Physics 215B: Quantum Field Theory, Part 2
Winter 2025

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

Last updated: 2025/03/12, 21:16:56

Contents

0.1 Introductory remarks

Here is a brief reminder about the big picture. Quantum field theory (QFT) is the quantum mechanics of *extensive degrees of freedom*. What I mean by this is that at each point of space, there's some stuff that can wiggle.

'Extensive degrees of freedom' are those 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*.

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 $\mathbf{H} = \sum_{\text{patches}} \mathbf{H}(\text{nearby patches})$.[1]

Such a starting point is useful as a description of the excitations of the vacuum (high energy physics), or (perhaps more obviously) as a description of various kinds of condensed matter. In the former case, there is an (emergent) Lorentz symmetry. As you saw a bit in 215A, the long-wavelength excitations of the coupled patches of space need have little in common with the microscopic variables. The big goal in this business is to connect these two descriptions; the framework in which we will do so is called the Renormalization Group (RG).

Sometimes there is a description in terms of weakly-coupled fields. In 215A, you saw that the low-energy excitations of weakly-coupled fields are *particles*. A consequence of relativity is that the number of particles isn't fixed. That is: there are processes where the number of particles changes in time. Sometimes this is used to motivate the study of QFT. It's a necessary consequence of Lorentz symmetry, but the converse is false: particle production and destruction can happen without relativity.

'Divergences'. It is tempting to try to define QFT directly in the continuum, rather than in a regulated way as I have above. Such a description is more symmetric, and hence easier to calculate with, no question. In practice, we will use the continuum a lot of the time. When we do so, we will encounter UV divergences, and it will be useful to remember the starting point above. It will also be useful to remember that we have no way of knowing that some version of the starting point above isn't correct

[1] Actually, the Hilbert space of a gauge theory is *not* of this form; rather, it is a subspace of such a space which satisfies the Gauss law. This is a source of a lot of confusion, which I hope to dispel.

at short-enough distances, and that anything we say about such short distance scales is a total fiction on which our answers to physics questions cannot depend. Part of our job is to understand why it is that our complete ignorance of short-distance physics doesn't prevent us from making progress.

Last quarter 215A ended right when you learned to compute amplitudes for simple processes in QED. There are many measurable quantities that can be computed using the formalism you developed, and the success of leading-order QED is a real high point of physics. I have a bit more to say about that success.

Lurking behind that success, however, is a dark cloud. The leading order of perturbation theory is given by tree diagrams; more complicated diagrams should be suppressed by more powers of $\frac{e^2}{4\pi} \equiv \alpha$. You might think that if the leading calculation worked so well, we should do even better by looking at the next term. But there is a surprise: naively calculating the next term in the continuum gives an infinite correction.

So an important job will be to explain the non-naive point of view on this calculation which allows us to extract finite, meaningful answers from perturbation theory.

Even more important will be to give a better point of view, from which we never would have encountered divergences in the first place. The (silly) name for this point of view is the renormalization group.

So here are some goals for the quarter, both practical and philosophical:

- A central theme this quarter will be coarse-graining in quantum systems with extensive degrees of freedom, aka the renormalization group (RG) in QFT.

  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. And the things we should care about the most are the biggest ones – the modes with the longest wavelength. So the 'better perspective' alluded to above is: we should try to understand QFT scale by scale.

  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[2]) 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

[2]It isn't. Grass is a really stupid plant for us to be growing.

(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.

- At some point we should talk about non-Abelian gauge theory, since it's the basis of the Standard Model of particle physics.

- There is more to QFT than perturbation theory about free fields in a Fock vacuum. There exist QFTs that we don't know how to build in this way at all.

- It is worthwhile to spend some 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!). Maybe this will wait until 215C.

- In addition to its importance in high energy physics, I want to emphasize that QFT is also quite central in many aspects of condensed matter physics. 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, theorist-centered perspective on QFT that 'begins from a classical lagrangian and quantizes it'.

- Given time, I would like to show that many fancy phenomena precious to particle physicists can emerge from 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.

0.2 Sources and acknowledgement

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*

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

Banks, *Modern Quantum Field Theory: A Concise Introduction*

Schwartz, *Quantum field theory and the standard model*

David Tong's lecture notes

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

Some other books that might be useful to us are:

Xiao-Gang Wen, *Quantum Field Theory of Many-Body Systems*

Sidney Coleman, *Aspects of Symmetry*

Alexander Polyakov, *Gauge Fields and Strings*

Eduardo Fradkin, *Field Theories of Condensed Matter Systems*

Eduardo Fradkin, *Quantum Field Theory, an Integrated Approach*

0.3 Conventions

Following most QFT books, I am going to use the $+---$ signature convention for the Minkowski metric. I am (somehow, still) 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 $i, j, k, ...$ for spatial indices.

The convention that repeated indices are summed is always in effect unless otherwise indicated. d is the number of space dimensions, D is the number of spacetime dimensions (it's bigger!).

$\equiv$ means 'equals by definition'. $A \stackrel{!}{=} B$ means we are demanding that $A = B$. $A \stackrel{?}{=} B$ means A probably doesn't equal B.

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 $đk \equiv \frac{dk}{2\pi}$. 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

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

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.
$\mathcal{L} \ni \mathcal{O}$ means the object $\mathcal{L}$ contains the term $\mathcal{O}$.

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.

1 Three Parables on the UV

At this point I believe you are able to use QED to compute the amplitudes and cross-sections for many physical processes involving electrons, photons and positrons. More precisely, you know how to compute the leading-order-in-α amplitudes, using Feynman diagrams that are trees – diagrams without loops. The natural next step is to look at the next terms in the perturbation expansion in α, which come from diagrams with one loop. This innocent-seeming step opens a big can of worms. The reason this is a big step is that a tree-level process is classical in a certain sense.

In a tree-level diagram, the quantum numbers (and in particular the momenta) of the intermediate states are fixed by the external states[3]. In contrast, once there is a loop, there are undetermined momenta which must be summed, and this sum includes, it seems, arbitrarily high momentum modes. Surely we have no information yet about these modes from our piddling low-energy experiments. (Perhaps this is an opportunity to learn about them?) What do we do?

In order to put ourselves in the right frame of mind to think about this stuff, we'll make a brief retreat to a parable about a system with finitely many degrees of freedom in §1.1. A second parable is motivated by the question: where do field theories come from? In §1.2, I want to convey that what looks to a long-wavelength observer like a QFT may actually be very simple microscopically; I will use this as an opportunity to introduce the path integral approach to QFT. A third useful parable in §1.3 will come from the Casimir effect – the additive constant in the energy of free fields. Then in the next chapter we'll apply these lessons to a simple field theory example, namely scalar field theory. Then we'll come back to perturbation theory in QED. Reading assignment for this chapter: Zee §III.

1.1 A parable from quantum mechanics on the breaking of scale invariance

Recall that the coupling constant in ϕ^4 theory in $D = 3 + 1$ spacetime dimensions is dimensionless, and the same is true of the electromagnetic coupling e in QED in $D = 3+1$ spacetime dimensions. In fact, the mass parameters are the only dimensionful quantities in those theories, at least in their classical avatars. This means that if we ignore the masses, for example because we are interested in physics at much higher energies, then these models seem to possess *scale invariance*: the physics is unchanged under zooming in or out.

[3] If multiple diagrams contribute to the same process, there can still be a small amount of quantum interference.

Here we will study a simple quantum mechanical example (that is: an example with a finite number of degrees of freedom)[4] with such (classical) scale invariance. It exhibits many interesting features that can happen in strongly interacting quantum field theory – asymptotic freedom, dimensional transmutation. Because the model is simple, we can understand these phenomena without resort to perturbation theory. They will nevertheless illuminate some ways of thinking that we'll need in examples where perturbating is our only option.

Consider the following ('bare') action for a QFT in $D = 0 + 1$ dimensions:

$$S[q] = \int dt \left(\frac{1}{2}\dot{\vec{q}}^2 + g_0 \delta^{(2)}(\vec{q}) \right) \equiv \int dt \left(\frac{1}{2}\dot{\vec{q}}^2 - V(\vec{q}) \right)$$

where $\vec{q} = (x, y)$ are two coordinates of a quantum particle, and the potential involves $\delta^{(2)}(\vec{q}) \equiv \delta(x)\delta(y)$, a Dirac delta function. I chose the sign so that $g_0 > 0$ is attractive. (Notice that I have absorbed the inertial mass m in $\frac{1}{2}mv^2$ into a redefinition of the variable q, $q \to \sqrt{m}q$.)

First, let's do dimensional analysis (always a good idea). Since $\hbar = 1$, all dimensionful quantites are some power of an energy. Let $[X]$ denote the number of powers of energy in the units of the quantity X; that is, if $X \sim (\text{energy})^{+\nu(X)}$ then we have $[X] = \nu(X)$, a number. We have:

$$[t] = [(\text{energy}/\hbar)^{-1}] = -1 \implies [dt] = -1.$$

The action appears in the exponent in the path integrand, and is therefore dimensionless (it has units of $\hbar$), so we had better have:

$$0 = [S] = [\hbar]$$

and this applies to each term in the action. We begin with the kinetic term:

$$0 = [\int dt \dot{\vec{q}}^2] \implies$$

$$[\dot{\vec{q}}^2] = +1 \implies [\dot{\vec{q}}] = +\frac{1}{2} \implies [\vec{q}] = -\frac{1}{2}.$$

Since $1 = \int dq \delta(q)$, we have $0 = [dq] + [\delta(q)]$ and

$$[\delta^d(\vec{q})] = -[q]d = \frac{d}{2}, \quad \text{and in particular } [\delta^2(\vec{q})] = 1.$$

This implies that when $d = 2$ the naive ("engineering") dimensions of the coupling constant g_0 are $[g_0] = 0$ – it is dimensionless. Classically, the theory does not have a special length scale; it is scale invariant.

[4] I learned this example from the late Marty Halpern.

The Hamiltonian associated with the Lagrangian above is

$$H = \frac{1}{2}\left(p_x^2 + p_y^2\right) + V(\vec{q}).$$

Now we treat this as a quantum system. Acting in the position basis, the quantum Hamiltonian operator is

$$\mathbf{H} = -\frac{\hbar^2}{2}\left(\partial_x^2 + \partial_y^2\right) - g_0\delta^{(2)}(\vec{q})$$

So in the Schrödinger equation $\mathbf{H}\psi = \left(-\frac{\hbar^2}{2}\nabla^2 + V(\vec{q})\right)\psi = E\psi$, the second term on the LHS is

$$V(\vec{q})\psi(\vec{q}) = -g_0\delta^{(2)}(\vec{q})\psi(0).$$

To make it look more like we are doing QFT, let's solve it in momentum space:

$$\psi(\vec{q}) \equiv \int \frac{d^2p}{(2\pi\hbar)^2}\, e^{\mathbf{i}\vec{p}\cdot\vec{q}/\hbar}\varphi(\vec{p}) \equiv \int \text{đ}^2p\, e^{\mathbf{i}\vec{p}\cdot\vec{q}}\varphi(\vec{p}) \tag{1.1}$$

where I used some notation from §0.3 and set $\hbar = 1$. The delta function is

$$\delta^{(2)}(q) = \int \frac{d^2p}{(2\pi\hbar)^2}\, e^{\mathbf{i}\vec{p}\cdot\vec{q}/\hbar}. \tag{1.2}$$

So the Schrödinger equation says

$$\begin{aligned}
\left(-\frac{1}{2}\nabla^2 - E\right)\psi(q) &= -V(q)\psi(q) \\
\int \text{đ}^2p e^{ip\cdot q}\left(\frac{p^2}{2} - E\right)\varphi(p) &= +g_0\delta^2(q)\psi(0) \\
&\overset{(1.2)}{=} +g_0\left(\int \text{đ}^2p e^{\mathbf{i}p\cdot q}\right)\psi(0)
\end{aligned} \tag{1.3}$$

where to get to the second line, we just plugged in (1.1). Integrating the both-hand side of (1.3) over q: $(\int \text{đ}^2q e^{-\mathbf{i}p\cdot q}\,((1.3))\,)$ says

$$\left(\frac{\vec{p}^2}{2} - E\right)\varphi(\vec{p}) = +g_0\underbrace{\int \frac{d^2p'}{(2\pi\hbar)^2}\varphi(\vec{p}')}_{=\psi(0)}$$

There are two cases to consider:

- $\psi(\vec{q}=0) = \int \text{đ}^2 p \varphi(\vec{p}) = 0$. Then this case is the same as a free theory, with the constraint that $\psi(0) = 0$,

$$\left(\frac{\vec{p}^2}{2} - E\right)\varphi(\vec{p}) = 0$$

i.e. plane waves which vanish at the origin, *e.g.* $\psi \propto \sin \frac{p_x x}{\hbar} e^{\pm \mathbf{i} p_y y/\hbar}$. These scattering solutions don't see the delta-function potential at all.

- $\psi(0) \equiv \alpha \neq 0$, some constant to be determined. This means $\vec{p}^2/2 - E \neq 0$, so we can divide by it :

$$\varphi(\vec{p}) = \frac{g_0}{\frac{\vec{p}^2}{2} - E}\left(\int \text{đ}^2 p' \varphi(\vec{p}')\right) = \frac{g_0}{\frac{\vec{p}^2}{2} - E}\alpha.$$

The integral of the RHS (for $\psi(0) = \alpha$) is a little problematic if $E > 0$, since then there is some value of p where $p^2 = 2E$. Avoid this singularity by going to the boundstate region: consider $E = -\epsilon_B < 0$. So:

$$\varphi(\vec{p}) = \frac{g_0}{\frac{\vec{p}^2}{2} + \epsilon_B}\alpha.$$

What happens if we integrate this $\int \text{đ}^2 p$ to check self-consistency – the LHS should give α again:

$$0 \stackrel{!}{=} \underbrace{\int \text{đ}^2 p \varphi(\vec{p})}_{=\psi(0)=\alpha\neq 0}\left(1 - \int \text{đ}^2 p \frac{g_0}{\frac{\vec{p}^2}{2} + \epsilon_B}\right)$$

$$\implies \qquad \boxed{\int \text{đ}^2 p \frac{g_0}{\frac{\vec{p}^2}{2} + \epsilon_B} = 1}$$

is a condition on the energy ϵ_B of possible boundstates.

But there's a problem: the integral on the LHS behaves at large p like

$$\int \frac{d^2 p}{p^2} = \infty \ .$$

At this point in an ordinary QM class, you would give up on this model. In QFT we don't have that luxury, because this kind of thing happens all over the place. Here's what we do instead.

We cut off the integral at some large $p = \Lambda$:

$$\int^{\Lambda} \frac{d^2 p}{p^2} \sim \log \Lambda \ .$$

This our first example of the general principle that a classically scale invariant system will exhibit 'logarithmic divergences' (meaning: logarithmic dependence on the cutoff). It's the only dependence on Λ allowed by dimensional analysis.

The introduction of the cutoff can be thought of in many ways: we could say there are no momentum states with $|p| > \Lambda$ (as in a lattice model), or maybe we could say that the potential is not really a delta function if we look more closely. The choice of narrative here had better not affect our answers to physics questions at energies far below the cutoff.

More precisely:

$$\int^{\Lambda} \frac{d^2p}{\frac{p^2}{2} + \epsilon_B} = 2\pi \int_0^{\Lambda} \frac{pdp}{\frac{p^2}{2} + \epsilon_B} = 2\pi \log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right) \ .$$

So in our cutoff theory, the boundstate condition is:

$$1 = g_0 \int^{\Lambda} \frac{\text{đ}^2p}{\frac{p^2}{2} + \epsilon_B} = \frac{g_0}{2\pi\hbar^2} \log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right) \ .$$

A solution only exists for $g_0 > 0$. This makes sense since only then is the potential attractive (recall that $V = -g_0\delta$), and so only then should it have boundstates.

Now here's a trivial-seeming step that offers a dramatic new vista: solve for ϵ_B.

$$\epsilon_B = \frac{\Lambda^2}{2} \frac{1}{e^{\frac{2\pi\hbar^2}{g_0}} - 1} \ . \tag{1.4}$$

As we remove the cutoff ($\Lambda \to \infty$), we see that $E = -\epsilon_B \to -\infty$, the boundstate becomes more and more bound – the potential is too attractive.

Suppose we insist that the boundstate energy ϵ_B is a fixed thing – imagine we've *measured* it to be 200 MeV[5]. We should express everything in terms of the measured quantity. Then, given some cutoff Λ, we should solve for $g_0(\Lambda)$ to get the boundstate energy we have measured:

$$g_0(\Lambda) = \frac{2\pi\hbar^2}{\log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right)}.$$

This is the crucial step: this silly symbol g_0 that appeared in our action doesn't mean anything to anyone (see Zee's dialogue with the S.E. in section III). *We are allowing $g_0 \equiv$ the bare coupling to be cutoff-dependent.* This despite the fact that the cutoff is a purely fictional notion.

Instead of a dimensionless coupling g_0, the useful theory contains an arbitrary *dimensionful* coupling constant (here ϵ_B). This phenomenon is called *dimensional*

[5]Spoiler alert: I picked this value of energy to stress the analogy with QCD.

transmutation. The cutoff is supposed to go away in observables, which depend on ϵ_B instead.

In QCD we expect that in an identical way, an arbitrary scale Λ_{QCD} will enter into physical quantities. (If QCD were the theory of the whole world, we would work in units where it was unity.) This can be taken to be the rest mass of some mesons – boundstates of quarks. Unlike this example, in QCD there are many boundstates, but their energies are dimensionless multiplies of the one dimensionful scale, Λ_{QCD}. Nature chooses $\Lambda_{QCD} \simeq 200$ MeV.

[This dimensional transmutation phenomenon was maybe first seen in a perturbative field theory in S. Coleman, E. Weinberg, *Phys Rev* **D7** (1973) 1898. Maybe we'll come back to their example.]

There are more lessons in this example. Go back to (1.4):

$$\epsilon_B = \frac{\Lambda^2}{2} \frac{1}{e^{\frac{2\pi\hbar^2}{g_0}} - 1} \overset{g_0 \to 0}{\simeq} e^{\frac{-2\pi\hbar^2}{g_0}} \neq \sum_{n=0}^{\infty} g_0^n f_n(\Lambda)$$

it is *not* analytic (*i.e.* a power series) in $g_0(\Lambda)$ near small g_0; rather, there is an essential singularity in g_0. (All derivatives of ϵ_B with respect to g_0 vanish at $g_0 = 0$.) You can't expand the dimensionful parameter in powers of the coupling. This means that you'll never see it in perturbation theory in g_0. Dimensional transmutation is an inherently non-perturbative phenomenon.

Notice furthermore that even for moderately weak coupling, $\epsilon_B \ll \Lambda^2$. For example, when $g_0 = .1$, $\epsilon_B/\Lambda^2 = 10^{-28}$! Dimensional transmutation generates a hierarchy of scales. This phenomenon explains why the critical temperature below which metals go superconducting is much less than their intrinsic energy scale (the Fermi energy).

Look at how the bare coupling depends on the cutoff in this example:

$$g_0(\Lambda) = \frac{2\pi\hbar^2}{\log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right)} \overset{\Lambda^2 \gg \epsilon_B}{\to} \frac{2\pi\hbar^2}{\log\left(\frac{\Lambda^2}{2\epsilon_B}\right)} \overset{\Lambda^2 \gg \epsilon_B}{\to} 0$$

– the bare coupling vanishes in this limit, since we are insisting that the parameter ϵ_B is fixed. This is called *asymptotic freedom* (AF): the bare coupling goes to zero (*i.e.* the theory becomes free) as the cutoff is removed. This also happens in QCD.

RG flow equations. Define the *beta function* as (minus) the logarithmic derivative of the bare coupling with respect to the cutoff:

$$\textbf{Def:} \qquad \beta(g_0) \equiv -\Lambda \frac{\partial}{\partial \Lambda} g_0(\Lambda) \ .$$

For this theory

$$\beta(g_0) = -\Lambda \frac{\partial}{\partial \Lambda} \left(\frac{2\pi\hbar^2}{\log\left(1 + \frac{\Lambda^2}{2\epsilon_B}\right)} \right) \overset{\text{calculate}}{=} + \frac{g_0^2}{\pi\hbar^2} \left(\underbrace{1}_{\text{perturbative}} - \underbrace{e^{-2\pi\hbar^2/g_0}}_{\text{not perturbative}} \right).$$

Notice that it's a function only of g_0, and not explicitly of Λ. Also, in this simple toy theory, the perturbation series for the beta function happens to stop at order g_0^2.

β measures the failure of the cutoff to disappear from our discussion – it signals a quantum mechanical violation of scale invariance. What's β for? Flow equations:

$$\dot{g}_0 = \beta(g_0).$$

[6] This is a tautology. The dot is

$$\dot{A} = \partial_s A, \quad s \equiv -\log \Lambda/\Lambda_0 \implies \partial_s = -\Lambda \partial_\Lambda.$$

(Λ_0 is some reference scale.) But forget for the moment that this is just a definition:

$$\dot{g}_0 = + \frac{g_0^2}{\pi\hbar^2} \left(1 - e^{-2\pi\hbar^2/g_0}\right) .$$

This equation tells you how g_0 changes as you change the cutoff. Think of it as a nonlinear dynamical system. Given a dynamical system, the first question you should ask is about its fixed points.

Def: A fixed point $g_0^\star$ of a flow is a value of the coupling g_0 where the flow stops:

$$\boxed{0 = \dot{g}_0|_{g_0^\star} = \beta(g_0^\star)} \ ,$$

a zero of the beta function. (Note: if we have many couplings g_i, then we have such an equation for each g: $\dot{g}_i = \beta_i(g)$. So β_i is (locally) a vector field on the space of couplings. In that case a fixed point is a simultaneous zero of every component of the vector field.)

Where are the fixed points in our example?

$$\beta(g_0) = + \frac{g_0^2}{\pi\hbar^2} \left(1 - e^{-2\pi\hbar^2/g_0}\right) .$$

[6] Warning: The sign in this definition carries a great deal of cultural baggage. With the definition given here, the flow (increasing s) is toward the IR, toward low energy. This is opposite the high-energy particle physics convention, where the flow goes from IR to UV, I guess with the idea that we learn more physics by going to higher energies. As we will see, there is a strong argument to be made for the sign choice we've made, that the flow should be regarded as going from UV to IR, since we lose information as we move in that direction – in fact, the IR behavior does not determine the UV behavior in general, but UV does determine IR.

There's only one: $g_0^\star = 0$, near which $\beta(g_0) \sim \frac{g_0^2}{\pi\hbar}$, the non-perturbative terms are small. What does the flow look like near this point? For $g_0 > 0$, $\dot{g}_0 = \beta(g_0) > 0$. With our (correct) definition of the direction of flow, $g_0 = 0$ is a repulsive (unstable) fixed point:

```
*->->->->->->->->->->->->----------------------- g_0
```

$g_0^\star = 0$.

We already knew this. It just says $g_0(\Lambda) \sim \frac{1}{\log \Lambda^2} \to 0$ at large Λ. A lesson is that in the vicinity of such an AF fixed point, the non-perturbatuve stuff $e^{\frac{-2\pi\hbar^2}{g_0}}$ is small. So we can get good results near the fixed point from the perturbative part of β. That is: we can reliably compute the behavior of the flow of couplings near an AF fixed point *perturbatively*, and be sure that it is an AF fixed point. This is the situation in QCD.

On the other hand, the dimensional transmutation phenomenon that we've shown here is something that we can't prove in QCD. However, the circumstantial evidence is very strong!

Another example where this happens is quantum mechanics in any number of variables with a central inverse-square potential $V = -\frac{g_0^2}{r^2}$. It is also classically scale invariant:

$$[r] = -\frac{1}{2}, \qquad \left[\frac{1}{r^2}\right] = +1 \qquad \Longrightarrow \qquad [g_0] = 0.$$

This model was studied in K.M. Case, *Phys Rev* **80** (1950) 797 and you will study it on the first homework. The resulting boundstates and dimensional transmutation phenomenon are called Efimov states; this model preserves a *discrete* scale invariance.

Here's a quote from Marty Halpern from his lecture on this subject:

I want you to study this set of examples very carefully, because it's the only time in your career when you will understand what is going on.

In my experience it's been basically true. For real QFTs, you get distracted by Feynman diagrams, gauge invariance, regularization and renormalization schemes, and the fact that you can only do perturbation theory.

[End of Lecture 1]

1.2 Field theory from balls and springs via the path integral

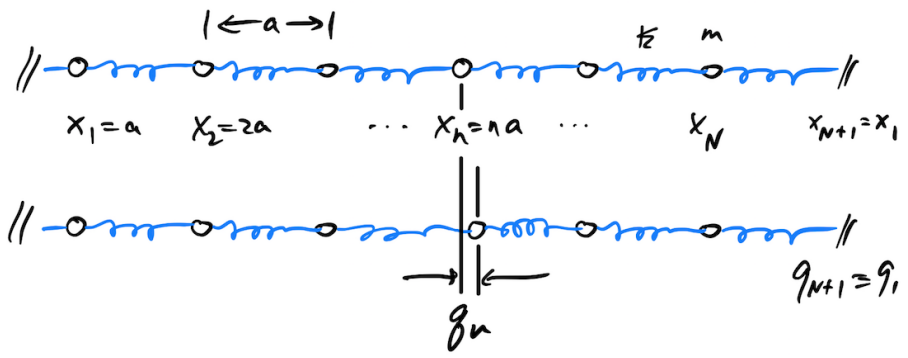
From particles to fields to particles again. Here is a way to discover QFT starting with some prosaic ingredients.

Consider a linear chain of particles of mass m, each connected to its neighbors by springs with spring constant κ. This is a model of a (one-dimensional) crystalline solid. When in equilibrium, the masses form a regular one-dimensional crystal lattice (equally spaced mass points). Now let q_n denote the displacement of the nth mass from its equilibrium position x_n and let p_n be the corresponding momentum. Assume there are N masses and (for simplicity) impose periodic boundary conditions: $q_{n+N} = q_n$. The equilibrium positions themselves are

$$x_n = na, n = 1, 2...N$$

where a is the lattice spacing. The Hamiltonian for the collection of particles is:

$$\mathbf{H} = \sum_{n=1}^{N} \left(\frac{\mathbf{p}_n^2}{2m} + \frac{1}{2}\kappa \left(\mathbf{q}_n - \mathbf{q}_{n-1}\right)^2 \right) + \lambda \mathbf{q}^4. \tag{1.5}$$



Notice that this is an ordinary quantum mechanics system, made of *particles*. In particular, the whole story below will take place within the fixed Hilbert space of the positions of the N particles.

I've included a token anharmonic term $\lambda \mathbf{q}^4$ to remind us that we are leaving stuff out; for example we might worry whether we could use this model to describe *melting*.

Set $\lambda = 0$ for a while. With $\lambda = 0$, the hamiltonian above describes a collection of coupled harmonic oscillators, with a matrix of spring constants $V = \kappa_{ab}\mathbf{q}_a\mathbf{q}_b$. If we diagonalize the matrix of spring constants, we will have a description in terms of decoupled oscillators, called *normal modes*. Because the chain is translation invariant, the normal modes are labelled by a wavenumber k, and the eigenvalues are $\omega_k^2 = \frac{2\kappa}{m}\sin^2\frac{ka}{2}$. (This step is purely a classical mechanics problem, or alternatively, the

problem of diagonalizing the matrix of spring constants κ_{ab}.) The change of variables that diagonalizes the matrix of spring constants is:

$$\mathbf{q}_n = \sqrt{\frac{\hbar}{2\mu}} \sum_k \frac{1}{\sqrt{\omega_k}} \left(e^{\mathbf{i}kx_n} \mathbf{a}_k + e^{-\mathbf{i}kx_n} \mathbf{a}_k^\dagger \right) + \frac{1}{\sqrt{N}} \mathbf{q}_0. \tag{1.6}$$

Then we can use our knowledge of the spectrum of a single SHO $H = \hbar\omega \left(a^\dagger a + \frac{1}{2} \right)$ to construct the whole spectrum of excitations of the chain,

$$H = \sum_k \hbar\omega_k \left(a_k^\dagger a_k + \frac{1}{2} \right) + \frac{p_0^2}{2m}.$$

(Here p_0 is the center-of-mass momentum of the chain.) The groundstate is $|0\rangle$, the state annihilated by all the annihilation operators $a_k |0\rangle = 0$, and excited states are built like $|k_1, k_2\rangle = a_{k_1}^\dagger a_{k_2}^\dagger |0\rangle$. In the context of an elastic solid, these excitations are called *phonons*.

Instead, let's use the path integral.

Path integral reminder in a box. [A useful reference is Zee §I.2] If we use the path integral description, some of these things (in particular the continuum, sound-wave limit) are more obvious-seeming.

Let's remind ourselves how the path integral formulation of QM works for a particle in one dimension with $\mathbf{H} = \frac{\mathbf{p}^2}{2m} + V(\mathbf{q})$. The basic statement is the following formula for the propagator – the amplitude to propagate from position eigenstate $|q_0\rangle$ to position eigenstate $|q\rangle$ during a time interval t is

$$\langle q | \, e^{-\mathbf{i}\mathbf{H}t} \, |q_0\rangle = \int_{q(0)=q_0}^{q(t)=q} [dq] e^{\mathbf{i} \int_0^t \mathrm{d}s \, \left(\frac{1}{2}\dot{q}^2 - V(q) \right)} \, .$$

Here $[dq] \equiv \mathcal{N} \prod_{l=1}^{M_t} dq(t_l)$ – the path integral measure is defined by a limiting procedure ($M_t \equiv \frac{t}{\Delta t} \to \infty, \Delta t \to 0, t$ fixed), and $\mathcal{N}$ is a normalization factor that always drops out of physical quantities so I don't need to tell you what it is.

Recall that the key step in the derivation of this statement is the evaluation of the propagator for an infinitesimal time step:

$$\langle q_2 | \, e^{-\mathbf{i}\mathbf{H}\Delta t} \, |q_1\rangle = \langle q_2 | \, e^{-\mathbf{i}\Delta t \frac{\mathbf{p}^2}{2m}} e^{-\mathbf{i}\Delta t V(\mathbf{q})} \, |q_2\rangle + \mathcal{O}(\Delta t^2) \, .$$

An integral expression for this involving only numbers (no operators, no states) can be obtained by inserting resolutions of the identity

$$\mathbb{1} = \mathbb{1}^2 = \left(\int đp |p\rangle\langle p| \right) \left(\int \mathrm{d}q |q\rangle\langle q| \right) = \int đp \int dq |p\rangle\langle q| e^{ipq}$$

in between the two exponentials. For a more extensive reminder, please see §2.4 of this document.

Two quick but invaluable applications of the path integral:

1. The path integral *explains* the origin of the variational principle and the special role of configurations that solve the equations of motion. They are points of stationary phase in the path integral, where

$$0 = \frac{\delta S}{\delta q(t)}. \tag{1.7}$$

In case you are not familiar with functional derivatives: From the definition of the path integral as a limit it should be clear what is meant by this expression. If we didn't take the limit, the stationary phase condition is that S is extremized in every direction $q_i = q(t_i)$:

$$0 = \frac{\partial S}{\partial q_i}.$$

The functional derivative in (1.7) means the same thing but looks nicer because we can pretend everything is smooth rather than discrete. Since q_i and q_j are independent variables,

$$\frac{\partial q_i}{\partial q_j} = \delta_{ij} \qquad \text{and therefore} \qquad \frac{\delta q(t)}{\delta q(s)} = \delta(t-s).$$

With the help of the chain rule, this is the only functional derivative you need to know how to take[7].

Part of this statement is a discovery of the true role of $\hbar$: its job is to make up the units so that the exponent in $e^{\mathbf{i}S[q]/\hbar}$ is dimensionless.

2. Above I wrote a formula for the real-time propagator. Euclidean path integrals are also very useful, because they compute ground-state expectation values. Here's why:

The vacuum can be prepared by starting in an arbitrary state and acting with e^{-TH} for some large T, and then normalizing (as usual when discussing path integrals, it's best to not worry about the normalization and only ask questions that don't depend on it),

$$|0\rangle = \mathcal{N} e^{-\mathbf{H}T} |\text{any}\rangle .$$

[7] If you are not yet comfortable with the machinery of functional derivatives, please work through pages 2-28 through 2-30 of this document right now.

To see this, just expand in the energy eigenbasis. This 'imaginary time evolution operator' $e^{-\mathbf{H}T}$ has a path integral representation just like the real time operator, by nearly the same calculation

$$e^{-\mathbf{H}T} = \int [Dq] e^{-\int_{-T}^{0} d\tau L(q(\tau),\dot{q}(\tau))} |q(0)\rangle\langle q(-T)|.$$

Doing the same thing to prepare $\langle 0|$, making a sandwich of $f(\mathbf{q})$ and taking $T \to \infty$ we can forget about the arbitrary states at the end, and we arrive at

$$\langle f(\mathbf{q})\rangle \equiv \frac{\langle 0| f(\mathbf{q}) |0\rangle}{\langle 0|0\rangle} = \frac{1}{Z}\int [Dq] e^{-S_E} f(q(0)) = \frac{1}{Z}\int \prod_i^{M_t} dq_i \; e^{-\sum_j L(q_j)} f(q_0) \tag{1.8}$$

where $Z \equiv \int \prod_i dq_i \; e^{-\sum_j L(q_j)}$ (here i, j are discrete time labels) and $S_E[q]$ is the euclidean action. Relative to the real time action, there is an important sign[8]:

$$S_E[q] = \int d\tau \left((\partial_\tau q)^2 + V(q) \right).$$

This sign makes good sense – it means that configurations with large V are *suppressed* in the euclidean path integral.

This trick of using imaginary time evolution to prepare the groundstate is a crucial one that we will use all the time below[9].

In the special case where L is quadratic in q and $\dot{q}$, this can be written as

$$Z = \int \prod_i dq_i \; e^{-q_i D_{ij} q_j}$$

where D_{ij} is the (real, symmetric) matrix which discretizes the action. Repeated indices are summed. This is a gaussian integral. This is the path integral point of view on why quadratic terms are special.

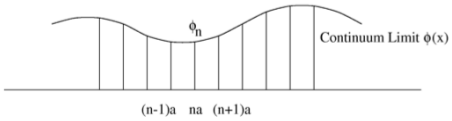
This discussion of path integrals has been for a single degree of freedom q. But as we've seen, we can make a field theory by coupling together a collection of such degrees of freedom q_n. The path integral measure for a field theory is just the product of the measures for each of these q_n – it's just a matter of sticking more labels on the integration variables.

[8] By the way, I use square brackets for $S[q]$ to remind myself that S is a function*al* of q, a machine that eats functions and spits out numbers – for each function $q(\tau)$, $S[q]$ is a number. But in fact my main point here is that for our purposes there is not such a big difference between a functional and a function of many variables (the values of the function at a discrete collection of points).

[9] We don't know how to do imaginary time evolution in the laboratory. Here is one recent attempt to do it using measurements. A reason to believe that it shouldn't be possible to do this *efficiently* is that it can be used to solve NP complete problems (by encoding their solution into the groundstate of a Hamiltonian).

Scalar field theory in one dimension. [Zee §1.3] The (real-time) path integral for our collection of oscillators is

$$Z = \int [dq_1 \cdots dq_N] e^{\mathbf{i}S[q]}$$

with $S[q] = \int dt \left(\sum_n \frac{1}{2} m_n \dot{q}_n^2 - V(\{q\}) \right) \equiv \int dt L(q, \dot{q})$. The potential is $V(\{q\}) = \sum_n \frac{1}{2} \kappa \left(q_{n+1} - q_n \right)^2$. Now suppose we have poor eyesight and can't resolve the individual atoms in the chain; rather we're only interested in the *long-wavelength* (small-wavenumber) physics. Sometimes this is described as taking a continuum limit, $a \to 0, N \to \infty$. Basically the only thing we need is to think of $q_n = q(x = na)$ as defining a smooth function: 

[Note that the continuum field is often called $\phi(x)$ instead of $q(x)$ for some reason. At least the letters $q(x)$ and $\phi(x)$ look similar.]

