2^{\frac{1}{2}} = \mathbf{T}^\dagger = \mathbf{T}^t$$

for convenience (so it's symmetric). This object is the *transfer matrix*. What's the trace over in (11.3)? It's a single two-state system – a single qbit (or quantum spin) that we've constructed from this chain of classical two-valued variables.

Even if we didn't care about quantum spins, this way of organizing the partition sum of the Ising chain does the sum for us (since the trace is basis-independent, and so we might as well evaluate it in the basis where $\mathbf{T}$ is diagonal):

$$Z_1 = \mathrm{tr} \mathbf{T}^{M_\tau} = \lambda_+^{M_\tau} + \lambda_-^{M_\tau}$$

where $\lambda_\pm$ are the two eigenvalues of the transfer matrix, $\lambda_+ \geq \lambda_-$:

$$\lambda_\pm = e^K \cosh h \pm \sqrt{e^{2K} \sinh^2 h + e^{-2K}} \overset{h \to 0}{\to} \begin{cases} 2 \cosh K \\ 2 \sinh K. \end{cases} \tag{11.4}$$

In the thermodynamic limit, $M_\tau \gg 1$, the bigger one dominates the free energy

$$e^{-F} = Z_1 = \lambda_+^{M_\tau} \left(1 + \left(\frac{\lambda_-}{\lambda_+} \right)^{M_\tau} \right) \sim \lambda_+^{M_\tau}.$$

Now I command you to think of the transfer matrix as

$$\mathbf{T} = e^{-\Delta\tau\mathbf{H}}$$

the propagator in euclidean time (by an amount $\Delta\tau$), where $\mathbf{H}$ is the *quantum* hamiltonian operator for a single qbit (note the boldface to denote quantum operators). So what's $\mathbf{H}$? To answer this, let's rewrite the parts of the transfer matrix in terms of paulis, thinking of $s = \pm$ as $\mathbf{Z}$-eigenstates. For $\mathbf{T}_2$, which is diagonal in the $\mathbf{Z}$ basis, this is easy:

$$\mathbf{T}_2 = e^{h\mathbf{Z}}.$$

To write $\mathbf{T}_1$ this way, stare at its matrix elements in the $\mathbf{Z}$ basis:

$$\langle s_1| \mathbf{T}_1 |s_2\rangle = \begin{pmatrix} e^K & e^{-K} \\ e^{-K} & e^K \end{pmatrix}_{s_1 s_2}$$

and compare them to those of

$$e^{a\mathbf{X}+b\mathbb{1}} = e^b e^{a\mathbf{X}} = e^b\left(\cosh a + \mathbf{X}\sinh a\right)$$

which are

$$\langle s_1| e^{a\mathbf{X}+b\mathbb{1}} |s_2\rangle = e^b \begin{pmatrix} \cosh a & \sinh a \\ \sinh a & \cosh a \end{pmatrix}_{s_1,s_2}$$

So we want $e^b\sinh a = e^{-K}, e^b\cosh a = e^K$ which is solved by

$$e^{-2K} = \tanh a \ . \tag{11.5}$$

So we want to identify

$$\mathbf{T}_1\mathbf{T}_2 = e^{b\mathbb{1}+a\mathbf{X}}e^{h\mathbf{Z}} \equiv e^{-\Delta\tau\mathbf{H}}$$

for small $\Delta\tau$. This requires that a, b, h scale like $\Delta\tau$, and so we can combine the exponents. Assuming that $\Delta\tau \ll E_0^{-1}, h^{-1}$, the result is

$$\mathbf{H} = E_0 - \frac{\Delta}{2}\mathbf{X} - \bar{h}\mathbf{Z} \ .$$

Here $E_0 = \frac{b}{\Delta\tau}, \bar{h} = \frac{h}{\Delta\tau}, \Delta = \frac{2a}{\Delta\tau}$. (Note that it's not surprising that the Hamiltonian for an isolated qbit is of the form $\mathbf{H} = d_0\mathbb{1} + \vec{d}\cdot\vec{\boldsymbol{\sigma}}$, since these operators span the set of hermitian operators on a qbit; but the relation between the parameters that we've found will be important.)

To recap, let's go backwards: consider the quantum system consisting of a single spin with $\mathbf{H} = E_0 - \frac{\Delta}{2}\mathbf{X} + \bar{h}\mathbf{Z}$. Set $\bar{h} = 0$ for a moment. Then Δ is the energy gap between the groundstate and the first excited state (hence the name). The thermal partition function is

$$Z_Q(T) = \mathrm{tr} e^{-\mathbf{H}/T} = \sum_{s=\pm} \langle s| e^{-\beta \mathbf{H}} |s\rangle \,, \tag{11.6}$$

where we've evaluated the trace in the $\mathbf{Z}$ basis, $\mathbf{Z}|s\rangle = s|s\rangle$. I emphasize that T here is the temperature to which we are subjecting our quantum spin; $\beta = \frac{1}{T}$ is the length of the euclidean time circle. Break up the euclidean time circle into M_τ intervals of size $\Delta\tau = \beta/M_\tau$. Insert many resolutions of unity (this is called 'Trotter decomposition')

$$Z_Q = \sum_{s_1 \dots s_{M_\tau}} \langle s_{M_\tau}| e^{-\Delta\tau \mathbf{H}} |s_{M_\tau - 1}\rangle \langle s_{M_\tau -1}| e^{-\Delta\tau \mathbf{H}} |s_{M_\tau - 2}\rangle \cdots \langle s_1| e^{-\Delta\tau \mathbf{H}} |s_{M_\tau}\rangle \ .$$

The RHS is the partition function of a classical Ising chain, Z_1 in (11.2), with $h = 0$ and K given by (11.5), which in the present variables is:

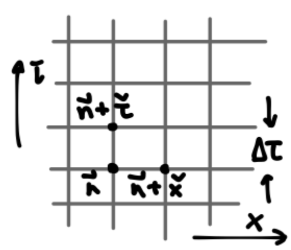
$$e^{-2K} = \tanh\left(\frac{\beta\Delta}{2M_\tau}\right) \ . \tag{11.7}$$

Notice that if our interest is in the quantum model with couplings E_0, Δ, we can use any M_τ we want – there are many classical models we could use[1]. For given M_τ, the couplings we should choose are related by (11.7).

A quantum system with just a single spin (for any $\mathbf{H}$ not proportional to $\mathbb{1}$) clearly has a unique groundstate; this statement means the absence of a phase transition in the 1d Ising chain.

More than one spin.[2] Let's do that procedure again, this time supposing the graph in question is a cubic lattice with more than one dimension, and let's think of one of the directions as euclidean time, τ. We'll end up with more than one spin.

We're going to rewrite the sum in (11.1) as a sum of products of (transfer) matrices. I will draw the pictures associated to a square lattice, but this is not a crucial limitation. Label points on the lattice by a vector $\vec{n}$ of integers; a unit vector in the time direction is $\check{\tau}$. First rewrite the classical action S in $Z_c = \sum e^{-S}$, using $s_j^2 = 1$, as



[1]If we include the $\mathbf{Z}$ term, we need to take $\Delta\tau$ small enough so that we can write

$$e^{-\Delta\tau \mathbf{H}} = e^{\Delta\tau \frac{\Delta}{2} \mathbf{X}} e^{-\Delta\tau\left(E_0 - \bar{h}\mathbf{Z}\right)} + \mathcal{O}(\Delta\tau^2)$$

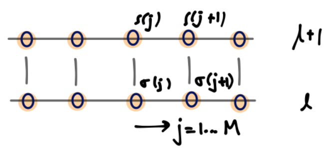
[2]This discussion comes from this paper of Fradkin and Susskind, and can be found in Kogut's review article.

$$
\begin{aligned}
S &= -\sum_{\vec n}\left(K s(\vec n+\check\tau)s(\vec n) + K_x s(\vec n+\check x)s(\vec n)\right)\\
&= K\sum_{\vec n}\left(\frac{1}{2}\left(s(\vec n+\check\tau)-s(\vec n)\right)^2 - 1\right) - K_x\sum_{\vec n} s(\vec n+\check x)s(\vec n)\\
&= \text{const} + \sum_{\text{rows at fixed time, } l} L(l+1,l)
\end{aligned}
\tag{11.8}
$$

with[3]

$$
L(s,\sigma) = \frac{1}{2}K\sum_j \left(s(j)-\sigma(j)\right)^2 - \frac{1}{2}K_x\sum_j \left(s(j+1)s(j)+\sigma(j+1)\sigma(j)\right).
$$

σ and s are the names for the spins on successive time slices, as in the figure at left.



The transfer matrix between successive time slices is a $2^M \times 2^M$ matrix:

$$
\langle s|\,\mathbf{T}\,|\sigma\rangle = T_{s\sigma} = e^{-L(s,\sigma)},
$$

in terms of which

$$
Z = \sum_{\{s\}} e^{-S} = \sum_{\{s\}}\prod_{l=1}^{M_\tau} T_{s(l+1,j),s(l,j)} = \mathrm{tr}_{\mathcal H}\mathbf{T}^{M_\tau}\ .
$$

This is just as in the one-site case; the difference is that now the hilbert space has a two-state system for every site on a fixed-l slice of the lattice. I will call this "space", and label these sites by an index j. (Note that nothing we say in this discussion requires space to be one-dimensional.) So $\mathcal H = \bigotimes_j \mathcal H_j$, where each $\mathcal H_j$ is a two-state system.

[End of Lecture 41]

The diagonal entries of $T_{s,\sigma}$ come from contributions where $s(l)=\sigma(l)$: they come with a factor of $\mathbf{T}_{s=\sigma} = e^{-L(0\text{ flips})}$ with

$$
L(0\text{ flips}) = -K_x\sum_j \sigma(j+1)\sigma(j).
$$

The one-off-the-diagonal terms come from

$$
\sigma(j)=s(j), \quad \text{except for one site where instead } \sigma(j)=-s(j).
$$

This gives a contribution

$$
L(1\text{ flips}) = \underbrace{\frac{1}{2}K(1-(-1))^2}_{=2K} - \frac{1}{2}K_x\sum_j\left(\sigma(j+1)\sigma(j)+s(j+1)s(j)\right).
$$

[3] Note that 'L' is for 'Lagrangian', so that $S=\int d\tau L$ and 'S' is for 'action'.

Similarly,

$$L(n \text{ flips}) = 2nK - \frac{1}{2}K_x \sum_j \left(\sigma(j+1)\sigma(j) + s(j+1)s(j)\right).$$

Now we need to figure out who is $\mathbf{H}$, as defined by

$$\mathbf{T} = e^{-\Delta\tau\mathbf{H}} \simeq 1 - \Delta\tau\mathbf{H} \ ;$$

we want to consider $\Delta\tau$ small and must choose K_x, K to make it so. We have to match the matrix elements $\langle s| \mathbf{T} |\sigma\rangle = T_{s\sigma}$:

$$\begin{aligned} T(0 \text{ flips})_{s\sigma} &= \delta_{s\sigma} e^{K_x \sum_j s(j)s(j+1)} &&\simeq \mathbb{1} - \Delta\tau\mathbf{H}|_{0 \text{ flips}} \\ T(1 \text{ flip})_{s\sigma} &= e^{-2K} e^{\frac{1}{2}K_x \sum_j (\sigma(j+1)\sigma(j)+s(j+1)s(j))} &&\simeq -\Delta\tau\mathbf{H}|_{1 \text{ flip}} \\ T(\text{n flips})_{s\sigma} &= e^{-2nK} e^{\frac{1}{2}K_x \sum_j (\sigma(j+1)\sigma(j)+s(j+1)s(j))} &&\simeq -\Delta\tau\mathbf{H}|_{\text{n flips}} \end{aligned} \tag{11.9}$$

From the first line, we learn that $K_x \sim \Delta\tau$; from the second we learn $e^{-2K} \sim \Delta\tau$; we'll call the ratio which we'll keep finite $g \equiv K_x^{-1} e^{-2K}$. To make τ continuous, we take $K \to \infty, K_x \to 0$, holding g fixed. Then we see that the n-flip matrix elements go like $e^{-nK} \sim (\Delta\tau)^n$ and can be ignored – the hamlitonian only has 0- and 1-flip terms.

To reproduce (11.9), we must take

$$\mathbf{H}_{\text{TFIM}} = -J\left(g\sum_j \mathbf{X}_j + \sum_j \mathbf{Z}_{j+1}\mathbf{Z}_j\right).$$

Here J is a constant with dimensions of energy that we pull out of $\Delta\tau$. The first term is the 'one-flip' term; the second is the 'zero-flips' term. The first term is a 'transverse magnetic field' in the sense that it is transverse to the axis along which the neighboring spins interact. So this is called the *transverse field ising model*. In $D = 1+1$ it can be understood completely, and I hope to say more about it later this quarter. As we'll see, it contains the *universal* physics of the 2d Ising model, including Onsager's solution. The word 'universal' requires some discussion.

Symmetry of the transverse field quantum Ising model: $\mathbf{H}_{\text{TFIM}}$ has a $\mathbb{Z}_2$ symmetry, generated by $\mathbf{S} = \prod_j \mathbf{X}_j$, which acts by

$$\mathbf{S}\mathbf{Z}_j = -\mathbf{Z}_j\mathbf{S}, \ \mathbf{S}\mathbf{X}_j = +\mathbf{X}_j\mathbf{S}, \ \ \forall j;$$

On $\mathbf{Z}$ eigenstates it acts as:

$$\mathbf{S}\left|\{s_j\}_j\right\rangle = \left|\{-s_j\}_j\right\rangle \quad .$$

It is a symmetry in the sense that:

$$[\mathbf{H}_{\text{TFIM}}, \mathbf{S}] = 0.$$

Notice that $\mathbf{S}^2 = \prod_j \mathbf{X}_j^2 = \mathbb{1},$, and $\mathbf{S} = \mathbf{S}^\dagger = \mathbf{S}^{-1}$.

By 'a $\mathbb{Z}_2$ symmetry,' I mean that the symmetry group consists of two elements $G = \{\mathbb{1}, \mathbf{S}\}$, and they satisfy $\mathbf{S}^2 = \mathbb{1}$, just like the group $\{1, -1\}$ under multiplication. This group is $G = \mathbb{Z}_2$. (For a bit of context, the group $\mathbb{Z}_N$ is realized by the Nth roots of unity, under multiplication.)

The existence of this symmetry of the quantum model is a direct consequence of the fact that the hamiltonian of the classical system (the action $S[s]$) was invariant under the operation $s_j \to -s_j, \forall j$. This meant that the matrix elements of the transfer matrix satisfy $T_{s,s'} = T_{-s,-s'}$ which implies the symmetry of $\mathbf{H}$. (Note that symmetries of the classical action do not so immediately imply symmetries of the associated quantum system if the system is not as well-regulated as ours is. This is the phenomenon called 'anomaly'.)

Quantum Ising in d space dimensions to classical Ising in $d+1$ dims

[Sachdev, 2d ed p. 75] Just to make sure it's nailed down, let's go backwards again. The partition function of the quantum Ising model at temperature T is

$$Z_Q(T) = \text{tr}_{\bigotimes_{j=1}^M \mathcal{H}_j} e^{-\frac{1}{T}\mathbf{H}_I} = \text{tr}\left(e^{-\Delta\tau \mathbf{H}_I}\right)^{M_\tau}$$

The transfer matrix here $e^{-\Delta\tau\mathbf{H}_I}$ is a $2^M \times 2^M$ matrix. We're going to take $\Delta\tau \to 0, M_\tau \to \infty$, holding $\frac{1}{T} = \Delta M_\tau$ fixed. Let's use the usual[4] 'split-step' trick of breaking up the non-commuting parts of $\mathbf{H}$:

$$e^{-\Delta\tau\mathbf{H}_I} \equiv \mathbf{T}_x\mathbf{T}_z + \mathcal{O}(\Delta\tau^2).$$

$$\mathbf{T}_x \equiv e^{Jg\Delta\tau\sum_j \mathbf{X}_j}, \quad \mathbf{T}_z \equiv e^{J\Delta\tau\sum_j \mathbf{Z}_j\mathbf{Z}_{j+1}} \quad .$$

Now insert a resolution of the identity in the $\mathbf{Z}$-basis,

$$\mathbb{1} = \sum_{\{s_j\}_{j=1}^M} |\{s_j\}\rangle\langle\{s_j\}|, \quad \mathbf{Z}_j|\{s_j\}\rangle = s_j|\{s_j\}\rangle, \quad s_j = \pm 1.$$

[4] By 'usual' I mean that this is just like in the path integral of a 1d particle, when we write

$$e^{-\Delta\tau\mathbf{H}} = e^{-\frac{\Delta\tau}{2m}\mathbf{p}^2}e^{-\Delta\tau V(\mathbf{q})} + \mathcal{O}(\Delta\tau^2).$$

many many times, one between each pair of transfer operators; this turns the transfer operators into transfer *matrices*. The $\mathbf{T}_z$ bit is diagonal, by design:

$$\mathbf{T}_z |\{s_j\}\rangle = e^{J\Delta\tau \sum_j s_j s_{j+1}} |\{s_j\}\rangle .$$

The $\mathbf{T}_x$ bit is off-diagonal, but only on a single spin at a time:

$$\langle\{s'_j\}| \mathbf{T}_x |\{s_j\}\rangle = \prod_j \underbrace{\langle s'_j| e^{Jg\Delta\tau \mathbf{X}_j} |s_j\rangle}_{2\times 2}$$

Acting on a single spin at site j, this 2×2 matrix is just the one from the previous discussion:

$$\langle s'_j| e^{Jg\Delta\tau \mathbf{X}_j} |s_j\rangle = e^{-b} e^{K s'_j s_j}, \quad e^{-b} = \frac{1}{2}\cosh(2Jg\Delta\tau), \quad e^{-2K} = \tanh(Jg\Delta\tau) \ .$$

Notice that it wasn't important to restrict to $1+1$ dimensions here. The only difference is in the $\mathbf{T}_z$ bit, which gets replaced by a product over all neighbors in higher dimensions:

$$\langle\{s'_j\}| \mathbf{T}_z |\{s_j\}\rangle = \delta_{s,s'} e^{J\Delta\tau \sum_{\langle jl\rangle} s_j s_l}$$

where $\langle jl\rangle$ denotes nearest neighbors, and the innocent-looking $\delta_{s,s'}$ sets the spins $s_j = s'_j$ equal for all sites.

Label the time slices by a variable $l = 1...M_\tau$.

$$Z = \mathrm{tr} e^{-\frac{1}{T}\mathbf{H}_I} = \sum_{\{s_j(l)\}} \prod_{l=1}^{M_\tau} \langle\{s_j(l+1)\}| \mathbf{T}_z \mathbf{T}_x |\{s_j(l)\}\rangle$$

The sum on the RHS runs over the 2^{MM_τ} values of $s_j(l) = \pm 1$, which is the right set of things to sum over in the $d+1$-dimensional classical ising model. The weight in the partition sum is

$$\begin{aligned} Z &= \underbrace{e^{-bM_\tau}}_{\substack{\text{unimportant} \\ \text{constant}}} \sum_{\{s_j(l)\}_{j,l}} \exp\left(\sum_{j,l}\left(\underbrace{J\Delta\tau s_j(l) s_{j+1}(l)}_{\text{space deriv, from } \mathbf{T}_z} + \underbrace{K s_j(l) s_j(l+1)}_{\text{time deriv, from } \mathbf{T}_x}\right)\right) \\ &= \sum_{\text{spins}} e^{-S_{\text{classical ising}}} \end{aligned}$$

except that the the couplings are a bit anisotropic: the couplings in the 'space' direction $K_x = J\Delta\tau$ are not the same as the couplings in the 'time' direction, which satisfy $e^{-2K} = \tanh(Jg\Delta\tau)$. (At the critical point $K = K_c$, this can be absorbed in a rescaling of spatial directions, as we'll see later.)

Dictionary. So this establishes a mapping between classical systems in $d+1$ dimensions and quantum systems in d space dimensions. Here's the dictionary:

statistical mechanics in $d+1$ dimensions	quantum system in d space dimensions
transfer matrix	euclidean-time propagator, $e^{-\Delta\tau\mathbf{H}}$
statistical 'temperature'	(lattice-scale) coupling K
free energy in infinite volume	groundstate energy: $e^{-F} = Z = \mathrm{tr}e^{-\beta\mathbf{H}} \overset{\beta\to 0}{\to} e^{-\beta E_0}$
periodicity of euclidean time L_τ	temperature: $\beta = \frac{1}{T} = \Delta\tau M_\tau$
statistical averages	groundstate expectation values of time-ordered operators

Note that this correspondence between classical and quantum systems is not an isomorphism. For one thing, we've seen that *many* classical systems are related to the same quantum system, which does not care about the lattice spacing in time. There is a set of physical quantities which agree between these different classical systems, called *universal*, which is the information in the quantum system. More on this below.

Consequences for phase transitions and quantum phase transitions.

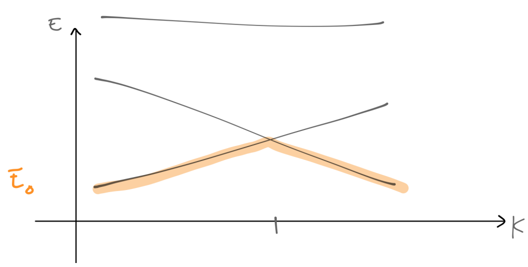
One immediate consequence is the following. Think about what happens at a *phase transition* of the classical problem. This means that the free energy $F(K, ...)$ has some kind of singularity at some value of the parameters, let's suppose it's the statistical temperature, *i.e.* the parameter we've been calling K. 'Singularity' means breakdown of the Taylor expansion, *i.e.* a disagreement between the actual behavior of the function and its Taylor series – a *non-analyticity*. First, this can *only* happen in the thermodynamic limit (at the very least $M_\tau \to \infty$), since otherwise there are only a finite number of terms in the partition sum and F is an analytic function of K (it's a polynomial in e^{-K}).

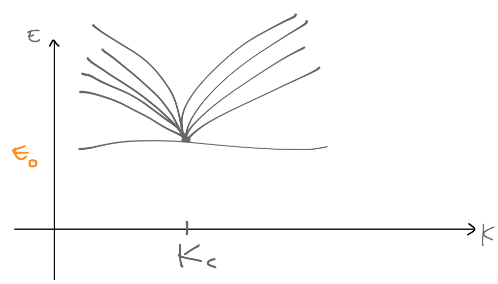
An important dichotomy is between *continuous* phase transitions (also called second order or higher) and first-order phase transitions; at the latter, $\partial_K F$ is discontinous at the transition, at the former it is not. This seems at first like an innocuous distinction, but think about it from the point of view of the transfer matrix for a moment. In the thermodynamic limit, $Z = \lambda_1(K)^{M_\tau}$, where $\lambda_1(K)$ is the largest eigenvalue of $\mathbf{T}(K)$. How can this have a singularity in K? There are two possibilities:

1. $\lambda_1(K)$ is itself a singular function of K. How can this happen? One way it can happen is if there is a level-crossing where two completely unrelated eigenvectors switch which is the smallest (while remaining separated from all the others).

   This is a first-order transition. A distinctive feature of a first order transition is a *latent heat*: although the free energies of the two phases are equal at the transition (they have to be in order to exchange dominance there), their entropies (and hence energies) are not: $S \propto \partial_K F$ jumps across the transition.

2. The other possibility is that the eigenvalues of $\mathbf{T}$ have an *accumulation point* at $K = K_c$, so that we can no longer ignore the contributions from the other eigenvalues to $\mathrm{tr}\mathbf{T}^{M_\tau}$, even when $M_\tau = \infty$. This is the exciting case of a continuous phase transition. In this case the critical point K_c is really special, and it has its own (euclidean) field theory which encodes all its intrinsic features.





Now translate those statements into statements about the corresponding quantum system. Recall that $\mathbf{T} = e^{-\Delta\tau \mathbf{H}}$ – eigenvectors of $\mathbf{T}$ are eigenvectors of $\mathbf{H}$! Their eigenvalues are related by

$$\lambda_a = e^{-\Delta\tau E_a} ,$$

so the largest eigenvalue of the transfer matrix corresponds to the smallest eigenvalue of $\mathbf{H}$: the groundstate. The two cases described above are:

1. As the parameter in $\mathbf{H}$ varies, two completely orthogonal states switch which one is the groundstate. This is a 'first-order quantum phase transition', but that name is a bit grandiose for this boring phenomenon, because the states on the two sides of the transition don't need to know anything about each other, and there is no interesting critical theory. For example, the *third* excited state need know nothing about the transition.

2. At a continuous transition in $F(K)$, the spectrum of $\mathbf{T}$ piles up at the top. This means that the spectrum of $\mathbf{H}$ is piling up at the bottom: the gap is closing. There is a gapless state which describes the physics in a whole neighborhood of the critical point.

Using the quantum-to-classical dictionary, the groundstate energy of the TFIM at the transition reproduces Onsager's tour-de-force free energy calculation.

Another failure mode of this correspondence: there are some quantum systems which when Trotterized produce a stat mech model with non-positive Boltzmann weights, *i.e.* $e^{-S} < 0$ for some configurations; this requires the classical hamiltonian S to be complex. These models are less familiar! An example where this happens is the spin-$\frac{1}{2}$ Heisenberg ($\equiv \mathsf{SU}(2)$-invariant) chain, as you'll see on the homework. This is a manifestation of a *sign problem*, which is a general term for a situation requiring adding up a bunch of numbers which aren't all positive, and hence may involve large cancellations. Sometimes such a problem can be removed by cleverness, sometimes it is a fundamental issue of computational complexity.

The quantum phase transitions of such quantum systems are not just ordinary finite-temperature transitions of familiar classical stat mech systems. So for the collector of QFTs, there is something to be gained by studying quantum phase transitions.

Correlation functions. [Sachdev, 2d ed p. 69] For now, let's construct correlation functions of spins in the classical Ising chain, (11.2), using the transfer matrix. (We'll study correlation functions in the TFIM later, I think.) Let

$$C(l, l') \equiv \langle s_l s_{l'} \rangle = \frac{1}{Z_1} \sum_{\{s_l\}_l} e^{-H_c} s_l s_{l'}$$

By translation invariance, this is only a function of the difference $C(l, l') = C(l - l')$. For simplicity, set the external field $h = 0$. Also, assume that $l' > l$ (as we'll see, this is *time-ordering* of the correlation function). In terms of the transfer matrix, it is:

$$C(l - l') = \frac{1}{Z} \mathrm{tr} \left(\mathbf{T}^{M_\tau - l'} \mathbf{Z} \mathbf{T}^{l'-l} \mathbf{Z} \mathbf{T}^{l} \right) . \tag{11.10}$$

Notice that there is only *one* operator $\mathbf{Z} = \boldsymbol{\sigma}^z$ here; it is the matrix

$$\mathbf{Z}_{ss'} = \delta_{ss'} s \ .$$

All the information about the index l, l' is encoded in the location in the trace.

Let's evaluate this trace in the basis of $\mathbf{T}$ eigenstates. When $h = 0$, we have $\mathbf{T} = e^{K} \mathbb{1} + e^{-K} \mathbf{X}$, so these are $\mathbf{X}$ eigenstates:

$$\mathbf{T} \left| \rightarrow \right\rangle = \lambda_+ \left| \rightarrow \right\rangle , \quad \mathbf{T} \left| \leftarrow \right\rangle = \lambda_- \left| \rightarrow \right\rangle \ .$$

Here $\left| \rightarrow \right\rangle \equiv \frac{1}{\sqrt{2}} \left(\left| \uparrow \right\rangle + \left| \downarrow \right\rangle \right)$.

In this basis

$$\langle\alpha|\,\mathbf{Z}\,|\beta\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{\alpha\beta}, \quad \alpha,\beta = \rightarrow \text{ or } \leftarrow .$$

So the trace (aka path integral) has two terms: one where the system spends $l' - l$ steps in the state $|\rightarrow\rangle$ (and the rest in $|\leftarrow\rangle$), and one where it spends $l' - l$ steps in the state $|\rightarrow\rangle$. The result (if we take $M_\tau \to \infty$ holding fixed $l' - l$) is

$$C(l'-l) = \frac{\lambda_+^{M_\tau - l' + l}\lambda_-^{l'-l} + \lambda_-^{M_\tau - l' + l}\lambda_+^{l'-l}}{\lambda_+^{M_\tau} + \lambda_-^{M_\tau}} \stackrel{M_\tau\to\infty}{\to} \tanh^{l'-l} K \ . \tag{11.11}$$

You should think of the insertions as

$$s_l = \mathbf{Z}(\tau), \ \tau = \Delta\tau l.$$

So what we've just computed is

$$C(\tau) = \langle \mathbf{Z}(\tau)\mathbf{Z}(0)\rangle = \tanh^l K = e^{-|\tau|/\xi} \tag{11.12}$$

where the *correlation time* ξ satisfies

$$\frac{1}{\xi} = \frac{1}{\Delta\tau}\ln\coth K \ . \tag{11.13}$$

Notice that this is the same as our formula for the gap, Δ, in (11.7).[5] This connection between the correlation length in euclidean time and the energy gap is general and important.

For large K, ξ is much bigger than the lattice spacing:

$$\frac{\xi}{\Delta\tau} \stackrel{K\gg 1}{\simeq} \frac{1}{2}e^{2K} \gg 1.$$

This is the limit we had to take to make the euclidean time continuous.

---

[5]Seeing this requires the following cool hyperbolic trig fact:

$$\text{If} \ \ e^{-2K} = \tanh X \ \ \text{then} \ \ e^{-2X} = \tanh K \tag{11.14}$$

(*i.e.* this equation is 'self-dual') which follows from algebra. Here (11.7) says $X = \frac{\Delta}{TM_\tau} = \Delta\tau\Delta$ while (11.13) says $X = \Delta\tau/\xi$. Actually this relation (11.14) can be made manifestly symmetric by writing it as

$$1 = \sinh 2X \sinh 2K \ .$$

(You may notice that this is the same combination that appears in the Kramers-Wannier self-duality condition.) I don't know a slick way to show this, but if you just solve this quadratic equation for e^{-2K} and boil it enough, you'll find $\tanh X$.

Notice that if we had taken $l < l'$ instead, we would have found the same answer with $l' - l$ replaced by $l - l'$.

[End of Lecture 42]

Continuum scaling limit and universality

[Sachdev, 2d ed §5.5.1, 5.5.2] Now we are going to grapple with the term 'universal'. Let's think about the Ising chain some more. We'll regard $M_\tau \Delta\tau$ as a physical quantity, the proper length of the chain. We'd like to take a continuum limit, where $M_\tau \to \infty$ or $\Delta\tau \to 0$ or maybe both. Such a limit is useful if $\xi \gg \Delta\tau$. This decides how we should scale K, h in the limit. More explicitly, here is the prescription: Hold fixed physical quantities (*i.e.* eliminate the quantities on the RHS of these expressions in favor of those on the LHS):

$$\begin{aligned} \text{the correlation length,} \quad & \xi && \simeq \Delta\tau \frac{1}{2} e^{2K}, \\ \text{the length of the chain,} \quad & L_\tau && = \Delta\tau M_\tau, \\ \text{physical separations between operators,} \quad & \tau && = (l - l')\Delta\tau, \\ \text{the applied field in the quantum system,} \quad & \bar{h} && = h/\Delta\tau. \end{aligned} \tag{11.15}$$

while taking $\Delta\tau \to 0, K \to \infty, M_\tau \to \infty$.

What physics of the various chains will agree? Certainly only quantities that don't depend explicitly on the lattice spacing; such quantities are called *universal*.

Consider the thermal free energy of the single quantum spin (11.6)[6]: The energy spectrum of our spin is $E_\pm = E_0 \pm \sqrt{(\Delta/2)^2 + \bar{h}^2}$, which means

$$F = -T \log Z_Q = E_0 - T \ln \left(2 \cosh \left(\beta \sqrt{(\Delta/2)^2 + \bar{h}^2} \right) \right)$$

(just evaluate the trace in the energy eigenbasis). In fact, this is just the behavior of the ising chain partition function in the scaling limit (11.15), since, in the limit (11.4) becomes

$$\lambda_\pm \simeq \sqrt{\frac{2\xi}{\Delta\tau}} \left(1 \pm \frac{\Delta\tau}{2\xi} \sqrt{1 + 4\bar{h}^2\xi^2} \right)$$

and so in the scaling limit (11.15)

$$F \simeq L_\tau \left(\underbrace{-\frac{K}{\Delta\tau}}_{\text{cutoff-dependent vac. energy}} - \frac{1}{L_\tau} \ln \left(2 \cosh \frac{L_\tau}{2} \sqrt{\xi^{-2} + 4\bar{h}^2} \right) \right),$$

which is the same (up to an additive constant) as the quantum formula under the previously-made identifications $T = \frac{1}{L_\tau}$, $\xi^{-1} = \Delta$.

[6] [Sachdev, 1st ed p. 19, 2d ed p. 73]

We can also use the quantum system to compute the correlation functions of the classical chain in the scaling limit (11.11). They are *time-ordered* correlation functions:

$$C(\tau_1 - \tau_2) = Z_Q^{-1} \mathrm{tr} e^{-\beta \mathbf{H}} \left(\theta(\tau_1 - \tau_2) \mathbf{Z}(\tau_1) \mathbf{Z}(\tau_2) + \theta(\tau_2 - \tau_1) \mathbf{Z}(\tau_2) \mathbf{Z}(\tau_1) \right)$$

where

$$\mathbf{Z}(\tau) \equiv e^{\mathbf{H}\tau} \mathbf{Z} e^{-\mathbf{H}\tau}.$$

This time-ordering is just the fact that we had to decide whether l' or l was bigger in (11.10).

For example, consider what happens to this when $T \to 0$. Then (inserting $\mathbb{1} = \sum_n |n\rangle \langle n|$, in an energy eigenbasis $\mathbf{H} |n\rangle = E_n |n\rangle$),

$$C(\tau)|_{T=0} = \sum_n | \langle 0| \mathbf{Z} |n\rangle |^2 e^{-(E_n - E_0)|\tau|}$$

where the $|\tau|$ is taking care of the time-ordering. This is a *spectral representation* of the correlator. For large τ, the contribution of $|n\rangle$ is exponentially suppressed by its energy, so the sum is approximated well by the lowest energy state for which the matrix element is nonzero. Assuming this is the first excited state (which in our two-state system it has no choice!), we have

$$C(\tau)|_{T=0} \overset{\tau \to \infty}{\simeq} e^{-\tau/\xi}, \quad \xi = 1/\Delta,$$

where Δ is the energy gap.

In these senses, the quantum theory of a single qbit is *the* universal theory of the Ising chain. For example, if we began with a chain that had in addition next-nearest-neighbor interactions, $\Delta H_c = K' \sum_j s(j) s(j+2)$, we could redo the procedure above. The scaling limit would not be exactly the same; we would have to scale K' somehow (it would also have to grow in the limit). But we would find the same 2-state quantum system, and when expressed in terms of physical variables, the $\Delta\tau$-independent terms in F would be identical, as would the form of the correlation functions, which is

$$C(\tau) = \langle \mathbf{Z}(\tau) \mathbf{Z}(0) \rangle = \frac{e^{-|\tau|/\xi} + e^{-(L_\tau - |\tau|)/\xi}}{1 + e^{-L_\tau/\xi}}.$$

(Note that in this expression we did not assume $|\tau| \ll L_\tau$ as we did before in (11.12), to which this reduces in that limit.)

11.2 Interlude on differential forms and algebraic topology

The next item of business is coherent state path integrals of all kinds. We are going to make a sneak attack on them.

[Zee section IV.4] We interrupt this physics discussion with a message from our mathematical underpinnings. This is nothing fancy, mostly just some book-keeping. It's some notation that we'll find useful. As a small payoff we can define some simple topological invariants of smooth manifolds.

Suppose we are given a smooth manifold X on which we can do calculus. For now, we don't even need a metric on X.

A p-form on X is a completely antisymmetric p-index tensor,

$$A \equiv \frac{1}{p!} A_{m_1 \ldots m_p} \mathrm{d}x^{m_1} \wedge \ldots \wedge \mathrm{d}x^{m_p}.$$

The coordinate one-forms are fermionic objects in the sense that $\mathrm{d}x^{m_1} \wedge \mathrm{d}x^{m_2} = -\mathrm{d}x^{m_2} \wedge \mathrm{d}x^{m_1}$ and $(\mathrm{d}x)^2 = 0$. The point in life of a p-form is that it can be integrated over a p-dimensional space. The order of its indices keeps track of the orientation (and it saves us the trouble of writing them). It is a geometric object, in the sense that it is something that can be (wants to be) integrated over a p-dimensional subspace of X, and its integral will only depend on the subspace, not on the coordinates we use to describe it.

Familiar examples include the gauge potential $A = A_\mu \mathrm{d}x^\mu$, and its field strength $F = \frac{1}{2} F_{\mu\nu} \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$. Given a curve C in X parameterized as $x^\mu(s)$, we have

$$\int_C A \equiv \int_C dx^\mu A_\mu(x) = \int ds \frac{dx^\mu}{ds} A_\mu(x(s))$$

and this would be the same if we chose some other parameterization or some other local coordinates.

The *wedge product* of a p-form A and a q-form B is a $p+q$ form

$$A \wedge B = A_{m_1 .. m_p} B_{m_{p+1} \ldots m_{p+q}} \mathrm{d}x^{m_1} \wedge \ldots \wedge \mathrm{d}x^{m_q} \ ,$$

[7] The space of p-forms on a manifold X is sometimes denoted $\Omega^p(X)$, especially when

[7] The components of $A \wedge B$ are then

$$(A \wedge B)_{m_1 \ldots m_{p+q}} = \frac{(p+q)!}{p!q!} A_{[m_1 \ldots m_p} B_{m_{p+1} \ldots m_{p+q}]}$$

where [..] means sum over permutations with a -1 for odd permutations. Try not to get caught up in the numerical prefactors.

it is to be regarded as a vector space (let's say over $\mathbb{R}$).

The *exterior derivative* d acts on forms as

$$\begin{array}{rcl} \mathrm{d}: \Omega^p(X) & \to & \Omega^{p+1}(X) \\ A & \mapsto & \mathrm{d}A \end{array}$$

by

$$\mathrm{d}A = \partial_{m_1} (A)_{m_2 \dots m_{p+1}} \mathrm{d}x^{m_1} \wedge \dots \wedge \mathrm{d}x^{m_{p+1}} \frac{1}{(p+1)!}.$$

You can check that

$$\mathrm{d}^2 = 0$$

basically because derivatives commute. Notice that $F = \mathrm{d}A$ in the example above. Denoting the boundary of a region D by ∂D, Stokes' theorem is

$$\int_D \mathrm{d}\alpha = \int_{\partial D} \alpha.$$

And notice that $\Omega^{p>\dim(X)}(X) = 0$ – there are no forms of rank larger than the dimension of the space.

A form ω_p is *closed* if it is killed by d: $\mathrm{d}\omega_p = 0$.

A form ω_p is *exact* if it is d of something: $\omega_p = \mathrm{d}\alpha_{p-1}$. That something must be a $(p-1)$-form.

Because of the property $\mathrm{d}^2 = 0$, it is possible to define *cohomology* – the image of one $\mathrm{d}: \Omega^p \to \Omega^{p+1}$ is in the kernel of the next $\mathrm{d}: \Omega^{p+1} \to \Omega^{p+2}$ (*i.e.* the Ω^ps form a *chain complex*). The pth de Rham cohomology group of the space X is defined to be

$$H^p(X) \equiv \frac{\text{closed } p\text{-forms on } X}{\text{exact } p\text{-forms on } X} = \frac{\ker(\mathrm{d}) \in \Omega^p}{\mathrm{Im}(\mathrm{d}) \in \Omega^p}.$$

That is, two closed p-forms are equivalent in cohomology if they differ by an exact form:

$$[\omega_p] - [\omega_p + \mathrm{d}\alpha_{p-1}] = 0 \in H^p(X),$$

where $[\omega_p]$ denotes the equivalence class. The dimension of this group is $b^p \equiv \dim H^p(X)$ called the pth betti number and is a topological invariant of X. The euler characteristic of X, which you can get by triangulating X and counting edges and faces and stuff is

$$\chi(X) = \sum_{p=0}^{d=\dim(X)} (-1)^p b^p(X).$$

Here's a very simple example, where $X = S^1$ is a circle. $x \simeq x + 2\pi$ is a coordinate; the radius will not matter since it can be varied continuously. An element of $\Omega^0(S^1)$ is a smooth *periodic* function of x. An element of $\Omega^1(S^1)$ is of the form $A_1(x)dx$ where A_1 is a smooth periodic function. Every such element is closed because there are no 2-forms on a 1d space. The exterior derivative on a 0-form is

$$\mathrm{d}A_0(x) = A_0' dx$$

Which 0-forms are closed? $A_0' = 0$ means A_0 is a constant. Which 1-forms can we make this way? The only one we can't make is dx itself, because x is not a periodic function. Therefore $b^0(S^1) = b^1(S^1) = 1$.

Now suppose we have a volume element on X, *i.e.* a way of integrating d-forms. This is guaranteed if we have a metric, since then we can integrate $\int \sqrt{\det g} ...$, but is less structure. Given a volume form, we can define the Hodge star operation $\star$ which maps a p-form into a $(d-p)$-form:

$$\star : \Omega^p \to \Omega^{d-p}$$

by

$$\left(\star A^{(p)}\right)_{\mu_1 \cdots \mu_{d-p}} \equiv \epsilon_{\mu_1 \cdots \mu_d} A^{(p)\ \mu_{d-p+1} \cdots \mu_d}$$

An application: consider the Maxwell action, $\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$. You can show that this is the same as $S[A] = \int F \wedge \star F$. (Don't trust my numerical prefactor.) You can derive the Maxwell EOM by $0 = \frac{\delta S}{\delta A}$. $\int F \wedge F$ is the θ term. The magnetic dual field strength is $\tilde{F} = \star F$. Many generalizations of duality can be written naturally using the Hodge $\star$ operation.

As you can see from the Maxwell example, the Hodge star gives an inner product on Ω^p: for two p-forms α, β $(\alpha, \beta) = \int \alpha \wedge \star \beta)$, $(\alpha, \alpha) \geq 0$. We can define the adjoint of d with respect to the inner product by

$$\int \mathrm{d}^\dagger \alpha \wedge \star \beta = (\mathrm{d}^\dagger \alpha, \beta) \equiv (\alpha, \mathrm{d}\beta) = \int \alpha \wedge \star \mathrm{d}\beta$$

Combining this relation with integration by parts, we find $\mathrm{d}^\dagger = \pm \star \mathrm{d} \star$.

We can make a Laplacian on forms by

$$\Delta = \mathrm{d}\mathrm{d}^\dagger + \mathrm{d}^\dagger \mathrm{d}.$$

This is a supersymmetry algebra, in the sense that $\mathrm{d}, \mathrm{d}^\dagger$ are grassmann operators.

Any cohomology class $[\omega]$ has a *harmonic* representative, $[\omega] = [\tilde{\omega}]$ where in addition to being closed $\mathrm{d}\omega = \mathrm{d}\tilde{\omega} = 0$, it is *co-closed*, $0 = \mathrm{d}^\dagger\tilde{\omega}$, and hence *harmonic* $\Delta\tilde{\omega} = 0$.

I mention this because it implies Poincare duality: $b^p(X) = b^{d-p}(X)$ if X has a volume form. This follows because the map $H^p \to H^{d-p}$ $[\omega_p] \mapsto [\star\omega_p]$ is an isomorphism. (Choose the harmonic representative, it has $\mathrm{d}\star\tilde{\omega}_p = 0$.)

The de Rham complex of X can be realized as the groundstates of a physical system, namely the supersymmetric nonlinear sigma model with target space X. The fermions play the role of the dx^μs. The states are of the form

$$|A\rangle = \sum_{p=1}^{d} A_{\mu_1\cdots\mu_p}(x)\psi^{\mu_1}\psi^{\mu_2}\cdots\psi^{\mu_p}\,|0\rangle$$

where ψ are some fermion creation operators. This shows that the hilbert space is the space of forms on X, that is $\mathcal{H} \simeq \Omega(X) = \oplus_p \Omega^p(X)$. The supercharges act like d and $\mathrm{d}^\dagger$ and therefore the supersymmetric groundstates are (harmonic representatives of) cohomology classes.

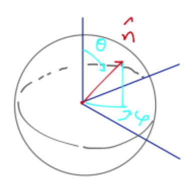
This machinery will be very useful to us. I use it all the time.

[End of Lecture 43]

11.3 Coherent state path integrals for spin systems

11.3.1 Geometric quantization and coherent state quantization of spin systems

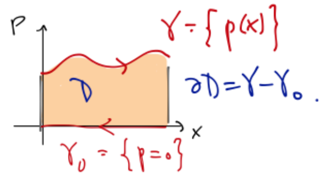
[Zinn-Justin, Appendix A3; XGW §2.3] We're going to spend some time talking about QFT in $D = 0+1$, then we'll work our way up to $D = 1+1$, and beyond. Consider the nice, round two-sphere. It has an area element which can be written



$$\omega = s\,\mathrm{d}\cos\theta\wedge\mathrm{d}\varphi \qquad \text{and satisfies} \qquad \int_{S^2}\omega = 4\pi s.$$

s is a number. Suppose we think of this sphere as the *phase space* of some dynamical system. We can use ω as the symplectic form. What is the associated quantum mechanics system?

Let me remind you what I mean by 'the symplectic form'. Recall the phase space formulation of classical dynamics. The action associated to a trajectory is



$$\mathcal{A}[x(t), p(t)] = \int_{t_1}^{t_2} \mathrm{d}t \left(p\dot{x} - H(x,p) \right) = \int_\gamma p(x)\mathrm{d}x - \int H \mathrm{d}t$$

where γ is the trajectory through the phase space. The first term is the area 'under the graph' in the classical phase space – the area between (p, x) and $(p = 0, x)$. We can rewrite it as

$$\int p(t)\dot{x}(t)\mathrm{d}t = \int_{\partial D} p\mathrm{d}x = \int_D \mathrm{d}p \wedge \mathrm{d}x$$

using Stokes' theorem; here ∂D is the closed curve made by the classical trajectory and some reference trajectory ($p = 0$) and it bounds some region D. Here $\omega = \mathrm{d}p \wedge \mathrm{d}x$ is the symplectic form. More generally, we can consider an $2n$-dimensional phase space with coordinates $u_\alpha, \alpha = 1..2n$ and symplectic form

$$\omega = \omega_{\alpha\beta}\mathrm{d}u^\alpha \wedge \mathrm{d}u^\beta$$

and action

$$\mathcal{A}[u] = \int_D \omega - \int_{\partial D} \mathrm{d}t H(u,t).$$

The symplectic form says who is canonically conjugate to whom. It's important that $\mathrm{d}\omega = 0$ so that the equations of motion resulting from $\mathcal{A}$ depend only on the trajectory $\gamma = \partial D$ and not on the interior of D. The equations of motion from varying u are

$$\omega_{\alpha\beta}\dot{u}^\beta = \frac{\partial H}{\partial u^\alpha}.$$

Locally, we can find coordinates p, x so that $\omega = \mathrm{d}(p\mathrm{d}x)$. Globally on the phase space this is not guaranteed – the symplectic form needs to be closed, but need not be exact.

So the example above of the two-sphere is one where the symplectic form is closed (there are no three-forms on the two sphere, so $\mathrm{d}\omega = 0$ automatically), but is not exact. One way to see that it isn't exact is that if we integrate it over the whole two-sphere, we get the area:

$$\int_{S^2} \omega = 4\pi s \ .$$

On the other hand, the integral of an exact form over a closed manifold (meaning a manifold without boundary, like our sphere) is zero:

$$\int_C \mathrm{d}\alpha = \int_{\partial C} \alpha = 0.$$

So there can't be a *globally defined* one-form α such that $\mathrm{d}\alpha = \omega$. Locally, we can find one; for example:

$$\alpha = s\cos\theta \mathrm{d}\varphi \ ,$$

but this is singular at the poles, where φ is not a good coordinate.

So: what I mean by "what is the associated quantum system..." is the following: let's construct a system whose path integral is

$$Z = \int [\mathrm{d}\theta \mathrm{d}\varphi] e^{\frac{\mathbf{i}}{\hbar}\mathcal{A}[\theta,\varphi]} \tag{11.16}$$

with the action above, and where $[\mathrm{d}x]$ denotes the path integral measure:

$$[\mathrm{d}x] \equiv \aleph \prod_{i=1}^{N} \mathrm{d}x(t_i)$$

where $\aleph$ involves lots of awful constants that drop out of ratios. It is important that the measure does not depend on our choice of coordinates on the sphere.

- Hint 1: the model has an action of $\mathsf{O}(3)$, by rotations of the sphere.

- Hint 2: We actually didn't specify the model yet, since we didn't choose the Hamiltonian. For definiteness, let's pick the hamiltonian to be

  $$H = -s\vec{h}\cdot\vec{n}$$

  where $\vec{n} \equiv (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$. WLOG, we can take the polar axis to be along the 'magnetic field': $\vec{h} = \hat{z}h$. The equations of motion are then

  $$0 = \frac{\delta\mathcal{A}}{\delta\theta(t)} = -s\sin\theta\left(\dot{\varphi} - h\right), \quad 0 = \frac{\delta\mathcal{A}}{\delta\varphi(t)} = -\partial_t\left(s\cos\theta\right)$$

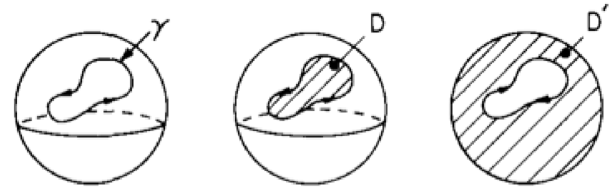
  which by rotation invariance can be written better as

  $$\partial_t\vec{n} = \vec{h}\times\vec{n}. \tag{11.17}$$

  This is a big hint about the answer to the question.

- Hint 3: Semiclassical expectations. Semiclassically, each patch of phase space of area $\hbar$ contributes one quantum state. Therefore we expect that if our whole phase space has area $4\pi s$, we should get approximately $\frac{4\pi s}{2\pi\hbar} = \frac{2s}{\hbar}$ states, at least at large $s/\hbar$. (Notice that s appears out front of the action.) This will turn out to be very close – the right answer is $2s+1$ (when the spin is measured in units with $\hbar = 1$)!

In QM we care that the action produces a well-defined phase – the action must be defined modulo additions of 2π times an integer. We should get the same answer whether we fill in one side D of the trajectory γ or the other D'. The difference between them is



[from Witten]

$$s\left(\int_D - \int_{D'}\right) \text{area} = s \int_{S^2} \text{area} \ .$$

So in this difference s multiplies $\int_{S^2} \text{area} = 4\pi$ (actually, this can be multiplied by an integer which is the number of times the area is covered). Our path integral will be well-defined (*i.e.* independent of our arbitrary choice of 'inside' and 'outside') only if $4\pi s \in 2\pi\mathbb{Z}$, that is if $\boxed{2s \in \mathbb{Z} \text{ is an integer}}$.

The conclusion of this discussion is that the coefficient of the area term must be an integer. We will interpret this integer below.

WZW term. We have a nice geometric interpretation of the 'area' term in our action $\mathcal{A}$ – it's the solid angle swept out by the particle's trajectory. But how do we write it in a manifestly SU(2) invariant way? We'd like to be able to write it, not in terms of the annoying coordinates θ, ϕ, but directly in terms of

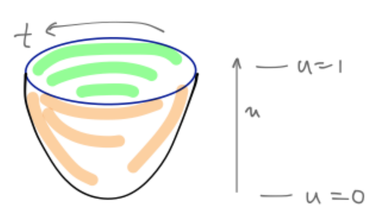
$$n^a \equiv (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)^a.$$

One way to do this is to add an extra dimension (!):

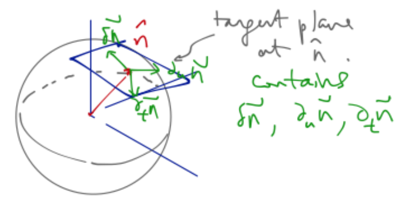
$$\frac{1}{4\pi}\int \mathrm{d}t\,(1-\cos\theta)\,\partial_t\phi = \frac{1}{8\pi}\int_0^1 \mathrm{d}u \int \mathrm{d}t\epsilon_{\mu\nu}n^a\partial_\mu n^b\partial_\nu n^c\epsilon^{abc} \equiv W_0[\vec{n}]$$

where $x^\mu = (t, u)$, and the ϵ tensors are completely antisymmetric in their indices with all nonzero entries 1 and -1.

In order to write this formula we have to extend the $\vec{n}$-field into the extra dimension whose coordinate is u. We do this in such a way that the real spin lives at $u = 1$: $\vec{n}(t, u = 1) = \vec{n}(t)$, and $\vec{n}(t, u = 0) = (0, 0, 1)$ – it goes to the north pole at the other end of the extra dimension for all t. If we consider periodic boundary conditions in time $n(\beta) = n(0)$, then this means that the space is really a disk with the origin at $u = 0$, and the boundary at $u = 1$. Call this disk B, its boundary ∂B is the real spacetime ('B' is for 'ball').



This WZW term has the property that its variation with respect to $\vec{n}$ depends only on the values at the boundary (that is: δW_0 is a total derivative). The crucial reason is that allowed variations $\delta\vec{n}$ lie on the 2-sphere, as do derivatives $\partial_\mu \vec{n}$; this means $\epsilon^{abc}\delta n^a \partial_\mu n^b \partial_\nu n^c = 0$, since they all lie in a two-dimensional tangent plane to the 2-sphere at $\vec{n}(t)$. Therefore:



$$
\begin{aligned}
\delta W_0 &= \int_0^1 \mathrm{d}u \int \mathrm{d}t \frac{1}{4\pi}\epsilon^{\mu\nu} n^a \partial_\mu \delta n^b \partial_\nu n^c \epsilon^{abc} = \int_B \frac{1}{4\pi} n^a \mathrm{d}\delta n^b \wedge \mathrm{d}n^c \epsilon^{abc} \\
&= \int_0^1 \mathrm{d}u \int \mathrm{d}t\ \partial_\mu \left(\frac{1}{4\pi}\epsilon^{\mu\nu} n^a \delta n^b \partial_\nu n^c \epsilon^{abc} \right) = \int_B \mathrm{d}\left(\frac{1}{4\pi} n^a \delta n^b \mathrm{d}n^c \epsilon^{abc} \right) \\
&\overset{\text{Stokes}}{=} \frac{1}{4\pi} \int \mathrm{d}t \delta\vec{n} \cdot \left(\dot{\vec{n}} \times \vec{n} \right). \qquad (11.18)
\end{aligned}
$$

(Note that $\epsilon^{abc} n^a m^b \ell^c = \vec{n} \cdot \left(\vec{m} \times \vec{\ell} \right)$. The right expressions in red in each line are a rewriting in terms of differential forms; notice how much prettier they are.) So the equations of motion coming from this term do not depend on how we extend it into the auxiliary dimension.

And in fact they are the same as the ones we found earlier:

$$
0 = \frac{\delta}{\delta\vec{n}(t)} \left(4\pi s W_0[n] + s\vec{h} \cdot \vec{n} + \lambda \left(\vec{n}^2 - 1 \right) \right) = s\partial_t \vec{n} \times \vec{n} + s\vec{h} + 2\lambda\vec{n}
$$

(λ is a Lagrange multiplier to enforce unit length.) The cross product of this equation with $\vec{n}$ is $\partial_t \vec{n} = \vec{h} \times \vec{n}$.

In QM we also care that the action produces a well-defined phase – the action must be defined modulo additions of 2π times an integer. There may be many ways to extend $\hat{n}$ into an extra dimension; another obvious way is shown in the figure above. The demand that the action is the same modulo $2\pi\mathbb{Z}$ gives the same quantization law as above for the coefficient of the WZW term. So the WZW term is topological in the sense that because of topology its coefficient must be quantized.

(This set of ideas generalizes to many other examples, with other fields in other dimensions. WZW stands for Wess-Zumino-Witten.)

Coherent quantization of spin systems. [Wen §2.3.1, Fradkin, Sachdev, *QPT*, chapter 13 and §2.2 of cond-mat/0109419] To understand more about the path integral we've just constructed, we now go in the opposite direction. Start with a spin one-half system, with

$$
\mathcal{H}_{\frac{1}{2}} \equiv \mathrm{span}\{ |\uparrow\rangle, |\downarrow\rangle \}.
$$

Define spin coherent states $|\vec{n}\rangle$ by[8]:

$$\vec{\boldsymbol{\sigma}} \cdot \vec{n} \, |\vec{n}\rangle = |\vec{n}\rangle \ .$$

These states form another basis for $\mathcal{H}_{\frac{1}{2}}$; they are related to the basis where $\boldsymbol{\sigma}^z$ is diagonal by:

$$|\vec{n}\rangle = z_1 |\uparrow\rangle + z_2 |\downarrow\rangle \,, \qquad \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} e^{\mathbf{i}\varphi/2} \cos \frac{\theta}{2} e^{\mathbf{i}\psi/2} \\ e^{-\mathbf{i}\varphi/2} \sin \frac{\theta}{2} e^{\mathbf{i}\psi/2} \end{pmatrix} \tag{11.19}$$

as you can see by diagonalizing $\vec{n} \cdot \vec{\boldsymbol{\sigma}}$ in the $\boldsymbol{\sigma}^z$ basis. Notice that

$$\vec{n} = z^\dagger \vec{\boldsymbol{\sigma}} z, \quad |z_1|^2 + |z_2|^2 = 1$$

and the phase of z_α does not affect $\vec{n}$ (this is the Hopf fibration $S^3 \to S^2$). In (11.19) I chose a representative of the phase. The space of independent states is a two-sphere:

$$S^2 = \{(z_1, z_2) | |z_1|^2 + |z_2|^2 = 1\} / (z_\alpha \simeq e^{\mathbf{i}\chi} z_\alpha).$$

It is just the ordinary Bloch sphere of pure states of a qbit.