By Taylor expanding $q(x_{n-1})$ about $q(x_n)$, we have

$$\left(q_n - q_{n-1} \right)^2 \simeq a^2 \left(\partial_x q \right)^2 |_{x=na}, \quad a \sum_n f(q_n) \simeq \int dx f(q(x)).$$

The path integral becomes:

$$Z = \int [dq] e^{\mathbf{i}S[q]}$$

with $[dq]$ now representing an integral over all configurations $q(t, x)$ (*defined* by this limit) and

$$S[q] = \int dt \int dx \frac{1}{2} \left(\mu \left(\partial_t q \right)^2 - \mu v_s^2 \left(\partial_x q \right)^2 - r q^2 - u q^4 - ... \right) \equiv \int dt \int dx \mathcal{L} \qquad (1.9)$$

where I've introduced some parameters μ, v_s, r, u determined from $m, \kappa, a...$ in some ways that we needn't worry about, except to say that they are finite in the continuum limit. The $\cdots$ includes terms like $a^4 \left(\partial_x q \right)^4$ that are small when $k \ll \frac{1}{a}$, so we ignore them. $\mathcal{L}$ is the Lagrangian *density* whose integral over space is the Lagrangian $L = \int dx \mathcal{L}$.

The equation of motion is the stationary phase condition,

$$0 = \frac{\delta S}{\delta q(x,t)} = -\mu \ddot{q} + \mu v_s^2 \partial_x^2 q - r q - 2 u q^3 - ...$$

In this expression I have written a functional derivative; with our lattice regulator, it is simply a(n extremely useful) shorthand notation for the collection of partial derivatives $\frac{\partial}{\partial q_n}$.

From the phonon problem, we automatically find $r = u = 0$, and the equation of motion is just the wave equation:

$$0 = -\mu\partial_t^2 q + v_s^2 \partial_x^2 q. \tag{1.10}$$

This happened because of the symmetry $q_n \to q_n + \epsilon$. This is the operation that *translates* the whole crystal. It guarantees low-energy phonons near $k = 0$ because it means $q(x)$ can only appear in S via its derivatives. (This is a general property of *Goldstone modes*; more on this later.)

These expressions are easy to generalize to arbitrary dimension, just let the field depend on more coordinates, and replace ∂_x with the gradient operator.

The lattice spacing a and the size of the box Na in the discussion above are playing very specific roles in *regularizing* our 1-dimensional scalar field theory. The lattice spacing a implies a maximum wavenumber or shortest wavelength and so is called an "ultraviolet (UV) cutoff", because the UV is the short-wavelength end of the visible light spectrum. The size of the box Na implies a maximum wavelength mode which fits in the box and so is called an "infrared (IR) cutoff".

If we also take the infinite volume limit, then the sums over k become integrals. In this limit we can make the replacement

$$\frac{1}{L^d}\sum_k \rightsquigarrow \int \text{đ}^d k, \quad L^d \delta_{kk'} \rightsquigarrow (2\pi)^d \delta^{(d)}(k - k').$$

A check of the normalization factors comes from combining these two rules

$$\forall k',\ 1 = \sum_k \delta_{k,k'} = \int \text{đ}^d k (2\pi)^d \delta^{(d)}(k - k').$$

Fields mediate forces. [Zee §1.3] To see a nice and simple application of the path integral, consider again our chain of balls on springs. Suppose a giant hand reaches in and pushes the atom at position x_n a little bit. This can be described by adding to the hamiltonian a term

$$\delta V(q) = -J_n(t) q_n(t)$$

which applies a force $J_n(t)$ to the nth atom. We can ask, in the presence of such a force, what is the amplitude to go from state I to state F in time T:

$$\langle F|\, e^{-\mathbf{i}\int_0^T dt H(t)}\, |I\rangle = \int_{F\leftarrow I} [D\phi] e^{\mathbf{i}\int dt d^d x \left(\frac{1}{2}(\partial\phi)^2 - V(\phi) + J(x)\phi(x)\right)}.$$

As you see, this is a quantity for which we have a path integral representation. Let's think in particular about the case where $|F\rangle = |I\rangle = |0\rangle$, the groundstate. In that case,

the boundary conditions on the path integral are: first do some euclidean evolution to prepare $|0\rangle$, then evolve in real time, then do some euclidean evolution to prepare $\langle 0|$.

Here's a reason we might care about this quantity: take the initial and final states to be the groundstate:

$$\langle 0|\, e^{-\mathbf{i}\int_0^T dt H(t)}\, |0\rangle \simeq e^{-\mathbf{i}\int_0^T dt E_{\rm gs}(J)}.$$

If the time-dependence is slow enough, the answer is obtained by the adiabatic approximation: just add up the instantaneous groundstate energy at each time step. So this computes the J-dependent groundstate energy.

Let's retreat to the case where the action is quadratic in ϕ, so that we can actually do the path integral:

$$\mathcal{L}(\phi) = \frac{1}{2}\left(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2\right) \overset{\text{IBP}}{=} -\frac{1}{2}\phi\left(\partial^2 + m^2\right)\phi + \text{total derivative}. \tag{1.11}$$

Going back to the lattice to make the integrals slightly less scary, we have

$$e^{\mathbf{i}W[J]} \equiv \int [D\phi] e^{\mathbf{i}\int(\mathcal{L}+J\phi)} = \int_{-\infty}^{\infty} \prod_{n,t}^{M_t,N} dq_{n,t} e^{\frac{\mathbf{i}}{2}q_x A_{xy} q_y + \mathbf{i}J_x q_x} = \sqrt{\frac{(2\pi\mathbf{i})^{NM_t}}{\det A}} e^{-\frac{\mathbf{i}}{2}J_x A^{-1}_{xy} J_y}.$$

Here repeated indices are summed as usual: $q_x A_{xy} q_y = \int dxdy\phi(x)A_{xy}\phi(y)$, etc... So you can see that the matrix A multiplying the quadratic term in this gaussian integral is $A_{xy} = -\delta^{d+1}(x-y)\left(\partial_x^2 + m^2\right)$. It is an $NM_t \times NM_t$ matrix. Its inverse A^{-1} satisfies by definition $A_{xz}A^{-1}_{zy} = \delta_{xy}$, which is the differential equation

$$-\left(\partial^2 + m^2\right)D(x-y) = \delta(x-y). \tag{1.12}$$

This equation says that D is a *Green's function* for the operator $-(\partial^2 + m^2)$. The fact that there is no special point in spacetime says $A^{-1}_{xy} = D(x-y)$ only depends on the difference of its arguments.

Does this integral actually converge? An integral of the form $\int_{\mathbb{R}} dq e^{-\frac{1}{2}qAq}$ surely converges if A is a positive matrix. Actually, this is overkill – it is enough to replace $m^2 \to m^2 - \mathbf{i}\epsilon$ to make all the integrals converge. Here ϵ is an *infinitesimal*, which means $\epsilon^2 = 0$ and $c\epsilon = \epsilon$ for any positive c. Then each $\int dq_{nt}$ will have a factor of $e^{-\epsilon\int q_{nt}^2}$ which suppresses the integrand in the dangerous large-field region[10].

The equation (1.12) is translation-invariant and linear so you should not be surprised that it is solved by going to Fourier space (in space and time):

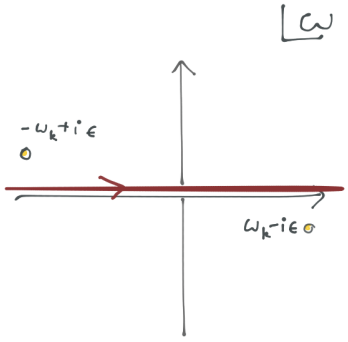
$$D(x) = \int \mathrm{đ}^{d+1}k\; e^{\mathbf{i}k_\mu x^\mu} D_k, \quad \delta^{d+1}(x) = \int \mathrm{đ}^{d+1}k\; e^{\mathbf{i}k_\mu x^\mu}.$$

[10]Here I have shown you *one* way to make the integral well-defined. You might worry that there could be others (there are). Another thing you might be bothered by is the boundary conditions on the fields and their relation to the initial and final states. These issues are closely related!

in terms of which (1.12) becomes the algebraic equation $1 = (k^2 - m^2 + \mathbf{i}\epsilon)D_k$. Hence

$$D(x) = \int \mathrm{d}^{d+1}k \frac{e^{\mathbf{i}kx}}{k^2 - m^2 + \mathbf{i}\epsilon}.$$

Notice that the shift by ϵ saves the day here: it keeps the integration contour from running right over the pole at $k^2 = m^2$, by moving slightly in the imaginary direction. More explicitly,



$$k^2 - m^2 + \mathbf{i}\epsilon = \omega^2 - \vec{k}^2 - m^2 + \mathbf{i}\epsilon$$

is zero when

$$\omega = \pm\sqrt{\vec{k}^2 + m^2 - \mathbf{i}\epsilon} \overset{\text{Taylor}}{=} \pm(\omega_k - \mathbf{i}\epsilon), \quad \omega_k \equiv \sqrt{\vec{k}^2 + m^2}.$$

In the second step I Taylor expanded $\sqrt{\omega_k^2 - \mathbf{i}\epsilon} = \sqrt{\omega_k^2} - \frac{\mathbf{i}\epsilon}{\omega_k} + \mathcal{O}(\epsilon)^2$ and used the facts that $\omega_k > 0$, and that anything positive times an infinitesmal is an infinitesimal.

We can then do the ω integral by contours[11]: if $t > 0$ ($t < 0$), we can close the contour in the UHP (LHP) since the integrand goes like $e^{-\mathrm{Im}\,\omega t}$, and the integral equals the residue of the pole at $\omega = \omega_k \mp \mathbf{i}\epsilon$ (times $2\pi\mathbf{i}$):

$$D(x) = -\mathbf{i}\int \mathrm{d}^d k \left(\theta(t)\frac{e^{-\mathbf{i}(\omega_k t - \vec{k}\cdot\vec{x})}}{2\omega_k} + \theta(-t)\frac{e^{\mathbf{i}(\omega_k t - \vec{k}\cdot\vec{x})}}{2\omega_k}\right). \tag{1.13}$$

We'll learn to call this *time-ordered* in a moment.

The propagator. Who is $D(x)$, besides some quantity in terms of which we did a Gaussian integral? The inverse matrix can be extracted via a two-point correlation function:

$$A^{-1} = \int dq q^2 e^{-\frac{1}{2}qAq}/Z.$$

Putting back all the labels, the same manipulations show that

$$D(x - y) \overset{?}{=} \langle 0|\phi(x)\phi(y)|0\rangle \tag{1.14}$$

– the amplitude to propagate an excitation created from the vacuum by $\phi(x)$ to be annihilated by $\phi(y)$. The propagator, for short.

(Notice that if the system is Lorentz invariant (which starting from (1.11) it is) then since $D(x)$ is a scalar quantity, it can only depend on x through Lorentz invariants made from x^μ, namely the proper distance $x^2 = t^2 - \vec{x}^2$, and the sign of t.)

[11] We are using the Cauchy residue theorem $\oint_C dz f(z) = 2\pi\mathbf{i}\sum_{z_j} \mathrm{Res}_{z=z_j} f$ where z_j are the poles of f. To remember the sign, consider a small circle C_0 counterclockwise around the origin and $f(z) = 1/z$, so $\oint_{C_0} \frac{dz}{z} = \mathbf{i}\int_0^{2\pi} d\theta = 2\pi\mathbf{i}$.

Why the '?' in (1.14)? For one thing, $\phi(x)$ and $\phi(y)$ are operators – the order matters. How do I know which order in which to write them? To reproduce (1.13) the thing to do is to *time-order* them:

$$\langle 0|\mathcal{T}\phi(x)\phi(y)|0\rangle \equiv \theta(x^0-y^0)\,\langle 0|\phi(x)\phi(y)|0\rangle + \theta(y^0-x^0)\,\langle 0|\phi(y)\phi(x)|0\rangle\,.$$

One way to see this is to plug in the mode expansion of the free scalar

$$\phi(x) = \int \frac{\text{đ}^d k}{\sqrt{2\omega_k}}\left(e^{\mathbf{i}k^\mu x_\mu}\mathbf{a}_k + e^{-\mathbf{i}k^\mu x_\mu}\mathbf{a}_k^\dagger\right)|_{k_0=\omega_{\vec{k}}}, \tag{1.15}$$

to see *e.g.*

$$\langle 0|\phi(x)\phi(y)|0\rangle = \int \frac{\text{đ}^d k\,\text{đ}^d q}{2\sqrt{\omega_k\omega_q}} e^{-\mathbf{i}kx+\mathbf{i}qy}\,\langle 0|\,a_k a_q^\dagger\,|0\rangle = \int \frac{\text{đ}^d k}{2\omega_k} e^{-\mathbf{i}k(x-y)}$$

(where $k^0=\omega_k, q^0=\omega_q$ to satisfy the KG equation), which reproduces the first term in (1.13)[12][13].

Now why should we care about the propagator? Look again at $W[J]$. We've learned that (up to terms independent of J),

$$W[J] = -\frac{1}{2}\int\int d^{d+1}x d^{d+1}y J(x)D(x-y)J(y) = -\frac{1}{2}\int \text{đ}^{d+1}k J_k^\star \frac{1}{k^2-m^2+\mathbf{i}\epsilon}J_k$$

Here $J(x)=\int \text{đ}^{d+1}k e^{\mathbf{i}kx}J_k$, $J_k^\star = J_{-k}$ (since $J(x)$ is real).

We get to pick $J(x)$. Let's choose $J=J_1+J_2$ to describe (in Zee's words) two lumps sitting still on the mattress: $J_a(x)=\delta^3(x-x_a), a=1,2$. Then $J_k = \int dx^0 e^{-\mathbf{i}k^0x^0}\left(e^{\mathbf{i}\vec{k}\cdot\vec{x}_1}+e^{\mathbf{i}\vec{k}\cdot\vec{x}_2}\right)$. The interaction between the two lumps *mediated* by the mattress field ϕ will then be described by the J_1J_2 cross-terms in $W[J]$:

$$W[J] = -\frac{2}{2}\int dx^0\int dy^0\int \text{đ}k^0 e^{\mathbf{i}k^0(x^0-y^0)}\int \text{đ}^3k\frac{e^{\mathbf{i}\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{k^2-m^2+\mathbf{i}\epsilon}+\ldots \tag{1.16}$$

$$= -\int dx^0 \underbrace{\left(\int \text{đ}k^0 2\pi\delta(k^0)\right)}_{=1}\int \text{đ}^3k\frac{e^{\mathbf{i}\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{k^2-m^2+\mathbf{i}\epsilon}+\ldots \tag{1.17}$$

$$= +\int dx^0\int \text{đ}^3k\frac{e^{\mathbf{i}\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{\vec{k}^2+m^2-\mathbf{i}\epsilon}+\ldots \tag{1.18}$$

[12] The other ways of making the path integral well-defined correspond to other ways of ordering the ϕs, and other initial and final states.

[13] In comparing to (1.13), it helps to notice that we can redefine the $\vec{k}$ integration variable to reverse the sign of the exponent of the spatial part, $\int \text{đ}^d k f(\vec{k}^2)e^{\mathbf{i}\vec{k}\cdot\vec{x}} = \int \text{đ}^d k f(\vec{k}^2)e^{-\mathbf{i}\vec{k}\cdot\vec{x}}$. (Thanks to Hung-Hwa Lin for help during lecture.)

(The ... indicate terms which don't depend on x_1, x_2, so let's ignore them.)

For this choice of J, the Hamiltonian is time-independent, and $e^{\mathbf{i}W} = \langle 0|\, e^{-\mathbf{i}HT}\, |0\rangle = e^{-\mathbf{i}E_{\rm gs}(J)T}$, so $W = -E_{\rm gs}(J)T$. We learn that

$$E_{\rm gs}(J) = -\int \text{đ}^d k \frac{e^{\mathbf{i}\vec{k}\cdot\vec{x}_{12}}}{\vec{k}^2 + m^2} + \text{const.}$$

Notice that we can drop the $\mathbf{i}\epsilon$ now, because this integrand is nonsingular for real $\vec{k}$. In $d = 1$, there are poles at $k = \pm\mathbf{i}m$, and we can close the contour in the UHP for free to get[14]

$$E_{\rm gs}(J) = -\frac{2\pi\mathbf{i}}{2\pi}\frac{e^{-mx}}{2\mathbf{i}m} = -\frac{e^{-mx}}{2m}.$$

Since x is the separation between the lumps, this means that our field has produced an *attractive* force between the lumps

$$F = -\partial_x E_{\rm gs}(J) = +\frac{1}{2}e^{-mx}$$

which falls off exponentially with the separation between the lumps. The range of the potential goes inversely with the mass of the 'force carrier' ϕ. The 3d version of this potential $\frac{e^{-mr}}{r}$ (see footnote 14) is called the Yukawa potential. [End of Lecture 2]

1.3 Casimir effect: vacuum energy is real

[A. Zee, *Quantum Field Theory in a Nutshell*, §I.9]

This subsection has two purposes. The main purpose is to give an object lesson in asking the right questions. In physics, the right question is often a question that can be answered by an experiment, at least in principle. The answers to such questions are less sensitive to our silly theoretical prejudices, *e.g.* about what happens to physics at very short distances.

[14] For convenience, here's the integral in 3d:

$$\begin{aligned}\int \text{đ}^3 k \frac{e^{\mathbf{i}\vec{k}\cdot\vec{x}}}{\vec{k}^2 + M^2} &\overset{y\equiv\cos\theta}{=} \frac{1}{(2\pi)^2}\int_0^\infty \frac{k^2 dk}{k^2+M^2}\underbrace{\int_{-1}^{1} dy e^{\mathbf{i}kyr}}_{=2\frac{\sin kr}{kr}} = \frac{1}{(2\pi)^2 r}\int_{-\infty}^{\infty}\frac{dk k\sin kr}{k^2+M^2} \\ &= \frac{1}{(2\pi)^2 r}\left(\frac{1}{2\mathbf{i}}\int_{-\infty}^{\infty} dk \frac{k e^{\mathbf{i}kr}}{k^2+M^2} + h.c.\right) \quad \text{close contour in UHP for free} \\ &= \frac{1}{(2\pi)^2 r}\frac{1}{2\mathbf{i}} 2\pi\mathbf{i}\frac{\mathbf{i}M e^{\mathbf{i}(\mathbf{i}M)r}}{2\mathbf{i}M}\cdot 2 = \frac{e^{-Mr}}{4\pi r}.\end{aligned}$$

A second purpose is to show that the energy of the vacuum of a free scalar field $\sum_k \hbar\omega_k$ (or indeed the $\frac{1}{2}\hbar\omega$ energy of the vacuum of the quantum harmonic oscillator) is real. Sometimes we can get rid of it by choosing the zero of energy (which doesn't matter unless we are studying dynamical gravity). But it is meaningful if we can *vary* ω (or the collection of ωs if we have many oscillators as for the radiation field) and compare the difference.

In the context of the bunch of oscillators making up the radiation field, we can change the spectrum of frequencies of these oscillators $\{\omega_k\}$ by putting it in a box and varying the size of the box. In particular, two parallel conducting plates separated by some distance d experience an attractive force from the change in the vacuum energy of the EM field resulting from their presence. The plates put boundary conditions on the field, and therefore on which normal modes are present.

To avoid some complications of E&M which are not essential for our point here, we're going to make two simplifications:

- we're going to solve the problem in 1+1 dimensions
- and we're going to solve it for a scalar field.

To avoid the problem of changing the boundary conditions *outside* the plates we use the following device with *three* plates:

$$| \leftarrow d \rightarrow | \longleftarrow \qquad L-d \qquad \longrightarrow |$$

(We will consider $L \gg d$, so we don't really care about the far right plate.) The 'perfectly conducting' plates impose the boundary condition that our scalar field $q(x)$ vanishes there. The normal modes of the scalar field $q(x)$ in the left cavity are then

$$q_j = \sin\left(j\pi x/d\right), \quad j = 1, 2, ...$$

with frequencies $\omega_j = \frac{\pi|j|}{d}c$. There is a similar expression for the modes in the right cavity, which we won't need. We're going to add up all the $\frac{1}{2}\hbar\omega$s for all the modes in both cavities to get the vacuum energy $E_0(d)$; the force on the middle plate is then $-\partial_d E_0$.

The vacuum energy in the whole region of interest between the outer plates is the sum of the vacuum energies of the two cavities

$$E_0(d) = f(d) + f(L-d)$$

where

$$f(d) = \frac{1}{2}\hbar\sum_{j=1}^{\infty}\omega_j = \hbar c\frac{\pi}{2d}\sum_{j=1}^{\infty} j \overset{?!?!!}{=} \infty.$$

We have done something wrong. What?

Our crime is hubris: we assumed that we knew what the modes of arbitrarily large mode number k (arbitrarily short wavelength, arbitrarily high frequency) are doing, and in particular we assumed that they cared about our silly plates. In fact, no metal in existence can put boundary conditions on the modes of large enough frequency – those modes don't care about d. The reason a conductor puts boundary conditions on the EM field is that the electrons move around to compensate for an applied field, but there is a limit on how fast the electrons can move (*e.g.* the speed of light). The resulting cutoff frequency is called the *plasma frequency* but we don't actually need to know about all these details. To parametrize our ignorance of what the high-frequency modes do, we must *cut off* (or *regularize*) the contribution of the high-frequency modes. Let's call modes with $\omega_j \gg \pi/a$ high frequency, where a is some short time[15], $a \ll d$. Replace

$$\begin{aligned} f(d) \rightsquigarrow f(a,d) &= \hbar \frac{\pi}{2d} \sum_{j=1}^{\infty} e^{-a\omega_j/\pi} j \\ &= -\frac{\pi\hbar}{2} \partial_a \underbrace{\left(\sum_{j=1}^{\infty} e^{-aj/d} \right)}_{= \frac{1}{1-e^{-a/d}} - 1} \\ &= +\frac{\pi\hbar}{2d} \frac{e^{a/d}}{\left(e^{a/d}-1\right)^2} \\ &\overset{a \ll d}{\simeq} \hbar \left(\underbrace{\frac{\pi d}{2a^2}}_{\to\infty \text{ as } a\to 0} - \frac{\pi}{24d} + \frac{\pi a^2}{480 d^3} + ... \right) \end{aligned} \tag{1.19}$$

Answers which don't depend on a have a chance of being meaningful. The thing we can measure is the force:

$$\begin{aligned} F &= -\partial_d E_0 = -\left(f'(d) - f'(L-d)\right) \\ &= -\hbar \left(\left(\frac{\pi}{2a^2} + \frac{\pi}{24d^2} + \mathcal{O}(a^2) \right) - \left(\frac{\pi}{2a^2} + \frac{\pi}{24\left(L-d\right)^2} + \mathcal{O}(a^2) \right) \right) \\ &\overset{a\to 0}{=} -\frac{\pi\hbar}{24} \left(\frac{1}{d^2} - \frac{1}{(L-d)^2} \right) \\ &\overset{L \gg d}{=} -\frac{\pi\hbar c}{24 d^2} \left(1 + \mathcal{O}(d/L)\right) \ . \end{aligned} \tag{1.20}$$

[15] You can think of a as the time it takes the waves to move by one lattice spacing. If we work in units where the velocity is $c = 1$, this is just the lattice spacing. I will do so for the rest of this discussion.

This is an attractive force between the plates. (I put the c back in the last line by dimensional analysis.)

The analogous force between real conducting plates, caused by the change of boundary conditions on the electromagnetic field, has been measured.

The string theorists will tell you that $\sum_{j=1}^{\infty} j = -\frac{1}{12}$, and our calculation above agrees with them in some sense. But what this foolishness means is that if we compute something that is not dependent on the cutoff we have to get the same answer no matter what cutoff we use. Notice that it is crucial to ask the right questions.

An important question is to what extent could we have picked a different cutoff function (instead of $e^{-\pi\omega/a}$) and gotten the same answer for the physics. This interesting question is answered affirmatively in Zee's wonderful book, 2d edition, section I.9 Appendix (available electronically here!). Go study this appendix now.

A comment about possible physical applications of the calculation we actually did: you could ask me whether there is such a thing as a Casimir force due to the vacuum fluctuations of *phonons*. Certainly it's true that the boundary of a chunk of solid puts boundary conditions on the phonon modes, which change when we change the size of the solid. The problem with the idea that this might produce a measurable force (which would lead the solid to want to shrink) is that it is hard to distinguish the 'phonon vacuum energy' from the rest of the *energy of formation* of the solid, that is, the energy difference between the crystalline configuration of the atoms and the configuration when they are all infinitely separated. Certainly the latter is not well-described in the harmonic approximation.

Lest this seem too disconnected from the other 'divergences' we will encounter in QFT associated with loop amplitude, let me emphasize that the vacuum energy *is* a loop amplitude in this free field theory (the only one). Specifically

$$E_0 = \langle 0|H_0|0\rangle = \langle 0| \int d^d x\ \phi(x) \left(-\partial_t^2 + \nabla^2 + m^2\right) \phi(x) \,|0\rangle =$$  .

A few comments about the 3+1 dimensional case of E&M. Assume the size of the plates is much larger than their separation L. Dimensional analysis shows that the force per unit area from vacuum fluctuations must be of the form

$$P = A\frac{\hbar c}{L^4}$$

where A is a numerical number. A is not zero!

Use periodic boundary conditions in the xy planes (along the plates). The allowed

wave vectors are then

$$\vec{k} = \left(\frac{2\pi n_x}{L_x}, \frac{2\pi n_y}{L_y} \right)$$

with n_x, n_y integers.

We have to do a bit of E&M here. Assume the plates are perfect conductors (this where the hubris about the high-frequency modes enters). This means that the transverse component of the electric field must vanish at the surface. Instead of plane waves in z, we get standing waves: $\phi(z) \propto \sin(n\pi z/L)$.

The frequencies of the associated standing waves are then

$$\omega_n(\vec{k}) = c\sqrt{\frac{\pi^2 n^2}{L^2} + \vec{k}^2}, n = 0, 1, 2, \cdots$$

Also, there is only one polarization state for $n = 0$.

So the zero-point energy is

$$E_0(L) = \frac{\hbar}{2} \left(2 \sum_{n,\vec{k}}{}' \omega_n(\vec{k}) \right)$$

where it's useful to define

$$\sum_{n,\vec{k}}{}' \equiv \frac{1}{2} \sum_{n=0,\vec{k}} + \sum_{n\geq 1,\vec{k}}$$

Now you can imagine introducing a regulator like the one we used above, and replacing

$$\sum_{n,\vec{k}}{}' \cdot \rightsquigarrow \sum_{n,\vec{k}}{}' e^{-a\omega_n(\vec{k})/\pi}.$$

and doing the sums and integrals and extracting the small-a behavior.

---

2 To infinity and beyond

2.1 A simple example of perturbative renormalization in QFT

[Zee §III.1, Schwartz §15.4] Now let's consider an actual, interacting field theory in $D > 0+1$ dimensions, but a simple one, namely the theory of a real scalar field in 3+1 dimensions, with

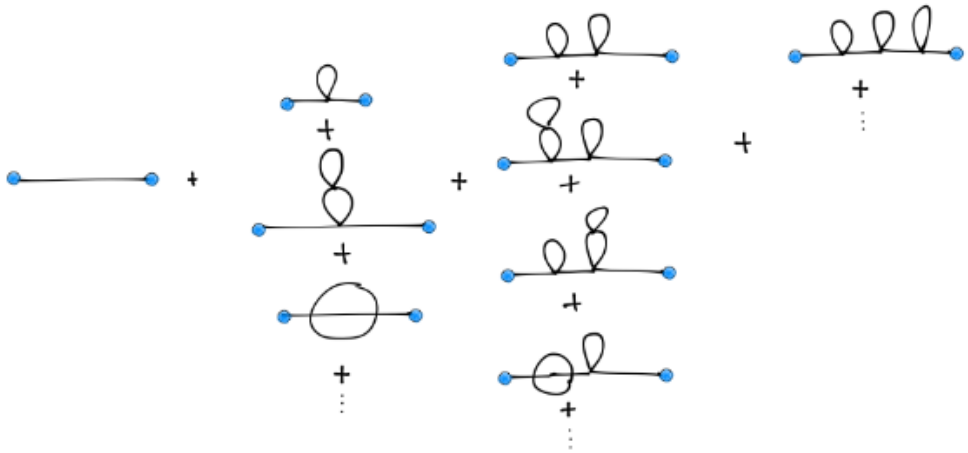
$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - \frac{1}{2}m_0^2\phi^2 - \frac{g}{4!}\phi^4. \tag{2.1}$$

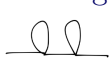
Demanding that the action is dimensionless means that $[\phi] = \frac{D-2}{2}$ so $[m_0] = 1$ and $[g] = \frac{4-D}{2}$, so g is dimensionless in $D = 4$. As above, this will mean *logarithms*!

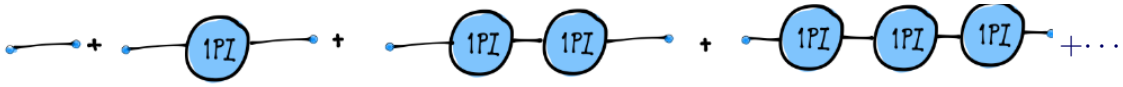
First let's think about the two-point function. Last quarter, you developed an expansion for the time-ordered two-point function of the field ϕ in terms of a sum of connected diagrams, ordered by the number of powers of g. Let's think about the momentum-space Green's function, and factor out the overall delta function that follows from translation symmetry by writing:

$$\tilde{G}^{(2)}(p_1, p_2) \equiv \not{\delta}^{d+1}(p_1 + p_2)\tilde{G}^{(2)}(p_1).$$

It will be useful to re-organize this sum, in the following way:

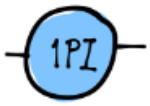


Here's the pattern: we define a diagram to be *one-particle irreducible* (1PI) if it cannot be disconnected by cutting through a single internal propagator. So for example,  is 1PI, but  is not; rather, the latter contributes to the bit with two 1PI insertions. Then

$\tilde{G}^{(2)}(p) =$ 

$+\cdots$

So that we may write equations without pictures, let

$$-\mathbf{i}\Sigma(p) \equiv$$ 

denote the 1PI two-point function. Σ being 1PI means that the external lines sticking out of it are 'nubbins,' placeholders where propagators may be attached. That's why there are no blue dots at the ends.

Now suppose we know Σ. It is known as the *self-energy*, for reasons we will see next. Then we can write

$$\begin{aligned}
\tilde{G}^{(2)}(p) &= \frac{\mathbf{i}}{p^2-m_0^2} + \frac{\mathbf{i}}{p^2-m_0^2}(-\mathbf{i}\Sigma(p))\frac{\mathbf{i}}{p^2-m_0^2} + \frac{\mathbf{i}}{p^2-m_0^2}(-\mathbf{i}\Sigma(p))\frac{\mathbf{i}}{p^2-m_0^2}(-\mathbf{i}\Sigma(p))\frac{\mathbf{i}}{p^2-m_0^2} + \cdots \\
&= \frac{\mathbf{i}}{p^2-m_0^2}\left(1 + \frac{\Sigma}{p^2-m_0^2} + \left(\frac{\Sigma}{p^2-m_0^2}\right)^2 + \cdots\right) \\
&= \frac{\mathbf{i}}{p^2-m_0^2}\frac{1}{1-\frac{\Sigma}{p^2-m_0^2}} = \frac{\mathbf{i}}{p^2-m_0^2-\Sigma(p)}.
\end{aligned} \tag{2.2}$$

A comment about summing this infinite series:

$$1 + x + x^2 + \cdots = \frac{1}{1-x}. \tag{2.3}$$

You probably know that a geometric series has a radius of convergence of 1. This is because the function to which it sums has a pole at $x = 1$, and the radius of convergence is at most the distance to the nearest singularity.

On the other hand, there is a theorem of complex analysis that if two functions analytic in an open set D agree in D then they are the same. This is the basis for analytic continuation.

The LHS of (2.2) is not a priori defined when $|x| > 1$. The relation (2.2) therefore gives a useful meaning to it in this regime.

We see that the self-energy shifts the m^2 of the particle – it moves the location of the pole in the propagator. In the interacting theory, $m^2 \equiv m_0^2 + \Sigma(p)|_{p^2=m^2}$ is the physical mass-squared (if this equation in fact has a solution!), while m_0 (what we've been calling m until just now) is deprecatingly called the 'bare mass'.

For $p^2 \sim m^2$, we will write

$$\tilde{G}^{(2)}(p) \equiv \left(\frac{\mathbf{i}Z}{p^2-m^2} + \text{regular bits}\right) \tag{2.4}$$

This equation defines the residue of the pole, Z, which is called the 'wavefunction renormalization factor'. It is 1 in the free theory, and represents the amplitude for the field to create a particle, and the other terms, which are not singular at $p^2 = m^2$, represent the amplitude for the field to do something else (such as create multiparticle states), and are absent in the free theory. Unitarity (positivity of norms of states) requires $Z < 1$; I will explain why in our section on consequences of unitarity. Notice that if we know Σ only to some order in perturbation theory, then (2.2) is still true, up to corrections at higher order.

Next let's do $2 \leftarrow 2$ scattering of ϕ particles.

$$\mathbf{i}\mathcal{M}_{2\leftarrow 2} = \quad + \quad + \quad + \quad + \mathcal{O}(g^3)$$

$$= -\mathbf{i}g \quad + \quad \mathbf{i}\mathcal{M}_s \quad + \quad \mathbf{i}\mathcal{M}_t \quad + \quad \mathbf{i}\mathcal{M}_u \quad + \quad \mathcal{O}(g^3)$$

where, in terms of $q_s \equiv k_1 + k_2$, the s-channel 1-loop amplitude is

$$\mathbf{i}\mathcal{M}_s = \frac{1}{2}(-\mathbf{i}g)^2 \int \mathrm{d}^4k \frac{\mathbf{i}}{k^2 - m^2 + \mathbf{i}\epsilon} \frac{\mathbf{i}}{(q_s - k)^2 - m^2 + \mathbf{i}\epsilon} \quad \sim \int^\Lambda \frac{d^4k}{k^4}.$$

Parametrizing ignorance. What is a scalar field? As we saw in the parable above, one way to discover scalar field theory is to start with a chain of masses connected by springs, like a mattress, and look at the long-wavelength (small-wavenumber) modes. So the coherent excitations of such a field are sound waves and the quanta of the field are called phonons. In the sum, $\int d^4k$, the region of integration that's causing the trouble is *not* the part where the system looks most like a field theory. That is: if we look closely enough (small enough $1/k$), we will see that the mattress is made of springs. In terms of the microscopic description with springs, there is a smallest wavelength, of order the inverse lattice spacing: the sum over k stops. [End of Lecture 3]

Field theories arise from many such models, which may differ dramatically in their short-distance physics. We'd like to not worry too much about which one, but rather say things which do not depend on this choice. Recall the discussion of the Casimir force: in that calculation, different choices of regulators for the mode sum corresponded (for example) to different material properties of the conducting plates. The Casimir force was independent of this choice; more generally, it is an important part of the physics problem to identify which quantities are UV sensitive and which are not.

If we had an actual lattice (like the chain of springs), we would replace the inverse propagator $p^2 - m^2 = \omega^2 - \vec{p}^2 - m^2$ with $\omega^2 - \omega_p^2 - m^2$, where ω_p is the dispersion relation (*e.g.* $\omega_p = 2t \sum_{i=1}^d (1 - \cos p_i a)$ for nearest-neighbor hopping on the cubic lattice), and

p is restricted to the Brillouin zone ($-\pi/a \le p_i < \pi/a$ for the cubic lattice). Instead, for simplicity, let's just impose a hard cutoff on the euclidean momentum $\sum_{i=0}^{d} p^2 \le \Lambda^2$.

Parametrizing ignorance is ubiquitous in science. In the context of field theory, at least in the high-energy community, it is called 'regularization'.

Now we need to talk about the integral a little more. The part that is causing the trouble is the bit with large k, which might as well be $|k| \sim \Lambda \gg m$, so let's set $m = 0$ for simplicity.

We'll spend lots of time learning to do integrals below. Here's the answer:

$$\mathbf{i}\mathcal{M} = -\mathbf{i}g + \mathbf{i}Cg^2\left(\log\frac{\Lambda^2}{s} + \log\frac{\Lambda^2}{t} + \log\frac{\Lambda^2}{u}\right) + \mathcal{O}(g^3)$$

If you must know, $C = \frac{1}{32\pi^2}$.

Observables can be predicted from other observables. Again, the boldface statement might sound like some content-free tweet from a boring philosophy-of-science twitter feed, but actually it's a very important thing to remember here.

What is g? As Zee's Smart Experimentalist says, it is just a letter in some theorist's lagrangian, and it doesn't help anyone to write physical quantities in terms of it. Much more useful would be to say what is the scattering amplitude in terms of things that can be measured. So, suppose someone scatters ϕ particles at some given $(s,t,u) = (s_0,t_0,u_0)$, and finds for the amplitude $\mathbf{i}\mathcal{M}(s_0,t_0,u_0) = -\mathbf{i}g_P$ where P is for 'physical'.[16] *This* we can relate to our theory letters:

$$-\mathbf{i}g_P = \mathbf{i}\mathcal{M}(s_0,t_0,u_0) = -\mathbf{i}g + \mathbf{i}Cg^2L_0 + \mathcal{O}(g^3) \tag{2.5}$$

where $L_0 \equiv \log\frac{\Lambda^2}{s_0} + \log\frac{\Lambda^2}{t_0} + \log\frac{\Lambda^2}{u_0}$. (Note that quantities like g_P are often called g_R where 'R' is for 'renormalized,' whatever that is.) I emphasize that this much we would have to do to make useful predictions, even if there were no specter of infinity or dependence on a fictitious cutoff.

Renormalization. Now here comes the big gestalt shift: Solve this equation (2.5) for the stupid letter g

$$-\mathbf{i}g = -\mathbf{i}g_P - \mathbf{i}Cg^2L_0 + \mathcal{O}(g^3) \tag{2.6}$$

$$= -\mathbf{i}g_P - \mathbf{i}Cg_P^2L_0 + \mathcal{O}(g_P^3). \tag{2.7}$$

[16] You might hesitate here about my referring to the amplitude $\mathcal{M}$ as an 'observable'. The difficult and interesting question of what can actually be measured in experiments can be decoupled a bit from this discussion. If you want to worry about this, see the beginning of Schwartz, chapter 18.

where to obtain (2.7) we used (2.6) in the g^2 term and in the remainder term: $\mathcal{O}(g^3) = \mathcal{O}(g_P^3)$. (We will use this trick all the time below.) Now we can eliminate g from the discussion:

$$\begin{aligned} \mathbf{i}\mathcal{M}(s,t,u) &= -\mathbf{i}g + \mathbf{i}Cg^2L + \mathcal{O}(g^3) \\ &\overset{(2.7)}{=} -\mathbf{i}g_P - \mathbf{i}Cg_P^2L_0 + \mathbf{i}Cg_P^2L + \mathcal{O}(g_P^3) \\ &= -\mathbf{i}g_P + \mathbf{i}Cg_P^2\left(\log\frac{s_0}{s} + \log\frac{t_0}{t} + \log\frac{u_0}{u}\right) + \mathcal{O}(g_P^3). \end{aligned} \tag{2.8}$$

This expresses the amplitude at any momenta (within the range of validity of the theory!) in terms of measured quantities, g_P, s_0, t_0, u_0. The cutoff Λ is gone! Just like in our parable in §1.1, it was eliminated by letting the coupling vary with it, $g = g(\Lambda)$, according to (2.7). We'll say a lot more about how to think about that dependence.

Renormalized perturbation theory. To slick up this machinery, consider the following Lagrangian density (in fact the same as (2.1), with $m = 0$ for simplicity):

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - \frac{g_P}{4!}\phi^4 - \frac{\delta_g}{4!}\phi^4 \tag{2.9}$$

but written in terms of the measured coupling g_P, and some as-yet-undetermined 'counterterm' δ_g. Then (with some foresight, we treat $\delta_g \sim g_P^2$, since its job is to cancel a term of this order)

$$\mathcal{M}(s,t,u) = -g_P - \delta_g - Cg_P^2\left(\log\frac{s}{\Lambda^2} + \log\frac{t}{\Lambda^2} + \log\frac{u}{\Lambda^2}\right) + \mathcal{O}(g_P^3).$$

If, in order to enforce the *renormalization condition* $\mathcal{M}(s_0,t_0,u_0) = -g_P$, we choose

$$\delta_g = -g_P^2C\left(\log\frac{s_0}{\Lambda^2} + \log\frac{t_0}{\Lambda^2} + \log\frac{u_0}{\Lambda^2}\right)$$

then we find

$$\mathcal{M}(s,t,u) = -g_P - Cg_P^2\left(\log\frac{s}{s_0} + \log\frac{t}{t_0} + \log\frac{u}{u_0}\right) + \mathcal{O}(g_P^3)$$

– all the dependence on the unknown cutoff is gone, we satisfy the observational demand $\mathcal{M}(s_0,t_0,u_0) = -g_P$, and we can predict the scattering amplitude (and others!) at any momenta.

The only price is that the 'bare coupling' g depends on the cutoff, and becomes infinite if we pretend that there is no cutoff. Happily, we didn't care about g anyway. We can just let it go.

The step whereby we were able to absorb all the dependence on the cutoff into the bare coupling constant involved some apparent magic. It is not so clear that the

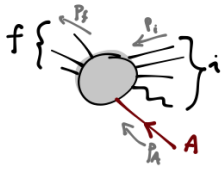
same magic will happen if we study the next order $\mathcal{O}(g_P^3)$ terms, or if we study other amplitudes. A QFT where all the cutoff dependence to all orders can be removed with a finite number of counterterms is called 'renormalizable'. As we will see, such a field theory is less useful because it allows us to pretend that it is valid up to arbitrarily high energies. The alternative, where we must add more counterterms (such as something like $\frac{\delta_6}{\Lambda^2}\phi^6$) at each order in perturbation theory, is called an *effective field theory*, which is a field theory that has the decency to predict its regime of validity.

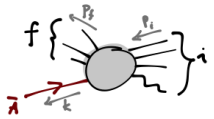
2.2 Towards quantum corrections to the Coulomb force law

As a prelude to studying loops in QED, and to make clear what is at stake, let me fill in some of the details of the leading-order calculation. By studying scattering of an electron from a heavy charged fermion (a muon is convenient) we will reconstruct the cross section for scattering off a Coulomb potential (named after Mott). This will emphasize the fact that the tree-level process is classical. Then we'll figure out how it is corrected by other QED processes.

Crossing symmetry. [Schwartz §13.3, 13.4, Peskin §5] If you look at a Feynman diagram on its side (for example because someone else fails to use the convention that time goes to the left) it is still a valid amplitude for some process. Similarly, dragging particles between the initial and final state also produces a valid amplitude. Making this relation precise can save us some work. The precise relation for dragging an incoming particle into the final state, so that it is an outgoing antiparticle, is:

$$\mathbf{i}\mathcal{M}_{f\leftarrow iA}(p_f;p_i,p_A) = \quad = \mathbf{i}\mathcal{M}_{\bar{A}f\leftarrow i}(p_f,k=-p_A;p_i) = \quad .$$





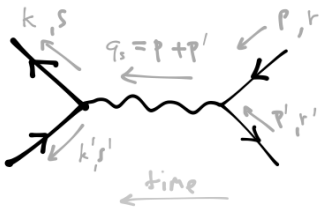
(If you must, note that this is another sense in which an antiparticle is a particle going backwards in time.) If A is a spinor particle, the sum relations for particles and antiparticles are different:

$$\sum_r u^r(p)\bar{u}^r(p) = \not{p} + m, \qquad \sum_r v^r(k)\bar{v}^r(k) = \not{k} - m = -(\not{p} + m)$$

– after accounting for $k = -p_A$, they differ by an overall sign. Hence we must also append a fermion sign factor $(-1)^{\text{number of fermions shuffled between in and out}}$ in the unpolarized scattering probability. Here is an example.

$\boxed{\mu^+\mu^- \leftarrow e^+e^-}$**.** For example, consider the process $\mu^+\mu^- \leftarrow e^+e^-$. To try to keep things straight, I'll call the electron momenta p, p' and the muon momenta k, k', since that won't change under crossing. The amplitude is

$$\mathbf{i}\mathcal{M}_{\mu^+\mu^-\leftarrow e^+e^-} =$$



$$= \left(-\mathbf{i}e\bar{u}^s(k)\gamma^\mu v^{s'}(k')\right)_{\text{muons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu q_\nu}{q^2}\right)}{q^2} \left(-\mathbf{i}e\bar{v}^{r'}(p')\gamma^\nu u^r(p)\right)_{\text{electrons}} \tag{2.10}$$

(with $q \equiv p + p' = k + k'$). If we don't keep track of the spins, then we must average over initial states and sum over final states, so the unpolarized scattering probability density is

$$\frac{1}{4}\sum_{\text{spins}} |\mathcal{M}|^2 \overset{\text{spinor traces}}{=} \frac{1}{4}\frac{e^4}{s^2} E^{\mu\nu} M_{\mu\nu} = \frac{2e^4}{s^2}\left(t^2 + u^2 + 4s(m_e^2 + m_\mu^2) - 2(m_e^2 + m_\mu^2)^2\right), \tag{2.11}$$

where the tensor objects $E^{\mu\nu}, M^{\mu\nu}$ come respectively from the electron and muon lines,

$$\begin{aligned}\frac{1}{4}E_{\mu\nu} &= p_\mu p'_\nu + p'_\mu p_\nu - \eta_{\mu\nu}(p \cdot p' + m_e^2)\\ \frac{1}{4}M_{\mu\nu} &= k_\mu k'_\nu + k'_\mu k_\nu - \eta_{\mu\nu}(k \cdot k' + m_\mu^2),\end{aligned}$$

and they are contracted by the photon line, with $s = q^2 = (p + p')^2$. The last step requires some algebra.

Spinor trace ninjutsu. In case you missed it, here is how to evaluate the spinor traces:

$$\text{The trace is cyclic:} \qquad \mathrm{tr}\,(AB\cdots C) = \mathrm{tr}\,(CAB\cdots). \tag{2.12}$$

$$\text{Our gamma matrices are } 4\times 4, \text{ so} \qquad \mathrm{tr}\mathbb{1} = 4.$$

$$\mathrm{tr}\gamma^\mu = \mathrm{tr}\left(\gamma^5\right)^2\gamma^\mu \overset{(2.12)}{=} \mathrm{tr}\gamma^5\gamma^\mu\gamma^5 \overset{\{\gamma^5,\gamma^\mu\}=0}{=} -\mathrm{tr}\gamma^\mu = 0. \tag{2.13}$$

The same trick shows that the trace of any odd number of gammas vanishes. The idea is that an odd number of gammas is a map between the L and R subspaces, so it has only off-diagonal terms in the Weyl basis.

$$\mathrm{tr}\gamma^\mu\gamma^\nu \overset{\text{clifford}}{=} -\mathrm{tr}\gamma^\nu\gamma^\mu + 2\eta^{\mu\nu}\mathrm{tr}\mathbb{1} \overset{(2.12)}{=} -\mathrm{tr}\gamma^\mu\gamma^\nu + 8\eta^{\mu\nu} \implies \mathrm{tr}\gamma^\mu\gamma^\nu = 4\eta^{\mu\nu}. \tag{2.14}$$

$$\mathrm{tr}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma = 4\left(\eta^{\mu\nu}\eta^{\rho\sigma} + \eta^{\sigma\mu}\eta^{\nu\rho} - \eta^{\mu\rho}\eta^{\nu\sigma}\right). \tag{2.15}$$

Why is this? The completely antisymmetric bit vanishes because it is proportional to γ^5 which is traceless (by the same argument as (2.13)). If any pair of indices is the same then the other two must be too by (2.14). If adjacent pairs are the same they can just square to one and we get $+1$; if alternating pairs are the same (and different from each other) then we must move them through each other with the anticommutator. If they are all the same we get 4 (in $D = 4$).

$$\mathrm{tr}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma^5 = -4\mathbf{i}\epsilon^{\mu\nu\rho\sigma}.$$

$\boxed{e^-\mu^- \leftarrow e^-\mu^-}$. To get from our previous work the amplitude (tree level, so far) for the process $e^-\mu^- \leftarrow e^-\mu^-$, we must move the incoming positron line to an outgoing

electron line, and move the outgoing antimuon line to an incoming muon line (hence the sign in σ will be $(-1)^{\text{number of fermions shuffled between in and out}} = (-1)^2 = 1$). Relative to the amplitude for $\mu^+\mu^- \leftarrow e^+e^-$ (2.10), we must replace the relevant vs with us for the initial/final antiparticles that were moved into final/initial particles, and we must replace $p' \to -p', k' \to -k'$:

$$\mathbf{i}\mathcal{M} = \quad = (-\mathbf{i}e\bar{u}(p')\gamma^\mu u(p)))_{\text{electrons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu^t q_\nu^t}{q_t^2}\right)}{q_t^2} (-\mathbf{i}e\bar{u}(k)\gamma^\nu u(k'))_{\text{muons}} \tag{2.16}$$

with $q_t \equiv p - p' = k' - k$, so that the Mandelstam variable is $t = q_t^2$. After the spin sum,

$$\begin{aligned}\frac{1}{4}\sum_{s,s',r,r'} |\mathcal{M}|^2 \equiv \frac{1}{4}\frac{e^4}{t^2} E^{\mu\nu} M_{\mu\nu} = \quad & 4\tfrac{e^4}{t^2}\left(-p^\mu p'^\nu - p'^\mu p^\nu - \eta^{\mu\nu}(-p\cdot p' + m_e^2)\right) \\ & \cdot\left(-k_\mu k'_\nu - k'_\mu k_\nu - \eta_{\mu\nu}(-k\cdot k' + m_\mu^2)\right) \\ = \tfrac{2e^4}{t^2}&\left(u^2 + s^2 + 4t(m_e^2 + m_\mu^2) - 2(m_e^2 + m_\mu^2)^2\right).\end{aligned} \tag{2.17}$$

Here $E_{\mu\nu}$ and $M_{\mu\nu}$ are the factors associated with the electron and muon lines respectively. The end result is obtained from the result for $e^+e^- \to \mu^+\mu^-$ by the permutation $(s,t,u) \to (t,u,s)$ of the Mandelstam variables, since

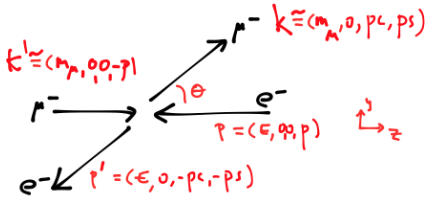
$$\begin{aligned} s &= (p+p')^2 \to (p-p')^2 = t \\ t &= (p-k)^2 \to (p-k)^2 = u \\ u &= (p-k')^2 \to (p+k')^2 = s\ . \end{aligned} \tag{2.18}$$

By this substitution we could get the last expression directly from (2.11) without the previous steps.

Payoff: the Mott formula. Recall other ways of figuring out the scattering cross section from a Coulomb potential from a point charge of charge ze.

We think about scattering from a fixed electrostatic potential $A_0 = \frac{ze}{r}$ and do classical mechanics. I can never remember how this goes. Instead, let's just scatter an electron off a heavy charge, such as a muon. If the charge of the heavy object were z times that of the electron, we would multiply the amplitude by z and the cross section by z^2.

Kinematics. 'Heavy' here means that we can approximate the CoM frame by its rest frame, and its initial and final energy as $k'_0 = m_\mu, k_0 = \sqrt{m_\mu^2 + \vec{k}^2} = m_\mu + \frac{1}{2}\vec{k}^2/m_\mu + \cdots \simeq m_\mu$. Also, this means the collision is approximately elastic. In the diagram of the kinematics at right, $c \equiv \cos\theta, s \equiv \sin\theta$.



This means that the muon-line tensor factor $M_{\mu\nu}$ in (2.18) simplifies dramatically:

$$-\frac{1}{4}M_{\mu\nu} \simeq k_\mu k'_\nu + k'_\mu k_\nu - \eta_{\mu\nu}\left(\underbrace{k\cdot k' - m_\mu^2}_{\simeq m_\mu^2 - m_\mu^2 = 0}\right) \simeq \delta_{\mu 0}\delta_{\nu 0} 2m_\mu^2.$$

In the electron line, we get

$$-p\cdot p' + m_e^2 = -E^2 + \vec{p}^2\cos\theta + m_e^2 = -\vec{p}^2(1-\cos\theta). \tag{2.19}$$

So

$$\begin{aligned}\frac{1}{4}E^{\mu\nu}M_{\mu\nu} = -2m_\mu^2 E^{00} &= 8m_\mu^2(2E^2 - \eta^{00}(p\cdot p' - m_e^2)) \\ &\overset{(2.19)}{=} 8m_\mu^2(2E^2 - \vec{p}^2(1-\cos\theta)) \\ &\overset{\text{trig}}{=} 8m_\mu^2 2(E^2 - \vec{p}^2\sin^2\frac{\theta}{2}) \overset{\beta^2\equiv\vec{p}^2/E^2}{=} 16m_\mu^2E^2(1-\beta^2\sin^2\frac{\theta}{2})\,.\end{aligned} \tag{2.20}$$

Note that $t = (p-p')^2 = -2\vec{p}^2(1-\cos\theta)$.

Now we can use the formula for the cross section with a two-body final state, in the CoM frame (for the derivation, see these notes, §4.8):

$$d\sigma = \frac{1}{v_{\text{rel}}}\frac{1}{2E_A 2E_B}\left(\frac{1}{4}\sum_{\text{spins}}|\mathcal{M}|^2\right)d\Pi_{LI}. \tag{2.21}$$

So the differential cross section is

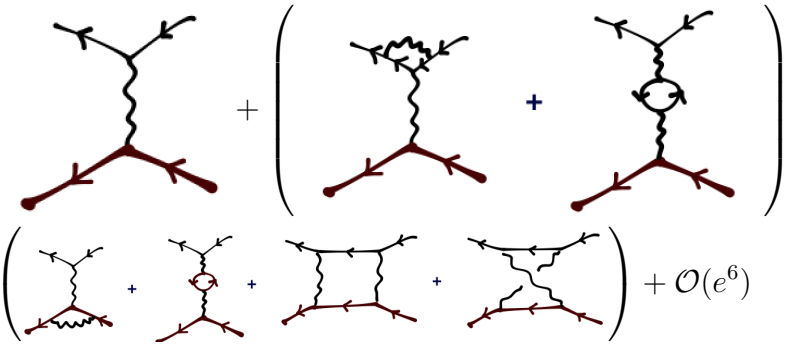
$$\begin{aligned} d\sigma &= \underbrace{\frac{1}{v_{\text{rel}}}}_{=\beta}\frac{1}{2E}\frac{1}{2m_\mu}\frac{z^2e^4}{t^2}16m_\mu^2E^2(1-\beta^2\sin^2\frac{\theta}{2})\frac{d\Omega}{16\pi^2}\frac{p}{E_{\text{total}}} \\ &\overset{E_{\text{total}}\sim m_\mu}{=} \frac{4Ep}{\beta}\frac{z^2\alpha^2(1-\beta^2\sin^2\frac{\theta}{2})}{t^2}d\Omega\end{aligned}$$

from which we get

$$\frac{d\sigma}{d\Omega}_{\text{Mott}} = z^2\frac{\alpha^2(1-\beta^2\sin^2\frac{\theta}{2})}{4\beta^2\vec{p}^2\sin^4\frac{\theta}{2}}. \tag{2.22}$$

- The Mott formula is independent of the muon mass, which may as well be infinite.
- If we take $\beta \ll 1$ in this formula we get the Rutherford formula.
- Notice that it blows up at $\theta \to 0$. This is a symptom of the long-range nature of the Coulomb potential, *i.e.* the masslessness of the photon.
- The calculation we've done characterizes the tree-level scattering of point particles, with no substructure. We can apply it to scattering of electrons off of protons, if E is small enough. For larger $E \gg m_e$, where we would have to go back to (2.22) and redo the kinematics take into account the recoil of the proton, but it would still describe scattering of point particles (for the result, see Eq. (13.103) of Schwartz's book). Actual experiments scattering electrons off protons reveal a high-energy regime that deviates from this behavior because the proton has substructure (it is made of quarks and gluons). It was a big surprise that at even higher energies the tree-level pointlike scattering formula applies again. This is a symptom of the asymptotic freedom of the quarks and gluons.

Radiative corrections. Now it's time to think about perturbative corrections to this cross section. Given that the leading-order calculation reproduced the classical physics of the Coulomb potential, you can think of what we are doing as effectively discovering (high-energy or short-distance) quantum corrections to the Coulomb law. The diagrams we must include are these (I made the muon lines thicker and also red):

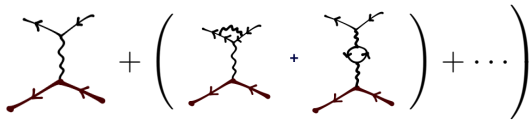
$\mathbf{i}\mathcal{M}_{e\mu\leftarrow e\mu} =$  $+ \mathcal{O}(e^6)$

• What do the one-loop diagrams in the second line have in common? They have an internal muon line. Why does this matter? When the energy going through the line is much smaller than the muon mass, then the propagator is $\frac{\mathbf{i}(\not{k}+m_\mu)}{k^2-m_\mu^2} \sim \frac{1}{m_\mu}$ and its relative contribution is down by $k/m_\mu \ll 1$. So let's neglect these for now.

• Why don't we include diagrams like  ? The LSZ formula tells us

that their effects on the S-matrix are accounted for by the wavefunction renormalization factors Z

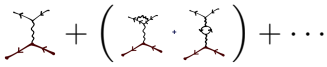
$$S_{e\mu\leftarrow e\mu} = \sqrt{Z_e}^2\sqrt{Z_\mu}^2 \Bigg(\quad + \Big(\quad \cdot \quad \Big) + \cdots \Bigg)_{\text{amputated, on-shell}}$$



and in determining the locations of the poles whose residues are the S-matrix elements. This is why $\mathcal{M}$ is a sum of *amputated* diagrams. So we'll take care of these when we talk about the electron self-energy. [End of Lecture 4]

• Notice that the one-loop *amplitudes* are suppressed relative to the tree level amplitude by two factors of e, hence one factor of the fine structure constant $\alpha = \frac{e^2}{4\pi}$. Their leading effects on the cross section come from

$$\sigma \sim Z_e Z_\mu \Big| \quad + \Big(\quad \cdot \quad \Big) + \cdots \Big|^2 \sim \sigma_{\text{tree}} + \mathcal{O}(\alpha^3)$$



from the cross term between the tree and one-loop amplitudes.

In the above discussion, we encounter all three 'primitive' one-loop divergent amplitudes of QED, which we'll study in turn:

- electron self-energy:  describing the difficulty for electron propagation posed by its emitting and reabsorbing photons
- vertex correction:  whose name is self-explanatory
- vacuum polarization (photon self-energy):  describing the difficulty for photon propagation posed by its turning into electron-positron pairs.

2.3 The LSZ reduction formula

Claim (the LSZ reduction formula):

$$S_{fi} \equiv \langle \vec{p}_1 \cdots \vec{p}_n | S | \vec{k}_1 \cdots \vec{k}_m \rangle = \prod_{a=1}^{n+m} \left(\lim_{P_a^0 \to E_{\vec{P}_a}} \frac{P_a^2 - m^2}{\mathbf{i}\sqrt{Z}} \right) \tilde{G}^{(n+m)} \left(k_1 \cdots k_m, -p_1 \cdots - p_n \right) \tag{2.23}$$

where $P_a \in \{p_i, k_i\}$. In words: the S-matrix elements are obtained from Green's functions by amputating the external legs, and putting the momenta on-shell. Notice

that choosing all the final momenta p_i different from all the initial momenta k_i goes a long way towards eliminating diagrams that are not fully connected.

This formula provides the bridge from time-ordered Green's functions (which we know how to compute in perturbation theory) and the S-matrix, which collects probability amplitudes for things to happen to particles, in terms of which we may compute cross sections and lifetimes. Let us spend just another moment inspecting the construction of this fine conveyance.

Why is LSZ true? Here's the argument I've found which best combines concision and truthiness. [It is mainly from the nice book by Maggiore §5.2; I also like Schwartz' chapter 6; Peskin's argument is in section 4.6.] The argument, which I write for a scalar field theory, has several steps. The field operators in this discussion are all in Heisenberg picture.

1. First, for a free field, the mode expansion (1.15) implies that we can extract the ladder operators by:

$$\begin{aligned}\sqrt{2\omega_k}a_k &= \mathbf{i}\int d^dx\; e^{\mathbf{i}kx}\left(-\mathbf{i}\omega_k + \partial_t\right)\phi_{\text{free}}(x)\\ \sqrt{2\omega_k}a_k^\dagger &= -\mathbf{i}\int d^dx\; e^{-\mathbf{i}kx}\left(+\mathbf{i}\omega_k + \partial_t\right)\phi_{\text{free}}(x)\;.\end{aligned} \tag{2.24}$$

   Notice that the LHS is independent of time, but the integrand of the RHS is not.

2. Here comes the tricky part. In a system with a well-defined S-matrix (the thing we're trying to compute here), the excitations are particles. Because, by assumption, the interactions are local, and the initial and final state involve well-separated particles, we can ignore the interactions in the initial and final states. This means we can label the initial and final states by the states of an auxiliary *free* theory[17]. This allows us to write the field in terms of some pretend free fields of mass m (not m_0!)

$$\phi(x)\begin{cases}\overset{t\to-\infty}{\rightsquigarrow} & Z^{\frac12}\phi_{\text{in}}(x)\\ \overset{t\to+\infty}{\rightsquigarrow} & Z^{\frac12}\phi_{\text{out}}(x)\end{cases}.$$

   The factors of $\sqrt{Z}$ are required to get the correct two point functions (2.4) near the mass shell. The mode operators for ϕ_{in} are called $a^{(\text{in})}$ etc. $\phi_{\text{in, out}}$ are free fields: their full hamiltonian is H_0. They are in Heisenberg picture, and the

[17] Here's why this is really bad: nearly everything we might scatter is a boundstate. For example: atoms, nuclei, nucleons etc... But if there are no interactions there are no boundstates. Please see the comments below this list for the explanation of why this works.

reference time for $\phi_{\text{in, out}}$ is $\pm\infty$ respectively. Since they are free fields, we can use (2.24) to write

$$\sqrt{2\omega_k}a^{(\text{in})\dagger} = -\mathbf{i}\int d^dx\, e^{-\mathbf{i}kx}\,(+\mathbf{i}\omega_k + \partial_t)\,\phi_{\text{in}}(x) = -\mathbf{i}Z^{-1/2}\int d^dx\, e^{-\mathbf{i}kx}\,(+\mathbf{i}\omega_k + \partial_t)\,\phi(x)|^{t\to-\infty}$$

where in the second step we used the independence on time in (2.24), even though $\phi(x)$ is not a free field. An expression for $a^{(\text{out})\dagger}$ obtains if we take $t \to +\infty$ instead.

3. Now make this expression covariant using the fundamental theorem of calculus:

$$\begin{aligned}\sqrt{2\omega_k}\left(a^{(\text{in})\dagger} - a^{(\text{out})\dagger}\right) &= \mathbf{i}Z^{-1/2}\int_{-\infty}^{\infty} dt\partial_t\left(\int d^dx\; e^{-\mathbf{i}kx}\,(\mathbf{i}\omega_k + \partial_t)\,\phi(x)\right)\\ &= \mathbf{i}Z^{-1/2}\int d^{d+1}x\left(e^{-\mathbf{i}kx}\partial_t^2\phi - \phi\cdot\underbrace{\partial_t^2\; e^{-\mathbf{i}k_\mu x^\mu}}_{\left(\vec{\nabla}^2 - m^2\right)e^{-\mathbf{i}kx}}\right)\\ &\overset{\text{IBP}}{=} \mathbf{i}Z^{-1/2}\int d^{d+1}x e^{-\mathbf{i}kx}\left(\Box + m^2\right)\phi(x) \qquad (2.25)\end{aligned}$$

In the last step we made a promise to only use wavepackets for external states, so that we can do IBP in space.

4. Now, here's where the S-matrix enters. Assume no subset of the incoming momenta $\sum_i k_i$ is the same as any subset of the outgoing momentum $\sum_j p_j$.

$$\begin{aligned}&\langle p_1\cdots p_n|\,S\,|k_1\cdots k_m\rangle\\ =&\prod_{p,k}\sqrt{2\omega}\,\langle\Omega|\prod a_p^{\text{out}}S\prod a_k^{\text{in}\dagger}\,|\Omega\rangle\\ =&\prod_{p,k}\sqrt{2\omega}\,\langle\Omega|\,\mathcal{T}\left(\prod a_p^{\text{out}}S\prod a_k^{\text{in}\dagger}\right)|\Omega\rangle \qquad a^{\text{out}}\text{ lives at } t=+\infty\\ =&\prod_{p,k}\sqrt{2\omega}\,\langle\Omega|\,\mathcal{T}\left(\prod a_p^{\text{out}}S\left(a_{k_1}^{\text{in}\dagger} - a_{k_1}^{\text{out}\dagger}\right)\prod_2^m a_k^{\text{in}\dagger}\right)|\Omega\rangle \qquad \text{since } p_i\neq k_j,\text{ use } \langle 0|\,a^{\text{out}\dagger} = 0\\ \overset{(2.25)}{=}&\prod_{p,k\backslash k_1}\sqrt{2\omega}\mathbf{i}Z^{-1/2}\int d^{d+1}x_1 e^{-\mathbf{i}k_1x_1}\,\langle\Omega|\,\mathcal{T}\left(\prod a_p^{\text{out}}S\left(\Box + m^2\right)\phi(x_1)\prod_2^m a_k^{\text{in}\dagger}\right)|\Omega\rangle\\ =&\prod_{p,k\backslash k_1}\sqrt{2\omega}\mathbf{i}Z^{-1/2}\int d^{d+1}x_1 e^{-\mathbf{i}k_1x_1}\left(\Box + m^2\right)\langle\Omega|\,\mathcal{T}\left(\prod a_p^{\text{out}}S\phi(x_1)\prod_2^m a_k^{\text{in}\dagger}\right)|\Omega\rangle + \text{X}\end{aligned}$$

In the last step, X comes from where the $\Box_{x_1}$ hits the time ordering symbol. This gives terms that will not matter when we take $k^2 \to m^2$, I promise.

5. Now do this for every particle to get

$$\begin{aligned}\langle p_1 \cdots p_n| \, S \, |k_1 \cdots k_m\rangle = \prod_{j=1}^{m} \int d^{d+1}y_j \; e^{+\mathbf{i}p_j y_j} \mathbf{i} Z^{-1/2} \left(\Box_j + m^2\right) \\ \prod_{i=1}^{n} \int d^{d+1}x_i \; e^{-\mathbf{i}k_i x_i} \mathbf{i} Z^{-1/2} \left(\Box_i + m^2\right) \; \langle \Omega| \, \mathcal{T}\phi(x_i) \cdots \phi(y_j) S \, |\Omega\rangle + \mathrm{X}\end{aligned}$$

The x and y integrals are just Fourier transforms, and this says that near the mass shell,

$$\tilde{G}^{(n+m)}(k_1 \cdots k_m, -p_1 \cdots - p_n) = \prod_{a}^{n+m} \frac{\mathbf{i}\sqrt{Z}}{P_a^2 - m^2} \, \langle p_1 \cdots p_n| \, S \, |k_1 \cdots k_m\rangle + \text{regular}$$

(where $P_a \in \{p_j, k_i\}$) which is the same as (2.23).

Comments about LSZ.

- An apparently important role is played in the derivation of LSZ by this idea of turning off the interactions in the initial and final states. The reason this crazy-seeming trick works is the locality of interactions. In a scattering experiment, we scatter well-localized wavepackets that start and end far away from each other, and therefore do not interact at the beginning and the end. Although we calculate S-matrix elements of plane waves (not localized at all), this is also just an expedient – if you forced me we could figure out how to write the formula for wavepackets instead. No physics conclusion would be different.

- The real content of the $\phi_{\text{in}}, \phi_{\text{out}}$ business is *adiabatic continuation* – we are labelling the (low-lying) eigenstates of the full Hamiltonian by those of the free theory. This does not always work! There are at least two failure modes.

  The most dramatic is when the eigenstates are not particle states at all, but rather some kind of soup of excitations. This is what happens in a conformal field theory (CFT), and is sometimes called *unparticles*. In this case, not only does our formula for the S-matrix not work, but the S-matrix itself is not so well-defined.

  Less dramatic is when there are asymptotic particle states, but they are not simply related to the quanta of the fields in the Lagrangian. This happens any time there are boundstates, like positronium in QED. But it happens even more dramatically in QCD, where our microscopic Lagrangian description is in terms of quarks and gluons, but the quarks and gluons are *confined*, and the asymptotic particle states are instead all colorless composites called hadrons (protons, neutrons, pions ...). The next item shows that LSZ also teaches us how to deal with this.