These states are not orthogonal (there are infinitely many of them and the Hilbert space is only 2-dimensional!):

$$\langle \check{n}_1 | \check{n}_2 \rangle = z_1^\dagger z_2$$

as you can see using the σ^z-basis representation (11.19). The (over-)completeness relation in this basis is:

$$\int \frac{\mathrm{d}^2 \vec{n}}{2\pi} |\vec{n}\rangle \langle \vec{n}| = 1\!\!1_{2\times 2}. \tag{11.20}$$

As always, we can construct a path integral representation of any amplitude by inserting many copies of $1\!\!1$ in between successive time steps. For example, we can construct such a representation for the propagator using (11.20) many times:

$$\begin{aligned} \mathbf{i}G(\vec{n}_f, \vec{n}_1, t) &\equiv \langle \vec{n}_f | e^{-\mathbf{i}\mathbf{H}t} |\vec{n}_1\rangle \\ &= \int \prod_{i=1}^{N \equiv \frac{t}{\mathrm{d}t}} \frac{\mathrm{d}^2 \vec{n}(t_i)}{2\pi} \lim_{\mathrm{d}t \to 0} \langle \vec{n}(t) | \vec{n}(t_N) \rangle \ldots \langle \vec{n}(t_2) | \vec{n}(t_1) \rangle \langle \vec{n}(t_1) | \vec{n}(0) \rangle \,. \end{aligned} \tag{11.21}$$

(Notice that $\mathbf{H} = 0$ here, so $\mathbf{U} \equiv e^{-\mathbf{i}\mathbf{H}t}$ is actually the identity.) The crucial ingredient is

$$\langle \vec{n}(t+\epsilon) | \vec{n}(t) \rangle = z^\dagger(\mathrm{d}t) z(0) = 1 - z^\dagger(\mathrm{d}t) \left(z(\mathrm{d}t) - z(0) \right) \approx e^{-z^\dagger \partial_t z \mathrm{d}t}.$$

[8]For more general spin representation with spin $s > \frac{1}{2}$, and spin operator $\vec{\mathbf{S}}$, we would generalize this equation to

$$\vec{\mathbf{S}} \cdot \vec{n} \, |\vec{n}\rangle = s \, |\vec{n}\rangle \,.$$

$$\mathbf{i}G(\vec{n}_2, \vec{n}_1, t) = \int \left[\frac{D\vec{n}}{2\pi}\right] e^{\mathbf{i}S_B[\vec{n}(t)]}, \qquad S_B[\vec{n}(t)] = \int_0^t \mathrm{d}t \mathbf{i} z^\dagger \dot{z} \ . \tag{11.22}$$

Notice how weird this is: even though the Hamiltonian of the spins was zero – whatever their state, they have no potential energy and no kinetic energy – the action in the path integral is not zero. This phase $e^{\mathbf{i}S_B}$ is a quantum phenomenon called a Berry phase. [End of Lecture 44]

Starting from the action S_B and doing the Legendre transform to find the Hamiltonian you will get zero. The first-derivative action says that $z^\dagger$ is the canonical momentum conjugate to z: the space with coordinates $(z, z^\dagger)$ becomes the phase space (just like position and momentum)! But this phase space is curved. In fact it is the two-sphere

$$S^2 = \{(z_1, z_2) | |z_1|^2 + |z_2|^2 = 1\} / (z_\alpha \simeq e^{\mathbf{i}\psi} z_\alpha).$$

In terms of the coordinates θ, φ above, we have

$$S_B[z] = S_B[\theta, \varphi] = \int dt \left(-\frac{1}{2}\cos\theta\dot{\phi} - \frac{1}{2}\dot{\phi}\right) = -4\pi s W_0[\hat{n}]|_{s=\frac{1}{2}}. \tag{11.23}$$

BIG CONCLUSION: This is the 'area' term that we studied above, with $s = \frac{1}{2}$! So the expression in terms of z in (11.22) gives another way to write the area term which is manifestly SU(2) invariant; this time the price is introducing these auxiliary z variables.

The Berry phase $S_B[n]$ is *geometric*, in the sense that it depends on the trajectory of the spin through time, but not on its parametrization, or speed or duration. It is called the Berry phase of the spin history because it is the phase acquired by a spin which follows the instantaneous groundstate (*i.e.* adiabatic evolution) $|\Psi_0(t)\rangle$ of $H(\check{n}(t), t) \equiv -h(t)\check{n}(t) \cdot \mathbf{S}$, with $h > 0$. This is Berry's adiabatic phase, $S_B = -\lim_{\partial_t h \to 0} \int dt \mathrm{Im} \langle \Psi_0(t)| \partial_t |\Psi_0(t)\rangle$.

Making different choices of for the phase ψ at different times can shift the constant in front of the second term in (11.23); as we observed earlier, this term is a total derivative. Different choices of ψ change the overall phase of the wavefunction, which doesn't change physics (recall that this is why the space of normalized states of a qbit is a two-sphere and not a three-sphere). Notice that $\mathcal{A}_t = z^\dagger \partial_t z$ is like the time component of a gauge field.

Since S_B is geometric, like integrals of differential forms, let's take advantage of this to make it pretty and relate it to familiar objects. Introduce a vector potential (the Berry connection) on the sphere $A^a, a = x, y, z$ so that

$$S_B = \oint d\tau \dot{n}_a A^a = \oint_\gamma A \overset{\text{Stokes}}{=} \int_D F$$

where $\gamma = \partial D$ is the trajectory. ($F = \mathrm{d}A$ is the Berry curvature.) What is the correct form? We must have $(\nabla \times A) \cdot \check{n} = \epsilon^{abc}\partial_{n^a} A^b n^c = 1$ (for spin half). This is a monopole field. Two choices which work are

$$A^{(1)} = -\cos\theta d\varphi, \quad \text{and} \quad A^{(2)} = (1 - \cos\theta) d\varphi.$$

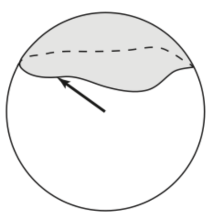
These two expressions differ by the gauge transformation $\mathrm{d}\varphi$, which is locally a total derivative. The first is singular at the N and S poles, $\check{n} = \pm\check{z}$. The second is singular only at the S pole. Considered as part of a 3d field configuration, this codimension two singularity is the 'Dirac string'. The demand of invisibility of the Dirac string quantizes the Berry flux.

If we redo the above coherent-state quantization for a spin-s system we'll get the expression with general s (see below). Notice that this only makes sense when $2s \in \mathbb{Z}$.

We can add a nonzero Hamiltonian for our spin; for example, we can put it in an external Zeeman field $\vec{h}$, which adds $\mathbf{H} = -\vec{h} \cdot \vec{\mathbf{S}}$. This will pass innocently through the construction of the path integral, adding a term to the action $S = S_B + S_h$,

$$S_h = \int \mathrm{d}t \left(s\vec{h} \cdot \vec{n} \right)$$

where s is the spin.



We are back at the system (11.16). We see that the system we get by 'geometric quantization' of the sphere is a quantum spin. The quantized coefficient of the area is $2s$: it determines the dimension of the spin space to be $2s+1$. Here the quantization of the WZW term is just quantization of angular momentum. (In higher-dimensional field theories, it is something else.)

Deep statement: the purpose in life of the WZW term is to enforce the commutation relation of the SU(2) generators, $[\mathbf{S}^i, \mathbf{S}^j] = \mathbf{i}\epsilon^{ijk}\mathbf{S}^k$. It says that the different components of the spin don't commute, and it says precisely what they don't commute to.

Incidentally, another way to realize this system whose action is proportional to the area of the sphere is to take a particle on the sphere, put a magnetic monopole in the center, and take the limit that the mass of the particle goes to zero. In that context, the quantization of $2s$ is Dirac quantization of magnetic charge. And the degeneracy of $2s + 1$ states is the degeneracy of states in the lowest Landau level for a charged particle in a magnetic field; the $m \to 0$ limit gets rid of the higher Landau levels (which are separated from the lowest by the cylotron frequency, $\frac{eB}{mc}$).

In the crucial step, we assumed the path $z(t)$ was smooth enough in time that we could do calculus, $z(t+\epsilon) - z(t) = \epsilon \dot{z}(t) + \mathcal{O}(\epsilon^2)$. Is this true of the important contributions to the path integral? Sometimes not, and we'll come back to this later.

Digression on $s > \frac{1}{2}$. [Auerbach, *Interacting Electrons and Quantum Magnetism*] I want to say something about larger-spin representations of $\mathsf{SU}(2)$, partly to verify the claim above that it results in a factor of $2s$ in front of the Berry phase term. Also, large s allows us to approximate the integral by stationary phase.

In general, a useful way to think about the coherent state $|\check{n}\rangle$ is to start with the maximal-spin eigenstate $|s,s\rangle$ of $\mathbf{S}^z$ (the analog of spin up for general s), and *rotate it* by the rotation that takes $\mathbf{S}^z$ to $\mathbf{S}\cdot\check{n}$:

$$|\check{n}\rangle = \mathcal{R}(\chi,\theta,\varphi)\,|s,s\rangle = e^{\mathbf{i}\mathbf{S}^z\varphi}e^{\mathbf{i}\mathbf{S}^y\theta}e^{\mathbf{i}\mathbf{S}^z\psi}\,|s,s\rangle = e^{\mathbf{i}s\psi}e^{\mathbf{i}\mathbf{S}^z\varphi}e^{\mathbf{i}\mathbf{S}^y\theta}\,|s,s\rangle\,.$$

Schwinger bosons. The following is a helpful device for spin matrix elements. Consider two copies of the harmonic oscillator algebra, with modes a,b satisfing $[a,a^\dagger] = 1 = [b,b^\dagger], [a,b] = [a,b^\dagger] = 0$. Then

$$\mathbf{S}^+ = a^\dagger b,\ \mathbf{S}^- = b^\dagger a,\ \mathbf{S}^z = a^\dagger a - b^\dagger b$$

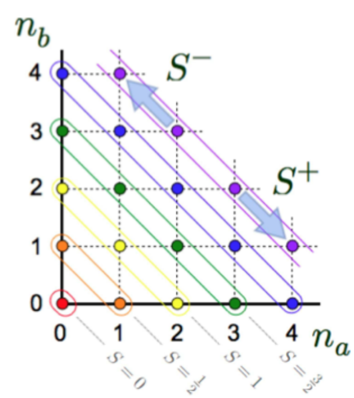
satisfy the $\mathsf{SU}(2)$ algebra. The no-boson state $|0\rangle$ is a singlet of this $\mathsf{SU}(2)$, and the one-boson states $\begin{pmatrix} a^\dagger|0\rangle \\ b^\dagger|0\rangle\end{pmatrix}$ form a spin-half doublet.

More generally, the states

$$\mathcal{H}_s \equiv \text{span}\{|n_a,n_b\rangle\,|a^\dagger a + b^\dagger b \equiv n_a + n_b = 2d\}$$

form a spin-s representation. Algebraic evidence for this is the fact that $\vec{S}^2 P_s = s(s+1)P_s$, where P_s is the projector onto $\mathcal{H}_s$. The spin-s eigenstates of $\mathbf{S}^z$ are

$$|s,m\rangle = \frac{(a^\dagger)^{s+m}}{\sqrt{(s+m)!}}\frac{(b^\dagger)^{s-m}}{\sqrt{(s-m)!}}\,|0\rangle\,.$$



[nice figure from Arovas and Auerbach, 0809.4836.]

The fact that $\begin{pmatrix} a^\dagger|0\rangle \\ b^\dagger|0\rangle\end{pmatrix}$ form a doublet means that $\begin{pmatrix} a^\dagger \\ b^\dagger\end{pmatrix}$ must be a doublet. But we know how a doublet transforms under a rotation, and this means we know how to write the coherent state:

$$|\check{n}\rangle = \mathcal{R}\,|s,s\rangle = \mathcal{R}\frac{(a^\dagger)^{2s}}{\sqrt{(2s)!}}\,|0\rangle = \mathcal{R}\frac{(a^\dagger)^{2s}}{\sqrt{(2s)!}}\mathcal{R}^{-1}\mathcal{R}\,|0\rangle = \frac{(a'^\dagger)^{2s}}{\sqrt{(2s)!}}\,|0\rangle = \frac{(z_1 a^\dagger + z_2 b^\dagger)^{2s}}{\sqrt{(2s)!}}\,|0\rangle\,.$$

Here $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} e^{\mathbf{i}\varphi/2}\cos\frac{\theta}{2}e^{\mathbf{i}\psi/2} \\ e^{-\mathbf{i}\varphi/2}\sin\frac{\theta}{2}e^{\mathbf{i}\psi/2} \end{pmatrix}$ as above[9].

But now we can compute the crucial ingredient in the coherent state path integral, the overlap of successive coherent states:

$$\langle\check{n}|\check{n}'\rangle = \frac{e^{-\mathbf{i}s(\psi-\psi')}}{(2s)!}\underbrace{\langle 0|\,(z_1^\star a + z_2^\star b)^{2s}(z_1' a^\dagger + z_2' b^\dagger)^{2s}\,|0\rangle}_{\overset{\text{Wick}}{=}(2s)!\left([z_1^\star a+z_2^\star b, z_1' a^\dagger + z_2' b^\dagger]\right)^{2s}} = e^{-\mathbf{i}s(\psi-\psi')}(z_1^\star z_1' + z_2^\star z_2')^{2s} = \left(e^{-\mathbf{i}(\psi-\psi')/2}z^\dagger\cdot z'\right)^{2s}.$$

Here's the point: this is the same as the spin-half answer, raised to the $2s$ power. This means that the Berry phase just gets multiplied by $2s$, $S_B^{(s)}[n] = 2sS_B^{(\frac{1}{2})}[n]$, as we claimed.

Semi-classical spectrum. Above we found a path integral representation for the Green's function of a spin as a function of time, $G(n_t, n_0; t)$. The information this contains about the spectrum of the hamiltonian can be extracted by Fourier transforming

$$G(n_t, n_0; E) \equiv -\mathbf{i}\int_0^\infty dt G(n_t, n_0; t)e^{\mathbf{i}(E+\mathbf{i}\epsilon)t}$$

and taking the trace

$$\Gamma(E) \equiv \int \frac{d^2n_0}{2\pi}G(n_0, n_0; E) = \operatorname{tr}\frac{1}{E - \mathbf{H} + \mathbf{i}\epsilon}.$$

This function has poles at the eigenvalues of $\mathbf{H}$. Its imaginary part is the spectral density, $\rho(E) = \frac{1}{\pi}\mathrm{Im}\,\Gamma(E) = \sum_\alpha \delta(E - E_\alpha)$.

The path integral representation is

$$\Gamma(E) = -\mathbf{i}\int dt \oint D\check{n}\; e^{\mathbf{i}((E+\mathbf{i}\epsilon)t + sS[n])}.$$

The $\oint$ indicates periodic boundary conditions, $\check{n}(0) = \check{n}(t)$, and $S[n] = S_B[n] - \int^t dt' H_{\mathrm{cl}}[n]/s$. Here $H_{\mathrm{cl}}[n] \equiv \langle\check{n}|\,\mathbf{H}\,|\check{n}\rangle$.

At large s, field configurations which vary too much in time are cancelled out by the rapidly oscillating phase, that is: we can try to do these integrals by stationary phase. The stationarity condition for the n integral is the equations of motion $0 = \dot{n}\times n - \partial_n H_{\mathrm{cl}}$. If $\mathbf{H} = \vec{h}\cdot\mathbf{S}$, this gives the Landau-Lifshitz equation (11.17) for precession. We keep only solutions periodic with $t = nT$ an integer multiple of the period T. The stationarity condition for the t integral is

$$0 = E + \partial_t S[n] = E - H_{\mathrm{cl}}[n].$$

[9] Sometimes (such as in lecture) you may see the notation $z_1 \equiv u, z_2 \equiv v$.

In the second equality we used the fact that the Berry phase is geometric, it depends only on the trajectory, not on t (how long it takes to get there). So the semiclassical trajectories are periodic solutions to the EOM with energy $E = H_{\text{cl}}[n^E]$. The exponent evaluated on such a trajectory is then just the Berry term. Denoting by n_1^E such trajectories which traverse once ('prime' orbits),

$$\Gamma(E) \sim \sum_{n_1^E} \sum_{n=0}^{\infty} e^{\mathbf{i}nsS_B[n]} = \sum_{n_1^E} \frac{e^{\mathbf{i}nsS_B[n]}}{1 - e^{\mathbf{i}nsS_B[n]}}.$$

This is an instance of the Gutzwiller trace formula. The locations of poles of this function approximate the eigenvalues of $\mathbf{H}$. They occur at $E = E_{sc}^m$ such that $S_B[\vec{n}^{E_m}] = \frac{2\pi m}{s}$. The actual eigenvalues are $E^m = E_{sc}^m + \mathcal{O}(1/s)$.

If the path integral in question were a 1d particle in a potential, with $S_B = \int p dx$, and $H_{\text{cl}} = p^2 + V(x)$, the semiclassical condition would reduce to

$$2\pi m = \oint_{x^{E_m}} p(x) dx = \int_{\text{turning points}} \sqrt{E_m - V(x)}$$

the Bohr-Sommerfeld condition.

[End of Lecture 45]

11.3.2 Ferromagnets and antiferromagnets.

[Zee §6.5] Now we'll try $D \geq 1+1$. Consider a chain of spins, each of spin $s \in \mathbb{Z}/2$, interacting via the Heisenberg hamiltonian:

$$\mathbf{H} = \sum_j J \vec{\mathbf{S}}_j \cdot \vec{\mathbf{S}}_{j+1}.$$

This hamiltonian is invariant under global spin rotations, $\mathbf{S}_j^a \to \mathcal{R}\mathbf{S}_j^a\mathcal{R}^{-1} = R_b^a \mathbf{S}_j^b$ for all j. For $J < 0$, this interaction is ferromagnetic, so it favors a state like $\left\langle \vec{\mathbf{S}}_j \right\rangle = s\hat{z}$. For $J > 0$, the neighboring spins want to anti-align; this is an antiferromagnet: $\left\langle \vec{\mathbf{S}}_j \right\rangle = (-1)^j s\hat{z}$. Note that I am lying about there being spontaneous breaking of a continuous symmetry in 1+1 dimensions. Really there is only short-range order because of the Coleman-Mermin-Wagner theorem. But that is enough for the calculation we want to do.[10]

[10] Even more generally, the consequence of short-range interactions of some particular sign for the groundstate is not so obvious. For example, antiferromagnetic interactions may be *frustrated*: If I want to disagree with both Kenenisa and Lasse, and Kenenisa and Lasse want to disagree with each other, then some of us will have to agree, or maybe someone has to withhold their opinion, $\langle S \rangle = 0$.

We can write down the action that we get by coherent-state quantization – it's just many copies of the above, where each spin plays the role of the external magnetic field for its neighbors:

$$L = \mathbf{i}s \sum_j z_j^\dagger \partial_t z_j - Js^2 \sum_j \vec{n}_j \cdot \vec{n}_{j+1}.$$

Spin waves in ferromagnets. Let's use this to find the equation of motion for small fluctuations $\delta\vec{n}_i = \vec{S}_i - s\hat{z}$ about the ferromagnetic state. Once we recognize the existence of the Berry phase term, this is the easy case. In fact the discussion is not restricted to $D = 1+1$. Assume the system is translation invariant, so we should Fourier transform. The condition that $\vec{n}_j^2 = 1$ means that $\delta n_z(k) = 0$.[11] Linearizing in $\delta\vec{n}$ (using (11.18)) and fourier transforming, we find

$$0 = \begin{pmatrix} h(k) & -\frac{\mathbf{i}}{2}\omega \\ \frac{\mathbf{i}}{2}\omega & h(k) \end{pmatrix} \begin{pmatrix} \delta n_x(k) \\ \delta n_y(k) \end{pmatrix}$$

with $h(k)$ determined by the exchange (J) term. It is the lattice laplacian in k-space. For example for the square lattice, it is $h(k) = 4s|J|\left(2 - \cos k_x a - \cos k_y a\right) \overset{k\to 0}{\simeq} 2s|J|a^2k^2$, with a the lattice spacing. For small k, the eigenvectors have $\omega \sim k^2$, a $z = 2$ dispersion (meaning that there is scale invariance near $\omega = k = 0$, but space and time scale differently: $k \to \lambda k, \omega \to \lambda^2\omega$. The two spin polarizations have their relative phases locked $\delta n_x(k) = \mathbf{i}\delta n_y(k)/h_k$, and so these modes describe precession of the spin about the ordering vector. These low-lying spin excitations are visible in neutron scattering and they dominate the low-temperature thermodynamics. Their thermal excitations produce a version of the blackbody spectrum with $z = 2$. We can determine the generalization of the Stefan-Boltzmann law by dimensional analysis: the free energy (or the energy itself) is extensive, so $F \propto L^d$, but it must have dimensions of energy, and the only other scale available is the temperature. With $z \neq 1$, temperature scales like $[T] = [L^{-z}]$. Therefore $F = cL^dT^{\frac{d+1}{z}}$. (For $z = 1$ this is the ordinary Stefan-Boltzmann law).

Notice that a ferromagnet is a bit special because the order parameter $Q^z = \sum_i \mathbf{S}_i^z$ is actually *conserved*, $[Q^z, \mathbf{H}] = 0$. This is actually what's responsible for the funny $z = 2$ dispersion of the goldstones, and the fact that although the groundstate breaks *two* generators Q^x and Q^y, there is only one gapless mode. If you are impatient to understand this connection, take a look at this paper.

[11] $1 = n_j^2 \forall j \implies n_j \cdot \delta n_j = 0, \forall j$ which means that for any k,

$$0 = \sum_j e^{\mathbf{i}kja} n_j \cdot \delta n_j = \sum_q n^z(k-q)\delta n^z(q) = \delta n_k^z.$$

Antiferromagnets. [Fradkin, 2d ed, p. 203] Now, let's study instead the equation of motion for small fluctuations about the antiferromagnetic state. The conclusion will be that there is a linear dispersion relation. This would be the conclusion we came to if we simply erased the WZW/Berry phase term and replaced it with an ordinary kinetic term

$$\frac{1}{2g^2}\sum_j \partial_t \vec{n}_j \cdot \partial_t \vec{n}_j \ .$$

How this comes about is actually a bit more involved! An important role will be played[12] by the *ferro*magnetic fluctuation $\vec{\ell}_j$ in

$$\vec{n}_j = (-1)^j \vec{m}_j + a\vec{\ell}_j \ .$$

$\vec{m}_j$ is the AF fluctuation; a is the lattice spacing; $s \in \mathbb{Z}/2$ is the spin. The constraint $\vec{n}^2 = 1$ tells us that $\vec{m}^2 = 1$ and $\vec{m}\cdot\vec{\ell} = 0$.

Why do we have to include both variables? Because $\vec{m}$ are the AF order-parameter fluctuations, but the total spin is conserved, and therefore its local fluctuations $\vec{\ell}$ still constitute a slow mode. This is an illustration of a general point: amongst the low-energy modes in our effective field theory, we should make sure we keep track of the conserved quantities, which can often move around but can never disappear. The name for this principle is *hydrodynamics*.

The exchange (J) term in the action is (using $\vec{n}_{2r}-\vec{n}_{2r-1} \approx a\left(\partial_x \vec{m}_{2r} + 2\ell_{2r}\right) + \mathcal{O}(a^2)$)

$$S_J[\vec{n}_j = (-1)^j\vec{m}_j + a\vec{\ell}_j] = -aJs^2 \int \mathrm{d}x\mathrm{d}t \left(\frac{1}{2}\left(\partial_x \vec{m}\right)^2 + 2\ell^2\right) \ .$$

The WZW terms evaluate to[13]

$$S_W = s\sum_{j=1}^N W_0[(-1)^j m_j + \ell_j] \overset{N\to\infty, a\to 0, Na \text{ fixed}}{\simeq} \int \mathrm{d}x\mathrm{d}t \ \left(\frac{s}{2}\vec{m}\cdot(\partial_t\vec{m}\times\partial_x\vec{m}) + s\vec{\ell}\cdot(\vec{m}\times\partial_t\vec{m})\right) .$$

[12] A pointer to the future: this story is very similar to the origin of the second order kinetic term for the Goldstone mode in a superfluid arises. The role of $\vec{\ell}$ there is played by ρ, the density. Naturally, we will discuss this when we do coherent state quantization of bosons in §11.5.

[13] The essential ingredient is

$$\delta W_0[n] = \int \mathrm{d}t \delta\vec{n}\cdot(\vec{n}\times\partial_t\vec{n}) \ .$$

So

$$W_0[n_{2r}] - W[n_{2r-1}] = -\frac{1}{2}\frac{\mathrm{d}x}{a}\frac{\delta W_0}{\delta n^i}\partial_x \hat{n}^i a = -\frac{1}{2}\mathrm{d}x\hat{n}\times\partial_t\hat{n}\cdot\partial_x\hat{n}.$$

Altogether, we find that ℓ is an auxiliary field with no time derivative:

$$L[m,\ell] = -2aJs^2\vec{\ell}^2 + s\vec{\ell}\cdot(\vec{m}\times\partial_t\vec{m}) + L[m]$$

so we can integrate out ℓ (this is the step analogous to what we'll do for ρ in the EFT of SF in §11.5) to find

$$S[\vec{m}] = \int \mathrm{d}x\mathrm{d}t\left(\frac{1}{2g^2}\left(\frac{1}{v_s}\left(\partial_t\vec{m}\right)^2 - v_s\left(\partial_x\vec{m}\right)^2\right) + \frac{\theta}{8\pi}\epsilon_{\mu\nu}\vec{m}\cdot\left(\partial_\mu\vec{m}\times\partial_\nu\vec{m}\right)\right), \quad (11.24)$$

with $g^2 = \frac{2}{s}$ and $v_s = 2aJs$, and $\theta = 2\pi s$. The equation of motion for small fluctuations of $\vec{m}$ therefore gives linear dispersion with velocity v_s. Notice that there are two independent gapless modes. Some of these fluctuations have wavenumber k close to π, since they are fluctuations of the AF order ($k = \pi$ means changing sign between each site), that is, $\omega \sim |k-\pi|$. (For a more microscopic treatment, see the book by Auerbach.)

The last ('theta') term in (11.24) is a total derivative. This means it doesn't affect the EOM, and it doesn't affect the Feynman rules. It is even more topological than the WZW term – its value only depends on the topology of the field configuration, and not on local variations. It is like the $\theta F\wedge F$ term in 4d gauge theory. You might think then that it doesn't matter. Although it doesn't affect small fluctuations of the fields, it does affect the path integral. Where have we seen this functional before? The integrand is the same as in our 2d representation of the WZW term in 0+1 dimensions: the object multiplying theta counts the winding number of the field configuration $\vec{m}$, the number of times Q the map $\vec{m}:\mathbb{R}^2\to S^2$ covers its image (we can assume that the map $\vec{m}(|x|\to\infty)$ approaches a constant, say the north pole). We can break up the path integral into sectors, labelled by this number $Q \equiv \frac{1}{8\pi}\int \mathrm{d}x\mathrm{d}t\ \epsilon_{\mu\nu}\vec{m}\cdot(\partial_\mu\vec{m}\times\partial_\nu\vec{m})$:

$$Z = \int[D\vec{m}]e^{\mathbf{i}S} = \sum_{Q\in\mathbb{Z}}\int[D\vec{m}]_Q e^{\mathbf{i}S_{\theta=0}}e^{\mathbf{i}\theta Q}\ .$$

θ determines the relative phase of different topological sectors (for $\theta=\pi$, this a minus sign for odd Q).

Actually, the theta term makes a huge difference. (Perhaps it is not so surprising if you think about the quantum mechanics of a particle constrained to move on a ring with magnetic flux through it?) The model with even s flows to a trivial theory in the IR, while the model with odd s flows to a nontrivial fixed point, called the $\mathrm{SU}(2)_1$ WZW model. It can be described in terms of one free relativistic boson. If you are impatient to understand more about this, the 2^{nd} edition of the book by Fradkin continues this discussion. Perhaps I can be persuaded to say more.

[End of Lecture 46]

Nonlinear sigma models in perturbation theory. Let us discuss what happens in perturbation theory in small g. A momentum-shell calculation integrating out fast modes (see the next subsection, §11.3.3) shows that

$$\frac{dg^2}{d\ell} = (D-2)g^2 + (n-2)K_D g^4 + \mathcal{O}(g^3) \tag{11.25}$$

where ℓ is the logarithmic RG time, and $\ell \to \infty$ is the IR. n is the number of components of $\hat{n}$, here $n = 3$, and $K_D = \frac{\Omega_{D-1}}{(2\pi)^D}$ as usual. Cultural remark: the second term is proportional to the *curvature* of the target space, here S^{n-1}, which has positive curvature for $n > 1$. For $n = 2$, we get S^1 which is one-dimensional and hence flat and there is no perturbative beta function. In fact, for $n = 2$, it's a free massless scalar. (But there is more to say about this innocent-looking scalar!)

The fact that the RHS of (11.25) is positive in $D = 2$ says that this model is *asymptotically free* – the coupling is weak in the UV (though this isn't so important if we are starting from a lattice model) and becomes strong in the IR. This is opposite what happens in QED; the screening of the charge in QED makes sense in terms of polarization of the vacuum by virtual charges. Why does this antiscreening happen here? There's a nice answer: the effect of the short-wavelength fluctuations is to make the spin-ordering vector $\vec{n}$ effectively *smaller*. It is like what happens when you do the block spin procedure, only this time don't use majority rule, but just average the spins. But rescaling the variable $\vec{n} \to a\vec{n}$ with $a \lesssim 1$ is the same as rescaling the coupling $g \to g/a$ – the coupling gets bigger. (Beware Peskin's comments about the connection between this result and the Coleman-Mermin-Wagner theorem: it's true that the logs in 2d enhance this effect, but in fact the model can reach a fixed point at finite coupling; in fact, this is what happens when $\theta = \pi$.)

Beyond perturbation theory. Like in QCD, this *infrared slavery* (the dark side of asymptotic freedom) means that we don't really know what happens in the IR from this calculation. From other viewpoints (Bethe ansatz solutions, many other methods), we know that (for integer s) there is an energy gap above the groundstate (named after Haldane) of order

$$\Lambda_H \sim \Lambda_0 e^{-\frac{c}{g_0^2}},$$

analogous to the QCD scale. Here g_0 is the value of g at the scale Λ_0; so Λ_H is roughly the energy scale where g becomes large. This is dimensional transmutation again.

For $s \in \mathbb{Z}$, for studying bulk properties like the energy gap, we can ignore the theta term since it only appears as $e^{2\pi \mathbf{i} n}$, with $n \in \mathbb{Z}$ in the path integral. [14] For half-integer

[14] $\theta = 2\pi n$ does, however, affect other properties, such as the groundstate wavefunction and the behavior in the presence of a boundary. $\theta = 2\pi$ is actually a different phase of matter than $\theta = 0$. It is an example of a SPT (symmetry-protected topological) phase, the first one discovered. See the homework for more on this.

s, there is destructive interference between the topological sectors. Various results (such as the paper by Read and Shankar, Nuclear Physics B336 (1990) 457-474, which contains an amazingly apt Woody Allen joke) show that this destroys the gap. This last sentence was a bit unsatisfying; more satisfying would be to understand the origin of the gap in the $\theta = 2\pi n$ case, and show that this interference removes that mechanism. This strategy is taken in this paper by Affleck.

11.3.3 The beta function for 2d non-linear sigma models

[Polyakov §3.2; Peskin §13.3; Auerbach chapter 13] I can't resist explaining the result (11.25). Consider this action for a $D = 2$ non-linear sigma model with target space S^{n+1}, of radius R:

$$S = \int \mathrm{d}^2 x R^2 \partial_\mu \hat{n} \cdot \partial^\mu \hat{n} \equiv \int \mathrm{d}^2 x R^2 \mathrm{d}n^2.$$

Notice that R is a coupling constant (it's what I called $1/g$ earlier). In the second step I made some compact notation.

Since not all of the components of $\hat{n}$ are independent (recall that $\hat{n} \cdot \hat{n} = 1$!), the expansion into slow and fast modes here is a little trickier than in our previous examples. Following Polyakov, let

$$n^i(x) \equiv n^i_<(x)\sqrt{1-\phi^2_>} + \sum_{a=1}^{n-1} \phi^>_a(x) e^i_a(x). \tag{11.26}$$

Here the slow modes are represented by the unit vector $n^i_<(x)$, $\hat{n}_< \cdot \hat{n}_< = 1$; the variables e^i_a are a basis of unit vectors spanning the $n-1$ directions perpendicular to $\vec{n}_<(x)$

$$n_< \cdot \hat{e}_a = 0, \hat{e}_a \cdot \hat{e}_a = 1; \tag{11.27}$$

they are not dynamical variables and how we choose them does not matter.

The fast modes are encoded in $\phi^>_a(x) \equiv \int_{\Lambda/s}^{\Lambda} đk e^{\mathbf{i}kx} \phi_k$, which only has fourier modes in a shell of momenta, and $\phi^2_> \equiv \sum_{a=1}^{n-1} \phi^>_a \phi^>_a$. Notice that differentiating the relations in (11.27) gives

$$\hat{n}_< \cdot \mathrm{d}\hat{n}_< = 0, \quad \hat{n}_< \cdot \mathrm{d}\hat{e}_a + \mathrm{d}\hat{n}_< \cdot \hat{e}_a = 0. \tag{11.28}$$

Below when I write ϕs, the $>$ symbol is implicit.

We need to plug the expansion (11.26) into the action, whose basic ingredient is

$$\mathrm{d}n^i = \mathrm{d}n^i_< \left(1-\phi^2\right)^{\frac{1}{2}} - n^i_< \frac{\phi \cdot \mathrm{d}\phi}{\sqrt{1-\phi^2}} + \mathrm{d}\phi \cdot e^i + \phi \cdot \mathrm{d}e^i.$$

So $S_{\text{eff}} = \int d^2x \, \mathcal{L}$ with

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2g^2} (\mathrm{d}\vec{n})^2 \\
&= \frac{1}{2g^2} \Bigg((\mathrm{d}n_<)^2 \left(1 - \phi^2\right) + \underbrace{\mathrm{d}\phi^2}_{\text{kinetic term for } \phi} + 2\phi_a \mathrm{d}\phi_b \vec{e}_a \cdot \mathrm{d}\vec{e}_b \\
&\qquad + \underbrace{\mathrm{d}\phi_a \mathrm{d}\vec{n}_< \cdot \vec{e}_a}_{\text{source for } \phi} + \phi_a \phi_b \mathrm{d}\vec{e}_a \cdot \mathrm{d}\vec{e}_b + \mathcal{O}(\phi^3) \Bigg)
\end{aligned} \tag{11.29}$$

So let's do the integral over ϕ, by treating the $\mathrm{d}\phi^2$ term as the kinetic term in a gaussian integral, and the rest as perturbations:

$$e^{-S_{\text{eff}}[n_<]} = \int [D\phi_>]^{\Lambda}_{\Lambda/s} e^{-\int L} = \int [D\phi_>]^{\Lambda}_{\Lambda/s} e^{-\frac{1}{2g^2}\int (\mathrm{d}\phi)^2} \,(\text{all the rest}) \equiv \langle \text{all the rest} \rangle_{>,0} \, Z_{>,0} \;.$$

The $\langle ... \rangle_{>,0}$s that follow are with respect to this measure.

$$\implies L_{\text{eff}}[n_<] = \frac{1}{2g^2} (\mathrm{d}n_<)^2 \left(1 - \left\langle \phi^2 \right\rangle_{>,0}\right) + \langle \phi_a \phi_b \rangle_{>,0} \mathrm{d}\vec{e}_a \cdot \mathrm{d}\vec{e}_b + \text{terms with more derivatives}$$

$$\langle \phi_a \phi_b \rangle_{>,0} = \delta_{ab} g^2 \int_{\Lambda/s}^{\Lambda} \frac{\mathrm{d}^2 k}{k^2} = g^2 K_2 \log(s) \delta_{ab}, \qquad K_2 = \frac{1}{2\pi}.$$

What to do with this $\mathrm{d}\vec{e}_a \cdot \mathrm{d}\vec{e}_b$ nonsense? Remember, $\vec{e}_a$ are just some arbitrary basis of the space perpendicular to $\hat{n}_<$; its variation can be expanded in our ON basis at x, $(n_<, e_c)$ as

$$\mathrm{d}\vec{e}_a = (\mathrm{d}e_a \cdot \hat{n}_<)\, \hat{n}_< + \sum_{c=1}^{n-1} \underbrace{(\mathrm{d}\vec{e}_a \cdot \vec{e}_c)}_{\overset{(11.28)}{=} -\mathrm{d}\hat{n}_< \cdot \vec{e}_a} \; \vec{e}_c$$

Therefore

$$\mathrm{d}\vec{e}_a \cdot \mathrm{d}\vec{e}_a = +\, (\mathrm{d}n_<)^2 + \sum_{c,a} (\vec{e}_c \cdot \mathrm{d}\vec{e}_a)^2$$

where the second term is a higher-derivative operator that we can ignore for our present purposes. Therefore

$$\begin{aligned}
L_{\text{eff}}[n] &= \frac{1}{2g^2} (\mathrm{d}\hat{n}_<)^2 \left(1 - ((N-1) - 1)\, g^2 K_2 \log s\right) + ... \\
&\simeq \frac{1}{2} \left(g^2 + \frac{g^4}{4\pi} (N-2) \log s + ... \right)^{-1} (\mathrm{d}\hat{n}_<)^2 + ...
\end{aligned} \tag{11.30}$$

Differentiating this running coupling with respect to s gives the one-loop term in the beta function quoted above. The tree-level (order g^2) term comes from engineering dimensions.

11.3.4 $\mathbb{CP}^1$ representation and Large-N

[Auerbach, *Interacting Electrons and Quantum Magnetism*, Polyakov, *Gauge fields and strings*] Above we used large spin as our small parameter to try to control the contributions to the path integral. Here we describe another route to a small parameter, which can be just as useful if we're interested in small spin (like spin-$\frac{1}{2}$).

Recall the relationship between the coherent state vector $\check{n}$ and the spinor components z: $n^a = z^\dagger \sigma^a z$. Imagine doing this at each point in space and time:

$$n^a(x) = z^\dagger(x)\sigma^a z(x). \tag{11.31}$$

We saw that the Berry phase term could be written nicely in terms of z as $\mathbf{i}z^\dagger \dot{z}$, what about the rest of the path integral?

First, some counting: $1 = \check{n}^2 \Leftrightarrow 1 = z^\dagger \cdot z = \sum_{m=\uparrow,\downarrow} |z_m|^2$. But this leaves only two components of n, and three components of z_m. The difference is made up by the fact that the rephasing

$$z_m(x) \to e^{\mathbf{i}\chi(x)} z_m(x) \tag{11.32}$$

doesn't change $\check{n}$. So it can't act on the physical Hilbert space. This is a (local, since $\chi(x)$ depends on x) $\mathsf{U}(1)$ gauge redundancy of the description in terms of z.

There two ways to proceed from here. One is via exact path integral tricks which are relatively straightforward in this case, but generally unavailable. The second is by the Landau method of knowing the answer: what else could it be.

Path integral manipulations. [Auerbach, chapter 14] First notice that the AF kinetic term is

$$\partial_\mu n^a \partial^\mu n^a = 4\left(\partial_\mu z^\dagger \partial^\mu z - \mathcal{A}_\mu \mathcal{A}^\mu\right) = 4\left(\partial_\mu z^\dagger \partial^\mu z - \mathcal{A}_\mu \mathcal{A}^\mu z^\dagger z\right). \tag{11.33}$$

where $\mathcal{A}_\mu \equiv -\frac{\mathbf{i}}{2}\left(z^\dagger \partial_\mu z - \partial_\mu z^\dagger z\right)$ is a connection one-form made from z itself. Notice that $\mathcal{A}_\mu \to \mathcal{A}_\mu - \partial\chi$ and the BHS of (11.33) is gauge invariant under (11.32). We must impose the constraint $|z(x)|^2 = 1$ at each site, which let's do it by a lagrange muliptlier $\delta[|z|^2 - 1] = \int D\lambda\ e^{\mathbf{i}\int d^dx \lambda(x)\left(|z|^2 - 1\right)}$. In the action, the $\mathcal{A}^2$ term is a self-interaction of the zs, which makes it difficult to do the integral. The standard trick for ameliorating this problem is the Hubbard-Stratonovich identity:

$$e^{c\mathcal{A}_\mu^2} = \sqrt{\frac{c}{\pi}} \int dA_\mu\ e^{-cA_\mu^2 + 2cA_\mu \mathcal{A}^\mu}.$$

The saddle point value of A is $\mathcal{A}$. This gives

$$e^{-\#\int dn^2} = \int [dA] e^{-\#\int |(\partial - \mathbf{i}A)z|^2}.$$

Finally, let's think about the measure at each point: $\int d^2n\delta(n^2-1)... = \int_0^\pi \sin\theta d\theta \int_0^{2\pi} d\varphi....$ Compare this to the integral over zs, parametrized as $z = \begin{pmatrix} \rho_1 e^{\mathbf{i}\phi_1} \\ \rho_2 e^{\mathbf{i}\phi_2} \end{pmatrix} = \begin{pmatrix} \cos\frac{\theta}{2} e^{\mathbf{i}\varphi/2} e^{\mathbf{i}\chi/2} \\ \sin\frac{\theta}{2} e^{-\mathbf{i}\varphi/2} e^{\mathbf{i}\chi/2} \end{pmatrix}$:

$$\int dz dz^\dagger \delta(|z|^2-1)... = \int \prod_{m=1,2} \rho_m d\rho_m d\phi_m \delta(\rho_1+\rho_2-1)...$$
$$= c\int d\rho\rho\sqrt{1-\rho^2}d\varphi d\chi ... = c'\int \sin\frac{\theta}{2}\cos\frac{\theta}{2} d\theta d\varphi d\chi...$$

which is the same as $\int dn$ except for the extra integral over χ: that's the gauge direction. The integral over χ is just a number at each point, as long as we integrate invariant objects (otherwise, it gives zero). Thinking of z as parametrizing an arbitrary normalized spinor $z = \mathcal{R}(\theta,\varphi,\chi)\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, so that $\mathcal{R}$ is an arbitrary element of $\mathsf{SU}(2)$, we've just shown the geometric equivalence between the round S^2 and $\mathbb{CP}^1 = \mathsf{SU}(2)/\mathsf{U}(1)$.

$$Z_{S^2} \simeq \int [dz dz^\dagger dA d\lambda] e^{-\int d^Dx \left(\frac{2\Lambda^{D-2}}{g^2}|(\partial - \mathbf{i}A)z|^2 - \lambda\left(|z|^2-1\right)\right)}. \qquad (11.34)$$

This is a $\mathsf{U}(1)$ gauge theory with $N=2$ charged scalars. It is called the $\mathbb{CP}^1$ sigma model. There are two slightly funny things: (1) the first is that the gauge field A lacks a kinetic term: in the microscopic description we are making here, it is infinitely strongly coupled. We'll see what the interactions with matter have to say about the coupling in the IR. (2) The second funny thing is that the scalars z have a funny interaction with this field λ which only appears linearly. If we add a λ^2/κ quadratic term, we can do the lambda integral and find $V(|z|^2) = \kappa(|z|^2-1)^2$, an ordinary quartic potential for $|z|$. This has the effect of replacing the delta function imposition with an energetic recommendation that $|z|^2 = 1$. This is called a *soft* constraint, and it shouldn't change the universal physics.

Alternatively, we could have arrived at this point

$$Z_{S^2} \simeq \int [dz dz^\dagger dA] e^{-\int d^Dx \left(\frac{2\Lambda^{D-2}}{g^2}|(\partial - \mathbf{i}A)z|^2 - \kappa\left(|z|^2-1\right)^2\right)}$$

by regarding (11.31) as a *slave-particle* or *parton* ansatz for a new set of variables. The demand of gauge invariance (11.32) is a strong constraint on the form of the interactions, and requires the inclusion of the gauge field A.

Other such ansatze are possible, such as one in terms of slave fermions $\vec{S} = \psi^\dagger\vec{\sigma}\psi$. In this case, this turns out to be *also* correct. (More later, after we discuss anomalies.) More generally, any given ansatz may not be useful to describe the relevant physics.

[End of Lecture 47]

Large N. This representation allows the introduction of another possible small parameter, namely the number of components of z. Suppose instead of two components, it has N

$$\sum_{m=1}^{N} |z_m|^2 = \frac{N}{2},$$

and let's think about the resulting $\mathbb{CP}^{N-1}$ sigma model (notice that $\mathbb{CP}^{N-1}$ and S^N are different generalizations of S^2, in the sense that for $N \to 1$ they are both S^2):

$$\begin{aligned} Z_{\mathbb{CP}^{N-1}} &= \int [dzdz^\dagger dAd\lambda] e^{-\int d^Dx \left(\frac{2\Lambda^{D-2}}{g^2} |(\partial - \mathbf{i}A)z|^2 - \lambda\left(|z|^2 - N/2\right)\right)} \\ &= \int [dAd\lambda] e^{-NS[A,\lambda]} \overset{N \gg 1}{\simeq} Z' e^{-NS[\underline{A},\underline{\lambda}]}. \end{aligned}$$

The z-integral is gaussian in the representation (11.34) even for $N = 2$, but the resulting integrals over A, λ are then horrible, with action

$$S[A, \lambda] = \operatorname{tr} \ln \left(- (\partial - \mathbf{i}A)^2 + \lambda \right) - \frac{\mathbf{i}\Lambda^{D-2}}{g^2} \int \lambda.$$

The role of large N is to make those integrals well-peaked about their saddle point. The saddle point equations are solved by $\underline{A} = 0$ (though there may sometimes be other saddles where $\underline{A} \neq 0$, which break various discrete symmetries). This leaves us with

$$S[0, \lambda] = \int đ^D k \ln(k^2 + \lambda) - \frac{\mathbf{i}}{g^2} V \lambda$$

(where V is the number of sites, the volume of space, and I've assumed constant λ), which is solved by $\lambda = -\mathbf{i}\underline{\lambda}$ satisfying

$$\int \frac{đ^D k}{k^2 + \underline{\lambda}} = \frac{1}{g^2}.$$

The solution of this equation depends on the number of dimensions D.

$$D = 1: \quad \frac{1}{g^2 \Lambda} = \int \frac{đk}{k^2 + \underline{\lambda}} = \frac{1}{\sqrt{\underline{\lambda}}} \int \frac{đ\underline{k}}{\underline{k}^2 + 1} \quad \Longrightarrow \underline{\lambda} = g^4 \Lambda^2 .$$

This says that the mass of the excitations is $m = g^2$. Where did that come from? $D = 1$ means we are studying the quantum mechanics of a particle contrained to move on $\mathbb{CP}^{N-1}$:

$$H = \frac{1}{2g^2\Lambda} \partial_z \partial_{\bar{z}} + \left(|z|^2 - N/2 \right)^2 .$$

The groundstate is the uniform state $\langle z|\text{groundstate}\rangle = \Psi(z) = \frac{1}{\text{vol}}$. QM of finite number of degrees of freedom on a compact space has a gap above the groundstate. This gap is determined by the kinetic energy and naturally goes like $g^2\Lambda$.

$$D=2: \quad g^{-2} = \int \frac{đ^2 k}{k^2+\underline{\lambda}} = -\frac{1}{4\pi}\ln\frac{\underline{\lambda}}{\Lambda^2} \implies \underline{\lambda} = \Lambda^2 e^{-\frac{4\pi}{g^2}}.$$

This is the case with asymptotic freedom; here we see again that asymptotic freedom is accompanied by dimensional transmutation: the interactions have generated a mass scale

$$m = \Lambda e^{-\frac{2\pi}{g^2}}$$

which is parametrically (in the bare coupling g) smaller than the cutoff.

$$D=3: \quad \frac{\Lambda}{g^2} = \int \frac{đ^3 k}{k^2+\underline{\lambda}} \implies |\sqrt{\underline{\lambda}}| = \frac{\Lambda}{2}\left(\frac{2}{\pi} - \frac{4\pi}{g^2}\right).$$

Notice that for $D \geq 3$ there is a critical value of g below which there is no solution. That means symmetry breaking: the saddle point is at $\underline{\lambda} = m^2 = 0$, and the z-fields are gapless Goldstone modes. This doesn't happen in $D \leq 2$. The critical coupling occurs when $g_c^{-2} = \int \frac{đ^D k}{k^2} \simeq \frac{\Lambda^{D-2}}{D-2}$. The rate at which the mass goes to zero as $g \to g_c$ from above is

$$m^2 \simeq \Lambda^2 \left(\frac{g^2-g_c^2}{g_c^2}\right)^{\frac{2}{D-2}}.$$

This is a universal exponent. (For more on critical exponents from large-N calculations, see Peskin p. 464-465.)

A quantity we'd like to be able to compute for $N=2$ is $S^{+-}(x) \equiv \langle S^+(0)S^-(x)\rangle$. We can write this in terms of the coherent state variables using the identity

$$\mathbf{S}^a = \mathcal{N}_s \int dn\, |\check{n}\rangle\langle\check{n}|\, n^a, \quad (\mathcal{N}_s = \frac{(s+1)(2s+1)}{4\pi}).$$

(Up to the constant factor, this identity follows from $\mathsf{SU}(2)$ invariance. The constant can be checked by looking at a convenient matrix element of the BHS.) Then:

$$S^{+-}(x) = \langle (n^x + \mathbf{i}n^y)(0)(n^x - \mathbf{i}n^y)(x)\rangle\,.$$

Recalling that $n^x + \mathbf{i}n^y = z^\dagger \sigma^+ z = z_1^\star z_2$, we can generalize this to large N as the four-point function

$$S^{m\neq m'}(x) = \langle z_m^\star(0) z_{m'}(0) z_m^\star(x) z_{m'}(x)\rangle \overset{N\gg 1}{\simeq} |G(x)|^2$$

which *factorizes* at leading order in large N. This phenomenon (large-N factorization) that at large-N the correlations are dominated by the disconnected bits is general. (Let me postpone the diagrammatic argument for a bit.) The factors are:

$$G(x) = \frac{1}{Z}\int [dz] z^\dagger(0) z(x) e^{-\frac{2\Lambda^{d-2}}{g^2}\int đ^d k\left(|k|^2+\underline{\lambda}\right) z_k^\dagger z_k - \frac{NV\Lambda^{d-2}}{g^2}\underline{\lambda}}$$

$$= \int đ^d k \frac{e^{-\mathbf{i}kx}}{|k|^2 + \underline{\lambda}} \simeq \frac{1}{|x|^{\frac{d-1}{2}}} e^{-|x|\sqrt{\underline{\lambda}}}.$$

This says that the correlation length for the spins in $S^{m\neq m'}(x) \overset{x>\xi}{\simeq} \frac{1}{|x|^{d-1}} e^{-\xi|x|}$ is $\xi = \frac{1}{2\sqrt{\underline{\lambda}}}$ depends variously on d. In $D = 1$, it is $\xi = \Lambda/g^2$, so large-N predicts a gap, growing with g. In $D = 2$, the correlation length is $\xi = \Lambda^{-1} e^{+\frac{2\pi}{g^2}}$ In $D = 3$, the correlation length diverges as $g \to g_c$ $\xi = \Lambda^{-1} \left(\frac{2}{\pi} - \frac{4\pi}{g^2}\right)^{-1}$, signaling the presence of gapless modes, which we interpret as Goldstones.

Exercise. Check that the other components of the spin such as $S^z = |z^m|^2 - |z^{m'}|^2$ have the same falloff, as they must by $\mathsf{SU}(N)$ symmetry.

A dynamical gauge field emerges. Finally, let me show you that a gauge field emerges. Let's expand the action $S_{\text{eff}}[A, \lambda]$ about the saddle point at $A = 0, \lambda = \underline{\lambda} \equiv m^2$:

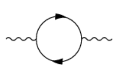
$$S[A = 0 + a, \lambda = m^2 + v] = W_0 + \underbrace{W_0}_{=0\text{by def}} + W_2 + \mathcal{O}(\delta^3)$$

where the interesting bit is the terms quadratic in the fluctuations:

$$W_2 = \frac{N}{2} \int đ^D q \left(v(q)\Pi(q)v(-q) + A_\mu(q)\Pi_{\mu\nu}(q)A_\nu(-q) \right)$$

where

$$\Pi(q) = ... = \int đ^D k \frac{1}{(k^2 + m^2)((k+q)^2 + m^2)} \tag{11.35}$$

$\Pi_{\mu\nu} =$  $+\text{diamagnetic diagram} = \int đ^D k \frac{(2k+q)_\mu (2k+q)_\nu}{(k^2 + m^2)((k+q)^2 + m^2)} - 2g_{\mu\nu} \int \frac{đ^D k}{k^2 + m^2}. \qquad (11.36)$

Familiarly, gauge invariance implies that $q^\mu \Pi_{\mu\nu}(q) = 0$ – it prevents a mass for the gauge field. For example, in $D = 2$, the long wavelength behavior is

$$\Pi_{\mu\nu}(q) \overset{q\to 0}{\sim} \frac{c}{m^2} \left(q^2 g_{\mu\nu} - q_\mu q_\nu \right)$$

which means that the effective action for the gauge fluctuation is

$$W_2 \sim \frac{N}{m^2} \int d^2 x F_{\mu\nu} F^{\mu\nu} + \text{more derivatives}.$$

It is a dynamical gauge field.

Another term we can add to the action for a 2d gauge field is

$$\theta \int \frac{F}{2\pi}$$

where we regard $F = dA$ as a two-form. This is the 2d theta term, analogous to $\int F \wedge F$ in $D = 4$ in that $F = dA$ is locally a total derivative, it doesn't affect the equations of motion, and it integrates to an integer on smooth configurations (we will show this when we study anomalies). This integer is called the *Chern number* of the gauge field configuration. What integer is it? On the homework you'll show that $F \propto \epsilon^{abc} n^a dn^b dn^c$. It's the skyrmion number! So the coefficient is $\theta = 2\pi s$.

11.3.5 Large-N diagrams.

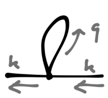
I think it will help to bring home some of the previous ideas by rederiving them using diagrams in a familiar context. So let's study the $\mathsf{O}(N)$ model:

$$L = \frac{1}{2}\partial\vec{\varphi}\cdot\partial\vec{\varphi} + \frac{g}{4N}\left(\vec{\varphi}\cdot\vec{\varphi}\right)^2 + \frac{m^2}{2}\vec{\varphi}\cdot\vec{\varphi}. \tag{11.37}$$

Let's do euclidean spacetime, D dimensions. The bare propagator is

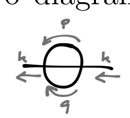
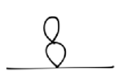
$$\langle \varphi_b(x)\varphi_a(0)\rangle = \int đ^D k\ \frac{e^{-\mathbf{i}kx}}{k^2+m^2} \equiv \int đ^D k\ \Delta_0(k).$$

The bare vertex is $-\frac{2g}{N}\left(\delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}\right)$. With this normalization, the leading correction to the propagator is



$$= -\frac{g}{4N}(4N+8)\delta_{ab}\int \frac{đq}{q^2+m^2} \overset{N\gg 1}{\simeq} -g\delta_{ab}\int đq\Delta_0(q)$$

of order N^0. This is the motivation for the normalization of the coupling in (11.37).

Which diagrams dominate at large N (and fixed g)? Compare two diagrams at the same order in λ with different topology of the index flow: eyeball  and cactus . The former has one index loop, and the latter has two, and therefore dominates. The general pattern is that: at large N cacti dominate the 1PI self-energy. Each extra pod we add to the cactus costs a factor of g/N but gains an index loop N. So the sum of cacti is a function of gN^0.

The full propagator, by the usual geometric series, is then

$$\Delta_F(k) = \frac{1}{k^2+m^2+\Sigma(k)}. \tag{11.38}$$

We can sum all the cacti by noticing that cacti are self-similar: if we replace Δ_0 by Δ_F in the propagator:

$$\Sigma(p) = g\int đ^D k\Delta_F(k) + \mathcal{O}(1/N). \tag{11.39}$$

The equations (11.38), (11.39) are integral equations for Δ_F; they are called Schwinger-Dyson equations,

OK, now notice the p-dependence in (11.39): the RHS is independent of p to leading order in N, so $\Sigma(p) = \delta m^2$ is just a mass shift.

Look at the position-space propagator

$$\langle \varphi_b(x)\varphi_a(y)\rangle = \delta_{ab}\int đ^D k e^{-\mathbf{i}k(x-y)}\Delta_F(k). \tag{11.40}$$

Let

$$y^2 \equiv \left\langle \frac{\sum_a \varphi_a(x)\varphi_a(x)}{N}\right\rangle = \left\langle \frac{\varphi^2}{N}\right\rangle;$$

it is independent of x by translation invariance. Now let $y \to x$ in (11.40):

$$y^2 = \int đ^D k \Delta_F(k) \overset{(11.39)}{=} g^{-1}\Sigma.$$

Now integrate the BHS of (11.38):

$$\int đ^D p \Delta_F(p) = \int đ^D \frac{1}{p^2+m^2+\Sigma}$$
$$y^2 = \int đ^D p \frac{1}{p^2+m^2+gy^2}.$$

This is an equation for the positive number y^2. Notice its similarity to the gap equation we found from saddle point.