- In our discussion of QFT, a special role has been played by certain fields called ϕ. Suppose we have some other (say hermitian) local operator $\mathcal{O}$ such that

$$\langle p| \mathcal{O}(x) |\Omega\rangle = Z_{\mathcal{O}} e^{\mathbf{i}px}$$

where $\langle p|$ is a one-particle state made by our friend ϕ (we could put some labels, *e.g.* for spin or polarization or flavor, on both the operator and the state, but let's not). Such an $\mathcal{O}$ is called an 'interpolating field' or 'interpolating operator'. And suppose we have information about the correlation functions of $\mathcal{O}$:

$$G^{(n)}_{\mathcal{O}}(1 \cdots n) \equiv \langle \Omega| \mathcal{T} \left(\mathcal{O}_1(x_1) \cdots \mathcal{O}_n(x_n)\right) |\Omega\rangle .$$

In this case, there is a more general statement of LSZ:

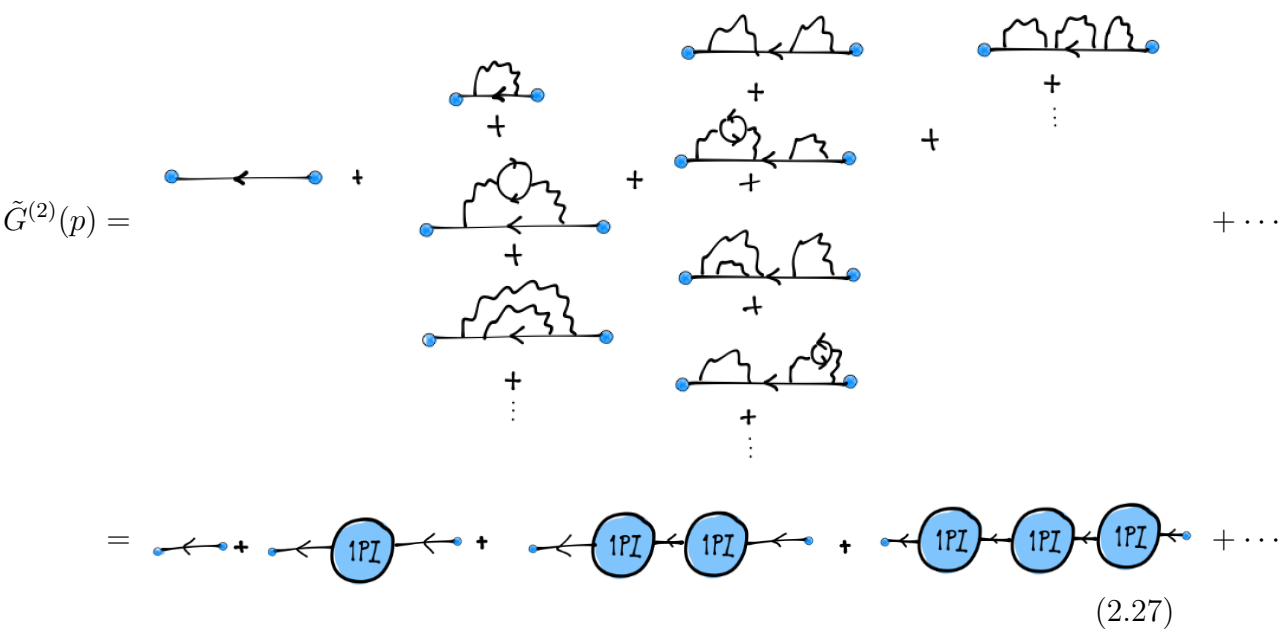
$$\begin{aligned} &\textstyle\prod_{a\in i} \left(Z_a^{-1/2} \mathbf{i} \int d^{d+1}x_a e^{-\mathbf{i}p_a x_a} \left(\Box_a + m_a^2\right)\right) \\ &\textstyle\prod_{b\in f} \left(Z_b^{-1/2} \mathbf{i} \int d^{d+1}x_b e^{+\mathbf{i}p_b x_b} \left(\Box_b + m_b^2\right)\right) \; G^{(n)}_{\mathcal{O}}(1 \cdots n) \\ &\qquad\qquad = \langle \{p_f\}| S |\{p_i\}\rangle \end{aligned} \tag{2.26}$$

This more general statement follows as above if we can write $\mathcal{O}_a \overset{t\to-\infty}{\rightsquigarrow} \sqrt{Z_a}\phi_{\text{in}}$. This more general formula allows us to scatter particles that are not 'elementary' in the sense that they are just the quanta of the fields in terms of which we write our Lagrangian.

(So, for the example of positronium, which is a boundstate of an electron and a positron, we can take $\mathcal{O}(x) = \bar{\psi}_e(x)\psi_e(x)$ where ψ_e is the field that annihilates an electron. This operator has some nonzero amplitude to create a state with a positronium atom from the vacuum: $\langle \text{positronium}, q| \mathcal{O}(x) |\Omega\rangle = Ze^{\mathbf{i}qx}$.)

2.4 Electron self-energy in QED

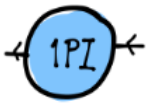
Let's think about the electron two-point function in momentum space:



(2.27)

We've grouped the diagrams according to their behavior when we divide input and output by cutting a single line. As in our discussion for scalars above, a diagram that cannot be divided by cutting a single line is called *one-particle irreducible* (1PI). The blue blob is defined to be the sum of all 1PI diagrams. We will denote the 1PI two-point function by

$$-\mathbf{i}\Sigma(p) \equiv$$



a blob with nubbins; for fermions with conserved particle number, the nubbins carry arrows indicating the particle number flow. Let me call the tree level propagator

$$\longleftarrow = \mathbf{i}S(p) \equiv \frac{\mathbf{i}(\not{p} + m_0)}{p^2 - m_0^2 + \mathbf{i}\epsilon} = \frac{\mathbf{i}}{\not{p} - m_0}$$

– notice that I added a demeaning subscript to the notation for the mass appearing in the Lagrangian. Foreshadowing.

Note that S, G, Σ are all matrices in the spinor space. The full momentum-space two point function is then:

$$\tilde{G}^{(2)}(p) = \mathbf{i}S + \mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S + \mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S\left(-\mathbf{i}\Sigma(p)\right)\mathbf{i}S + \cdots$$

$$= \mathbf{i}S\left(1 + \Sigma S + \Sigma S \Sigma S + \cdots\right) = \mathbf{i}S\frac{1}{1-\Sigma S}$$
$$= \frac{\mathbf{i}}{\not{p} - m_0}\frac{1}{1-\Sigma\frac{1}{\not{p}-m_0}} = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(\not{p})}. \tag{2.28}$$

Are you worried about these manipulations because Σ and S are matrices in the spinor indices? Don't be: they are both made entirely from the matrix $\not{p}$, and therefore they commute; we could do these manipulations in the eigenbasis of $\not{p}$.

So let's write $\Sigma = \Sigma(\not{p})$. This fully corrected propagator has a pole at

$$\not{p} = m\mathbb{1} \equiv m_0\mathbb{1} + \Sigma(m\mathbb{1}) \ . \tag{2.29}$$

This means that the actual mass of the particle is this new quantity m. But what is m (it is called the 'renormalized mass')? To figure it out, we need to know about Σ.

In QED we must study Σ in perturbation theory. As you can see from (2.27), the leading (one-loop) contribution is

$$-\mathbf{i}\Sigma_2(p) = \quad = (-\mathbf{i}e)^2\int \mathrm{d}^4k\ \gamma^\mu \frac{\mathbf{i}(\not{k}+m_0)}{k^2-m_0^2+\mathbf{i}\epsilon}\gamma^\nu \frac{-\mathbf{i}\eta_{\mu\nu}}{(p-k)^2-\mu^2+\mathbf{i}\epsilon} \ . \tag{2.30}$$

Notice that I am relying on the Ward identity to enforce the fact that only the transverse bit of the photon propagator matters. Also, I added a mass μ for the photon as an IR regulator. We must keep the external momentum p arbitrary, since we don't even know where the mass-shell is! Note that it is OK to use the bare propagator in the integrand of (2.30) since we are neglecting other corrections at the next order.

Finally, I can't put it off any longer: how are we going to do this loop-momentum integral?

Step 1: Feynman parameter trick. It is a good idea to consider the integral

$$\int_0^1 dx \frac{1}{(xA+(1-x)B)^2} = \int_0^1 dx \frac{1}{(x(A-B)+B)^2} = \frac{1}{A-B}\frac{-1}{x(A-B)+B}\Big|_{x=0}^{x=1}$$
$$= \frac{1}{A-B}\left(-\frac{1}{A}+\frac{1}{B}\right) = \frac{1}{AB}.$$

This allows us to combine the denominators into one:

$$\mathcal{I} = \underbrace{\frac{1}{k^2-m_0^2+\mathbf{i}\epsilon}}_{B}\underbrace{\frac{1}{(p-k)^2-\mu^2+\mathbf{i}\epsilon}}_{A} = \int_0^1 dx \frac{1}{\left(x\left((p^2-2pk+k^2)-\mu^2+\mathbf{i}\epsilon\right)+(1-x)(k^2-m_0^2+\mathbf{i}\epsilon)\right)^2}$$

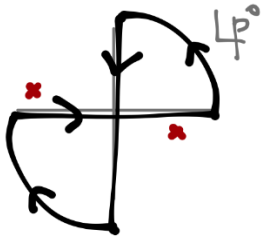
Step 2: Now we can complete the square

$$\mathcal{I} = \int_0^1 dx \frac{1}{\left(\underbrace{(k-px)}_{\equiv \ell}{}^2 - \Delta + \mathbf{i}\epsilon \right)^2}$$

with

$$\ell^\mu \equiv k^\mu - p^\mu x, \quad \Delta \equiv +p^2x^2 + x\mu^2 - xp^2 + (1-x)m_0^2 = x\mu^2 + (1-x)m_0^2 - x(1-x)p^2.$$

Step 3: Wick rotate. Because of the $\mathbf{i}\epsilon$ we've been dutifully carrying around, the poles of the ℓ^0 integral don't occur in the first and third quadrants of the complex ℓ^0 plane. (And the integrand decays at large $|\ell^0|$.) This means that we can rotate the contour to euclidean time for free: $\ell^0 \equiv \mathbf{i}\ell^4$. Equivalently: the integral over the contour at right vanishes, so the real time contour gives the same answer as the (upward-directed) Euclidean contour.



Notice that $\ell^2 = -\ell_E^2$. Altogether

$$-\mathbf{i}\Sigma_2(p) = -e^2 \int \mathrm{đ}^4\ell \int_0^1 dx \frac{N}{(\ell^2 - \Delta + \mathbf{i}\epsilon)^2} = -e^2 \int_0^1 dx \mathbf{i} \int \mathrm{đ}^4\ell_E \frac{N}{(\ell_E^2 + \Delta)^2}$$

where the numerator is

$$N = \gamma^\mu \left(\not{\ell} + x\not{p} + m_0 \right) \gamma_\mu = -2\left(\not{l} + x\not{p} \right) + 4m_0.$$

Here I used two Clifford algebra facts: $\gamma^\mu\gamma_\mu = 4$ and $\gamma^\mu \not{p} \gamma_\mu = -2\not{p}$. Think about the contribution from the term with $\not{\ell}$ in the numerator: everything else is invariant under rotations of ℓ [End of Lecture 5]

$$\mathrm{đ}^4\ell_E = \frac{1}{(2\pi)^4} d\Omega_3 \ell^3 d\ell = \frac{d\Omega_3}{(2\pi)^4} \ell^2 \frac{d\ell^2}{2},$$

so this averages to zero. The rest is of the form (using $\int_{S^3} d\Omega_3 = 2\pi^2$)

$$\begin{aligned} \Sigma_2(p) &= e^2 \int_0^1 dx \int \frac{\ell^2 d\ell^2}{2} \frac{(2\pi^2)}{(2\pi)^4} \frac{2(2m_0 - x\not{p})}{(\ell^2 + \Delta)^2} \\ &= \frac{e^2}{8\pi^2} \int_0^1 dx (2m_0 - x\not{p}) \mathcal{J} \end{aligned} \tag{2.31}$$

with

$$\mathcal{J} = \int_0^\infty d\ell^2 \frac{\ell^2}{(\ell^2 + \Delta)^2}.$$

In the large ℓ part of the integrand this is

$$\int^{\Lambda} \frac{d\ell^2}{\ell^2} \sim \log \Lambda. \tag{2.32}$$

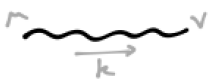
You knew this UV divergence was coming. To be more precise, let's add zero:

$$\begin{aligned}\mathcal{J} &= \int d\ell^2 \left(\frac{\ell^2+\Delta}{(\ell^2+\Delta)^2} - \frac{\Delta}{(\ell^2+\Delta)^2} \right) \\ &= \int_0^{\infty} d\ell^2 \left(\frac{1}{\ell^2+\Delta} - \frac{\Delta}{(\ell^2+\Delta)^2} \right) = \ln(\ell^2+\Delta)\big|_{\ell^2=0}^{\infty} + \frac{\Delta}{\ell^2+\Delta}\bigg|_{\ell^2=0}^{\infty} = \ln(\ell^2+\Delta)\big|_{\ell^2=0}^{\infty} - 1.\end{aligned}$$

Recall that

$$\Delta = x\mu^2 + (1-x)m_0^2 - x(1-x)p^2 \equiv \Delta(\mu^2).$$

Pauli-Villars regularization. Here is a convenient fiction about the UV: when you exchange a photon, you also exchange a very heavy particle, with mass $m^2 = \Lambda^2$, with an extra (-1) in front of its propagator. This means that (in this Pauli-Villars (PV) regulation scheme) the Feynman rule for the wiggly line is instead



$$\begin{aligned}&\rightsquigarrow -\mathbf{i}\eta_{\mu\nu}\left(\frac{1}{k^2-\mu^2+\mathbf{i}\epsilon} - \frac{1}{k^2-\Lambda^2+\mathbf{i}\epsilon} \right) \\ &= -\mathbf{i}\eta_{\mu\nu}\left(\frac{\mu^2-\Lambda^2}{(k^2-\mu^2+\mathbf{i}\epsilon)\,(k^2-\Lambda^2+\mathbf{i}\epsilon)} \right)\end{aligned}$$

This goes like $\frac{1}{k^4}$ at large k, so the integrals are more convergent. Yay.

Notice that the contribution from the Pauli-Villars photon to tree-level amplitudes goes like $|\frac{1}{k^2-\Lambda^2}| \overset{\Lambda \gg k}{\sim} \frac{1}{\Lambda^2}$ (where k is the momentum going through the photon line, determined by the external momenta), which innocuously vanishes as $\Lambda \to \infty$.

Remembering that the residue of the pole in the propagator is the probability for the field operator to create a particle from the vacuum, you might worry that this is a negative probability, and unitarity is called into question. This PV particle is a *ghost*. However, we will choose Λ so large that the pole in the propagator at $k^2 = \Lambda^2$ will never be accessed and we'll never have external Pauli-Villars particles. We are using this as a device to define the theory in a regime of energies much less than Λ.

To accomplish the change we made requires adding a new vector field with mass Λ and a wrong-sign kinetic term (to give the wrong-sign propagator), with the same coupling to charged particles as the photon field. The quanta of this field are the heavy weird-sign photon particles. You shouldn't take the regulated theory too seriously. For example, the wrong-sign kinetic terms for the PV fields means that very wiggly

configurations will be energetically *favored* rather than suppressed by the Hamiltonian. It will not make much sense non-perturbatively.

I emphasize that this regulator is one possibility of many. They each have their drawbacks. They all break scale invariance. Nice things about PV are that it is Lorentz invariant and gauge invariant; the bad thing is it's not unitary. What's wrong with the hard cutoff on Euclidean momenta that we wrote in (2.32)? As we'll discuss further below, it violates gauge invariance. This is a disaster because a gauge transformation is supposed to relate two descriptions of the same state; violating this invariance means adding degrees of freedom. A class of regulators that make perfect sense non-perturbatively is the lattice (as in the model with masses on springs). The price is that it really messes up the spacetime symmetries.

Regulator	Lorentz invariance	Gauge invariance	Unitarity	Scale invariance
Pauli-Villars	yes	yes	no	no
Euclidean cutoff	yes	no	yes	no
Lattice	very much no	yes	yes	no
Dim reg	yes	yes	yes?	no

Applying PV to the self-energy integral amounts to the replacement

$$\begin{aligned}\mathcal{J} &\rightsquigarrow \mathcal{J}_{\Delta(\mu^2)} - \mathcal{J}_{\Delta(\Lambda^2)} \\ &= \left[\left(\ln\left(\ell^2+\Delta(\mu^2)\right)-1\right) - \left(\ln\left(\ell^2+\Delta(\Lambda^2)\right)-1\right)\right]\Big|_{\ell=0}^{\infty} \\ &= \ln\frac{\ell^2+\Delta(\mu^2)}{\ell^2+\Delta(\Lambda^2)}\Big|_{\ell=0}^{\infty} \\ &= \ln 1/1 - \ln\frac{\Delta(\mu^2)}{\Delta(\Lambda^2)} = \ln\frac{\Delta(\Lambda^2)}{\Delta(\mu^2)}.\end{aligned}$$

Notice that we can take advantage of our ignorance of the microphysics to make the cutoff (the PV scale Λ) as big as we like and thereby simplify our lives:

$$\Delta(\Lambda^2) = x\Lambda^2 + (1-x)m_0^2 - x(1-x)p^2 \overset{\Lambda\gg\text{everyone}}{\approx} x\Lambda^2.$$

Finally then

$$\Sigma_2(\not{p})_{PV} = \frac{\alpha}{2\pi}\int_0^1 dx(2m_0\mathbb{1} - x\not{p})\ln\frac{x\Lambda^2}{x\mu^2+(1-x)m_0^2-x(1-x)p^2}. \tag{2.33}$$

(Note that $p^2 = \not{p}^2$, so it is appropriate to regard this as just a function of $\not{p}$.)

Mass renormalization. Having arrived at this regulated expression for the self-energy we need to "impose a renormalization condition," *i.e.* introduce some observable

physics in terms of which to parametrize our answers. As a first step, we return to (2.29): the shift in the mass as a result of this one-loop self-energy is

$$\begin{aligned}
\delta m \equiv m - m_0 &= \Sigma_2(\not{p} = m) + \mathcal{O}(e^4) = \Sigma_2(\not{p} = m_0) + \mathcal{O}(e^4) \\
&= \frac{\alpha}{2\pi}\int_0^1 dx \; \mathbb{1}(2-x)m_0 \ln \frac{x\Lambda^2}{\underbrace{x\mu^2 + (1-x)m_0^2 - x(1-x)m_0^2}_{=x\mu^2+(1-x)^2m_0^2 \equiv f(x,m_0,\mu)}} \\
&= \frac{\alpha}{2\pi}\int_0^1 dx \; (2-x)m_0 \left(\underbrace{\ln \frac{\Lambda^2}{m_0^2}}_{\text{divergent}} + \underbrace{\ln \frac{xm_0^2}{f(x,m_0,\mu)}}_{\text{relatively small}} \right) \\
&\approx \frac{\alpha}{2\pi}\left(2 - \frac{1}{2}\right) m_0 \ln \frac{\Lambda^2}{m_0^2} = \frac{3\alpha}{4\pi} m_0 \ln \frac{\Lambda^2}{m_0^2}.
\end{aligned} \tag{2.34}$$

In the penultimate step (with the $\approx$), we've neglected the finite bit (labelled 'relatively small') compared to the logarithmically divergent bit: we've already assumed $\Lambda \gg$ all other scales in the problem.

Now the physics input: The mass of the electron is 511 keV (you can ask how we measure it and whether the answer we get depends on the resolution of the measurement, and indeed there is more to this story; this is a low-energy answer, for example we could make the electron go in a known magnetic field and measure the radius of curvature r of its orbit and its velocity v and set $m_e v^2/r = evB/c$), so

$$511 \text{ keV} \approx m_e = m_0 \left(1 + \frac{3\alpha}{4\pi} \ln \frac{\Lambda^2}{m_0^2}\right) + \mathcal{O}(\alpha^2). \tag{2.35}$$

In this equation, the LHS is a measured quantity. In the correction on the RHS $\alpha \approx \frac{1}{137}$ is small, but it is multiplied by $\ln \frac{\Lambda^2}{m_0}$ which is arbitrarily large. This means that the bare mass m_0, which is going to absorb the cutoff dependence here, must actually be really small. (Notice that actually I've lied a little here: the α we've been using is still the bare charge; we will need to renormalize that one, too, before we are done.) I emphasize: m_0 and the other fake, bare parameters in $\mathcal{L}$ depend on Λ and the order of perturbation theory to which we are working and other theorist bookkeeping garbage; m_e does not. At each order in perturbation theory, we eliminate m_0 using (2.35) (with the RHS computed to the appropriate order), and write our predictions in terms of m_e. It is not too surprising that the mass of the electron includes such contributions: it must be difficult to travel through space if you are constantly emitting and re-absorbing photons.

Wavefunction renormalization. The actual propagator for the electron, near

the electron pole is

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(\not{p})} \overset{p\sim m}{\simeq} \frac{\mathbf{i}Z}{\not{p} - m} + \text{regular terms.} \tag{2.36}$$

The residue of the pole at the electron mass is no longer equal to one, but rather Z. To see what Z actually is at this order in e^2, Taylor expand near the pole

$$\begin{aligned}\Sigma(\not{p}) &\overset{\text{Taylor}}{=} \Sigma(\not{p} = m) + \frac{\partial\Sigma}{\partial\not{p}}|_{\not{p}=m}(\not{p} - m) + \mathcal{O}(\not{p} - m)^2 \\ &\overset{\text{pert th}}{=} \Sigma_2(\not{p} = m) + \frac{\partial\Sigma_2}{\partial\not{p}}|_{\not{p}=m_0}(\not{p} - m) + \mathcal{O}(e^4, \not{p} - m)^2)\end{aligned}$$

So then (2.36) becomes

$$\tilde{G}^{(2)}(p) \overset{p\sim m}{\sim} \frac{\mathbf{i}}{\not{p} - m - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}(\not{p} - m)} = \frac{\mathbf{i}}{(\not{p} - m)\left(1 - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}\right)} \tag{2.37}$$

So that

$$Z = \frac{1}{1 - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}} \simeq 1 + \frac{\partial\Sigma}{\partial\not{p}}|_{m_0} \equiv 1 + \delta Z$$

and at leading nontrivial order

$$\begin{aligned}\delta Z = \left.\frac{\partial\Sigma_2}{\partial\not{p}}\right|_{\not{p}=m_0} &\overset{(2.33)}{=} \frac{\alpha}{2\pi}\mathbb{1}\int_0^1 dx \left(-x\ln\frac{x\Lambda^2}{f(x, m_0, \mu)} + (2-x)m_0^2\frac{-2x(1-x)}{f(x, m_0, \mu)}\right) \\ &= -\frac{\alpha}{4\pi}\left(\ln\frac{\Lambda^2}{m_0^2} + \text{finite}\right).\end{aligned} \tag{2.38}$$

Here $f = f(x, m_0, \mu)$ is the same quantity defined in the second line of (2.34). Notice that $\delta Z < 0$, so $Z < 1$ as promised.

We'll see below (in §2.6.3) that the cutoff-dependence in δZ plays a crucial role in making the S matrix (for example for the $e\mu \to e\mu$ process we've been discussing) cutoff-independent and finite, when written in terms of physical variables.

2.5 Big picture interlude

OK, I am having a hard time just pounding away at one-loop QED. Let's take a break and think about the self-energy corrections in scalar field theory. Then we will step back and think about the general structure of short-distance senstivity in (relativistic) QFT, before returning to the QED vertex correction and vacuum polarization.

2.5.1 Self-energy in ϕ^4 theory

[Zee §III.3] Let's return to the ϕ^4 theory in $D = 3+1$ for a moment. The $\mathcal{M}_{\phi\phi\leftarrow\phi\phi}$ amplitude is not the only place where the cutoff appears.

Above we added a counterterm of the same form as the ϕ^4 term in the Lagrangian. Now we will see that we need counterterms for everybody:

$$\mathcal{L} = -\frac{1}{2}\left(\phi\Box\phi + m_P^2\phi^2\right) - \frac{g_P}{4!}\phi^4 - \frac{\delta_g}{4!}\phi^4 - \frac{1}{2}\delta Z\phi\Box\phi - \frac{1}{2}\delta_{m^2}\phi^2.$$

(I may sometimes forget to put the subscripts $_P$ on the physical quantities below.)

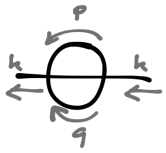
Here is a way in which ϕ^4 theory is weird: At one loop there is no wavefunction renormalization. That is,

$$\delta\Sigma_1(k) = \quad = -\mathbf{i}g\int^\Lambda \mathrm{đ}^4q \frac{\mathbf{i}}{q^2 - m^2 + \mathbf{i}\epsilon} = \delta\Sigma_1(k=0) \sim g\Lambda^2$$

which is certainly quadratically divergent, but totally independent of the external momentum. This means that when we Taylor expand in k (as we just did in (2.37)), this diagram only contributes to the mass renormalization. Demanding that the pole in the propagator occurs at $p^2 = m^2$, we must set $\delta_{m^2} = -\delta\Sigma_1$.

So let's see what happens if we keep going:

$$\delta\Sigma_2(k) = \quad = \frac{(-\mathbf{i}g)^2}{3!}\int \mathrm{đ}^4p \int \mathrm{đ}^4q \mathbf{i}D_0(p)\mathbf{i}D_0(q)\mathbf{i}D_0(k-p-q) \equiv I(k^2, m, \Lambda).$$



Here $\mathbf{i}D_0(p) \equiv \frac{\mathbf{i}}{p^2 - m^2 + \mathbf{i}\epsilon}$ is the free propagator (the factor of $\mathbf{i}$ is for later convenience), and we've defined I by this expression. The fact that I depends only on k^2 is a consequence of Lorentz invariance. Counting powers of the loop momenta, the short-distance bit of this integral is of the schematic form $\int^\Lambda \frac{d^8P}{P^6} \sim \Lambda^2$, also quadratically divergent, but this time k^2-dependent, so there will be a nonzero $\delta Z \propto g^2$. As we just

did for the electron self-energy, we should Taylor expand in k. (We'll learn more about why and when the answer is analytic in k^2 at $k=0$ later.) The series expansion in k^2 (let's do it about $k^2=0\sim m^2$ to look at the UV behavior) is

$$\delta\Sigma_2(k^2) = A_0 + k^2 A_1 + k^4 A_2 + \cdots$$

where $A_0 = I(k^2=0) \sim \Lambda^2$. In contrast, dimensional analysis says $A_1 = \frac{\partial}{\partial k^2} I|_{k^2=0} \sim \int \frac{d^8P}{P^8} \sim \Lambda^{0^+} \sim \ln\Lambda$ has two fewer powers of the cutoff. After that it's clear sailing: $A_2 = \left(\frac{\partial}{\partial k^2}\right)^2 I|_{k^2=0} \sim \int^\Lambda \frac{d^8P}{P^{10}} \sim \Lambda^{-2}$ is finite as we remove the cutoff, and so are all the later coefficients.

If instead the physical pole were at a nonzero value of the mass, we should Taylor expand about $k^2 = m_P^2$ instead[18]:

$$D^{-1}(k) = D_0^{-1}(k) - \Sigma(k) = k^2 - m_0^2 - \underbrace{\left(\delta\Sigma_1(m_P^2) + A_0\right)}_{\equiv a \sim \Lambda^2} - (k^2-m_P^2)A_1 - (k^2-m_P^2)^2 A_2 + \cdots \tag{2.39}$$

where now $A_n \equiv \frac{1}{n!}\left(\frac{\partial}{\partial k^2}\right)^n \Sigma_2(k^2)|_{k^2=m_P^2}$. The $\cdots$ here includes both higher orders in g ($\mathcal{O}(g^3)$) and higher powers of k^2, *i.e.* higher derivative terms.

Therefore, the propagator is

$$D(k) = \frac{1}{(1-A_1)(k^2-m_P^2)} + \cdots = \frac{Z}{k^2-m_P^2} + \cdots$$

with

$$Z = \frac{1}{1-A_1}, \quad m_P^2 = m^2 + a$$

where a was defined in (2.39).

Some points to notice: • $\delta Z = A_1$. (The higher-order bits we have no right to keep, since they are the same order as the 3-loop correction.)

[End of Lecture 6]

• The contributions $A_{n\geq 2}(k^2)^n$ can be reproduced by counterterms of the form $A_n \phi \Box^n \phi$. Had they been cutoff dependent we would have needed to add such (cutoff-dependent) counterterms.

• The mass-squared of the scalar field in $D = 3+1$ is quadratically divergent, while the mass of the spinor field (in QED, and also in the Yukawa theory) is only log

[18] Actually there is another diagram at two loops, namely  . Like the one-loop diagram, this does not depend on the external momentum, and so just renormalizes the mass. It should be included in A_0.

divergent. This UV sensitivity of scalar fields is ubiquitous[19] (see the homework) and is the source of many headaches.

• On the term 'wavefunction renormalization': who is ϕ? Also just a theorist's letter. Sometimes (in condensed matter) it is defined by some relation to observation (like the height of a wave in a mattress), in high energy theory not so much. Classically, we fixed its (multiplicative) normalization by setting the coefficient of $\phi\Box\phi$ to one. If we want to restore that convention after renormalization, we can make a redefinition of the field $\phi_P \equiv Z^{-1/2}\phi$. This is the origin of the term 'wavefunction renormalization'. A slightly better name would be 'field renormalization', but even better would be just 'kinetic term renormalization'.

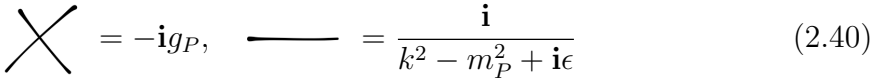
Renormalized perturbation theory revisited. The full story for the renormalized perturbation expansion in ϕ^4 theory is then

$$\mathcal{L} = \frac{1}{2}\left(\partial\phi\right)^2 - \frac{1}{2}m_P^2\phi^2 - \frac{g_P}{4!}\phi^4 + \mathcal{L}_{ct}$$

with

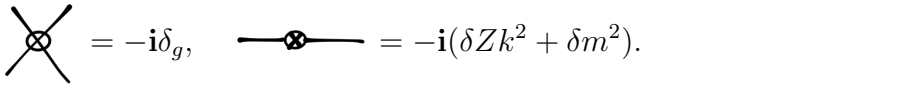
$$\mathcal{L}_{ct} = \frac{1}{2}\delta Z\left(\partial\phi\right)^2 - \frac{1}{2}\delta m^2\phi^2 - \frac{\delta_g}{4!}\phi^4.$$

Here are the instructions for using it: The Feynman rules are as before: the coupling and propagator are



(2.40)

but the terms in $\mathcal{L}_{ct}$ (the counterterms) are treated as new vertices, and treated perturbatively:



All integrals are regulated, in the same way (whatever it is). The counterterm couplings $\delta_g, \delta Z, \delta m^2$ are determined iteratively, as follows: given the δ_{N-1}s up to $\mathcal{O}(g_P^N)$, we fix each one $\delta = \delta_{N-1} + g_P^N\Delta\delta_N + \mathcal{O}(g_P^{N+1})$ by demanding that (2.40) are actually true up to $\mathcal{O}(g_P^{N+1})$. This pushes the cutoff dependence back into the muck a bit further.

I say this is the full story, but wait: we didn't try to compute amplitudes with more than four ϕs (such as $3 \leftarrow 3$ scattering of ϕ quanta). How do we know those don't require new counterterms (like a ϕ^6 term, for example)?

[19] At least for most regulators. We'll see that dim reg is special.

2.5.2 Where is the UV sensitivity?

[still Zee §III.3, Peskin ch. 10. We'll follow Zee's discussion pretty closely for a bit.] Given some process in a relativistic, perturbative QFT, how do we know if it will depend on the cutoff? We'd like to be able answer this question for a theory with scalars, spinors, vectors. Here's how: First, look at each diagram $\mathcal{A}$ (at some order in the loop expansion). Define the 'superficial' degree of divergence of $\mathcal{A}$ to be

$$D_{\mathcal{A}} \equiv [\mathcal{A}|_{\text{all coupling constants} = 1}], \tag{2.41}$$

its engineering dimension, ignoring the contributions from coupling constants. In the limit that $\Lambda \gg$ all other scales, we must then have $\mathcal{A} \sim \Lambda^{D_{\mathcal{A}}}$ (in the absence of cancellations). A log divergent amplitude has $D_{\mathcal{A}} = 0$ (sometimes it's called $D_{\mathcal{A}} = 0^+$).

Let's start simple, and study the ϕ^4 theory in $D = 4$. Consider a connected diagram $\mathcal{A}$ with B_E external scalar lines. I claim that $D_{\mathcal{A}} = 4 - B_E$.

Why doesn't it (explicitly) depend on any other data of the diagram, such as

$B_I \equiv$ # of internal scalar lines (*i.e.*, propagators)
$V \equiv$ # of ϕ^4 vertices
$L \equiv$ # of loops

$B_I = 8$
$B_E = 4$
$V = 5$
$L = 4$

? We can understand this better using two facts of graph theory and some power counting. I recommend checking my claims below with an example, such as the one at right.

Graph theory fact #1: These quantities are not all independent. For a connected graph,

$$L = B_I - (V - 1). \tag{2.42}$$

Proof[20]: Imagine placing the vertices on the page and adding the propagators one at a time. You need $V - 1$ internal lines just to connect up all V vertices. After that, each *internal* line you add necessarily adds one more loop. ■

Another way to think about this fact makes clear that $L =$ # of loops = # of momentum integrals. Before imposing momentum conservation at the vertices, each internal line has a momentum that we must integrate: $\prod_{\alpha=1}^{B_I} \int \mathrm{d}^D q_\alpha$. We then stick a $\delta^{(D)}(\sum q)$ for each vertex, but one of these gives the overall momentum conservation

[20] I learned this one from my class-mate M.B. Schulz.

$\delta^{(D)}(k_T)$, so we have $V-1$ fewer momentum integrals. For the example above, (2.42) says $4 = 8 - (5-1)$.

Graph theory fact #2: Each external line comes out of one vertex. Each internal line connects two vertices. Altogether, the number of ends of lines sticking out of vertices in a connected Feynman diagram in ϕ^4 theory is

$$B_E + 2B_I = 4V$$

where the RHS comes from noting that each vertex has four lines coming out of it (in ϕ^4 theory). In the example, this is $4 + 2 \cdot 8 = 4 \cdot 5$. So we can eliminate

$$B_I = 2V - B_E/2. \tag{2.43}$$

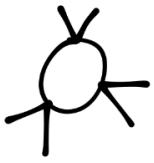
Now we count powers of momenta:

$$\mathcal{A} \overset{\Lambda \gg \text{everyone}}{\sim} \prod_{a=1}^{L} \int^{\Lambda} \text{đ}^D k_a \prod_{\alpha=1}^{B_I} \frac{1}{k_\alpha^2}.$$

Since we are about to apply the definition of the superficial degree of divergence, we ignore the factors from coupling constants. Since we are interested in the UV structure, I've set the mass to zero, as well as all the external momenta. The only scale left in the problem is the cutoff, so the dimensions of $\mathcal{A}$ must be made up by the cutoff:

$$\begin{aligned} D_{\mathcal{A}} = [\mathcal{A}] &= DL - 2B_I \\ &\overset{(2.42)}{=} B_I(D-2) - D(V-1) \\ &\overset{(2.43)}{=} D + \frac{2-D}{2} B_E + V(D-4). \end{aligned}$$

If we set $D = 3+1 = 4$, we get $D_{\mathcal{A}} = 4 - B_E$, as claimed. Notice that with $B_E = 2$ we indeed reproduce $D_{\mathcal{A}} = 2$, the quadratic divergence in the mass renormalization, and with $B_E = 4$ we get $D_{\mathcal{A}} = 0$, the log divergence in the $2 \leftarrow 2$ scattering. This pattern continues: with more than four external legs, $D_{\mathcal{A}} = 4 - B_E < 0$, which means the cutoff dependence must go away when $\Lambda \to 0$. This is illustrated by the following diagram with $B_E = 6$:



$$\sim \int^{\Lambda} \frac{\text{đ}^4 P}{P^6} \sim \Lambda^{-2}.$$

So indeed we don't need more counterterms for higher-point interactions in this theory.

Why is the answer independent of V in $D = 4$? This has the dramatic consequence that once we fix up the cutoff dependence in the one-loop diagrams, the higher orders have to work out, *i.e.* it strongly suggests that the theory is renormalizable. [21]

Before we answer this, let's explore the pattern a bit more. Suppose we include also a fermion field ψ in our field theory, and suppose we couple it to our scalar by a Yukawa interaction:

$$S_{\text{bare}}[\phi, \psi] = -\int d^D x \left(\frac{1}{2}\phi \left(\Box + m_\phi^2\right) \phi + \bar{\psi} \left(-\partial\!\!\!/ + m_\psi\right) \psi + y\phi\bar{\psi}\psi + \frac{g}{4!}\phi^4 \right).$$

To find the degree of divergence in an amplitude in this model, we have to independently keep track of the number fermion lines F_E, F_I, since a fermion propagator has dimension $[\frac{1}{\not{p}}] = -1$, so that $D_{\mathcal{A}} = [\mathcal{A}] = DL - 2B_I - F_I$. The number of ends-of-fermion-lines is $2V_y = F_E + 2F_I$ and the number of ends-of-boson-lines is $V_y + 4V_g = B_E + 2B_I$. The number of loops is $L = B_I + F_I - (V_y + V_g - 1)$. Putting these together (I used Mathematica) we get[22]

$$D_{\mathcal{A}} = D + (D-4)\left(V_g + \frac{1}{2}V_y\right) + B_E\left(\frac{2-D}{2}\right) + F_E\left(\frac{1-D}{2}\right). \tag{2.45}$$

Again in $D = 4$ the answer is independent of the number of vertices! Is there something special about four spacetime dimensions?