[End of Lecture 48]

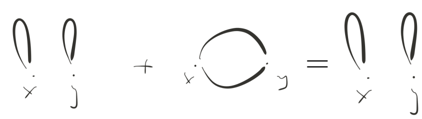
Large-N factorization. [Halpern] The fact that the fluctuations about the saddle point are suppressed by powers of N has consequences for the structure of the correlation functions in a large-N field theory. A basic example is

$$\langle \mathcal{I}(x)\mathcal{I}(y)\rangle = \langle \mathcal{I}(x)\rangle\langle \mathcal{I}(y)\rangle + \mathcal{O}\left(N^{-1}\right)$$

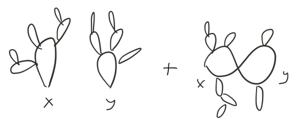
where $\mathcal{I}$ are any invariants of the large-N group (*i.e.* $\mathsf{O}(N)$ in the $\mathsf{O}(N)$ model (naturally) and $\mathsf{SU}(N)$ in the $\mathbb{CP}^{N-1}$ model), and $\langle ...\rangle$ denotes either euclidean vacuum expectation value or time-ordered vacuum expectation value. Consider, for example, in the $\mathsf{O}(N)$ model, normalized as above

$$\left\langle \frac{\varphi^2(x)}{N}\frac{\varphi^2(y)}{N}\right\rangle.$$

In the free theory, $g=0$, there are two diagrams

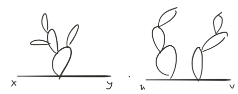
$$\left\langle \frac{\varphi^2(x)}{N}\frac{\varphi^2(y)}{N}\right\rangle_{\text{free}} =$$  $$+\,\mathcal{O}\left(N^{-1}\right)$$

– the disconnected diagram dominates, because it has one more index loop and the same number of interactions (zero). With interactions, representative diagrams are

$$\left\langle \frac{\varphi^2(x)}{N}\frac{\varphi^2(y)}{N} \right\rangle =$$

$$= \left\langle \frac{\varphi^2(x)}{N} \right\rangle \left\langle \frac{\varphi^2(y)}{N} \right\rangle + \mathcal{O}\left(N^{-1}\right) = y^4 + \mathcal{O}\left(N^{-1}\right)$$

– it is independent of $x - y$ to leading order.

The same phenomenon happens for correlators of non-local singlet operators:

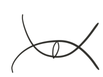
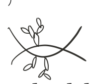
$$\left\langle \frac{\varphi(x)\cdot\varphi(y)}{N}\frac{\varphi(u)\cdot\varphi(v)}{N} \right\rangle =$$

$$= \left\langle \frac{\varphi(x)\cdot\varphi(y)}{N} \right\rangle \left\langle \frac{\varphi(u)\cdot\varphi(v)}{N} \right\rangle + \mathcal{O}\left(N^{-1}\right)$$

The basic statement is that mean field theory works for singlets. At large N, the entanglement follows the flavor lines.

We can still ask: what processes dominate the connected (small) bit at large N? And what about non-singlet operators? Consider (no sum on b, a):

$$G_{4,c}^{b\neq a} = \langle \varphi_b(p_4)\varphi_b(p_3)\varphi_a(p_2)\varphi_a(p_1)\rangle =$$

$$+ \mathcal{O}\left(N^{-2}\right)$$

The answer is: *bubbles*. More specifically chains of bubbles, propagating in the s-channel. What's special about the s-channel, here? It's the channel in which we can make $\mathsf{O}(N)$ singlets. Other candidates are eyeballs:  and ladders:  but as you can see, these go like N^{-2}. However, bubbles can have cactuses growing on them, like this:  To sum all of these, we just use the full propagator in the internal lines of the bubbles, $\Delta_0 \to \Delta_F$.

I claim that the bubble sum is a geometric series:

$$G_{4,c}^{b\neq a} = -\left(\Delta_0(\text{external})\right)^4 \frac{2}{N}\frac{g}{1 + gL(p_1 + p_2)} + \mathcal{O}\left(N^{-2}\right) \tag{11.41}$$

where L is the loop integral $L(p) \equiv \int đ^D k \Delta_F(k)\Delta_F(p + k)$. You can see this by being careful about the symmetry factors.


$$= \Delta_0\left(\text{external}\right)^4 \left(\frac{g}{4N}\right) \cdot 2 \cdot 4$$


$$= \Delta_0\left(\text{external}\right)^4 \left(\frac{g}{4N}\right)^2 \cdot 2\cdot 4\cdot 8 \cdot \underbrace{\frac{1}{2!}}_{Dyson} L = \Delta_0\left(\text{external}\right)^4 \frac{2}{N}\left(g\right)^2 L.$$

Similarly, the chain of two bubbles is $\frac{2}{N}g^3L^2$, etc.

Here's how we knew this had to work without worrying about the damn symmetry factors: the bubble chain *is* the σ propagator! At the saddle, $\sigma \simeq \varphi^a\varphi^a$, which is what is going in and out of this amplitude. And the effective action for sigma (after integrating out φ) is

$$S_{\text{eff}}[\sigma] = \int \frac{\sigma^2}{g} + \operatorname{tr}\ln\left(\partial^2 + m^2 + \sigma\right).$$

The connected two-point function means we subract of $\langle\underline{\sigma}\rangle\langle\underline{\sigma}\rangle$, which is the same as considering the two point function of the deviation from saddle value. This is

$$\langle\sigma_1\sigma_2\rangle = \left(\frac{\delta^2}{\delta\sigma_1\delta\sigma_2}S_{\text{eff}}[\sigma]\right)^{-1} = \left(\frac{1}{g^{-1} + \left(\frac{1}{\partial^2+m^2+\sigma}\right)^2}\right)^{-1}$$

which becomes exactly the expression above if we write it in momentum space.

Two comments: (1) We were pretty brash in integrating out all the φ variables and keeping the σ variable: how do we know which are the slow ones and which are the fast ones? This sort of non-Wilsonian strategy is common in the literature on large-N, where physicists are so excited to see an integral that they can actually do that they don't pause to worry about slow and fast. But if we did run afoul of Wilson, at least we'll know it, because the action for σ will be nonlocal.

(2) $\sigma \sim \varphi^2$ is a composite operator. Nevertheless, the sigma propagator we've just derived can have poles at some $p^2 = m^2$ (likely with complex m). These would produce particle-like resonances in a scattering experiment (such as $2-2$ scattering of φs of the same flavor) which involved sigmas propagating in the s-channel. Who is to say what is fundamental.

Now that you believe me, look again at (11.41); it is of the form

$$G^{b\neq a}_{4,c} = -\left(\Delta_0(\text{external})\right)^4 \frac{2}{N} g_{\text{eff}}(p_1+p_2) + \mathcal{O}\left(N^{-2}\right)$$

where now

$$g_{\text{eff}}(p) = \frac{g}{1 + g\int đ^D k \Delta_F(k)\Delta_F(p+k))}$$

is a momentum-dependent effective coupling, just like one dreams of when talking about the RG.

11.4 Coherent state path integrals for fermions

We'll need these for our discussion of anomalies, and if we ever get to perturbative QCD (which differs from Yang-Mills theory by the addition of fermionic quarks).

[Shankar, *Principles of QM*, path integrals revisited. In this chapter of his great QM textbook, Shankar sneaks in lots of insights useful for modern condensed matter physics]

Consider the algebra of a single fermion mode operator[15]:

$$\{\mathbf{c}, \mathbf{c}\} = 0, \quad \{\mathbf{c}^\dagger, \mathbf{c}^\dagger\} = 0, \quad \{\mathbf{c}, \mathbf{c}^\dagger\} = 1 \ .$$

With a single mode, the general Hamiltonian is

$$\mathbf{H} = \mathbf{c}^\dagger \mathbf{c} \, (\omega_0 - \mu)$$

(ω_0 and μ are (redundant when there is only one mode) constants). This algebra is represented on a two-state system $|1\rangle = \mathbf{c}^\dagger |0\rangle$. We might be interested in its thermal partition function

$$Z = \text{tr } e^{-\frac{\mathbf{H}}{T}} \ .$$

(In this example, it happens to equal $Z = 1 + e^{-\frac{\omega_0 - \mu}{T}}$, as you can see by computing the trace in the eigenbasis of $\mathbf{n} = \mathbf{c}^\dagger \mathbf{c}$. But never mind that; the one mode is a proxy for many, where it's not quite so easy to sum.) How do we trotterize this? That is, what is 'the' corresponding classical system? (One answer is to use the (0d) Jordan-Wigner map which relates spins and fermions. Perhaps more about that later. Here's another, different, answer.) We can do the Trotterizing using any resolution of the identity on $\mathcal{H}$, so there can be many very-different-looking answers to this question.

Let's define coherent states for fermionic operators:

$$\mathbf{c} |\psi\rangle = \psi |\psi\rangle \, . \tag{11.42}$$

Here ψ is a c-number (not an operator), but acting twice with $\mathbf{c}$ we see that we must have $\psi^2 = 0$. ψ is a grassmann number. These satisfy

$$\psi_1 \psi_2 = -\psi_2 \psi_1, \psi \mathbf{c} = -\mathbf{c} \psi \tag{11.43}$$

– they anticommute with each other and with fermionic operators, and commute with ordinary numbers and bosons. They seem weird but they are easy. We'll need to

[15] For many modes,

$$\{\mathbf{c}_i, \mathbf{c}_j\} = 0, \quad \{\mathbf{c}_j^\dagger, \mathbf{c}_j^\dagger\} = 0, \quad \{\mathbf{c}_j, \mathbf{c}_j^\dagger\} = \mathbb{1}\delta_{ij} \ .$$

consider multiple grassmann numbers when we have more than one fermion mode, where $\{\mathbf{c}_1, \mathbf{c}_2\} = 0$ will require that they anticommute $\{\psi_1, \psi_2\} = 0$ (as in the definition (11.43)); note that we will be simultaneously diagonalizing operators which *anti*commute.

The solution to equation (11.42) is very simple:

$$|\psi\rangle = |0\rangle - \psi |1\rangle$$

where as above $|0\rangle$ is the empty state ($\mathbf{c}|0\rangle = 0$) and $|1\rangle = \mathbf{c}^\dagger |0\rangle$ is the filled state. (Check: $\mathbf{c}|\psi\rangle = \mathbf{c}|0\rangle - \mathbf{c}\psi|1\rangle = +\psi\mathbf{c}|1\rangle = \psi|0\rangle = \psi|\psi\rangle$.)

Similarly, the left-eigenvector of the creation operator is

$$\langle\bar{\psi}| \mathbf{c}^\dagger = \langle\bar{\psi}| \bar{\psi}, \quad \langle\bar{\psi}| = \langle 0| - \langle 1| \bar{\psi} = \langle 0| + \bar{\psi} \langle 1| .$$

Notice that these states are weird in that they are elements of an enlarged hilbert space with grassmann coefficients (usually we just allow complex numbers). Also, $\bar{\psi}$ is *not* the complex conjugate of ψ and $\langle\bar{\psi}|$ is not the adjoint of $|\psi\rangle$. Rather, their overlap is

$$\langle\bar{\psi}|\psi\rangle = 1 + \bar{\psi}\psi = e^{\bar{\psi}\psi}.$$

Grassmann calculus summary. In the last expression we have seen an example of the amazing simplicity of Taylor's theorem for grassmann functions:

$$f(\psi) = f_0 + f_1\psi \; .$$

Integration is just as easy and its the same as taking derivatives:

$$\int \psi d\psi = 1, \qquad \int 1 d\psi = 0.$$

With more than one grassmann we have to worry about the order:

$$1 = \int \bar{\psi}\psi d\psi d\bar{\psi} = -\int \bar{\psi}\psi d\bar{\psi} d\psi.$$

The only integral, really, is the gaussian integral:

$$\int e^{-a\bar{\psi}\psi} d\bar{\psi} d\psi = a.$$

Many of these give

$$\int e^{-\bar{\psi}\cdot A\cdot\psi} d\bar{\psi} d\psi = \det A.$$

Here $\bar{\psi} \cdot A \cdot \psi \equiv (\bar{\psi}_1, \cdots, \bar{\psi}_M) \begin{pmatrix} A_{11} & A_{12} & \cdots \\ A_{21} & \ddots & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_M \end{pmatrix}$. One way to get this expression is to change variables to diagonalize the matrix A.

$$\langle \bar{\psi}\psi \rangle \equiv \frac{\int \bar{\psi}\psi e^{-a\bar{\psi}\psi} d\bar{\psi} d\psi}{\int e^{-a\bar{\psi}\psi} d\bar{\psi} d\psi} = -\frac{1}{a} = -\langle \psi\bar{\psi} \rangle .$$

If for many grassman variables we use the action $S = \sum_i a_i \bar{\psi}_i \psi_i$ (diagonalize A above) then

$$\langle \bar{\psi}_i \psi_j \rangle = \frac{\delta_{ij}}{a_i} \equiv \langle \bar{i} j \rangle \tag{11.44}$$

and Wick's theorem here is

$$\langle \bar{\psi}_i \bar{\psi}_j \psi_k \psi_l \rangle = \langle \bar{i} l \rangle \langle \bar{j} k \rangle - \langle \bar{i} k \rangle \langle \bar{j} l \rangle .$$

Back to quantum mechanics: The resolution of $\mathbb{1}$ in this basis is

$$\mathbb{1} = \int d\bar{\psi} d\psi \; e^{-\bar{\psi}\psi} |\psi\rangle \langle \bar{\psi}| \tag{11.45}$$

And if $\mathbf{A}$ is a bosonic operator (made of an even number of grassmann operators),

$$\mathrm{tr}\mathbf{A} = \int d\bar{\psi} d\psi \; e^{-\bar{\psi}\psi} \langle -\bar{\psi}| \mathbf{A} |\psi\rangle \;\; .$$

(Note the minus sign; it will lead to a deep statement.) So the partition function is:

$$Z = \int d\bar{\psi}_0 d\psi_0 \; e^{-\bar{\psi}_0\psi_0} \langle -\bar{\psi}_0| \underbrace{\underbrace{e^{-\frac{\mathbf{H}}{T}}}_{=(1-\Delta\tau\mathbf{H})\cdots(1-\Delta\tau\mathbf{H})}}_{M \text{ times}} |\psi_0\rangle$$

Now insert (11.45) in between each pair of Trotter factors to get

$$Z = \int \prod_{l=0}^{M-1} d\bar{\psi}_l d\psi_l e^{-\bar{\psi}_l\psi_l} \langle \bar{\psi}_{l+1}| (1-\Delta\tau\mathbf{H}) |\psi_l\rangle \;\; .$$

Because of the $-\bar{\psi}$ in (11.45), to get this nice expression we had to define an extra letter

$$\bar{\psi}_M = -\bar{\psi}_0, \quad \psi_M = -\psi_0 \tag{11.46}$$

so we could replace $\langle -\bar{\psi}_0| = \langle \bar{\psi}_M|$. [End of Lecture 49]

Now we use the coherent state property to turn the matrix elements into grassmann-valued functions:

$$\langle\bar\psi_{l+1}|\left(1-\Delta\tau H(\mathbf{c}^\dagger,\mathbf{c})\right)|\psi_l\rangle = \langle\bar\psi_{l+1}|\left(1-\Delta\tau H(\bar\psi_{l+1},\psi_l)\right)|\psi_l\rangle \overset{\Delta\tau\to 0}{=} e^{\bar\psi_{l+1}\psi_l}e^{-\Delta\tau H(\bar\psi_{l+1},\psi_l)}.$$

It was important that in $\mathbf{H}$ all $\mathbf{c}$s were to the right of all $\mathbf{c}^\dagger$s, *i.e.* that $\mathbf{H}$ was normal ordered.)

So we have

$$\begin{aligned} Z &= \int \prod_{l=0}^{M-1} d\bar\psi_l d\psi_l e^{-\bar\psi_l\psi_l}e^{\bar\psi_{l+1}\psi_l}e^{-\Delta\tau H(\bar\psi_{l+1},\psi_l)} \\ &= \int \prod_{l=0}^{M-1} d\bar\psi_l d\psi_l \exp\left(\Delta\tau\left(\underbrace{\frac{\bar\psi_{l+1}-\bar\psi_l}{\Delta\tau}}_{=\partial_\tau\bar\psi}\psi_l - H(\bar\psi_{l+1},\psi_l)\right)\right) \\ &\simeq \int [D\bar\psi D\psi]\exp\left(\int_0^{1/T} d\tau\ \bar\psi(\tau)\left(-\partial_\tau-\omega_0+\mu\right)\psi(\tau)\right) = \int [D\bar\psi D\psi]e^{-S[\bar\psi,\psi]}. \end{aligned} \tag{11.47}$$

Points to note:

- In the penultimate step we defined, as usual, continuum fields

$$\psi(\tau_l = \Delta\tau l) \equiv \psi_l, \qquad \bar\psi(\tau_l=\Delta\tau l) \equiv \bar\psi_l.$$

- We elided the difference $H(\bar\psi_{l+1},\psi_l) = H(\bar\psi_l,\psi_l)+\mathcal{O}(\Delta\tau)$ in the last expression. This difference is usually negligible and sometimes helpful (an example where it's helpful is the discussion of the number density below).

- The APBCs (11.46) on $\psi(\tau+\frac{1}{T}) = -\psi(\tau)$ mean that in its fourier representation[16]

$$\psi(\tau) = T\sum_n \psi(\omega)e^{-\mathbf{i}\omega_n\tau}, \qquad \bar\psi(\tau) = T\sum_n \bar\psi(\omega)e^{\mathbf{i}\omega_n\tau} \tag{11.48}$$

  the *Matsubara frequencies*

$$\omega_n = (2n+1)\pi T, \quad n\in\mathbb{Z}$$

  are half-integer multiples of πT.

- The measure $[D\bar\psi D\psi]$ is defined by this equation, just as in the bosonic path integral.

[16] $\bar\psi$ is still not the complex conjugate of ψ but the relative sign is convenient.

- The derivative of a grassmann function is also defined by this equation; note that $\psi_{l+1} - \psi_l$ is not 'small' in any sense.
- In the last step we integrated by parts, *i.e.* relabeled terms in the sum, so

  $$\sum_l \left(\bar{\psi}_{l+1} - \bar{\psi}_l\right) \psi_l = \sum_l \bar{\psi}_{l+1}\psi_l - \sum_l \bar{\psi}_l\psi_l = \sum_{l'=l-1} \bar{\psi}_{l'}\psi_{l-1} - \sum_l \bar{\psi}_l\psi_l = -\sum_l \bar{\psi}_l \left(\psi_l - \psi_{l-1}\right).$$

  Note that no grassmanns were moved through each other in this process.

The punchline of this discussion for now is that the euclidean action is

$$\boxed{S[\bar{\psi}, \psi] = \int d\tau \left(\bar{\psi}\partial_\tau\psi + H(\bar{\psi}, \psi)\right)}\ .$$

The first-order kinetic term we've found $\bar{\psi}\partial_\tau\psi$ is sometimes called a 'Berry phase term'. Note the funny-looking sign.

Continuum limit warning (about the red $\simeq$ in (11.47)). The Berry phase term is *actually*

$$\sum_{l=0}^{N-1} \bar{\psi}_{l+1}\left(\psi_{l+1} - \psi_l\right) = T \sum_{\omega_n} \bar{\psi}(\omega_n) \left(1 - e^{\mathbf{i}\omega_n\tau}\right) \psi(\omega_n)$$

and in (11.47) we have kept only the leading nonzero term:

$$\left(1 - e^{\mathbf{i}\omega_n\tau}\right) \to \mathbf{i}\omega_n\tau.$$

Clearly this replacement is just fine if

$$\omega_n\tau \ll 1$$

for all ω_n which matter. Which ω_n contribute? I claim that if we use a reasonable $\mathbf{H} = \mathbf{H}_{\text{quadratic}} + \mathbf{H}_{\text{int}}$, reasonable quantities like Z, $\left\langle \mathcal{O}^\dagger\mathcal{O}\right\rangle$, are dominated by $\omega_n \ll \tau^{-1}$.

There's more we can learn from what we've done here that I don't want to pass up. Let's use this formalism to compute the fermion density at $T = 0$:

$$\langle \mathbf{N} \rangle = \frac{1}{Z}\text{tr}e^{-\mathbf{H}/T}\mathbf{c}^\dagger\mathbf{c}.$$

This is an example where the annoying $\Delta\tau$s in the path integral not only matter, but are extremely friendly to us.

Frequency space, $T \to 0$.

Let's change variables to frequency-space fields, which diagonalize S. The Jacobian is 1 (since fourier transform is unitary):

$$D\bar{\psi}(\tau)D\psi(\tau) = \prod_n d\bar{\psi}(\omega_n)d\psi(\omega_n) \overset{T\to 0}{\to} D\bar{\psi}(\omega)D\psi(\omega).$$

The partition function is

$$Z = \int D\bar{\psi}(\omega)D\psi(\omega)\ \exp\left(T\sum_{\omega_n}\bar{\psi}(\omega_n)\left(\mathbf{i}\omega_n - \omega_0 + \mu\right)\psi(\omega_n)\right).$$

Notice that in the zero-temperature limit

$$T\sum_{\omega_n} \mapsto \int \frac{d\omega}{2\pi} \equiv \int đ\omega.$$

(This is the same fact as $V\sum_k \mapsto \int đ^d k$ in the thermodynamic limit.) So the zero-temperature partition function is

$$Z \overset{T\to 0}{=} \int D\bar{\psi}(\omega)D\psi(\omega)\ \exp\left(\int_{-\infty}^{\infty} đ\omega\bar{\psi}(\omega)\left(\mathbf{i}\omega - \omega_0 + \mu\right)\psi(\omega)\right).$$

Using the gaussian-integral formula (11.44) you can see that the *propagator* for ψ is

$$\left\langle\bar{\psi}(\omega_1)\psi(\omega_2)\right\rangle = \underbrace{\frac{\delta_{\omega_1,\omega_2}}{T}}_{\overset{T\to 0}{\to}\delta(\omega_1-\omega_2)} \frac{2\pi}{\mathbf{i}\omega_1 - \omega_0 + \mu}. \tag{11.49}$$

In particular $\left\langle\bar{\psi}(\omega)\psi(\omega)\right\rangle = \frac{2\pi/T}{\mathbf{i}\omega-\omega_0+\mu}$. $\delta(\omega = 0) = 1/T$ is the 'volume' of the time direction.

Back to the number density. Using the same strategy as above, we have

$$\langle\mathbf{N}\rangle = \frac{1}{Z}\int\prod_{l=0}^{M-1+1}\left(d\bar{\psi}_l d\psi_l e^{-\bar{\psi}_l\psi_l}\right)\prod_{l=1}^{M-1}\left\langle\bar{\psi}_{l+1}|(1-\Delta\tau\mathbf{H}(\mathbf{c}^\dagger\mathbf{c}))|\psi_l\right\rangle \underbrace{\left\langle\bar{\psi}_{N+1}\right|\mathbf{c}^\dagger\mathbf{c}\left|\psi_N\right\rangle}_{=\bar{\psi}_{N+1}\psi_N=\bar{\psi}(\tau_N+\Delta\tau)\psi(\tau_N)},$$

where τ_N is any of the time steps. This formula has a built-in point-splitting of the operators!

$$\begin{aligned}\langle\mathbf{N}\rangle &= \frac{1}{Z}\int D\bar{\psi}D\psi\ e^{-S[\bar{\psi},\psi]}\bar{\psi}(\tau_N+\Delta\tau)\psi(\tau_N)\\ &= \int_{-\infty}^{\infty}đ\omega\frac{e^{\mathbf{i}\omega\Delta\tau}}{\mathbf{i}\omega-\omega_0+\mu} = \theta(\mu-\omega_0).\end{aligned} \tag{11.50}$$

Which is the right answer: the mode is occupied in the groundstate only if $\omega_0 < \mu$. In the last step we used the fact that $\Delta\tau > 0$ to close the contour in the UHP; so we only pick up the pole if it is in the UHP. Notice that this quantity is very *UV sensitive*: if we put a frequency cutoff on the integral, $\int^\Lambda \frac{d\omega}{\omega} \sim \log \Lambda$, the integral diverges logarithmically. For most calculations the $\Delta\tau$ can be ignored, but here it told us the right way to treat the divergence. [17]

Where do topological terms come from? [Abanov ch 7] Here is a quick application of fermionic path integrals related to the previous subsection §11.3. Consider a 0+1 dimensional model of *spinful* fermions $\mathbf{c}_\alpha, \alpha = \uparrow, \downarrow$ coupled to a single spin s, $\vec{\mathbf{S}}$. Let's couple them in an $\mathsf{SU}(2)$-invariant way:

$$H_K = M \left(\mathbf{c}^\dagger \vec{\sigma} \mathbf{c} \right) \cdot \vec{\mathbf{S}}$$

by coupling the spin of the fermion $\mathbf{c}^\dagger_\alpha \vec{\sigma}_{\alpha\beta} \mathbf{c}_\beta$ to the spin. 'K' is for 'Kondo'. Notice that M is an energy scale. (Ex: find the spectrum of H_K.)

Now apply both of the previous coherent state path integrals that we've learned to write the (say euclidean) partition sum as

$$Z = \int [D\psi D\bar{\psi} D\vec{n}] e^{-S_0[n] - \int_0^T dt \bar{\psi}(\partial_t - M\vec{n}\cdot\vec{\sigma})\psi}$$

where $\psi = (\psi_\uparrow, \psi_\downarrow)$ is a two-component Grassmann spinor, and $\vec{\sigma}$ are Pauli matrices acting on its spinor indices. $\vec{n}^2 = 1$. Let $S_0[n] = \int K\dot{n}^2 - (2s+1)\pi \mathbf{i} W_0[n]$, where I've added a second-order kinetic term for fun.

First of all, consider a fixed, say static, configuration of $\check{n}$. What does this do to the propagation of the fermion? I claim that it gaps out the fermion excitations, in the sense that

$$\left\langle \mathbf{c}^\dagger_\alpha(t) \mathbf{c}_\beta(0) \right\rangle \equiv \left\langle \bar{\psi}_\alpha(t) \psi_\beta(0) \right\rangle$$

will be short-ranged in time. Let's see this using the path integral.

We can do the (gaussian) integral over the fermion:

$$Z = \int [D\vec{n}] e^{-S_{\text{eff}}[\vec{n}]}$$

[17] The calculation between the first and second lines of (11.50) is familiar to us – it is a single Wick contraction, and can be described as a feynman diagram with one line between the two insertions. More prosaically, it is

$$\left\langle \bar{\psi}(\tau_N + \Delta\tau)\psi(\tau_N) \right\rangle \overset{(11.48)}{=} T^2 \sum_{nm} e^{\mathbf{i}(\omega_n - \omega_m)\tau + \mathbf{i}\omega_n \Delta\tau} \left\langle \bar{\psi}(\omega_n)\psi(\omega_m) \right\rangle \overset{(11.49)}{=} T \sum_m \frac{e^{\mathbf{i}\omega_n \Delta\tau}}{\mathbf{i}\omega_n - \omega_0 + \mu} \overset{T\to 0}{\to} \int đ\omega \frac{e^{\mathbf{i}\omega\Delta\tau}}{\mathbf{i}\omega - \omega_0 + \mu}.$$

with

$$S_{\text{eff}}[\vec{n}] = S_0[\vec{n}] - \log\det\left(\partial_t - M\vec{n}\cdot\vec{\sigma}\right) \equiv -\log\det D.$$

The variation of the effective action under a variation of $\vec{n}$ is:

$$\delta S_{\text{eff}} = -\text{tr}\left(\delta D D^{-1}\right) = -\text{tr}\left(\delta D D^\dagger \left(DD^\dagger\right)^{-1}\right)$$

where $D^\dagger = -\partial_t + M\vec{n}\cdot\vec{\sigma}$. This is

$$\delta S_{\text{eff}} = M\text{tr}\left(\delta\vec{n}\cdot\vec{\sigma}\left(\partial_t + M\vec{n}\cdot\vec{\sigma}\right)\left(\underbrace{-\partial_t^2 + M^2 - M\dot{\vec{n}}\cdot\vec{\sigma}}_{=DD^\dagger}\right)^{-1}\right). \tag{11.51}$$

We can expand the denominator in $\dot{\vec{n}}/M$ (and use $n^2 = 1$) to get

$$\delta S_{\text{eff}} = \int \mathrm{d}t\left(-\frac{M}{|M|}\frac{1}{2}\delta\vec{n}\cdot\left(\vec{n}\times\dot{\vec{n}}\right) + \frac{1}{4M}\delta\dot{\vec{n}}\dot{\vec{n}} + ...\right)$$

where ... is higher order in the expansion and we ignore it. But we know this is the variation of

$$S_{\text{eff}} = -2\pi\frac{M}{|M|}W_0 + \int_0^T \mathrm{d}t\left(\frac{1}{8M}\dot{\vec{n}}^2\right) + \mathcal{O}\left(\frac{\dot{n}}{M}\right)^2$$

where W_0 is the WZW term. Integrating out the fermions has shifted the coefficient of the WZW term from $s \to s \mp \frac{1}{2}$ depending on the sign of M. This is satisfying: we are adding angular momenta, $s\otimes\frac{1}{2} = \left(s - \frac{1}{2}\right)\oplus\left(s + \frac{1}{2}\right)$. If $M > 0$, it is an antiferromagnetic interaction whose groundstates will be the ones with smaller eigenvalue of $\vec{S}^2$. If $M < 0$, it is ferromagnetic, and the low-energy manifold grows.

The second term in S_{eff} is a shift of K. Higher-order terms are suppressed by more powers of $\frac{\dot{n}}{M}$, so for $\dot{n} \ll M$, this is a local action. That means that the coupling to n must have gapped out the fermions. That the term proportional to M is a funny mass term for the fermions is clear from the expression for $DD^\dagger$ in (11.51): when n is static, $DD^\dagger = -\partial_t^2 + M^2$, so that the fermion propagator is

$$\left\langle\bar{\psi}_\alpha(t)\psi_\beta(0)\right\rangle = \left(\frac{1}{D}\right)_t = \left(\frac{D}{DD^\dagger}\right)_t = \int đ\omega\frac{e^{\mathbf{i}\omega t}\left(\omega + \mathbf{i}M\vec{n}\cdot\sigma_{\alpha\beta}\right)}{\omega^2 + M^2} \sim e^{-Mt}$$

which is short-ranged in time. So indeed the fermions are fast modes in the presence of the coupling to the n-field.

Such topological terms are one way in which some (topological) information from short distances can persist in the low energy effective action. Being quantized, they can't change under the continuous RG evolution. Here the WZW term manages to be independent of M, the mass scale of the fermions. Here the information is that the system is made of fermions (or at least a half-integer spin representation of $\mathsf{SU}(2)$).

The above calculation generalizes well to higher dimensions. For many examples of its application, see this paper (the context for this paper will become clearer in §14.3).

11.5 Coherent state path integrals for bosons

[Wen §3.3] We can do the same thing for bosons, using ordinary SHO (simple harmonic oscillator) coherent states. What I mean by 'bosons' is a many-body system whose Hilbert space can be written as $\mathcal{H} = \otimes_k \mathcal{H}_k$ where k is a label (could be real space, could be momentum space) and

$$\mathcal{H}_k = \text{span}\{|0\rangle_k, \mathbf{a}_k^\dagger |0\rangle_k, \frac{1}{2}\left(\mathbf{a}_k^\dagger\right)^2 |0\rangle_k, ...\} = \text{span}\{|n\rangle_{\vec{k}}, n = 0, 1, 2...\}$$

is the SHO Hilbert space. Assume the modes satisfy

$$[\mathbf{a}_{\vec{k}}, \mathbf{a}_{\vec{k}'}^\dagger] = \delta^d(\vec{k} - \vec{k}').$$

A good example hamiltonian to keep in mind is the free one,

$$\mathbf{H}_0 = \sum_{\vec{k}} \left(\epsilon_{\vec{k}} - \mu\right) \mathbf{a}_{\vec{k}}^\dagger \mathbf{a}_{\vec{k}} \ .$$

The object $\epsilon_{\vec{k}} - \mu$ determines the energy of the state with one boson of momentum $\vec{k}$: $\mathbf{a}_{\vec{k}}^\dagger |0\rangle$. The chemical potential μ shifts the energy of any state by an amount proportional to

$$\left\langle \sum_{\vec{k}} \mathbf{a}_{\vec{k}}^\dagger \mathbf{a}_{\vec{k}} \right\rangle = N$$

the number of bosons.

For each normal mode $\mathbf{a}$, coherent states are[18]

$$\mathbf{a}|\phi\rangle = \phi|\phi\rangle; \qquad |\phi\rangle = \mathcal{N} e^{\phi \mathbf{a}^\dagger} |0\rangle .$$

The eigenbra of $\mathbf{a}^\dagger$ is $\langle\phi|$, with

$$\langle\phi| \mathbf{a}^\dagger = \langle\phi| \phi^\star, \quad \langle\phi| = \langle 0| e^{+\phi^\star \mathbf{a}} \mathcal{N}.$$

(In this case, this equation is the adjoint of the previous one.) Their overlap is[19]:

$$\langle\phi_1|\phi_2\rangle = e^{\phi_1^\star \phi_2}.$$

[18] The right equation is true because

$$\mathbf{a} e^{\phi \mathbf{a}^\dagger} |0\rangle = \sum_{n=0}^{\infty} \frac{\phi^n}{n!} \underbrace{\mathbf{a}\left(\mathbf{a}^\dagger\right)^n |0\rangle}_{n(\mathbf{a}^\dagger)^{n-1}|0\rangle} = \sum_{m=n-1} \frac{\phi^{m+1}}{m!} \left(\mathbf{a}^\dagger\right)^m |0\rangle .$$

So, as Jonathan Lam points out, a better name for these would be *Hilbert hotel states*.

[19] Check this by expanding the coherent states in the number basis and doing the integrals

$$\int \frac{d\phi d\phi^\star}{\pi} e^{-\phi\phi^\star} \phi^n \left(\phi^\star\right)^{n'} = \int_0^{2\pi} \frac{d\theta}{2\pi} e^{\mathbf{i}(n-n')\theta} \int_0^\infty du e^{-u} u^{\frac{n+n'}{2}}$$

to get $\mathbb{1} = \sum_n |n\rangle\langle n|$.

If we choose $\mathcal{N} = e^{-|\phi|^2/2}$, they are normalized, but it is more convenient to set $\mathcal{N} = 1$. The overcompleteness relation on $\mathcal{H}_k$ is

$$\mathbb{1}_k = \int \frac{d\phi d\phi^\star}{\pi} e^{-|\phi|^2} |\phi\rangle \langle\phi| .$$

It will be convenient to arrange all our operators into sums of *normal-ordered* operators:

$$: \mathbf{a}_k \mathbf{a}_l^\dagger :=: \mathbf{a}_l^\dagger \mathbf{a}_k := \mathbf{a}_l^\dagger \mathbf{a}_k$$

with all annihilation operators to the right of all creation operators. Coherent state expectation values of such operators can be built from the monomials

$$\langle\phi| \prod_k \left(\mathbf{a}_k^\dagger\right)^{M_k} (\mathbf{a}_k)^{N_k} |\phi\rangle = \prod_k (\phi_k^\star)^{M_k} (\phi_k)^{N_k} .$$

[End of Lecture 50]

Let the Hamiltonian be $\mathbf{H} = H(\{\mathbf{a}_k^\dagger\}, \{\mathbf{a}_k\}) =: \mathbf{H} :$, normal ordered. By now you know how to derive the path integral using this resolution of the identity $\mathbb{1} = \prod_{\vec{k}} \mathbb{1}_{\vec{k}}$,

$$\begin{aligned} Z &= \mathrm{tr} e^{-\mathbf{H}/T} \\ &= \int_{\phi_{N+1}=\phi_0} \prod_{l=0}^{N} d\phi_l \; e^{-\sum_{l=0}^{N}\left(\phi_{l+1}(\phi_{l+1}-\phi_l) - \Delta\tau H(\phi_{l+1}^\star \phi_l)\right)} \\ &\simeq \int_{\phi(0)=\phi(1/T)} [D\phi] \; e^{-\int_0^{1/T} d\tau (\phi^\star \partial_\tau \phi + H(\phi^\star, \phi))}. \end{aligned} \tag{11.52}$$

Putting back the mode labels, this is

$$Z = \int [Da] e^{\int dt \sum_{\vec{k}} \left(\frac{1}{2}\left(a_{\vec{k}}^\star \dot{a}_{\vec{k}} - a_{\vec{k}} \dot{a}_{\vec{k}}^\star\right) - \left(\epsilon_{\vec{k}} - \mu\right) a_{\vec{k}}^\star a_{\vec{k}}\right)}.$$

In real space $a_{\vec{k}} \equiv \int \mathrm{d}^{D-1}x e^{\mathbf{i}\vec{k}\cdot\vec{x}} \psi(\vec{x})$, Taylor expanding $\epsilon_{\vec{k}} - \mu = -\mu + \frac{\vec{k}^2}{2m} + \mathcal{O}(k^4)$, this is

$$Z = \int [D\psi] e^{\int \mathrm{d}^d\vec{x}\mathrm{d}t\left(\frac{1}{2}(\psi^\star \partial_t \psi - \psi \partial_t \psi^\star) - \frac{1}{2m} \vec{\nabla}\psi^\star \cdot \vec{\nabla}\psi - \mu \psi^\star \psi\right)}.$$

Real time. If you are interested in real-time propagation, rather than euclidean time, just replace the euclidean propagator $e^{-\tau\mathbf{H}} \mapsto e^{-\mathbf{i}t\mathbf{H}}$. The result, for example, for the amplitude to propagate from one bose coherent state to another is

$$\langle\phi_f, t_f| e^{-\mathbf{i}t\mathbf{H}} |\psi_0, t_0\rangle = \int_{\phi(t_0)=\phi_0}^{\phi(t_f)=\phi_f} D\phi^\star D\phi \; e^{\frac{\mathbf{i}}{\hbar} \int_{t_0}^{t_f} dt (\mathbf{i}\hbar \phi^\star \partial_t \phi - H(\phi, \phi^\star))}.$$

Note a distinguishing feature of the Berry phase term that it produces a complex term in the real-time action.

This is the non-relativistic field theory we found in 215A by taking the $E \ll m$ limit of a relativistic scalar field. Notice that the field ψ is actually the coherent state eigenvalue!

If instead we had an interaction term in H, say $\Delta H = \int d^dx \int d^dy \frac{1}{2}\psi^\star(x,t)\psi(x,t)V(x-y)\psi^\star(y,t)\psi(y,t)$, it would lead to a term in the path integral action

$$S_i = -\int \mathrm{d}t \int d^dx \int d^dy \frac{1}{2}\psi^\star(x,t)\psi(x,t)V(x-y)\psi^\star(y,t)\psi(y,t) \ .$$

In the special case $V(x-y) = V(x)\delta^d(x-y)$, this is the local quartic interaction we considered briefly earlier.

Non-relativistic scalar fields

[Zee §III.5, V.1, Kaplan nucl-th/0510023 §1.2.1] In the previous discussion of the EFT for a superconductor (at the end of 215B), I spoke as if the complex scalar were relativistic.

In superconducting materials, it is generally not. In real superconductors, at least. How should we think about a non-relativistic field? A simple answer comes from realizing that a relativistic field which can make a boson of mass m can certainly make a boson of mass m which is moving slowly, with $v \ll c$. By taking a limit of the relativistic model, then, we can make a description which is useful for describing the interactions of an indefinite number of bosons moving slowly in some Lorentz frame. A situation that calls for such a description is a large collection of ^{4}He atoms.

Reminder: non-relativistic limit of a relativistic scalar field. A non-relativistic particle in a relativistic theory (consider massive ϕ^4 theory) has energy

$$E = \sqrt{p^2 + m^2} \overset{\text{if } v \ll c}{=} m + \frac{p^2}{2m} + \ldots$$

This means that the field that creates and annihilates it looks like

$$\phi(\vec{x},t) = \sum_{\vec{k}} \frac{1}{\sqrt{2E_{\vec{k}}}} \left(a_{\vec{k}} e^{-\mathbf{i}E_{\vec{k}}t - \mathbf{i}\vec{k}\cdot\vec{x}} + h.c. \right)$$

In particular, we have

$$\dot{\phi}^2 \simeq m^2\phi^2$$

and the BHS of this equation is large. To remove this large number let's change variables:

$$\phi(x,t) \equiv \frac{1}{\sqrt{2m}} \left(e^{-\mathbf{i}mt} \underbrace{\Phi(x,t)}_{\text{complex}, \dot{\Phi} \ll m\Phi} + h.c. \right) \ .$$

Notice that Φ is complex, even if ϕ is real.

Let's think about the action governing this NR sector of the theory. We can drop terms with unequal numbers of Φ and $\Phi^\star$ since such terms would come with a factor of $e^{\mathbf{i}mt}$ which gives zero when integrated over time. Starting from $(\partial\phi)^2 - m^2\phi^2 - \lambda\phi^4$ we get:

$$L_{\text{real time}} = \Phi^\star \left(\mathbf{i}\partial_t + \frac{\vec{\nabla}^2}{2m} \right) \Phi - g^2 \left(\Phi^\star\Phi\right)^2 + ... \tag{11.53}$$

with $g^2 = \frac{\lambda}{4m^2}$.

Notice that Φ is a complex field and its action has a U(1) symmetry, $\Phi \to e^{\mathbf{i}\alpha}\Phi$, even though the full theory did not. The associated conserved charge is the number of particles:

$$j_0 = \Phi^\star\Phi, j_i = \frac{\mathbf{i}}{2m}\left(\Phi^\star\partial_i\Phi - \partial_i\Phi^\star\Phi\right), \quad \partial_t j_0 - \nabla\cdot\vec{j} = 0 \ .$$

Notice that the 'mass term' $\Phi^\star\Phi$ is then actually the chemical potential term, which encourages a nonzero density of particles to be present.

This is an example of an *emergent* symmetry: a symmetry of an EFT that is not a symmetry of the microscopic theory. The ... in (11.53) include terms which break this symmetry, but they are irrelevant. (Particle physics folks sometimes call such a symmetry 'accidental', which is a terrible name. An example of an emergent symmetry in the Standard Model is baryon number.)

To see more precisely what we mean by irrelevant, let's think about scaling. To keep this kinetic term fixed we must scale time and space differently:

$$x \to \tilde{x} = sx, \ t \to \tilde{t} = s^2 t, \ \Phi \to \tilde{\Phi}(\tilde{x},\tilde{t}) = \zeta\Phi(sx, s^2t) \ .$$

A fixed point with this scaling rule has dynamical exponent $z = 2$. The scaling of the bare action (with no mode elimination step) is

$$\begin{aligned} S_E^{(0)} &= \int \underbrace{dtd^d\vec{x}}_{=s^{d+z}d\tilde{t}d^d\tilde{x}} \left(\Phi^\star\left(sx, s^2t\right) \underbrace{\left(\partial_t - \frac{\vec{\nabla}^2}{2m}\right)}_{=s^{-2}\left(\tilde{\partial}_t - \frac{\tilde{\vec{\nabla}}^2}{2m}\right)} \Phi(sx, s^2t) - g^2\left(\Phi^\star\Phi(sx,s^2t)\right)^2 + ... \right) \\ &= \underbrace{s^{d+z-2}\zeta^{-2}}_{\overset{!}{=}1 \implies \zeta = s^{-3/2}} \int \mathrm{d}\tilde{t}\mathrm{d}^d\tilde{x} \left(\tilde{\Phi}^\star \left(\tilde{\partial}_t - \frac{\tilde{\vec{\nabla}}^2}{2m}\right)\tilde{\Phi} - \zeta^{-2}g^2\left(\tilde{\Phi}^\star\tilde{\Phi}(\tilde{x},\tilde{t})\right)^2 + ... \right) \end{aligned} \tag{11.54}$$

From this we learn that $\tilde{g} = s^{-3+2=-1}g \to 0$ in the IR – the quartic term is irrelevant in $D = d + 1 = 3 + 1$ with nonrelativistic scaling! Where does it become marginal? Recall the delta function potential for a particle in two dimensions.

Number and phase angle. In the NR theory, the canonical momentum for Φ is just $\frac{\partial L}{\partial \dot{\Phi}} \sim \Phi^\star$, with no derivatives. This statement becomes more shocking if we change variables to $\Phi = \sqrt{\rho}e^{\mathbf{i}\theta}$. This is a useful change of variables, if for example we knew ρ didn't want to be zero, as would happen if we add to (11.53) a term of the form $-\mu\Phi^\star\Phi$. So consider the action density

$$L = L_{\text{real time}} = \Phi^\star \left(\mathbf{i}\partial_t + \frac{\vec{\nabla}^2}{2m} \right) \Phi - V(\Phi^\star\Phi), \quad V(\Phi^\star\Phi) \equiv g^2 \left(\Phi^\star\Phi\right)^2 - \mu\Phi^\star\Phi.$$

In polar coordinates this is

$$L = \frac{\mathbf{i}}{2}\partial_t\rho - \rho\partial_t\theta - \frac{1}{2m}\left(\rho \left(\nabla\theta\right)^2 + \frac{1}{4\rho}\left(\nabla\rho\right)^2 \right) - V(\rho). \tag{11.55}$$

The first term is a total derivative. The second term says that the canonical momentum for the phase variable θ is $\rho = \Phi^\star\Phi = j_0$, the particle number density. Quantumly, then:

$$[\hat{\rho}(\vec{x}, t), \hat{\theta}(\vec{x}', t)] = \mathbf{i}\delta^d(\vec{x} - \vec{x}'). \tag{11.56}$$

Number and phase are canonically conjugate variables. If we fix the phase, the amplitude is maximally uncertain.

If we integrate over space, $N \equiv \int d^dx\rho(\vec{x}, t)$ gives the total number of particles, which is time independent, and satisfies $[N, \theta] = \mathbf{i}$.

What is the term $\mu\Phi^\star\Phi = \mu\rho$? It is a chemical potential for the boson number.

This relation (11.56) explains why there's no Higgs boson in most non-relativistic superconductors and superfluids (in the absence of some extra assumption of particle-hole symmetry). In the NR theory with first order time derivative, the would-be amplitude mode which oscillates about the minimum of $V(\rho)$ is actually just the conjugate momentum for the goldstone boson!

Superfluids. [Zee §V.1] Let me amplify the previous remark. A superconductor is just a superfluid coupled to an external U(1) gauge field, so we've already understood something about superfluids.

The effective field theory has the basic lagrangian (11.55), with $\langle\rho\rangle = \bar{\rho} \neq 0$. This nonzero density can be accomplished by adding an appropriate chemical potential to (11.55); up to an uninteresting constant, this is

$$L = \frac{\mathbf{i}}{2}\partial_t\rho - \rho\partial_t\theta - \frac{1}{2m}\left(\rho \left(\nabla\theta\right)^2 + \frac{1}{4\rho}\left(\nabla\rho\right)^2 \right) - g^2\left(\rho - \bar{\rho}\right)^2.$$

Expand around such a condensed state in small fluctuations $\sqrt{\rho} = \sqrt{\bar{\rho}} + h$, $h \ll \sqrt{\bar{\rho}}$:

$$L = -2\sqrt{\bar{\rho}}h\partial_t\theta - \frac{\bar{\rho}}{2m}\left(\vec{\nabla}\theta\right)^2 - \frac{1}{2m}\left(\vec{\nabla}h\right)^2 - 4g^2\bar{\rho}h^2 + \ldots$$

Notice that h, the fluctuation of the amplitude mode, is playing the role of the canonical momentum of the goldstone mode θ. The effects of the fluctuations can be incorporated by doing the gaussian integral over h (What suppresses self-interactions of h?), and the result is

$$\begin{aligned} L &= \sqrt{\bar{\rho}}\partial_t\theta \frac{1}{4g^2\bar{\rho} - \frac{\nabla^2}{2m}} \sqrt{\bar{\rho}}\partial_t\theta - \frac{\bar{\rho}}{2m}\left(\vec{\nabla}\theta\right)^2 \\ &= \frac{1}{4g^2}\left(\partial_t\theta\right)^2 - \frac{\bar{\rho}}{2m}\left(\nabla\theta\right)^2 + ... \end{aligned} \tag{11.57}$$

where in the second line we are expanding in the small wavenumber k of the modes, that is, we are constructing an action for Goldstone modes whose wavenumber is $k \ll \sqrt{9g^2\bar{\rho}m}$ so we can ignore higher gradient terms.

The linearly dispersing mode in this superfluid that we have found is sometimes called the phonon. This is a good name because the wave involves oscillations of the density:

$$\underline{h} = \frac{1}{4g^2\bar{\rho} - \frac{\nabla^2}{2m}} \sqrt{\bar{\rho}}\partial_t\theta$$

is the saddle point solution for h. The phonon has dispersion relation

$$\omega^2 = \frac{2g^2\bar{\rho}}{m}\vec{k}^2.$$

This mode has an emergent Lorentz symmetry with a lightcone with velocity $v_c = g\sqrt{2\bar{\rho}/m}$. The fact that the sound velocity involves g – which determined the steepness of the walls of the wine-bottle potential – is a consequence of the non-relativistic dispersion of the bosons. In the relativistic theory, we have $L = \partial_\mu\Phi^\star\partial^\mu\Phi - g\left(\Phi^\star\Phi - v^2\right)^2$ and we can take $g \to \infty$ fixing v and still get a linearly dispersing mode by plugging in $\Phi = e^{\mathbf{i}\theta}v$.

The importance of the linearly dispersing phonon mode of the superfluid is that there is *no other* low energy excitation of the fluid. With a classical pile of (*e.g.* non interacting) bosons, a chunk of moving fluid can donate some small momentum $\vec{k}$ to a single boson at energy cost $\frac{(\hbar\vec{k})^2}{2m}$. A quadratic dispersion means more modes at small k than a linear dispersion (the density of states is $N(E) \propto k^{D-1}\frac{dk}{dE}$). With only a linearly dispersing mode at low energies, there is a critical velocity below which a non-relativistic chunk of fluid cannot give up any momentum [Landau]: conserving momentum $M\vec{v} = M\vec{v}' + \hbar\vec{k}$ says the change in energy (which must be negative for this to happen on its own) is

$$\frac{1}{2}M(v')^2 + \hbar\omega(k) - \frac{1}{2}Mv^2 = -\hbar kv + \frac{(\hbar k)^2}{2m} + \hbar\omega(k) = (-v + v_c)\hbar k + \frac{(\hbar k)^2}{2m} \ .$$

For small k, this is only negative when $v > v_c = \partial_k\omega|_{k=0}$.

You can ask: an ordinary liquid also has a linearly dispersing sound mode; why doesn't Landau's argument mean that it has superfluid flow? The answer is that it has *other* modes with softer dispersion (so more contribution at low energies), in particular diffusion modes, with $\omega \propto k^2$ (there is an important factor of $\mathbf{i}$ in there).

The Goldstone boson has a compact target space, $\theta(x) \equiv \theta(x) + 2\pi$, since, after all, it is the phase of the boson field. This is significant because it means that as the phase wanders around in space, it can come back to its initial value after going around the circle – such a loop encloses a *vortex*. Somewhere inside, we must have $\Phi = 0$. There is much more to say about this.

12 Anomalies

[Zee §IV.7; Polyakov, *Gauge Fields and Strings*, §6.3; K. Fujikawa, *Phys. Rev. Lett.* **42** (1979) 1195; Argyres, 1996 lectures on supersymmetry §14.3; Peskin, chapter 19]

Topology means the study of quantities which can't vary smoothly, but can only jump. Like quantities which must be integers. But the Wilson RG is a smooth process. Therefore topological information in a QFT is something the RG can't wash away – information which is RG invariant. An example we've seen already is the integer coefficients of WZW terms, which encode commutation relations. Another class of examples (in fact they are related) is *anomalies*.

Suppose we have in our hands a classical field theory in the continuum which has some symmetry. Is there a well-defined QFT whose classical limit produces this classical field theory and preserves that symmetry? The path integral construction of QFT offers some insight here. The path integral involves two ingredients: (1) an action, which is shared with the classical field theory, and (2) a path integral measure. It is possible that the action is invariant but the measure is not. This is called an anomaly. It means that the symmetry is broken, and its current conservation is violated by a known amount, and this often has many other consequences that can be understood by humans. [End of Lecture 51]

Notice that here I am speaking about actual, global symmetries. I am not talking about gauge redundancies. If you think that two field configurations are equivalent but the path integral tells you that they would give different contributions, you are doing something wrong. An anomaly in a 'gauge symmetry' means that the system has more degrees of freedom than you thought. (In particular, it does not mean that the world is inconsistent. For a clear discussion of this, please see Preskill, 1990.)

We have already seen a dramatic example of an anomaly: the violation of classical scale invariance (*e.g.* in massless ϕ^4 theory, or in massless QED) by quantum effects.

Notice that the name 'anomaly' betrays the bias that we construct a QFT by starting with a continuum action for a classical field theory; you would never imagine that *e.g.* scale invariance was an exact symmetry if you started from a well-defined quantum lattice model.

The example we will focus on here is the *chiral anomaly*. This is an equation for the violation of the chiral (aka axial) current for fermions coupled to a background gauge field. The chiral anomaly was first discovered in perturbation theory, by computing a certain Feynman diagram with a triangle; the calculation was motivated by the experimental observation of the process $\pi^0 \to \gamma\gamma$, which would not happen if the chiral current were conserved.

I will outline a derivation of this effect which is more illuminating than the triangle diagram. It shows that the one-loop result is exact – there are no other corrections. It shows that the quantity on the right hand side of the continuity equation for the would-be current integrates to an integer. It gives a proof of the *index theorem*, relating numbers of solutions to the Dirac equation in a background field configuration to a certain integral of field strengths. It butters your toast.

12.0.1 Chiral anomaly

Chiral symmetries. In even-dimensional spacetimes, the Dirac representation of $\mathrm{SO}(D-1,1)$ is reducible. This is because

$$\gamma^5 \equiv \prod_{\mu=0}^{D-1} \gamma^\mu \neq 1, \quad \text{satisfies} \quad \{\gamma^5, \gamma^\mu\} = 0, \forall \mu$$

which means that γ^5 commutes with the Lorentz generators

$$[\gamma^5, \Sigma^{\mu\nu}] = 0, \qquad \Sigma^{\mu\nu} \equiv \frac{1}{2}[\gamma^\mu, \gamma^\nu].$$

A left- or right-handed Weyl spinor is an irreducible representation of $\mathrm{SO}(D-1,1)$, $\psi_{L/R} \equiv \frac{1}{2}\left(1 \pm \gamma^5\right)\psi$. This allows the possibility that the L and R spinors can transform differently under a symmetry; such a symmetry is a chiral symmetry.

Note that in $D = 4k$ dimensions, if ψ_L is a left-handed spinor in representation $\mathbf{r}$ of some group G, then its image under CPT, $\psi_L^{CPT}(t, \vec{x}) \equiv \mathbf{i}\gamma^0 \left(\psi_L(-t, -\vec{x})\right)^\star$, is right-handed and transforms in representation $\bar{\mathbf{r}}$ of G. Therefore chiral symmetries arise when the Weyl fermions transform in *complex representations* of the symmetry group, where $\bar{\mathbf{r}} \neq \mathbf{r}$. (In $D = 4k + 2$, CPT maps left-handed fields to left-handed fields. For more detail on discrete symmetries and Dirac fields, see Peskin §3.6.)

Some more explicit words about chiral fermions in $D = 3+1$, mostly notation. Recall Peskin's *Weyl* basis of gamma matrices in 3+1 dimensions, in which γ^5 is diagonal:

$$\gamma^\mu = \begin{pmatrix} 0 & \bar{\boldsymbol{\sigma}}^\mu \\ \boldsymbol{\sigma}^\mu & 0 \end{pmatrix}, \quad \boldsymbol{\sigma}^\mu \equiv (\mathbb{1}, \vec{\boldsymbol{\sigma}})^\mu, \quad \bar{\boldsymbol{\sigma}}^\mu \equiv (\mathbb{1}, -\vec{\boldsymbol{\sigma}})^\mu, \quad \gamma^5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

This makes the reducibility of the Dirac representation of $\mathrm{SO}(3,1)$ manifest, since the Lorentz generators are $\propto [\gamma^\mu, \gamma^\nu]$ block diagonal in this basis. The gammas are a map from the $(\mathbf{1}, \mathbf{2_R})$ representation to the $(\mathbf{2_L}, \mathbf{1})$ representation. It is sometimes useful to denote the $\mathbf{2_R}$ indices by $\alpha, \beta = 1, 2$ and the $\mathbf{2_L}$ indices by $\dot{\alpha}, \dot{\beta} = 1, 2$. Then we can define two-component Weyl spinors $\psi_{L/R} = P_{L/R}\psi \equiv \frac{1}{2}\left(1 \pm \gamma^5\right)\psi$ by simply forgetting about the other two components. The conjugate of a L spinor $\chi = \psi_L$ (L means $\gamma^5\chi = \chi$) is right-handed:

$$\bar{\chi} = \chi^\dagger\gamma^0, \quad \bar{\chi}\gamma^5 = \chi^\dagger\gamma^0\gamma^5 = -\chi^\dagger\gamma^5\gamma^0 = -\chi^\dagger\gamma^0 = -\bar{\chi}.$$

We can represent any system of Dirac fermions in terms of a collection of twice as many Weyl fermions.

For a continuous symmetry G, we can be more explicit about the meaning of a complex representation. The statement that ψ is in representation $\mathbf{r}$ means that its transformation law is

$$\delta\psi_a = \mathbf{i}\epsilon^A \left(t^A_{\mathbf{r}}\right)_{ab}\psi_b$$

where t^A, $A = 1..\dim G$ are generators of G in representation $\mathbf{r}$; for a compact lie group G, we may take the t^A to be Hermitian. The conjugate representation, by definition, is one with which you can make a singlet of G – it's the way $\psi^{\star T}$ transforms:

$$\delta\psi^{\star T}_a = -\mathbf{i}\epsilon^A \left(t^A_{\mathbf{r}}\right)^T_{ab}\psi^{\star T}_b \; .$$

So:

$$t^A_{\bar{\mathbf{r}}} = -\left(t^A_{\mathbf{r}}\right)^T .$$

The condition for a complex representation is that this is different from $t^A_{\mathbf{r}}$ (actually we have to allow for relabelling of the generators). The simplest case is $G = \mathrm{U}(1)$, where t is just a number indicating the charge. In that case, any nonzero charge gives a complex representation.

Consider the effective action produced by integrating out Dirac fermions coupled to a *background* gauge field (the gauge field is just going to sit there for this whole calculation):

$$e^{\mathbf{i}S_{\text{eff}}[A]} \equiv \int [D\psi D\bar{\psi}] \; e^{\mathbf{i}S[\psi,\bar{\psi},A]} \; .$$

We must specify how the fermions coupled to the gauge field. The simplest example is if A is a $U(1)$ gauge field and ψ is minimally coupled:

$$S[\psi,\bar{\psi},A] = \int \mathrm{d}^D x \bar{\psi}\mathbf{i}\not{D}\psi, \qquad \not{D}\psi \equiv \gamma^\mu\left(\partial_\mu + \mathbf{i}A_\mu\right)\psi.$$

We will focus on this example, but you could imagine instead that A_μ is a non-Abelian gauge field for the group G, and ψ is in a representation R, with gauge generators $T^A(R)$ ($A = 1...\mathrm{dim}G$), so the coupling would be

$$\bar{\psi}\mathbf{i}\not{D}\psi = \bar{\psi}_a\gamma^\mu\left(\partial_\mu\delta_{ab} + \mathbf{i}A^A_\mu T^A(R)_{ab}\right)\psi_b \; . \tag{12.1}$$

Much of the discussion below applies for any even D.

In the absence of a mass term, the action (in the Weyl basis) involves no coupling between L and R:

$$S[\psi, \bar{\psi}, A] = \int \mathrm{d}^D x \left(\psi_L^\dagger \mathbf{i} \sigma^\mu D_\mu \psi_L + \psi_R^\dagger \mathbf{i} \bar{\sigma}^\mu D_\mu \psi_R \right)$$

and therefore is invariant under the global chiral rotation

$$\psi \to e^{\mathbf{i}\alpha\gamma^5}\psi, \quad \psi^\dagger \to \psi^\dagger e^{-\mathbf{i}\alpha\gamma^5}, \quad \bar{\psi} \to \bar{\psi} e^{+\mathbf{i}\alpha\gamma^5}. \quad \text{That is: } \psi_L \to e^{\mathbf{i}\alpha}\psi_L, \quad \psi_R \to e^{-\mathbf{i}\alpha}\psi_R.$$

(The mass term couples the two components

$$L_m = \bar{\psi}\left(\mathrm{Re}m + \mathrm{Im}m\gamma^5\right)\psi = m\psi_L^\dagger \psi_R + h.c.;$$

notice that the mass parameter is complex.) The associated Noether current is $j^5_\mu = \bar{\psi}\bar{\gamma}^5\gamma_\mu\psi$, and it seems like we should have $\partial^\mu j^5_\mu \stackrel{?}{=} 0$. This follows from the massless (classical) Dirac equation $0 = \gamma^\mu\partial_\mu\psi$. (With the mass term, we would have instead $\partial^\mu j^5_\mu \stackrel{?}{=} 2\mathbf{i}\bar{\psi}\left(\mathrm{Re}m\gamma^5 + \mathrm{Im}m\right)\psi$.)

Notice that there is another current $j^\mu = \bar{\psi}\gamma^\mu\psi$. j^μ is the current which is coupled to the gauge field, $L \ni A_\mu j^\mu$. The conservation of this current is required for gauge invariance of the effective action

$$S_{\text{eff}}[A_\mu] \stackrel{!}{=} S_{\text{eff}}[A_\mu + \partial_\mu\lambda] \sim \log\left\langle e^{\mathbf{i}\int\lambda(x)\partial_\mu j^\mu}\right\rangle + S_{\text{eff}}[A_\mu].$$

No matter what happens we can't find an anomaly in j^μ. The anomalous one is the other one, the *axial current*.

To derive the conservation law we can use the Noether method. This amounts to substituting $\psi'(x) \equiv e^{\mathbf{i}\alpha(x)\gamma^5}\psi(x)$ into the action:

$$S_F[\psi'] = \int \mathrm{d}^D x \bar{\psi} e^{+\mathbf{i}\alpha\gamma^5}\mathbf{i}\not{D} e^{\mathbf{i}\alpha\gamma^5}\psi = \int \mathrm{d}^D x \left(\bar{\psi}\mathbf{i}\not{D}\psi + \bar{\psi}\mathbf{i}\gamma^5\left(\not{\partial}\alpha\right)\psi\right) \stackrel{\text{IBP}}{=} S_F[\psi] - \mathbf{i}\int \alpha(x)\partial^\mu \mathrm{tr}\bar{\psi}\gamma^5\gamma_\mu\psi.$$

Then we can completely get rid of $\alpha(x)$ if we can change integration variables, *i.e.* if $[D\psi'] \stackrel{?}{=} [D\psi]$. Usually this is true, but here we pick up an interesting Jacobian.

Claim:

$$e^{\mathbf{i}S_{\text{eff}}[A]} = \int [D\psi' D\bar{\psi}'] e^{\mathbf{i}S_F[\psi']} = \int [D\psi D\bar{\psi}] e^{\mathbf{i}S_F[\psi] + \int \mathrm{d}^D x \alpha(x)\left(\partial_\mu j^\mu_5 - \mathcal{A}(x)\right)}$$

where

$$\mathcal{A}(x) = \sum_n \mathrm{tr}\bar{\xi}_n\gamma^5\xi_n \tag{12.2}$$

where ξ_n are a basis of eigenspinors of the Dirac operator. The contribution to $\mathcal{A}$ can be attributed to zeromodes of the Dirac operator.

The expression above is actually independent of α, since the path integral is invariant under a change of variables. For a conserved current, α would multiply the divergence of the current and this demand would imply current conservation. Here this implies that instead of current conservation we have a specific violation of the current:

$$\partial^\mu j^5_\mu = \mathcal{A}(x).$$

What is the anomaly. [Polyakov §6.3] An alternative useful (perhaps more efficient) perspective is that the anomaly arises from trying to define the axial current operator, which after all is a composite operator. Thus we should try to compute

$$\langle \partial_\mu j^\mu_5 \rangle = \partial_\mu \left\langle \bar\psi(x)\gamma^\mu\gamma^5\psi(x)\right\rangle$$

– the coincident operators on the RHS need to be regulated.

Consider Dirac fermions coupled to a *background* gauge field configuration $A_\mu(x)$, with action

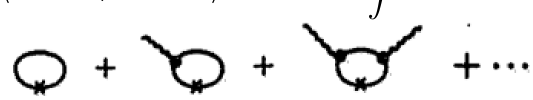
$$S = \int \mathrm{d}^D x \bar\psi \left(\mathbf{i}\gamma^\mu \left(\partial_\mu + \mathbf{i}A_\mu\right)\right)\psi.$$

For a while the discussion works in any even dimension, where $\gamma^5 = \prod_{\mu=0}^{D-1}\gamma^\mu$ satisfies $\{\gamma^\mu, \gamma^5\} = 0$ and is not the identity. (The discussion that follows actually works also for non-Abelian gauge fields.) The classical Dirac equation immediately implies that the axial current is conserved

$$\partial_\mu \left(\mathbf{i}\bar\psi\gamma^\mu\gamma^5\psi\right) \overset{?}{=} 0.$$

Consider, on the other hand, the (Euclidean vacuum) expectation value

$$J^5_\mu \equiv \left\langle \mathbf{i}\bar\psi(x)\gamma_\mu\gamma^5\psi(x)\right\rangle \equiv Z^{-1}[A]\int [D\psi D\bar\psi]e^{-S_F[\psi]}j^5_\mu$$

$=$ 

$$= -\mathbf{i}\mathrm{Tr}\,{}_\gamma\gamma_\mu\gamma^5 G^{[A]}(x,x) \tag{12.3}$$

where G is the Green's function of the Dirac operator in the gauge field background (and the figure is from Polyakov's book). We can construct it out of eigenfunctions of $\mathbf{i}\not{D}$:

$$\mathbf{i}\not{D}\xi_n(x) = \epsilon_n\xi_n(x), \qquad \bar\xi_n(x)\mathbf{i}\gamma^\mu\left(-\overleftarrow{\partial}_\mu + \mathbf{i}A_\mu\right) = \epsilon_n\bar\xi_n \tag{12.4}$$

in terms of which[20]

$$G(x,x') = \sum_n \frac{1}{\epsilon_n} \xi_n(x) \bar{\xi}_n(x'). \tag{12.5}$$

(I am suppressing spinor indices all over the place, note that here we are taking the outer product of the spinors.)

We want to define the coincidence limit, as $x' \to x$. The problem with this limit arises from the large $|\epsilon_n|$ eigenvalues; the contributions of such short-wavelength modes are local and most of them can be absorbed in renormalization of couplings. It should not (and does not) matter how we regulate them, but we must pick a regulator. A convenient choice here is heat-kernel regulator:

$$G_s(x,x') \equiv \sum_n e^{-s\epsilon_n^2} \frac{1}{\epsilon_n} \xi_n(x) \bar{\xi}_n(x')$$

and

$$J_\mu^5(x) = \sum_n e^{-s\epsilon_n^2} \frac{1}{\epsilon_n} \bar{\xi}_n(x) \gamma^5 \gamma_\mu \xi_n(x) \ .$$

The anomaly is

$$\partial^\mu J_\mu^5 = \partial^\mu \left\langle j_\mu^5 \right\rangle = \sum_n \mathbf{i} \partial^\mu \left(\bar{\xi}_n \gamma_\mu \gamma^5 \xi_n \right) \frac{e^{-s\epsilon_n^2}}{\epsilon_n}.$$

The definition (12.4) says

$$\mathbf{i} \partial^\mu \left(\bar{\xi}_n \gamma_\mu \gamma^5 \xi_n \right) = -2\epsilon_n \bar{\xi}_n \gamma_5 \xi_n$$

using $\{\gamma^5, \gamma^\mu\} = 0$. (Notice that the story would deviate dramatically here if we were studying the vector current which lacks the γ^5.) This gives

$$\partial^\mu J_\mu^5 = 2 \mathrm{Tr}\ {}_\alpha \gamma^5 e^{-s\left(\mathbf{i}\cancel{D}\right)^2}$$

with

$$\left(\mathbf{i}\cancel{D}\right)^2 = -\left(\gamma_\mu \left(\partial_\mu + \mathbf{i} A_\mu\right)\right)^2 = -\left(\partial_\mu + A_\mu\right)^2 - \frac{\mathbf{i}}{2} \Sigma_{\mu\nu} F^{\mu\nu}$$

where $\Sigma_{\mu\nu} \equiv \frac{1}{2}[\gamma_\mu, \gamma_\nu]$ is the spin Lorentz generator. This is (12.2), now better defined by the heat kernel regulator. We've shown that in any even dimension,

$$\partial^\mu \left\langle j_\mu^5(x) \right\rangle = 2 \mathrm{Tr}\ {}_\alpha \gamma^5 e^{s\cancel{D}^2} \tag{12.6}$$

This can now be expanded in small s, which amounts to an expansion in powers of A, F. If there is no background field, $A = 0$, we get

$$\left\langle x | e^{-s\left(\mathbf{i}\cancel{\partial}\right)^2} | x \right\rangle = \int đ^D p \ e^{-sp^2} = \underbrace{K_D}_{= \frac{\Omega_{D-1}}{(2\pi)^D} \text{ as before}} \frac{1}{s^{D/2}} \overset{D=4}{=} \frac{1}{16\pi^2 s^2}. \tag{12.7}$$

[20]Actually, this step is full of danger. (Polyakov has done it to me again. Thanks to Sridip Pal for discussions of this point.) See §12.0.2 below.

This term will renormalize the charge density

$$\rho(x) = \left\langle \psi^\dagger \psi(x) \right\rangle = \mathrm{tr}\gamma^0 G(x,x),$$

for which we must add a counterterm (in fact, it is accounted for by the counterterm for the gauge field kinetic term, *i.e.* the running of the gauge coupling). But it will not affect the axial current conservation which is proportional to

$$\mathrm{tr}\left(\gamma^5 G(x,x)\right)|_{A=0} \propto \mathrm{tr}\gamma^5 = 0.$$

Similarly, bringing down more powers of $(\partial + A)^2$ doesn't give something nonzero since the γ^5 remains.

In $D = 4$, the first term from expanding $\Sigma_{\mu\nu}F^{\mu\nu}$ is still zero from the spinor trace. (Not so in $D = 2$.) The first nonzero term comes from the next term:

$$\mathrm{tr}\left(\gamma_5 e^{-s\left(\mathbf{i}\not{D}\right)^2}\right)_{xx} = \underbrace{\left\langle x|e^{-s(\mathbf{i}D)^2}|x\right\rangle}_{\overset{(12.7)}{=}\frac{1}{16\pi^2 s^2}+\mathcal{O}(s^{-1})} \cdot \frac{s^2}{8}\cdot(\mathbf{i}^2)\underbrace{\mathrm{tr}\left(\gamma^5\Sigma^{\mu\nu}\Sigma^{\rho\lambda}\right)}_{=4\epsilon^{\mu\nu\rho\lambda}}\cdot\underbrace{\mathrm{tr}_c}_{\text{color}}\left(F_{\mu\nu}F_{\rho\lambda}\right) + \mathcal{O}(s^1)\ .$$

In the abelian case, just ignore the trace over color indices, tr_c. The terms that go like positive powers of s go away in the continuum limit. Therefore

$$\partial_\mu J_5^\mu = -2\cdot\frac{1}{16\pi s^2}\cdot\frac{s^2}{8}\cdot 4\epsilon^{\mu\nu\rho\lambda}\mathrm{tr}_c F_{\mu\nu}F_{\rho\lambda} + \mathcal{O}(s^1) = -\frac{1}{8\pi^2}\mathrm{tr}F_{\mu\nu}\left(\star F\right)^{\mu\nu}\ . \tag{12.8}$$

(Here $(\star F)^{\mu\nu} \equiv \frac{1}{8}\epsilon^{\mu\nu\rho\lambda}F_{\rho\lambda}$.) This is the chiral anomaly formula. It can also be usefully written as:

$$\partial_\mu J_5^\mu = -\frac{1}{8\pi^2}\mathrm{tr}F\wedge F = -\frac{1}{32\pi^2}\vec{E}\cdot\vec{B}.$$

- This object on the RHS is a total derivative. In the abelian case it is

  $$F\wedge F = \mathrm{d}\left(A\wedge F\right)\ .$$

  Its integral over spacetime is a topological (in fact $16\pi^2$ times an integer) characterizing the gauge field configuration. How do I know it is an integer? The anomaly formula! The change in the number of left-handed fermions minus the number of right-handed fermions during some time interval is:

  $$\Delta Q_A \equiv \Delta\left(N_L - N_R\right) = \int \mathrm{d}t\partial_t J_0^5 = \int_{M_4}\partial^\mu J_\mu^5 = 2\int_{M_4}\frac{F\wedge F}{16\pi^2}$$

  where M_4 is the spacetime region under consideration. If nothing is going on at the boundaries of this spacetime region (*i.e.* the fields go to the vacuum, or there is no boundary, so that no fermions are entering or leaving), we can conclude that the RHS is an integer.

- Look back at the diagrams in (12.3). Which term in that expansion gave the nonzero contribution to the axial current violation? In $D = 4$ it is the diagram with three current insertions, the ABJ triangle diagram. So in fact we did end up computing the triangle diagram. But this calculation also shows that nothing else contributes, even non-perturbatively. [End of Lecture 52]

- We chose a particular regulator above. The answer we got did not depend on the cutoff; in fact whatever regulator we used, we would get this answer.

- Consider what happens if we redo this calculation in other dimensions. We only consider even dimensions because in odd dimensions there is no analog of γ^5 – the Dirac spinor representation is irreducible. In $2n$ dimensions, we need n powers of F to soak up the indices on the epsilon tensor.

- If we had kept the non-abelian structure in (12.1) through the whole calculation, the only difference is that the trace in (12.8) would have included a trace over representations of the gauge group; and we could have considered also a non-abelian flavor transformation

$$\psi_I \to \left(e^{\mathbf{i}\gamma^5 g^a \tau^a}\right)_{IJ} \psi_J$$

for some flavor rotation generator τ^a. Then we would have found:

$$\partial^\mu j^{5a}_\mu = \frac{1}{16\pi^2}\epsilon^{\mu\nu\rho\lambda} F^A_{\mu\nu} F^B_{\rho\lambda} \mathrm{tr}_{c,a}\left(T^A T^B \tau^a\right).$$

A similar statement applies to the case of multiple species of fermion fields: their contributions to the anomaly add. Sometimes they can cancel; the Electroweak gauge interactions are an example of this.

12.0.2 Zeromodes of the Dirac operator

Do you see now why I said that the step involving the fermion Green's function was full of danger? The danger arises because the Dirac operator (whose inverse is the Green's function) can have *zeromodes*, eigenspinors with eigenvalue $\epsilon_n = 0$. In that case, $\mathbf{i}\not{D}$ is *not* invertible, and the expression (12.5) for G is ambiguous. This factor of ϵ_n is about to be cancelled when we compute the divergence of the current and arrive at (12.2). Usually this kind of thing is not a problem because we can lift the zeromodes a little and put them back at the end. But here it is actually hiding something important. The zeromodes cannot just be lifted. This is true because nonzero modes of $\mathbf{i}\not{D}$ must come in left-right pairs: this is because $\{\gamma^5, \mathbf{i}\not{D}\} = 0$, so $\mathbf{i}\not{D}$ and γ^5 cannot be simultaneously diagonalized in general. That is: if $\mathbf{i}\not{D}\xi = \epsilon\xi$ then $(\gamma^5\xi)$ is also an eigenvector of $\mathbf{i}\not{D}\xi$, with eigenvalue $-\epsilon$. Only for $\epsilon = 0$ does this fail, so zeromodes can come by themselves. So you can't just smoothly change the eigenvalue of some ξ_0 from zero unless it has a partner with whom to pair up. By taking linear combinations

$$\chi_n^{L/R} = \frac{1}{2}\left(1 \pm \gamma^5\right)\xi_n$$

these two partners can be arranged into a pair of simultaneous eigenvectors of $(\mathbf{i}\not{D})^2$ (with eigenvalue ϵ_n^2) and of γ^5 with $\gamma^5 = \pm$ respectively.

This leads us to a deep fact, called the (Atiyah-Singer) *index theorem*: *only* zeromodes can contribute to the anomaly. Any mode ξ_n with nonzero eigenvalue has a partner with the opposite sign of γ^5; hence they cancel exactly in

$$\sum_n \bar{\xi}_n \gamma^5 \xi_n e^{-s\epsilon_n^2} \quad !$$

So the anomaly equation tells us that the number of zeromodes of the Dirac operator, weighted by handedness (*i.e.* with a + for L and - for R) is equal to

$$N_L - N_R = \int d^D x \mathcal{A}(x) = \int \frac{1}{16\pi^2} F \wedge F.$$

A practical consequence for us is that it makes manifest that the result is independent of the regulator s.

12.0.3 The physics of the anomaly

[Polyakov, page 102; Kaplan 0912.2560 §2.1; Alvarez-Gaumé] Consider non-relativistic free (*i.e.* no 4-fermion interactions) fermions in 1+1 dimensions, *e.g.* with 1-particle dispersion $\omega_k = \frac{1}{2m}\vec{k}^2$. The groundstate of N such fermions is described by filling the N lowest-energy single particle levels, up the Fermi momentum: $|k| \leq k_F$ are filled. We must introduce an infrared regulator so that the levels are discrete – put them in a box of length L, so that $k_n = \frac{2\pi n}{L}$. (In Figure 1, the red circles are possible 1-particle states, and the green ones are the occupied ones.) The lowest-energy excitations of this groundstate come from taking a fermion just below the Fermi level $|k_1| \lesssim k_F$ and putting it just above $|k_2| \gtrsim k_F$; the energy cost is

$$E_{k_1-k_2} = \frac{1}{2m}\left(k_F + k_1\right)^2 - \frac{1}{2m}\left(k_F - k_2\right)^2 \simeq \frac{k_F}{m}\left(k_1 - k_2\right)$$

– we get relativistic dispersion with velocity $v_F = \frac{k_F}{m}$. The fields near these Fermi points in k-space satisfy the Dirac equation[21]:

$$\left(\omega - \delta k\right)\psi_L = 0, \quad \left(\omega + \delta k\right)\psi_R = 0.$$

[21] This example is worthwhile for us also because we see the relativistic Dirac equation is emerging from a non-relativistic model; in fact we could have started from an even more distant starting point

It would therefore seem to imply a conserved axial current – the number of left moving fermions minus the number of right moving fermions. But the fields ψ_L and ψ_R are not independent; with high-enough energy excitations, you reach the bottom of the band (near $k = 0$ here) and you can't tell the difference. This means that the numbers are *not* separately conserved.

We can do better in this 1+1d example and show that the amount by which the axial current is violated is given by the anomaly formula. Consider subjecting our poor 1+1d free fermions to an electric field $E_x(t)$ which is constant in space and slowly varies in time. Suppose we gradually turn it on and then turn it off; here gradually means slowly enough that the process is adiabatic. Then each particle experiences a force $\partial_t p = eE_x$ and its net change in momentum is

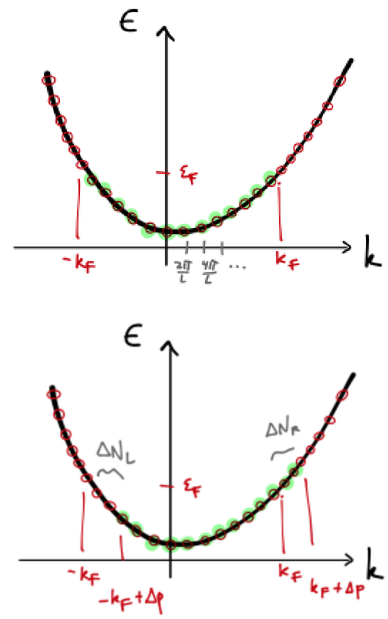


Figure 1: Green dots represent occupied 1-particle states. Top: In the groundstate. Bottom: After applying $E_x(t)$.

$$\Delta p = e \int \mathrm{d}t E_x(t).$$

This means that the electric field puts the fermions in a state where the Fermi surface $k = k_F$ has shifted to the right by Δp, as in the figure. Notice that the total number of fermions is of course the same – charge is conserved.

Now consider the point of view of the low-energy theory at the Fermi points. This theory has the action

$$S[\psi] = \int \mathrm{d}x\mathrm{d}t \bar{\psi} \left(\mathbf{i}\gamma^\mu \partial_\mu\right) \psi \ ,$$

where γ^μ are 2×2 and the upper/lower component of ψ creates fermions near the left/right Fermi point. In the process above, we have added N_R right-moving particles and taken away N_L left-moving particles, that is *added* N_L left-moving holes (aka anti-particles). The axial charge of the state has changed by

$$\Delta Q_A = \Delta(N_L - N_R) = 2\frac{\Delta p}{2\pi/L} = \frac{L}{\pi}\Delta p = \frac{L}{\pi}e \int \mathrm{d}t E_x(t) = \frac{e}{\pi}\int \mathrm{d}t\mathrm{d}x E_x = \frac{e}{2\pi}\int \epsilon_{\mu\nu}F^{\mu\nu}$$

On the other hand, the LHS is $\Delta Q_A = \int \partial^\mu J^A_\mu$. We can infer a local version of this

– *e.g.* from a lattice model, like

$$H = -t\sum_n c_n^\dagger c_{n+1} + h.c.$$

where the dispersion would be $\omega_k = -2t\left(\cos ka - 1\right) \sim \frac{1}{2m}k^2 + \mathcal{O}(k^4)$ with $\frac{1}{2m} = ta^2$.

equation by letting E vary slowly in space as well, and we conclude that

$$\partial_\mu J^\mu_A = \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu}.$$

This agrees exactly with the anomaly equation in $D = 1{+}1$ produced by the calculation above in (12.6) (see the homework).

13 Saddle points, non-perturbative field theory and resummations

13.1 Instantons in the Abelian Higgs model in $D = 1+1$

[Coleman p. 302-307] Consider the $\mathbb{CP}^{N-1}$ model in $D = 1+1$ again. What is the force between two distant (massive) z-particles? According to (11.35), the force from σ exchange is short-ranged: $\Pi(q \to 0) = \frac{4\pi}{N}$. But the Coulomb force, from A in $D = 1+1$ is independent of separation (*i.e.* the potential $\int đp \, \frac{e^{ipx}}{p^2} \sim x$ is linear). This means confinement.

Let's think more about abelian gauge theory in $D = 1+1$. Consider the case of $N = 1$. This could be called the $\mathbb{CP}^0$ model, but it is usually called the Abelian Higgs model.

$$L = \frac{1}{4e^2}F^2 + D_\mu z^\dagger D^\mu z + \frac{\kappa}{4}(z^\dagger z)^2 + \frac{\mu^2}{2}z^\dagger z + \theta\frac{F}{2\pi}.$$

What would a classical physicist say is the phase diagram of this model as we vary μ^2? For $\mu^2 > 0$, it is 2d scalar QED. There is no propagating photon, but (as we just discussed) the model confines because of the Coulomb force. The spectrum is made of boundstates of zs and $z^\dagger$s, which are stable because there is no photon for them to decay into. For $\mu^2 < 0$, it looks like the potential wants $|z|^2 = \mu^2/\kappa \equiv v^2$ in the groundstate. This would mean that A_μ eats the phase of z, gets a mass (a massive vector in $D = 1+1$ has a propagating component); the radial excitation of z is also massive. In such a Higgs phase, external charges don't care about each other, the force is short-ranged.

Not all of the statements in the classical, shaded box are correct quantumly. In fact, even at $\mu^2 < 0$, external charges are still confined (but with a different string tension than $\mu^2 > 0$). Non-perturbative physics makes a big difference here.

Let's try to do the euclidean path integral at $\mu^2 < 0$ by saddle point. This means we have to find minima of

$$S_E^0 \equiv \int \left(\frac{1}{4e^2}F^2 + D_\mu z^\dagger D^\mu z + \frac{\kappa}{4}(z^\dagger z - v^2)^2\right) d^2x.$$

(Ignore θ for now, since it doesn't affect the EOM.) Where have you seen this before?

This is exactly the functional we had to minimize in §10.1 to find the (Abrikosov-Nielsen-Olesen) vortex solution of the Abelian Higgs model. There we were thinking about a 3+1 D field theory, and we found a static configuration, translation invariant in one spatial direction, localized in the two remaining directions. Here we have only two dimensions. The same solution of the equations now represents an *instanton* – a solution of the euclidean equations of motion, localized in euclidean spacetime. Here's a quick review of the solution: Choosing polar coordinates about some origin (more on this soom), the solution has (in order that $V(\rho)$ goes to zero at large r)

$$z(r,\theta) \overset{r\to\infty}{\to} g(\theta)v,$$

where $g(\theta)$ is a phase. We can make the $|Dz|^2$ term happy by setting

$$A \overset{r\to\infty}{\to} -\mathbf{i}g\partial_\mu g + \mathcal{O}(r^{-2}).$$

Then the F^2 term is automatically happy.

What are the possible $g(\theta)$? g is a map from the circle at infinity to the circle of phases $g : S^1 \to S^1$. Such maps are classified by a winding number, $Q \in \mathbb{Z}$. A representative of each class is $g(\theta) = e^{\mathbf{i}Q\theta}$. This function gives

$$\int_{\text{spacetime}} \frac{F}{2\pi} = \oint \frac{A}{2\pi} = -\mathbf{i}\int_0^{2\pi} đ\theta e^{-\mathbf{i}Q\theta}(+\mathbf{i}Q)e^{\mathbf{i}Q\theta} = Q.$$

The winding number determines the flux.

This means the partition function is

$$Z = \int [dAdz]e^{-S[A,z]} = \sum_{Q_T\in\mathbb{Z}} e^{\mathbf{i}\theta Q_T} Z_{Q_T} \simeq \sum_{Q_T} e^{\mathbf{i}\theta Q_T} e^{-S_0^{Q_T}} \frac{1}{\det S''_{Q_T}}.$$

In the last step I made a caricature of the saddle point approximation. Notice the dependence of the instanton ($Q \neq 0$) contributions: if we scale out an overall coupling (by rescaling fields) and write the action as $S[\phi] = \frac{1}{g^2}S[\phi, \text{ratios of couplings}]$, then $e^{-S_0} = e^{-\frac{1}{g^2}S[\underline{\phi},\text{ratios}]}$ is non-analytic at $g = 0$ – all the terms of its taylor expansion vanish at $g = 0$. This is not something we could ever produce by perturbation series, it is *non-perturbative*. Notice that it is also *small* at weak coupling. However, sometimes it is the leading contribution, *e.g.* to the energy of a metastable vacuum. (For more on this, see Coleman.)

To do better, we need to understand the saddle points better.

1. First, in the instanton solution we found, we picked a center, the location of the core of the vortex. But in fact, there is a solution for any center x_0^μ, with the same action. This means the determinant of S'' actually has a zero! The resolution is simple: There is actually a *family* of saddles, labelled by the *collective coordinate* x_0^μ. We just have to *do* the integral over these coordinates. The result is simple:

it produces a factor of $\int d^D x_0 = VT$ where VT is the volume of spacetime. The contribution of one instanton to the integral is then

$$Ke^{-S_0} e^{\mathbf{i}\theta} VT$$

for some horrible constant K.

2. Second, since the vortex solution is localized, we can make arbitrarily-close-to-solutions by introducing multiple vortices with their respective centers arbitrarily far from each other. The Q_T is actually the sum of the instanton numbers. If they are far enough apart, their actions also add. Each center has its own collective coordinate and produces its own factor of VT.

3. We can also have anti-instantons. This just means that individual Qs can be negative.

So we are going to approximate our integral by a dilute gas of instantons and anti-instantons. Their actions add. A necessary condition for this to be a good idea is that $VT \gg (\text{core size})^2$. $e^{\mathbf{i}\theta}$ is the instanton fugacity.

$$\begin{aligned} Z &= \text{Tr } e^{-TH} \overset{T\to\infty}{\simeq} \sum_{n,\bar{n}} \left(Ke^{-S_0}\right)^{n+\bar{n}} (VT)^{n+\bar{n}} e^{\mathbf{i}(n-\bar{n})\theta} \frac{1}{n!\bar{n}!} \\ &= \left(\sum_n \frac{1}{n!} \left(Ke^{-S_0} VT\right)^n e^{\mathbf{i}n\theta} \right) \times (h.c.) \\ &= e^{VTKe^{-S_0} e^{\mathbf{i}\theta} + h.c.} = e^{VT2Ke^{-S_0} \cos\theta}. \end{aligned} \tag{13.1}$$

We should be happy about this answer. Summing over the dilute gas of instantons gives an extensive contribution to the free energy. The free energy per unit time in the euclidean path integral is the groundstate energy density:

$$Z = \text{Tr } e^{-TH} \overset{T\to\infty}{\simeq} e^{-TV\mathcal{E}(\theta)}, \quad \Longrightarrow \quad \mathcal{E}(\theta) = -2K\cos\theta e^{-S_0}.$$

[End of Lecture 53]

We can also calculate the expected flux:

$$\left\langle \frac{\int F}{2\pi} \right\rangle = \sum_Q Q e^{\mathbf{i}\theta Q} e^{-S} \sum_Q e^{\mathbf{i}\theta Q} e^{-S} = -\mathbf{i}\partial_\theta \ln Z(\theta) = 2KV \sin\theta e^{-S_0}.$$

Therefore, when $\theta \neq 0 \bmod \pi$, there is a nonzero electric field in the vacuum: $\langle F_{01} \rangle = E \neq 0$. It is uniform.

A small variation of this calculation gives the force between external charges:

$$\left\langle W\left[\begin{array}{l}\square \updownarrow T' \\ \overset{\leftrightarrow}{L}' \\ \\ \end{array}\right]\right\rangle = \left\langle e^{\mathbf{i}\frac{q}{e}\oint_\square A_\mu dx^\mu}\right\rangle = \left\langle e^{\mathbf{i}\frac{q}{e}\int_\blacksquare F}\right\rangle$$

This has the effect of shifting the value of θ on the *inside* of the loop to $\theta_{\text{in}} \equiv \theta + \frac{q}{e}2\pi$. So the answer in the dilute instanton gas approximation is

$$\left\langle W\left[\begin{array}{l}\square \updownarrow T' \\ \overset{\leftrightarrow}{L}' \\ \\ \end{array}\right]\right\rangle = \frac{\exp\left(2Ke^{S_0}\left(\underbrace{(LT - L'T')\cos\theta}_{\text{inside}} + \underbrace{L'T'\cos\theta_{\text{in}}}_{\text{outside}}\right)\right)}{e^{2Ke^{-S_0}LT\cos\theta}} = e^{-T'V(L')}$$

with

$$V(L') = L'2Ke^{-S_0}\left(\cos\theta - \cos\left(\theta + 2\pi\frac{q}{e}\right)\right)$$

which is linear in the separation between the charges – linear confinement, except when $q = ne, n \in \mathbb{Z}$.

Here's how to think about this result. For small $\theta, q/e$, the potential between charges is

$$V(L') \overset{\theta \ll 1}{\simeq} L'Ke^{-S_0}\left(\left(\theta + 2\pi\frac{q}{e}\right)^2 - \theta^2\right)$$

and the energy and flux are

$$\mathcal{E}(\theta) \overset{\theta \ll 1}{\simeq} 2Ke^{-S_0}\theta^2 + \text{const}, \qquad \langle F\rangle \overset{\theta \ll 1}{\simeq} 4\pi Ke^{-S_0}\theta.$$

θ is like the charge on a pair of parallel capacitor plates at $x = \infty$. Adding charge and anticharge changes the electric field in between, and the energy density is quadratic in the field, $U \propto E^2$. But what happens when $q = ne$? Notice that the potential is actually periodic in $q \to q + ne$. If $L' > \frac{1}{2\mu}$ (μ is the mass of the z excitations), then the energy can be decreased by pair-creating a z and $z^\dagger$, which then fly to the capacitor plates and discharge them, changing $\theta \to \theta - 2\pi$.

Comments about $D = 4$. Some of the features of this story carry over to gauge theory in $D = 3 + 1$. Indeed there is a close parallel between the $\theta \int_2 F$ term and the $\theta \int_4 F \wedge F$ term. In 4d, too, there are solutions of the euclidean equations (even in pure Yang-Mills theory) which are localized in spacetime. (The word *instanton* is sometimes used to refer to these solutions, even when they appear in other contexts than euclidean saddle points. These solutions were found by Belavin, Polyakov, Schwartz and Tyupin.) Again, the gauge field looks like a gauge transformation at ∞:

$$A \overset{r\to\infty}{\to} -\mathbf{i}g\partial_\mu g + \mathcal{O}(r^{-\#}).$$

Now g is a map from the 3-sphere at infinity (in euclidean 4-space) to the gauge group, $g : S^3 \to \mathsf{G}$. Any simple Lie group has an $\mathsf{SU}(2) \simeq S^3$ inside, and there is an integer classification of such maps. So again there is a sum over $Q \in \mathbb{Z}$. However: the calculation leading to confinement does *not* go through so simply. The 4d θ term does not produce a nonzero electric field in the vacuum, and an external charge isn't like a capacitor plate. As Coleman says, whatever causes confinement in 4d gauge theory, it's not instantons.

13.2 Blobology (aka Large Deviation Theory)

Many bits of the following discussion are already familiar, but I like the organization.

Feynman diagrams from the path integral. Now that we are using path integrals all the time, the diagrammatic expansion is much less mysterious (perhaps we should have started here, like Zee does? maybe next time). Much of what we have to say below is still interesting for QFT in $0+0$ dimensions, which means integrals. If everything is positive, this is probability theory. Suppose we want to do the integral

$$Z(J) = \int_{-\infty}^{\infty} dq\; e^{-\frac{1}{2}m^2q^2 - \frac{g}{4!}q^4 + Jq} \equiv \int dq\; e^{-S(q)} \; . \tag{13.2}$$

It is the path integral for ϕ^4 theory with fewer labels. For $g = 0$, this is a gaussian integral which we know how to do. For $g \neq 0$ it's not an elementary function of its arguments. We can develop a (non-convergent!) series expansion in g by writing it as

$$Z(J) = \int_{-\infty}^{\infty} dq\; e^{-\frac{1}{2}m^2q^2 + Jq}\left(1 - \frac{g}{4!}q^4 + \frac{1}{2}\left(-\frac{g}{4!}q^4\right)^2 + \cdots\right)$$

and integrating term by term. And the term with q^{4n} (that is, the coefficient of g^n) is

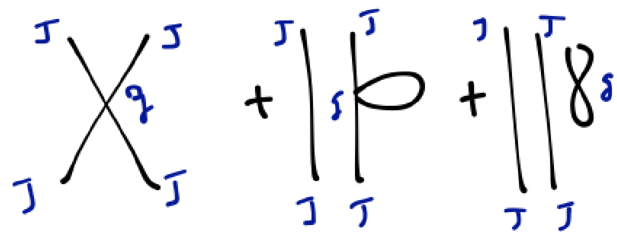
$$\int_{-\infty}^{\infty} dq\; e^{-\frac{1}{2}m^2q^2 + Jq}q^{4n} = \left(\frac{\partial}{\partial J}\right)^{4n}\int_{-\infty}^{\infty} dq\; e^{-\frac{1}{2}m^2q^2 + Jq} = \left(\frac{\partial}{\partial J}\right)^{4n} e^{\frac{1}{2}J\frac{1}{m^2}J}\sqrt{\frac{2\pi}{m^2}} \; .$$

So:

$$Z(J) = \sqrt{\frac{2\pi}{m^2}}e^{-\frac{g}{4!}\left(\frac{\partial}{\partial J}\right)^4}e^{\frac{1}{2}J\frac{1}{m^2}J} \; .$$

This is a double expansion in powers of J and powers of g. The process of computing the coefficient of $J^n g^m$ can be described usefully in terms of diagrams. There is a factor of $1/m^2$ for each line (the propagator), and a factor of $(-g)$ for each 4-point vertex (the coupling), and a factor of J for each external line (the source). For example, the coefficient of gJ^4 comes from:

$$\sim \left(\frac{1}{m^2}\right)^4 gJ^4.$$



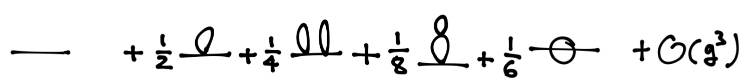
$$\propto \left(\frac{1}{m^2}\right)^4 J^4 g$$

There is a symmetry factor which comes from expanding the exponential: if the diagram has some symmetry preserving the external labels, the multiplicity of diagrams does not completely cancel the $1/n!$.

As another example, consider the analog of the two-point function:

$$G \equiv \langle q^2 \rangle |_{J=0} = \frac{\int dq\ q^2\ e^{-S(q)}}{\int dq\ e^{-S(q)}} = -2\frac{\partial}{\partial m^2} \log Z(J=0).$$

In perturbation theory this is:

$G \simeq$ 

$$= m^{-2}\left(1 \quad - \frac{1}{2} g m^{-2} \quad + \quad \frac{2}{3} g^2 m^{-4} \quad + \mathcal{O}(g^3)\right) \tag{13.3}$$

Brief comments about large orders of perturbation theory.

- How do I know the perturbation series about $g = 0$ doesn't converge? One way to see this is to notice that if I made g even infinitesimally negative, the integral itself would not converge (the potential would be unbounded below), and $Z_{g=-|\epsilon|}$ is not defined. Therefore Z_g as a function of g cannot be analytic in a neighborhood of $g = 0$. This argument is due to Dyson.

- The expansion of the exponential in the integrand is clearly convergent for each q. The place where we went wrong is exchanging the order of integration over q and summation over n.

- The integral actually does have a name – it's a Bessel function:

$$Z(J=0) = \frac{2}{\sqrt{m^2}} \sqrt{\rho} e^{\rho} K_{\frac{1}{4}}(\rho), \qquad \rho \equiv \frac{3m^4}{4g}$$

  (for $\mathrm{Re}\sqrt{\rho} > 0$), as `Mathematica` will tell you. Because we know about Bessel functions, in this case we can actually figure out what happens at strong coupling, when $g \gg m^4$, using the asymptotics of the Bessel function.

- In this case, the perturbation expansion too can be given a closed form expression:

$$Z(0) \simeq \sqrt{\frac{2\pi}{m^2}} \sum_n \frac{(-1)^n}{n!} \frac{2^{2n+\frac{1}{2}}}{(4!)^n} \Gamma\left(2n + \frac{1}{2}\right) \left(\frac{g}{m^4}\right)^n. \tag{13.4}$$

- The expansion for G is of the form

$$G \simeq m^{-2} \sum_{n=0}^{\infty} c_n \left(\frac{g}{m^4}\right)^n.$$

  When n is large, the coefficients satisfy $c_{n+1} \overset{n \gg 1}{\simeq} -\frac{2}{3} n c_n$ (you can see this by looking at the coefficients in (13.4)) so that $|c_n| \sim n!$. This factorial growth of the number of diagrams is general in QFT and is another way to see that the series does not converge.

- The fact that the coefficients c_n grow means that there is a *best* number of orders to keep. The errors start getting bigger when $c_{n+1}\left(\frac{g}{m^4}\right) \sim c_n$, that is, at order $n \sim \frac{3m^4}{2g}$. So if you want to evaluate G at this value of the coupling, you should stop at that order of n.

- A technique called *Borel resummation* can sometimes produce a well-defined function of g from an asymptotic series whose coefficients diverge like $n!$. The idea is to make a new series

$$B(z) \equiv \sum_{m=0} \frac{c_m}{n!} z^m$$

  whose coefficients are ensmallened by $n!$. Then to get back $Z(g)$ we use the identity

$$1 = \frac{1}{n!} \int_0^{\infty} dz e^{-z} z^n$$

  and do the Laplace transform of $B(z)$:

$$\int_0^{\infty} dz B(z) e^{-z/g} = \sum_{m=0} c_m \frac{\int_0^{\infty} dz e^{-z/g} z^m}{m!} = g \sum_{m=0}^{\infty} c_m g^m = g Z(g).$$

  This procedure requires both that the series in $B(z)$ converges and that the Laplace transform can be done. In fact this procedure works in this case.

  The existence of saddle-point contributions to $Z(g)$ which go like $e^{-a/g}$ imply that the number of diagrams at large order grows like $n!$. This is because they are associated with singularities of $B(z)$ at $z = a$; such a singularity means the sum of $\frac{c_n}{n!} z^n$ must diverge at $z = a$. (More generally, non-perturbative effects which go like $e^{-a/g^{1/p}}$ (larger if $p > 1$) are associated with (faster) growth like $(pn)!$. See this classic work.)

- The function $G(g)$ can be analytically continued in g away from the real axis, and can in fact be defined on the whole complex g plane. It has a branch cut on the negative real axis, across which its discontinuity is related to its imaginary part. The imaginary part goes like $e^{-\frac{a}{|g|}}$ near the origin and can be computed by a tunneling calculation.

  How did we know Z has a branch cut? One way is from the asymptotics of the Bessel function. But, better, *why* does Z satisfy the Bessel differential equation as a function of the couplings? The answer, as you'll check on the homework, is that the Bessel equation is a Schwinger-Dyson equation,

$$0 = \int_{-\infty}^{\infty} \frac{\partial}{\partial q} \left(\text{something} e^{-S(q)} \right)$$

  which results from demanding that we can change integration variables in the path integral.

For a bit more about this, you might look at sections 3 and 4 of this recent paper from which I got some of the details here. See also the giant book by Zinn-Justin. There is a deep connection between the large-order behavior of the perturbation series about the trivial saddle point and the contributions of non-trivial saddle points. The keywords for this connection are *resurgence* and *trans-series* and a starting references is here.

The Feynman diagrams we've been drawing all along are the same but with more labels. Notice that each of the qs in our integral could come with a label, $q \to q_a$. Then each line in our diagram would be associated with a matrix $(m^{-2})_{ab}$ which is the inverse of the quadratic term $q_a m^2_{ab} q_b$ in the action. If our diagrams have loops we get free sums over the label. If that label is conserved by the interactions, the vertices will have some delta functions. In the case of translation-invariant field theories we can label lines by the conserved momentum k. Each comes with a factor of the free propagator $\frac{\mathbf{i}}{k^2+m^2+\mathbf{i}\epsilon}$, each vertex conserves momentum, so comes with $\mathbf{i}g\delta^D\left(\sum k\right)(2\pi)^D$, and we must integrate over momenta on internal lines $\int đ^D k$.

Next, three general organizing facts about the diagrammatic expansion, two already familiar. In thinking about the combinatorics below, we will represent collections of Feynman diagrams by blobs with legs sticking out, and think about how the blobs combine. Then we can just renormalize the appropriate blobs and be done.

The following discussion will look like I am talking about a field theory with a single scalar field. But really each of the ϕs is a collection of fields and all the indices are too small to see. This is yet another example of coarse-graining.

1. **Disconnected diagrams exponentiate.** [Zee, I.7, Banks, chapter 3] Recall that the Feynman rules come with a (often annoying, here crucial) statement about symmetry factors: we must divide the contribution of a given diagram by the order of the symmetry group of the diagram (preserving various external labels). For a diagram with k identical disconnected pieces, this symmetry group includes the permutation group S_k which permutes the identical pieces and has $k!$ elements. (Recall that the origin of the symmetry factors is that symmetric feynman diagrams fail to completely cancel the $1/n!$ in the Dyson formula. For a reminder about this, see *e.g.* Peskin p. 93.) Therefore:

$$Z = \sum (\text{all diagrams}) = e^{\sum(\text{connected diagrams})} = e^{\mathbf{i}W}.$$

You can go a long way towards convincing yourself of this by studying the case where there are only two connected diagrams $A+B$ (draw whatever two squiggles you want) and writing out e^{A+B} in terms of disconnected diagrams with symmetry factors.

Notice that this relationship is just like that of the partition function to the (Helmholtz) free energy $Z = e^{-\beta F}$ (modulo the factor of $\mathbf{i}$) in statistical mechanics (and is the same as that relationship when we study the euclidean path integral with periodic boundary conditions in euclidean time). This statement is extremely general. It remains true if we include external sources:

$$Z[J] = \int [D\phi] e^{\mathbf{i}S[\phi]+\mathbf{i}\int \phi J} = e^{\mathbf{i}W[J]}.$$

Now the diagrams have sources J at which propagator lines can terminate; (the perturbation theory approximation to) $W[J]$ is the sum of all connected such diagrams. For example

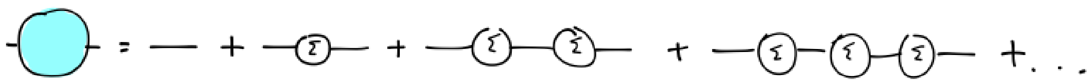
$$\langle \phi(x) \rangle = \frac{1}{Z}\frac{\delta}{\mathbf{i}\delta J(x)} Z = \frac{\delta}{\mathbf{i}\delta J(x)} \log Z = \frac{\delta}{\delta J(x)} W$$

$$\langle \mathcal{T} \phi(x)\phi(y) \rangle = \frac{\delta}{\mathbf{i}\delta J(x)}\frac{\delta}{\mathbf{i}\delta J(y)} \log Z = \frac{\delta}{\mathbf{i}\delta J(x)}\frac{\delta}{\mathbf{i}\delta J(y)} \mathbf{i}W \ .$$

(Note that here $\langle \phi \rangle \equiv \langle \phi \rangle_J$ depends on J. You can set it to zero if you want, but the equation is true for any J.) If you forget to divide by the normalization Z, and instead look at just $\frac{\delta}{\delta J(x)}\frac{\delta}{\delta J(y)} Z$, you get *disconnected* quantities like $\langle \phi \rangle \langle \phi \rangle$ (the terminology comes from the diagrammatic representation). [22] The point in life of W is that by differentiating it with respect to J we can construct all the connected Green's functions.

[22]More precisely: $\frac{\delta}{\delta J(x)}\frac{\delta}{\delta J(y)} Z = \frac{\delta}{\delta J(x)} \left(\langle \phi(x) \rangle_J Z \right) = \langle \phi(x) \rangle_J \langle \phi(y) \rangle_J Z + \langle \phi(x)\phi(y) \rangle_J Z.$

2. **Propagator corrections form a geometric series.** This one I don't need to say more about:



3. **The sum of all connected diagrams is the Legendre transform of the sum of the 1PI diagrams.**

   [Banks, 3.8; Zee IV.3; Schwarz §34, Srednicki §21] A simpler way to say our third fact is

$$\sum(\text{connected diagrams}) = \sum(\text{connected } \textit{tree} \text{ diagrams with 1PI vertices})$$

   where a *tree* diagram is one with no loops. But the description in terms of Legendre transform will be extremely useful. Along the way we will show that the perturbation expansion is a semi-classical expansion. And we will construct a useful object called the 1PI effective action Γ. The basic idea is that we can construct the actual correct correlation functions by making *tree* diagrams ($\equiv$ diagrams with no loops) using the 1PI effective action as the action.

   Notice that this is a very good reason to care about the notion of 1PI: if we sum all the tree diagrams using the 1PI blobs, we clearly are including all the diagrams. Now we just have to see what machinery will pick out the 1PI blobs. The answer is: Legendre transform. There are many ways to go about showing this, and all involve a bit of complication. Bear with me for a bit; we will learn a lot along the way.

   Def'n of ϕ_c, the 'classical field'. Consider the functional integral for a scalar field theory:

$$Z[J] = e^{\mathbf{i}W[J]} = \int [D\phi] e^{\mathbf{i}\left(S[\phi] + \int J\phi\right)} . \tag{13.5}$$

   Define

$$\phi_c(x) \equiv \frac{\delta W[J]}{\delta J(x)} = \frac{1}{Z} \int [D\phi] e^{\mathbf{i}\left(S[\phi]+\int J\phi\right)} \phi(x) = \langle 0| \hat{\phi}(x) |0\rangle \ . \tag{13.6}$$

   This is the vacuum expectation value of the field operator, in the presence of the source J. Note that $\phi_c(x)$ is a functional of J.

Warning: we are going to use the letter ϕ for many conceptually distinct objects here: the functional integration variable ϕ, the quantum field operator $\hat{\phi}$, the classical field ϕ_c. I will not always use the hats and subscripts.

[End of Lecture 54]

Legendre Transform. Next we recall the notion of Legendre transform and extend it to the functional case: Given a function L of $\dot{q}$, we can make a new function H of p (the Legendre transform of L with respect to $\dot{q}$) defined by:

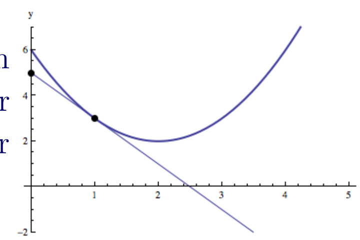
$$H(p,q) = p\dot{q} - L(\dot{q},q).$$

On the RHS here, $\dot{q}$ must be eliminated in favor of p using the relation $p = \frac{\partial L}{\partial \dot{q}}$. You've also seen this manipulation in thermodynamics using these letters:

$$F(T,V) = E(S,V) - TS, \quad T = \frac{\partial E}{\partial S}|_V .$$

The point of this operation is that it relates the free energies associated with different ensembles in which different variables are held fixed.

More mathematically, it encodes a function (at least one with nonvanishing second derivative, *i.e.* one which is convex or concave) in terms of its envelope of tangents. For further discussion of this point of view, look here.



Now the functional version: Given a functional $W[J]$, we can make a new associated functional Γ of the conjugate variable ϕ_c:

$$\Gamma[\phi_c] \equiv W[J] - \int J\phi_c.$$

Again, the RHS of this equation defines a functional of ϕ_c implicitly by the fact that J can be determined from ϕ_c, using (13.6)[23].

Interpretation of ϕ_c. How to interpret ϕ_c? It's some function of spacetime, which depends on the source J. **Claim:** It solves

$$-J(x) = \frac{\delta\Gamma[\phi_c]}{\delta\phi_c(x)} \tag{13.7}$$

So, in particular, when $J = 0$, it solves

$$0 = \frac{\delta\Gamma[\phi_c]}{\delta\phi_c(x)}|_{\phi_c = \langle\phi\rangle} \tag{13.8}$$

[23]Come back later and worry about what happens if J is not determined uniquely.

– the extremum of the effective action is $\langle\phi\rangle$. This gives a classical-like equation of motion for the field operator expectation value in QFT.

Proof of (13.7): $$\frac{\delta\Gamma[\phi_c]}{\delta\phi_c(x)} = \frac{\delta}{\delta\phi_c(x)}\left(W[J] - \int dy J(y)\phi_c(y)\right)$$

What do we do here? We use the functional product rule – there are three places where the derivative hits:

$$\frac{\delta\Gamma[\phi_c]}{\delta\phi_c(x)} = \frac{\delta W[J]}{\delta\phi_c(x)} - J(x) - \int dy \frac{\delta J(y)}{\delta\phi_c(x)}\phi_c(y)$$

In the first term we must use the functional chain rule:

$$\frac{\delta W[J]}{\delta\phi_c(x)} = \int dy \frac{\delta J(y)}{\delta\phi_c(x)}\frac{\delta W[J]}{\delta J(y)} = \int dy \frac{\delta J(y)}{\delta\phi_c(x)}\phi_c(y).$$

So we have:

$$\frac{\delta\Gamma[\phi_c]}{\delta\phi_c(x)} = \int dy \frac{\delta J(y)}{\delta\phi_c(x)}\phi_c(y) - J(x) - \int dy \frac{\delta J(y)}{\delta\phi_c(x)}\phi_c(y) = -J(x). \tag{13.9}$$

Now $\phi_c|_{J=0} = \langle\phi\rangle$. So if we set $J = 0$, we get the equation (13.8) above. So (13.8) replaces the action principle in QFT – to the extent that we can calculate $\Gamma[\phi_c]$. (Note that there can be more than one extremum of Γ. That requires further examination.)

Next we will build towards a demonstration of the diagrammatic interpretation of the Legendre transform; along the way we will uncover important features of the structure of perturbation theory.

Semiclassical expansion of path integral. Recall that the Legendre transform in thermodynamics is the leading term you get if you compute the partition function by saddle point – the classical approximation. In thermodynamics, this comes from the following manipulation: the thermal partition function is:

$$Z = e^{-\beta F} = \mathrm{tr} e^{-\beta \mathbf{H}} = \int dE \underbrace{\Omega(E)}_{(\text{density of states with energy } E) = e^{S(E)}} e^{-\beta E} \overset{\text{saddle}}{\approx} e^{S(E_\star) - \beta E_\star}|_{E_\star \text{ solves } \partial_E S = \beta} .$$

The log of this equation then says $F = E - TS$ with S eliminated in favor of T by $T = \frac{1}{\partial_E S}|_V = \partial_S E|_V$, *i.e.* the Legendre transform we discussed above. In simple thermodynamics the saddle point approx is justified by the thermodynamic limit: the quantity in the exponent is extensive, so the saddle point is well-peaked. This part of the analogy will not always hold, and we will need to think about fluctuations about the saddle point.

Let's go back to (13.5) and think about its semiclassical expansion. If we were going to do this path integral by stationary phase, we would solve

$$0 = \frac{\delta}{\delta\phi(x)}\left(S[\phi] + \int \phi J\right) = \frac{\delta S}{\delta\phi(x)} + J(x) \ . \tag{13.10}$$

This determines some function ϕ which depends on J; let's denote it here as $\phi^{[J]}(x)$. In the semiclassical approximation to $Z[J] = e^{\mathbf{i}W[J]}$, we would just plug this back into the exponent of the integrand:

$$W_c[J] = \frac{1}{g^2\hbar}\left(S[\phi^{[J]}] + \int J\phi^{[J]}\right).$$

So in this approximation, (13.10) is exactly the equation determining ϕ_c. This is just the Legendre transformation of the original bare action $S[\phi]$ (I hope this manipulation is also familiar from stat mech, and I promise we're not going in circles).

Let's think about expanding $S[\phi]$ about such a saddle point $\phi^{[J]}$ (or more correctly, a point of stationary phase). The stationary phase (or semi-classical) expansion familiar from QM is an expansion in powers of $\hbar$ (WKB):

$$Z = e^{\mathbf{i}W/\hbar} = \int dx\, e^{\frac{\mathbf{i}}{\hbar}S(x)} = \int dx e^{\frac{\mathbf{i}}{\hbar}\left(S(x_0)+(x-x_0)\underbrace{S'(x_0)}_{=0}+\frac{1}{2}(x-x_0)^2S''(x_0)+\ldots\right)} = e^{\mathbf{i}W_0/\hbar+\mathbf{i}W_1+\mathbf{i}\hbar W_2+\ldots}$$

with $W_0 = S(x_0)$, and W_n comes from (the exponentiation of) diagrams involving n contractions of $\delta x = x - x_0$, each of which comes with a power of $\hbar$: $\langle \delta x \delta x\rangle \sim \hbar$.

Expansion in $\hbar$ = expansion in coupling. Is this semiclassical expansion the same as the expansion in powers of the coupling? Yes, if there is indeed a notion of "*the* coupling", *i.e.* only one for each field. Then by a rescaling of the fields we can put all the dependence on the coupling in front:

$$S = \frac{1}{g^2}s[\phi]$$

so that the path integral is

$$\int [D\phi]\, e^{\mathbf{i}\frac{s[\phi]}{\hbar g^2}+\int \phi J}.$$

(It may be necessary to rescale our sources J, too.) For example, suppose we are talking about a QFT of a single field $\tilde{\phi}$ with action

$$S[\tilde{\phi}] = \int \left(\left(\partial\tilde{\phi}\right)^2 - \lambda\tilde{\phi}^p\right).$$

Then define $\phi \equiv \tilde{\phi}\lambda^\alpha$ and choose $\alpha = \frac{1}{p-2}$ to get

$$S[\phi] = \frac{1}{\lambda^{\frac{2}{p-2}}} \int \left((\partial\phi)^2 - \phi^p \right) = \frac{1}{g^2} s[\phi].$$

with $g \equiv \lambda^{\frac{1}{p-2}}$, and $s[\phi]$ independent of g. Then the path-integrand is $e^{\frac{\mathbf{i}}{\hbar g^2} s[\phi]}$ and so g and $\hbar$ will appear only in the combination $g^2\hbar$. (If we have more than one coupling term, this direct connection must break down; instead we can scale out some overall factor from all the couplings and that appears with $\hbar$.)