To temper your enthusiasm, consider adding a four-fermion interaction: $G(\bar{\psi}\psi)(\bar{\psi}\psi)$ (or maybe $G_V(\bar{\psi}\gamma^\mu\psi)(\bar{\psi}\gamma_\mu\psi)$ or $G_A(\bar{\psi}\gamma^\mu\gamma^5\psi)(\bar{\psi}\gamma_\mu\gamma^5\psi)$ or any other pile of gamma matrices in between, with the indices contracted). When you redo this calculation on the homework, you'll find that in $D = 4$ a diagram (for simplicity, one with no ϕ^4 or

[21] Why isn't it a proof of renormalizability? Consider the following integral:

$$\mathcal{I} = \int^\Lambda \frac{d^4p}{(p^2+m^2)^5} \int^\Lambda d^4k.$$

According to our method of counting, we would say $D_{\mathcal{I}} = 4 + 4 - 10 = -2$ and declare this finite and cutoff-independent. On the other hand, it certainly does depend on the physics at the cutoff, since the first factor has a finite part $\sim m^{-6}$. The rest of the work involving 'nested divergences' and forests is in showing that the extra structure in the problem prevents things like $\mathcal{I}$ from being Feynman amplitudes.

[22] A better way to write this formula is

$$D_{\mathcal{A}} = D + [g]V_g + [y]V_y - B_E[\phi] - F_E[\phi]\ . \tag{2.44}$$

From this way of writing it, you can see how it generalizes to any theory. See the homework. Thanks to Haoran Sun for pointing this out.

Yukawa interactions) has

$$D_{\mathcal{A}} = 4 - (1)B_E - \left(\frac{3}{2}\right) F_E + 2V_G,$$

where V_G is the number of insertions of the 4-fermion term. This dependence on the number of four-fermi vertices means that there are worse and worse divergences as we look at higher-order corrections to a given process. Even worse, it means that for any number of external lines F_E no matter how big, there is a large enough order in perturbation theory in G where the cutoff will appear! This means we need $\delta_n(\bar{\psi}\psi)^n$ counterterms for every n, which we'll need to fix with physical input. This is a bit unappetizing, and such an interaction is called "non-renormalizable". However, when we remember that we only need to make predictions to a given precision (so that we only need to go to a finite order in this process) we will see that such theories are nevertheless quite useful. (In fact, they predict their own demise. A realization of such a relativistic theory with 4-fermion interactions is Fermi's theory of the weak interactions. It tells us that it breaks down at the energy scale where we find the W and Z bosons.)

So why were those other examples independent of V? It's because the couplings were dimensionless. Those theories were classically scale invariant (except for the mass terms).

2.5.3 Naive scale invariance in scalar field theory

[Halpern] Consider a field theory of a scalar field ϕ in D spacetime dimensions, with an action of the form

$$S[\phi] = \int d^D x \left(\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - g\phi^p\right)$$

for some constants p, g. Which value of p makes this scale invariant? (That is: when is g dimensionless, and hence possibly the coupling for a renormalizable interaction.)

Naive mass dimensions:

$$[S] = [\hbar] = 0, \quad [x] \equiv -1, \quad [d^D x] = -D, \quad [\partial] = 1$$

The kinetic term tells us the engineering dimensions of ϕ:

$$0 = [S_{\text{kinetic}}] = -D + 2 + 2[\phi] \quad \Longrightarrow \quad [\phi] = \frac{D-2}{2}.$$

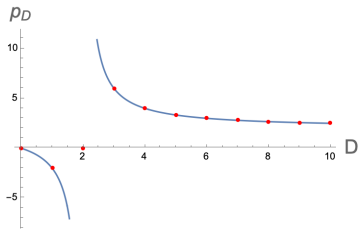
Notice that the $D = 1$ case agrees with our quantum mechanics counting from §1.1. Quantum field theory in $D = 1$ spacetime dimensions *is* quantum mechanics.

Then the self-interaction term has dimensions

$$0 = [S_{\text{interaction}}] = -D + [g] + p[\phi] \implies [g] = D - p[\phi] = D + p\frac{2-D}{2}$$

We expect scale invariance when $[g] = 0$, which happens when

$$\boxed{p = p_D \equiv \frac{2D}{D-2}},$$



i.e. the scale invariant scalar-field self-interaction in D spacetime dimensions is $\phi^{\frac{2D}{D-2}}$.

D	1	2	3	4	5	6	...	D	∞
$[\phi]$	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	3/2	2	...	$\frac{D-2}{2}$	∞
scale-inv't $p \equiv p_D$	-2	$\infty\star$	6	4	10/3	3	...	$\frac{2D}{D-2}$	2

$\star$ What is happening in $D = 2$? The field is dimensionless, and so *any* power of ϕ is naively scale invariant, as are more complicated interactions like $g_{ij}(\phi)\partial_\mu\phi^i\partial^\mu\phi^j$, where the coupling $g(\phi)$ is a function of ϕ. This allows for scale-invariant *non-linear sigma models*, where the fields are coordinates on a curved manifold with metric $ds^2 = g_{ij}(\phi)d\phi^i d\phi^j$.

In dimensions where we get fractional powers, this isn't so nice.

Notice that the mass term $\Delta S = \int d^D x \frac{m^2}{2}\phi^2$ gives

$$0 = -D + 2[m] + 2[\phi] \quad \implies \quad [m] = 1 \quad \forall D < \infty$$

– it's a mass, yay.

What are the consequences of this engineering-dimensions calculation in QFT? For $D > 2$, an interaction of the form $g\phi^p$ has

$$[g] = D \cdot \frac{p_D - p}{p_D} \begin{cases} < 0 \text{ when } p > p_D, & \text{non-renormalizable or irrelevant} \\ = 0 \text{ when } p = p_D, & \text{renormalizable or marginal} \\ > 0 \text{ when } p < p_D, & \text{super-renormalizable or relevant.} \end{cases} \tag{2.46}$$

Consider the 'non-renormalizable' case. Suppose we calculate in QFT some quantity f (say a scattering amplitude) with $[f]$ as its naive dimension, in perturbation theory in g, *e.g.* by Feynman diagrams. We'll get:

$$f = \sum_{n=0}^{\infty} g^n c_n$$

with c_n independent of g. So

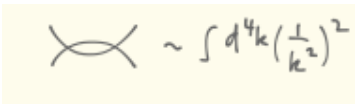
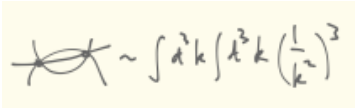
$$[f] = n[g] + [c_n] \implies [c_n] = [f] - n[g]$$

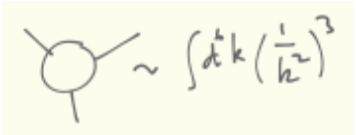
So if $[g] < 0$, c_n must have more and more powers of some mass (inverse length) as n increases. What dimensionful quantity makes up the difference? Sometimes it is masses or external momenta. But generically, it gets made up by UV divergences (if everything is infinite, dimensional analysis can fail, nothing is real, I am the walrus). More usefully, in a meaningful theory with a UV cutoff, Λ_{UV}, the dimensions get made up by the UV cutoff, which has $[\Lambda_{UV}] = 1$. Generically (*i.e.* in the absence of cancellations): $c_n = \tilde{c}_n \left(\Lambda_{UV}\right)^{-n[g]}$, where $\tilde{c}_n$ is dimensionless, and $n[g] < 0$ – it's higher and higher powers of the cutoff.

Consider the renormalizable (classically scale invariant) case: $[c_n] = [f]$, since $[g] = 0$. But in fact, what you'll get is something like

$$c_n = \tilde{c}_n \log^{\nu(n)}\left(\frac{\Lambda_{UV}}{\Lambda_{IR}}\right) \ ,$$

where Λ_{IR} is an infrared cutoff or a mass or external momentum, $[\Lambda_{IR}] = 1$. Some classically scale invariant examples (so that $m = 0$ and the bare propagator is $1/k^2$) where you can see that we get logs from loop amplitudes:

ϕ^4 in $D = 4$:  ϕ^6 in $D = 3$: 

ϕ^3 in $D = 6$:  In $D = 2$, even the propagator for a massless scalar field has logs:

$$\langle \phi(x)\phi(0)\rangle = \int \mathrm{d}^2k \frac{e^{-\mathbf{i}kx}}{k^2} \sim \log\frac{|x|}{\Lambda_{UV}}.$$

The terms involving 'renormalizable' in (2.46) are somewhat old-fashioned and come from a high-energy physics point of view where the short-distance physics is unknown, and we want to get as far as we can in that direction with our limited knowledge (in which case the condition 'renormalizability' lets us get away with this indefinitely – it lets us imagine we know everything). The latter terms are natural in the opposite situation (like condensed matter physics) where we know some basically correct microscopic description but want to know what happens at low energies. Then an operator like $\frac{1}{M^{24}}\phi^{28}$ whose coefficient is suppressed by some large mass scale M is *irrelevant* for physics at energies far below that scale. Inversely, an operator like $m^2\phi^2$ gives a mass to the ϕ particles, and matters very much (is relevant) at energies $E < m$. In the marginal case, the quantum corrections have a chance to make a big difference.

[End of Lecture 7]

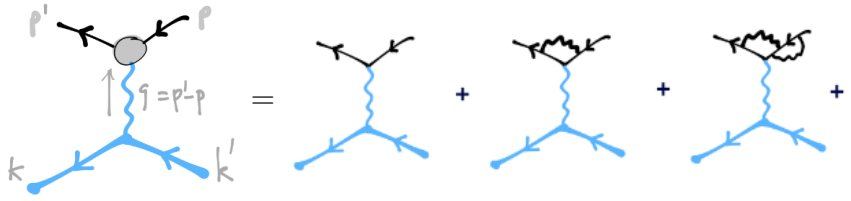
2.6 Vertex correction in QED and soft photons

[Peskin chapter 6, Schwartz chapter 17, Zee chapter III.6] Back to work on QED. The vertex correction  has some great physics payoffs:

- We'll cancel the cutoff dependence we found in the S matrix from δZ.
- We'll compute $g-2$ (the anomalous magnetic moment) of the electron, the locus of some of the most precise agreement between theory and experiment. (Actually the agreement is so good that it's used as the *definition* of the fine structure constant. But a similar calculation gives the leading anomalous magnetic moment of the muon.)
- We'll see that the exclusive differential cross section $\left(\frac{d\sigma}{d\Omega}\right)_{e\mu\leftarrow e\mu}$ that we've been considering is not really an observable. Actually it is infinity![23] The key word here is 'exclusive,' which means that we demand that the final state is *exactly* one electron and one muon and *absolutely nothing else*. Think for a moment about how you might do that measurement.

  This is an example of an IR divergence. While UV divergences mean we're overstepping our bounds (by taking too seriously our Lagrangian parameters or our knowledge of short distances), IR divergences mean we are *asking the wrong question*.

To get started, consider the following class of diagrams.



$$\equiv \mathbf{i}\mathcal{M} \qquad = \mathbf{i}e^2\left(\bar{u}(p')\Gamma^\mu(p,p')u(p)\right)\frac{1}{q^2}\bar{u}(K')\gamma_\mu u(K) \tag{2.47}$$

The shaded blob is the *vertex function* Γ. The role of the light blue factors is just to make and propagate the photon that hits our electron; let's forget about them. Denote the photon momentum by $q = p' - p$. We'll assume that the electron momenta p, p' are on-shell, but q^μ is not, as in the $e\mu$ scattering process. Then $q^2 = 2m^2 - 2p'\cdot p$.

[23] More accurately, the exclusive cross section is zero; the one-loop correction is minus infinity, which is perturbation theory's clumsy attempt to correct the finite tree level answer to make it zero.

Before calculating the leading correction to the vertex $\Gamma^\mu = \gamma^\mu + \mathcal{O}(e^2)$, let's think about what the answer can be. It is a vector made from $p^\mu, (p')^\mu, \gamma^\mu$ and m, e and numbers. It can't have any γ^5 or $\epsilon^{\mu\nu\rho\sigma}$ by parity symmetry of QED. So on general grounds we can organize it as

$$\Gamma^\mu(p,p') = A\gamma^\mu + B(p+p')^\mu + C(p-p')^\mu \tag{2.48}$$

where A, B, C are Lorentz-invariant functions of $p^2 = (p')^2 = m^2$, $p\cdot p'$, $\not{p}, \not{p}'$. But, for example, $\not{p}\gamma^\mu u(p) = (m\gamma^\mu - p^\mu)u(p)$ which just mixes up the terms; really A, B, C are just functions of the momentum transfer q^2. Gauge invariance, in the form of the Ward identity, says that contracting the photon line with the photon momentum should give zero:

$$0 \overset{\text{Ward}}{=} q_\mu \bar{u}(p')\Gamma^\mu u(p) \overset{(2.48)}{=} \bar{u}(p')\left(A \underbrace{\not{q}}_{=\not{p}'-\not{p}\overset{\bar{u}(p')\ldots u(p)}{=} m-m=0} + B\underbrace{(p+p')\cdot(p-p')}_{=m^2-m^2=0} + Cq^2 \right) u(p)$$

Therefore $0 = Cq^2\bar{u}(p')u(p)$ for general q^2 and general spinors, so $C = 0$. This is the moment for the Gordon identity to shine:

$$\bar{u}(p')\gamma^\mu u(p) = \bar{u}(p')\left(\frac{p^\mu + p'^\mu}{2m} + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}\right)u(p)$$

(where $\sigma^{\mu\nu} \equiv \frac{\mathbf{i}}{2}[\gamma^\mu, \gamma^\nu]$) can be used to eliminate the $p+p'$ term[24]. The Gordon identity shows that the QED interaction vertex $\bar{u}(p')\gamma^\mu u(p)A_\mu$ contains a magnetic moment bit in addition to the $p + p'$ term (which is there for a charged scalar field).

It is then convenient (and conventional) to parametrize the vertex in terms of the two *form factors* $F_{1,2}$:

$$\Gamma^\mu(p,p') = \gamma^\mu F_1(q^2) + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}F_2(q^2). \tag{2.49}$$

This little monstrosity has the complete information about the coupling of the electron to the electromagnetic field, such as for example a background electromagnetic field. It is a parametrization of the matrix elements of the current between two one-electron states, incorporating the facts of gauge invariance and Lorentz invariance.

[24] Actually this is why we didn't include a $\sigma^{\mu\nu}$ term. You could ask: what about a term like $\sigma^{\mu\nu}(p+p')^\nu$? Well, there's another Gordon identity that relates that to things we've already included:

$$\bar{u}_2\sigma_{\mu\nu}(p_1+p_2)^\nu u_1 = \mathbf{i}\bar{u}_2\left(q_\mu - (m_1-m_2)\gamma_\mu\right)u_1.$$

It is proved the same way: just use the Dirac equation $\not{p}_1 u_1 = m_1 u_1, \bar{u}_2\not{p}_2 = \bar{u}_2 m_2$ and the Clifford algebra. We are interested here in the case where $m_1 = m_2$.

The first term at zero momentum $eF_1(q^2=0)$ is the electric charge of the electron (if you don't believe it, use the vertex (2.49) to calculate the Coulomb field of the electron; there are some details on page 186 of Peskin). Since the tree-level bit of F_1 is 1, if by the letter e here we mean the actual charge, then we'd better include counterterms ($\mathcal{L}_{ct} \ni \bar{\psi}\delta_e\gamma^\mu A_\mu\psi$) to make sure it isn't corrected: $F_1(0)=1$.

The magnetic moment of the electron is the coefficient $\vec{\mu}$ of $\tilde{V}(q) = -\vec{\mu}\cdot\vec{B}(q) + ...$ in the non-relativistic effective potential. Comparing the non-relativistic limit of $\bar{u}(p')\Gamma^i u(p)A_i(q) = -\vec{\mu}\cdot\vec{B}(q) + ...$, shows that (see Peskin p. 187)

$$\vec{\mu} = g\frac{e}{2m}\vec{S},$$

where $\vec{S} \equiv \xi^\dagger\frac{\vec{\sigma}}{2}\xi$ is the electron spin. Comparing with the vertex function, this says that the g factor is

$$g = 2(F_1(0)+F_2(0)) = 2+2F_2(0) = 2+\mathcal{O}(\alpha).$$

We see that the anomalous magnetic moment of the electron is $g-2 = 2F_2(q^2=0)$.

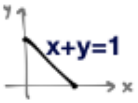
Now that we have some expectation about the form of the answer, and some ideas about what it's for, we sketch the evaluation of the one-loop QED vertex correction:

$$= -(e)^3\int \mathrm{d}^4k\,\bar{u}(p')\gamma^\nu\frac{\not{k}'+m_e}{(k')^2-m_e^2}\gamma^\mu\frac{\not{k}+m_e}{k^2-m_e^2}\gamma^\rho u(p)\cdot\frac{\eta_{\nu\rho}}{(p-k)^2-m_\gamma^2}$$

with $k' \equiv k+q$.

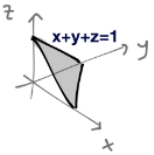
Step (1) Feynman parameters again. The one we showed before can be rewritten more symmetrically as:

$$\frac{1}{AB} = \int_0^1 dx\int_0^1 dy\;\delta(x+y-1)\frac{1}{(xA+yB)^2}$$



Now how can you resist the generalization[25]:

$$\frac{1}{ABC} = \int_0^1 dx \int_0^1 dy \int_0^1 dz\ \delta(x+y+z-1)\frac{2}{(xA+yB+zC)^3}$$



So, set $A = (k')^2 - m_e^2, B = k^2 - m_e^2, C = (p-k)^2 - m_\gamma^2$ (with the appropriate $\mathbf{i}\epsilon$s), so that the integral we have to do is

$$\int \frac{\mathrm{đ}^4 k N^\mu}{(k^2 + k\cdot(\cdots) + \cdots)^3}.$$

Step (2) Complete the square, $\ell = k - zp + xq$ to get $\int \frac{\mathrm{đ}^4 \ell N^\mu}{(\ell^2-\Delta)^3}$ where

$$\Delta = -xyq^2 + (1-z)^2 m^2 + z m_\gamma^2. \tag{2.51}$$

The ℓ-dependence in the numerator is either 1 or ℓ^μ or $\ell^\mu\ell^\nu$. In the integral over ℓ, the second averages to zero, and the third averages to $\eta^{\mu\nu}\ell^2\frac{1}{4}$. As a result, the momentum integrals we need are just

$$\int \frac{\mathrm{đ}^D \ell}{(\ell^2-\Delta)^m} \quad \text{and} \quad \int \frac{\mathrm{đ}^D \ell\ \ell^2}{(\ell^2-\Delta)^m}\ .$$

Right now we only need $D = 4$ and $m = 3$, but it turns out to be quite useful to think about them all at once. Like in our discussion of the electron self-energy diagram, we

[25] Peskin outlines a proof by induction of the whole family of such identities on page 190. But here's a simpler proof using *Schwinger parameters*. You'll agree that

$$\frac{1}{A} = \int_0^\infty ds\ e^{-sA}. \tag{2.50}$$

Applying this identity to each factor gives

$$\frac{1}{A_1 A_2 \cdots A_n} = \int_0^\infty ds_1 \cdots \int_0^\infty ds_n\ e^{-\sum_{i=1}^n s_i A_i}.$$

Now use scaling to set $\tau \equiv \sum_{i=1}^n s_i$, and $x_i \equiv s_i/\tau$. Then

$$\frac{1}{A_1 A_2 \cdots A_n} = \int_0^\infty d\tau \tau^{n-1} \prod_{i=1}^n \int_0^1 dx_i \delta\left(\sum_{i=1}^n x_i - 1\right)\ e^{-\tau\sum_i x_i A_i}.$$

Now do the integral over τ, using $\int_0^\infty d\tau\tau^{n-1}e^{-\tau X} = \frac{(n-1)!}{X^n}$ (differentiate (2.50) wrt A), to arrive at

$$\frac{1}{A_1 A_2 \cdots A_n} = \prod_{i=1}^n \int_0^1 dx_i \delta\left(\sum_{j=1}^n x_j - 1\right) \frac{(n-1)!}{\left(\sum_j x_j A_j\right)^n}.$$

can evaluate them by
Step (3) Wick rotating (which changes the denominator to $\ell_E^2 + \Delta$) and going to polar coordinates. This gives:

$$\int \frac{\mathrm{đ}^D \ell}{\left(\ell^2 - \Delta\right)^m} = (-1)^m \frac{\mathbf{i}}{(4\pi)^{D/2}} \frac{\Gamma\left(m - \frac{D}{2}\right)}{\Gamma(m)} \left(\frac{1}{\Delta}\right)^{m - \frac{D}{2}}. \tag{2.52}$$

$$\int \frac{\mathrm{đ}^D \ell \; \ell^2}{\left(\ell^2 - \Delta\right)^m} = (-1)^{m-1} \frac{D}{2} \frac{\mathbf{i}}{(4\pi)^{D/2}} \frac{\Gamma\left(m - \frac{D}{2} - 1\right)}{\Gamma(m)} \left(\frac{1}{\Delta}\right)^{m - \frac{D}{2} - 1}. \tag{2.53}$$

Notice that these integrals are not equal to infinity when the parameter D is not an integer. This is the idea behind *dimensional regularization*.

Step (0) But for now let's persist in using the Pauli Villars regulator. (I call this step (0) instead of (4) because it should have been there all along.) Here this means we subtract from the amplitude the same quantity with m_γ^2 replaced by Λ^2. The dangerous bit comes from the ℓ^2 term we just mentioned, since $m - D/2 - 1 = 3 - 4/2 - 1 = 0$ means logs.

The numerator is

$$\begin{aligned} N^\mu &= \bar{u}(p')\gamma^\nu \left(\not{k} + \not{q} + m_e\right) \gamma^\mu \left(\not{k} + m_e\right) \gamma_\nu u(p) \\ &= -2 \left(\mathcal{A}\bar{u}(p')\gamma^\mu u(p) + \mathcal{B}\bar{u}(p')\sigma^{\mu\nu} q_\nu u(p) + \mathcal{C}\bar{u}(p') q^\mu u(p)\right) \end{aligned} \tag{2.54}$$

where

$$\begin{aligned} \mathcal{A} &= -\frac{1}{2}\ell^2 + (1-x)(1-y)q^2 + (1 - 4z + z^2)m^2 \\ \mathcal{B} &= \mathbf{i} m z (1-z) \\ \mathcal{C} &= m(z-2)(y-x) \; . \end{aligned} \tag{2.55}$$

Great effort was expended to arrive at these simple expressions (actually most of the algebra is done explicitly on page 319 of Schwartz). Now you say: but you promised there would be no term like $\mathcal{C}$ because of the Ward identity. Indeed I did and indeed there isn't, because $\mathcal{C}$ is odd in $x \leftrightarrow y$ while everything else is even, so this term integrates to zero.

The first term (with $\mathcal{A}$) is a correction to the charge of the electron and will be UV divergent. More explicitly, we get, using Pauli-Villars,

$$\int \mathrm{đ}^4 \ell \left(\frac{\ell^2}{\left(\ell^2 - \Delta_{m_\gamma}\right)^3} - \frac{\ell^2}{\left(\ell^2 - \Delta_\Lambda\right)^3} \right) = \frac{\mathbf{i}}{(4\pi)^2} \ln \frac{\Delta_\Lambda}{\Delta_{m_\gamma}}.$$

The other bits are finite, and we ignore the terms that go like negative powers of Λ. More on this cutoff dependence soon. But first something wonderful:

2.6.1 Anomalous magnetic moment

The second term $\mathcal{B}$ contains the anomalous magnetic moment:

$$\begin{aligned} F_2(q^2) &= \frac{2m}{e} \cdot (\text{the term with } \mathcal{B}\ ) \\ &= \frac{2m}{e} 4e^3 \cdot m \cdot \int dxdydz\delta(x+y+z-1) z(1-z) \underbrace{\int \frac{\text{đ}^4\ell}{(\ell^2-\Delta)^3}}_{=\frac{-\mathbf{i}}{32\pi^2\Delta}} \\ &= \frac{\alpha}{\pi} m^2 \int dxdydz\delta(x+y+z-1) \frac{z(1-z)}{(1-z)^2 m^2 - xyq^2}. \end{aligned} \tag{2.56}$$

The correction to the magnetic moment is the long-wavelength bit of this:

$$F_2(q^2=0) = \frac{\alpha}{\pi} m^2 \int_0^1 dz \int_0^{1-z} dy \frac{z}{(1-z)m^2} = \frac{\alpha}{2\pi}.$$

$$\boxed{g = 2 + \frac{\alpha}{\pi} + \mathcal{O}(\alpha^2).}$$

A rare opportunity for me to plug in numbers: $g = 2.00232$.

2.6.2 IR divergences mean wrong questions.

There is a term in the numerator from the $\mathcal{A}\gamma^\mu$ bit

$$\int \frac{\text{đ}^4\ell}{(\ell^2-\Delta)^3} = c\frac{1}{\Delta}$$

(with $c = -\frac{\mathbf{i}}{32\pi^2}$ again), but without the factor of $z(1-z)$ we had in the magnetic moment calculation. It looks like we've gotten away without having to introduce a UV regulator here, too (so far). But now look at what happens when we try to do the Feynman parameter integrals. For example, at $q^2 = 0$, we get (if we had set $m_\gamma = 0$)

$$\begin{aligned} \int dxdydz\delta(x+y+z-1) \frac{m^2(1-4z+z^2)}{\Delta} &= m^2 \int_0^1 dz \int_0^{1-z} dy \frac{-2+2(1-z)+(1-z)^2}{(1-z)^2 m^2} \\ &= \int_0^1 dz \frac{-2}{(1-z)} + \text{finite}, \end{aligned} \tag{2.57}$$

which diverges at the upper limit of integration. In fact it's divergent even when $q^2 \neq 0$. This is a place where we actually need to include the photon mass, m_γ, for our own safety. The fact that restoring $m_\gamma \neq 0$ in (2.51) regulates this divergence is one way to see that it is indeed an IR divergence.

The (IR singular bit of the) vertex (to $\mathcal{O}(\alpha)$) is of the form

$$\Gamma^\mu = \gamma^\mu \left(1 - \frac{\alpha}{2\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{{m_\gamma}^2}\right)\right) + \text{ stuff that is finite as } m_\gamma \to 0. \tag{2.58}$$

Notice that the IR divergent stuff depends on the electron momenta p, p' only through q, the momentum of the photon. So it looks like we are led to conclude

$$\left(\frac{d\sigma}{d\Omega}\right)_{\mu e \leftarrow \mu e} = \left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} \left(1 - \frac{\alpha}{\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{{m_\gamma}^2}\right)\right) + \mathcal{O}\left(\alpha^2\right)$$

which blows up when we remove the fake photon mass $m_\gamma \to 0$. Notice that for t-channel exchange, $-q^2 > 0$, so the argument of the log is positive, the cross-section is real. But notice that the one-loop correction is not only infinite, but *negative* infinity, which simply cannot happen from the definition of the cross section. This is perturbation theory's way of telling us that the answer is $1 - \alpha \cdot \infty + \mathcal{O}(\alpha^2) \simeq 0$ – the putatively small corrections from radiative effects are actually trying to make the answer *zero*.

[Schwartz §20.1] I wanted to just quote the above result for (2.58) but I lost my nerve, so here is a bit more detail leading to it. The IR dangerous bit comes from the second term in $\mathcal{A}$ above. That is,

$$F_1(q^2) = 1 + f(q^2) + \delta_e + \mathcal{O}(\alpha^2)$$

with

$$f(q^2) = \frac{e^2}{8\pi^2} \int_0^1 dx dy dz \delta(x+y+z-1) \left(\ln \frac{z\Lambda^2}{\Delta} + \frac{q^2(1-x)(1-y) + m_e^2(1-4z+z^2)}{\Delta}\right).$$

δ_e here is a counterterm coefficient for the $\Psi \gamma^\mu A_\mu \Psi$ vertex.

We can be more explicit if we consider $-q^2 \gg m_e^2$ so that we can ignore the electron mass everywhere. Then we would choose the counterterm δ_e so that

$$1 = F_1(0) \implies \delta_e = -f(0) \overset{m_e/q \to 0}{\to} -\frac{e^2}{8\pi^2} \frac{1}{2} \ln \frac{\Lambda^2}{m_\gamma^2}.$$

And the form of $f(q^2)$ is

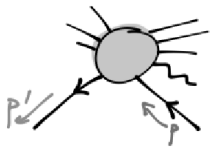
$$f(q^2)|_{m_e=0} = \frac{e^2}{8\pi^2} \int dx dy dz \delta(x+y+z-1) \left(\underbrace{\ln \frac{(1-x-y)\Lambda^2}{\Delta}}_{\text{IR finite}} + \frac{q^2(1-x)(1-y)}{-xyq^2 + (1-x-y)m_\gamma^2}\right)$$

$$F_1(q^2)|_{m_e=0} = 1 - \frac{e^2}{16\pi^2} \left(\ln^2 \frac{-q^2}{m_\gamma^2} + 3 \ln \frac{-q^2}{m_\gamma^2}\right) + \text{finite}.$$

In doing the integrals, we had to remember the $\mathbf{i}\epsilon$ in the propagators, which can be reproduced by the replacement $q^2 \to q^2 + \mathbf{i}\epsilon$. This $\ln^2(q^2/m_\gamma)$ is called a *Sudakov double logarithm*. Notice that taking differences of these at different q^2 will not make it finite.

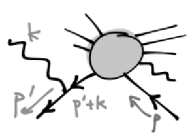
Diversity and inclusion to the rescue. Before you throw up your hands in despair, I would like to bring to your attention another consequence of the masslessness of the photon: It means that real (as opposed to virtual) photons can be made with arbitrarily low energy. But a detector has a minimum triggering energy: the detector works by particles doing some physical something to stuff in the detector, and it has a finite energy resolution – it takes a finite amount of energy E_c for those particles to do the stuff. This means that a process with exactly one e and one μ in the final state  cannot be distinguished from a process ending in $e\mu$ plus a photon of arbitrarily small energy, such as would result from  (final-state radiation) or  (initial-state radiation). This ambiguity is present for *any* process with external charged particles.

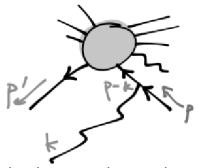
Being more inclusive, then, we cannot distinguish amplitudes of the form

$$\bar{u}(p')\mathcal{M}_0(p',p)u(p) \equiv -\mathbf{i}\left(\right.$$


$$\left. \right) ,$$

from more inclusive amplitudes like

$$-\mathbf{i}\left(\right.$$


$$+$$


$$\left. \right)$$

$$= \bar{u}(p')\gamma^\mu \tfrac{e}{\not{p}'+\not{k}-m_e}\mathcal{M}_0(p',p)u(p)\epsilon^\star_\mu(k) + \bar{u}(p')\mathcal{M}_0(p',p)\tfrac{e}{\not{p}-\not{k}-m_e}\gamma^\mu u(p)\epsilon^\star_\mu(k) \ .$$

Now, by assumption the extra outgoing photon is real (k^2=0) and it is *soft*, in the sense that $k^0 < E_c$, the detector cutoff. Notice that this condition is not Lorentz invariant, but depends on a choice of frame. If we boost a soft photon to a frame at .99999c, it is no longer soft. But the experiment we are trying to describe is also not Lorentz invariant: the condition $k^0 < E_c$ is imposed in the rest frame of the detector.

Because the photon is soft, $k^0 < E_c$, we can approximate the numerator of the second term as

$$\left(\not{p} - \not{k} + m_e\right)\gamma^\mu u(p) \simeq \left(\not{p} + m_e\right)\gamma^\mu u(p) \overset{\text{Clifford}}{=} (2p^\mu + \gamma^\mu \underbrace{\left(-\not{p} + m_e\right)}_{=0})u(p) = 2p^\mu u(p).$$

In the denominator we have *e.g.* $(p-k)^2 - m_e^2 = p^2 - m_e^2 - 2p\cdot k + k^2 \sim -2p\cdot k$ since the electron is on shell $p^2 = m_e^2$ and so is the photon $k^2 = 0$. Therefore

$$\mathcal{M}\left(e\mu + \text{one soft } \gamma \leftarrow e\mu\right) = e\bar{u}(p')\mathcal{M}_0(p',p)u(p)\left(\frac{p'\cdot\epsilon^\star}{p'\cdot k + \mathbf{i}\epsilon} - \frac{p\cdot\epsilon^\star}{p\cdot k - \mathbf{i}\epsilon}\right) \tag{2.59}$$

This is bremsstrahlung. Before we continue this calculation to find the inclusive amplitude that a real detector actually measures, let's pause to relate the previous expression to some physics we know. [End of Lecture 8]

Where have we seen this kind of expression

$$\frac{p'^\mu}{p'\cdot k + \mathbf{i}\epsilon} - \frac{p^\mu}{p\cdot k - \mathbf{i}\epsilon} \equiv \frac{1}{\mathbf{i}e}\tilde{j}^\mu(k)$$

before? Notice that the $\mathbf{i}\epsilon$ are different because one comes from final state and one from initial. Well, this object is the Fourier transform $\tilde{j}^\mu(k) = \int d^4x\, e^{+\mathbf{i}kx} j^\mu(x)$ of the current

$$j^\mu(x) = e\int d\tau \frac{dy^\mu}{d\tau}\delta^{(4)}(x - y(\tau))$$

associated with a particle which executes a piecewise linear motion [26]

$$y^\mu(\tau) = \begin{cases} \frac{p^\mu}{m}\tau, & \tau < 0 \\ \frac{p'^\mu}{m}\tau, & \tau > 0 \end{cases}.$$

This is a good approximation to the motion a free particle that experiences a sudden acceleration; sudden means that the duration of the pulse is short compared to ω^{-1} for any frequency we're going to measure. The electromagnetic radiation that such an accelerating charge produces is given classically by Maxwell's equation: $\tilde{A}^\mu(k) = -\frac{1}{k^2}\tilde{j}^\mu(k)$.

I claim further that the factor $\frac{\alpha}{\pi}f_{IR}(q^2) = \frac{\alpha}{\pi}\ln\left(\frac{-q^2}{m^2}\right)$ (which entered our lives in (2.58)) arises classically as the *number* of soft photons produced by such a process in each decade of wavenumber. You can figure this out by plugging $\tilde{A}^\mu(k) = -\frac{1}{k^2}\tilde{j}^\mu(k)$ into the electromagnetic energy $\frac{1}{2}\int d^3x\,(E^2 + B^2) = \int \mathrm{d}^3k\hbar\omega_k n_k$. (Note that the integral over k here actually diverges; this is an artifact of the approximation that the momentum change is instantaneous.) See Peskin §6.1 for help.