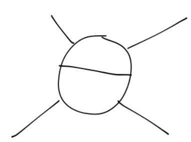
Loop expansion = expansion in coupling. Now I want to convince you that this is also the same as the loop expansion. The first correction in the semi-classical expansion comes from

$$S_2[\phi_0, \delta\phi] \equiv \frac{1}{g^2} \int dxdy\delta\phi(x)\delta\phi(y) \frac{\delta^2 s}{\delta\phi(x)\delta\phi(y)}|_{\phi=\phi_0}.$$

For the accounting of powers of g, it's useful to define $\Delta = g^{-1}\delta\phi$, so the action is

$$g^{-2}s[\phi] = g^{-2}s[\phi_0] + S_2[\Delta] + \sum_n g^{n-2}V_n[\Delta].$$

With this normalization, the power of the field Δ appearing in each term of the action is correlated with the power of g in that term. And the Δ propagator is independent of g.

So use the action $s[\phi]$, in an expansion about $\phi_\star$ to construct Feynman rules for correlators of Δ: the propagator is $\langle \mathcal{T}\Delta(x)\Delta(y)\rangle \propto g^0$, the 3-point vertex comes from V_3 and goes like $g^{3-2=1}$, and so on. Consider a diagram that contributes to an E-point function (of Δ) at order g^n, for example this contribution to the $(E = 4)$-point function at order $n = 6 \cdot (3-2) = 6$:  With our normalization of Δ, the powers of g come only from the vertices; a degree k vertex contributes $k-2$ powers of g; so the number of powers of g is

$$n = \sum_{\text{vertices, } i} (k_i - 2) = \sum_i k_i - 2V \tag{13.11}$$

where

$$
\begin{aligned}
V &= \#\text{ of vertices } \ (\text{This does not include external vertices.})\\
&\quad \text{We also define:}\\
n &= \#\text{ of powers of } g\\
L &= \#\text{ of loops } = \#\text{of independent internal momentum integrals}\\
I &= \#\text{ of internal lines } = \#\text{ of internal propoagators}\\
E &= \#\text{ of external lines}
\end{aligned}
$$

Facts about graphs:

- The total number of lines leaving all the vertices is equal to the total number of lines:
$$\sum_{\text{vertices, } i} k_i = E + 2I. \tag{13.12}$$
So the number of internal lines is
$$I = \frac{1}{2}\left(\sum_{\text{vertices, } i} k_i - E\right). \tag{13.13}$$

- For a connected graph, the number of loops is
$$L = I - V + 1 \tag{13.14}$$
since each loop is a sequence of internal lines interrupted by vertices. (This fact is probably best proved inductively. The generalization to graphs with multiple disconnected components is $L = I - V + C$.)

We conclude that[24]

$$L \overset{(13.14)}{=} I - V + 1 \overset{(13.13)}{=} \frac{1}{2}\left(\sum_i k_i - E\right) - V + 1 = \frac{n-E}{2} + 1 \overset{(13.11)}{=} \frac{n-E}{2} + 1.$$

This equation says:

$L = \frac{n-E}{2} + 1$: More powers of g means (linearly) more loops.

[24]You should check that these relations are all true for some random example, like the one above, which has $I = 7, L = 2, \sum k_i = 18, V = 6, E = 4$. You will notice that Banks has several typos in his discussion of this in §3.4. His Es should be $E/2$s in the equations after (3.31).

Diagrams with a fixed number of external lines and more loops are suppressed by more powers of g. (By rescaling the external field, it is possible to remove the dependence on E.)

We can summarize what we've learned by writing the sum of connected graphs as

$$W[J] = \sum_{L=0}^{\infty} \left(g^2\hbar\right)^{L-1} W_L$$

where W_L is the sum of connected graphs with L loops. In particular, the order-$\hbar^{-1}$ (classical) bit W_0 comes from *tree graphs*, graphs without loops. Solving the classical equations of motion sums up the tree diagrams.

Diagrammatic interpretation of Legendre transform. $\Gamma[\phi]$ is called the 1PI effective action[25]. And as its name suggests, Γ has a diagrammatic interpretation: it is the sum of just the 1PI connected diagrams. (Recall that $W[J]$ is the sum of *all* connected diagrams.) Consider the (functional) Taylor expansion Γ_n in ϕ

$$\Gamma[\phi] = \sum_n \frac{1}{n!} \int \Gamma_n(x_1...x_n)\phi(x_1)...\phi(x_n) d^D x_1 \cdots d^D x_n \ .$$

The coefficients Γ_n are called 1PI Green's functions (we will justify this name presently). To get the full connected Green's functions, we sum all tree diagrams with the 1PI Green's functions as vertices, using the full connected two-point function as the propagators.

Perhaps the simplest way to arrive at this result is to consider what happens if we try to use Γ as the action in the path integral instead of S.

$$Z_{\Gamma,\hbar}[J] \equiv \int [D\phi] e^{\frac{\mathbf{i}}{\hbar}\left(\Gamma[\phi] + \int J\phi\right)}$$

By the preceding arguments, the expansion of $\log Z_\Gamma[J]$ in powers of $\hbar$, in the limit $\hbar \to 0$ is

$$\lim_{\hbar\to 0} \log Z_{\Gamma,\hbar}[J] = \sum_L \left(g^2\hbar\right)^{L-1} W_L^\Gamma \ .$$

The leading, tree level term in the $\hbar$ expansion, is obtained by solving

$$\frac{\delta\Gamma}{\delta\phi(x)} = -J(x)$$

[25]The 1PI effective action Γ must be distinguished from the *Wilsonian effective action* – the difference is that here we integrated over everybody, whereas the Wilsonian action integrates only high-energy modes. The different effective actions correspond to different choices about what we care about and what we don't, and hence different choices of what modes to integrate out.

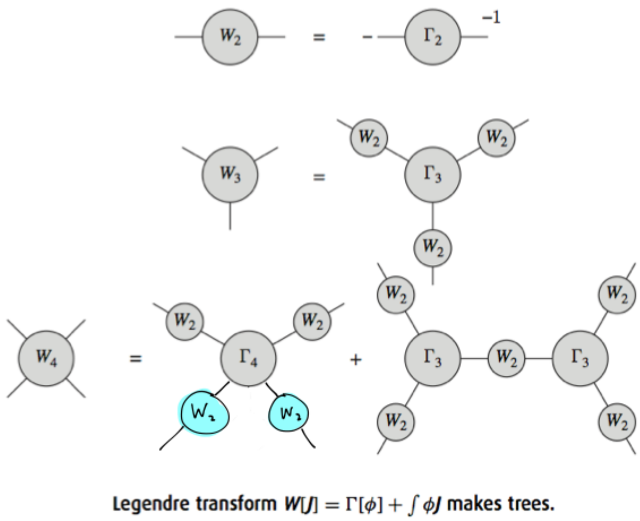


Figure 2: [From Banks, *Modern Quantum Field Theory*, slightly improved] W_n denotes the connected n-point function, $\left(\frac{\partial}{\partial J}\right)^n W[J] = \langle \phi^n \rangle$.

and plugging the solution into Γ; the result is

$$\left(\Gamma[\phi] + \int \phi J \right)_{\frac{\partial \Gamma}{\partial \phi(x)} = -J(x)} \overset{\text{inverse Legendre transf}}{\equiv} W[J].$$

This expression is the definition of the inverse Legendre transform, and we see that it gives back $W[J]$: the generating functional of connected correlators! On the other hand, the counting of powers above indicates that the only terms that survive the $\hbar \to 0$ limit are tree diagrams where we use the terms in the Taylor expansion of $\Gamma[\phi]$ as the vertices. This is exactly the statement we were trying to demonstrate: the sum of all connected diagrams is the sum of tree diagrams made using 1PI vertices and the exact propagator (by definition of 1PI). Therefore Γ_n are the 1PI vertices.

For a more arduous but more direct proof of this statement, see the problem set and/or Banks §3.5. There is an important typo on page 29 of Banks' book; it should say:

$$\frac{\delta^2 W}{\delta J(x) \delta J(y)} = \frac{\delta \phi(y)}{\delta J(x)} = \left(\frac{\delta J(x)}{\delta \phi(y)} \right)^{-1} \overset{(13.9)}{=} - \left(\frac{\delta^2 \Gamma}{\delta \phi(x) \delta \phi(y)} \right)^{-1}. \tag{13.15}$$

(where $\phi \equiv \phi_c$ here). You can prove this from the definitions above. Inverse here means in the sense of integral operators: $\int d^D z K(x,z) K^{-1}(z,y) = \delta^D(x-y)$. So we can write the preceding result more compactly as:

$$W_2 = -\Gamma_2^{-1}.$$

Here's two ways to think about why we get an inverse here: (1) diagrammatically, the 1PI blob is defined by removing the external propagators; but these external propagators are each W_2; removing two of them from one of them leaves -1 of them. You're on your own for the sign. (2) In the expansion of $\Gamma = \sum_n \int \Gamma_n \phi^n$ in powers of the field, the second term is $\int\int \phi\Gamma_2\phi$, which plays the role of the kinetic term in the effective action (which we're instructed to use to make tree diagrams). The full propagator is then the inverse of the kinetic operator here, namely Γ_2^{-1}. Again, you're on your own for the sign.

The idea to show the general case in Fig. 2 is to just compute W_n by taking the derivatives starting from (13.15): Differentiate again wrt J and use the matrix differentiation formula $dK^{-1} = -K^{-1}dKK^{-1}$ and the chain rule to get

$$W_3(x,y,z) = \int dw_1 \int dw_2 \int dw_3 W_2(x,w_1)W_2(y,w_2)W_2(z,w_3)\Gamma_3(w_1,w_2,w_3) \ .$$

To get the rest of the W_n requires an induction step.

This business is useful in at least two ways. First it lets us focus our attention on a much smaller collection of diagrams when we are doing our perturbative renormalization.

Secondly, this notion of effective action is extremely useful in thinking about the vacuum structure of field theories, and about spontaneous symmetry breaking. In particular, we can expand the functional in the form

$$\Gamma[\phi_c] = \int d^D x \left(-V_{\text{eff}}(\phi_c) + Z(\phi_c)\left(\partial\phi_c\right)^2 + ...\right)$$

(where the ... indicate terms with more derivatives of ϕ). In particular, in the case where ϕ_c is constant in spacetime we can minimize the *function* $V_{\text{eff}}(\phi_c)$ to find the vacuum. This is a lucrative endeavor which you get to do for homework.

13.3 Coleman-Weinberg potential

[Zee §IV.3, Xi Yin's notes §4.2] Let us now take seriously the lack of indices on our field ϕ, and see about actually evaluating more of the semiclassical expansion of the path integral of a scalar field (eventually we will specify $D = 3+1$):

$$Z[J] = e^{\frac{\mathbf{i}}{\hbar}W[J]} = \int [D\phi] e^{\frac{\mathbf{i}}{\hbar}\left(S[\phi] + \int J\phi\right)} \ . \tag{13.16}$$

To add some drama to this discussion consider the following: if the potential V in $S = \int \left(\frac{1}{2}\left(\partial\phi\right)^2 - V(\phi)\right)$ has a *minimum* at the origin, then we expect that the vacuum has $\langle\phi\rangle = 0$. If on the other hand, the potential has a *maximum* at the origin, then the field will find a minimum somewhere else, $\langle\phi\rangle \neq 0$. If the potential has a discrete symmetry under $\phi \to -\phi$ (no odd powers of ϕ in V), then in the latter case ($V''(0) < 0$) this symmetry will be broken. If the potential is flat ($V''(0) = 0$) near the origin, what happens? Quantum effects matter.

The configuration of stationary phase is $\phi = \phi_\star$, which satisfies

$$0 = \frac{\delta\left(S + \int J\phi\right)}{\delta\phi(x)}|_{\phi=\phi_\star} = -\partial^2\phi_\star(x) - V'(\phi_\star(x)) + J(x) \ . \tag{13.17}$$

Change the integration variable in (13.16) to $\phi = \phi_\star + \varphi$, and expand in powers of the fluctuation φ:

$$\begin{aligned}
Z[J] &= e^{\frac{\mathbf{i}}{\hbar}\left(S[\phi_\star]+\int J\phi_\star\right)}\int [D\varphi] e^{\frac{\mathbf{i}}{\hbar}\int d^Dx \frac{1}{2}\left((\partial\varphi)^2 - V''(\phi_\star)\varphi^2 + \mathcal{O}(\varphi^3)\right)} \\
&\stackrel{\text{IBP}}{=} e^{\frac{\mathbf{i}}{\hbar}\left(S[\phi_\star]+\int J\phi_\star\right)}\int [D\varphi] e^{-\frac{\mathbf{i}}{\hbar}\int d^Dx \frac{1}{2}\left(\varphi\left(\partial^2 + V''(\phi_\star)\right)\varphi + \mathcal{O}(\varphi^3)\right)} \\
&\approx e^{\frac{\mathbf{i}}{\hbar}\left(S[\phi_\star]+\int J\phi_\star\right)} \frac{1}{\sqrt{\det\left(\partial^2 + V''(\phi_\star)\right)}} \\
&= e^{\frac{\mathbf{i}}{\hbar}\left(S[\phi_\star]+\int J\phi_\star\right)} e^{-\frac{1}{2}\operatorname{tr}\log\left(\partial^2 + V''(\phi_\star)\right)}.
\end{aligned}$$

In the second line, we integrated by parts to get the φ integral to look like a souped-up version of the fundamental formula of gaussian integrals – just think of $\partial^2 + V''$ as a big matrix – and in the third line, we did that integral. In the last line we used the matrix identity $\operatorname{tr}\log = \log\det$. Note that all the $\phi_\star$s appearing in this expression are functionals of J, determined by (13.17).

So taking logs of the BHS of the previous equation we have the generating functional:

$$W[J] = S[\phi_\star] + \int J\phi_\star + \frac{\mathbf{i}\hbar}{2}\operatorname{tr}\log\left(\partial^2 + V''(\phi_\star)\right) + \mathcal{O}(\hbar^2) \ .$$

To find the effective potential, we need to Legendre transform to get a functional of ϕ_c:

$$\phi_c(x) = \frac{\delta W}{\delta J(x)} \stackrel{\text{chain rule}}{=} \int d^Dz \frac{\delta\left(S[\phi_\star] + \int J\phi_\star\right)}{\delta\phi_\star(z)} \frac{\delta\phi_\star(z)}{\delta J(x)} + \phi_\star(x) + \mathcal{O}(\hbar) \stackrel{(13.17)}{=} \phi_\star(x) + \mathcal{O}(\hbar) \ .$$

The 1PI effective action is then:

$$\Gamma[\phi_c] \equiv W - \int J\phi_c = S[\phi_c] + \frac{\mathbf{i}\hbar}{2}\operatorname{tr}\log\left(\partial^2 + V''(\phi_c)\right) + \mathcal{O}(\hbar^2).$$

To leading order in $\hbar$, we just plug in the solution; to next order we need to compute the sum of the logs of the eigenvalues of a differential operator. This is challenging in general. In the special case that we are interested in ϕ_c which is constant in spacetime, it is doable. This case is also often physically relevant if our goal is to solve (13.8) to find the groundstate, which often preserves translation invariance (gradients cost energy). If $\phi_c(x) = \phi$ is spacetime-independent then we can write

$$\Gamma[\phi_c(x) = \phi] \equiv \int d^D x \; V_{\text{eff}}(\phi).$$

The computation of the trace-log is doable in this case because it is translation invariant, and hence we can use fourier space. We do this next.

13.3.1 The one-loop effective potential

The tr in the one-loop contribution is a trace over the space on which the differential operator ($\equiv$big matrix) acts; it acts on the space of scalar fields φ:

$$\left(\left(\partial^2 + V''(\phi)\right)\varphi\right)_x = \sum_y \left(\partial^2 + V''(\phi)\right)_{xy} \varphi_y \equiv \left(\partial_x^2 + V''(\phi)\right)\varphi(x)$$

with matrix element $\left(\partial^2 + V''\right)_{xy} = \delta^D(x-y)\left(\partial_x^2 + V''\right)$. (Note that in these expressions, we've assumed ϕ is a background field, not the same as the fluctuation φ – this operator is linear. Further we've assumed that that background field ϕ is a constant, which greatly simplifies the problem.) The trace can be represented as a position integral:

$$\text{tr}\bullet = \int d^D x \, \langle x | \bullet | x \rangle$$

so

$$\begin{aligned}
\text{tr}\log\left(\partial^2 + V''(\phi)\right) &= \int d^D x \, \langle x | \log\left(\partial^2 + V''\right) | x \rangle \\
&= \int d^D x \int đ^D k \int đ^D k' \, \langle x | k' \rangle \langle k' | \log\left(\partial^2 + V''\right) | k \rangle \langle k | x \rangle \qquad \left(\mathbb{1} = \int đ^D k \, | k \rangle \langle k |\right) \\
&= \int d^D x \int đ^D k \int đ^D k' \, \langle x | k' \rangle \langle k' | \log\left(-k^2 + V''\right) | k \rangle \langle k | x \rangle \\
&\qquad \left(\langle k' | \log\left(-k^2 + V''\right) | k \rangle = \delta^D(k - k') \log\left(-k^2 + V''\right)\right) \\
&= \int d^D x \int đ^D k \log\left(-k^2 + V''\right), \qquad \left(\| \langle x | k \rangle \|^2 = 1\right)
\end{aligned}$$

The $\int d^D x$ goes along for the ride and we conclude that

$$V_{\text{eff}}(\phi) = V(\phi) - \frac{\mathbf{i}\hbar}{2} \int đ^D k \log\left(k^2 - V''(\phi)\right) + \mathcal{O}(\hbar^2).$$

What does it mean to take the log of a dimensionful thing? It means we haven't been careful about the additive constant (constant means independent of ϕ). And we don't need to be (unless we're worried about dynamical gravity); so let's choose the constant so that

$$V_{\text{eff}}(\phi) = V(\phi) - \frac{\mathbf{i}\hbar}{2}\int đ^D k \log\left(\frac{k^2 - V''(\phi)}{k^2}\right) + \mathcal{O}(\hbar^2). \tag{13.18}$$

[End of Lecture 55]

$\boxed{V_{\mathbf{1\ loop}} = \sum_{\vec{k}} \frac{1}{2}\hbar\omega_{\vec{k}}}$. Here's the interpretation of the 1-loop potential: $V''(\phi)$ is the mass^2 of the field when it has the constant value ϕ; the one-loop term $V_{1\text{ loop}}$ is the vacuum energy $\int d^{D-1}\vec{k}\frac{1}{2}\hbar\omega_{\vec{k}}$ from the gaussian fluctuations of a field with that mass^2; it depends on the field because the mass depends on the field.

[Zee II.5.3] Why is $V_{1\text{ loop}}$ the vacuum energy? Recall that $k^2 \equiv \omega^2 - \vec{k}^2$ and $đ^D k = đ\omega đ^{D-1}\vec{k}$. Consider the integrand of the spatial momentum integrals: $V_{1\text{ loop}} = -\mathbf{i}\frac{\hbar}{2}\int đ^{D-1}\vec{k}\mathcal{I}$, with

$$\mathcal{I} \equiv \int đ\omega \log\left(\frac{k^2 - V''(\phi) + \mathbf{i}\epsilon}{k^2 + \mathbf{i}\epsilon}\right) = \int đ\omega \log\left(\frac{\omega^2 - \omega_k^2 + \mathbf{i}\epsilon}{\omega^2 - \omega_{k'}^2 + \mathbf{i}\epsilon}\right)$$

with $\omega_k = \sqrt{\vec{k}^2 + V''(\phi)}$, and $\omega_{k'} = |\vec{k}|$. The $\mathbf{i}\epsilon$ prescription is as usual inherited from the euclidean path integral. Notice that the integral is convergent – at large ω, the integrand goes like

$$\log\left(\frac{\omega^2 - A}{\omega^2 - B}\right) = \log\left(\frac{1 - \frac{A}{\omega^2}}{1 - \frac{B}{\omega^2}}\right) = \log\left(1 - \frac{A - B}{\omega^2} + \mathcal{O}\left(\frac{1}{\omega^4}\right)\right) \simeq \frac{A - B}{\omega^2}\ .$$

Integrate by parts:

$$\begin{aligned}\mathcal{I} = \int đ\omega \log\left(\frac{k^2 - V''(\phi) + \mathbf{i}\epsilon}{k^2 + \mathbf{i}\epsilon}\right) &= -\int đ\omega\omega\partial_\omega \log\left(\frac{\omega^2 - \omega_k^2}{\omega - \omega_{k'}}\right) \\ &= -2\int đ\omega\omega\left(\frac{\omega}{\omega^2 - \omega_k^2 + \mathbf{i}\epsilon} - (\omega_k \to \omega_{k'})\right) \\ &= -\mathbf{i}2\omega_k^2\left(\frac{1}{-2\omega_k}\right) - (\omega_k \to \omega_{k'}) = \mathbf{i}\left(\omega_k - \omega_{k'}\right).\end{aligned}$$

This is what we are summing (times $-\mathbf{i}\frac{1}{2}\hbar$) over all the modes $\int đ^{D-1}\vec{k}$.

13.3.2 Renormalization of the effective action

So we have a cute expression for the effective potential (13.18). Unfortunately it seems to be equal to infinity. The problem, as usual, is that we assumed that the parameters in the bare action $S[\phi]$ could be finite without introducing any cutoff. Let us parametrize (following Zee §IV.3) the action as $S = \int d^D x \mathcal{L}$ with

$$\mathcal{L} = \frac{1}{2}\left(\partial\phi\right)^2 - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4!}\lambda\phi^4 - A\left(\partial\phi\right)^2 - B\phi^2 - C\phi^4$$

and we will think of A, B, C as *counterterms*, in which to absorb the cutoff dependence.

So our effective potential is actually:

$$V_{\text{eff}}(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4!}\lambda\phi^4 + B(\Lambda)\phi^2 + C(\Lambda)\phi^4 + \frac{\hbar}{2}\int^{\Lambda} đ^D k_E \log\left(\frac{k_E^2 + V''(\phi)}{k_E^2}\right) ,$$

(notice that A drops out in this special case with constant ϕ). We rotated the integration contour to euclidean space. This permits a nice regulator, which is just to limit the integration region to $\{k_E | k_E^2 \leq \Lambda^2\}$ for some big (Euclidean) wavenumber Λ.

Now let us specify to the case of $D = 4$, where the model with $\mu = 0$ is classically scale invariant. The integrals are elementary[26]

$$V_{\text{eff}}(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4!}\lambda\phi^4 + B(\Lambda)\phi^2 + C(\Lambda)\phi^4 + \frac{\Lambda^2}{32\pi^2}V''(\phi) - \frac{\left(V''(\phi)\right)^2}{64\pi^2}\log\frac{\sqrt{e}\Lambda^2}{V''(\phi)}.$$

Notice that the leading cutoff dependence of the integral is Λ^2, and there is also a subleading logarithmically-cutoff-dependent term. ("log divergence" is certainly easier to say.)

Luckily we have two counterterms. Consider the case where V is a quartic polynomial; then V'' is quadratic, and $(V'')^2$ is quartic. In that case the two counterterms are in just the right form to absorb the Λ dependence. On the other hand, if V were sextic (recall that this is in the non-renormalizable category according to our dimensional analysis), we would have a fourth counterterm $D\phi^6$, but in this case $(V'')^2 \sim \phi^8$, and we're in trouble (adding a bare ϕ^8 term would produce $(V'')^2 \sim \phi^{12}$... and so on). We'll need a better way to think about such non-renormalizable theories. The better way (which we will return to in the next section) is simply to recognize that in non-renormalizable theories, *the cutoff is real* – it is part of the definition of the field theory. In renormalizable theories, we may pretend that it is not (though it usually is real there, too).

[26] This is not the same as 'easy'. The expressions here assume that $\Lambda \gg V''$.

Renormalization conditions. Return to the renormalizable case, $V = \lambda\phi^4$ where we've found

$$V_{\text{eff}} = \phi^2 \left(\frac{1}{2}\mu^2 + B + \lambda\frac{\Lambda^2}{64\pi^2}\right) + \phi^4 \left(\frac{1}{4!}\lambda + C + \frac{\lambda^2}{16\pi^2}\log\frac{\phi^2}{\Lambda^2}\right) + \mathcal{O}(\lambda^3) \ .$$

(I've absorbed an additive $\log\sqrt{e}$ in C.) The counting of counterterms works out, but how do we determine them? We need to impose *renormalization conditions*, *i.e.* specify some observable quantities to parametrize our model, in terms of which we can eliminate the silly letters in the lagrangian. We need two of these. Of course, what is observable depends on the physical system at hand. Let's suppose that we can measure some properties of the effective potential. For example, suppose we can measure the mass^2 when $\phi = 0$:

$$\mu^2 = \frac{\partial^2 V_{\text{eff}}}{\partial\phi^2}|_{\phi=0} \qquad \Longrightarrow \text{ we should set } B = -\lambda\frac{\Lambda^2}{64\pi^2}.$$

For example, we could consider the case $\mu = 0$, when the potential is flat at the origin. With $\mu = 0$, have

$$V_{\text{eff}}(\phi) = \left(\frac{1}{4!}\lambda + \frac{\lambda^2}{(16\pi)^2}\log\frac{\phi^2}{\Lambda^2} + C(\Lambda)\right)\phi^4 + \mathcal{O}(\lambda^3) \ .$$

And for the second renormalization condition, suppose we can measure the quartic term

$$\lambda_M = \frac{\partial^4 V_{\text{eff}}}{\partial\phi^4}|_{\phi=M} \ . \tag{13.19}$$

Here M is some arbitrarily chosen quantity with dimensions of mass. We run into trouble if we try to set it to zero because of $\partial_\phi^4\left(\phi^4\log\phi\right) \sim \log\phi$. So the coupling depends very explicitly on the value of M at which we set the renormalization condition. Let's use (13.19) to eliminate C:

$$\lambda(M) \stackrel{!}{=} 4!\left(\frac{\lambda}{4!} + C + \left(\frac{\lambda}{16\pi}\right)^2\left(\log\frac{\phi^2}{\Lambda^2} + c_1\right)\right)|_{\phi=M} \tag{13.20}$$

(where c_1 is a numerical constant that you should determine) to get

$$V_{\text{eff}}(\phi) = \frac{1}{4!}\lambda(M)\phi^4 + \left(\frac{\lambda(M)}{16\pi}\right)^2\left(\log\frac{\phi^2}{M^2} - c_1\right)\phi^4 + \mathcal{O}(\lambda(M)^3).$$

Here I used the fact that we are only accurate to $\mathcal{O}(\lambda^2)$ to replace $\lambda = \lambda(M) + \mathcal{O}(\lambda(M)^2)$ in various places. We can feel a sense of victory here: the dependence on the cutoff has disappeared. Further, the answer for V_{eff} does not depend on our renormalization point M:

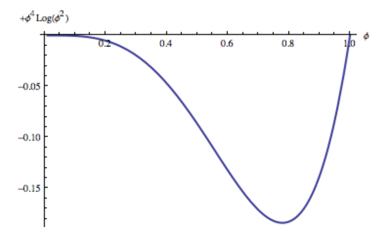
$$M\frac{d}{dM}V_{\text{eff}} = \frac{1}{4!}\phi^4\left(M\partial_M\lambda - \frac{2}{M}\frac{\lambda^2}{(16\pi^2)} + \mathcal{O}(\lambda^3)\right) = \mathcal{O}(\lambda^3) \tag{13.21}$$

which vanishes to this order from the definition of $\lambda(M)$ (13.20), which implies

$$M\partial_M\lambda(M) = \frac{3}{16\pi^2}\lambda(M)^2 + \mathcal{O}(\lambda^3) \equiv \beta(\lambda).$$

The fact (13.21) is sometimes called the *Callan-Symanzik equation*, the condition that $\lambda(M)$ must satisfy in order that physics be independent of our choice of renormalization point M. [End of Lecture 56]

So: when $\mu = 0$ is the $\phi \to -\phi$ symmetry broken by the groundstate? The effective potential looks like the figure at right for $\phi < M$. Certainly it looks like this will push the field away from the origin.



However, the minima lie in a region where our approximations aren't so great. In particular, the next correction looks like:

$$\lambda\phi^4\left(1 + \lambda\log\phi^2 + \left(\lambda\log\phi^2\right)^2 + ...\right)$$

– the expansion parameter is really $\lambda\log\phi$. (I haven't shown this yet, it is an application of the RG, below.) The apparent minimum lies in a regime where the higher powers of $\lambda\log\phi$ are just as important as the one we've kept.

RG-improvement. How do I know the good expansion parameter is actually $\lambda\log\phi/M$? The RG. Define $t \equiv \log\phi_c/M$ and $V_{eff}(\phi_c) = \frac{\phi_c^4}{4!}U(t,\lambda)$. We'll regard U as a running coupling, and t as the RG scaling parameter. Our renormalization conditions are $U(0,\lambda) = \lambda, Z(\lambda) = 1$, these provide initial conditions. At one loop in ϕ^4 theory, there are no anomalous dimensions, $\gamma(\lambda) = \frac{\partial}{\partial M}Z = \mathcal{O}(\lambda^2)$. This makes the RG equations quite simple. The running coupling U satisfies (to this order)

$$\frac{dU}{dt} = \beta(U) = \frac{3U^2}{16\pi^2}$$

which (with the initial condition $U(0,\lambda) = \lambda$) is solved by

$$U(\lambda,t) = \frac{\lambda}{1 - \frac{3\lambda t}{16\pi^2}}.$$

Therefore, the RG-improved effective potential is

$$V_{eff}(\phi_c) = \frac{\phi_c^4}{4!} U(t, \lambda) = \frac{1}{4!} \frac{\lambda \phi_c^4}{1 - \frac{3\lambda}{32\pi^2} \log \frac{\phi_c^2}{M^2}}.$$

The good news: this is valid as long as U is small, and it agrees with our previous answer, which was valid as long as $\lambda \ll 1$ *and* $\lambda t \ll 1$. The bad news is that there is no sign of the minimum we saw in the raw one-loop answer.

By the way, in nearly every other example, there will be wavefunction renormalization. In that case, the Callan-Syzmanzik (CS) equation we need to solve is

$$(-\partial_t + \beta \partial_\lambda + 4\gamma) U(t, \lambda) = 0$$

whose solution is

$$U(t, \lambda) = f(U(t, \lambda)) \exp \left(\int_0^t dt' 4\gamma(U(t', \lambda)) \right), \quad \partial_t U(t, \lambda) = \beta(U), \ U(0, \lambda) = \lambda.$$

f can be determined by studying the CS equation at $t = 0$. For more detail, see E. Weinberg's thesis.

We can get around this issue by studying a system where the fluctuations producing the extra terms in the potential for ϕ come from some *other* field whose mass depends on ϕ. For example, consider a fermion field whose mass depends on ϕ:

$$S[\psi, \phi] = \int d^D x \bar{\psi} \left(\mathbf{i} \partial\!\!\!/ - m - g\phi \right) \psi$$

– then $m_\psi = m + g\phi$. The $\sum \frac{1}{2} \hbar\omega$s from the fermion will now depend on ϕ (the also have the opposite sign because they come from fermions), and we get a reliable answer for $\langle \phi \rangle \neq 0$ from this phenomenon of *radiative symmetry breaking*. In $D = 1 + 1$ this is a field theory description of the Peierls instability of a 1d chain of fermions (ψ) coupled to phonons (ψ). Notice that when ϕ gets an expectation value it gives a mass to the fermions. The microscopic picture is that the translation symmetry is spontaneously broken to a twice-as-big lattice spacing, alternating between strong and weak hopping matrix elements. This produces a gap in the spectrum of the tight-binding model. (For a little more, see Zee page 300.)

A second example where radiative symmetry breaking happens is scalar QED. There we can play the gauge coupling and the scalar self-coupling off each other. I'll say a bit more about this example as it's realized in condensed matter below.

Another example which has attracted a lot of attention is the Standard Model Higgs. Its expectation value affects the masses of many fields, and you might imagine this might produce features in its effective potential. Under various (strong) assumptions about what lies beyond the Standard Model, there is some drama here; I recommend Schwarz's discussion on page 748-750.

13.3.3 Useful properties of the effective action

[For a version of this discussion which is better in just about every way, see Coleman, *Aspects of Symmetry* §5.3.7. I also highly recommend all the preceding sections! And the ones that come after. This book is available electronically from the UCSD library.]

V_{eff} as minimum energy with fixed ϕ. Recall that $\langle\phi\rangle$ is the configuration of ϕ_c which extremizes the effective action $\Gamma[\phi_c]$. Even away from its minimum, the effective potential has a useful physical interpretation. It is the natural extension of the interpretation of the potential in classical field theory, which is: $V(\phi)$ = the value of the energy density if you fix the field equal to ϕ everywhere. Consider the space of states of the QFT where the field has a given expectation value:

$$|\Omega\rangle \quad \text{such that} \qquad \langle\Omega|\,\phi(x)\,|\Omega\rangle = \phi_0(x) \quad ; \tag{13.22}$$

one of them has the smallest energy. I claim that its energy is $V_{\text{eff}}(\phi_0)$. This fact, which we'll show next, has some useful consequences.

Let $|\Omega_{\phi_0}\rangle$ be the (normalized) state of the QFT which minimizes the energy subject to the constraint (13.22). The familiar way to do this (familiar from QM, associated with Rayleigh and Ritz)[27] is to introduce Lagrange multipliers to impose (13.22) and the normalization condition and extremize without constraints the functional

$$\langle\Omega|\,\mathbf{H}\,|\Omega\rangle - \alpha\left(\langle\Omega|\Omega\rangle - 1\right) - \int d^{D-1}\vec{x}\beta(\vec{x})\left(\langle\Omega|\,\phi(\vec{x},t)\,|\Omega\rangle - \phi_0(\vec{x})\right)$$

with respect to $|\Omega\rangle$ and the functions on space α, β. [28]

[27] The more familiar thing is to find the state which extremizes $\langle a|\,\mathbf{H}\,|a\rangle$ subject to the normalization condition $\langle a|a\rangle = 1$. To do this, we vary $\langle a|\,\mathbf{H}\,|a\rangle - E\left(\langle a|a\rangle - 1\right)$ with respect to both $|a\rangle$ and the Lagrange multiplier E. The equation from varying $|a\rangle$ says that the extremum occurs when $(\mathbf{H} - E)\,|a\rangle = 0$, *i.e.* $|a\rangle$ is an energy eigenstate with energy E. Notice that we could just as well have varied the simpler thing

$$\langle a|\,(\mathbf{H} - E)\,|a\rangle$$

and found the same answer.

[28] Here is the QM version (*i.e.* the same thing without all the labels): we want to find the extremum of $\langle a|\mathbf{H}|a\rangle$ with $|a\rangle$ normalized *and* $\langle a|\mathbf{A}|a\rangle = A_c$ some fixed number. Then we introduce two Lagrange multipliers E, J and vary without constraint the quantity

$$\langle a|\,(\mathbf{H} - E - J\mathbf{A})\,|a\rangle$$

(plus irrelevant constants). The solution satisfies

$$(\mathbf{H} - E - J\mathbf{A})\,|a\rangle = 0$$

so $|a\rangle$ is an eigenstate of the perturbed hamiltonian $\mathbf{H} - J\mathbf{A}$, with energy E. J is an auxiliary thing, Clearly the extremum with respect to α, β imposes the desired constraints. Extremizing with respect to $|\Omega\rangle$ gives:

$$\mathbf{H}|\Omega\rangle = \alpha|\Omega\rangle + \int d^{D-1}\vec{x}\beta(\vec{x})\phi(\vec{x},t)|\Omega\rangle \tag{13.23}$$

or

$$\left(\mathbf{H} - \int d^{D-1}\vec{x}\beta(\vec{x})\phi(\vec{x},t)\right)|\Omega\rangle = \alpha|\Omega\rangle \tag{13.24}$$

Note that α, β are functionals of ϕ_0. We can interpret the operator $\mathbf{H}_\beta \equiv \mathbf{H} - \int d^{D-1}\vec{x}\beta(\vec{x})\phi(\vec{x},t)$ on the LHS of (13.24) as the hamiltonian with a source β; and α is the groundstate energy in the presence of that source. (Note that that source is *chosen* so that $\langle\phi\rangle = \phi_0$ – it is a functional of ϕ_0.)

This groundstate energy is related to the generating functional $W[J = \beta]$ as we've seen several times – $e^{\mathbf{i}W[\beta]}$ is the vacuum persistence amplitude in the presence of the source

$$e^{\mathbf{i}W[\beta]} = \left\langle 0|\mathcal{T}e^{\mathbf{i}\int\beta\phi}|0\right\rangle = \langle 0_\beta|e^{-\mathbf{i}T\mathbf{H}_\beta}|0_\beta\rangle = e^{-\mathbf{i}\alpha T} \tag{13.25}$$

where T is the time duration. (If you want, you could imagine that we are adiabatically turning on the interactions for a time duration T.)

The actual extremal energy (of the unperturbed hamiltonian, with constrained expectation value of ϕ) is obtained by taking the overlap of (13.23) with $\langle\Omega|$ (really all the Ωs below are Ω_{ϕ_0}s):

$$\begin{aligned}
\langle\Omega|\mathbf{H}|\Omega\rangle &= \alpha\langle\Omega|\Omega\rangle + \int d^{D-1}\vec{x}\beta(\vec{x})\langle\Omega|\phi(\vec{x},t)|\Omega\rangle \\
&= \alpha + \int d^{D-1}\vec{x}\beta(\vec{x})\phi_0(\vec{x}) \\
&\overset{(13.25)}{=} \frac{1}{T}\left(-W[\beta] + \int d^D x\beta(\vec{x})\phi_0(\vec{x})\right) \\
&\overset{\text{Legendre}}{=} -\frac{1}{T}\Gamma[\phi_0] \overset{\phi=\phi_0,\ \text{const}}{=} \int d^{D-1}\vec{x}V_{\text{eff}}(\phi_0).
\end{aligned}$$

Cluster decomposition. The relationship (13.25) between the generating functional $W[J]$ (for time-independent J) and the energy in the presence of the source is

---

which really depends on our choice A_c, via

$$A_c = \langle a|\mathbf{A}|a\rangle = -\frac{dE}{dJ}.$$

(If you like, we used the Feynman-Hellmann theorem, $\frac{dE}{dJ} = \langle\frac{d\mathbf{H}}{dJ}\rangle$.) The quantity we extremized is

$$\langle a|\mathbf{H}|a\rangle = E + JA_c = E - J\frac{dE}{dJ}.$$

This Legendre transform is exactly (the QM analog of) the effective potential.

very useful. (You've previously used it on the homework to compute the potential between static sources, and to calculate the probability for pair creation in an electric field.) Notice that it gives an independent proof that W only gets contributions from connected amplitudes. Amplitudes with n connected components, $\underbrace{\langle ....\rangle \langle ...\rangle \langle ...\rangle}_{n \text{ of these}}$, go like T^n (where T is the time duration) at large T. Since $W = -E_J T$ goes like T^1, we conclude that it has one connected component (terms that went like $T^{n>1}$ would dominate at large T and therefore must be absent). This *extensivity* of W in T is of the same nature as the extensivity in volume of the free energy in thermodynamics.

[Brown, 6.4.2] Another important reason why W must be connected is called the *cluster decomposition property*. Consider a source which has the form $J(x) = J_1(x) + J_2(x)$ where the two parts have support in widely-separated (spacelike separated) spacetime regions. If all the fields are massive, 'widely-separated' means precisely that the distance between the regions is $R \gg 1/m$, much larger than the range of the interactions mediated by ϕ. In this case, measurements made in region 1 cannot have any effect on those in region 2, and they should be uncorrelated. If so, the probability amplitude factorizes

$$Z[J_1 + J_2] = Z[J_1]Z[J_2]$$

which by the magic of logs is the same as

$$W[J_1 + J_2] = W[J_1] + W[J_2].$$

If W were not connected, it would not have this additive property.

There are actually some exceptions to cluster decomposition arising from situations where we prepare an initial state (it could be the groundstate for some hamiltonian) in which there are correlations between the excitations in the widely separated regions. Such a thing happens in situations with spontaneous symmetry breaking, where the value of the field is the same everywhere in space, and therefore correlates distant regions.

Convexity of the effective potential. Another important property of the effective potential is $V''_{\text{eff}}(\phi) > 0$ – the effective potential is *convex* (sometimes called 'concave up'). We can see this directly from our previous work. Most simply, recall that the functional Taylor coefficients of $\Gamma[\phi]$ are the 1PI Green's functions; V_{eff} is just Γ evaluated for constant ϕ, *i.e.* zero momentum; therefore the Taylor coefficients of V_{eff} are the 1PI Green's functions at zero momentum. In particular, $V''_{\text{eff}}(\phi) = \langle \phi_{k=0}\phi_{k=0}\rangle$: the ground state expectation value of the square of a hermitian operator, which is positive.[29] [30]

On the other hand, it seems that if $V(\phi)$ has a maximum, or even any region of field space where $V''(\phi) < 0$, we get a *complex* one-loop effective potential (from the log of a negative V''). What gives? One resolution is that in this case the minimum energy state with fixed $\langle\phi\rangle$ is not a ϕ eigenstate.

For example, consider a quartic potential $\frac{1}{2}m^2\phi^2 + \frac{g}{4!}\phi^4$ with $m^2 < 0$, with minima at $\phi_\pm \equiv \pm\sqrt{\frac{6|m|^2}{g}}$. Then for $\langle\phi\rangle \in (\phi_-, \phi_+)$, rather we can lower the energy below $V(\phi)$ by considering a state

$$|\Omega\rangle = c_+ |\Omega_+\rangle + c_- |\Omega_-\rangle\,, \quad \langle\Omega|\phi|\Omega\rangle = |c_+|^2\phi_+ + |c_-|^2\phi_-.$$

The one-loop effective potential at ϕ only knows about some infinitesimal neighborhood of the field space near ϕ, and fails to see this non-perturbative stuff. In fact, the correct effective potential is exactly flat in between the two minima. More generally, if the two minima have unequal energies, we have

$$V_{\text{eff}} = \langle\Omega|\,\mathbf{H}\,|\Omega\rangle = |c_+|^2 V(\phi_+) + |c_-|^2 V(\phi_-)$$

– the potential interpolates *linearly* between the energies of the two surrounding minima.

The imaginary part of $V_{\mathbf{1\ loop}}$ is a decay rate. If we find that the (perturbative approximation to) effective potential $E \equiv V_{1\text{ loop}}$ is complex, it means that the amplitude for our state to persist is not just a phase:

$$\mathcal{A} \equiv \langle 0|\, e^{-\mathbf{i}T\mathbf{H}}\, |0\rangle = e^{-\mathbf{i}E\mathcal{V}T}$$

[29] More explicitly: Begin from $V_{\text{eff}} = -\frac{\Gamma}{\mathcal{V}}$.

$$\frac{\partial}{\partial\phi_0} V_{\text{eff}}(\phi_0) = -\int \frac{d^Dx}{\mathcal{V}} \frac{\delta}{\delta\phi(x)} \frac{\Gamma[\phi]}{\mathcal{V}}|_{\phi(x)=\phi_0} \overset{(13.7)}{=} -\frac{1}{\mathcal{V}}\int \frac{d^Dx}{\mathcal{V}}\,(-J(x))\,|_{\phi(x)=\phi_0}.$$

In the first expression here, we are averaging over space the functional derivative of Γ. The second derivative is then

$$\left(\frac{\partial}{\partial\phi_0}\right)^2 V_{\text{eff}}(\phi_0) = \frac{1}{\mathcal{V}}\int \frac{d^Dy}{\mathcal{V}} \frac{\delta}{\delta\phi(y)} \int \frac{d^Dx}{\mathcal{V}}\,(J(x))\,|_{\phi(x)=\phi_0} = +\frac{1}{\mathcal{V}^3}\int_y\int_x \frac{\delta J(x)}{\delta\phi(y)}|_{\phi(x)=\phi_0}$$

Using (13.15), this is

$$V''_{\text{eff}} = +\frac{1}{\mathcal{V}^3}\int_y\int_x \left(W_2^{-1}\right)_{xy}$$

– the inverse is in a matrix sense, with x, y as matrix indices. But W_2 is a positive operator – it is the groundstate expectation value of the square of a hermitian operator.

[30] In fact, the whole effective action $\Gamma[\phi]$ is a convex functional: $\frac{\delta^2\Gamma}{\delta\phi(x)\delta\phi(y)}$ is a positive integral operator. For more on this, I recommend Brown, *Quantum Field Theory*, Chapter 6.

has a modulus different from one ($\mathcal{V}$ is the volume of space). Notice that the $|0\rangle$ here is our perturbative approximation to the groundstate of the system, which is *wrong* in the region of field space where $V'' < 0$. The modulus of this object is

$$P_{\text{no decay}} = \| \mathcal{A} \|^2 = e^{-\mathcal{V}T 2\text{Im}\, E}$$

– we can interpret $2\text{Im}\, E$ as the (connected!) decay probability of the state in question per unit time per unit volume. (Notice that this relation means that the imaginary part of $V_{\text{1-loop}}$ had *better* be positive, so that the probability stays less than one! In the one-loop approximation, this is guaranteed by the correct $\mathbf{i}\epsilon$ prescription.)

For more on what happens when the perturbative answer becomes complex and non-convex, and how to interpret the imaginary part, see this paper by E. Weinberg and Wu.

---

14 Duality

14.1 XY transition from superfluid to Mott insulator, and T-duality

In this subsection and the next we're going to think about ways to think about bosonic field theories with a $\mathsf{U}(1)$ symmetry, and dualities between them, in $D = 1+1$ and $D = 2+1$.

[This discussion is from Ashvin Vishwanath's lecture notes.] Consider the Bose-Hubbard model (in any dimension, but we'll specify to $D = 1+1$ at some point)

$$\mathbf{H}_{BH} = -\tilde{J}\sum_{\langle ij\rangle}\left(\mathbf{b}_i^\dagger \mathbf{b}_j + h.c.\right) + \frac{U}{2}\sum_i \mathbf{n}_i\left(\mathbf{n}_i - 1\right) - \mu\sum_i \mathbf{n}_i$$

where the $\mathbf{b}^\dagger$s and $\mathbf{b}$ are bosonic creation and annihilation operators at each site: $[\mathbf{b}_i, \mathbf{b}_j^\dagger] = \delta_{ij}$. $\mathbf{n}_i \equiv \mathbf{b}_i^\dagger \mathbf{b}_i$ counts the number of bosons at site i. The last Hubbard-U term is zero if $\mathbf{n}_j^b = 0, 1$, but exacts an energetic penalty $\Delta E = U$ if a single site j is occupied by two bosons.

The Hilbert space which represents the boson algebra has a useful *number-phase* representation in terms of

$$[\mathbf{n}_i, \phi_j] = -\mathbf{i}\delta_{ij}, \quad \phi_i \equiv \phi_i + 2\pi, \quad \mathbf{n}_i \in \mathbb{Z}$$

(where the last statement pertains to the eigenvalues of the operator). The bosons are

$$\mathbf{b}_i = e^{-\mathbf{i}\phi_i}\sqrt{\mathbf{n}_i}, \quad \mathbf{b}_i^\dagger = \sqrt{\mathbf{n}_i}e^{+\mathbf{i}\phi_i} \;;$$

these expressions have the same algebra as the original $\mathbf{b}$s. In terms of these operators, the hamiltonian is

$$\mathbf{H}_{BH} = -\tilde{J}\sum_{\langle ij\rangle}\left(\sqrt{\mathbf{n}_i}e^{\mathbf{i}(\phi_i-\phi_j)}\sqrt{\mathbf{n}_j} + h.c.\right) + \frac{U}{2}\sum_i \mathbf{n}_i\left(\mathbf{n}_i - 1\right) - \mu\sum_i \mathbf{n}_i.$$

If $\langle \mathbf{n}_i\rangle = n_0 \gg 1$, so that $\mathbf{n}_i = n_0 + \Delta\mathbf{n}_i, \Delta\mathbf{n}_i \ll n_0$ then $\mathbf{b}_i = e^{-\mathbf{i}\phi}\sqrt{\mathbf{n}_i} \simeq e^{-\mathbf{i}\phi_i}\sqrt{n_0}$ and

$$\mathbf{H}_{BH} \simeq -\underbrace{2\tilde{J}n_0}_{\equiv J}\sum_{\langle ij\rangle}\cos\left(\phi_i - \phi_j\right) + \frac{U}{2}\sum_i\left(\Delta\mathbf{n}_i\right)^2 \equiv \mathbf{H}_{\text{rotors}}$$

where we set $n_0 \equiv \mu/U \gg 1$. This is a *rotor model*.

This model has two phases:
$\boxed{U \gg J}$: then we must satisfy the U term first and the number is locked, $\Delta\mathbf{n} = 0$ in

the groundstate. This is a Mott insulator, with a gap of order U. Since $\mathbf{n}$ and ϕ are conjugate variables, definite number means wildly fluctuating phase.

$\boxed{U \ll J}$: then we must satisfy the J term first and the phase is locked, $\phi = 0$ in the groundstate, or at least it will try. This is the superfluid (SF). That is, we can try to expand the cosine potential[31]

$$\mathbf{H}_{\text{rotors}} = U \sum_i \mathbf{n}_i^2 - J \sum_{\langle ij \rangle} \cos(\phi_i - \phi_j) \simeq U \sum_i \mathbf{n}_i^2 - J \sum_{\langle ij \rangle} \left(1 - \frac{1}{2}(\phi_i - \phi_j)^2 + \ldots\right) \tag{14.1}$$

which is a bunch of harmonic oscillators and can be solved by Fourier: $\phi_i = \frac{1}{\sqrt{N^d}} \sum_i e^{-\mathbf{i}k \cdot x_i} \phi_k$, so

$$\mathbf{H} \simeq \sum_k \left(U \pi_k \pi_{-k} + J(1 - \cos ka) \phi_k \phi_{-k}\right)$$

This has gapless *phonon* modes at $k = 0$, whose existence is predicted by Nambu-Goldstone. I have written the hamiltonian in 1d notation but nothing has required it so far. The low energy physics is described by the continuum lagrangian density

$$L_{\text{eff}} = \frac{\rho_s}{2} \left(\frac{(\partial_\tau \phi)^2}{c} + c\left(\vec{\nabla}\phi\right)^2 \right) \tag{14.2}$$

with $\rho_s = \sqrt{J/U}, c = \sqrt{JU}$. ρ_s is called the superfluid stiffness. This is a free massless scalar theory. The demand of the $\mathsf{U}(1)$ symmetry $\phi \to \phi + \alpha$ forbids interactions which would be relevant; the only allowed interactions are derivative interactions (as you can see by keeping more terms in the Taylor expansion (14.1)) such as $(\partial\phi)^4$.

Now 1d comes in: In $d > 1$, there is long range order – the bosons condense and spontaneously break the phase rotation symmetry $\phi \to \phi + \alpha$; the variable ϕ is a Goldstone boson. In 1d there is no long-range order. The two phases are still distinct however, since one has a gap and the other does not. The correlators of the boson operator $b_i \sim e^{\mathbf{i}\phi_i}$ diagnose the difference. In the Mott phase they have exponential decay. In the "SF" they have

$$\left\langle e^{\mathbf{i}\phi(x)} e^{-\mathbf{i}\phi(y)} \right\rangle = \frac{c_0}{r^\eta}, \quad \eta = \frac{1}{2\pi\rho_s} = \frac{1}{2K}.$$

This is *algebraic long range order*. This is a sharp distinction between the two phases we've discussed, even though the IR fluctuations destroy the $\langle b \rangle$.

[End of Lecture 57]

[31] From now on the background density n_0 will not play a role and I will write $\mathbf{n}_i$ for $\Delta\mathbf{n}_i$.

Massless scalars in $D = 1 + 1$ and T-duality-invariance of the spectrum. A lot of physics is hidden in the innocent-looking theory of the superfluid goldstone boson. Consider the following (real-time) continuum action for a free massless scalar field in 1+1 dimensions:

$$S[\phi] = \frac{T}{2} \int dt \int_0^L dx \left((\partial_0 \phi)^2 - (\partial_x \phi)^2 \right) = 2T \int dx dt \partial_+ \phi \partial_- \phi \ . \tag{14.3}$$

I have set the velocity of the bosons to $c = 1$ by rescaling t. Here $x^\pm \equiv t \pm x$ are lightcone coordinates; the derivatives are $\partial_\pm \equiv \frac{1}{2}(\partial_t \pm \partial_x)$. Space is a circle: the point labelled x is the same as the point labelled $x + L$.

We will assume that the field space of ϕ itself is periodic:

$$\phi(x,t) \equiv \phi(x,t) + 2\pi, \quad \forall x, t \ .$$

So the *field* space is a circle S^1 with (angular) coordinate ϕ. It can be useful to think of the action (14.3) as describing the propagation of a string, since a field configuration describes an embedding of the real two dimensional space into the *target space*, which here is a circle. This is a simple special case of a nonlinear sigma model. The name T-duality comes from the literature on string theory. The worldsheet theory of a string propagating on a circle of radius $R = \sqrt{\rho_s}$ is governed by the Lagrangian (14.2). To see this, recall that the action of a 2d nonlinear sigma model with target space metric $g_{\mu\nu}\phi^\mu\phi^\nu$ is $\frac{1}{\alpha'}\int d^2\sigma g_{\mu\nu}\partial\phi^\mu\partial\phi^\nu$. Here $\frac{1}{\alpha'}$ is the tension (energy per unit length) of the string; work in units where this disappears from now on. Here we have only one dimension, with $g_{\phi\phi} = \rho_s$.

Notice that we could rescale $\phi \to \lambda\phi$ and change the radius; but this would change the periodicity of $\phi \equiv \phi + 2\pi$. The proper length of the period is $2\pi R$ and is invariant under a change of field variables. This proper length distinguishes different theories because the operators $: e^{\alpha\phi} :$ (which you saw on a previous homework were good operators of definite scaling dimension in the theory of the free boson (unlike ϕ itself)) must be periodic; this determines the allowed values of α.

First a little bit of classical field theory. The equations of motion for ϕ are

$$0 = \frac{\delta S}{\delta \phi(x,t)} \propto \partial^\mu \partial_\mu \phi \propto \partial_+ \partial_- \phi$$

which is solved by

$$\phi(x,t) \equiv \phi_L(x^+) + \phi_R(x^-) \ .$$

In euclidean time, $\phi_{L,R}$ depend (anti-)holomorphically on the complex coordinate $z \equiv \frac{1}{2}(x + \mathbf{i}\tau)$ and the machinery of complex analysis becomes useful.

Symmetries: Since $S[\phi]$ only depends on ϕ through its derivatives, there is a simple symmetry $\phi \to \phi + \epsilon$. By the Nöther method the associated current is

$$j_\mu = T\partial_\mu \phi \ . \tag{14.4}$$

This symmetry is translations in the target space, and I will sometimes call the associated conserved charge 'momentum'.

There is another symmetry which is less obvious. It comes about because of the topology of the target space. Since $\phi(x,t) \equiv \phi(x,t) + 2\pi m, m \in \mathbb{Z}$ describe the same point (it is a redundancy in our description, in fact a discrete gauge redundancy), we don't need $\phi(x+L,t) = \phi(x,t)$. It is enough to have

$$\phi(x+L,t) = \phi(x,t) + 2\pi m, \quad m \in \mathbb{Z}$$

The number m cannot change without the string breaking: it is a topological charge, a winding number:

$$m = \frac{1}{2\pi}\phi(x,t)|_{x=0}^{x=L} \overset{\text{FTC}}{=} \frac{1}{2\pi}\int_0^L dx \partial_x \phi \ . \tag{14.5}$$

The associated current whose charge density is $\frac{1}{\pi}\partial_x\phi$ (which integrates over space to the topological charge) is

$$\tilde{j}_\mu = \frac{1}{2\pi}(\partial_x\phi, -\partial_0\phi)_\mu = \frac{1}{2\pi}\epsilon^{\mu\nu}\partial_\nu\phi.$$

This is conserved because of the equality of the mixed partials: $\epsilon^{\mu\nu}\partial_\mu\partial_\nu = 0$.

Let's expand in normal modes: $\phi = \phi_L + \phi_R$ with

$$\begin{aligned}
\phi_L(t+x) &= q_L + \underbrace{(p+w)}_{\equiv \frac{1}{2T}p_L}(t+x) - \mathbf{i}\sqrt{\frac{L}{4\pi T}}\sum_{n\neq 0}\frac{\rho_n}{n}e^{\mathbf{i}n(t+x)\frac{2\pi}{L}}, \\
\phi_R(t-x) &= q_R + \underbrace{(p-w)}_{\equiv \frac{1}{2T}p_R}(t-x) - \mathbf{i}\sqrt{\frac{L}{4\pi T}}\sum_{n\neq 0}\frac{\tilde{\rho}_n}{n}e^{\mathbf{i}n(t-x)\frac{2\pi}{L}},
\end{aligned} \tag{14.6}$$

The factor of $\frac{1}{n}$ is a convention whose origin you will appreciate below, as are the other normalization factors. Real ϕ means $\rho_n^\dagger = \rho_{-n}$ (If we didn't put the $\mathbf{i}$ it would have been $-\rho_{-n}$).

Here $q \equiv \frac{1}{L}\int_0^L dx\phi(x,t) = q_L - q_R$ is the center-of-mass position of the string. The canonical momentum for ϕ is $\pi(x,t) = T\partial_0\phi(x,t) = T\left(\partial_+\phi_L + \partial_-\phi_R\right)$.

QM. Now we'll do quantum mechanics. Recall that a quantum mechanical particle on a circle has momentum quantized in units of integers over the period. Since ϕ is periodic, the wavefunction(al)s must be periodic in the center-of-mass coordinate q with period 2π, and this means that the total (target-space) momentum must be an integer

$$\mathbb{Z} \ni j = \pi_0 \equiv \int_0^L dx \pi(x,t) = T \int_0^L dx \partial_t \phi \overset{(14.6)}{=} LT2p$$

So our conserved charges are quantized according to

$$p = \frac{j}{2LT}, \qquad w \overset{(14.6)(14.5)}{=} \frac{\pi m}{L}, \qquad j, m \in \mathbb{Z} .$$

(Don't confuse the target-space momentum j with the 'worldsheet momentum' n!)

(Note that this theory is scale-free. We could use this freedom to choose units where $L = 2\pi$.)