---

[26] Check it:

$$\int d^4x j^\mu(x)e^{+\mathbf{i}kx} = e\int d\tau \frac{dy^\mu(\tau)}{d\tau}e^{\mathbf{i}k\cdot y(\tau)} = e\int_0^\infty d\tau \frac{p'^\mu}{m}e^{\mathbf{i}\left(\frac{k\cdot p'}{m} + \mathbf{i}\epsilon\right)\tau} + e\int_{-\infty}^0 d\tau \frac{p^\mu}{m}e^{\mathbf{i}\left(\frac{k\cdot p}{m} - \mathbf{i}\epsilon\right)\tau} = \tilde{j}^\mu(k).$$

Notice that the $\mathbf{i}\epsilon$ are convergence factors in the Fourier transforms.

The cross section for $e\mu$ to go to $e\mu$ plus a single soft photon is then

$$\left(\frac{d\sigma}{d\Omega}\right)^{E_\gamma<E_c}_{\mu e\gamma_{\rm soft}\leftarrow\mu e} = \left(\frac{d\sigma}{d\Omega}\right)_{\rm Mott} e^2 \underbrace{\int_0^{E_c} \frac{\mathrm{d}^3k}{2E_k}}_{\gamma\text{ phase space}} \left|\frac{p\cdot\epsilon^\star}{p\cdot k} - \frac{p'\cdot\epsilon^\star}{p'\cdot k}\right|^2 \overset{E_k=|\vec{k}|}{\sim} \int_0 \frac{d^3k}{k^3} = \infty. \quad (2.60)$$

This is *another* IR divergence. (One divergence is bad news, but two is an opportunity for hope.) Just like we must stick to our UV regulators like religious zealots, we must cleave tightly to the consistency of our IR regulators: we need to put back the photon mass:

$$E_k = \sqrt{\vec{k}^2 + m_\gamma^2}$$

which means that the lower limit of the k integral gets cut off at m_γ:

$$\int_0^{E_c} \frac{dk}{E_k} = \left(\int_0^{m_\gamma} + \int_{m_\gamma}^{E_c}\right) \frac{dk}{\sqrt{k^2+{m_\gamma}^2}} \sim \underbrace{\int_0^{m_\gamma} \frac{dk}{m_\gamma}}_{=1} + \underbrace{\int_{m_\gamma}^{E_c} \frac{dk}{k}}_{\ln\frac{E_c}{m_\gamma}}.$$

Being careful about the factors, the actual cross section measured by a detector with energy resolution E_c is[27]

$$\begin{aligned}
\left(\frac{d\sigma}{d\Omega}\right)^{\rm observed} &= \left(\frac{d\sigma}{d\Omega}\right)_{e\mu\leftarrow\mu e} + \left(\frac{d\sigma}{d\Omega}\right)^{E_\gamma<E_c}_{\mu e\gamma_{\rm soft}\leftarrow\mu e} + \mathcal{O}(\alpha^3) \\
&= \left(\frac{d\sigma}{d\Omega}\right)_{\rm Mott} \left(1 \underbrace{-\frac{\alpha}{\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{{m_\gamma}^2}\right)}_{\text{vertex correction}} + \underbrace{\frac{\alpha}{\pi} f_{IR}(q^2) \ln\left(\frac{E_c^2}{{m_\gamma}^2}\right)}_{\text{soft photons}} \right) \\
&= \left(\frac{d\sigma}{d\Omega}\right)_{\rm Mott} \left(1 - \frac{\alpha}{\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{E_c^2}\right)\right)
\end{aligned}$$

The thing we can actually measure is independent of the IR regulator photon mass m_γ, and finite when we remove it. On the other hand, it depends on the detector resolution. Like in the plot of some kind of Disney movie, an apparently-minor character whom you may have been tempted to disregard as an ugly detail has saved the day.

[27] Notice that we add the cross-sections, not the amplitudes, for these processes with different final states. Here's why: even though we don't measure the existence of the photon, *something* does: it gets absorbed by some part of the apparatus or the rest of the world and therefore becomes entangled with some of its degrees of freedom; when we fail to distinguish between those states, we trace over them, and this erases the interference terms we would get if we summed the amplitudes.

Alternatively, a final state with no photon is orthogonal to any state with a photon. So the two terms in the wavefunction cannot interfere. Thanks to Eric Michelsen for the comment.

I didn't show explicitly that the coefficient of the log is the same function $f_{IR}(q^2)$. In fact this function is $f_{IR}(q^2) = \log(-q^2/m^2)$, so the product $f_{IR} \ln q^2 \sim \ln^2 q^2$ is the Sudakov double logarithm. A benefit of the calculation that shows that the same f_{IR} appears in both places (Peskin chapter 6.5) is that it also shows that this pattern persists at higher order in α: there is a $\ln^2(q^2/{m_\gamma}^2)$ dependence in the two-loop vertex correction, and a matching $-\ln^2(E_c^2/{m_\gamma}^2)$ term in the amplitude to emit *two* soft photons. There is a $\frac{1}{2!}$ from Bose statistics of these photons. The result exponentiates, and we get

$$e^{-\frac{\alpha}{\pi} f \ln(-q^2/{m_\gamma}^2)} e^{+\frac{\alpha}{\pi} f \ln(E_c^2/{m_\gamma}^2)} = e^{-\frac{\alpha}{\pi} f \ln(-q^2/E_c^2)}.$$

You may be bothered that I've made all this discussion about the corrections from the electron line, but said nothing about the muon line. But the theory should make sense even if the electron and muon charges Q_e, Q_m were different, so the calculation should make sense term-by-term in an expansion in Q_m.

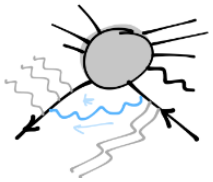
Some relevant names for future reference: The name for the guarantee that this always works in QED is the *Bloch-Nordsieck theorem*. Closely-related but more serious issues arise in QCD, the theory of quarks and gluons; this is the beginning of the story of *jets* (a jet is some IR-cutoff dependent notion of a QCD-charged particle plus the cloud of stuff it carries with it) and *parton distribution functions*.

Sketch of exponentiation of soft photons. [Peskin §6.5] In the following we will just keep track of the bits that diverge when $m_\gamma \to 0$. Consider a diagram with n soft external photons with momenta $\{k_\alpha\}_{\alpha=1}^n$, summed over ways of distributing them on an initial and final electron line:

$$\sum_{n_f=0}^{n} \quad = \bar{u}(p') \mathbf{i}\mathcal{M}_0 u(p) e^n \prod_{\alpha=1}^{n} \left(\frac{p'^{\mu_\alpha}}{p' \cdot k_\alpha} - \frac{p^{\mu_\alpha}}{p \cdot k_\alpha} \right) \epsilon^{\alpha\star}_{\mu_\alpha} \equiv \mathcal{A}_n.$$

Here the two terms in each factor are just as in (2.59), one term from initial and one from final-state emission; expanding the product gives the sum over $n_f = n - n_i$ (the number coming from the final-state line) on the RHS. From this expression, we can make a diagram with a soft-photon *loop* by picking an initial line α and a final line β, setting $k_\alpha = -k_\beta \equiv k$, erasing their polarization vectors, tying them together with a

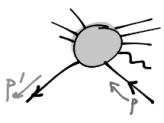
photon propagator, and summing over k:



$$= \mathcal{A}_{n-2}\frac{e^2}{2}\int \mathrm{d}\!\!^{-4}k \frac{-\mathbf{i}\eta_{\rho\sigma}}{k^2-m_\gamma^2}\left(\frac{p'}{p'\cdot k}-\frac{p}{p\cdot k}\right)^\rho\left(\frac{p'}{-p'\cdot k}-\frac{p}{-p\cdot k}\right)^\sigma \tag{2.61}$$

The factor of $\frac{1}{2}$ accounts for the symmetry under exchange of $\alpha \leftrightarrow \beta$. The last two factors are the approximate electron propagators attached to the new internal photon line. For the case of $n = 2$, this is the whole story, and this is

$$\bar{u}\mathbf{i}\mathcal{M}_0 u\cdot\mathbf{X} = \left(\quad\right)\cdot\left(\quad\right)_{\text{soft part}}$$

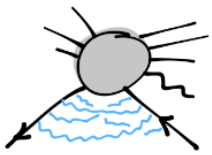


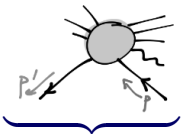
(where here 'soft part' means the part that is singular in m_γ) from which we conclude that

$$\mathbf{X} = -\frac{\alpha}{2\pi}f_{IR}(q^2)\ln\left(\frac{-q^2}{m_\gamma^2}\right) + \text{finite}.$$

The integral is the same as the one in (2.60) (and it is done in Peskin, page 201). Taking the most IR-divergent bit with m virtual soft photons (order α^m) for each m gives

$$\mathcal{M}_{\text{virtual soft}} = \sum_{m=0}^{\infty}\left(\quad\right) = \underbrace{\left(\quad\right)}_{\mathbf{i}\mathcal{M}_0}\underbrace{\sum_m \frac{1}{m!}\mathbf{X}^m}_{e^{\mathbf{X}}}$$





where the $1/m!$ is a symmetry factor from interchanging the virtual soft photons. Notice that this verifies my claim that the $-\infty$ in the one-loop answer is perturbation theory's way of trying to make the cross-section zero: since $\mathbf{X} \overset{m_\gamma\to 0}{\to} -\infty$, $d\sigma_{\text{exclusive}} \propto e^{2\mathbf{X}} \overset{m_\gamma\to 0}{\to} 0$.

Now consider the case of one real external soft ($E \in [m_\gamma, E_c]$) photon in the final state. The cross section is

$$\begin{aligned} d\sigma_{1\gamma} &= \int d\Pi \underbrace{\sum_{\text{pols}}\epsilon^\mu\epsilon^{\star\nu}}_{=-\eta^{\mu\nu}}\mathcal{M}_\mu\mathcal{M}_\nu^\star \\ &\overset{(2.59)}{=} \int d\Pi_0 |\bar{u}(p')\mathcal{M}_0 u(p)|^2 \int^{E_c}\frac{\mathrm{d}\!\!^3k}{2E_k}(-\eta_{\mu\nu})\,e^2\left(\frac{p'}{p'\cdot k}-\frac{p}{p\cdot k}\right)^\mu\left(\frac{p'}{-p'\cdot k}-\frac{p}{-p\cdot k}\right)^\nu \end{aligned}$$

$$
\begin{aligned}
&\equiv d\sigma_0 \mathbf{Y}, \\
\mathbf{Y} &= \frac{\alpha}{\pi} f_{IR}(q^2) \ln\left(\frac{E_c^2}{m_\gamma^2}\right).
\end{aligned}
$$

(This was actually exactly the same integral as in the virtual-photon calculation in (2.61).) Therefore, the exclusive cross section, including contributions of soft real photons gives

$$
\sum_{n=0}^{\infty} d\sigma_{n\gamma} = d\sigma_0 \sum_n \frac{1}{n!} \mathbf{Y}^n = d\sigma_0 e^{\mathbf{Y}}.
$$

Here the $n!$ is because the final state contains n identical bosons.

Putting the two effects together gives the promised cancellation of m_γ dependence to all orders in α:

$$
\begin{aligned}
d\sigma &= d\sigma_0 e^{2\mathbf{X}} e^{\mathbf{Y}} \\
&= d\sigma_0 \exp\left(-\frac{\alpha}{\pi} f_{IR}(q^2) \ln\frac{-q^2}{m_\gamma^2} + \frac{\alpha}{\pi} f_{IR}(q^2) \ln\frac{E_c^2}{m_\gamma^2}\right) \\
&= d\sigma_0 \exp\left(-\frac{\alpha}{\pi} f_{IR}(q^2) \ln\frac{-q^2}{E_c^2}\right) .
\end{aligned}
$$

This might seem pretty fancy, but unpacking the sum we did, the basic statement is that the probability of finding n photons with energy in a given (low-energy) range $[E_-, E_+]$ is

$$
P_{[E_-,E_+]} = \frac{1}{n!}\lambda^n e^{-\lambda}, \qquad \lambda = \frac{\alpha}{\pi} f_{IR}(q^2) \ln\frac{E_+}{E_-} = \langle n \rangle = \langle n^2 \rangle - \langle n \rangle^2
$$

a Poisson distribution. This is just what one finds in a coherent state of the radiation field.

2.6.3 Some magic from gauge invariance of QED

We found that the self-energy of the electron gave a wavefunction renormalization factor

$$
Z_e = 1 + \frac{\partial \Sigma}{\partial \not{p}}\Big|_{\not{p}=m_0} + \mathcal{O}(e^4) = 1 - \frac{\alpha}{4\pi} \ln\frac{\Lambda^2}{m^2} + \text{finite} + \mathcal{O}(\alpha^2).
$$

We care about this because there is a factor of Z_e in the LSZ formula for an S-matrix element with two external electrons. On the other hand, we found a cutoff-dependent correction to the vertex $e\gamma^\mu F_1(q^2)$ of the form

$$
F_1(q^2) = 1 + \frac{\alpha}{4\pi} \ln\frac{\Lambda^2}{m^2} + \text{finite} + \mathcal{O}(\alpha^2).
$$

Combining these together

$$S_{e\mu\leftarrow e\mu} = \left(\sqrt{Z_e}\right)^2 \left(\text{  } + \left(\text{  } \cdot \text{  } \right) + \cdots \right)$$
$$= \left(1 - \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{m^2} + \cdots\right) e^2 \bar{u}(p') \left(\gamma^\mu \left(1 + \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{m^2} + \cdots\right) + \frac{\alpha}{\pi} \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} \right) u(p)$$

the UV divergence from the vertex cancels the one in the self-energy. Why did this have to happen? During our discussion of the IR divergences, I mentioned a counterterm δ_e for the vertex. But how many counterterms do we get here? Is there a point of view that makes this cancellation obvious? Notice that the $\cdots$ multiplying the γ^μ term still contain the vacuum polarization diagram, which is our next subject, and which may be (is) cutoff dependent. Read on.

2.7 Vacuum polarization

[Zee, III.7] We've been writing the QED lagrangian as

$$\mathcal{L} = \bar{\psi} \left(\not{\partial} + \mathbf{i} e \not{\tilde{A}} - m \right) \psi - \frac{1}{4} \tilde{F}_{\mu\nu} \tilde{F}^{\mu\nu}.$$

I've put tildes on the photon field because of what's about to happen: Suppose we rescale the definition of the photon field $e\tilde{A}_\mu \equiv A_\mu, e\tilde{F}_{\mu\nu} \equiv F_{\mu\nu}$. Then the coupling e moves to the photon kinetic term:

$$\mathcal{L} = \bar{\psi} \left(\not{\partial} + \mathbf{i} \not{A} - m \right) \psi - \frac{1}{4e^2} F_{\mu\nu} F^{\mu\nu}.$$

With this normalization, instead of measuring the coupling between electrons and photons, the coupling constant e measures the difficulty a photon has propagating through space:

$$\langle A_\mu A_\nu \rangle \sim \frac{-\mathbf{i}\eta_{\mu\nu} e^2}{q^2}.$$

None of the physics is different, since each internal photon line still has two ends on a $\bar{\psi}\not{A}\psi$ vertex.

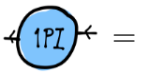
But from this point of view it is clear that the magic of the previous subsection is a consequence of gauge invariance, here's why: the demand of gauge invariance relates the coefficients of the $\bar{\psi}\not{\partial}\psi$ and $\bar{\psi}\not{A}\psi$ terms[28]. Therefore, any counterterm we need for the $\bar{\psi}\not{\partial}\psi$ term (which comes from the electron self-energy correction and is traditionally

[28] Notice that the gauge transformation of the rescaled A_μ is $A_\mu \to A_\mu + \partial_\mu \lambda(x), \psi(x) \to e^{\mathbf{i}q\lambda(x)}\psi(x)$ so that $D_\mu \psi \equiv (\partial + q\mathbf{i}A)_\mu \psi \to e^{\mathbf{i}q\lambda} D_\mu \psi$ where q is the charge of the field ($q = -1$ for the electron). This is to be contrasted with the transformation of $\tilde{A}_\mu \to \tilde{A}_\mu - \partial_\mu \lambda(x)/e$.

called δZ_2) must be the *same* as the counterterm for the $\bar{\psi}\not{A}\psi$ term (which comes from the vertex correction and is called δZ_1). No magic, just gauge invariance.

A further virtue of this reshuffling of the factors of e (emphasized by Zee on page 205) arises when we couple more than one species of charged particle to the electromagnetic field, *e.g.* electrons and muons or, more numerously, protons: once we recognize that charge renormalization is a property of the photon itself, it makes clear that quantum corrections cannot mess with the ratio of the charges. A deviation from -1 of the ratio of the charges of electron and proton as a result of interactions might seem plausible given what a mess the proton is, and would be a big deal for atoms. Gauge invariance forbids it. [End of Lecture 9]

An important comment about radiative corrections and symmetry. I've mentioned several times in lecture the perspective that we can add a counterterm for every term in the Lagrangian. Why don't we need counterterms of other operators? For example, in ϕ^4 theory, why don't we need a ϕ^3 counterterm? The validity of this perspective relies on a crucial fact: the terms that are generated in perturbation theory respect the symmetries of the theory. ϕ^4 theory (with no ϕ^3 term) has a $\mathbb{Z}_2$ symmetry under which $\phi \to -\phi$; this forbids a ϕ^3 term from being generated at any order in perturbation theory. An apparent counterexample to this perspective can happen that we forget to add some terms to our Lagrangian that respect the symmetries of the theory. For example, in the Yukawa theory I could forget the ϕ^4 term; this is generated by loops from the Yukawa interaction because it has the same symmetries. So what I really have in mind is that in our Lagrangian we always add all the terms that respect some set of symmetries. Except that we actually haven't been doing that because, for example, the term $\frac{1}{M^2}\phi\phi$ has the same symmetries as the ϕ^4 theory, but we don't include it. On the other hand, as you can see by dimensional analysis, it is an irrelevant perturbation of the ϕ^4 theory, so its effects will be small in the IR, for energies $E \ll M$. This is the reason we don't include it.

Just as we defined the electron self-energy (amputated 2-point function) as 

$= -\mathbf{i}\Sigma(\not{p})$ (with two spinor indices implied), we define the photon self-energy as

$$+\mathbf{i}\Pi_{\mu\nu}(q^2) \equiv$$ 

$$= + \mathcal{O}(e^4)$$

(the diagrams on the RHS are amputated). It is a function of q^2 by Lorentz symmetry. (The reason for the difference in sign is that the electron propagator is $\frac{+\mathbf{i}}{\not{p}-m}$ while the photon propagator is $\frac{-\mathbf{i}\eta_{\mu\nu}}{q^2}$.) Because of Lorentz symmetry, we can parametrize the

answer as

$$\Pi^{\mu\nu}(q^2) = A(q^2)\eta^{\mu\nu} + B(q^2)q^\mu q^\nu.$$

The Ward identity says

$$0 = q_\mu \Pi^{\mu\nu}(q^2) \quad \Longrightarrow \quad 0 = Aq^\nu + Bq^2 q^\nu \quad \Longrightarrow \quad B = -A/q^2.$$

Let $A \equiv \Pi q^2$ so that

$$\Pi^{\mu\nu}(q^2) = \Pi(q^2) q^2 \underbrace{\left(\eta^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right)}_{=\Delta_T^{\mu\nu}}.$$

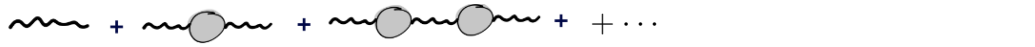
This object $\Delta_T^{\mu\nu}$ is a projector

$$\Delta_{T\rho}^{\mu} \Delta_{T\nu}^{\rho} = \Delta_{T\nu}^{\mu} \tag{2.62}$$

onto modes transverse to q^μ. Recall that we can take the bare propagator to be

 $= \dfrac{-\mathbf{i}\Delta_T}{q^2}$

without changing any gauge-invariant physics. This is useful because then

$$\begin{aligned}
\tilde{G}^{(2)}(q) \quad = \quad & \text{  } \\
\overset{(2.62)}{=} \quad & \frac{-\mathbf{i}\Delta_T}{q^2}\left(1 + \mathbf{i}\Pi q^2 \Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right) + \mathbf{i}\Pi q^2 \Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right)\mathbf{i}\Pi q^2 \Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right) + \cdots\right) \\
\overset{\Delta_T^2 = \Delta_T}{=} \quad & \frac{-\mathbf{i}\Delta_T}{q^2}\left(1 + \Pi\Delta_T + \Pi^2\Delta_T + \cdots\right) = \frac{-\mathbf{i}\Delta_T}{q^2 - q^2\Pi(q^2)}.
\end{aligned} \tag{2.63}$$

Does the photon get a mass? If $\Pi(q^2)$ is regular at $q^2 = 0$, then the photon remains massless. Only if the thing I called A above $q^2\Pi(q^2) \overset{q^2\to 0}{\to} A_0 \neq 0$ (that is, if $\Pi(q^2) \sim \frac{A_0}{q^2}$ or worse), then $\tilde{G} \overset{q^2\to 0}{\sim} \frac{1}{q^2 - A_0}$ does not have a pole at $q^2 = 0$. In order to get such a singularity in the photon self energy $\Pi(q^2) \sim \frac{A_0}{q^2}$ we need a process like $\delta\Pi \sim$  , where the intermediate state is a massless boson with propagator $\sim \frac{A_0}{q^2}$. As I will explain below, this is the Anderson-Higgs mechanism (not the easiest way to understand it).

The Ward identity played an important role here. Why does it work for the vacuum polarization?

$$q_\mu \Pi_2^{\mu\nu}(q^2) = q_\mu \text{  } \propto e^2 \int \mathrm{d}^4 p \,\mathrm{tr} \frac{1}{\not{p} + \not{q} - m} \not{q} \frac{1}{\not{p} - m} \gamma^\nu.$$

But here is an identity:

$$\frac{1}{\not{p} + \not{q} - m} \not{q} \frac{1}{\not{p} - m} = \frac{1}{\not{p} - m} - \frac{1}{\not{p} + \not{q} - m}. \tag{2.64}$$

Now, if we can redefine the integration variable $p \to p + q$ in the second term, the two terms cancel.

Why do I say 'if'? If the integral is UV divergent, this shift is not innocuous. So we have to address the cutoff dependence. In particular, if we just put a hard cutoff on the momentum, the integral would not be invariant under this change of variables. This is an explicit demonstration of the failure of gauge invariance of the hard cutoff.

In addition to the (lack of) mass renormalization, we've figured out the electromagnetic field strength renormalization. That is, the photon propagator near the pole at $q^2 = 0$ has the form $\tilde{G}^{(2)}(q) \simeq Z_\gamma \frac{-\mathbf{i}\Delta_T}{q^2}$ with

$$Z_\gamma \equiv Z_3 = \frac{1}{1 - \Pi(0)} \sim 1 + \Pi_2(0) + \mathcal{O}(e^4)$$

where Π_2 stands for the $\mathcal{O}(e^2)$ (one-loop) contribution. We need Z_γ for example for the S-matrix for processes with external photons, like Compton scattering.

Claim: If we do it right[29], the cutoff dependence looks like:

$$\Pi_2(q^2) = \frac{\alpha_0}{2\pi}\left(-\frac{2}{3}\ln\frac{\Lambda^2}{m_e^2} + \underbrace{2D(q^2)}_{\text{finite}}\right) \tag{2.65}$$

where Λ is the UV scale of ignorance. The photon propagator gets corrected to

$$\frac{e_0^2\Delta_T}{q^2} \rightsquigarrow \frac{Z_\gamma e_0^2 \Delta_T}{q^2},$$

and $Z_\gamma = \frac{1}{1-\Pi(0)}$ blows up logarithmically if we try to remove the cutoff. You see that the fine structure constant $\alpha_0 = \frac{e_0^2}{4\pi}$ has acquired the subscript of deprecation: we can make the photon propagator sensible while removing the cutoff if we are willing to recognize that the letter e_0 we've been carrying around is a fiction, and write everything in terms of $e \equiv \sqrt{Z_\gamma} e_0$ where $\frac{e^2}{4\pi} = \frac{1}{137}$ is the measured fine structure constant (at low energy). To this order, then, we write (ignoring the finite terms because we assume $\Lambda \gg$ everyone)

$$e_0^2 = e^2\left(1 + \frac{\alpha_0}{4\pi}\frac{2}{3}\ln\frac{\Lambda^2}{m_e^2}\right) + \mathcal{O}(\alpha^2). \tag{2.66}$$

[29] What I mean here is: if we do it in a way that respects the gauge invariance and hence the Ward identity. The simple PV regulator we've been using does not quite do that. However, an only slightly more involved implementation, explained in Zee (pages 202-204), does. (More details in §2.7.1.) Alternatively, we could use dimensional regularization everywhere.

$$m_0 = m + \mathcal{O}(\alpha_0) = m + \mathcal{O}(\alpha). \tag{2.67}$$

Since the difference between α_0 and α is higher order (in either), our book-keeping is unchanged. Inverting the relationship perturbatively, the renormalized charge is

$$e^2 = e_0^2 \left(1 - \frac{\alpha_0}{2\pi}\frac{2}{3}\ln\frac{\Lambda^2}{m_e^2} + \mathcal{O}(\alpha^2)\right)$$

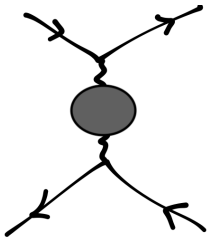
– in QED, the quantum fluctuations *reduce* the charge, as you might expect from the interpretation of this phenomenon as dielectric screening by virtual e^+e^- pairs.

In the example of $e\mu \leftarrow e\mu$ scattering, the full one-loop UV cutoff dependence then looks like

$$\begin{aligned} S_{e\mu\leftarrow e\mu} = &\left(1 - \frac{\alpha_0}{4\pi}\ln\Lambda^2 + \frac{\alpha_0}{2\pi}A(m_0)\right) e_0^2 \\ &L_\mu \bar{u}(p') \left[\gamma^\mu \left(1 + \frac{\alpha_0}{4\pi}\ln\Lambda^2 + \frac{\alpha_0}{2\pi}(B+D) + \frac{\alpha_0}{2\pi}\left(-\frac{2}{3}\ln\Lambda^2\right)\right) + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}\frac{\alpha_0}{2\pi}C(q^2, m_0)\right] u(p) \\ = &\, e^2 L_\mu \bar{u}(p') \left[\gamma^\mu \left(1 + \frac{\alpha}{2\pi}(A+B+D)\right) + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}\frac{\alpha}{2\pi}C\right] u(p) + \mathcal{O}(\alpha^2) \end{aligned} \tag{2.68}$$

where L_μ is the stuff from the muon line, and A, B, C, D are finite functions of m, q^2. In the second step, two things happened: (1) we cancelled the UV divergences from the Z-factor and from the vertex correction: this had to happen because there was no possible counterterm. (2) we used (2.66) and (2.67) to write everything in terms of the measured e, m. This removes the remaining cutoff dependence.

Claim: this works for all processes to order α^2. For example, Bhabha scattering gets a contribution of the form

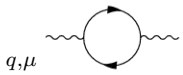


$$\propto e_0 \frac{1}{1 - \Pi(0)} e_0 = e^2.$$

In order to say what is $A+B+D$ we need to specify more carefully a renormalization scheme (other combinations of A, B, D can be changed by gauge transformations and field redefinitions). To do that, I need to give a bit more detail about the integral.

2.7.1 Under the hood

The vacuum-polarization contribution of a fermion of mass m and charge e at one loop is



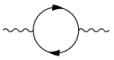
$$= -\int \mathrm{đ}^D k \mathrm{tr}\left((\mathbf{i}e\gamma^\mu) \frac{\mathbf{i}\,(\not{k}+m)}{k^2-m^2} (\mathbf{i}e\gamma^\nu) \frac{\mathbf{i}\,(\not{q}+\not{k}+m)}{(q+k)^2-m^2} \right)$$

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, q. As Zee says, our job in this course is not to train to be professional integrators. The result of this boiling can be written

$$\mathbf{i}\Pi_2^{\mu\nu}(q) = -e^2 \int \mathrm{đ}^D \ell \int_0^1 dx \frac{N^{\mu\nu}}{\left(\ell^2-\Delta\right)^2}$$

with $\ell = k + xq$ is a new integration variable, $\Delta \equiv m^2 - x(1-x)q^2$, and the numerator (for $D = 4$) is

$$N^{\mu\nu}/4 = 2\ell^\mu\ell^\nu - \eta^{\mu\nu}\ell^2 - 2x(1-x)q^\mu q^\nu + \eta^{\mu\nu}\left(m^2 + x(1-x)q^2\right) + \text{terms linear in } \ell^\mu \quad .$$

At this point I can illustrate explicitly why we can't use the euclidean momentum cutoff in gauge theory. With a euclidean momentum cutoff, the diagram  gives something of the form

$$\mathbf{i}\Pi_2^{\mu\nu} \propto e^2 \int^\Lambda d^4\ell_E \frac{\ell_E^2 \eta^{\mu\nu}}{\left(\ell_E^2+\Delta\right)^2} + \ldots \propto e^2\Lambda^2\eta^{\mu\nu}$$

This is *NOT* of the form $\Pi^{\mu\nu} = \Delta_T^{\mu\nu}\Pi(p^2)$ promised by the Ward identity; rather it produces a correction to the photon mass proportional to the cutoff. What happened? Our cutoff was not gauge invariant. Oops.[30]

Fancier PV regularization. [Zee page 202] We can fix the problem by adding also heavy Pauli-Villars electron ghosts. Suppose we add a bunch of them with masses

[30] Two points: How could we have predicted that the cutoff on euclidean momentum $\ell_E^2 < \Lambda^2$ would break gauge invariance? Its violation of the Ward identity here is a proof, but involved some work. The idea is that the momentum of a charged field shifts under a gauge transformation. Second: it *is* possible to construct a gauge invariant regulator with an explicit UV cutoff, using a lattice. The price, however, is that the gauge field enters only via the link variables $U(x,\hat{e}) = e^{\mathbf{i}\int_x^{x+\hat{e}} A}$ where x is a site in the lattice and $\hat{e}$ is the direction to a neighboring site in the lattice. For more, look up 'lattice gauge theory' in Zee's index. More on this later.

m_a and couplings $\sqrt{c_a}e$ to the photon. Then the vacuum polarization is that of the electron itself plus

$$-\sum_a c_a \int \mathrm{đ}^D k \mathrm{tr}\left((\mathbf{i}e\gamma^\mu)\frac{\mathbf{i}}{\not q+\not k-m_a}(\mathbf{i}e\gamma^\nu)\frac{\mathbf{i}}{\not q-m_a}\right) \sim \int^\Lambda \mathrm{đ}^4 k\left(\frac{\sum_a c_a}{k^2}+\frac{\sum_a c_a m_a^2}{k^4}+\cdots\right).$$

So, if we take $\sum_a c_a = -1$ we cancel the Λ^2 term, and if we take $\sum_a c_a m_a^2 = -m^2$, we also cancel the $\ln\Lambda$ term. This requires at least two PV electron fields, but so what? Once we do this, the momentum integral converges, and the Ward identity applies, so the answer will be of the promised form $\Pi^{\mu\nu} = q^2\Pi\Delta_T^{\mu\nu}$. After some more boiling[31], the answer, for $q^2 \ll m_a^2$ (the regime of interest for physics) is

$$\Pi_2(q^2) = -\frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\ln\frac{M^2}{m^2-x(1-x)q^2}$$

where $\ln M^2 \equiv \sum_a c_a \ln m_a^2$. This M plays the role of the UV scale of ignorance thenceforth. As promised, if we set q^2 here, we get (2.65), with a finite part $D(q^2)$ that you can read off.

Notice that this is perfectly consistent with our other two one-loop PV calculations: in those, the extra PV electrons never get a chance to run. At higher loops, we would have to make sure to be consistent.

Dimensional regularization. A regulator that is more automatically gauge invariant is dimensional regularization (dim reg). I have already been writing many of the integrals in D dimensions. One small difference when we are considering this as a regulator for an integral of fixed dimension is that we don't want to violate dimensional analysis, so we should really replace

$$\int d^4\ell \longrightarrow \int \frac{d^{4-\epsilon}\ell}{\bar\mu^{-\epsilon}}$$

where $D = 4-\epsilon$ and $\bar\mu$ is an arbitrary mass scale that will appear in the regulated answers, which we put here to preserve dim'l analysis – *i.e.* the couplings in dim reg will have the same engineering dimensions they had in the unregulated theory (dimensionless couplings remain dimensionless). $\bar\mu$ will parametrize our RG, *i.e.* play

[31] Some useful integrals for this purpose:

$$\int_0^\Lambda \frac{d\ell\ell^3}{(\ell^2+\Delta)^2} = \frac{1}{2}\left(\log\frac{\Lambda^2}{\Delta}-1\right)+\mathcal{O}\left(\frac{\Delta}{\Lambda^2}\right).$$

$$\int_0^\Lambda \frac{d\ell\ell^5}{(\ell^2+\Delta)^2} = \frac{1}{2}\Lambda^2+\frac{\Delta}{2}\left(1-2\log\frac{\Lambda^2}{\Delta}\right)+\mathcal{O}\left(\frac{\Delta}{\Lambda^2}\right).$$

the role of the RG scale. (It is often called μ at this step and then suddenly replaced by something also called μ; I will instead call this $\bar{\mu}$ and relate it to the thing that ends up being called μ.)

[Zinn-Justin 4th ed page 233] Dimensionally regularized integrals can be defined systematically with a few axioms indicating how the D-dimensional integrals behave under

1. translations $\int \text{đ}^D p f(p+q) = \int \text{đ}^D p f(p)$ [32]
2. scaling $\int \text{đ}^D p f(sp) = |s|^{-D} \int \text{đ}^D p f(p)$
3. factorization $\int \text{đ}^D p \int \text{đ}^D q f(p) g(q) = \int \text{đ}^D p f(p) \int \text{đ}^D q g(q)$

The (obvious?) third axiom implies the following formula for the sphere volume as a continuous function of D:

$$\left(\frac{\pi}{a}\right)^{D/2} = \int d^D x e^{-a\vec{x}^2} = \Omega_{D-1} \int_0^\infty x^{D-1} dx e^{-ax^2} = \frac{1}{2} a^{-\frac{D}{2}} \Gamma\left(\frac{D}{2}\right) \Omega_{D-1} \,. \tag{2.69}$$

This defines Ω_{D-1} for general D. [End of Lecture 10]

In dim reg, the one-loop vacuum polarization correction *does* satisfy the gauge-invariance Ward identity $\Pi^{\mu\nu} = \Delta_T^{\mu\nu} q^2 \Pi_2(q^2)$. A peek at the tables of dim reg integrals shows that Π_2 is:

$$\begin{aligned} \Pi_2(q^2) &\overset{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2-D/2)}{\Delta^{2-D/2}} \bar{\mu}^\epsilon \\ &\overset{D\to 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left(\frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right) + \mathcal{O}(\epsilon)\right) \end{aligned} \tag{2.70}$$

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, which appears in the Taylor expansion of the Euler gamma function; we define μ in this way so that, like Rosencrantz and Guildenstern in *Hamlet*, γ_E both appears and disappears from the discussion in this one scene.

In the second line of (2.70), we expanded the Γ-function about $D = 4$. Notice that what was a log divergence, becomes a $\frac{1}{\epsilon}$ pole in dim reg. There are other singularities of this function at other integer dimensions. It is an interesting question to ponder why

[32] Note that this rule fails for the euclidean momentum cutoff. Also note that this is the property we needed to demonstrate the Ward identity for the vertex correction using (2.64).

the integrals have such nice behavior as a function of D. That is: they only have simple poles. A partial answer is that in order to have worse (*e.g.* essential) singularities at some D, the perturbative field theory would have to somehow fail to make sense at larger D.