Now I put the mode coefficients in boldface:

$$\begin{aligned}
\boldsymbol{\phi}_L(x^+) &= \mathbf{q}_L + \frac{1}{2T}\mathbf{p}_L x^+ - \mathbf{i}\sqrt{\frac{L}{4\pi T}} \sum_{n \neq 0} \frac{\boldsymbol{\rho}_n}{n} e^{\mathbf{i}\frac{2\pi}{L} n x^+}, \\
\boldsymbol{\phi}_R(x^-) &= \mathbf{q}_R + \frac{1}{2T}\mathbf{p}_R x^- - \mathbf{i}\sqrt{\frac{L}{4\pi T}} \sum_{n \neq 0} \frac{\tilde{\boldsymbol{\rho}}_n}{n} e^{\mathbf{i}\frac{2\pi}{L} n x^-},
\end{aligned} \tag{14.7}$$

The nonzero canonical equal-time commutators are

$$[\boldsymbol{\phi}(x), \boldsymbol{\pi}(x')] = -\mathbf{i}\delta(x - x')$$

which determines the commutators of the modes (this was the motivation for the weird normalizations)

$$[\mathbf{q}_L, \mathbf{p}_L] = [\mathbf{q}_R, \mathbf{p}_R] = \mathbf{i}, \qquad [\boldsymbol{\rho}_n, \boldsymbol{\rho}_{n'}^\dagger] = n\delta_{n,n'}, \text{or } \; [\boldsymbol{\rho}_n, \boldsymbol{\rho}_{n'}] = n\delta_{n+n'} \;,$$

and the same for the rightmovers with twiddles. This is one simple harmonic oscillator for each $n \geq 1$ (and each chirality); the funny normalization is conventional.

$$\begin{aligned}
\mathbf{H} &= \int dx \left(\boldsymbol{\pi}(x)\dot{\boldsymbol{\phi}}(x) - \mathcal{L} \right) = \frac{1}{2} \int dx \left(\frac{\boldsymbol{\pi}^2}{T} + T\left(\partial_x \boldsymbol{\phi}\right)^2 \right) \\
&= \underbrace{L\frac{1}{4T}\left(\mathbf{p}_L^2 + \mathbf{p}_R^2\right)}_{\frac{\boldsymbol{\pi}_0^2}{2T} + \frac{T}{2}\mathbf{w}^2} + \pi \sum_{n=1}^\infty \left(\boldsymbol{\rho}_{-n}\boldsymbol{\rho}_n + \tilde{\boldsymbol{\rho}}_{-n}\tilde{\boldsymbol{\rho}}_n \right) + \mathfrak{a} \\
&= \frac{1}{2L}\left(\frac{j^2}{T} + T\left(2\pi m\right)^2 \right) + \pi \sum_{n=1}^\infty n\left(\mathbf{N}_n + \tilde{\mathbf{N}}_n \right) + \mathfrak{a}
\end{aligned} \tag{14.8}$$

Here $\mathfrak{a}$ is a (UV sensitive) constant which will not be important for us (it is very important in string theory), which is the price we pay for writing the hamiltonian as a sum of normal-ordered terms – the modes with negative indices are to the right and they annihilate the vacuum:

$$\boldsymbol{\rho}_n |0\rangle = 0, \quad \tilde{\boldsymbol{\rho}}_n |0\rangle = 0, \quad \text{for } n > 0 \ .$$

Energy eigenstates can be labelled by a target-momentum j and a winding m. Notice that there is an operator $\mathbf{w}$ whose eigenvalues are w, and it has a conjugate momentum $\mathbf{p}_L - \mathbf{p}_R$ which increments its value. So when I write $|0\rangle$ above, I really should label a vacuum of the oscillator modes with p, w.

$\mathbf{N}_n \equiv \frac{1}{n}\boldsymbol{\rho}_{-n}\boldsymbol{\rho}_n$ is the number operator; if we redefine $\mathbf{a}_n \equiv \sqrt{n}^{-1}\boldsymbol{\rho}_n (n > 0)$, we have $[\mathbf{a}_n, \mathbf{a}_m^\dagger] = \delta_{nm}$ and $\mathbf{N}_n = \mathbf{a}_n^\dagger \mathbf{a}_n$ is the ordinary thing.

Notice that (14.4) means that there are separately-conserved left-moving and right-moving currents:

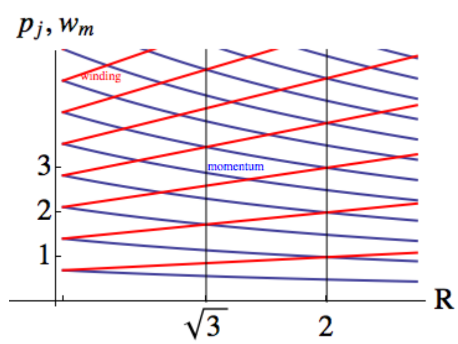
$$(j_L)^\mu = (j_L^z, j_L^{\bar{z}})^\mu \equiv (j_+, 0)^\mu$$
$$(j_R)^\mu = (j_R^z, j_R^{\bar{z}})^\mu \equiv (0, j_-)^\mu$$

Here j_L only depends on the modes $\boldsymbol{\rho}_n$, and j_R only depends on the modes $\tilde{\boldsymbol{\rho}}_n$:

$$j_+ = \partial_+\phi = \partial_+\phi(x^+) = \mathbf{p} + \mathbf{w} + \sqrt{\frac{\pi}{LT}} \sum_{n\neq 0} \boldsymbol{\rho}_n e^{\mathbf{i}\frac{2\pi}{L}nx^+}$$

$$j_- = \partial_-\phi = \partial_-\phi(x^-) = \mathbf{p} - \mathbf{w} + \sqrt{\frac{\pi}{LT}} \sum_{n\neq 0} \tilde{\boldsymbol{\rho}}_n e^{\mathbf{i}\frac{2\pi}{L}nx^-}$$

Here's an **Observation (T-duality)**: At large T (think of this as a large radius of the target space), the momentum modes are closely-spaced in energy, and exciting the winding modes is costly, since the string has a tension, it costs energy-per-unit-length T to stretch it. But the spectrum (14.8) is invariant under the operation



$$m \leftrightarrow j, \quad T \leftrightarrow \frac{1}{(2\pi)^2 T}$$

which takes the radius of the circle to its inverse and exchanges the momentum and winding modes. This is called *T-duality*. The required duality map on the fields is

$$\phi_L + \phi_R \leftrightarrow \phi_L - \phi_R \ .$$

(The variable R in the plot is $R \equiv \sqrt{\pi T}$.)

T-duality says string theory on a large circle is the same as string theory on a small circle. On the homework you'll get to see a derivation of this statement in the continuum which allows some generalizations.

Vertex operators. It is worthwhile to pause for another moment and think about the operators which create the winding modes. They are like vortex creation operators. Since ϕ has logarithmic correlators, you might think that exponentiating it is a good idea. First let's take advantage of the fact that the ϕ correlations split into left and right bits to write $\phi(z,\bar{z}) = \phi_L(z) + \phi_R(\bar{z})$:

$$\langle \phi_L(z)\phi_L(0)\rangle = -\frac{1}{\pi T}\log\frac{z}{a}, \quad \langle \phi_R(\bar{z})\phi_R(0)\rangle = -\frac{1}{\pi T}\log\frac{\bar{z}}{a}\ , \quad \langle \phi_L(z)\phi_R(0)\rangle = 0\ . \tag{14.9}$$

A set of operators with definite scaling dimension is:

$$\mathcal{V}_{\alpha,\beta}(z,\bar{z}) =: e^{\mathbf{i}(\alpha\phi_L(z)+\beta\phi_R(\bar{z}))} :\ .$$

This is a composite operator which we have defined by normal-ordering. The normal ordering prescription is: $\boxed{q,p,-,+}$, that is: positive-momentum modes (lowering operators) go on the right, and p counts as a lowering operator, so in particular using the expansion (please beware my factors here): $\phi_L(z) = \mathbf{q}_L + \mathbf{p}_L z + \mathbf{i}\sum_{n\neq 0}\frac{\boldsymbol{\rho}_n}{n}w^n$, we have

$$: e^{\mathbf{i}\alpha\phi_L(z)} :\equiv e^{\mathbf{i}\alpha\mathbf{q}_L} e^{\mathbf{i}\alpha\mathbf{p}_L z} e^{\mathbf{i}\alpha\sum_{n<0}\frac{\boldsymbol{\rho}_n}{n}w^n} e^{\mathbf{i}\alpha\sum_{n>0}\frac{\boldsymbol{\rho}_n}{n}w^n}$$

(I used the definition $w \equiv e^{2\pi i z/L}$.)

How should we think about this operator ? In the QM of a free particle, the operator $e^{\mathbf{i}p\mathbf{x}}$ inserts momentum p – it takes a momentum-space wavefunction $\psi(p_0) = \langle p_0|\psi\rangle$ and gives

$$\langle p_0|\, e^{\mathbf{i}p\mathbf{x}}\, |\psi\rangle = \psi(p_0 + p).$$

It's the same thing here, with one more twist.

In order for $\mathcal{V}_{\alpha,\beta}$ to be well-defined under $\phi \to \phi + 2\pi$, we'd better have $\alpha + \beta \in \mathbb{Z}$ – momentum is quantized, just like for the particle (the center of mass is just a particle). Let's consider what the operator $\mathcal{V}_{\alpha,\beta}$ does to a winding and momentum eigenstate $|w,p\rangle$ (with no oscillator excitations, $\boldsymbol{\rho}_n|p,w\rangle = 0, n < 0$):

$$\mathcal{V}_{\alpha\beta}(0)\,|w,p\rangle = e^{\mathbf{i}(\alpha+\beta)\mathbf{q}_0} e^{\mathbf{i}(\alpha-\beta)\tilde{\boldsymbol{\phi}}_0} e^{\mathbf{i}\alpha\sum_{n<0}\boldsymbol{\rho}_n} e^{\mathbf{i}\alpha\sum_{n>0}\boldsymbol{\rho}_n}\,|w,p\rangle = e^{\mathbf{i}\alpha\sum_{n<0}\boldsymbol{\rho}_n}\,|w+\alpha-\beta, p+\alpha+\beta\rangle \tag{14.10}$$

The monster in front here creates oscillator excitations. I wrote $\mathbf{q}_0 \equiv \mathbf{q}_L + \mathbf{q}_R$ and $\tilde{\boldsymbol{\phi}}_0 \equiv \mathbf{q}_L - \mathbf{q}_R$. The important thing is that the winding number has been incremented by $\alpha - \beta$; this means that $\alpha - \beta$ must be an integer, too. We conclude that

$$\alpha + \beta \in \mathbb{Z}, \quad \alpha - \beta \in \mathbb{Z} \tag{14.11}$$

so they can both be half-integer, or they can both be integers.

By doing the gaussian integral (or moving the annihilation operators to the right) their correlators are

$$\langle \mathcal{V}_{\alpha,\beta}(z,\bar{z})\mathcal{V}_{\alpha',\beta'}(0,0)\rangle = \frac{D_0}{z^{\frac{\alpha^2}{\pi T}}\bar{z}^{\frac{\beta^2}{\pi T}}} \ . \tag{14.12}$$

The zeromode prefactor D_0 is:

$$D_0 = \left\langle e^{\mathbf{i}((\alpha+\alpha')\mathbf{q}_L+(\beta+\beta')\mathbf{q}_R)}\right\rangle_0 = \delta_{\alpha+\alpha'}\delta_{\beta+\beta'}.$$

This is charge conservation.

We conclude that the operator $\mathcal{V}_{\alpha,\beta}$ has scaling dimension

$$(h_L, h_R) = \frac{1}{2\pi T}(\alpha^2, \beta^2).$$

Notice the remarkable fact that the exponential of a dimension-zero operator manages to have nonzero scaling dimension. This requires that the multiplicative prefactor depend on the cutoff a to the appropriate power (and it is therefore nonuniversal). We could perform a multiplicative renormalization of our operators $\mathcal{V}$ to remove this cutoff dependence from the correlators.

The values of α, β allowed by single-valuedness of ϕ and its wavefunctional are integers. We see (at least) three special values of the parameter T:

- **The $\mathsf{SU}(2)$ radius**: When $2\pi T = 1$, the operators with $(n, m) = 1$ are *marginal*. Also, the operators with $(n, m) = (1, 0)$ and $(n, m) = (0, 1)$ have the scaling behavior of currents, and by holomorphicity are in fact conserved.

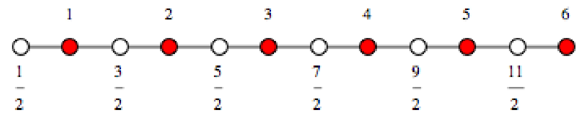
- **The free fermion radius**: when $2\pi T = 2$, $\mathcal{V}_{1,0}$ has dimension $(\frac{1}{2}, 0)$, which is the dimension of a left-moving free fermion, with action $\int dtdx\bar{\psi}\partial_+\psi$. In fact the scalar theory with this radius is equivalent to a massless Dirac fermion! This equivalence is an example of *bosonization*. In particular, the radius-changing deformation of the boson maps to a marginal four-fermion interaction: by studying free bosons we can learn about interacting fermions. Should I say more about this?

- **The supersymmetric radius**: when $2\pi T = \frac{2}{3}$, $\mathcal{V}_{1,0}$ has dimension $(\frac{3}{2}, 0)$ and represents a supersymmetry current.

After this detour, let's turn to the drama of the bose-Hubbard model. Starting from large J/U, where we found a superfluid, what happens as U grows and makes the phase fluctuate more? Our continuum description in terms of harmonic oscillators hides (but does not ignore) the fact that $\phi \simeq \phi + 2\pi$. The system admits vortices, aka winding modes.

Lattice T-duality. To see their effects let us do T-duality on the lattice.

The dual variables live on the bonds, labelled by $\bar{i} = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}....$



Introduce

$$\mathbf{m}_{\bar{i}} \equiv \frac{\phi_{i+1} - \phi_i}{2\pi}, \quad \Theta_{\bar{i}} \equiv \sum_{j<\bar{i}} 2\pi \mathbf{n}_j \tag{14.13}$$

which together imply

$$[\mathbf{m}_{\bar{i}}, \Theta_{\bar{j}}] = -\mathbf{i}\delta_{\bar{i}\bar{j}}.$$

To understand where these expressions come from, notice that the operator

$$e^{\mathbf{i}\Theta_{\bar{i}}} = e^{\mathbf{i}\sum_{j<\bar{i}} 2\pi \mathbf{n}_j}$$

rotates the phase of the boson on all sites to the left of $\bar{i}$ (by 2π). It inserts a vortex in between the sites i and $i+1$. The rotor hamiltonian is

$$\begin{aligned} \mathbf{H}_{\text{rotors}} &= \frac{U}{2}\sum_{\bar{i}} \left(\frac{\Theta_{\bar{i}+1} - \Theta_{\bar{i}}}{2\pi}\right)^2 - J\sum_{\bar{i}} \cos 2\pi \mathbf{m}_{\bar{i}} \\ &\overset{SF}{\simeq} \sum_{\bar{i}} \left(\frac{U}{2}\left(\frac{\Delta\Theta}{2\pi}\right)^2 + \frac{J}{2}\left(2\pi \mathbf{m}_{\bar{i}}\right)^2\right) \end{aligned} \tag{14.14}$$

where in the second step, we assumed we were in the SF phase, so the phase fluctuations and hence $\mathbf{m}_{\bar{i}}$ are small. This looks like a chain of masses connected by springs again, but with the roles of kinetic and potential energies reversed – the second term should be regarded as a π^2 kinetic energy term. BUT: we must not forget that $\Theta \in 2\pi\mathbb{Z}$! It's oscillators with discretized positions. We can rewrite it in terms of continuous Θ at the expense of imposing the condition $\Theta \in 2\pi\mathbb{Z}$ energetically by adding a term $-\lambda \cos\Theta$[32]. The resulting model has the action

$$L_{\text{eff}} = \frac{1}{2(2\pi)^2 \rho_s}\left(\partial_\mu \Theta\right)^2 - \lambda \cos\Theta. \tag{14.15}$$

[32]This step seems scary at first sight, since we're adding degrees of freedom to our system, albeit gapped ones. $\Theta_{\bar{i}}$ is the number of bosons to the left of $\bar{i}$ (times 2π). An analogy that I find useful is to the fact that the number of atoms of air in the room is an integer. This constraint can have some important consequences, for example, were they to solidify. But in our coarse-grained description of the fluid phase, we use variables (the continuum number density) where the number of atoms (implicitly) varies continuously. The nice thing about this story (both for vortices and for air) is that the system tells us when we can't ignore this quantization constraint.

Ignoring the λ term, this is the T-dual action, with ρ_s replaced by $\frac{1}{(2\pi)^2\rho_s}$. The coupling got inverted here because in the dual variables it's the J term that's like the π^2 inertia term, and the U term is like the restoring force. This $\Theta = \phi_L - \phi_R$ is therefore T-dual variable, with ETCRs

$$[\phi(x), \Theta(y)] = 2\pi \mathsf{i} \mathsf{sign}(x-y). \tag{14.16}$$

This commutator follows directly from the definition of Θ (14.13). (14.16) means that the operator $\cos\Theta(x)$ jumps the SF phase variable ϕ by 2π – it inserts a 2π vortex, as we designed it to do. So λ is like a chemical potential for vortices.

This system has two regimes, depending on the scaling dimension of the vortex insertion operator:

• If λ is an irrelevant coupling, we can ignore it in the IR and we get a superfluid, with algebraic LRO.

• If the vortices are relevant, $\lambda \to \infty$ in the IR, and we pin the dual phase, $\Theta_{\bar{i}} = 0, \forall \bar{i}$. This is the Mott insulator, since $\Theta_{\bar{i}} = 0$ means $\mathbf{n}_i = 0$ – the number fluctuations are frozen.

When is λ relevant? Expanding around the free theory,

$$\left\langle e^{\mathbf{i}\Theta(x)} e^{-\mathbf{i}\Theta(0)} \right\rangle = \frac{c}{x^{2\pi\rho_s}}$$

this has scaling dimension $\Delta = \pi\rho_s$ which is relevant if $2 > \Delta = \pi\rho_s$. Since the bose correlators behave as $\left\langle b^\dagger b \right\rangle \sim x^{-\eta}$ with $\eta = \frac{1}{2\pi\rho_s}$, we see that only if $\eta < \frac{1}{4}$ do we have a stable SF phase. (Recall that $\rho_s = \sqrt{J/U}$.) If $\eta > \frac{1}{4}$, the SF is unstable to proliferation of vortices and we end up in the Mott insulator, where the quantization of particle number matters. A lesson: we can think of the Mott insulator as a condensate of vortices. [End of Lecture 58]

Note: If we think about this euclidean field theory as a 2+0 dimensional stat-mech problem, the role of the varying ρ_s is played by temperature, and this transition we've found of the XY model, where by varying the radius the vortices become relevant, is the Kosterlitz-Thouless transition. Most continuous phase transitions occur by tuning the coefficient of a relevant operator to zero (recall the general $\mathsf{O}(n)$ transition, where we have to tune $r \to r_c$ to get massless scalars). This is *not* what happens in the 2d XY model; rather, we are varying a *marginal* parameter and the dimensions of other operators depend on it and become relevant at some critical value of that marginal parameter. This leads to very weird scaling near the transition, of the form $e^{-\frac{a}{\sqrt{K-K_c}}}$ (for example, in the correlation length, the exponential arises from inverting expressions involving $G_R(z) = -\frac{1}{4\pi K}\log z$) – it is sometimes called an 'infinite order' phase transition, because all derivatives of such a function are continuous.

14.2 (2+1)-d XY is dual to (2+1)d electrodynamics

14.2.1 Mean field theory

Earlier (during our discussion of boson coherent states) I made some claims about the phase diagram of the Bose-Hubbard model

$$H_{BH} = \sum_i \left(-\mu n_i + U n_i(n_i - 1)\right) + \sum_{ij} b_i^\dagger w_{ij} b_j$$

which I would like to clarify.

[Sachdev] Consider a variational approach to the BH model. We'll find the best product-state wavefunction $|\Psi_{\text{var}}\rangle = \otimes_i |\psi_i\rangle$, and minimize the BH energy $\langle\Psi_{\text{var}}| H_{BH} |\Psi_{\text{var}}\rangle$ over all ψ_i. We can parametrize the single-site states as the groundstates of the mean-field hamiltonian:

$$H_{\text{MF}} = \sum_i h_i = \sum_i \left(-\mu n_i + U n_i(n_i - 1) - \Psi^\star b_i - \Psi b_i^\dagger\right).$$

Here Ψ is an effective field which incorporates the effects of the neighboring sites. Notice that nonzero Ψ breaks the $\mathsf{U}(1)$ boson number conservation: particles can hop out of the site we are considering. This also means that nonzero Ψ will signal SSB.

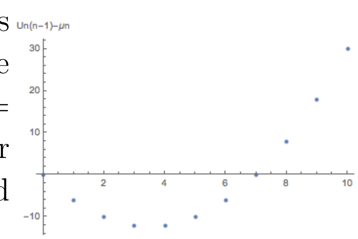
What does this simple approximation give up? For one, it assumes the groundstate preserves the lattice translation symmetry, which doesn't always happen. More painfully, it also gives up on any entanglement at all in the groundstate. Phases for which entanglement plays an important role will not be found this way.

We want to minimize over Ψ the quantity

$$\begin{aligned}\mathcal{E}_0 &\equiv \frac{1}{M} \langle\Psi_{\text{var}}| H_{BH} |\Psi_{\text{var}}\rangle = \frac{1}{M}\left(\langle\Psi_{\text{var}}| \left(\underbrace{H_{BH} - H_{MF}}_{=w\sum b^\dagger b + \Psi b + h.c.} + H_{MF}\right) |\Psi_{\text{var}}\rangle\right) \\ &= \frac{1}{M} E_{MF}(\Psi) - zw \left\langle b^\dagger\right\rangle \langle b\rangle + \langle b\rangle \Psi^\star + \left\langle b^\dagger\right\rangle \Psi. \end{aligned} \tag{14.17}$$

Here z is the coordination number of the lattice (the number of neighbors of a site, which we assume is the same for every site), and $\langle ..\rangle \equiv \langle\Psi_{\text{var}}| .. |\Psi_{\text{var}}\rangle$.

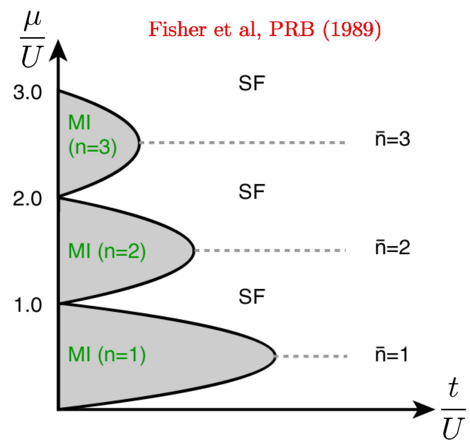
First consider $w = 0$, no hopping. Then $\Psi_B = 0$ (neighbors don't matter), and the single-site state is a number eigenstate $|\psi_i\rangle = |n_0(\mu/U)\rangle$, where $n_0(x) = 0$ for $x < 0$, and $n_0(x) = \lceil x \rceil$, (the ceiling of x, *i.e.* , the next integer larger than x), for $x > 0$. Precisely when μ/U is an integer, there is a twofold degeneracy per site.



This degeneracy is broken by a small hopping term. Away from the degenerate points, within a single Mott plateau, the hopping term does very little (even away from mean field theory). This is because there is an energy gap, and $[N, H_{BH}] = 0$, which means that a small perturbation has no other states to mix in which might have other eigenvalues of N. Therefore, within a whole open set, the particle number remains fixed. This means $\partial_\mu \langle N \rangle = 0$, the system is incompressible.

We can find the boundaries of this region by expanding $\mathcal{E}_0$ in Ψ, following Landau: $\mathcal{E}_0 = \mathcal{E}_0^0 + r|\Psi|^2 + \mathcal{O}(|\Psi|^4)$. We can compute the coefficients in perturbation theory, and this produces the following picture.

Mean field theory gives the famous picture at right, with lobes of different Mott insulator states with different (integer!) numbers of bosons per site. (The hopping parameter w is called t in the figure.)



(Fig credit: Roman Lutchyn)

14.2.2 Coherent state path integral

Actually we can do a bit better; some of our hard work will pay off. Consider the coherent state path integral for the Euclidean partition sum

$$Z = \int [d^2 b] e^{-\int_0^{1/T} d\tau \mathcal{L}_b}$$

$$\text{with } \; \mathcal{L}_b = \sum_i \left(b_i^\dagger \partial_\tau b_i - \mu b_i^\dagger b_i + U b_i^\dagger b_i^\dagger b_i b_i \right) - \sum_{ij} b_i^\dagger w_{ij} b_j$$

where we introduced the hopping matrix $w_{ij} = w$ if $\langle ij \rangle$ share a link, otherwise zero. Here the bs are numbers, coherent state eigenvalues. Here is another application of the Hubbard-Stratonovich transformation:

$$Z = \int [d^2 b][d^2 \Psi] e^{-\int_0^{1/T} d\tau \mathcal{L}_b'}$$

$$\text{with } \; \mathcal{L}_b' = \sum_i \left(b_i^\dagger \partial_\tau b_i - \mu b_i^\dagger b_i + U b_i^\dagger b_i^\dagger b_i b_i - \Psi b_i^\dagger - \Psi^\star b_i \right) + \sum_{ij} \Psi_i w_{ij}^{-1} \Psi_j .$$

(Warning: if w has negative eigenvalues, so that the gaussian integral over Ψ is well-defined, we need to add a big constant to it, and subtract it from the single-particle terms.) Now integrate out the b fields. It's not gaussian, but notice that the resulting action for Ψ is the connected generating function $W[J]$: $\int [d^2 b] e^{-S[b] + \int \Psi b + h.c.} = e^{-W[\Psi, \Psi^\star]}$. More specifically,

$$Z = \int [d^2 \Psi] e^{-\frac{V}{T}\mathcal{F}_0 - \int_0^{1/T} d\tau \mathcal{L}_B}$$

$$\text{with } \mathcal{L}_B = K_1 \Psi^\star \partial_\tau \Psi + K_2 |\partial_\tau \Psi|^2 + K_3 |\vec{\nabla} \Psi|^2 + \tilde{r} |\Psi|^2 + u |\Psi|^4 + \cdots$$

Here $V = Ma^d$ is the volume of space, and $\mathcal{F}_0$ is the mean-field free energy. The coefficients K etc are connected Green's functions of the bs. The choice of which terms I wrote was dictated by Landau, and the order in which I wrote them should have been determined by Wilson. The Mott-SF transition occurs when $\tilde{r}$ changes sign, that is, the condition $\tilde{r} = 0$ determines the location of the Mott-SF boundaries. You can see that generically we have $z = 2$ kinetic terms. Less obvious is that $\tilde{r}$ is proportional to the mean field coefficient r.

Here's the payoff. I claim that the coefficients in the action for Ψ are related by

$$K_1 = -\partial_\mu \tilde{r}. \tag{14.18}$$

This means that $K_1 = 0$ precisely when the boundary of the lobe has a vertical tangent. This means that right at those points (the ends of the dashed lines in the figure) the second-order kinetic term is the leading one, and we have $z = 1$.

Here's the proof of (14.18). $\mathcal{L}_B$ must have the same symmetries as $\mathcal{L}_b$. One such invariance is

$$b_i \to b_i e^{\mathbf{i}\phi(\tau)}, \;\; \Psi_i \to \Psi_i e^{\mathbf{i}\phi(\tau)}, \mu \to \mu + \mathbf{i}\partial_\tau \phi.$$

This is a funny transformation which acts on the couplings, so doesn't produce Noether currents. It is still useful though, because it implies

$$0 = \delta_\phi \left(K_1 \Psi^\star \partial_\tau \Psi + \tilde{r} |\Psi|^2 + ... \right) = K_1 |\Psi|^2 \mathbf{i} \partial_\tau \phi + \partial_\mu \tilde{r} \mathbf{i} \partial \phi |\Psi|^2 + ...$$

14.2.3 Duality

We have seen above (in §14.1) that the prevention of vortices is essential to superfluidity, which is the condensation of bosons. In $D = 1 + 1$, vortices are events in spacetime. In $D = 2 + 1$, vortices are actual particles, *i.e.* localizable objects, around which the superfluid phase variable winds by 2π (times an integer).

More explicitly, if the boson field which condenses is $b(x) = ve^{\mathbf{i}\phi}$, and we choose polar coordinates in space $x + \mathbf{i}y \equiv Re^{\mathbf{i}\varphi}$, then a vortex is a configuration of the order parameter field of the form $b(x) = f(R)e^{\mathbf{i}\varphi}$, where $f(R) \overset{R\to\infty}{\to} v$ far away: the phase of the order parameter winds around. Notice that the phase is ill-defined in the core of the vortex where $f(R) \overset{R\to 0}{\to} 0$. (This is familiar from our discussion of the Abelian Higgs model.)

To see the role of vortices in destroying superfluidity more clearly, consider superfluid flow in a 2d annulus geometry, with the same polar coordinates $x + \mathbf{i}y = Re^{\mathbf{i}\varphi}$. If the superfluid phase variable is in the configuration $\phi(R, \varphi) = n\varphi$, then the current is

$$\vec{J}(R, \varphi) = \rho_s \vec{\nabla}\phi = \check{\varphi}\rho_s \frac{n}{2\pi R}.$$

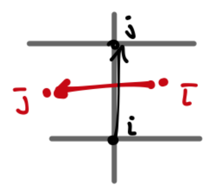
The current only changes if the integer n changes. This happens if vortices enter from the outside; removing the current (changing n to zero) requires n vortices to tunnel all the way through the sample, which if they are gapped and the sample is macroscopic can take a cosmologically long time.

There is a dual statement to the preceding three paragraphs: a state where the bosons themselves are gapped and localized – that is, a Mott insulator – can be described starting from the SF phase by the condensation of vortices. To see this, let us consider again the (simpler-than-Bose-Hubbard) $2+1$d rotor model

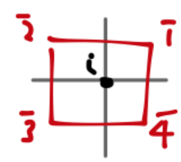
$$\mathbf{H}_{\text{rotors}} = U\sum_i \mathbf{n}_i^2 - J\sum_{\langle ij\rangle} \cos\left(\phi_i - \phi_j\right)$$

and introduce dual variables. Introduce a dual lattice whose sites are (centered in) the faces of the original (direct) lattice; each link of the dual lattice crosses one link of the direct lattice.

- First let $e_{\bar{i}\bar{j}} \equiv \frac{\phi_i - \phi_j}{2\pi}$. Here we define $\bar{i}\bar{j}$ by the right hand rule: $ij \times \bar{i}\bar{j} = +\check{z}$ (ij denotes the unit vector pointing from i to j). This is a lattice version of $\vec{e} = \check{z} \times \vec{\nabla}\phi\frac{1}{2\pi}$. Defining lattice derivatives $\Delta_x\phi_i \equiv \phi_i - \phi_{i+\check{x}}$, the definition is $e_x = -\frac{\Delta_y\phi}{2\pi}, e_y = \frac{\Delta_x\phi}{2\pi}$. It is like an electric field vector.



- The conjugate variable to the electric field is $a_{\bar{i}\bar{j}}$, which must therefore be made from the conjugate variable of ϕ_i, namely $\mathbf{n}_i$: $[\mathbf{n}_i, \phi_j] = -\mathbf{i}\delta_{ij}$. Acting with $\mathbf{n}_i$ translates ϕ_i, which means that it shifts all the $e_{\bar{i}\bar{j}}$ from the surrounding plaquettes. More precisely:



$$2\pi\mathbf{n}_i = a_{\bar{1}\bar{2}} + a_{\bar{2}\bar{3}} + a_{\bar{3}\bar{4}} + a_{\bar{4}\bar{1}}.$$

This is a lattice, integer version of $n \sim \frac{1}{2\pi}\vec{\nabla} \times a \cdot \check{z}$. In terms of these variables,

$$\mathbf{H}_{\text{rotors}} = \frac{U}{2}\sum_i \left(\frac{\Delta \times a}{2\pi}\right)^2 - J\sum_{\langle \bar{i}\bar{j}\rangle} \cos\left(2\pi e_{\bar{i}\bar{j}}\right)$$

with the following *constraint*. If it were really true that $\vec{e} = \frac{1}{2\pi}\check{z} \times \vec{\nabla}\phi$, with single-valued ϕ, then $\vec{\nabla} \cdot \vec{e} = \vec{\nabla} \cdot \left(\check{z} \times \vec{\nabla}\phi\right) = 0$. But there are vortices in the world, where ϕ is not single valued. The number of vortices $n_v(R)$ in some region R with $\partial R = C$ is determined by the winding number of the phase around C:

$$2\pi n_v(R) = \oint_C d\vec{\ell} \cdot \vec{\nabla}\phi \overset{\text{Stokes}}{=} 2\pi \int_R d^2x \vec{\nabla} \cdot \vec{e}$$

(More explicitly, $2\pi\vec{\nabla}\cdot\vec{e} = \epsilon_{zij}\partial_i\partial_j\phi = [\partial_x, \partial_y]\phi$ clearly vanishes if ϕ is single-valued.) Since this is true for any region R, we have

$$\vec{\nabla}\cdot\vec{e} = 2\pi\delta^2(\text{vortices}).$$

Actually, the lattice version of the equation has more information (and is true) because it keeps track of the fact that the number of vortices is an integer:

$$\Delta_x e_x + \Delta_y e_y \equiv \vec{\Delta}\cdot\vec{e}(\bar{i}) = 2\pi n_v(\bar{i}), \qquad n_v(\bar{i}) \in \mathbb{Z}.$$

It will not escape your notice that this is Gauss' law, with the density of *vortices* playing the role of the charge density.

Phases of the 2d rotors. Since $\vec{e} \sim \vec{\nabla}\phi$ varies continuously, *i.e.* electric flux is not quantized, this is called *noncompact* electrodynamics. Again we will impose the integer constraint $a \in 2\pi\mathbb{Z}$ energetically, *i.e.* let $a \in \mathbb{R}$ and add (something like) $\Delta\mathbf{H} \stackrel{?}{=} -t\cos a$ and see what happens when we make t finite. The expression in the previous sentence is not quite right, yet, however: This operator does not commute with our constraint $\vec{\Delta}\cdot\vec{e} - 2\pi n_v = 0$ – it jumps $\vec{e}$ but not n_v[33].

We can fix this by introducing explicitly the variable which creates vortices, $e^{-\mathbf{i}\chi}$, with:

$$[n_v(\bar{i}), \chi(\bar{j})] = -\mathbf{i}\delta_{\bar{i}\bar{j}} \ .$$

Certainly our Hilbert space contains states with different number of vortices, so we can introduce an operator which maps these sectors. Its locality might be an issue: certainly it is nonlocal with respect to the original variables, but we will see that we can treat it as a local operator (except for the fact that it carries gauge charge) in the dual description. Since $n_v \in \mathbb{Z}$, $\chi \simeq \chi + 2\pi$ lives on a circle. So:

$$\mathbf{H} \sim \sum_{\bar{i}} \left(\frac{U}{2}\left(\frac{\Delta\times a}{2\pi}\right)^2 + \frac{J}{2}\left(2\pi e\right)^2 - t\cos\left(\Delta\chi - a\right) \right)$$

still subject to the constraint $\vec{\Delta}\cdot\vec{e} = 2\pi n_v$.

Two regimes:

$\boxed{J \gg U, t}$: This suppresses e and its fluctuations, which means a fluctuates. The fluctuating a is governed by the gaussian hamiltonian

$$\mathbf{H} \sim \sum \left(\vec{e}^2 + \vec{b}^2\right)$$

[33] A set of words which has the same meaning as the above: $\cos a$ is not gauge invariant. Understanding these words requires us to think of the operator $G(\bar{i}) \equiv \vec{\Delta}\cdot\vec{e} - 2\pi n_v$ as the generator of a transformation,

$$\delta\mathcal{O} = \sum_{\bar{i}} s(\bar{i})[G(\bar{i}), \mathcal{O}].$$

It can be a useful picture.

with $b \equiv \frac{\Delta \times a}{2\pi}$, which should look familiar. This *deconfined* phase has a gapless photon; a $2+1$d photon has a single polarization state. This is the goldstone mode, and this regime describes the superfluid phase (note that the parameters work out right in the original variables). The relation between the photon a and the original phase variable, in the continuum is

$$\epsilon_{\mu\nu\rho}\partial_\nu a_\rho = \partial_\mu \phi.$$

$\boxed{t \gg U, J}$: In this regime we must satisfy the cosine first. Like in $D = 1+1$, this can be described as the statement that vortices condense. Expanding around its minimum, the cosine term is

$$\mathbf{h} \ni t\,(a - \partial\chi)^2$$

– the photon gets a mass by eating the phase variable χ. There is an energy gap. This is the Mott phase.

If the vortices carry other quantum numbers, the (analog of the) Mott phase can be more interesting, as we'll see in section 14.3.

Compact electrodynamics in $D = 2+1$. Note that this free photon phase of $D = 2+1$ electrodynamics is not accessible if e is quantized (so-called compact electrodynamics) where monopole instantons proliferate and gap out the photon. This is the subject of §14.2.5.

14.2.4 Particle-vortex duality in the continuum

The above is easier to understand (but a bit less precise) in the continuum. Consider a quantum system of bosons in $D = 2+1$ with a $\mathsf{U}(1)$ particle-number symmetry (a real symmetry, not a gauge redundancy). Let's focus on a complex, non-relativistic bose field b with action

$$S[b] = \int \mathrm{d}t\mathrm{d}^2x \left(b^\dagger \left(\mathbf{i}\partial_t - \vec{\nabla}^2 - \mu \right) b - U(b^\dagger b)^2 \right). \tag{14.19}$$

By Noether's theorem, the symmetry $b \to e^{\mathbf{i}\theta} b$ implies that the current

$$j_\mu = (j_t, \vec{j})_\mu = (b^\dagger b, \mathbf{i} b^\dagger \vec{\nabla} b + h.c.)_\mu$$

satisfies the continuity equation $\partial^\mu j_\mu = 0$.

This system has two phases of interest here. In the ordered/broken/superfluid phase, where the groundstate expectation value $\langle b \rangle = \sqrt{\rho_0}$ spontaneously breaks the U(1) symmetry, the goldstone boson θ in $b \equiv \sqrt{\rho_0} e^{\mathbf{i}\theta}$ is massless

$$S_{\text{eff}}[\theta] = \frac{\rho_0}{2} \int \left(\dot{\theta}^2 - \left(\vec{\nabla}\theta \right)^2 \right) \mathrm{d}^2x\mathrm{d}t, \quad j_\mu = \rho_0 \partial\theta\ .$$

In the disordered/unbroken/Mott insulator phase, $\langle b \rangle = 0$, and there is a mass gap. A dimensionless parameter which interpolates between these phases is $g = \mu/U$; large g encourages condensation of b.

We can 'solve' the continuity equation by writing

$$j^\mu = \epsilon^{\mu\cdot\cdot}\partial_\cdot a_\cdot \tag{14.20}$$

where $a_\cdot$ is a gauge potential. The time component of this equation says that the boson density is represented by the magnetic flux of a. The spatial components relate the boson charge current to the electric flux of a. The continuity equation for j is automatic – it is the Bianchi identity for a – as long as a is single-valued. That is: as long as there is no magnetic charge present. A term for this condition which is commonly used in the cond-mat literature is: "a is non-compact." (More on the other case below.)

The relation (14.20) is the basic ingredient of the duality, but it is not a complete description: in particular, how do we describe the boson itself in the dual variables? In the disordered phase, adding a boson is a well-defined thing which costs a definite energy. The boson is described by a localized clump of magnetic flux of a. Such a configuration is energetically favored if a participates in a superconductor – i.e. if a is coupled to a condensate of a charged field. The Meissner effect will then ensure that its magnetic flux is bunched together. So this suggests that we should introduce into the dual description a scalar field, call it Φ, minimally coupled to the gauge field a:

$$S[b] \leftrightsquigarrow S_{\text{dual}}[a, \Phi] \ .$$

And the disordered phase should be dual to a phase where $\langle \Phi \rangle \neq 0$, which gives a mass to the gauge field by the Anderson-Higgs mechanism.

Who is Φ? More precisely, what is the identity in terms of the original bosons of the particles it creates? When Φ is not condensed and its excitations are massive, the gauge field is massless. This the Coulomb phase of the Abelian Higgs model $S[a, \Phi]$; at low energies, it is just free electromagnetism in $D = 2 + 1$. These are the properties of the ordered phase of b. (This aspect of the duality is explained in Wen, §6.3.) The photon has one polarization state in $D = 2 + 1$ and is dual to the goldstone boson. This is the content of (14.20) in the ordered phase: $\epsilon^{\mu\cdot\cdot}\partial_\cdot a_\cdot = \rho_0 \partial_\mu \theta$ or $\star \mathrm{d}a = \rho_0 \mathrm{d}\theta$.

Condensing Φ gives a mass to the Goldstone boson whose masslessness is guaranteed by the broken $\mathsf{U}(1)$ symmetry. Therefore Φ is a disorder operator: its excitations are vortices in the bose condensate, which are gapped in the superfluid phase. The transition to the insulating phase can be described as a condensation of these vortices.

The vortices have relativistic kinetic terms, *i.e.* particle-hole symmetry. This is the statement that in the ordered phase of the time-reversal invariant bose system, a vortex and an antivortex have the same energy. An argument for this claim is the following. We may create vortices by rotating the sample, as was done in the figure at right. With time-reversal symmetry, rotating the sample one way will cost the same energy as rotating it the other way.

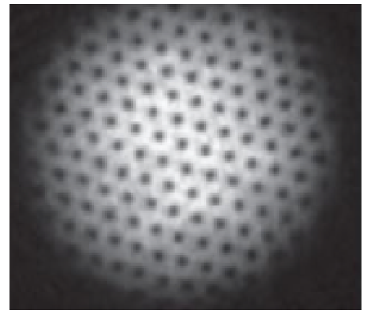


Fig: M. Zwierlein.

This means that the mass of the vortices $m_V^2 \Phi^\dagger \Phi$ is distinct from the vortex chemical potential $\mu_V \rho_V = \mu_V \mathbf{i} \Phi^\dagger \partial_t \Phi + h.c.$. The vortex mass[2] maps under the duality to the boson chemical potential. Taking it from positive to negative causes the vortices to condense and disorder (restore) the $\mathsf{U}(1)$ symmetry.

To what does the vortex chemical potential map? It is a term which breaks time-reversal, and which encourages the presence of vortices in the superfluid order. It's an external magnetic field for the bosons. (This also the same as putting the bosons into a rotating frame.)

To summarize, a useful dual description is the Abelian Higgs model

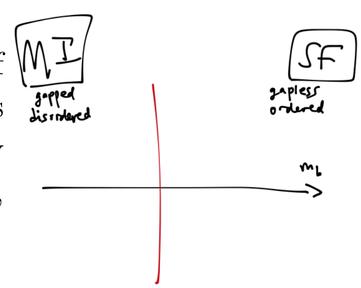
$$S[a, \Phi] = \int \mathrm{d}^2x \mathrm{d}t \left(\Phi^\dagger \left(\left(\mathbf{i}\partial_t - \mathbf{i}A_t - \mu \right)^2 + \left(\vec{\nabla} + \vec{A} \right)^2 \right) \Phi - \frac{1}{e^2} f_{\mu\nu} f^{\mu\nu} - V(\Phi^\dagger \Phi) \right).$$

We can parametrize V as

$$V = \lambda \left(\Phi^\dagger \Phi - v \right)^2$$

– when $v < 0$, $\langle \Phi \rangle = 0$, Φ is massive and we are in the Coulomb phase. When $v > 0$ Φ condenses and we are in the Anderson-Higgs phase.

The description above is valid near the boundary of one of the MI phases. At the tips of the lobes are special points where the bosons b themselves have particle-hole symmetry (*i.e.* relativistic kinetic terms). For more on this diagram, see *e.g.* chapter 9 of Sachdev.



In the previous discussion I have been assuming that the vortices of b have unit charge under a and are featureless bosons, *i.e.* do not carry any non-trivial quantum numbers under any other symmetry. If *e.g.* the vortices have more-than-minimal charge under a, say charge q, then condensing them leaves behind a $\mathbb{Z}_q$ gauge theory and produces a state with topological order. If the vortices carry some charge under some other symmetry (like lattice translations or rotations) then condensing them breaks that symmetry. If the vortices are minimal-charge fermions, then they can only condense in pairs, again leaving behind an unbroken $\mathbb{Z}_2$ gauge theory.

[End of Lecture 59]

14.2.5 Compact electrodynamics in $D = 2 + 1$

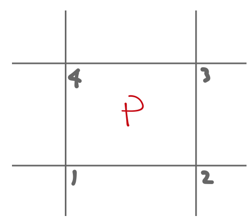
Consider a quantum system on a two-dimensional lattice (say, square) with rotors $\Theta_l \equiv \Theta_l + 2\pi m$ on the links l. (Think of this as the phase of a boson or the direction of an easy-plane spin.) The conjugate variable $\mathbf{n}_l$ is an integer

$$[\mathbf{n}_l, \Theta_{l'}] = -\mathbf{i}\delta_{l,l'}.$$

Here $\mathbf{n}_{ij} = \mathbf{n}_{ji}, \Theta_{ij} = \Theta_{ji}$ – we have not oriented our links (yet). We also impose the *Gauss' law constraint*

$$\mathbf{G}_s \equiv \sum_{l \in v(s)} \mathbf{n}_l = 0 \ \ \forall \text{ sites } s,$$

where the notation $v(s)$ means the set of links incident upon the site s ('v' is for 'vicinity').



We'll demand that the Hamiltonian is 'gauge invariant', that is, that $[\mathbf{H}, \mathbf{G}_s] = 0 \forall s$. Any terms which depend only on $\mathbf{n}$ are OK. The natural single-valued object made from Θ is $e^{\mathbf{i}\Theta_l}$, but this is not gauge invariant. A combination which is gauge invariant is the *plaquette operator*, associated to a face p of the lattice:

$$\prod_{l \in \partial p} e^{(-1)^y \mathbf{i}\Theta_l} \equiv e^{\mathbf{i}(\Theta_{12} - \Theta_{23} + \Theta_{34} - \Theta_{41})}$$

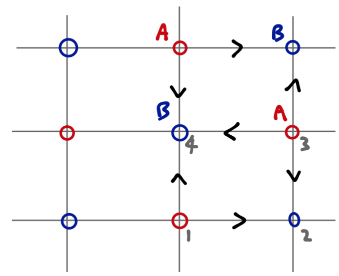
– we put a minus sign on the horizontal links. ∂p denotes the links running around the boundary of p. So a good hamiltonian is

$$\mathbf{H} = \frac{U}{2} \sum_l \mathbf{n}_l^2 - K \sum_{\square} \cos \left(\sum_{l \in \partial \square} (-1)^y \Theta_l \right).$$

Local Hilbert space. The space of gauge-invariant states is not a tensor product over local Hilbert spaces. This sometimes causes some confusion. Notice, however, that we can arrive at the gauge-theory hilbert space by imposing the Gauss' law constraint energetically (as in the toric code): Start with the following Hamiltonian acting on the full unconstrained rotor Hilbert space:

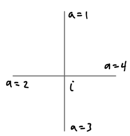
$$H_{\text{big}} = +\Gamma_\infty \sum_i \mathbf{G}_i + \mathbf{H}.$$

True to its name, the coefficient Γ_∞ is some huge energy scale which penalizes configurations which violate Gauss' law (if you like, such configurations describe some matter with rest mass Γ_∞). So, states with energy $\ll \Gamma_\infty$ all satisfy Gauss' law. Then further, we want $\mathbf{H}$ to act *within* this subspace, and not create excitations of enormous energies like Γ_∞. This requires $[\mathbf{G}_i, \mathbf{H}] = 0, \forall i$, which is exactly the condition that $\mathbf{H}$ is gauge invariant.



A useful change of variables gets rid of these annoying signs. Assume the lattice is bipartite: made of two sublattices A, B each of which only touches the other. Then draw arrows from A sites to B sites, and let

$$\begin{matrix} \mathbf{e}_{ij} \equiv \eta_i \mathbf{n}_{ij} \\ \mathbf{a}_{ij} \equiv \eta_i \Theta_{ij} \end{matrix}, \quad \eta_i \equiv \begin{cases} +1, & i \in A \\ -1, & i \in B \end{cases}.$$



Then the Gauss constraint now reads

$$0 = \mathbf{e}_{\bar{i}\bar{1}} + \mathbf{e}_{\bar{i}\bar{2}} + \mathbf{e}_{\bar{i}\bar{3}} + \mathbf{e}_{\bar{i}\bar{4}} \equiv \Delta \cdot \mathbf{e}(\bar{i}).$$

This is the lattice divergence operation. The plaquette term reads

$$\cos\left(\Theta_{12} - \Theta_{23} + \Theta_{34} - \Theta_{41}\right) = \cos\left(\mathbf{a}_{12} + \mathbf{a}_{23} + \mathbf{a}_{34} + \mathbf{a}_{41}\right) \equiv \cos\left(\Delta \times \mathbf{a}\right)$$

– the lattice curl (more precisely, it is $(\Delta \times \mathbf{a}) \cdot \check{z}$). In these variables,

$$\mathbf{H} = \frac{U}{2} \sum_l \mathbf{e}_l^2 - K \sum_\square \cos\left((\Delta \times \mathbf{a}) \cdot \check{n}_\square\right)$$

(in the last term we emphasize that this works in $D \geq 2+1$ if we remember to take the component of the curl normal to the face in question). This is (compact) lattice $\mathsf{U}(1)$ gauge theory, with no charges. The word 'compact' refers to the fact that the charge is quantized; the way we would add charge is by modifying the Gauss' law to

$$\underbrace{\Delta \cdot \mathbf{e}(\bar{i})}_{\in \mathbb{Z}} = \underbrace{\text{charge at } \bar{i}}_{\implies \in \mathbb{Z}}$$

where the charge must be quantized because the LHS is an integer. (In the noncompact electrodynamics we found dual to the superfluid, it was the continuous angle variable which participated in the Gauss' law, and the discrete variable which was gauge variant.)

What is it that's compact in compact QED?

The operator appearing in Gauss' law

$$\mathbf{G}(x) \equiv \left(\vec{\nabla} \cdot \vec{\mathbf{e}}(x) - 4\pi \mathbf{n}(x)\right)$$

(here $\mathbf{n}(x)$ is the density of charge) is the generator of gauge transformations, in the sense that a gauge transformation acts on any operator $\mathcal{O}$ by

$$\mathcal{O} \mapsto e^{-\mathbf{i}\sum_x \alpha(x)\mathbf{G}(x)} \mathcal{O} e^{\mathbf{i}\sum_x \alpha(x)\mathbf{G}(x)} \tag{14.21}$$

This is a fact we've seen repeatedly above, and it is familiar from ordinary QED, where using the canonical commutation relations

$$[\mathbf{a}^i(x), \mathbf{e}^j(y)] = -\mathbf{i}\delta^{ij}\delta(x-y), \quad [\phi(x), \mathbf{n}(y)] = -\mathbf{i}\delta(x-y)$$

(ϕ is the phase of a charged field, $\Phi = \rho e^{\mathbf{i}\phi}$) in (14.21) reproduce the familiar gauge transformations

$$\vec{\mathbf{a}} \to \vec{\mathbf{a}} + \vec{\nabla}\alpha, \qquad \phi \to \phi + \alpha \quad .$$

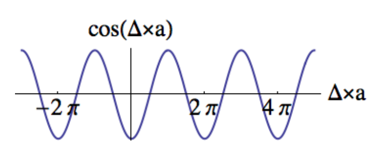
SO: if all the objects appearing in Gauss' law are integers (which is the case if charge is quantized and electric flux is quantized), it means that the gauge parameter α itself only enters mod 2π, which means the gauge transformations live in $\mathsf{U}(1)$, as opposed to $\mathbb{R}$. So it's the gauge group that's compact.

This distinction is very important, because (in the absence of matter) this model does not have a deconfined phase! To see this result (due to Polyakov), first consider strong coupling:

$\boxed{U \gg K}$: The groundstate has $\mathbf{e}_{\bar{l}} = 0, \ \forall \bar{l}$. (Notice that this configuration satisfies the constraint.) There is a gap to excitations where some link has an *integer* $\mathbf{e} \neq 0$, of order U. (If $\mathbf{e}$ were continuous, there would not be a gap!) In this phase, electric flux is confined, *i.e.* costs energy and is generally unwanted.

$\boxed{U \gg K}$: The surprising thing is what happens when we make the gauge coupling weak.

Then we should first minimize the magnetic flux term: minimizing $-\cos(\Delta \times \mathbf{a})$ means $\Delta \times \mathbf{a} \in 2\pi\mathbb{Z}$. Near each minimum, the physics looks like Maxwell, $\mathbf{h} \sim \mathbf{e}^2 + \mathbf{b}^2 + \cdots$. BUT: it turns out to be a colossally bad idea to ignore the tunnelling *between* the minima. To see this, begin by solving the Gauss law constraint $\Delta \cdot \mathbf{e} = 0$ by introducing



$$\mathbf{e}_{\bar{1}\bar{2}} \equiv \frac{1}{2\pi}(\chi_2 - \chi_1) \tag{14.22}$$

($i.e.$ $\vec{\mathbf{e}} = \check{z} \cdot \Delta\chi\frac{1}{2\pi}$.) χ is a (discrete!) 'height variable'. Then the operator

$$e^{\mathbf{i}(\Delta\times\mathbf{a})(\bar{i})}$$

increases the value of $\mathbf{e}_{\bar{i}\bar{a}}$ for all neighboring sites $\bar{a}$, which means it jumps $\chi_{\bar{i}} \to \chi_{\bar{i}} + 2\pi$. So we should regard

$$(\Delta \times \mathbf{a})(\bar{i}) \equiv \Pi_\chi(\bar{i})$$

as the conjugate variable to χ, in the sense that

$$[\Pi_\chi(r), \chi(r')] = -\mathbf{i}\delta_{rr'}.$$

Notice that this is consistent with thinking of χ as the dual scalar related to the gauge field by our friend the (Hodge) duality relation

$$\partial_\mu \chi = \epsilon_{\mu\nu\rho}\partial_\nu a_\rho.$$

The spatial components i say $\partial_i \chi = \epsilon_{ij} f_{0j}$, which is the continuum version of (14.22). The time component says $\dot{\chi} = \epsilon_{ij} f_{ij} = \nabla \times a$, which indeed says that (if χ has quadratic kinetic terms), the field momentum of χ is the magnetic flux. So χ is the would-be transverse photon mode.

The hamiltonian is now

$$\mathbf{H} = \frac{U}{2}\sum_l (\Delta\chi)^2 - K\sum_r \cos\Pi_\chi(r)$$

with no constraint. In the limit $U \gg K$, the spatial gradients of χ are forbidden – χ wants to be uniform. From the definition (14.22), uniform χ means there are no electric field lines, this is the confined phase. Deconfinement limit should be $K \gg U$, in which case it looks like we can Taylor expand the cosine $\cos\Pi_\chi \sim 1 - \frac{1}{2}\Pi_\chi^2$ about one of its minima, and get harmonic oscillators. But: tunneling between the neighboring vacua of $\Delta \times \mathbf{a}$ is accomplished by the *flux-insertion operator* (or *monopole operator*)

$$e^{\mathbf{i}\chi}, \quad \text{which satisfies} \quad [e^{\mathbf{i}\chi(r)}, (\Delta \times \mathbf{a})(r')] = e^{\mathbf{i}\chi(r)}\delta_{rr'}$$

– that is, $e^{\mathbf{i}\chi}$ is a raising operator for $\Delta \times \mathbf{a}$. To analyze whether the Maxwell limit survives this, let's go to the continuum and study perturbations of the free hamiltonian

$$\mathbf{H}_0 = \int \left(\frac{U}{2}\left(\vec{\nabla}\chi\right)^2 + \frac{K}{2}\Pi_\chi^2\right)$$

by

$$\mathbf{H}_1 = -\int V_0 \cos\chi \ .$$

This operator introduces tunneling events by $\Pi_\chi \to \Phi_\chi \pm 2\pi$ with rate V_0. Alternatively, notice that again we can think of the addition of this term as energetically imposing the condition that $\chi \in 2\pi\mathbb{Z}$.

So: is V_0 irrelevant? Very much no. In fact

$$\langle \cos\chi(r)\cos\chi(0)\rangle_0 \sim \text{const} \tag{14.23}$$

has constant amplitude at large r! That means that the operator has dimension zero, and the perturbation in the action has $[S_1 = -\int V_0 \cos\chi d^2 x d\tau] \sim L^3$, very relevant. The result is that it pins the χ field (the would-be photon mode) to an integer, from which it can't escape. This result is due to Polyakov.

To see (14.23) begin with the gaussian identity

$$\left\langle e^{\mathbf{i}s\chi(x)}e^{\mathbf{i}s'\chi(0)}\right\rangle = e^{-\frac{ss'}{2}\langle\chi(x)\chi(0)\rangle} ,$$

with $s, s' = \pm$. The required object is

$$\begin{aligned}
\langle\chi(x)\chi(0)\rangle &= \frac{\mathbf{i}}{T}\int đ^3 p \frac{e^{\mathbf{i}\vec{p}\cdot\vec{x}}}{p^2} = \mathbf{i}\frac{2\pi}{(2\pi)^3 T}\int_0^\infty dp \underbrace{\int_{-1}^1 d\cos\theta e^{\mathbf{i}px\cos\theta}}_{=\frac{2\sin px}{px}} \\
&= \mathbf{i}\frac{2}{(2\pi)^2 T}\int_0^\infty dp \frac{\sin px}{px} \\
&= \mathbf{i}\frac{2}{2\pi T}\frac{1}{x}\frac{1}{2}\underbrace{\int_{-\infty}^\infty d\bar{p}\frac{\sin\bar{p}}{\bar{p}}}_{=\pi} \\
&= \frac{\mathbf{i}}{2Tx}.
\end{aligned} \tag{14.24}$$

(I have set the velocity of propagation to 1, and $T \equiv U/K$ is the coefficient in front of the Lagrangian, $S = T\int d^3 x \partial_\mu\chi\partial^\mu\chi$.) So

$$\left\langle e^{\mathbf{i}s\chi(x)}e^{\mathbf{i}s'\chi(0)}\right\rangle = e^{-\mathbf{i}\frac{ss'}{4xT}} .$$

And

$$\langle\cos\chi(x)\cos\chi(0)\rangle = \cos\frac{1}{4Tx}$$

which does not decay at long distance, and in fact approaches a constant.

- The fact that the would-be-transverse-photon χ is massive means confinement of the gauge theory. To see that external charge is confined, think as usual about the big rectangular Wilson loop $\langle W(\Box)\rangle = \left\langle e^{\mathbf{i}\oint_\Box A}\right\rangle \overset{\text{euclidean}}{\sim} e^{-E(R)T}$ as an order parameter for confinement. In term of χ,

$$\oint_\Box A = \int_\blacksquare F_{12} = \int_\blacksquare g\dot{\chi}$$

(I've absorbed a factor of the gauge coupling into χ to make the dimensions work nicely, $\epsilon_{\mu\nu\rho}\partial_\nu A_\rho = g\partial_\mu\chi$) and the expectation is

$$\langle W(\Box)\rangle = Z^{-1}\int[d\chi]e^{-S_\chi + g\mathbf{i}\int_\blacksquare \dot{\chi}} \sim e^{-cg^2 m_\chi\cdot\text{area}(\blacksquare)}.$$

In the last step we did the gaussian integral from small χ fluctuations. This area-law behavior proportional to m_χ means that the mass for χ confines the gauge theory. This is the same (Polyakov) effect we saw in the previous section, where the monopole tunneling events produced the mass.

- Adding matter helps to produce a deconfined phase! In particular, the presence of enough massless charged fermions can render the monopole operator irrelevant. I recommend this paper by Tarun Grover for more on this.

- Think about the action of $e^{\mathbf{i}\chi(x,t)}$ from the point of view of $2+1$d spacetime: it inserts 2π magnetic flux at the spacetime point x, t. From that path integral viewpoint, this is an event localized in three dimensions which is a source of magnetic flux – a magnetic monopole. In Polyakov's paper, he uses a UV completion of the abelian gauge theory (not the lattice) in which the magnetic monopole is a smooth solution of field equations (the 't Hooft-Polyakov monopole), and these solutions are instanton events. The $\cos\chi$ potential we have found above arises from, that point of view, by the same kind of dilute instanton gas sum that we did in the $D = 1 + 1$ Abelian Higgs model.

14.3 Deconfined Quantum Criticality

[The original papers are this and this; this treatment follows Ami Katz' BU Physics 811 notes.] Consider a square lattice with quantum spins (spin half) at the sites, governed by the Hamiltonian

$$H_{JQ} \equiv J \sum_{\langle ij \rangle} \vec{S}_i \cdot \vec{S}_j + Q \sum_{[ijkl]} \left(\vec{S}_i \cdot \vec{S}_j - \frac{1}{4} \right) \left(\vec{S}_k \cdot \vec{S}_l - \frac{1}{4} \right).$$

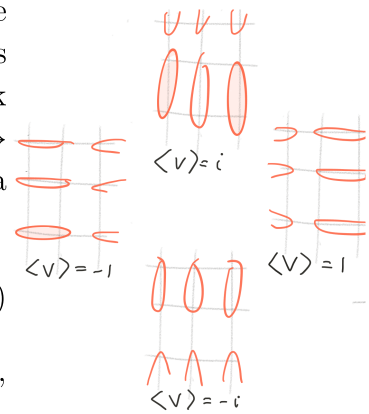
Here $\langle ij \rangle$ denotes pairs of sites which share a link, and $[ijkl]$ denotes groups of four sites at the corners of a plaquette. This JQ-model is a somewhat artificial model designed to bring out the following competition which also exists in more realistic models:

$\boxed{J \gg Q}$: the groundstate is a Neel antiferromagnet (AFM), with local order parameter $\vec{n} = \sum_i (-1)^{x_i+y_i} \vec{S}_i$, whose expectation value breaks the spin symmetry $\mathsf{SU}(2) \to \mathsf{U}(1)$. Hence, the low-energy physics is controlled by the (two) Nambu-Goldstone modes. This is well-described by the field theory we studied in §11.3.

$\boxed{Q \gg J}$: The Q-term is designed to favor configurations where the four spins around each square form a pair of singlets. A single Q-term has a two-fold degenerate groundstate, which look like $|{=}\rangle$ and $|{||}\rangle$. The sum of all of them has four groundstates, which look like ... These are called *valence-bond solid* (VBS) states. The VBS order parameter on the square lattice is

$$V = \sum_i \left((-1)^{x_i} \vec{S}_i \cdot \vec{S}_{i+x} + \mathbf{i}(-1)^{y_i} \vec{S}_i \cdot \vec{S}_{i+y} \right) \in \mathbb{Z}_4.$$

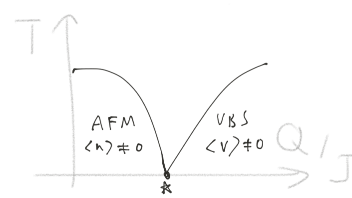
In the four solid states, it takes the values $1, \mathbf{i}, -1, -\mathbf{i}$. Notice that they are related by multiplication by $\mathbf{i} = e^{\mathbf{i}\pi/2}$. V is a singlet of the spin $\mathsf{SU}(2)$, but the VBS states do break spacetime symmetries: a lattice rotation acts by $R_{\pi/2} : V \to -\mathbf{i}V$ (the Neel order $\vec{n}$ is invariant), while a translation by a single lattice site acts by



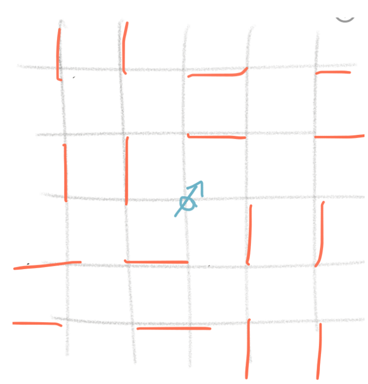
$$T_{x,y} : \vec{n} \to -\vec{n}, \; T_x : V \to -V^\dagger, T_y : V \to V^\dagger. \tag{14.25}$$

The VBS phase is gapped (it only breaks discrete symmetries, so no goldstones).

Claim: There seems to be a continuous transition between these two phases as a function of Q/J. (If it is first order, the latent heat is very small.) Here's why this is weird and fascinating: naively, the order parameters break totally different symmetries, and so need have nothing to do with each other. Landau then predicts that generically there should be a region where both are nonzero or where both are zero. Why should the transitions coincide? What are the degrees of freedom at $\star$?



To get a big hint, notice that the VBS order parameter is like a discrete rotor: if we had a triangular lattice it would be in $\mathbb{Z}_6$ and would come closer to approximating a circle-valued field. In any case, we can consider vortex configurations, where the phase of V rotates (discretely, between the four quadrants) as we go around a point in space. Such a vortex looks like the picture at right.



Notice that inside the core of the vortex, there is necessarily a spin which is not paired with another spin: The vortex carries spin: it transforms as a doublet under the spin $\mathsf{SU}(2)$. Why do we care about such vortices? I've been trying to persuade you for the past two sections that the way to think about destruction of (especially $\mathsf{U}(1)$) ordered phases is by proliferating vortex defects. Now think about proliferating this kind of VBS vortex. Since it carries spin, it necessarily must break the $\mathsf{SU}(2)$

symmetry, as the Neel phase does. This is why the transitions happen at the same point.

To make this more quantitative, let's think about it from the AFM side: how do we make V from the degrees of freedom of the low energy theory? It's not made from n since it's a spin singlet which isn't 1 (spin singlets made from n are even under a lattice translation). What about the $\mathbb{CP}^1$ version, aka the Abelian Higgs model, aka scalar QED (now in $D = 2+1$)?

$$L = -\frac{1}{4g^2} F^2 + |Dz|^2 - m^2|z|^2 - \frac{\lambda}{4}|z|^4$$

where $z = \begin{pmatrix} z_\uparrow \\ z_\downarrow \end{pmatrix}$, and $D_\mu z = (\partial_\mu - \mathbf{i} A_\mu) z$ as usual. Let's think about the phases of this model.

$\boxed{m^2 < 0}$: Here z condenses and breaks $\mathsf{SU}(2) \to \mathsf{U}(1)$, and A_μ is higgsed. A gauge invariant order parameter is $\vec{n} = z^\dagger \vec{\sigma} z$, and there are two goldstones associated with its rotations. This is the AFM. The cautionary tale I told you about this phase in $D = 1+1$ doesn't happen because now the vortices are *particles* rather than instanton events. More on these particles below.

$\boxed{m^2 > 0}$: Naively, in this phase, z are uncondensed and massive, leaving at low energies only $L_{\text{low-E}} \stackrel{?}{=} -\frac{1}{4g^2} F^2$, Maxwell theory in $D = 2+1$. This looks innocent but it will occupy us for quite a few pages starting now. This model has a conserved current (conserved by the Bianchi identity)

$$J_F^\mu \equiv \epsilon^{\mu\nu\rho} F_{\nu\rho}.$$

The thing that's conserved is the lines of magnetic flux. We can follow these more effectively by introducing the dual scalar field by a by-now-familiar duality relation:

$$J_F^\mu \equiv \epsilon^{\mu\nu\rho} F_{\nu\rho} \equiv g \partial^\mu \chi. \tag{14.26}$$

You can think of the last equation here as a *solution* of the conservation law $\partial_\mu J_F^\mu = 0$. The symmetry acts on χ by shifts: $\chi \to \chi + \text{constant}$. In terms of χ, the Maxwell action is

$$L_{\text{low-E}} \stackrel{?}{=} -\frac{1}{4g^2} F^2 = \frac{1}{2} \partial_\mu \chi \partial^\mu \chi.$$

But this is a massless scalar, a gapless theory. And what is the $\chi \to \chi + c$ symmetry in terms of the spin system? I claim that it's the rotation of the phase of the VBS order parameter, which is explicitly broken by the squareness of the square lattice. An improvement would then be

$$L_{\text{low-E}} = \frac{1}{2} \partial_\mu \chi \partial^\mu \chi - \frac{1}{2} m_\chi^2 \chi^2 + \cdots$$

where $m_\chi \sim \frac{1}{a^2}$ comes from the lattice breaking the rotation invariance (a is the lattice spacing).

To see that shifts of χ are VBS rotations, let's reproduce the lattice symmetries in the Abelian Higgs model. Here's the action of lattice translations $T \equiv T_x$ or T_y (take a deep breath.): $T : n^a \to -n^a$ but $n^a = z^\dagger \sigma^a z$, so on z we must have $T : z \to \mathbf{i}\sigma^2 z^\star$. The gauge current is $j_\mu = \mathbf{i} z^\dagger \partial_\mu z + h.c. \to -j_\mu$ which means we must have $A_\mu \to -A_\mu$ and $F_{\mu\nu} \to -F_{\mu\nu}$. Therefore by (14.26) we must have $T : \partial\chi \to -\partial\chi$ which means that

$$T_{x,y} : \chi \to -\chi + g\alpha_{x,y}$$

where $\alpha_{x,y}$ are some so-far-undetermined numbers, and g is there on dimensional grounds. Therefore, by choosing $T_{x,y}\chi \to -\chi \pm g\pi/2, R_{\pi/2} : \chi \to \chi - g\pi/2$ we can reproduce the transformation (14.25) by identifying

$$V = ce^{\mathbf{i}\chi/g}$$

(up to an undetermined overall complex number). Notice for future reference the canonical commutation relation between the flux current density ($J^0_F = g\dot{\chi} = \frac{g}{\mathbf{i}}\frac{\delta}{\delta\chi}$) and V:

$$[J^0_F(x), V(0)] = V(0)\delta^2(x). \tag{14.27}$$

It creates flux.

So χ is like the phase of the bosonic operator V which is condensed in the VBS phase; lattice effects break the $\mathsf{U}(1)$ symmetry down to some discrete subgroup ($\mathbb{Z}_4$ for the square lattice, $\mathbb{Z}_6$ for triangular, $\mathbb{Z}_3$ for honeycomb), with a potential of the form $\mathcal{V}(V^k) = m^3_\chi \cos(4\chi/g) + \cdots$, where $k = 4, 6, 3...$ depends on the lattice, which has k minima, corresponding to the k possible VBS states. By (14.27), such a potential has charge k under J_F.

Consider this phase from the point of view of the gauge theory now. Notice that χ is the same (up to a factor) dual variable we introduced in our discussion of compact QED, and the Wilson loop will again produce an area law if χ is massive, as with the Polyakov effect.

In order for this story to make sense, we need that $M, g^2 \ll \frac{1}{a^2}$, so that χ is actually a low-energy degree of freedom. The idea is that the critical point from tuning J/Q to the critical value is reached by taking $m_\chi \to 0$. What is the nature of this critical theory? It has emergent deconfined gauge fields, even though the phases on either side of the critical point do not (they are confined $m > 0$ and Higgsed $m < 0$ respectively). Hence the name deconfined quantum criticality.

The conjecture (which would explain the phase diagram above) is that this gauge theory is a critical theory (in fact a conformal field theory) with only one relevant operator (the one which tunes us through the phase transition, the mass for χ) which is a singlet under all the symmetries. Recall that $e^{\mathbf{i}k\chi}$ has charge k under the J_F symmetry, and the square lattice preserves a $\mathbb{Z}_4 \subset \mathsf{U}(1)$ subgroup, so only allows the 4-vortex-insertion operator $e^{\mathbf{i}4\chi}$. What is the dimension of this operator? The conjecture is that it has dimension larger than 3.

[End of Lecture 60]

Insanely brief sketch of a check at large N. Actually, this can be checked very explicitly in a large-N version of the model, with N component z fields, so that the spin is $\phi^A = z^\dagger T^A z, A = 1..N^2 - 1$. This has $\mathsf{SU}(N)$ symmetry. When $m^2 < 0$, it is broken to $\mathsf{SU}(N-1)$, with $2(N-1)$ goldstone bosons. (Actually there is a generalization of the lattice model which realizes this – just make the spins into $N \times N$ matrices.)

Introducing an H-S field σ to decouple the $|z|^4$ interaction, we can make the z integrals gaussian, and find (this calculation is just like our earlier analysis in §11.3.4)

$$S[A,\sigma] = \int đp \left(\frac{1}{4} F_{\mu\nu}(p) \left(\frac{1}{g_{UV}^2} + \frac{c_1 N}{\mathbf{i}p} \log \frac{2m + \mathbf{i}p}{2m - \mathbf{i}p} \right) F^{\mu\nu}(-p) + \sigma(p) \left(-\frac{1}{\lambda} + \frac{c_2 N}{\mathbf{i}p} \log \frac{2m + \mathbf{i}p}{2m - \mathbf{i}p} \right) \sigma(-p) \right)$$

In the IR limit, $m \ll p \ll g_{UV}^2 N, \lambda N$, this is a scale-invariant theory with $\langle FF \rangle \sim p, \langle \sigma\sigma \rangle \sim p$ so that both F and σ have dimension near 2. (Actually the dimension of F is fixed at 2 by flux conservation.) z doesn't get any anomalous dimension at leading order in N.

This is all consistent with the claim so far. What is the dimension of $V_4 = e^{\mathbf{i}4\chi}$? To answer this question, we use a powerful tool of conformal field theory called *radial quantization*. Consider the theory on a cylinder, $S^2 \times \mathbb{R}$, where the last factor we can interpret as time. In a conformal field theory there is a one-to-one map between local operators and states of the theory on $S^d \times \mathbb{R}$. The state corresponding to an operator $\mathcal{O}$ is just $\mathcal{O}(0)\left|0\right\rangle$. The energy of the state on the sphere is the scaling dimension of the operator. (For an explanation of this, I refer to §4 of these notes.)

The state created by acting with $V_k(0)$ on the vacuum maps by this transformation to an initial state with flux k spread over the sphere (think of it as the 2-sphere surrounding the origin in spacetime): this state has charge k under $Q_F = \int_{S^2} J_F^0 = \int_{S^2} F_{12}$. The dimension of V_k is the energy of the lowest-energy state with $Q_F = k$. We can compute this by euclidean-time path integral:

$$Z_k = \mathrm{tr}_{Q_F = k} e^{-TH_{\text{cyl}}} \overset{T\to\infty}{\to} e^{-T\Delta_k}.$$

This is

$$Z_k = \int [dA] \delta \left(\int F - k \right) \int [dz dz^\dagger] e^{-S[z,A]} \equiv e^{-F_k}$$

which at large-N we can do by saddle point. The dominant configuration of the gauge field is the charge-k magnetic monopole $\underline{A}_\varphi = \frac{k}{2}(1-\cos\varphi)$, and we must compute

$$\int [d^2 z] e^{z^\dagger\left(-D^\dagger_{\underline{A}} D_{\underline{A}} + m^2\right) z} = \det\left(-D^\dagger_{\underline{A}} D_{\underline{A}} + m^2\right)^{-N/2} = e^{-\frac{N}{2}\operatorname{tr}\log\left(-D^\dagger_{\underline{A}} D_{\underline{A}} + m^2\right)}$$

The free energy is then a sum over eigenstates of this operator

$$\left(-\partial_\tau^2 - \vec{D}^2_{\underline{A}}\right) f_\ell e^{\mathbf{i}\omega\tau} = \left(\omega^2 + \lambda_\ell(k)\right) f_\ell e^{\mathbf{i}\omega\tau}$$

$$F_k = NT \int đ\omega \sum_\ell (2\ell+1)\log(\omega^2 + \lambda_\ell(k) + m^2).$$

The difference $F_k - F_0$ is UV finite and gives $\Delta_k = N c_k,\ c_1 \sim .12, c_4 \sim .82$. Unitary requires $\Delta_1 \geq \frac{1}{2}$ (= the free scalar dimension), so don't trust this for $N < 4$.

Pure field theory description. We've been discussing a theory with $\mathsf{U}(1)_{\text{VBS}} \times \mathsf{SU}(2)_{\text{spin}}$ symmetry. Lattice details aside, how can we encode the way these two symmetries are mixed up which forces the order parameter of one to be the disorder operator for the other? To answer this, briefly consider enlarging the symmetry to $\mathsf{SO}(5) \subset \mathsf{U}(1)_{\text{VBS}} \times \mathsf{SU}(2)_{\text{spin}}$, and organize $(\operatorname{Re} V, \operatorname{Im} V, n^1, n^2, n^3) \equiv n^a$ into a 5-component mega-voltron-spin vector. We saw that in $D = 0+1$, we could make a WZW term with a 3-component spin

$$\mathcal{W}_0[(n^1, n^2, n^3)] = \int_{B_2} \epsilon^{abc} n^a dn^b \wedge dn^c.$$

Its point in life was to impose the spin commutation relations at spin s when the coefficient is $2s$. In $D = 1+1$, we can make a WZW term with a 4-component spin, which can have $\mathsf{SO}(4)$ symmetry

$$\mathcal{W}_1[(n^1, n^2, n^3, n^4)] = \int_{B_3} \epsilon^{abcd} n^a dn^b \wedge dn^c \wedge dn^d.$$

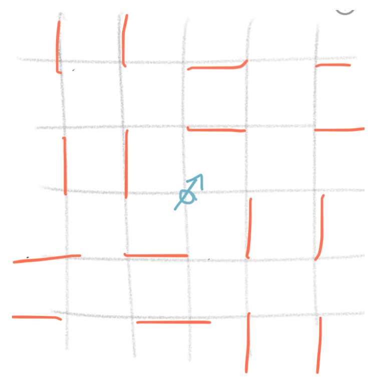
[34] Once we've got this far, how can you resist considering

$$\mathcal{W}_2[(n^1, n^2, n^3, n^4, n^5)] = \int_{B_4} \epsilon^{abcde} n^a dn^b \wedge dn^c \wedge dn^d \wedge dn^e.$$

[34]In fact the $D = 1+1$ version of this is extremely interesting. A few brief comments: (1) involves a *real* VBS order parameter n^4.) (2) The $D = 1+1$ term has the same number of derivatives (in the EOM) as the kinetic term $\partial n^a \partial n^a$. This means they can compete at a fixed point. The resulting CFTs are called WZW models. (3) The above is in fact a description of the spin-half chain, which previously we've described by an $\mathsf{O}(3)$ sigma model at $\theta = \pi$.

What does this do? Break the $\mathsf{SO}(5) \to \mathsf{U}(1) \times \mathsf{SU}(2)$ and consider a vortex configuration of V at $x^2 = x^3 = 0$. Suppose our action contains the term $k\mathcal{W}_2[n]$ with $k = 1$. Evaluate this in the presence of the vortex:

$$k\mathcal{W}_2[(n^1, n^2, n^3, n^4, n^5)|_{\text{vortex of } n^1 + \mathbf{i}n^2 \text{ at } x^2 = x^3 = 0}] = \int_{B_2|x^2=x^3=0} \epsilon^{abc} n^a dn^b \wedge dn^c = k\mathcal{W}_0[(n^1, n^2, n^3)].$$



This says the remaining three components satisfy the spin-half commutation relations: there is a spin in the core of the vortex, just as in the lattice picture at right.

15 Effective field theory

[Some nice lecture notes on effective field theory can be found here: J. Polchinski, A. Manohar, D. B. Kaplan, H. Georgi.]

Diatribe about 'renormalizability'. Having internalized Wilson's perspective on renormalization – namely that we should include all possible operators consistent with symmetries and let the dynamics decide which are important at low energies – we are led immediately to the idea of an *effective field theory* (EFT). There is no reason to demand that a field theory that we have found to be relevant for physics in some regime should be a valid description of the world to arbitrarily short (or long!) distances. This is a happy statement: there can always be new physics that has been so far hidden from us. Rather, an EFT comes with a regime of validity, and with necessary cutoffs. As we will discuss, in a useful implementation of an EFT, the cutoff implies a small parameter in which we can expand (and hence compute).

Caring about renormalizibility is pretending to know about physics at arbitrarily short distances. Which you don't.

Even when theories are renormalizable, this apparent victory is often false. For example, QED requires only two independent counterterms (mass and charge of the electron), and is therefore by the old-fashioned definition renormalizable, but it is superseded by the electroweak theory above 80GeV. Also: the coupling in QED actually increases logarithmically at shorter distances, and ultimately reaches a *Landau pole* at SOME RIDICULOUSLY HIGH ENERGY (of order $e^{+\frac{c}{\alpha}}$ where $\alpha \sim \frac{1}{137}$ is the fine structure constant (*e.g.* at the scale of atomic physics) and c is some numerical number. Plugging in numbers gives something like 10^{330} GeV, which is quite a bit larger than the Planck scale). This is of course completely irrelevant for physics and even in principle because of the previous remark about electroweak unification. And if not because of that, because of the Planck scale. A heartbreaking historical fact is that Landau and many other smart people gave up on QFT as a whole because of this silly fantasy about QED in an unphysical regime.

We will see below that even in QFTs which are non-renormalizable in the strict sense, there is a more useful notion of renormalizability: effective field theories come with a parameter (often some ratio of mass scales), in which we may expand the action. A useful EFT requires a finite number of counterterms at each order in the expansion.

Furthermore, I claim that this is *always* the definition of renormalizability that we are using, even if we are using a theory which is renormalizable in the traditional sense, which allows us to pretend that there is no cutoff. That is, there could always be corrections of order $\left(\frac{E}{E_{\text{new}}}\right)^n$ where E is some energy scale of physics that we are doing and E_{new} is some UV scale where new physics might come in; for large enough n, this is too small for us to have seen. The property of renormalizibility that actually matters is that we need a finite number of counterterms at each order in the expansion in $\frac{E}{E_{\text{new}}}$.

Renormalizable QFTs are in some sense *less* powerful than non-renormalizable ones – the latter have the decency to tell us when they are giving the wrong answer! That is, they tell us at what energy new physics *must* come in; with a renormalizable theory we may blithely pretend that it is valid in some ridiculously inappropriate regime like 10^{330} GeV.

Notions of EFT. There is a dichotomy in the way EFTs are used. Sometimes one knows a lot about the UV theory (*e.g.*

- electroweak gauge theory,
- QCD,
- electrons in a solid,
- water molecules

...) but it is complicated and unwieldy for the questions one wants to answer, so instead one develops an effective field theory involving just the appropriate and important dofs (*e.g.*, respectively,

- Fermi theory of weak interactions,
- chiral lagrangian (or HQET or SCET or ...),
- Landau Fermi liquid theory (or the Hubbard model or a topological field theory or ...),
- hydrodynamics (or some theory of phonons in ice or ...)

...). As you can see from the preceding lists of examples, even a single UV theory can have many different IR EFTs depending on what phase it is in, and depending on what question one wants to ask. The relationship between the pairs of theories above is always coarse-graining from the UV to the IR, though exactly what plays the role of the RG parameter can vary wildly. For example, in the example of the Fermi liquid theory, the scaling is $\omega \to 0$, and momenta scale towards the Fermi surface, not $\vec{k} = 0$.

A second situation is when one knows a description of some low-energy physics up to some UV scale, and wants to try to infer what the UV theory might be. This is a common situation in physics! Prominent examples include: the Standard Model, and quantized Einstein gravity. Occasionally we (humans) actually learn some physics and an example of an EFT from the second category moves to the first category.

Summary of basic EFT logic. Answer the following questions:

1. what are the dofs?
2. what are the symmetries?
3. where is the cutoff on its validity?

Then write down all interactions between the dofs which preserve the symmetry in an expansion in derivatives, with higher-dimension operators suppressed by more powers of the UV scale.

I must also emphasize two distinct usages of the term 'effective field theory' which are common, and which the discussion above is guilty of conflating (this (often slippery) distinction is emphasized in the review article by Georgi linked at the beginning of this subsection). The Wilsonian perspective advocated above produces a low-energy description of the physics which is really just a way of solving (if you can) the original model; very reductively, it's just a physically well-motivated order for doing the integrals. If you really integrate out the high energy modes exactly, you will get a non-local action for the low energy modes. This is to be contrasted with the local actions one uses in practice, by truncating the derivative expansion. It is the latter which is really the action of the effective field theory, as opposed to the *full theory*, with some of the integrals done already. The latter will give correct answers for physics below the cutoff scale, and it will give them much more easily.

Some interesting and/or important examples of EFT that we will not discuss explicitly, and where you can learn about them:

- Hydrodynamics [Kovtun]
- Fermi liquid theory [J. Polchinski, R. Shankar, *Rev. Mod. Phys.* **66** (1994) 129]
- chiral perturbation theory [D. B. Kaplan, §4]
- heavy quark effective field theory [D. B. Kaplan, §1.3]
- random surface growth (KPZ) [Zee, chapter VI]

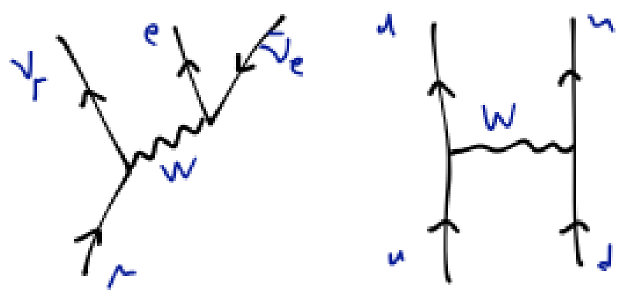
- color superconductors [D. B. Kaplan, §5]
- gravitational radiation [Goldberger, Rothstein]
- soft collinear effective theory [Becher, Stewart]
- magnets [Zee, chapter VI.5, hep-ph/9311264v1]
- effective field theory of cosmological inflation [Senatore et al, Cheung et al]
- effective field theory of dark matter direct detection [Fitzpatrick et al]

There are many others, the length of this list was limited by how long I was willing to spend digging up references. Here is a longer list.

15.1 Fermi theory of Weak Interactions

[from §5 of A. Manohar's EFT lectures] As a first example, let's think about part of the Standard Model.

$$L_{EW} \ni -\frac{\mathbf{i}g}{\sqrt{2}}\bar{\psi}_i\gamma^\mu P_L\psi_j W_\mu V_{ij} \quad + \text{terms involving } Z \text{ bosons} \tag{15.1}$$



Some things intermediate Ws can do: μ decay, $\Delta S = 1$ processes, neutron decay

If we are asking questions with external momenta less than M_W, we can integrate out W and make our lives simpler:

$$\delta S_{eff} \sim \left(\frac{\mathbf{i}g}{\sqrt{2}}\right)^2 V_{ij}V^\star_{k\ell}\int đ^D p \frac{-\mathbf{i}g_{\mu\nu}}{p^2 - M_W^2}\left(\bar{\psi}_i\gamma^\mu P_L\psi_j\right)(p)\left(\bar{\psi}_k\gamma^\nu P_L\psi_\ell\right)(-p)$$

(I am lying a little bit about the W propagator in that I am not explicitly projecting out the fourth polarization with the negative residue. Also hidden in my notation is the fact that the W carries electric charge, so the charges of $\bar{\psi}_i$ and ψ_j in (15.1) must differ by one.) This is non-local at scales $p \gtrsim M_W$ (recall our discussion in §8 (215B) with the two oscillators). But for $p^2 \ll M_W^2$,

$$\frac{1}{p^2 - M_W^2} \overset{p^2 \ll M_W^2}{\simeq} -\frac{1}{M_W^2}\left(1 + \underbrace{\frac{p^2}{M_W^2} + \frac{p^4}{M_W^4} + ...}_{\text{derivative couplings}}\right) \tag{15.2}$$

$$S_F = -\frac{4G_F}{\sqrt{2}} V_{ij} V_{kl}^\star \int d^4x \left(\bar\psi_i \gamma^\mu P_L \psi_j\right)(x) \left(\bar\psi_k \gamma_\mu P_L \psi_\ell\right)(x) + \mathcal{O}\left(\frac{1}{M_W^2}\right) + \text{kinetic terms for fermions} \tag{15.3}$$

where $G_F/\sqrt{2} \equiv \frac{g^2}{8M_W^2}$ is the Fermi coupling. We can use this (Fermi's) theory to compute the amplitudes above, and it is much simpler than the full electroweak theory (for example I don't have to lie about the form of the propagator of the W-boson like I did above).

On the other hand, this theory is *not the same* as the electroweak theory; for example it is not renormalizable, while the EW theory is. Its point in life is to help facilitate the expansion in $1/M_W$. There is something about the expression (15.3) that should make you nervous, namely the big red 1 in the $1/M_W^2$ corrections: what makes up the dimensions? This becomes an issue when we ask about ...

15.2 Loops in EFT

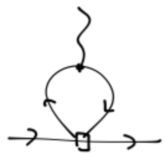
I skipped this subsection in lecture. Skip to §15.3. Suppose we try to define the Fermi theory S_F with a euclidean momentum cutoff $|k_E| < \Lambda$, like we've been using for most of our discussion so far. We expect that we'll have to set $\Lambda \sim M_W$. A simple example which shows that this is problematic is to ask about radiative corrections in the 4-Fermi theory to the coupling between the fermions and the Z (or the photon).

We are just trying to estimate the magnitude of this correction, so don't worry about the factors and the gamma matrices:

$$\sim I \equiv \underbrace{\frac{1}{M_W^2}}_{\propto G_F} \underbrace{\int^\Lambda đ^4k \frac{1}{k}\frac{1}{k} \text{tr}\left(\gamma ...\right)}_{\sim \int^\Lambda k dk \sim \Lambda^2 \sim M_W^2} \sim \mathcal{O}(1).$$

Even worse, consider what happens if we use the vertex coming from the $\left(\frac{p^2}{M_W^2}\right)^\ell$

correction in (15.2)



$$\sim I_\ell \equiv \frac{1}{M_W^2} \int^\Lambda \text{đ}^4 k \frac{1}{k^2} \left(\frac{k^2}{M_W^2} \right)^\ell \sim \mathcal{O}(1)$$

– it's also unsuppressed by powers of ... well, anything. This is a problem.

Fix: A way to fix this is to use a "mass-independent subtraction scheme", such as dimensional regularization and minimal subtraction ($\overline{\text{MS}}$). The crucial feature is that the dimensionful cutoff parameter appears only inside logarithms ($\log \mu$), and not as free-standing powers (μ^2).

With such a scheme, we'd get instead

$$I \sim \frac{m^2}{M_W^2} \log \mu \qquad I_\ell \sim \left(\frac{m^2}{M_W^2} \right)^{\ell+1} \log \mu$$

where m is some mass scale *other* than the RG scale μ (like a fermion mass parameter, or an external momentum, or a dynamical scale like Λ_{QCD}).

We will give a more detailed example next. The point is that in a mass-independent scheme, the regulator doesn't produce new dimensionful things that can cancel out the factors of M_W in the denominator. It respects the 'power counting': if you see 2ℓ powers of $1/M_W$ in the coefficient of some term in the action, that's how many powers will suppress its contributions to amplitudes. This means that the EFT is like a renormalizable theory *at each order in the expansion* (here in $1/M_W$), in that there is only a finite number of allowed vertices that contribute at each order (counterterms for which need to be fixed by a renormalization condition). The insatiable appetite for counterterms is still insatiable, but it eats only a finite number at each order in the expansion. Eventually you'll get to an order in the expansion that's too small to care about, at which point the EFT will have eaten only a finite number of counterterms.

There is a price for these wonderful features of mass-independent schemes, which has two aspects:

- Heavy particles (of mass m) don't decouple when $\mu < m$. For example, in a mass-independent scheme for a gauge theory, heavy charged particles contribute to the beta function for the gauge coupling even at $\mu \ll m$.

- Perturbation theory will break down at *low* energies, when $\mu < m$; in the example just mentioned this happens because the coupling keeps running.

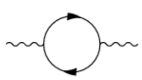
We will show both these properties very explicitly in the next subsection. The solution of both these problems is to integrate out the heavy particles by hand at $\mu = m$, and make a new EFT for $\mu < m$ which simply omits that field. Processes for which we should set $\mu < m$ don't have enough energy to make the heavy particles in external states anyway. (For some situations where you should still worry about them, see Aneesh Manohar's notes linked above.)

15.2.1 Comparison of schemes, case study

The case study we will make is the contribution of a charged fermion of mass m to the running of the QED gauge coupling.

Recall that the QED Lagrangian is

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \bar{\psi}\left(\mathbf{i}\not{D} - m\right)\psi$$

with $D_\mu = \partial_\mu - \mathbf{i}eA_\mu$. By redefining the field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ by a constant factor we can move around where the e appears, *i.e.* by writing $\tilde{A} = eA$, we can make the gauge kinetic term look like $\frac{1}{4e^2}\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu}$. This means that the charge renormalization can be seen *either* in the vacuum polarization, the correction to the photon propagator: . I will call this diagram $\mathbf{i}\Pi_{\mu\nu}$.

So the information about the running of the coupling is encoded in the gauge field two-point function:

$$\Pi_{\mu\nu} \equiv \langle A_\mu(p)A_\nu(q)\rangle = \left(p_\mu p_\nu - p^2 g_{\mu\nu}\right)\not{\delta}(p+q)\Pi(p^2)\ .$$

The factor $P_{\mu\nu} \equiv p_\mu p_\nu - p^2 g_{\mu\nu}$ is guaranteed to be the polarization structure by the gauge invariance Ward identity: $p^\mu \langle A_\mu(p)A_\nu(q)\rangle = 0$. That is: $p^\mu P_{\mu\nu} = 0$, and there is no other symmetric tensor made from p^μ which satisfies this. This determines the correlator up to a function of p^2, which we have called $\Pi(p^2)$.

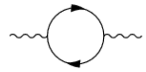
The choice of scheme shows up in our choice of renormalization condition to impose on $\Pi(p^2)$:

Mass-dependent scheme: subtract the value of the graph at $p^2 = -M^2$ (a very off-shell, euclidean, momentum). That is, we impose a renormalization condition which says

$$\Pi(p^2 = -M^2) \overset{!}{=} 1 \tag{15.4}$$

(which is the tree-level answer with the normalization above).

The contribution of a fermion of mass m and charge e is (factoring out the momentum-conserving delta function):

p,μ  p,ν

$$= -\int đ^D k \,\mathrm{tr}\left((-\mathbf{i}e\gamma^\mu) \frac{-\mathbf{i}\left(\not k + m\right)}{k^2 - m^2} (-\mathbf{i}e\gamma^\nu) \frac{-\mathbf{i}\left(\not p + \not k + m\right)}{(p+k)^2 - m^2} \right)$$

The minus sign out front is from the fermion loop. Some boiling, which you can find in Peskin (page 247) or Zee (§III.7), reduces this to something manageable. The steps involved are: (1) a trick to combine the denominators, like the Feynman trick $\frac{1}{AB} = \int_0^1 dx \left(\frac{1}{(1-x)A + xB}\right)^2$. (2) some Dirac algebra, to turn the numerator into a polynomial in k, p. As Zee says, our job in this course is not to train to be professional integrators. The result of this boiling can be written

$$\mathbf{i}\Pi^{\mu\nu} = -e^2 \int đ^D \ell \int_0^1 dx \frac{N^{\mu\nu}}{\left(\ell^2 - \Delta\right)^2}$$

with $\ell = k + xp$ is a new integration variable, $\Delta \equiv m^2 - x(1-x)p^2$, and the numerator is

$$N^{\mu\nu} = 2\ell^\mu \ell^\nu - g^{\mu\nu}\ell^2 - 2x(1-x)p^\mu p^\nu + g^{\mu\nu}\left(m^2 + x(1-x)p^2\right) + \text{terms linear in } \ell^\mu \quad .$$

In dim reg, the one-loop vacuum polarization correction satisfies the gauge invaraince Ward identity $\Pi^{\mu\nu} = P^{\mu\nu}\delta\Pi_2$ (unlike the euclidean momentum cutoff which is not gauge invariant). A peek at the tables of dim reg integrals shows that $\delta\Pi_2$ is:

$$\begin{aligned} \delta\Pi_2(p^2) &\overset{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2 - D/2)}{\Delta^{2-D/2}} \bar\mu^\epsilon \\ &\overset{D\to 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left(\frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right)\right) \end{aligned} \tag{15.5}$$

where we have introduced the heralded μ:

$$\mu^2 \equiv 4\pi\bar\mu^2 e^{-\gamma_E}$$

where γ_E is the Euler-Mascheroni constant. In the second line of (15.5), we expanded the Γ-function about $D = 4$; there are other singularities at other integer dimensions.

Mass-dependent scheme: Now back to our discussion of schemes. I remind you that in a mass-independent scheme, we demand that the counterterm cancels $\delta\Pi_2$ when we set the external momentum to $p^2 = -M^2$, so that the whole contribution at order e^2 is :

$$0 \overset{(15.4)!}{=} \Pi_2^{(M)}(p^2 = -M^2) = \underbrace{\delta_{F^2}^{(M)}}_{\text{counterterm coefficient for } \frac{1}{4}F_{\mu\nu}F^{\mu\nu}} + \delta\Pi_2$$

$$\implies \Pi_2^{(M)}(p^2) = \frac{e^2}{2\pi^2}\int dx x(1-x)\log\left(\frac{m^2 - x(1-x)p^2}{m^2 + x(1-x)M^2}\right).$$

Notice that the μs go away in this scheme.

Mass-Independent scheme: This is to be contrasted with what we get in a mass-independent scheme, such as $\overline{\text{MS}}$, in which Π is defined by the rule that we *subtract the* $1/\epsilon$ *pole*. This means that the counterterm is

$$\delta_{F^2}^{(\overline{\text{MS}})} = -\frac{e^2}{2\pi^2}\frac{2}{\epsilon}\underbrace{\int_0^1 dx x(1-x)}_{=1/6}.$$

(Confession: I don't know how to state this in terms of a simple renormalization condition on Π_2. Also: the bar in $\overline{\text{MS}}$ refers to the (not so important) distinction between $\bar{\mu}$ and μ.) The resulting vacuum polarization function is

$$\Pi_2^{(\overline{\text{MS}})}(p^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\log\left(\frac{m^2 - x(1-x)p^2}{\mu^2}\right).$$

Next we will talk about beta functions, and verify the claim above about the failure of decoupling. First let me say some words about what is failing. What is failing – the price we are paying for our power counting – is the basic principle of the RG, namely that physics at low energies shouldn't care about physics at high energies, except for small corrections to couplings. An informal version of this statement is: you don't need to know about nuclear physics to make toast. A more formal version is the *Appelquist-Carazzone Decoupling Theorem*, which I will not state (Phys. Rev. D11, 28565 (1975)). So it's something we must and will fix.

Beta functions. $\boxed{M}$: First in the mass-dependent scheme. Demanding that physics is independent of our made-up RG scale, we find

$$0 = M\frac{d}{dM}\Pi_2^{(M)}(p^2) = \left(M\frac{\partial}{\partial M} + \beta_e^{(M)} e\frac{\partial}{\partial e}\right)\Pi_2^{(M)}(p^2) = \left(M\frac{\partial}{\partial M} + \beta_e^{(M)}\underbrace{\cdot 2}_{\text{to this order}}\right)\Pi_2^{(M)}(p^2)$$

where I made the high-energy physics definition of the beta function[35]:

$$\beta_e^{(M)} \equiv \frac{1}{e}\left(M\partial_M e\right) = -\frac{\partial_\ell e}{e}, \quad M \equiv e^{-\ell}M_0 .$$

Here ℓ is the RG time again, it grows toward the IR. So we find

$$\beta_e^{(M)} = -\frac{1}{2}\left(\frac{e^2}{2\pi}\right)\int_0^1 dx x(1-x)\left(\frac{-2M^2 x(1-x)}{m^2 + M^2 x(1-x)}\right) + \mathcal{O}(e^3)$$

[35] I've defined these beta functions to be dimensionless, *i.e.* they are $\partial_{\log M}\log(g)$; this convention is not universally used.

$$
\begin{cases}
\stackrel{m\ll M}{\simeq} \frac{e^2}{2\pi^2}\int_0^1 dxx(1-x) = \frac{e^2}{12\pi^2} \\
\\
\stackrel{m\gg M}{\simeq} \frac{e^2}{2\pi^2}\int_0^1 dxx(1-x)\frac{M^2x(1-x)}{m^2} = \frac{e^2}{60\pi^2}\frac{M^2}{m^2}
\end{cases} . \tag{15.6}
$$

$$
\boxed{\overline{\text{MS}}}: \quad 0 = \mu\frac{d}{d\mu}\Pi_2^{(\overline{\text{MS}})}(p^2) = \left(\mu\frac{\partial}{\partial\mu} + \beta_e^{(\overline{\text{MS}})}e\frac{\partial}{\partial e}\right)\Pi_2^{(\overline{\text{MS}})}(p^2) = \left(\mu\frac{\partial}{\partial\mu} + \beta_e^{(\overline{\text{MS}})}\underbrace{\cdot 2}_{\text{to this order}}\right)\Pi_2^{(\overline{\text{MS}})}(p^2)
$$

$$
\begin{aligned}
\implies \beta_e^{(\overline{\text{MS}})} &= -\frac{1}{2}\frac{e^2}{2\pi^2}\underbrace{\int_0^1 dxx(1-x)}_{=1/6}\underbrace{\mu\partial_\mu \log\frac{m^2-p^2x(1-x)}{\mu^2}}_{=-2} \\
&= \frac{e^2}{12\pi^2}.
\end{aligned} \tag{15.7}
$$

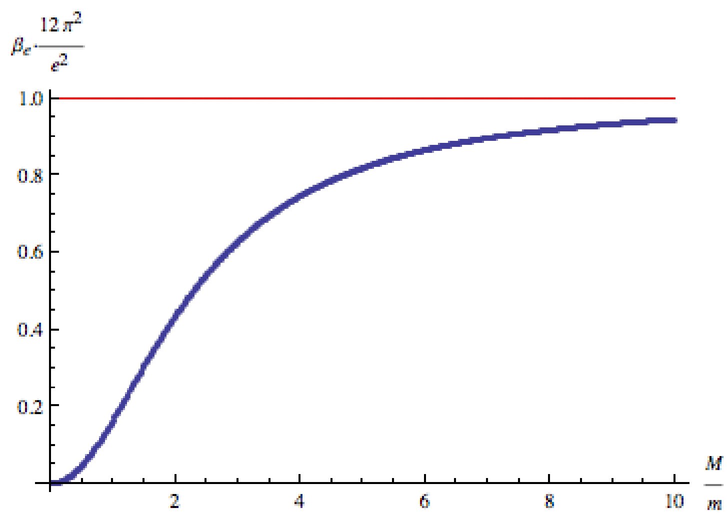
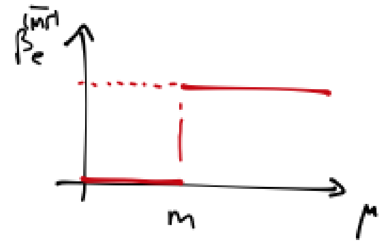


Figure 3: The blue curve is the mass-dependent-scheme beta function; at scales $M \ll m$, the mass of the heavy fermion, the fermion sensibly stops screening the charge. The red line is the $\overline{\text{MS}}$ beta function, which is just a constant, pinned at the UV value.

Also, the $\overline{\text{MS}}$ vacuum polarization behaves for small external momenta like

$$
\Pi_2(p^2 \ll m^2) \simeq -\frac{e^2}{2\pi^2}\int_0^1 dxx(1-x) \quad \underbrace{\log\frac{m^2}{\mu^2}}_{\gg 1,\text{for } \mu\ll m!\ \text{bad!}}
$$

As I mentioned, the resolution of both these problems is simply to define a new EFT for $\mu < m$ which omits the heavy field. Then the strong coupling problem goes away and the heavy fields do decouple. The price is that we have to do this by hand, and the beta function jumps at $\mu = m$; the coupling is continuous, though.



15.3 The Standard Model as an EFT.

The Standard Model. [Schwartz, §29]

	$L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	e_R	ν_R	$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	u_R	d_R	H
$\mathsf{SU}(3)$	-	-	-	$\square$	$\square$	$\square$	-
$\mathsf{SU}(2)$	$\square$	-	-	$\square$	-	-	$\square$
$\mathsf{U}(1)_Y$	$-\frac{1}{2}$	-1	0	$\frac{1}{6}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{2}$

Table 1: The Standard Model fields and their quantum numbers under the gauge group. $\square$ indicates fundamental representation, - indicates singlet. Except for the Higgs, each row is copied three times. Except for the Higgs all the other fields are Weyl fermions of the indicated handedness. Gauge fields as implied by the gauge groups. (Some people might leave out the right-handed neutrino, ν_R.)

Whence the values of the charges under the $\mathsf{U}(1)$ ("hypercharge")? The condition $Y_L + 3Y_Q = 0$ (where Y is the hypercharge) is required by anomaly cancellation. This implies that electrons and protons $p = \epsilon_{ijk} u_i u_j d_k$ have exactly opposite charges of the same magnitude.

The Lagrangian is just all the terms which are invariant under the gauge group $\mathsf{SU}(3) \times \mathsf{SU}(2) \times \mathsf{U}(1)$ with dimension less than or equal to four – all renormalizable terms. This includes a potential for the Higgs, $V(|H|) = m_H^2 |H|^2 + \lambda |H|^4$, where it turns out that $m_H^2 \leq 0$. The resulting Higgs vacuum expectation value breaks the Electroweak part of the gauge group

$$\mathsf{SU}(2) \times \mathsf{U}(1)_Y \overset{\langle H \rangle}{\rightsquigarrow} \mathsf{U}(1)_{EM}.$$

The broken gauge bosons get masses from the Higgs kinetic term

$$|D_\mu H|^2|_{H = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}} \quad \text{with} \quad D_\mu H = \left(\partial_\mu - \mathbf{i} g W_\mu^a \tau^a - \frac{1}{2} \mathbf{i} g' Y_\mu \right) H$$

where Y_μ is the hypercharge gauge boson, and $W^a, a = 1, 2, 3$ are the $\mathsf{SU}(2)$ gauge bosons. The photon and Z boson are

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} W^3_\mu \\ Y_\mu \end{pmatrix}.$$

There are also two massive W-bosons with electric charge ± 1.

Fermion masses come from Yukawa couplings

$$\mathcal{L}_{\text{Yukawa}} = -Y^\ell_{ij}\bar{L}_i H e^j_R - Y^u_{ij}\bar{Q}^i H d^j_R - Y^d_{ij}\bar{Q}^i \left(\mathbf{i}\tau^2 H^\star\right) u^j_R + h.c.$$

The contortion with the τ^2 is required to make a hypercharge invariant. Plugging in the Higgs vev to *e.g.* the lepton terms gives $-m_e\bar{e}_L e_R + h.c.$ with $m_e = y_e v/\sqrt{2}$. There's lots of drama about the matrices Y which can mix the generations. the mass for the ν_R (which maybe could not exist – it doesn't have any charges at all) you'll figure out on the homework.

Here is a useful mnemonic for remembering the table of quantum numbers (possibly it is more than that): There are larger simple Lie groups that contain the SM gauge group as subgroups:

$$\begin{array}{ccccc} \mathsf{SU}(3)\times\mathsf{SU}(2)\times\mathsf{U}(1)_Y & \subset & \mathsf{SU}(5) & \subset & \mathsf{SO}(10) \\ \text{one generation} & = & \mathbf{10}\oplus\bar{\mathbf{5}}\oplus\mathbf{1} & = & \mathbf{16} \end{array}$$

The singlet of $\mathsf{SU}(5)$ is the right-handed neutrino, but if we include it, one generation is an irreducible (spinor) representation of $\mathsf{SO}(10)$. This idea is called grand unification. It is easy to imagine that another instance of the Higgs mechanism accomplishes the breaking down to the Standard Model. Notice that this means leptons and quarks are in the same representations – they can turn into each other. This predicts that the proton should not be perfectly stable. Next we'll say more about this.

Beyond the Standard Model with EFT. At what energy does the Standard Model stop working? Because of the annoying feature of renormalizibility, it doesn't tell us. However, we have experimental evidence against a cutoff on the Standard Model (SM) at energies less than something like 10 TeV. The evidence I have in mind is the absence of interactions of the form

$$\delta L = \frac{1}{M^2}\left(\bar{\psi}A\psi\right)\cdot\left(\bar{\psi}B\psi\right)$$

(where ψ represent various SM fermion fields and A, B can be various gamma and flavor matrices) with $M \lesssim 10$ TeV. Notice that I am talking now about interactions *other* than the electroweak interactions, which as we've just discussed, for energies above $M_W \sim 80$GeV cannot be treated as contact interactions – you can see the Ws propagate!

If such operators were present, we would have found different answers for experiments at LEP. But such operators would be present if we consider new physics in addition to the Standard Model (in most ways of doing it) at energies less than 10 TeV. For example, many interesting ways of coupling in new particles with masses that make them accessible at the LHC would have generated such operators.

A little more explicitly: the Standard Model Lagrangian L_0 contains all the renormalizable (*i.e.* engineering dimension ≤ 4) operators that you can make from its fields (though the coefficients of the dimension 4 operators do vary through quite a large range, and the coefficients of the two relevant operators – namely the identity operator which has dimension zero, and the Higgs mass, which has engineering dimension two, are strangely small, and so is the QCD θ angle).

To understand what lies beyond the Standard Model, we can use our knowledge that whatever it is, it is probably heavy (it could also just be very weakly coupled, which is a different story), with some intrinsic scale Λ_{new}, so we can integrate it out and include its effects by corrections to the Standard Model:

$$L = L_0 + \frac{1}{\Lambda_{\text{new}}}\mathcal{O}^{(5)} + \frac{1}{\Lambda_{\text{new}}^2}\sum_i c_i \mathcal{O}_i^{(6)}$$

where the $\mathcal{O}$s are made of SM fields, and have the indicated engineering dimensions, and preserve the necessary symmetries of the SM.

In fact there is only one kind of operator of dimension 5:

$$\mathcal{O}^{(5)} = c_5 \epsilon_{ij} \left(\bar{L}^c\right)^i H^j \epsilon_{kl} L^k H^l$$

where $H^i = (h^+, h^0)^i$ is the $\mathrm{SU}(2)_{EW}$ Higgs doublet and $L^i = (\nu_L, e_L)^i$ is an $\mathrm{SU}(2)_{EW}$ doublet of left-handed leptons, and $\bar{L}^c \equiv L^T C$ where C is the charge conjugation matrix. (I say 'kind of operator' because we can have various flavor matrices in here.) On the problem set you get to see from whence such an operator might arise, and what it does if you plug in the higgs vev $\langle H \rangle = (0, v)$. This term violates lepton number.

At dimension 6, there are operators that directly violate baryon number, such as

$$\epsilon_{\alpha\beta\gamma} (\bar{u}_R)^c_\alpha (u_R)_\beta \, (\bar{u}_R)^c_\gamma \, e_R.$$

You should read the above tangle of symbols as '$qqq\ell$' – it turns three quarks into a lepton. The epsilon tensor makes a color SU(3) singlet; this thing has the quantum numbers of a baryon. The long lifetime of the proton (you can feel it in your bones – see Zee p. 413) then directly constrains the scale of new physics appearing in front of this operator.

Two more comments about this:

- If we didn't know about the Standard Model, (but after we knew about QM and GR and EFT (the last of which people didn't know before the SM for some reason)) we should have made the estimate that dimension-5 Planck-scale-suppressed operators like $\frac{1}{M_{\text{Planck}}} p\mathcal{O}$ would cause proton decay (into whatever $\mathcal{O}$ makes). This predicts $\Gamma_p \sim \frac{m_p^3}{M_{\text{Planck}}^2} \sim 10^{-13} s^{-1}$ which is *not* consistent with our bodies not glowing. Actually it is a remarkable fact that there are no gauge-invariant operators made of SM fields of dimension less than 6 that violate baryon number. This is an *emergent* symmetry, expected to be violated by the UV completion.
- Surely nothing can prevent $\Delta L \sim \left(\frac{1}{M_{\text{Planck}}}\right)^2 qqq\ell$. Happily, this is consistent with the observed proton lifetime.

There are $\sim 10^2$ dimension 6 operators that preserve baryon number, and therefore are not as tightly constrained[36] (Those that induce flavor-changing processes in the SM are more highly constrained and must have $\Lambda_{\text{new}} > 10^4$ TeV.) Two such operators are considered equivalent if they differ by something which vanishes by the tree-level SM equations of motion. This is the right thing to do, even for off-shell calculations (like green's functions and for fields running in loops). You know this from a previous problem set: the EOM are true as operator equations – Ward identities resulting from being free to change integration variables in the path integral[37].

15.4 Pions

[Schwartz §28.1] Below the scale of electroweak symmetry breaking, we can forget the W and Z bosons. Besides the 4-Fermi interactions, the remaining drama is QCD and electromagnetism:

$$\mathcal{L}_{QCD_2} = -\frac{1}{4}F_{\mu\nu}^2 + \mathbf{i} \sum_{\alpha=L,R} \sum_f \bar{q}_{\alpha f} \not{D} q_{\alpha f} - m\bar{q}Mq.$$

Here f is a sum over quark flavors, which includes the electroweak doublets, u and d. Let's focus on just these two lightest flavors, u and d. We can diagonalize the

[36]Recently, humans have gotten better at counting these operators. See this paper.

[37] There are a few meaningful subtleties here, as you might expect if you recall that the Ward identity is only true up to contact terms. The measure in the path integral can produce a Jacobian which renormalizes some of the couplings; the changes in source terms will drop out of S-matrix elements (recall our discussion of changing field variables in the Consequences of Unitarity section.) but can change the form of Green's functions. For more information on the use of eom to eliminate redundant operators in EFT, see Arzt, hep-ph/9304230 and Georgi, "On-Shell EFT".

mass matrix by a field redefinition (this is what makes the CKM matrix meaningful): $M = \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix}$. If it were the case that $m_u = m_d$, we would have *isospin* symmetry

$$\begin{pmatrix} u \\ d \end{pmatrix} \to U \begin{pmatrix} u \\ d \end{pmatrix}, \quad U \in \mathsf{SU}(N_f = 2).$$

If, further, there were no masses $m = 0$, then L and R decouple and we also have chiral symmetry, $q \to e^{\mathbf{i}\gamma_5 \alpha} q$, *i.e.*

$$q_L \to V q_L, q_R \to V^{-1} q_R,\ V \in \mathsf{SU}(N_f = 2).$$

Why do I restrict to $\mathsf{SU}(2)$ and not $\mathsf{U}(2)$? The central bit of the axial symmetry $\mathsf{U}(1)_A$ is anomalous – it's divergence is proportional to the gluon $F \wedge F$, which has all kinds of nonzero matrix elements. It's not a symmetry (see Peskin page 673 for more detail). The central bit of the vectorlike transformation $q \to e^{\mathbf{i}\alpha} q$ is *baryon number*, B. (Actually this is anomalous under the full electroweak symmetry, but $B - L$ is not).

The vacuum of QCD is mysterious, because of infrared slavery. Apparently it is the case that

$$\langle \bar{q}_f q_f \rangle = V^3$$

independent of flavor f. This condensate breaks

$$\mathsf{SU}(2)_L \times \mathsf{SU}(2)_R \to \mathsf{SU}(2)_{\text{isospin}}, \tag{15.8}$$

the diagonal combination. $\begin{pmatrix} u \\ d \end{pmatrix}$ is a doublet. Since $p = u_\alpha u_\beta d_\gamma \epsilon_{\alpha\beta\gamma}, n = u_\alpha d_\beta d_\gamma \epsilon_{\alpha\beta\gamma}$, this means that $\begin{pmatrix} p \\ n \end{pmatrix}$ is also a doublet. This symmetry is weakly broken by the difference of the masses $m_d = 4.7\text{MeV} \neq m_u = 2.15\text{MeV}$ and by the electromagnetic interactions, since $q_d = -1/3 \neq q_u = 2/3$.

This symmetry-breaking structure enormously constrains the dynamics of the color singlets which are the low-energy excitations above the QCD vacuum (hadrons). Let us use the EFT strategy. We know that the degrees of freedom must include (pseudo-)Goldstone bosons for the symmetry breaking (15.8) ('pseudo' because of the weak explicit breaking).