Now we are in a position to choose a renormalization condition (also known as a *renormalization scheme*), which will specify how much of the finite bit of Π gets subtracted by the $F_{\mu\nu}F^{\mu\nu}$ counterterm. One possibility is to demand that the photon propagator is not corrected at $q = 0$, *i.e.* demand $Z_\gamma = 1$. Then the resulting one-loop shift is

$$\delta\Pi_2(q^2) \equiv \Pi_2(q^2) - \Pi_2(0) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\log\left(\frac{m^2 - x(1-x)q^2}{m^2}\right).$$

We'll use this choice below.

Another popular choice, about which more later, is called the $\overline{\text{MS}}$ scheme, 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: MS here stands for 'minimal subtraction'; the bar in $\overline{\text{MS}}$ refers to the (not so important) distinction between $\bar{\mu}$ and μ.) The resulting vacuum polarization function is

$$\delta\Pi_2^{(\overline{\text{MS}})}(q^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\log\left(\frac{m^2 - x(1-x)q^2}{\mu^2}\right).$$

2.7.2 Physics from vacuum polarization

We began this discussion of QED loop effects by promising to compute quantum corrections to the Coulomb force. One class of physical effects of vacuum polarization arise from attaching the corrected photon propagator to a static delta-function charge source. The resulting effective Coulomb potential is the fourier transform of

$$\tilde{V}(q) = \frac{1}{q^2}\frac{e^2}{1-\Pi(q^2)} \equiv \frac{e^2_{\text{eff}}(q)}{q^2}. \tag{2.71}$$

This has consequences in both IR and UV.

IR: In the IR ($q^2 \ll m^2$), it affects the spectra of atoms. The leading correction is

$$\delta\Pi_2(q) = \frac{e^2}{2\pi^2}\int dx x(1-x)\ln\left(1-\frac{q^2}{m^2}x(1-x)\right) \overset{q\ll m}{\simeq} \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\left(-\frac{q^2}{m^2}x(1-x)\right) = -\frac{q^2}{60\pi^2 m^2}$$

which means

$$\tilde{V}(q) \overset{q\ll m}{\simeq} \frac{e^2}{q^2} + \frac{e^2}{q^2}\left(-\frac{q^2}{60\pi^2 m^2}\right) + \cdots$$

and hence

$$V(r) \simeq -\frac{e^2}{4\pi r} - \frac{e^4}{60\pi^2 m^2}\delta(r) + \cdots \equiv V + \Delta V.$$

This shifts the energy levels of hydrogen s-orbitals (the ones with support at the origin) by $\Delta E_s = \langle s|\Delta V|s\rangle$ which contributes to lowering the $2S$ state relative to the $2P$ state. This is the Lamb shift, whose observation played an important role in spurring people to study radiative corrections in QED.

This delta function is actually a long-wavelength approximation to what is called the Uehling potential; its actual range is $1/m_e$, which is the scale on which Π_2 varies . The delta function approximation is a good idea for atomic physics, since $\frac{1}{m_e} \ll a_0 = \frac{1}{\alpha m_e}$, the Bohr radius. See Schwartz p. 311 for a bit more on this.

UV: In the UV limit ($q^2 \gg m^2$), we can approximate $\ln\left(1-\frac{q^2}{m^2}x(1-x)\right) \simeq \ln\left(-\frac{q^2}{m^2}x(1-x)\right) \simeq \ln\left(-\frac{q^2}{m^2}\right)$ to get[33]

$$\Pi_2(q^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\ln\left(1-\frac{q^2}{m^2}x(1-x)\right) \simeq \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\ln\left(-\frac{q^2}{m^2}\right) = \frac{e^2}{12\pi^2}\ln\left(-\frac{q^2}{m^2}\right).$$

Therefore, the effective charge in (2.71) at high momentum photon exchange is

$$e^2_{\text{eff}}(q^2) \overset{q^2\gg m_e^2}{\simeq} \frac{e^2}{1-\frac{e^2}{12\pi^2}\ln\left(-\frac{q^2}{m^2}\right)}. \tag{2.72}$$

(Remember that $q^2 < 0$ for t-channel exchange, as in the static potential, so the argument of the log is positive and this is real.)

Two things: if we make q^2 big enough, we can make the loop correction as big as the 1. This requires $|q| \sim 10^{286}$ eV. Good luck with that. This is called a *Landau pole*. It is a strange (and, I think, sad) fact of history that the existence of this ridiculously high energy scale where the QED coupling blows up led Landau (and hence his school) to give up on QFT. If even Landau's view of the world was obstructed by such weird prejudices, what wonderful things are we missing?

[33]The last step is safe since the $x(1-x)$ suppresses the contributions of the endpoints of the x integral, so we can treat $x(1-x)$ as finite.

The second thing is: this perspective of a scale-dependent coupling is very valuable, and is a crucial ingredient in the renormalization group. The value $\alpha = \frac{1}{137}$ is the extreme IR value, for $q \ll m_e$.

3 Consequences of unitarity

Next I would like to fulfill my promise to show that conservation of probability guarantees that some things are positive (for example, Z and $1-Z$, where Z is any wavefunction renormalization factor). We will show that amplitudes develop an imaginary part when the virtual particles become real. (Someone should have put an extra factor of $\mathbf{i}$ in the definition to resolve this infelicity.) We will discuss the notion of *density of states* in QFT (this should be a positive number!), and in particular the notion of the *density of states contributing to a correlation function* $G = \langle \mathcal{O}\mathcal{O} \rangle$, also known as the *spectral density* of G (or of the operator $\mathcal{O}$). In high-energy physics this idea is associated with the names Källen-Lehmann and is part of a program of trying to use complex analysis to make progress in QFT. These quantities are also ubiquitous in the theory of condensed matter physics and participate in various *sum rules*. This discussion will be a break from perturbation theory; we will say things that are true with a capital 't'.

3.1 Spectral density

[Zee III.8, Appendix 2; Peskin §7.1; Xi Yin's notes for Harvard Physics 253b] In the following we will consider a (time-ordered) two-point function of an operator $\mathcal{O}$. We will make hardly any assumptions about this operator. We will assume it is a scalar under rotations, and will assume translation invariance in time and space. But we need not assume that $\mathcal{O}$ is 'elementary'. This is an extremely loaded term, a useful definition for which is: a field governed by a nearly-quadratic action. Also: try to keep an eye out for where (if anywhere) we assume Lorentz invariance.

So, let

$$-\mathbf{i}\mathcal{D}(x) \equiv \langle 0 | \mathcal{T}\mathcal{O}(x)\mathcal{O}^\dagger(0) | 0 \rangle .$$

Notice that we do not assume that $\mathcal{O}$ is hermitian. Here $|0\rangle$ denotes the true vacuum, which I called $|\Omega\rangle$ before, not our shabby perturbative approximation to it. Use translation invariance to move the left operator to the origin: $\mathcal{O}(x) = e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)e^{-\mathbf{i}\mathbf{P}x}$. This follows from the statement that $\mathbf{P}$ generates translations [34]

$$\partial_\mu \mathcal{O}(x) = \mathbf{i}[\mathbf{P}_\mu, \mathcal{O}(x)] .$$

[34] Note that $\mathbf{P}$ here is a D-component vector of operators

$$\mathbf{P}_\mu = (\mathbf{H}, \vec{\mathbf{P}})_\mu$$

which includes the Hamiltonian – we are using relativistic notation – but we haven't actually required any assumption about the action of boosts.

And let's unpack the time-ordering symbol:

$$-\mathbf{i}\mathcal{D}(x) = \theta(t)\,\langle 0|\, e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)e^{-\mathbf{i}\mathbf{P}x}\mathcal{O}^\dagger(0)\,|0\rangle + \theta(-t)\,\langle 0|\,\mathcal{O}^\dagger(0)e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)e^{-\mathbf{i}\mathbf{P}x}\,|0\rangle\,. \quad (3.1)$$

Now we need a resolution of the identity operator on the *entire* QFT $\mathcal{H}$:

$$\mathbb{1} = \sum_n |n\rangle\,\langle n|\,. \quad (3.2)$$

This innocent-looking summation variable n is hiding an enormous sum! Let's also assume that the groundstate $|0\rangle$ is translation invariant:

$$\mathbf{P}\,|0\rangle = 0.$$

We can label each state $|n\rangle$ by its total momentum (since the components of $\mathbf{P}^\mu$ commute with each other):

$$\mathbf{P}^\mu\,|n\rangle = p_n^\mu\,|n\rangle\,.$$

Let's examine the first term in (3.1); sticking the $\mathbb{1}$ in the form (3.2) in a suitable place:

$$\langle 0|\, e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)\mathbb{1}e^{-\mathbf{i}\mathbf{P}x}\mathcal{O}^\dagger(0)\,|0\rangle = \sum_n \langle 0|\,\mathcal{O}(0)\,|n\rangle\,\langle n|\,e^{-\mathbf{i}\mathbf{P}x}\mathcal{O}^\dagger(0)\,|0\rangle = \sum_n e^{-\mathbf{i}p_n x}\|\,\mathcal{O}_{0n}\,\|^2\,,$$

with $\mathcal{O}_{0n} \equiv \langle 0|\,\mathcal{O}(0)\,|n\rangle$ the matrix element of our operator between the vacuum and the state $|n\rangle$. Notice the absolute value: unitarity of our QFT requires this to be positive and this will have valuable consequences.

Next we work on the time-ordering symbol. I claim that :

$$\theta(x^0) = \theta(t) = -\mathbf{i}\int \text{đ}\omega \frac{e^{+\mathbf{i}\omega t}}{\omega - \mathbf{i}\epsilon}\quad ; \qquad \theta(-t) = +\mathbf{i}\int \text{đ}\omega \frac{e^{+\mathbf{i}\omega t}}{\omega + \mathbf{i}\epsilon}\,.$$

Just like in the discussion of the Feynman contour, the point of the $\mathbf{i}\epsilon$ is to push the pole inside or outside the integration contour. The half-plane in which we must close the contour depends on the sign of t. There is an important sign related to the orientation with which we circumnavigate the pole. Here is a check that we got the signs and factors right:

$$\frac{d\theta(t)}{dt} = -\mathbf{i}\partial_t \int \text{đ}\omega \frac{e^{\mathbf{i}\omega t}}{\omega - \mathbf{i}\epsilon} = \int \text{đ}\omega e^{\mathbf{i}\omega t} = \delta(t).$$

Consider now the fourier transform of $\mathcal{D}(x)$ (for simplicity, I've assumed $\mathcal{O} = \mathcal{O}^\dagger$ here, so that $\|\,\mathcal{O}_{n0}\,\| = \|\,\mathcal{O}_{0n}\,\|$):

$$-\mathbf{i}\mathcal{D}(q) \equiv \int d^D x e^{\mathbf{i}qx}\,(-\mathbf{i}\mathcal{D}(x)) = -\mathbf{i}(2\pi)^{D-1}\sum_n \|\,\mathcal{O}_{0n}\,\|^2\left(\frac{\delta^{(D-1)}(\vec{q}-\vec{p}_n)}{q^0 - p_n^0 + \mathbf{i}\epsilon} - \frac{\delta^{(D-1)}(\vec{q}+\vec{p}_n)}{q^0 + p_n^0 - \mathbf{i}\epsilon}\right)\,. \quad (3.3)$$

With this expression in hand, you could imagine measuring the $\mathcal{O}_{0n}$s and using that to determine $\mathcal{D}$.

Now suppose that our operator $\mathcal{O}$ is capable of creating a single particle (for example, suppose, if you must, that $\mathcal{O} = \phi$, a perturbative quantum field). Such a state is labelled only by its spatial momentum (assuming for simplicity that $\mathcal{O}$ doesn't carry any other conserved quantities): $\left|\vec{k}\right\rangle$ (here I briefly retreat to non-relativistic normalization of states $\left\langle \vec{k}|\vec{k}'\right\rangle = \delta^{D-1}(\vec{k}-\vec{k}')$). The statement that $\mathcal{O}$ can create this state from the vacuum means

$$\left\langle \vec{k}\right| \mathcal{O}(0)\left|0\right\rangle = \frac{Z^{\frac{1}{2}}}{\sqrt{(2\pi)^{D-1}\,2\omega_{\vec{k}}}} \tag{3.4}$$

where $Z \neq 0$ and $\omega_{\vec{k}}$ is the energy of the particle as a function of $\vec{k}$. For a Lorentz invariant theory, we can parametrize this as

$$\omega_{\vec{k}} \overset{\text{Lorentz!}}{\equiv} \sqrt{\vec{k}^2 + m^2}$$

in terms of m, the *mass* of the particle. [35] What is Z? From (3.4) and the axioms of QM, you can see that it's the probability that $\mathcal{O}$ creates this 1-particle state from the vacuum. In the free field theory it's 1, and it's positive because it's a probability. $1-Z$ measures the extent to which $\mathcal{O}$ does anything *besides* create this 1-particle state.

The identity of the one-particle Hilbert space (relatively tiny!) $\mathcal{H}_1$ is

$$\mathbb{1}_1 = \int \mathrm{đ}^{D-1}\vec{k}\left|\vec{k}\right\rangle\left\langle\vec{k}\right|, \quad \left\langle\vec{k}|\vec{k}'\right\rangle = \delta^{(D-1)}(\vec{k}-\vec{k}').$$

This is a summand in the whole horrible resolution:

$$\mathbb{1} = |0\rangle\langle 0| + \mathbb{1}_1 + \cdots .$$

[35] It's been a little while since we spoke explicitly about free fields, so let's remind ourselves about the appearance of $\omega^{-\frac{1}{2}}$ in (3.4), recall the expansion of a free scalar field in creation an annihilation operators:

$$\phi(x) = \int \frac{\mathrm{đ}^{D-1}\vec{p}}{\sqrt{2\omega_{\vec{p}}}}\left(\mathbf{a}_{\vec{p}}e^{-\mathbf{i}px} + \mathbf{a}^\dagger_{\vec{p}}e^{\mathbf{i}px}\right) \quad .$$

For a free field $\left|\vec{k}\right\rangle = \mathbf{a}^\dagger_{\vec{k}}\left|0\right\rangle$, and $\left\langle\vec{k}\right|\phi(0)\left|0\right\rangle = \frac{1}{\sqrt{(2\pi)^{D-1}2\omega_{\vec{k}}}}$. The factor of $\omega^{-\frac{1}{2}}$ is required by the ETCRs:

$$[\phi(\vec{x}),\pi(\vec{x}')] = \mathbf{i}\delta^{D-1}(\vec{x}-\vec{x}'), \quad [\mathbf{a}_{\vec{k}},\mathbf{a}^\dagger_{\vec{k}'}] = \delta^{D-1}(\vec{k}-\vec{k}') \ ,$$

where $\pi = \partial_t\phi$ is the canonical field momentum. It is just like in the simple harmonic oscillator, where

$$\mathbf{q} = \sqrt{\frac{\hbar}{2m\omega}}\left(\mathbf{a}+\mathbf{a}^\dagger\right), \quad \mathbf{p} = \mathbf{i}\sqrt{\frac{\hbar\omega}{2}}\left(\mathbf{a}-\mathbf{a}^\dagger\right).$$

I mention this because it lets us define the part of the horrible $\sum_n$ in (3.3) that comes from 1-particle states:

$$\begin{aligned}\implies \quad \mathbf{i}\mathcal{D}(q) &= \ldots\mathbf{i}(2\pi)^{D-1}\int \mathrm{d}^{D-1}\vec{k}\frac{Z}{(2\pi)^{D-1}2\omega_k}\left(\frac{\delta^{D-1}(\vec{q}-\vec{k})}{q^0-\omega_{\vec{k}}+\mathbf{i}\epsilon}-(\omega_k\to-\omega_k)\right)\\ &= \ldots\mathbf{i}\frac{Z}{2\omega_q}\left(\frac{1}{q^0-\omega_q+\mathbf{i}\epsilon}-\frac{1}{q^0+\omega_q+\mathbf{i}\epsilon}\right)\\ &\overset{\text{Lorentz}}{=} \ldots\mathbf{i}\frac{Z}{q^2-m^2+\mathbf{i}\epsilon}\end{aligned}$$

(Here again ... is contributions from states involving something else, *e.g.* more than one particle.) The big conclusion here is that even in the interacting theory, even if $\mathcal{O}$ is composite and complicated, if $\mathcal{O}$ can create a 1-particle state with mass m with probability Z, then its 2-point function has a pole at the right mass, and the residue of that pole is Z. [36]

The imaginary part of $\mathcal{D}$ is called the *spectral density* $\boldsymbol{\rho}$ (beware that different physicists have different conventions for the factor of $\mathbf{i}$ in front of the Green's function; the spectral density is not always the imaginary part, but it's always positive (in unitary theories)!

Using

$$\mathrm{Im}\,\frac{1}{Q\mp\mathbf{i}\epsilon}=\pm\pi\delta(Q),\qquad\text{(for } Q \text{ real)}. \tag{3.5}$$

in (3.3) we have (for real operators), $\mathrm{Im}\,\mathcal{D}(q) =$

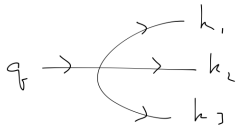
$$\mathrm{Im}\,\mathbf{i}\int d^Dx\, e^{\mathbf{i}qx}\left\langle 0\right|\mathcal{T}\mathcal{O}(x)\mathcal{O}(0)\left|0\right\rangle=\pi\,(2\pi)^{D-1}\sum_n\|\,\mathcal{O}_{0n}\,\|^2\left(\delta^D(q-p_n)+\underbrace{\delta^D(q+p_n)}_{=0\text{ for }q^0>0\text{ since }p_n^0>0}\right). \tag{3.6}$$

The second term on the RHS vanishes when $q^0 > 0$, since states in $\mathcal{H}$ have energy bigger than the energy of the groundstate. Therefore, (with Lorentz symmetry) the contribution of a 1-particle state to the spectral density is:

$$\mathrm{Im}\,\mathcal{D}(q^2)=\ldots+\pi Z\delta(q^2-m^2).$$

This quantity $\mathrm{Im}\,\mathcal{D}(q)$ (the spectral density of $\mathcal{O}$) is positive because it is the *number* of states (with D-momentum in an infinitesimal neighborhood of q), weighted by the modulus of their overlap with the state engendered by the operator on the groundstate.

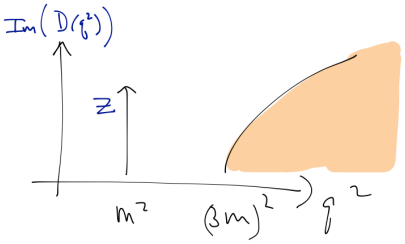
[36] If we hadn't assumed Lorentz invariance, this would be replaced by the statement: if the operator $\mathcal{O}$ can create a state with energy ω from the vacuum with probability Z, then its Green's function has a pole at that frequency, with residue Z.

Now what about multiparticle states? The associated sum over such states involves multiple (spatial) momentum integrals, not fixed by the total momentum. *e.g.* in ϕ^4 theory, ϕ can make a 3-particle state:  and the three particles must share the momentum q. In this case the sum over all 3-particle states is

$$\sum_{n,\text{ 3-particle states with momentum } q} \propto \int d\vec{k}_1 d\vec{k}_2 d\vec{k}_3 \delta^D(k_1+k_2+k_3-q)$$

(Note that I am not saying that a single *real* ϕ particle is decaying to three real ϕ particles; that can't happen if they are massive. Rather, in the diagram  you should think of the particle with momentum q as virtual.)

Now instead of an isolated pole, we have a whole collection of poles right next to each other. This is a branch cut. In this example, the branch cut begins at $q^2=(3m)^2$. $3m$ is the lowest energy q^0 at which we can produce three particles of mass m (in which case they have to be at rest).



Note that in ϕ^3 theory, we would instead find that the particle can decay into two particles, and the sum over two particle states would look like

$$\sum_{n,\text{ 2-particle states with momentum } q} \propto \int d\vec{k}_1 d\vec{k}_2 \delta^D(k_1+k_2-q)$$

so the continuum would start at $q^2=(2m)^2$.

Recall that for real x the imaginary part of a function of one variable with a branch cut, (like $\text{Im}(x+\mathbf{i}\epsilon)^\nu = \frac{1}{2}\left((x+\mathbf{i}\epsilon)^\nu - (x-\mathbf{i}\epsilon)^\nu\right)$) is equal to (half) the *discontinuity* of the function $((x)^\nu)$ across the branch cut. The discontinuity goes to zero as we approach the branch point. Near the multi-particle continuum, the Green's function has such a branch cut.

Now we recall some complex analysis, in the form of the *Kramers-Kronig (or dispersion) relations*:

$$\text{Re}G(z) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega \frac{\text{Im}G(\omega)}{\omega - z}$$

(valid if $\text{Im}G(\omega)$ is analytic in the UHP of ω and falls off faster than $1/\omega$). These equations, which I think we were supposed to learn in E&M but no one seems to, and

which relate the real and imaginary parts of an analytic function by an integral equation, can be interpreted as the statement that the imaginary part of a complex integral comes from the singularities of the integrand, and conversely that those singularities completely determine the function.

An even more dramatic version of these relations (whose imaginary part is the previous eqn) is

$$f(z) = \frac{1}{\pi}\int dw \frac{\rho(w)}{w-z} \ , \qquad \rho(w) \equiv \mathrm{Im}\, f(w + \mathbf{i}\epsilon).$$

The imaginary part determines the whole function.

Comments:

- The spectral density $\mathrm{Im}\mathcal{D}(q)$ *determines* $\mathcal{D}(q)$. When people get excited about this it is called the "S-matrix program" or something like that.

- The result we've shown protects physics from our caprices in choosing field variables. If someone else uses a different field variable $\eta \equiv Z^{\frac{1}{2}}\phi + \alpha\phi^3$, the result above with $\mathcal{O} = \eta$ shows that

  $$\int d^D x e^{\mathbf{i}qx} \langle \mathcal{T}\eta(x)\eta(0)\rangle$$

  still has a pole at $q^2 = m^2$ and a cut starting at the three-particle threshold, $q^2 = (3m)^2$. [End of Lecture 11]

- A sometimes-useful fact that we've basically already shown (for real operators):

  $$\mathrm{Im}\,\mathcal{D}(q) = \frac{1}{2}(2\pi)^D \sum_n \| \mathcal{O}_{0n} \|^2 \left(\delta^D(q-p_n) + \delta^D(q+p_n)\right) \overset{q^0\geq 0}{=} \frac{1}{2}\int d^D x e^{\mathbf{i}qx} \langle 0| \left[\mathcal{O}(x), \mathcal{O}(0)\right] |0\rangle \ .$$

  This shows that the retarded Green's function is also determined from the spectral density.

We can summarize what we've learned in the Lorentz-invariant case as follows: In a Lorentz invariant theory, the spectral density ρ for a scalar operator ϕ is a scalar function of p^μ with

$$\sum_s \delta^D(p-p_s) \| \langle 0|\, \phi(0)\, |s\rangle \, \|^2 = \frac{\theta(p^0)}{(2\pi)^{D-1}}\rho(p^2) \ .$$

Claims:

- $\rho(s) = \mathcal{N}\mathrm{Im}\,\mathcal{D}$ for some number $\mathcal{N}$ (I believe $\mathcal{N} = \pi$ here), when $s > 0$.
- $\rho(s) = 0$ for $s < 0$. There are no states for spacelike momenta.
- $\rho(s) \geq 0$ for $s \geq 0$. The density of states for timelike momenta is positive or zero.
- With our assumption about one-particle states, $\rho(s)$ has a delta-function singularity at $s = m^2$, with weight Z. More generally we have shown that

$$\mathcal{D}(k^2) = \int ds\ \rho(s) \frac{1}{k^2 - s + \mathbf{i}\epsilon}. \tag{3.7}$$

This is called the Källen-Lehmann spectral representation of the propagator; it represents it as a sum of *free* propagators with different masses, determined by the spectral density.

In particular, this result (3.7) implies that $\mathcal{D}(z = k^2)$ is an analytic function in the complex z-plane away from the support of ρ, *i.e.* away from the momenta where physical states live. Singularities of amplitudes come only from physics.

One consequence (assuming unitarity and Lorentz symmetry) is that at large $|k^2|$, the Green's function is bigger than $\frac{1}{k^2}$, since each term in the integral goes like $\frac{1}{k^2}$ and $\rho(s) \geq 0$ means that there cannot be cancellations between each $\frac{1}{k^2 - s}$ contribution. This means that if the kinetic term for your scalar field has more derivatives, something must break at short distances. Breaking Lorentz symmetry is the easiest way out, for example on a lattice; in a Lorentz-invariant theory, this is an indication that non-renormalizable terms imply more degrees of freedom at high energy. (More on this in subsection §3.2.) For example, consider the theory with Lagrangian $L = (\partial\phi)^2 + \frac{1}{\Lambda^2}(\partial^2\phi)^2$. It's quadratic in ϕ so we can solve it, and the (Källen-Lehman representation of the) propagator is

$$\frac{1}{k^2 + k^4/\Lambda^2} = \frac{1}{k^2} - \frac{1}{k^2 - \Lambda^2}$$

which as you can see looks just like a Pauli-Villars regulator. That is, we've added in a ghost field whose pole has a negative residue. As we've seen above, the residue of the pole in the propagator is a *probability*, and hence in a unitary theory had better be positive.

Taking into account our assumption about single-particle states, (3.7) is

$$\mathcal{D}(k^2) = \frac{Z}{k^2 - m^2 + \mathbf{i}\epsilon} + \int_{(3m)^2}^{\infty} ds\ \rho_c(s) \frac{1}{k^2 - s + \mathbf{i}\epsilon}$$

where ρ_c is just the continuum part. The pole at the particle-mass^2 survives interactions, with our assumption. (The value of the mass need not be the same

as the bare mass!) Actually in this expression I've left out the possibility of boundstates, which arise as poles at energy less than the 2-particle (or 3-particle) threshold.

- **Sum rule.** Finally, suppose that the field ϕ in question is a canonical (but not necessarily free) field, in the sense that

$$[\phi(x,t), \partial_t \phi(y,t)] = \mathbf{i}\delta^{(d)}(x-y).$$

This is a statement both about the normalization of the field, and that its canonical momentum is its time derivative. Then[37]

$$1 = \int_0^\infty ds \rho(s). \tag{3.9}$$

If we further assume that ϕ can create a one-particle state with mass m, so that $\rho(s) = Z\delta(s-m^2) + \rho_c(s)$ where $\rho_c(s) \geq 0$ is the contribution from the continuum of $\geq$ 2-particle states, then

$$1 = Z + \int_{\text{threshold}}^\infty ds \rho_c(s)$$

is a sum rule. It shows that $Z \in [0,1]$ and is just the statement that if the field doesn't create a single particle, it must do something else. The LHS is the probability that *something* happens.

The idea of spectral representation and spectral density is more general than the Lorentz-invariant case. In particular, the spectral density of a Green's function is an important concept in the study of condensed matter. For example, the spectral density for the electron 2-point function is the thing that actually gets measured in angle-resolved photoemission experiments (ARPES).

[37] Here's how to see this. For free fields (chapter 3) we have

$$\langle 0|[\phi(x), \phi(y)]|0\rangle_{\text{free}} = \Delta_+(x-y, m^2) - \Delta_+(y-x, m^2),$$

where $\Delta_+(x) = \int \frac{\mathrm{d}^d p}{2\omega_{\vec{p}}} e^{-\mathbf{i}p\cdot x}|_{p^0=\omega_{\vec{p}}}$. For an interacting canonical field, we have instead a spectral representation (by exactly the methods above):

$$\langle \Omega|[\phi(x), \phi(y)]|\Omega\rangle = \int d\mu^2 \rho(\mu^2) \left(\Delta_+(x-y, \mu^2) - \Delta_+(y-x, \mu^2)\right), \tag{3.8}$$

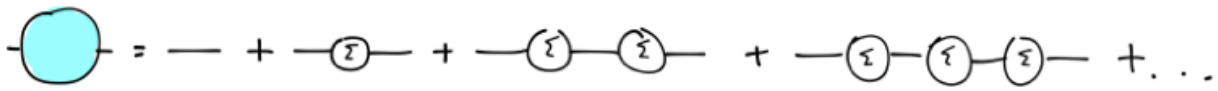
where ρ is the same spectral density as above. Now take $\partial_{x^0}|_{x^0=y^0}$ of the BHS of (3.8) and use $\partial_t \Delta_+(x-y;\mu^2)|_{x^0=y^0} = -\frac{\mathbf{i}}{2}\delta^{(d)}(\vec{x}-\vec{y})$.

3.2 Cutting rules and optical theorem

[Zee §III.8] So, that may have seemed a bit formal. What does this mean when we have in our hands a perturbative QFT? Consider the two point function of heavy relativistic scalar field Φ (of mass M) that has a perturbative cubic coupling to some lighter field ϕ (of mass m):

$$S = \int d^D x \left(\frac{1}{2} \left((\partial\phi)^2 - m^2\phi^2 + (\partial\Phi)^2 - M^2\Phi^2 \right) - \frac{g}{2!}\phi^2\Phi \right).$$

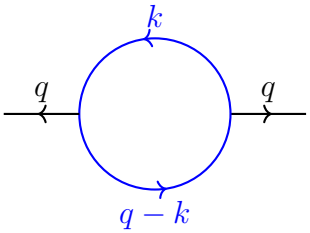
Sum the geometric series of 1PI insertions to get



$$\mathbf{i}\mathcal{D}_\phi(q) = \frac{\mathbf{i}}{q^2 - m^2 - \Sigma(q) + \mathbf{i}\epsilon}$$

where $\Sigma(q)$ is the 1PI two point vertex.

The leading contribution to Σ comes from the one loop diagram at right (with the light field (in blue) running in the loop) and is



$$\mathbf{i}\Sigma_{1\text{ loop}}(q^2) = \frac{1}{2}(\mathbf{i}g)^2 \int \mathrm{đ}^D k \frac{\mathbf{i}}{k^2 - m^2 + \mathbf{i}\epsilon} \frac{\mathbf{i}}{(q-k)^2 - m^2 + \mathbf{i}\epsilon}.$$

The $\frac{1}{2}$ is a symmetry factor from exchanging the two internal lines of the loop. Consider this function for real q, for which there are actual states of the scalar field – timelike q^μ, with $q^0 > m$. The real part of Σ shifts the mass. But what does it mean if this function has an imaginary part?

Claim: $\mathrm{Im}\,\Sigma/M$ is a decay rate.

It moves the energy of the particle off of the real axis from M (in its rest frame) to

$$\sqrt{M^2 + \mathbf{i}\mathrm{Im}\,\Sigma(M^2)} \overset{\text{small } \mathrm{Im}\Sigma \sim g^2}{\simeq} M + \mathbf{i}\frac{\mathrm{Im}\,\Sigma(M^2)}{2M}.$$

The fourier transform to real time is an amplitude for propagation in time of a state with complex energy $\mathcal{E}$: its wavefunction evolves like $\psi(t) \sim e^{-\mathbf{i}\mathcal{E}t}$ and has norm

$$\| \psi(t) \|^2 \sim \| e^{-\mathbf{i}\left(E - \mathbf{i}\frac{1}{2}\Gamma\right)t} \|^2 = e^{-\Gamma t}.$$

In our case, we have $\Gamma \sim \mathrm{Im}\,\Sigma(m^2)/m$ (I'll be more precise below), and we interpret that as the rate of decay of the norm of the single-particle state. There is a nonzero

probability that the state turns into something else as a result of time evolution in the QFT: the single particle must decay into some other state – generally, multiple particles. (We will see next how to figure out into what it decays.)

The absolute value of the Fourier transform of this quantity $\psi(t)$ is the kind of thing you would measure in a scattering experiment. This is

$$F(\omega) = \int dt\ e^{-\mathbf{i}\omega t}\psi(t) = \int_0^\infty dt\ e^{-\mathbf{i}\omega t} e^{\mathbf{i}\left(M - \frac{1}{2}\mathbf{i}\Gamma\right)t} = \frac{1}{\mathbf{i}\,(\omega - M) - \frac{1}{2}\Gamma}$$

$$\| F(\omega) \|^2 = \frac{1}{(\omega - M)^2 + \frac{1}{4}\Gamma^2}$$

is a Lorentzian in ω with width Γ. So Γ is sometimes called a *width*.

So: what is $\mathrm{Im}\,\Sigma_{1\ \mathrm{loop}}$ in this example?

On the intermediate propagators, we will use the identity

$$\frac{1}{k^2 - m^2 + \mathbf{i}\epsilon} = \mathcal{P}\frac{1}{k^2 - m^2} - \mathbf{i}\pi\delta(k^2 - m^2) \equiv \mathcal{P} - \mathbf{i}\Delta$$

where $\mathcal{P}$ denotes 'principal part'. Then

$$\mathrm{Im}\,\Sigma_{1\ \mathrm{loop}}(q) = -\frac{1}{2}g^2 \int d\Phi\ (\mathcal{P}_1\mathcal{P}_2 - \Delta_1\Delta_2)$$

with $d\Phi = \text{đ}^D k_1 \text{đ}^D k_2 (2\pi)^D \delta^D(k_1 + k_2 - q)$.

This next trick, to get rid of the principal part bit, is from Zee's book (the second edition on p.214; he also does the calculation by brute force in the appendix to that section). We can find a representation for the 1-loop self-energy in terms of real-space propagators: it's the Fourier transform of the amplitude to create two ϕ excitations at the origin at time zero with a single ϕ field (this is $-ig$), to propagate them both from 0 to x (this is $(\mathbf{i}\mathcal{D}(x))^2$) and then destroy them both with a single ϕ field (this is $-ig$ again). This is just the position-space diagram: 

Altogether:

$$\begin{aligned}\mathbf{i}\Sigma(q) &= \frac{1}{2}\int d^D x\ e^{\mathbf{i}qx}\,(-\mathbf{i}g)^2\,\mathbf{i}\mathcal{D}(x)\mathbf{i}\mathcal{D}(x) \\ &= \frac{1}{2}g^2 \int d\Phi \frac{1}{k_1^2 - m_1^2 + \mathbf{i}\epsilon}\frac{1}{k_2^2 - m_2^2 + \mathbf{i}\epsilon}\end{aligned} \tag{3.10}$$

In the bottom expression, the $\mathbf{i}\epsilon$s are designed to produce the *time-ordered* $\mathcal{D}(x)$s. Consider instead the strange combination

$$0 = \frac{1}{2}\int d^D x\ e^{\mathbf{i}qx}\,(\mathbf{i}g)^2\,\mathbf{i}\mathcal{D}_{\mathrm{adv}}(x)\mathbf{i}\mathcal{D}_{\mathrm{ret}}(x)$$

$$= \frac{1}{2}g^2 \int d\Phi \frac{1}{k_1^2 - m_1^2 - \sigma_1 \mathbf{i}\epsilon} \frac{1}{k_2^2 - m_2^2 + \sigma_2 \mathbf{i}\epsilon} \tag{3.11}$$

where $\sigma_{1,2} \equiv \text{sign}(k^0_{1,2})$. This expression vanishes because the integrand is identically zero: there is no value of t for which both the advanced and retarded propagators are nonzero (one has a $\theta(t)$ and the other has a $\theta(-t)$, and this is what's accomplished by the red σs). Therefore, we can add the imaginary part of zero

$$\text{Im}\,(\mathbf{i}0) = \frac{1}{2}g^2 \int d\Phi \,(\mathcal{P}_1\mathcal{P}_2 + \sigma_1\sigma_2\Delta_1\Delta_2)$$

to our expression for $\text{Im}\,\Sigma_{\text{1-loop}}$ to cancel the annoying principal part bits:

$$\text{Im}\,\Sigma_{\text{1-loop}} = \frac{1}{2}g^2 \int d\Phi \,((1 + \sigma_1\sigma_2)\,\Delta_1\Delta_2)\,.$$

The quantity $(1 + \sigma_1\sigma_2)$ is only nonzero (equal to 2) when k_1^0 and k_2^0 have the same sign; but in $d\Phi$ is a delta function that sets $q^0 = k_1^0 + k_2^0$. WLOG we can take $q^0 > 0$ since we only care about the propagation of positive-energy states. Therefore both k_1^0 and k_2^0 must be positive.