Effective field theory. Since QCD is strongly coupled in this regime, let's use the knowing-the-answer trick: the low energy theory must include some fields which represent the breaking of the symmetry (15.8). One way to do this is to introduce a field Σ which transforms like

$$\mathsf{SU}(2)_L \times \mathsf{SU}(2)_R : \Sigma \to g_L \Sigma g_R^\dagger, \quad \Sigma^\dagger \to g_R \Sigma^\dagger g_L^\dagger$$

(this will be called a *linear* sigma model, because Σ transforms linearly) and we can make singlets (hence an action) out of $|\Sigma|^2 = \Sigma_{ij}\Sigma^\dagger_{ji} = \text{tr}\Sigma\Sigma^\dagger$:

$$\mathcal{L} = |\partial_\mu \Sigma|^2 + m^2|\Sigma|^2 - \frac{\lambda}{4}|\Sigma|^4 + \cdots \tag{15.9}$$

which is designed to have a minimum at $\langle\Sigma\rangle = \frac{V}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$, with $V = 2m/\sqrt{\lambda}$, which preserves $\mathsf{SU}(2)_{\text{isospin}}$. We can parametrize the fluctuations about this configuration as

$$\Sigma(x) = \frac{V + \sigma(x)}{\sqrt{2}} e^{\frac{2\mathbf{i}\pi^a(x)\tau^a}{F_\pi}}$$

where F_π will be chosen to give $\pi^a(x)$ canonical kinetic terms. Under $g_{L/R} = e^{\mathbf{i}\theta_{L/R}\tau^a}$, the pion field transforms as

$$\pi^a \to \pi^a + \underbrace{\frac{F_\pi}{2}\left(\theta^a_L - \theta^a_R\right)}_{\text{nonlinear realization of } \mathsf{SU}(2)_{\text{axial}}} - \underbrace{\frac{1}{2} f^{abc}\left(\theta^a_L + \theta^a_R\right)\pi^c}_{\text{linear realiz'n (adj rep) of } \mathsf{SU}(2)_{\text{isospin}}} .$$

The fields $\pi^\pm, \pi^0$ create pions, they transform in the adjoint representation of the diagonal $\mathsf{SU}(2)_{\text{isospin}}$, and they *shift* under the broken symmetry. This shift symmetry forbids mass terms π^2. The radial excitation σ, on the other hand, is a fiction which we've introduced in (15.9), and which has no excuse to stick around at low energies (and does not). We can put it out of its misery by taking $m \to \infty, \lambda \to \infty$ fixing F_π. In the limit, the useful field to use is

$$U(x) \equiv \frac{\sqrt{2}}{V}\Sigma(x)|_{\sigma=0} = e^{\frac{2\mathbf{i}\pi^a\tau^a}{F_\pi}}$$

which is unitary $UU^\dagger = U^\dagger U = \mathbb{1}$. This last identity means that all terms in an action for U require derivatives, so (again) no mass for π. The most general Lagrangian for U can be written as an expansion in derivatives, and is called the *chiral Lagrangian*:

$$\mathcal{L}_\chi = \frac{F_\pi^2}{4}\text{tr}D_\mu U D^\mu U^\dagger + L_1 \text{tr}\left(D_\mu U D^\mu U^\dagger\right)^2 + L_2 \text{tr}D_\mu U D_\nu U^\dagger \text{tr}D^\nu U^\dagger D_\mu U + L_3 \text{tr}D_\mu U D^\mu U^\dagger D_\nu U D^\nu U^\dagger + \cdots \tag{15.10}$$

In terms of π, the leading terms are

$$L_\chi = \frac{1}{2}\partial_\mu\pi^a\partial^\mu\pi^a + \frac{1}{F_\pi^2}\left(-\frac{1}{3}\pi^0\pi^0 D_\mu\pi^+ D^\mu\pi^- + \cdots\right) + \frac{1}{F_\pi^4}\left(\frac{1}{18}\left(\pi^-\pi^+\right)^2 D_\mu\pi^0 D^\mu\pi^0 + \cdots\right)$$

This fixes the relative coefficients of many irrelevant interactions, all with two derivatives, suppressed by powers of F_π. The expansion of the L_i terms have four derivatives, and are therefore suppressed by further powers of E/F_π.

Pion masses. The pions aren't actually massless: $m_{\pi^\pm} \sim 140\text{MeV}$. In terms of quarks, one source for such a thing is the quark mass term $\mathcal{L} \ni \bar{q}Mq$. This breaks the isospin symmetry if the eigenvalues of M aren't equal. But an *invariance* of $\mathcal{L}$ is

$$q_{L/R} \to g_{L/R}q_{L/R}, \; M \to g_L M g_R^\dagger. \tag{15.11}$$

Think of M as a background field (such a thing is sometimes called a *spurion*). If M were an actual dynamical field, then (15.11) would be a symmetry. In the effective action which summarizes all the drama of strong-coupling QCD in terms of pions, the field M should still be there, and if we transform it as in (15.11), it should still be an invariance. Maybe we're going to do the path integral over M later. (This is the same strategy we used when deriving the vertical-tangents condition in the Bose-Hubbard phase diagram.)

So the chiral lagrangian $\mathcal{L}_\chi$ should depend on M and (15.11) should be an invariance. This determines

$$\Delta\mathcal{L}_\chi = \frac{V^3}{2}\text{tr}\left(MU + M^\dagger U^\dagger\right) + \cdots = V^3(m_u + m_d) - \frac{V^3}{2F_\pi^2}(m_u+m_d)\sum_a \pi_a^2 + \mathcal{O}(\pi^2).$$

The coefficient V^3 is chosen so that the first term matches $\langle \bar{q}Mq\rangle = V^3(m_u+m_d)$. The second term then gives

$$m_\pi^2 \simeq \frac{V^3}{F_\pi^2}(m_u+m_d)$$

which is called the Gell-Mann Oakes Renner relation.

Electroweak interactions. You may have noticed that I used covariant-looking Ds in (15.10). That's because the $\mathsf{SU}(2)_L$ symmetry we've been speaking about is actually gauged by W_μ^a. (The electroweak gauge boson kinetic terms are in the $\cdots$ of (15.10).) Recall that

$$\mathcal{L}_{\text{Weak}} \ni \frac{g}{2}W_\mu^a\left(\underbrace{J_\mu^a - J_\mu^{5a}}_{\text{`V' - `A'}}\right) = W_\mu^a\left(V_{ij}\bar{Q}_i\gamma^\mu(1-\gamma^5)\tau^a Q_j + \bar{L}_i\gamma^\mu\tau^a(1-\gamma^5)L_i\right)$$

where $Q_1 = \begin{pmatrix} u \\ d\end{pmatrix}, L_1 = \begin{pmatrix} e \\ \nu_e\end{pmatrix}$ are doublets of $\mathsf{SU}(2)_L$.

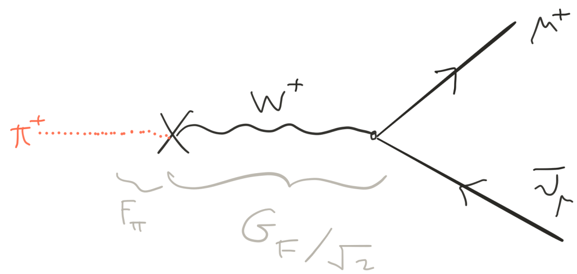
Now, in equations, the statement "a pion is a Goldstone boson for the axial $\mathsf{SU}(2)$" is:

$$\langle 0|\, J_\mu^{5a}(x)\left|\pi^b(p)\right\rangle = \mathbf{i}p_\mu F_\pi e^{-\mathbf{i}p\cdot x}\delta^{ab}.$$

If the vacuum were invariant under the symmetry transformation generated by J_μ, the BHS would vanish. The momentum dependence implements the fact that a global rotation does not change the energy. Contracting the BHS with p^μ and using current conservation gives $0 = p^2 F_\pi^2 = m_\pi^2 F_\pi^2$, a massless dispersion for the pions.

Combining the previous two paragraphs, we see that the following process can happen

$$\pi \overset{\text{Goldstone}}{\to} J_\mu^5 \overset{\text{electroweak interaction}}{\to} \text{leptons}$$



(15.12)

and in fact is responsible for the dominant decay channel of charged pions. (Time goes from left to right in these diagrams, sorry.)

$$\mathcal{M}(\pi^+ \to \mu^+ \nu_\mu) = \frac{G_F}{\sqrt{2}} F_\pi p^\mu \bar{v}_{\nu_\mu} \gamma^\mu (1 - \gamma^5) u_\mu$$

where the Fermi constant $G_F \sim 10^{-5} GeV^{-2}$ (known from *e.g.* $\mu^- \to e^- \bar{\nu}_e \nu_\mu$) is a good way to parametrize the Weak interaction amplitude. Squaring this and integrating over two-body phase space gives the decay rate

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

(You can see from the answer why the decay to muons is more important than the decay to electrons, since $m_\mu / m_e \sim 200$. This is called *helicity suppression* – the decay of the helicity-zero π^- into back-to-back spin-half particles by the weak interaction (which only produces L particles and R antiparticles) can't happen if helicity is conserved – the mass term is required to flip the e_L into an e_R.) This contributes most of $\tau_{\pi^+} = \Gamma^{-1} = 2.6 \cdot 10^{-8} s$.

Knowing further the mass of the muon $m_\mu = 106$MeV then determines $F_\pi = 92$MeV which fixes the leading terms in the chiral Lagrangian. This is why F_π is called the *pion decay constant*. This gives a huge set of predictions for *e.g.* pion scattering $\pi^0 \pi^0 \to \pi^+ \pi^-$.

Note that the neutral pion can decay by an anomaly into two photons:

$$q_\mu \langle p, k| J_\mu^{5, a=3}(q) |0\rangle = -\frac{e^2}{4\pi^2} \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta$$

where $\langle p, k|$ is a state with two photons, and this is a matrix element of the $J_e J_e J_{\text{isospin}}$ anomaly,

$$\partial_\mu J^{\mu 5a} = -\frac{e^2}{16\pi^2}\epsilon^{\nu\lambda\alpha\beta}F_{\nu\lambda}F_{\alpha\beta}\text{tr}\left(\tau^a Q^2\right)$$

where $Q = \begin{pmatrix} 2/3 & 0 \\ 0 & -1/3 \end{pmatrix}$ is the quark charge matrix.

SU(3) and baryons. The strange quark mass is also pretty small $m_s \sim 95\text{MeV}$, and $\langle \bar{s}s\rangle \sim V^3$. This means the approximate invariance and symmetry breaking pattern is actually $\mathsf{SU}(3)_L \times \mathsf{SU}(3)_R \to \mathsf{SU}(3)_{\text{diag}}$, meaning that there are $16 - 8 = 8$ pseudo NGBs. Besides $\pi^{\pm,0}$, the others are the kaons $K^{\pm,0}$ and η. It's still only the $\mathsf{SU}(2)_L$ that's gauged.

We can also include baryons $B = \epsilon_{\alpha\beta\gamma}q_\alpha q_\beta q_\gamma$. Since $q \in 3$ of $\mathsf{SU}(3)$, the baryons are in the representation

$$\begin{aligned} 3 \otimes 3 \otimes 3 \ &= (6 \oplus \bar{3}) \otimes 3 \ &= 10 \oplus 8 \oplus 8 \oplus 1 \\ \ydiagram{1} \otimes \ydiagram{1} \otimes \ydiagram{1} &= \left(\ydiagram{2} \oplus \ydiagram{1,1}\right) \otimes \ydiagram{1} &= \ydiagram{3} \oplus \ydiagram{2,1} \oplus \ydiagram{2,1} \oplus \ydiagram{1,1,1} \end{aligned} \tag{15.13}$$

The proton and neutron are in one of the octets. This point of view brought some order (and some predictions) to the otherwise-bewildering zoo of hadrons.

Returning to the two-flavor $\mathsf{SU}(2)$ approximation, We can include the nucleons $N_{L/R} = \begin{pmatrix} p \\ n \end{pmatrix}_{L/R}$ and couple them to pions by the symmetric coupling

$$\mathcal{L} \ni \lambda_{NN\pi}\bar{N}_L \Sigma N_R.$$

The expectation value for Σ gives a nucleon mass: $m_N = \lambda_{NN\pi}F_\pi$. This is a cheap version of the *Goldberger-Treiman relation*; for a better one see Peskin pp. 670-672.

WZW terms in the chiral Lagrangian. Finally, I would be remiss not to mention that the chiral Lagrangian must be supplemented by WZW terms to have the right realization of symmetries (in order to encode all the effects of anomalies, and in order to violate $\pi \to -\pi$ which is not a symmetry of QCD). This is where those terms were first discovered, and where it was realized that their coefficients are quantized. In particular the coefficient of the WZW term $W_4[U]$ here is N_c, the number of colors, as Witten shows by explicitly coupling to electromagnetism, and finding the term that encodes $\pi^0 \to \gamma\gamma$. One dramatic consequence here is that the chiral Lagrangian (with some higher-derivative terms) has a topological soliton solution (the skyrmion) which is a *fermion* if the number of colors of QCD is odd. It is a fermion because the WZW term evaluates to π on a spacetime trajectory where the soliton makes a 2π rotation. The baryon number of this configuration comes from the anomalous (WZW) contribution to the baryon number current $B_\mu = \frac{\epsilon_{\mu\nu\alpha\beta}}{24\pi^2}\text{tr}U^{-1}\partial_\nu U U^{-1}\partial_\alpha U U^{-1}\partial_\beta U$ whose conserved charge $\int_{\text{space}} B_0$ is the winding number of the map from space (plus the point at infinity) to the space of goldstones $S^3 \to \mathsf{SU}(3) \times \mathsf{SU}(3)/\mathsf{SU}(3)_{\text{preserved}} \simeq \mathsf{SU}(3)_{\text{broken}}$.

[End of Lecture 61]

15.5 Quantum Rayleigh scattering

We didn't have lecture time for the remaining sections, but you might still enjoy them.

[from hep-ph/9606222 and nucl-th/0510023] Why is the sky blue? Basically, it's because the blue light from the sun scatters in the atmosphere more than the red light, and you (I hope) only look at the scattered light.

Here is an understanding of this fact using the EFT logic. Consider the scattering of photons off atoms at low energies. Low energy means that the photon does not have enough energy to probe the substructure of the atom – it can't excite the electrons or the nuclei. This means that the atom is just a particle, with some mass M.

The dofs are just the photon field and the field that creates an atom.

The symmetries are Lorentz invariance and charge conjugation invariance and parity. We'll use the usual redundant description of the photon which has also gauge invariance.

The cutoff is the energy ΔE that it takes to excite atomic energy levels we've left out of the discussion. We allow no inelastic scattering. This means we require

$$E_\gamma \ll \Delta E \sim \frac{\alpha}{a_0} \ll a_0^{-1} \ll M_{\text{atom}}$$

Because of this separation of scales, we can also ignore the recoil of the atom, and treat it as infinitely heavy.

Since there are no charged objects in sight – atoms are neutral – gauge invariance means the Lagrangian can depend on the field strength $F_{\mu\nu}$. Let's call the field which destroys an atom with velocity v ϕ_v. $v^\mu v_\mu = 1$ and $v_\mu = (1,0,0,0)_\mu$ in the atom's rest frame. The Lagrangian can depend on v^μ. We can write a Lagrangian for the free atoms as

$$L_{\text{atom}} = \phi_v^\dagger \mathbf{i} v^\mu \partial_\mu \phi_v \quad .$$

This action is related by a boost to the statement that the atom at rest has zero energy – in the rest frame of the atom, the eom is just $\partial_t \phi_{v=(1,\vec{0})} = 0$.

So the Lagrangian density is

$$L_{\text{Maxwell}}[A] + L_{\text{atom}}[\phi_v] + L_{\text{int}}[A, \phi_v]$$

and we must determine L_{int}. It is made from local, Hermitian, gauge-invariant, Lorentz invariant operators we can construct out of $\phi_v, F_{\mu\nu}, v_\mu, \partial_\mu$ (It can only depend on $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and not A_μ directly, by gauge invariance.). It should actually only depend on the combination $\phi_v^\dagger \phi_v$ since we will not create and destroy atoms. Therefore

$$L_{\text{int}} = c_1 \phi_v^\dagger \phi_v F_{\mu\nu} F^{\mu\nu} + c_2 \phi_v^\dagger \phi_v v^\sigma F_{\sigma\mu} v_\lambda F^{\lambda\mu} + c_3 \phi_v^\dagger \phi_v \left(v^\lambda \partial_\lambda\right) F_{\mu\nu} F^{\mu\nu} + \dots$$

$\dots$ indicates terms with more derivatives and more powers of velocity (*i.e.* an expansion in $\partial \cdot v$). Which are the most important terms at low energies? Demanding that the Maxwell term dominate, we get the power counting rules (so time and space should scale the same way):

$$[\partial_\mu] = 1, \quad [F_{\mu\nu}] = 2$$

This then implies $[\phi_v] = 3/2, [v] = 0$ and therefore

$$[c_1] = [c_2] = -3, [c_3] = -4 \ .$$

Terms with more partials are more irrelevant.

What makes up these dimensions? They must come from the length scales that we have integrated out to get this description – the size of the atom $a_0 \sim \alpha m_e$ and the energy gap between the ground state and the electronic excited states $\Delta E \sim \alpha^2 m_e$. For $E_\gamma \ll \Delta E, a_0^{-1}$, we can just keep the two leading terms.

In the rest frame of the atom, these two leading terms $c_{1,2}$ represent just the scattering of E and B respectively. To determine their coefficients one would have to do a matching calculation to a more complete theory (compute transition rates in a theory that does include extra energy levels of the atom). But a reasonable guess is just that the scale of new physics (in this case atomic physics) makes up the dimensions: $c_1 \simeq c_2 \simeq a_0^3$. (In fact the magnetic term c_2 comes with extra factor of v/c which suppresses it.) The scattering cross section then goes like $\sigma \sim c_i^2 \sim a_0^6$; dimensional analysis ($[\sigma] = -2$ is an area, $[a_0^6] = -6$) then tells us that we have to make up four powers with the only other scale around:

$$\sigma \propto E_\gamma^4 a_0^6.$$

(The factor of E_γ^2 in the amplitude arises from $\vec{E} \propto \partial_t \vec{A}$.) Blue light, which has about twice the energy of red light, is therefore scattered 16 times as much.

The leading term that we left out is the one with coefficient c_3. The size of this coefficient determines when our approximations break down. We might expect this to come from the next smallest of our neglected scales, namely ΔE. That is, we expect

$$\sigma \propto E_\gamma^4 a_0^6 \left(1 + \mathcal{O}\left(\frac{E_\gamma}{\Delta E}\right)\right).$$

The ratio in the correction terms is appreciable for UV light.

15.6 Superconductors

Recall from 215B our effective (Landau-Ginzburg) description of superconductors which reproduces the Meissner effect, the Abelian Higgs model:

$$\mathcal{F} = \frac{1}{4}F_{ij}F_{ij} + |D_i\Phi|^2 + a|\Phi|^2 + \frac{1}{2}b|\Phi|^4 + ... \tag{15.14}$$

with $D_i\Phi \equiv (\partial_i - 2e\mathbf{i}A_i)\,\Phi$.

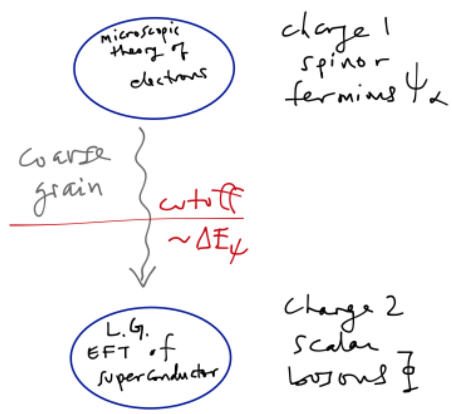
I want to make two more comments about this:

Symmetry breaking by fluctuations (Coleman-Weinberg) revisited. [Zee problem IV.6.9.] What happens near the transition, when $a = 0$ in (15.14)? Quantum fluctuations can lead to symmetry breaking. This is just the kind of question we discussed earlier, when we introduced the effective potential. Here it turns out that we can trust the answer (roughly because in this scalar electrodynamics, there are two couplings: e and the quartic self-coupling b).

New IR dofs. A feature of this example that I want you to notice: the microscopic description of real superconductor involves electrons – charge $1e$ spinor fermions, created by some fermionic operator ψ_α, $\alpha =\uparrow, \downarrow$.

We are describing the low-energy physics of a system of electrons in terms of a bosonic field, which (in simple 's-wave' superconductors) is roughly related to the electron field by

$$\Phi \sim \psi_\alpha \psi_\beta \epsilon^{\alpha\beta} \;; \tag{15.15}$$



Φ is called a Cooper pair field. At least, the charges and the spins and the statistics work out. The details of this relationship are not the important point I wanted to emphasize. Rather I wanted to emphasize the dramatic difference in the correct choice of variables between the UV description (spinor fermions) and the IR description (scalar bosons). One reason that this is possible is that it costs a large energy to make a fermionic excitation of the superconductor. This can be understood roughly as follows: The microscopic theory of the electrons looks something like

$$S[\psi] = S_2[\psi] + \int \mathrm{d}t\mathrm{d}^dx\; u\psi^\dagger\psi\psi^\dagger\psi + h.c. \tag{15.16}$$

where

$$S_2 = \int \mathrm{d}t \int đ^d k \psi_k^\dagger \left(\mathbf{i}\partial_t - \epsilon(k)\right)\psi_k.$$

Notice the strong similarity with the XY model action in our discussion of the RG (in fact this similarity was Shankar's motivation for explaining the RG for the XY model in the (classic) paper I cited there). A mean field theory description of the condensation of Cooper pairs (15.15) is obtained by replacing the quartic term in (15.16) by expectation values:

$$\begin{aligned} S_{MFT}[\psi] &= S_2[\psi] + \int \mathrm{d}t\mathrm{d}^d x \; u \langle \psi\psi \rangle \psi^\dagger \psi^\dagger + h.c. \\ &= S_2[\psi] + \int \mathrm{d}t\mathrm{d}^d x \; u\Phi\psi^\dagger\psi^\dagger + h.c. \end{aligned} \tag{15.17}$$

So an expectation value for Φ is a mass for the fermions. It is a funny kind of symmetry-breaking mass, but if you diagonalize the quadratic operator in (15.17) (actually it is done below) you will find that it costs an energy of order $\Delta E_\psi = u\langle\Phi\rangle$ to excite a fermion. That's the cutoff on the LG EFT.

A general lesson from this example is: the useful degrees of freedom at low energies can be very different from the microscopic dofs.

15.6.1 Lightning discussion of BCS.

I am sure that some of you are nervous about the step from $S[\psi]$ to $S_{MFT}[\psi]$ above. To make ourselves feel better about it, I will say a few more words about the steps from the microscopic model of electrons (15.16) to the LG theory of Cooper pairs (these steps were taken by Bardeen, Cooper and Schreiffer (BCS)).

First recall the *Hubbard-Stratonovich transformation* aka *completing the square.* in 0+0 dimensional field theory:

$$e^{-\mathbf{i}ux^4} = \sqrt{2\pi u}\int_{-\infty}^{\infty} \mathrm{d}\sigma \; e^{-\frac{1}{\mathbf{i}u}\sigma^2 - 2\mathbf{i}x^2\sigma} \; . \tag{15.18}$$

At the cost of introducing an extra field σ, we turn a quartic term in x into a quadratic term in x. The RHS of (15.18) is gaussian in x and we know how to integrate it over x. (The version with $\mathbf{i}$ is relevant for the real-time integral.) Notice the weird extra factor of $\mathbf{i}$ lurking in (15.18). This can be understood as arising because we are trying to use a scalar field σ, to mediate a repulsive interaction (which it is, for positive u) (see Zee p. 193, 2nd Ed).

Actually, we'll need a complex H-S field:

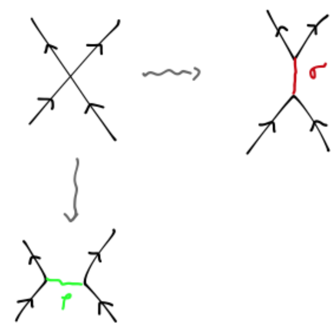
$$e^{-\mathbf{i}ux^2\bar{x}^2} = 2\pi u^2 \int_{-\infty}^{\infty} \mathrm{d}\sigma \int_{-\infty}^{\infty} \mathrm{d}\bar{\sigma} \; e^{-\frac{1}{\mathbf{i}u}|\sigma|^2 - \mathbf{i}x^2\bar{\sigma} - \mathbf{i}\bar{x}^2\sigma} \; . \tag{15.19}$$

(The field-independent prefactor is, as usual, not important for path integrals.)

We can use a field theory generalization of (15.19) to 'decouple' the 4-fermion interaction in (15.16):

$$Z = \int [D\psi D\psi^\dagger] e^{\mathbf{i}S[\psi]} = \int [D\psi D\psi^\dagger D\sigma D\sigma^\dagger] e^{\mathbf{i}S_2[\psi] + \mathbf{i}\int d^D x (\bar{\sigma}\psi\psi + h.c.) - \int \mathrm{d}^D x \frac{|\sigma|^2(x)}{\mathbf{i}u}} . \quad (15.20)$$

The point of this is that now the fermion integral is gaussian. At the saddle point of the σ integral (which is exact because it is gaussian), σ is the Cooper pair field, $\sigma_{\text{saddle}} = u\psi\psi$.



Notice that we made a choice here about in which 'channel' to make the decoupling – we could have instead introduces a different auxiliary field ρ and written $S[\rho, \psi] = \int \rho\psi^\dagger\psi + \int \frac{\rho^2}{2u}$, which would break up the 4-fermion interaction in the t-channel (as an interaction of the fermion density $\psi^\dagger\psi$) instead of the s (BCS) channel (as an interaction of Cooper pairs ψ^2). At this stage both are correct, but they lead to different mean-field approximations below. That the BCS mean field theory wins is a consequence of the RG.

How can you resist doing the fermion integral in (15.20)? Let's study the case where the single-fermion dispersion is $\epsilon(k) = \frac{\vec{k}^2}{2m} - \mu$.

$$I_\psi[\sigma] \equiv \int [D\psi D\psi^\dagger] e^{\mathbf{i}\int \mathrm{d}t\mathrm{d}^d x \left(\psi^\dagger\left(\frac{\nabla^2}{2m} - \mu\right)\psi + \psi\bar{\sigma}\psi + \bar{\psi}\bar{\psi}\sigma\right)}$$

The action here can be written as the integral of

$$L = \begin{pmatrix}\bar{\psi} & \psi\end{pmatrix} \begin{pmatrix} \mathbf{i}\partial_t - \epsilon(-\mathbf{i}\nabla) & \sigma \\ \bar{\sigma} & -\left(\mathbf{i}\partial_t - \epsilon(-\mathbf{i}\nabla)\right) \end{pmatrix} \begin{pmatrix}\psi \\ \bar{\psi}\end{pmatrix} \equiv \begin{pmatrix}\bar{\psi} & \psi\end{pmatrix} M \begin{pmatrix}\psi \\ \bar{\psi}\end{pmatrix}$$

so the integral is

$$I_\psi[\sigma] = \det M = e^{\operatorname{tr}\log M(\sigma)}.$$

The matrix M is diagonal in momentum space, and the integral remaining to be done is

$$\int [D\sigma D\sigma^\dagger] e^{-\int \mathrm{d}^D x \frac{|\sigma(x)|^2}{2\mathbf{i}u} + \int \text{đ}^D k \log\left(\omega^2 - \epsilon_k^2 - |\sigma_k|^2\right)}.$$

It is often possible to do this integral by saddle point. This can justified, for example, by the largeness of the volume of the Fermi surface, $\{k|\epsilon(k) = \mu\}$, or by large N number of species of fermions. The result is an equation which determines σ, which as we saw earlier determines the fermion gap.

$$0 = \frac{\delta \text{exponent}}{\delta\bar{\sigma}} = \mathbf{i}\frac{\sigma}{2u} + \int \text{đ}\omega \text{đ}^d k \frac{2\sigma}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + \mathbf{i}\epsilon} .$$

We can do the frequency integral by residues:

$$\int đ\omega \frac{1}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + \mathbf{i}\epsilon} = \frac{1}{2\pi} 2\pi \mathbf{i} \frac{1}{2\sqrt{\epsilon_k^2 + |\sigma|^2}}.$$

The resulting equation is naturally called the *gap equation*:

$$1 = -2u \int đ^d p' \frac{1}{\sqrt{\epsilon(p')^2 + |\sigma|^2}} \tag{15.21}$$

which you can imagine solving self-consistently for σ. Plugging back into the action (15.20) says that σ determines the energy cost to have electrons around; more precisely, σ is the energy required to break a Cooper pair.

Comments:

- If we hadn't restricted to a delta-function 4-fermion interaction $u(p, p') = u_0$ at the outset, we would have found a more general equation like

$$\sigma(\vec{p}) = -\frac{1}{2} \int đ^d p' \frac{u(p, p')\sigma(\vec{p}')}{\sqrt{\epsilon(p')^2 + |\sigma(p')|^2}} \,.$$

- Notice that a solution of (15.21) requires $u < 0$, an *attractive* interaction. Superconductivity happens because the u that appears here is not the bare interaction between electrons, which is certainly repulsive (and long-ranged). This is where the phonons come in in the BCS discussion.

- I haven't included here effects of the fluctuations of the fermions. In fact, they make the four-fermion interaction which leads to Cooper pairing marginally *relevant*. This breaks the degeneracy in deciding how to split up the $\psi\psi\psi^\dagger\psi^\dagger$ into *e.g.* $\psi\psi\sigma$ or $\psi^\dagger\psi\rho$. BCS wins. This is explained beautifully in Polchinski, lecture 2, and R. Shankar. If there were time, I would summarize the EFT framework for understanding this in §15.7.

- A conservative perspective on the preceding calculation is that we have made a variational ansatz for the groundstate wavefunction, and the equation we solve for σ is minimizing the variational energy – finding the best wavefunction within the ansatz.

- I've tried to give the most efficient introduction I could here. I left out any possibility of k-dependence or spin dependence of the interactions or the pair field, and I've conflated the pair field with the gap. In particular, I've been sloppy about the dependence on k of σ above.

- You studied very closely related manipulation on a previous problem set, in an example (the Gross-Neveu model) where the saddle point is justified by large N.

15.7 Effective field theory of Fermi surfaces

[Polchinski, lecture 2, and R. Shankar] Electrically conducting solids are a remarkable phenomenon. An arbitrarily small electric field $\vec{E}$ leads to a nonzero current $\vec{j} = \sigma\vec{E}$. This means that there must be gapless modes with energies much less than the natural cutoff scale in the problem.

Scales involved: The Planck scale of solid state physics (made by the logic by which Planck made his quantum gravity energy scale, namely by making a quantity with dimensions of energy out of the available constants) is

$$E_0 = \frac{1}{2}\frac{e^4 m}{\hbar^2} = \frac{1}{2}\frac{e^2}{a_0} \sim 13\text{eV}$$

(where $m \equiv m_e$ is the electron mass and the factor of 2 is an abuse of outside information) which is the energy scale of *chemistry*. Chemistry is to solids as the melting of spacetime is to particle physics. There are other scales involved however. In particular a solid involves a lattice of nuclei, each with $M \gg m$ (approximately the proton mass). So m/M is a useful small parameter which controls the coupling between the electrons and the lattice vibrations. Also, the actual speed of light $c \gg v_F$ can generally also be treated as ∞ to first approximation. v_F/c suppresses spin orbit couplings (though large atomic numbers enhance them: $\lambda_{\text{SO}} \propto Z v_F/c$).

Let us attempt to construct a Wilsonian-natural effective field theory of this phenomenon. The answer is called Landau Fermi Liquid Theory. What are the right low-energy degrees of freedom? Let's make a guess that they are like electrons – fermions with spin and electric charge. They will not have exactly the properties of free electrons, since they must incorporate the effects of interactions with all their friends. The 'dressed' electrons are called quasielectrons, or more generally quasiparticles.

Given the strong interactions between so many particles, why should the dofs have anything at all to do with electrons? Landau's motivation for this description (which is not always correct) is that we can imagine starting from the free theory and adiabatically turning up the interactions. If we don't encounter any phase transition along the way, we can follow each state of the free theory, and use the same labels in the interacting theory.

We will show that there is a nearly-RG-stable fixed point describing gapless quasi-electrons. Notice that we are not trying to match this description directly to some microscopic lattice model of a solid; rather we will do bottom-up effective field theory.

Having guessed the necessary dofs, let's try to write an action for them consistent with the symmetries. A good starting point is the free theory:

$$S_{\text{free}}[\psi] = \int dt\, đ^d p \left(\mathbf{i}\psi_\sigma^\dagger(p)\partial_t\psi_\sigma(p) - (\epsilon(p) - \epsilon_F)\, \psi_\sigma^\dagger(p)\psi_\sigma(p) \right)$$

where σ is a spin index, ϵ_F is the Fermi energy (zero-temperature chemical potential), and $\epsilon(p)$ is the single-particle dispersion relation. For non-interacting non-relativistic electrons in free space, we have $\epsilon(p) = \frac{p^2}{2m}$. It will be useful to leave this as a general function of p. [38] [39]

The groundstate is the filled Fermi sea:

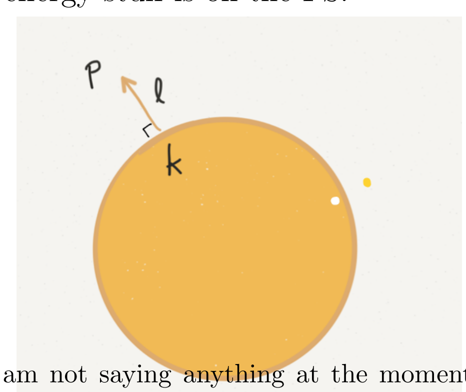
$$|\text{gs}\rangle = \prod_{p|\epsilon(p)<\epsilon_F} \psi_p^\dagger |0\rangle\,, \quad \psi_p |0\rangle = 0, \quad \forall p.$$

(If you don't like continuous products, put the system in a box so that p is a discrete label.) The Fermi surface is the set of points in momentum space at the boundary of the filled states:

$$\text{FS} \;\equiv \{p|\epsilon(p) = \epsilon_F\}.$$

The low-lying excitations are made by adding an electron just above the FS or removing an electron (creating a hole) just below.

We would like to define a scaling transformation which focuses on the low-energy excitations. We scale energies by a factor $E \to bE, b < 1$. In relativistic QFT, $\vec{p}$ scales like E, toward zero, $\vec{p} \to b\vec{p}$, since all the low-energy stuff is near $\vec{p} = 0$. Here the situation is much more interesting because the low-energy stuff is on the FS.



One way to implement this is to introduce a hierarchical labeling of points in momentum space, by breaking the momentum space into *patches* around the FS. (An analogous strategy of labeling is also used in heavy quark EFT and in SCET.)

We'll use a slightly different strategy, following Polchinski. To specify a point $\vec{p}$, we pick the

[38]Notice that we are assuming translation invariance. I am not saying anything at the moment about whether translation invariance is discrete (the ions make a periodic potential) or continuous.

[39]We have chosen the normalization of ψ to fix the coefficient of the ∂_t term (this rescaling may depend on p).

nearest point $\vec{k}$ on the FS, $\epsilon(\vec{k}) = \epsilon_F$ (draw a line perpendicular to the FS from $\vec{p}$), and let

$$\vec{p} = \vec{k} + \vec{\ell}.$$

So $d-1$ of the components are determined by $\vec{k}$ and one is determined by ℓ. (Clearly there are some exceptional cases if the FS gets too wiggly. Ignore these for now.)

$$\epsilon(p) - \epsilon_F = \ell v_F(\vec{k}) + \mathcal{O}(\ell^2), \qquad v_F \equiv \partial_p \epsilon|_{p=k}.$$

So a scaling rule which accomplishes our goal of focusing on the FS is

$$E \to bE, \quad \vec{k} \to \vec{k}, \quad \vec{l} \to b\vec{\ell}.$$

This implies

$$dt \to b^{-1} dt, \quad d^{d-1}\vec{k} \to d^{d-1}\vec{k}, \quad d\vec{\ell} \to bd\vec{\ell}, \quad \partial_t \to b\partial_t$$

$$S_{\text{free}} = \int \underbrace{dt\, d^{d-1}\vec{k}\, d\vec{\ell}}_{\sim b^0} \left(\mathbf{i}\psi^\dagger(p) \underbrace{\partial_t}_{\sim b^1} \psi(p) - \underbrace{\ell v_F(k)}_{\sim b^1} \psi^\dagger(p)\psi(p) \right)$$

In order to make this go like b^0 we require $\psi \to b^{-\frac{1}{2}}\psi$ near the free fixed point.

Next we will play the EFT game. To do so we must enumerate the symmetries we demand of our EFT:

1. Particle number, $\psi \to e^{\mathbf{i}\theta}\psi$
2. Spatial symmetries: either (a) continuous translation invariance and rotation invariance (as for *e.g.* liquid ^{3}He) or (b) lattice symmetries. This means that momentum space is periodically identified, roughly $p \simeq p + 2\pi/a$ where a is the lattice spacing (the set of independent momenta is called the Brillouin zone (BZ)) and p is only conserved modulo an inverse lattice vector $2\pi/a$; the momentum There can also be some remnant of rotation invariance preserved by the lattice. Case (b) reduces to case (a) if the Fermi surface does not go near the edges of the BZ.
3. Spin rotation symmetry, $\mathsf{SU}(n)$ if $\sigma = 1..n$. In the limit with $c \to \infty$, this is an internal symmetry, independent of rotations.
4. Let's assume that $\epsilon(p) = \epsilon(-p)$, which is a consequence of *e.g.* parity invariance.

Now we enumerate all terms analytic in ψ (since we are assuming that there are no other low-energy operators integrating out which is the only way to get non-analytic terms in ψ) and consistent with the symmetries; we can order them by the number of fermion operators involved. Particle number symmetry means every ψ comes with a $\psi^\dagger$. The possible **quadratic terms** are:

$$\int \underbrace{dt\ d^{d-1}\vec{k}\ d\vec{\ell}}_{\sim b^0}\ \mu(k)\ \underbrace{\psi^\dagger_\sigma(p)\psi_\sigma(p)}_{\sim b^{-1}} \sim b^{-1}$$

is relevant. This is like a mass term. But don't panic: it just shifts the FS around. The *existence* of a Fermi surface is Wilson-natural; any precise location or shape (modulo something enforced by symmetries, like roundness) is not.

Adding one extra ∂_t or factor of ℓ costs a b^1 and makes the operator marginal; those terms are already present in S_{free}. Adding more than one makes it irrelevant.

Quartic terms:

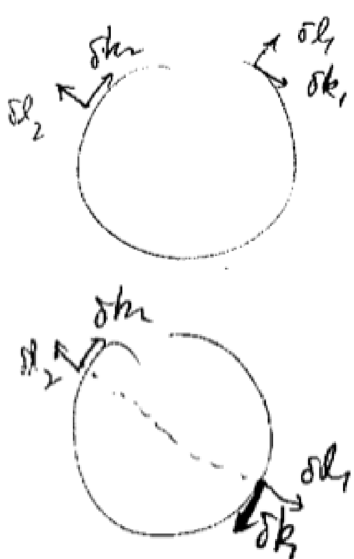
$$S_4 = \int \underbrace{dt \prod_{i=1}^{4} d^{d-1}\vec{k}_i d\vec{\ell}_i}_{\sim b^{-1+4-4/2}} u(4\cdots 1)\psi^\dagger_\sigma(p_1)\psi_\sigma(p_3)\psi^\dagger_{\sigma'}(p_2)\psi_{\sigma'}(p_4)\delta^d(\vec{p}_1+\vec{p}_2-\vec{p}_3-\vec{p}_4)$$

Note the similarity with the discussion of the XY model in §**??**. The minus signs on $p_{3,4}$ is because $\psi(p)$ *removes* a particle with momentum p. We assume u depends only on k, σ, so does not scale – this will give the most relevant piece. How does the delta function scale?

$$\delta^d\left(\vec{p}_1+\vec{p}_2-\vec{p}_3-\vec{p}_4\right) = \delta^d\left(k_1+k_2-k_3-k_4+\ell_1+\ell_2-\ell_3-\ell_4\right) \overset{?}{\simeq} \delta^d\left(k_1+k_2-k_3-k_4\right)$$

In the last (questioned) step, we used the fact that $\ell \ll k$ to ignore the contributions of the ℓs. If this is correct then the delta function does not scale (since ks do not), and $S_4 \sim b^1$ is irrelevant (and quartic interactions with derivatives are moreso). If this were correct, the free-fixed point would be exactly stable.

There are two important subtleties: (1) there exist phonons. (2) the questioned equality above is questionable because of kinematics of the Fermi surface. We will address these two issues in reverse order.



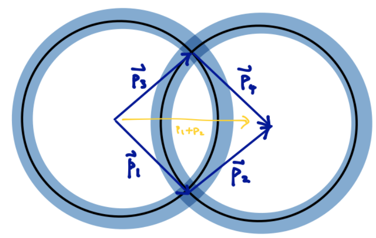
The **kinematic subtlety** in the treatment of the scaling of $\delta(p_1+p_2-p_3-p_4)$ arises because of the geometry of the Fermi surface. Consider scattering between two points on the FS, where (in the labeling convention above)

$$p_3 = p_1 + \delta k_1 + \delta \ell_1, \quad p_4 = p_2 + \delta k_2 + \delta \ell_2,$$

in which case the momentum delta function is

$$\delta^d(p_1 + p_2 - p_3 - p_4) = \delta^d(\delta k_1 + \delta \ell_1 + \delta k_2 + \delta \ell_2).$$

For generic choices of the two points $p_{1,2}$ (top figure at left), δk_1 and δk_2 are linearly independent and the $\delta\ell$s can indeed be ignored as we did above. However, for two points with $p_1 = -p_2$ (they are called *nested*, as depicted in the bottom figure at left), then one component of $\delta k_1 + \delta k_2$ is automatically zero, revealing the tiny $\delta\ell$s to the force of (one component of) the delta function. In this case, $\delta(\ell)$ scales like b^{-1}, and for this particular kinematic configuration the four-fermion interaction is (classically) marginal. Classically marginal means quantum mechanics has a chance to make a big difference.



A useful visualization is at right ($d = 2$ with a round FS is shown; this is what's depicted on the cover of the famous book by Abrikosov-Gorkov-Dzyaloshinski): the blue circles have radius k_F; the yellow vector is the sum of the two initial momenta $p_1 + p_2$, both of which are on the FS; the condition that $p_3 + p_4$, each also on the FS, add up to the same vector means that p_3 must lie on the intersection of the two circles (spheres in $d > 2$). But when $p_1 + p_2 = 0$, the two circles are on top of each other so they intersect everywhere! Comments:

1. We assumed that both p_1 and $-p_2$ were actually *on* the FS. This is automatic if $\epsilon(p) = \epsilon(-p)$, *i.e.* if ϵ is only a function of p^2.

2. This discussion works for any $d > 1$.

3. **Forward scattering.** There is a similar phenomenon for the case where $p_1 = p_3$ (and hence $p_2 = p_4$). This is called *forward scattering* because the final momenta are the same as the initial momenta. (We could just as well take $p_1 = p_4$ (and hence $p_2 = p_3$).) In this case too the delta function will constrain the ℓs and will therefore scale.

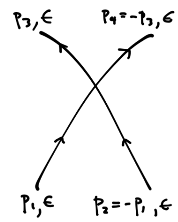
The tree-level-marginal 4-Fermi interactions at special kinematics leads to a *family* of fixed points labelled by 'Landau parameters'. In fact there is whole *function*s worth of fixed points. In 2d, the fixed point manifold is parametrized by the forward-scattering function

$$F(\theta_1, \theta_2) \equiv u(\theta_4 = \theta_2, \theta_3 = \theta_1, \theta_2, \theta_1)$$

(Fermi statistics implies that $u(\theta_4 = \theta_1, \theta_3 = \theta_2, \theta_2, \theta_1) = -F(\theta_1, \theta_2)$.) and the BCS-channel interaction:

$$V(\theta_1, \theta_3) = u(\theta_4 = -\theta_3, \theta_3, \theta_2 = -\theta_1, \theta_1).$$

Now let's think about what decision the fluctuations make about the fate of the nested interactions. The first claim, which I will not justify here, is that F is not renormalized at one loop. The interesting bit is the renormalization of the BCS interaction:



The electron propagator, obtained by inverting the kinetic operator S_{free}, is

$$G(\epsilon, p = k + l) = \frac{1}{\epsilon(1 + \mathbf{i}\eta) - v_F(k)\ell + \mathcal{O}(\ell)^2}$$

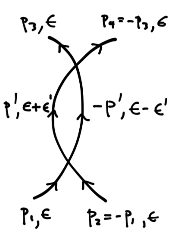
where I used $\eta \equiv 0^+$ for the infinitesimal specifying the contour prescription. (To understand the contour prescription for the *hole* propagator, it is useful to begin with

$$G(t, p) = \langle \epsilon_F | c_p^\dagger(t) c_p(0) | \epsilon_F \rangle, \quad c_p^\dagger(t) \equiv e^{-\mathbf{i}\mathbf{H}t} c_p^\dagger e^{\mathbf{i}\mathbf{H}t}$$

and use the free-fermion fact $[\mathbf{H}, c_p^\dagger] = \epsilon_p c_p^\dagger$.)

Let's assume rotation invariance. Then $V(\theta_3, \theta_1) = V(\theta_3 - \theta_1)$, $V_l = \int đ\theta e^{\mathbf{i}l\theta} V(\theta)$. Different angular momentum sectors decouple from each other at one loop.

We will focus on the s-wave bit of the interaction, so V is independent of momentum. We will integrate out just a shell in energy (depicted by the blue shaded shell in the Fermi surface figures) The interesting contribution comes from the following diagram:

$\delta^{(1)} V =$ 

$$= \mathbf{i} V^2 \int_{b\epsilon_0}^{\epsilon_0} \frac{d\epsilon' d^{d-1}k' d\ell'}{(2\pi)^{d+1}} \frac{1}{(\epsilon + \epsilon' - v_F(k')\ell')(\epsilon - \epsilon' - v_F(k')\ell')}$$

do $\int d\ell'$ by residues

$$= \mathbf{i} V^2 \int \frac{d\epsilon' d^{d-1}k'}{(2\pi)^{d+1}} \frac{1}{v_F(k')} \left(\underbrace{\epsilon - \epsilon' - (\epsilon + \epsilon')}_{=-2\epsilon'} \right)^{-1}$$

$$= -V^2 \underbrace{\int_{b\epsilon_0}^{\epsilon_0} \frac{d\epsilon'}{\epsilon'}}_{=\log(1/b)} \underbrace{\int \frac{d^{d-1}k'}{(2\pi)^d v_F(k')}}_{\text{dos at FS}} \tag{15.22}$$

Between the first and second lines, we did the ℓ' integral by residues. The crucial point is that we are interested in external energies $\epsilon \sim 0$, but we are integrating out a shell near the cutoff, so $|\epsilon'| > |\epsilon|$ and the sign of $\epsilon + \epsilon'$ is opposite that of $\epsilon - \epsilon'$; therefore there is a pole on either side of the real ℓ axis and we get the same answer by closing the contour either way. On one side the pole is at $\ell' = \frac{1}{v_F(k')}(\epsilon + \epsilon')$. (In the t-channel diagram (what Shankar calls ZS), the poles are on the *same* side and it therefore does not renormalize the four-fermion interaction.)

The result to one-loop is then

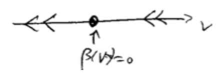
$$V(b) = V - V^2 N \log(1/b) + \mathcal{O}(V^3)$$

with $N \equiv \int \frac{d^{d-1}k'}{(2\pi)^d v_F(k')}$ is the density of states at the Fermi surface. From this we derive the beta function

$$b\frac{d}{db}V(b) = \beta_V = NV^2(b) + \mathcal{O}(V^3)$$

and the solution of the flow equation at $E = bE_1$ is

$$V(E) = \frac{V_1}{1 + NV_1 \log(E_1/E)} \begin{cases} \to 0 & \text{in IR for } \; V_1 > 0 \text{ (repulsive)} \\ \to -\infty & \text{in IR for } \; V_1 < 0 \text{ (attractive)} \end{cases} \tag{15.23}$$

There is therefore a very significant dichotomy depending on the sign of the coupling at the microscopic scale E_1, as in this phase diagram: 

The conclusion is that if the interaction starts attractive at some scale it flows to large attractive values. The thing that is decided by our perturbative analysis is that (if $V(E_1) > 0$) the decoupling we did with σ ('the BCS channel') wins over the decoupling with ρ ('the particle-hole channel'). What happens at $V \to -\infty$? Here we need non-perturbative physics.

The non-perturbative physics is in general hard, but we've already done what we can in §15.6.1.

The remaining question is: Who is V_1 and why would it be attractive (given that Coulomb interactions between electrons, while screened and therefore short-ranged, are repulsive)? The answer is:

Phonons. The lattice of positions taken by the ions making up a crystalline solid spontaneously break many spacetime symmetries of their governing Hamiltonian. This implies a collection of gapless Goldstone modes in any low-energy effective theory of such a solid[40]. The Goldstone theorem is satisfied by including a field

$$\vec{D} \propto \text{(local) displacement } \delta\vec{r} \text{ of ions from their equilibrium positions}$$

Most microscopically we have a bunch of coupled springs:

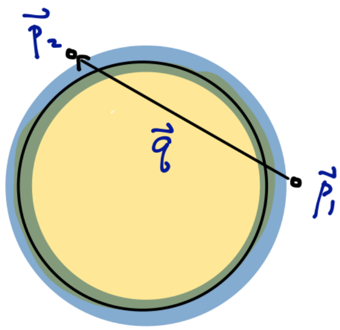
$$L_{\text{ions}} \sim \frac{1}{2}M\left(\dot{\delta\vec{r}}\right)^2 - k_{ij}\delta r^i \delta r^j + \ldots$$

[40] Note that there is a subtlety in counting Goldstone modes from spontaneously broken spacetime symmetries: there are more symmetry generators than Goldstones. Basically it's because the associated currents differ only by functions of spacetime; but a localized Goldstone particle is anyway made by a current times a function of spacetime, so you can't sharply distinguish the resulting particles. Some useful references on this subject are Low-Manohar and most recently Watanabe-Murayama.

with spring constants k independent of the nuclear mass M. It is useful to introduce a canonically normalized field in terms of which the action is

$$S[\vec{D} = (M)^{1/2}\,\delta\vec{r}] = \frac{1}{2}\int dt d^d q \left(\partial_t D_i(q)\partial_t D_i(-q) - \omega_{ij}^2(q) D_i(q) D_j(-q)\right).$$

Here $\omega^2 \propto M^{-1}$. Their status as Goldstones means that the eigenvalues of $\omega_{ij}^2(q) \sim |q|^2$ at small q: moving everyone by the same amount does not change the energy. This also constrains the coupling of these modes to the electrons: they can only couple through derivative interactions.



For purposes of their interactions with the electrons, a nonzero q which keeps the e^- on the FS must scale like $q \sim b^0$. Therefore

$$dt d^d q \left(\partial_t D\right)^2 \sim b^{+1+2[D]} \qquad \implies D \sim b^{-\frac{1}{2}}$$

and the restoring force $dtdqD^2\omega^2(q) \sim b^{-2}$ is relevant, and dominates over the ∂_t^2 term for

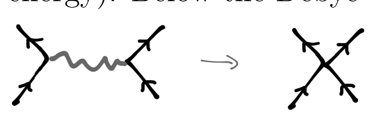
$$E < E_D = \sqrt{\frac{m}{M}} E_0 \qquad \text{the Debye energy.}$$

This means that phonons mediate *static* interactions below E_D – we can ignore retardation effects, and their effects on the electrons can be fully incorporated by the four-fermion interaction we used above (with some $\vec{k}$ dependence). How do they couple to the electrons?

$$\begin{aligned} S_{\text{int}}[D,\psi] &= \int dt q^3 q d^2 k_1 d\ell_1 d^2 k_2 d\ell_2 \; M^{-\frac{1}{2}} g_i(q,k_1,k_2) D_i(q) \psi_\sigma^\dagger(p_1)\psi_\sigma(p_2)\delta^3(p_1 - p_2 - q) \\ &\sim b^{-1+1+1-3/2} = b^{-1/2} \end{aligned} \tag{15.24}$$

– here we took the delta function to scale like b^0 as above. This is relevant when we use the $\dot{D}^2$ scaling for the phonons; when the restoring force dominates we should scale D differently and this is irrelevant for generic kinematics. This is consistent with our previous analysis of the four-fermion interaction.

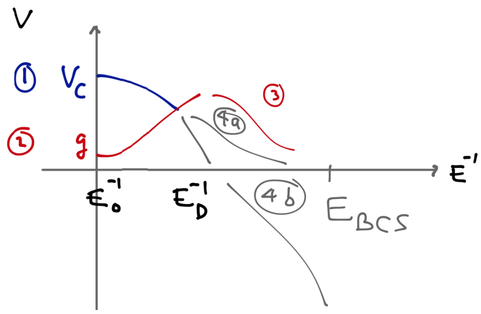
The summary of this discussion is: phonons do not destroy the Fermi surface, but they *do* produce an attractive contribution to the 4-fermion interaction, which is relevant in some range of scales (above the Debye energy). Below the Debye energy, it amounts to an addition to V that goes like $-g^2$:

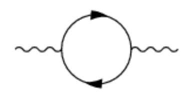


Notice that the scale at which the coupling V becomes strong ($V(E_{\text{BCS}}) \equiv 1$ in (15.23)) is

$$E_{\text{BCS}} \sim E_D e^{-\frac{1}{NV_D}}.$$

Two comments about this: First, it is non-perturbative in the interaction V_D. Second, it provides some verification of the role of phonons, since $E_D \sim M^{-1/2}$ can be varied by studying the same material with different isotopes and studying how the critical superconducting temperature ($\sim E_{\text{BCS}}$) scales with the nuclear mass.



Here's the narrative, proceeding as a function of decreasing energy scale, beginning at E_0, the Planck scale of solids: (1) Electrons repel each other by the Coulomb interaction. However, in a metal, this interaction is *screened* by processes like this:  (the intermediate state is an electron-hole pair) and is short-ranged. It is still repulsive, however. As we coarse-grain more and more, we see more and more electron-hole pairs and the force weakens. (2) While this is happening, the electron-phonon interaction is relevant and growing. This adds an attractive bit to V. This lasts until E_D. (3) At E_D the restoring force term in the phonon lagrangian dominates (for the purposes of their interactions with the electrons) and we can integrate them out. (4) What happens next depends on the sign of $V(E_D)$. If it's positive, V flows harmlessly to zero. If it's negative, it becomes moreso until we exit the perturbative analysis at E_{BCS}, and vindicate our choice of Hubbard-Stratonovich channel above.

Further brief comments, for which I refer you to Shankar:

1. Putting back the possible angular dependence of the BCS interaction, the result at one loop is

$$\frac{dV(\theta_1 - \theta_3)}{d\ell} = -\frac{1}{8\pi^2}\int_0^{2\pi} \bar{d}\theta V(\theta_1 - \theta)V(\theta - \theta_3)$$

   or in terms of angular momentum components,

$$\frac{dV_l}{d\ell} = -\frac{V_l^2}{4\pi}.$$

2. This example is interesting and novel in that it is a (family of) fixed point(s) characterized by a dimensionful quantity, namely k_F. This leads to a phenomenon called *hyperscaling violation* where thermodynamic quantities need not have their naive scaling with temperature.

3. The one loop analysis gives the right answer to all loops in the limit that $N \equiv k_F/\Lambda \gg 1$, where Λ is the UV cutoff on the momentum.

4. The forward scattering interaction (for any choice of function $F(\theta_{13})$) is not renormalized at one loop. This means it is exactly marginal at leading order in N.

5. Like in ϕ^4 theory, the sunrise diagram at two loops is the first appearance of wavefunction renormalization. In the context of the Fermi liquid theory, this leads to the renormalization of the effective mass which is called $m^\star$.

Another consequence of the FS kinematics which I should emphasize more: it allows the quasiparticle to be stable. The leading contribution to the decay rate of a one-quasiparticle state with momentum k can be obtained applying the optical theorem to the following process.

The intermediate state is two electrons with momenta $k' + q$ and $k - q$, and one *hole* with momentum k'. The hole propagator has the opposite $\mathbf{i}\eta$ prescription. After doing the frequency integrals by residues, we get

$$\Sigma(k, \epsilon) = \int đq\, đk' \frac{|u_q|^2}{D - \mathbf{i}\eta}$$

$$D \equiv \epsilon_k(1 + \mathbf{i}\eta) + \epsilon_{k'}(1 - \mathbf{i}\eta) - \epsilon_{k'+q}(1 + \mathbf{i}\eta) - \epsilon_{k-q}(1 + \mathbf{i}\eta)$$

(Notice that this is the eyeball diagram which gives the lowest-order contribution to the wavefunction renormalization of a field with quartic interactions.) By the optical theorem, its imaginary part is the (leading contribution to the) inverse-lifetime of the quasiparticle state with fixed k:

$$\tau^{-1}(k) = \operatorname{Im}\Sigma(k, \epsilon) = \pi \int đq\, đk' \delta(D)|u_q|^2 f(-\epsilon_{k'}) f(\epsilon_{k'+q}) f(\epsilon_{k-q})$$

where

$$f(\epsilon) = \lim_{T \to 0} \frac{1}{e^{\frac{\epsilon - \epsilon_F}{T}} + 1} = \theta(\epsilon < \epsilon_F)$$

is the Fermi function. This is just the demand that a particle can only scatter into an empty state and a hole can only scatter into a filled state. These constraints imply that all the energies are near the Fermi energy: both $\epsilon_{k'+q}$ and $\epsilon_{k'}$ lie in a shell of radius ϵ about the FS; the answer is proportional to the density of possible final states, which is thus

$$\tau^{-1} \propto \left(\frac{\epsilon}{\epsilon_F}\right)^2 .$$

So the width of the quasiparticle resonance is

$$\tau^{-1} \propto \epsilon^2 \ll \epsilon$$

much smaller than its frequency – it is a sharp resonance, a well-defined particle.