The result is that the only values of k on the RHS that contribute are ones with *positive* energy, and which satisfy all the momentum conservation constraints:

$$\text{Im}\,\Sigma = \frac{1}{2}g^2 \int d\Phi 2\theta(k_1^0)\theta(k_2^0)\Delta_1\Delta_2 = \frac{1}{2}g^2 2 \int d\Phi\theta(k_1^0)\theta(k_2^0)\pi\delta(k_1^2 - m^2)\pi\delta(k_2^2 - m^2)$$

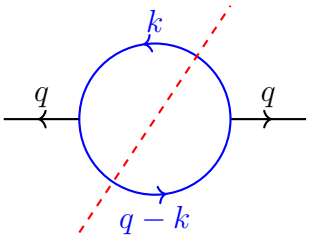
$$= \frac{g^2}{2}\frac{1}{2}\int \frac{\text{đ}^{D-1}\vec{k}_1}{2\omega_{\vec{k}_1}} \frac{\text{đ}^{D-1}\vec{k}_2}{2\omega_{\vec{k}_2}} (2\pi)^D \delta^D(k_1 + k_2 - q)\ .$$

In the last step we used the identity $\theta(k^0)\delta(k^2 - m^2) = \theta(k^0)\frac{\delta(k^0 - \omega_k)}{2\omega_k}$. But this is exactly (half) the density of actual final states into which the thing can decay! In summary:

$$\text{Im}\,\Sigma = \frac{1}{2} \sum_{\substack{\text{actual states } n \text{ of 2 particles} \\ \text{into which } \phi \text{ can decay}}} \| \mathcal{A}_{\phi \to n} \|^2 = m\Gamma. \tag{3.12}$$

In this example the decay amplitude $\mathcal{A}$ is just $\mathbf{i}g$. And the $\frac{1}{2}$ symmetry factor matches the factor that accounts for identical particles in the final state. (The other factor of two is part of the optical theorem, as we'll see next.) In the last step we compared to our expression for the decay rate (p. 109 of my 215A notes)[38]

[38] Actually there is a subtlety in this argument which is sometimes important: the optical theorem



This result is generalized by the *Cutkosky cutting rules* for finding the imaginary part of a Feynman diagram describing a physical process. The rough rules are the following. Assume the diagram is amputated – leave out the external propagators. Then any line drawn through the diagram that separates initial and final states (as at right) will 'cut' through some number of internal propagators; to find a contribution to the imaginary part, replace each of the cut propagators by $\theta(p^0)2\pi\delta(p^2-m^2) = \theta(p^0)\frac{2\pi\delta(p_0-\epsilon_p)}{2\epsilon_p}$. As Tony Zee says: the amplitude becomes imaginary when the intermediate particles become real (as opposed to virtual), aka 'go on-shell'. This is a place where the $\mathbf{i}\epsilon$s are crucial.

Why did I study this complicated theory with two scalars rather than just a single ϕ field? A single ϕ particle of mass m *cannot* decay into two ϕ particles each of mass m – the kinematics of this example do not allow any final state phase space. In that case the ϕ particle is stable (as it is in the theory above, too) as you can see by repeating the above calculation. What we learn from the calculation above is the simple fact that the heavy particle is unstable.

Optical Theorem. The general form of (3.12) is a general consequence of unitarity. Recall that the S-matrix is

$$\mathcal{S}_{fi} = \langle f|\, e^{-\mathbf{i}\mathbf{H}T}\,|i\rangle \equiv (\mathbb{1} + \mathbf{i}\mathcal{T})_{fi}\,.$$

$$\mathbf{H} = \mathbf{H}^\dagger \implies \mathbb{1} = \mathcal{S}^\dagger\mathcal{S} \implies 2\mathrm{Im}\,\mathcal{T} \equiv \mathbf{i}\left(\mathcal{T}^\dagger - \mathcal{T}\right) \overset{\mathbb{1}=\mathcal{S}^\dagger\mathcal{S}}{=} \mathcal{T}^\dagger\mathcal{T}.$$

This is called the optical theorem and it is the same as the one taught in some QM classes. The only difference here as that the matrices are much bigger. In terms of matrix elements:

$$2\mathrm{Im}\,\mathcal{T}_{fi} = \sum_n \mathcal{T}^\dagger_{fn}\mathcal{T}_{ni} \tag{3.13}$$

Here we've inserted a resolution of the identity (again on the QFT Hilbert space, the same scary sum) in between the two $\mathcal{T}$ operators. In the one-loop approximation, in the ϕ^3 theory here, the intermediate states that can contribute to $\sum_n$ are two-particle states, so that $\sum_n$ will turn into $\int \frac{\mathrm{d}^d\vec{k}_1}{2\omega_{k_1}}\,\frac{\mathrm{d}^d\vec{k}_2}{2\omega_{k_2}}\not{\phi}^D(k_T)$, the two-particle density of states.

A bit more explicitly, introducing a basis of scattering states

$$\langle f|\,\mathcal{T}\,|i\rangle = \mathcal{T}_{fi} = \not{\delta}^4(p_f - p_i)\mathcal{M}_{fi}, \quad \mathcal{T}^\dagger_{fi} = \not{\delta}^4(p_f - p_i)\mathcal{M}^\star_{if}, \quad \text{(recall that } \not{\delta}^d \equiv (2\pi)^d\delta^d\text{)}$$

is a statement about S-matrix elements, meaning that the external states must be on-shell. Above, we have focused on part of a diagram, namely the propagator for the unstable Φ particle, and its self energy – the external Φ lines are *not* on-shell here.

we have (denoting by N the number of particles in the intermediate state)

$$\begin{aligned}\langle F|\,\mathcal{T}^\dagger \mathbb{1}\mathcal{T}\,|I\rangle &= \sum_N \langle F|\,\mathcal{T}^\dagger \prod_{f=1}^N \int \frac{\text{đ}^3 q_f}{2E_f}\,|\{q_f\}\rangle\,\langle\{q_f\}|\,\mathcal{T}\,|I\rangle \\ &= \sum_N \prod_{f=1}^N \int \frac{\text{đ}^3 q_f}{2E_f} \text{đ}^4(p_F - \sum_f q_f)\mathcal{M}^\star_{\{q_f\}F}\text{đ}^4(p_I - \sum_f q_f)\mathcal{M}_{\{q_f\}I}\end{aligned}$$

Now notice that in this basis we have a $\text{đ}^4(p_F - p_I)$ on both sides of (3.13), and

$$\prod_{f=1}^N \int \frac{\text{đ}^3 q_f}{2E_f}\text{đ}^4(p_F - \sum_f q_f) = \int d\Pi_N$$

is the final-state phase space of the N particles. Therefore, the optical theorem says

$$\mathbf{i}\,(\mathcal{M}^\star_{IF} - \mathcal{M}_{FI}) = \sum_N \int d\Pi_N \mathcal{M}^\star_{\{q_f\}F}\mathcal{M}_{\{q_f\}I}.$$

Now consider *forward scattering*, $I = F$ (notice that here it is crucial that $\mathcal{M}$ is the transition matrix, $S = \mathbb{1} + \mathbf{i}\mathcal{T} = \mathbb{1} + \mathbf{i}\text{đ}(p_T)\mathcal{M}$):

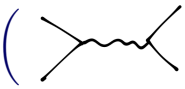
$$2\text{Im}\,\mathcal{M}_{II} = \sum_N \int d\Pi_N |\mathcal{M}_{\{q_f\}I}|^2.$$

For the special case of 2-particle scattering, we can relate the RHS to the total cross section for $2 \to$ anything in the CoM frame:

$$\text{Im}\,\mathcal{M}(k_1, k_2 \leftarrow k_1, k_2) = 2E_{cm}p_{cm}\sigma(\text{anything} \leftarrow k_1, k_2). \tag{3.14}$$

In more complicated examples (such as a box diagram contributing to 2-2 scattering), there can be more than one way to cut the diagram. Different ways of cutting the diagram correspond to discontinuities in different kinematical variables. To get the whole imaginary part, we have to add these up. A physical cut is a way of separating all initial-state particles from all final-state particles by cutting only internal lines. So for example, a t-channel tree-level diagram (like ) never has any imaginary part; this makes sense because the momentum of the exchanged particle is spacelike.

Resonances. A place where this technology is useful is when we want to study short-lived particles. In our formula for transition rates and cross sections we assumed plane waves for our external states. Some particles don't live long enough for

separately producing them:  and then watching them decay:  ; instead we must find them as resonances in scattering amplitudes of other particles: Im  .

So, consider the case $\mathbf{i}\mathcal{M}_{1\to1}\not{\delta}^D(p_I - p_F) = \langle F|\,\mathbf{i}\mathcal{T}\,|I\rangle$ where both I and F are one-particle states. A special case of the LSZ formula says

$$\mathbf{i}\mathcal{M}_{1\to1} = \left(\sqrt{Z}\right)^2 (\text{amputated 1-1 amplitudes}) = Z\left(G^{-1}GG^{-1}\right) = ZG^{-1} = -\mathbf{i}Z\left(p^2 - M_0^2 - \Sigma(p)\right) \tag{3.15}$$

where $-\mathbf{i}\Sigma$ is the self-energy, sum of all connected and amputated 1PI diagrams with one particle in and one particle out. Let $\Sigma(p) = A(p^2)+\mathbf{i}B(p^2)$ (not standard notation), so that near the pole in question, the propagator looks like

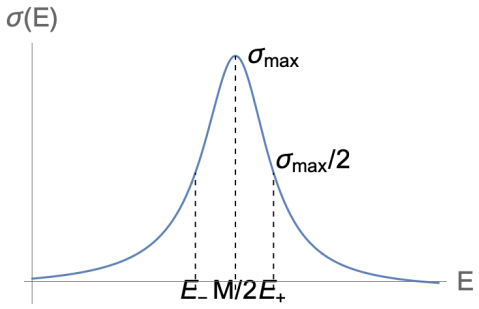
$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{p^2 - M_0^2 - \Sigma(p)} \simeq \frac{\mathbf{i}}{(p^2 - M^2)\underbrace{\left(1 - \partial_{p^2}A|_{M^2}\right)}_{=Z^{-1}} - \mathbf{i}B} = \frac{\mathbf{i}Z}{(p^2 - M^2) - \mathbf{i}BZ}. \tag{3.16}$$

In terms of the *particle width* $\Gamma_w \equiv -ZB(m^2)/M$, this is

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}Z}{(p^2 - M^2) + \mathbf{i}M\Gamma_w}.$$

So, if we can make the particle whose propagator we're discussing in the s-channel, the cross-section will be proportional to

$$\left|\tilde{G}^{(2)}(p)\right|^2 = \left|\frac{\mathbf{i}Z}{(p^2 - m^2) - \mathbf{i}m\Gamma_w}\right|^2 = \frac{Z^2}{(p^2 - M^2)^2 + M^2\Gamma_w^2}$$



a Lorentzian or Breit-Wigner distribution: In the COM frame, $p^2 = 4E^2$, and the cross section $\sigma(E)$ has a resonance peak at $2E = M$, with width Γ_w. It is the width in the sense that the function is half its maximum when $E = E_\pm = \sqrt{\frac{M(M\pm\Gamma_w)}{4}} \simeq \frac{M}{2} \pm \frac{\Gamma_w}{4}$.

This width is the same as the decay rate, because of the optical theorem:

$$\Gamma_w = -\frac{BZ}{M} \overset{(3.15)}{=} -\frac{1}{M}\left(-\operatorname{Im}\mathcal{M}_{1\to1}\right) \overset{\text{optical}}{=} \frac{1}{M}\frac{1}{2}\sum_n \int_f d\Pi_n |\mathcal{M}_{\{q_f\}1}|^2 = \Gamma$$

the last equation of which is exactly our formula for the decay rate. If it is not the case that $\Gamma \ll M$, *i.e.* if the resonance is too broad, the Taylor expansion of the inverse propagator we did in (3.16) may not be such a good idea. [End of Lecture 12]

Unitarity and high-energy physics. Two comments: (1) there had better not be any cutoff dependence in the imaginary part. If there is, we'll have trouble cancelling it by adding counterterms – an imaginary part of the action may destroy unitarity. This is elaborated a bit in Zee's discussion.

(2) Being bounded by 1, probabilities can't get too big. Cross sections are also bounded: there exist precise bounds from unitarity on the growth of cross sections with energy, such as the Froissart bound, $\sigma_{\text{total}}(s) \leq C \ln^2 s$ for a constant C. Xi Yin's notes describe a proof.

On the other hand, consider an interaction whose coupling G has mass dimension k. The cross section to which G contributes has dimensions of area, and comes from squaring an amplitude proportional to G, so comes with at least two powers of G. At $E \gg$ anything else, these dimensions must be made up with powers of E:

$$\sigma(E \gg ...) \sim G^2 E^{-2-2k}. \tag{3.17}$$

This means that if $k \leq -1$, the cross section grows at high energy, faster than the Froissart bound allows. In such a case, something else must happen to 'restore unitarity'. One example is Fermi's theory of Weak interactions, which involves a 4-fermion coupling $G_F \sim M_W^{-2}$. Here we know what happens, namely the electroweak theory, about which more soon. In gravity, $G_N \sim M_{\text{Pl}}^{-2}$, we can't say yet.

3.3 How to study hadrons with perturbative QCD

[Peskin §18.4] Here is a powerful physics application of both the optical theorem and the spectral representation. Consider the total inclusive cross section for e^+e^- scattering at energies $s = (k + k_+)^2 \gg m_e^2$, so that we can treat the electron as massless. In that limit $2E_{\text{cm}}p_{\text{cm}} = s$, and (3.14) becomes

$$\sigma^{\text{anything}\leftarrow e^+e^-} \overset{\text{optical thm}}{=} \frac{1}{s}\text{Im}\,\mathcal{M}_{\text{forward}}(e^+e^- \leftarrow e^+e^-) \tag{3.18}$$

where on the RHS, $\mathcal{M}_{\text{forward}}$ is the forward scattering amplitude (meaning that the initial and final electrons have the same momenta and spin)[39]. We've learned a bit about the contributions of electrons and muons to the BHS of this expression, what about QCD?

To leading order in α (small), but to all orders in the strong coupling α_s (big at

[39]Note that this fixes a factor of two in Peskin 18.80. Thanks to Aneesh Manohar for finding the problem.

low energies), the contributions of QCD look like

$$\mathbf{i}\mathcal{M}_h = \quad = (-\mathbf{i}e)^2 \bar{u}(k)\gamma_\mu v(k_+)\frac{-\mathbf{i}}{s}\mathbf{i}\Pi_h^{\mu\nu}(q)\frac{-\mathbf{i}}{s}\bar{v}(k_+)\gamma_\nu u(k)$$

with

$$= \mathbf{i}\Pi_h^{\mu\nu}(q) \overset{\text{Ward}}{=} \mathbf{i}(q^2\eta^{\mu\nu} - q^\mu q^\nu)\Pi_h(q^2)$$

the hadronic contribution to the vacuum polarization. We can pick out the contribution of the strong interactions by just keeping these bits on the BHS of (3.18):

$$\sigma^{\text{hadrons}\leftarrow e^+e^-} = \frac{1}{4}\sum_{\text{spins}}\frac{\text{Im}\,\mathcal{M}_h}{2s} \overset{s\gg m_e^2}{\simeq} -\frac{4\pi\alpha}{s}\text{Im}\,\Pi_h(s). \tag{3.19}$$

(The initial and final spins are equal and I've averaged the BHS over initial spins. We can ignore the longitudinal term $q^\mu q^\nu$ by the Ward identity. The spinor trace is $\sum_{\text{spins}} \bar{u}(k)\gamma_\mu v(k_+)\bar{v}(k_+)\gamma^\mu u(k) = -8k\cdot k_+ - 16m^2 \overset{s\gg m_e^2}{\simeq} -4s$.) As a reality check, consider the contribution from one loop of a heavy lepton of mass $M^2 \gg m_e^2$:

$$\text{Im}\,\Pi_L(s+\mathbf{i}\epsilon) = -\frac{\alpha}{3}F(M^2/s)$$

and

$$\sigma^{L^+L^-\leftarrow e^+e^-} = \frac{4\pi}{3}\frac{\alpha^2}{s}F(M^2/s)$$

with

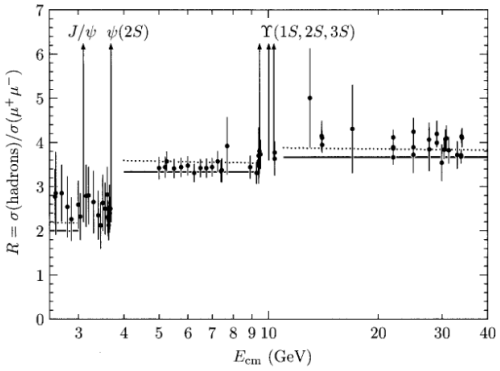
$$F(M^2/s) = \begin{cases} 0, & s < (2M)^2 \\ \sqrt{1-\frac{4M^2}{s}}\left(1+\frac{2M^2}{s}\right) = 1 + \mathcal{O}\left(M^2/s\right), & s > (2M)^2 \end{cases}.$$

For our purposes here, the important dofs of QCD are the quark fields q_f. They are Dirac spinors, with Lagrangian $L_q = \sum_f \bar{q}_f\left(\mathbf{i}\not{D} - m_f\right)q_f$, $D_\mu = \partial_\mu - \mathbf{i}Q_f A_\mu + ...$, where the ... is the coupling to the gluon field which we'll discuss soon enough. They have a color index that runs from 1 to 3 and which I've suppressed. The important point here is that they look just like electrons or muons in their coupling to the photon.

In perturbative QCD, the leading order result is therefore just the same contribution as above, a term from each quark with small enough mass:

$$\sigma_0^{\text{quarks}\leftarrow e^+e^-} = \underbrace{3}_{\text{colors}}\sum_{\text{flavors, f}} Q_f^2\frac{4\pi}{3}\frac{\alpha^2}{s}F(m_f^2/s). \tag{3.20}$$

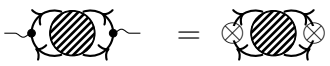
This expression (3.20) actually does remarkably well as a crude approximation to the measured $\sigma(\text{hadrons} \leftarrow e^+e^-)$ – see Fig. 5.3 of Peskin, at right. (This figure does not appear in the paper Peskin cites, I'm not sure of the correct provenance.) The key point is that the ratio of the hadronic cross section to that for muons in the final state jumps at $E = 2m_f$ where m_f is the mass of each new quark flavor (you can see $m_c \sim 1.3$ GeV and $m_b \sim 4.5$ GeV in the figure). See Peskin pp 139-141 for more. Furthermore, the size of the jump has information about the number of colors of quarks and their electric charges.



But Q: why is a perturbative analysis of QCD relevant here? You might think asymptotic freedom means QCD perturbation theory is good at high energy or short distances, and that seems to be borne out by noticing that Π_h is a two-point function of the quark contributions to the EM current:

$$\mathbf{i}\Pi_h^{\mu\nu}(q) = -e^2 \int d^4x \; e^{-\mathbf{i}q\cdot x} \langle\Omega| \mathcal{T} J^\mu(x) J^\nu(0) |\Omega\rangle \,, \quad J^\mu(x) \equiv \sum_f Q_f \bar{q}_f(x) \gamma^\mu q_f(x).$$

In diagrams what I am saying is:



where the $\otimes$ represents the quark part of the electromagnetic current operator $J^\mu(x)$. Maybe it looks like we are taking $x \to 0$ and therefore studying short distances. But if we are interested in large *timelike* q^μ here, that means that dominant contributions to the x integral are when the two points are timelike separated, and in the resolution of the identity in between the two Js includes physical states of QCD with lots of *real* hadrons. In contrast, the limit where we can do (maybe later we will learn how) perturbative QCD is when $q^2 = -Q_0^2 < 0$ is *spacelike* and large. (Preview: We can use the operator product expansion of the two currents.)

How can we use this knowledge to find the answer in the physical regime of $q^2 > 0$? The fact that Π_h is a two-point function of a gauge-invariant operator means that it has a spectral representation. It is analytic in the complex q^2 plane except for a branch cut on the positive real axis coming from production of real intermediate states, exactly where we want to know the answer. One way to encode the information we know is to

package it into moments:

$$I_n \equiv -\frac{4\pi\alpha}{n!}\left(\partial_{q^2}\right)^n \Pi_h|_{q^2=-Q_0^2} = -4\pi\alpha \oint_{C_{Q_0}} \frac{dq^2}{2\pi\mathbf{i}} \frac{\Pi_h(q^2)}{(q^2+Q_0^2)^{n+1}}$$

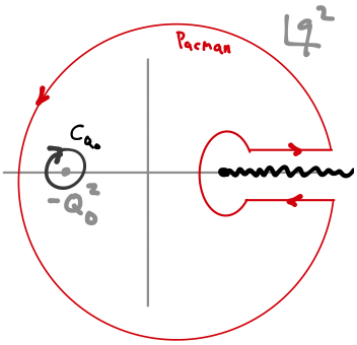
Here C_{Q_0} is a tiny contour around the point $q^2 = -Q_0^2$, a large spacelike momentum. The idea here is that this quantity can be computed by perturbative QCD.

But non-analyticities in $\Pi(q^2)$ only come from physical reasons – intermediate states going on-shell. Since the physical states all have $q^2 > 0$, we can deform the contour freely, away from the branch cut.

On the other hand, we know from the (appropriate generalization to currents of the) spectral representation sum rule (3.9) that $\Pi_h(q^2) \overset{|q|\gg\ldots}{\lesssim} \log(q^2)$, so for $n \geq 1$, the contour at infinity can be ignored.

Therefore (at the last step we also use $\mathrm{Disc}(\Pi_h) = 2\mathbf{i}\mathrm{Im}\,\Pi_h$)

$$\begin{aligned} I_n &= -4\pi\alpha \oint_{\text{Pacman}} \frac{dq^2}{2\pi\mathbf{i}} \frac{\Pi_h(q^2)}{(q^2+Q_0^2)^{n+1}} \\ &= -4\pi\alpha \int \frac{dq^2}{2\pi\mathbf{i}} \frac{\mathrm{Disc}\Pi_h}{(q^2+Q_0^2)^{n+1}} \\ &\overset{(3.19)}{=} \frac{1}{\pi}\int_{s_{\text{threshhold}}}^{\infty} ds \frac{s}{(s+Q_0^2)^{n+1}} \sigma^{\text{hadrons}\leftarrow e^+e^-}(s). \end{aligned}$$



On the RHS is (moments of) the measurable (indeed, measured) cross-section, and on the LHS is things we can calculate (later). If the convergence of these integrals were uniform in n, we could invert this relation and directly try to predict the cross section to hadrons. But it is not, and the correct cross section varies about the leading QCD answer more and more at lower energies, culminating at various Breit-Wigner resonance peaks at $\bar{q}q$ boundstates (such as J/ψ and Υ which you can see in the plot I took from Peskin's book above). These only exist because of QCD interactions that we've ignored in our crude approximation σ_0, (3.20).

4 Gauge theory

4.1 The Anderson-Higgs Mechanism and superconductors

Landau-Ginzburg EFT of superconductors: Massive vector fields as gauge fields. Consider a massive vector field B_μ with Lagrangian density

$$\mathcal{L}_B = -\frac{1}{4e^2}(dB)_{\mu\nu}(dB)^{\mu\nu} + \frac{1}{2}m^2 B_\mu B^\mu \tag{4.1}$$

where $(dB)_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu$. (Note the funny-looking sign of the mass term which comes from $B^\mu B_\mu = B_0^2 - B_i^2$.) The mass term is not invariant under $B_\mu \to B_\mu + \partial_\mu \lambda$, the would-be gauge transformation. We can understand the connection between massive vector fields and gauge theory by the 'Stueckelberg trick' of pretending that the gauge parameter is a field: Let $B_\mu \equiv A_\mu - \partial_\mu \theta$ where θ is a new degree of freedom. Since B is invariant under the transformation

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \lambda(x), \;\; \theta(x) \to \theta(x) + \lambda(x),$$

so is any functional of B. Notice that the fake new field θ transforms non-linearly (*i.e.* its transformation is affine). This was just a book-keeping step, but something nice happens:

$$(dB)_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu}$$

is the field strength of A. The mass term becomes

$$B_\mu B^\mu = (A_\mu - \partial_\mu \theta)(A^\mu - \partial^\mu \theta).$$

This contains a kinetic term for θ. We can think of this term as (energetically) setting θ equal to the longitudinal bit of the gauge field. One nice thing about this reshuffling is that the $m \to 0$ limit decouples the longitudinal bits. Furthermore, if we couple a conserved current ($\partial^\mu j_\mu = 0$) to B, then

$$\int d^D x \; j_\mu B^\mu = \int d^D x j_\mu A^\mu$$

it is the same as coupling to A_μ.

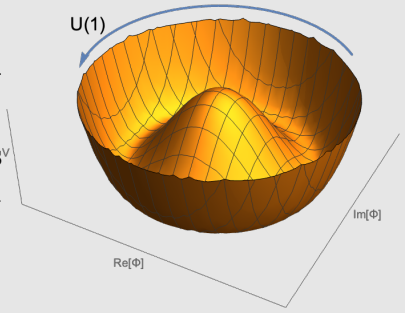
Who is θ? Our previous point of view was that it is fake and we can just choose the gauge parameter $\lambda(x)$ to get rid of it, and set $\theta(x) = 0$ (in which case, $B = A$). This is called *unitary gauge*, and gives us back the Proca theory of $B = A$, (4.1).

Consider, as an aside, the following model of a single complex scalar:

$$\mathcal{L}_{\text{global}} \equiv +\frac{1}{2}|\partial_\mu \Phi|^2 - V(|\Phi|)$$

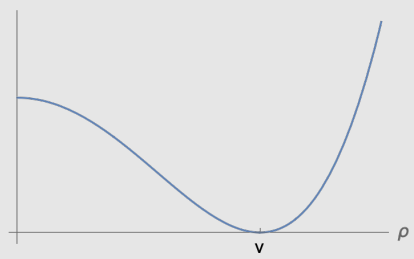
and let's take

$$V(|\Phi|) = \kappa(|\Phi|^2 - v^2)^2$$



for some couplings κ, v. This potential has a $\mathsf{U}(1)$ symmetry $\Phi \to e^{\mathbf{i}\lambda}\Phi$, and a circle of minima at $|\Phi|^2 = v^2$ (if $v^2 > 0$, which we'll assume). When $v^2 > 0$, the mass term about the origin has the wrong sign. In polar coordinates in field space, $\Phi \equiv \rho e^{\mathbf{i}\theta}$, the Lagrangian is

$$\mathcal{L}_{\text{global}} = +\frac{1}{2}\rho^2(\partial_\mu\theta)^2 + \frac{1}{2}(\partial\rho)^2 - V(\rho) \overset{E \ll V''(v)}{\simeq} \frac{1}{2}v^2(\partial\theta)^2.$$



In the last step, we observed that the excitations of the ρ field are have mass-squared $V''(v) = 8\kappa v^2$ about the minimum; below that energy scale, we can integrate it out and ignore it. The θ field is the massless Goldstone boson, which parametrizes the circle of minima.

This is a long-wavelength description of a (relativistic, because of the kinetic terms) *superfluid*. Here I am making a distinction between superfluid and superconductor; the former spontaneously breaks a continuous global symmetry and has a Goldstone boson in its spectrum.

In the groundstate of this system, the field Φ is said to condense, in the sense that it has a expectation value. We can also say that the particles created by Φ condense, in the sense that the field that creates them has an expecation value.

Now consider the following theory, related to the previous by *gauging* the $\mathsf{U}(1)$ symmetry:

$$\mathcal{L}_h \equiv -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}|D_\mu\Phi|^2 - V(|\Phi|)$$

where Φ is a complex, charged scalar field whose covariant derivative is $D_\mu\Phi = (\partial_\mu - \mathbf{i}qA_\mu)\,\Phi$, with the same V as above. This is called an *Abelian Higgs model*. The $\mathsf{U}(1)$ symmetry is *gauged*, in the sense that $A_\mu \to A_\mu + \partial_\mu\lambda(x), \Phi(x) \to e^{\mathbf{i}q\lambda(x)}\Phi(x)$ is an invariance of the action, and we've learned to regard such a *local* invariance as a redundancy of the description.

In polar coordinates in field space, $\Phi \equiv \rho e^{\mathbf{i}\theta}$, the Lagrangian is now

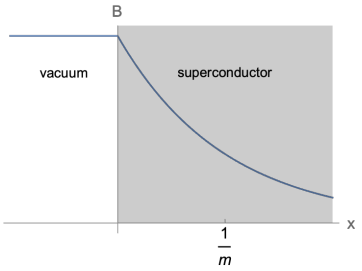
$$\mathcal{L}_h = -\frac{1}{4e^2} F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}q^2\rho^2(A_\mu - \partial_\mu\theta)^2 + \frac{1}{2}(\partial\rho)^2 - V(\rho).$$

This differs from the action for B written in terms of A, θ only in the addition of the *Higgs mode* ρ. Again we can go to unitary gauge and set $\theta = 0$. We find a massive gauge field A, plus a massive scalar ρ whose mass (expanding $V(\rho)$ about $\rho = v$) is [End of Lecture 13]

$$\partial_\rho^2 V|_{\rho=v} = m_\rho^2 = 8\kappa v^2 \overset{\kappa\gg 1}{\gg} m_A^2 = q^2\langle\rho\rangle^2 = q^2v^2.$$

That is: in the limit of large κ, the excitations of ρ are hard to make, and we get back $\mathcal{L}_B$. For any value of κ, we can say that the gauge field eats the would-be Goldstone boson θ and becomes heavy, in a manner consistent with gauge invariance[40]. This is the Anderson-Higgs mechanism.

The description of massive gauge fields in terms of $\mathcal{L}_h$ via the Anderson-Higgs mechanism is more useful than $\mathcal{L}_B$ for thinking about the renormalization of massive gauge fields: for example it is renormalizable, even if we couple A to other charged fields (*e.g.* Dirac fermions). This mechanism also works in the case of non-Abelian gauge fields and is an important ingredient in the (electroweak sector of the) Standard Model.



It is also a description of what happens to the EM field in a superconductor: the photon gets a mass; the resulting expulsion of magnetic flux is called the Meissner effect. For example, if we immerse a region $x > 0$ with $\Phi = v$ in an external constant magnetic field B_0,

$$0 = \partial_\mu F^{\mu\nu} - m^2 A^\nu \implies B(x) = Be^{-mx}. \qquad (4.2)$$

Another consequence of the mass is that if we do manage to sneak some magnetic flux into a superconductor, the flux lines will bunch up into a localized string. This can be shown by solving the equations of motion of the model above (see the homework). This is called a vortex (or vortex string in $3d$) because of what Φ does in this configuration: its phase winds around the defect. In a superconductor, the role of Φ is played by the Cooper pair field (which has electric charge $q = 2$). The fact that Φ has charge two is visible in the flux quantization of the vortices (this is part of the homework problem). I hope to say more about its origins in terms of electrons later.

I mention here the Meissner effect and the associated collimation of flux lines partly because it is helpful for developing a picture of confinement. In particular: think

[40] You can check that the mixing with θ is exactly what's required to make $\Pi(q)$ singular enough at $q = 0$ to give A a mass consistent with the Ward identity, as in our discussion at (2.63).

about the energetics of a magnetic monopole (suppose we had one available[41]) in a superconductor. If we try to insert it into a superconductor, it will trail behind it a vortex string along which all of its exiting magnetic flux is localized. This string has a finite tension (energy per unit length), as you'll study on the homework. If we make the superconducting region larger and larger, the energy of the monopole configuration grows linearly in the size – it is not a finite energy object in the thermodynamic limit. If monopoles were dynamical excitations of rest mass M_m, it would eventually become energetically favorable to pop an antimonopole out of the vacuum, so that the flux string connects the monopole to the antimonopole – this object can have finite energy inside the superconductor. But notice that in a region where electric charge is condensed, a single monopole is *confined* by the magnetic flux string it must carry around. A confining state of a gauge theory is like this for the *electric* charges, because *magnetic* charge is condensed. The two pictures are related by electromagnetic duality, which acts by

$$\vec{E} \to \vec{B}, \vec{B} \to -\vec{E}, e \to \frac{1}{e}, j_e^\mu \leftrightarrow j_m^\mu \tag{4.3}$$

where $j_{e,m}$ are the electric and magnetic current densities.

4.2 Festival of gauge invariance

Consider a collection of N complex scalar fields (we could just as well consider spinors) with, for definiteness, an action of the form

$$\mathcal{L} = \sum_{\alpha=1}^{N} \partial_\mu \Phi_\alpha^\star \partial^\mu \Phi_\alpha - V(\Phi_\alpha^\star \Phi_\alpha) \tag{4.4}$$

(or $\mathcal{L} = \bar{\Psi}_\alpha(\not{\partial} - m)\Psi_\alpha + (\bar{\Psi}_\alpha \Gamma \Psi_\alpha)^2$). So far, the model actually has an $\mathsf{O}(2N)$ symmetry except that for kicks I grouped the scalars into pairs, and wrote the potential in terms of the combination $\sum_{\alpha=1}^{N} \Phi_\alpha^\star \Phi_\alpha$.

Lighting review of Lie groups and Lie algebras. (4.4) is invariant under the $\mathsf{U}(N)$ transformation

$$\Phi_\alpha \mapsto \Lambda_{\alpha\beta} \Phi_\beta, \quad \Lambda^\dagger \Lambda = \mathbb{1}. \tag{4.5}$$

Any such $\mathsf{U}(N)$ matrix Λ can be parametrized as

$$\Lambda = \Lambda(\lambda) = e^{\mathbf{i} \sum_{A=1}^{N^2-1} \lambda^A T^A} e^{\mathbf{i}\lambda^0}.$$

λ^0 parametrizes a $\mathsf{U}(1)$ factor that commutes with everyone; we already know something about $\mathsf{U}(1)$ gauge theory from QED, so we won't focus on that. We'll focus on the

[41] Here is the paper about the only one that's been detected by humans so far.

non-abelian part: the T^A are the generators of $\mathsf{SU}(N)$, and are traceless, so $\mathsf{SU}(N) \ni \Lambda(\lambda^0 = 0)$ has $\det \Lambda(\lambda^0 = 0) = 1$. Here the index $A = 1...N^2 - 1 = \dim(\mathsf{SU}(N))$; the matrices T^A (and hence also Λ) are $N \times N$, and satisfy the Lie algebra relations

$$[T^A, T^B] = \mathbf{i} f_{ABC} T^C \tag{4.6}$$

where f_{ABC} are the structure constants of the Lie algebra $\mathsf{SU}(N)$. (4.6) is true because the $\mathbf{i}$ times the commutator of two hermitian $N \times N$ matrices is again a hermitian $N \times N$ matrix, and so must be some linear combination of the T^A. For the case of $\mathsf{SU}(2)$, $T^A = \frac{1}{2}\sigma^A, A = 1, 2, 3$, and $f_{ABC} = \epsilon_{ABC}$. The infinitesimal version of (4.5), with Λ close to the identity, is

$$\Phi_\alpha \mapsto \Phi_\alpha + \mathbf{i}\lambda^A T^A_{\alpha\beta}\Phi_\beta. \tag{4.7}$$

The $N \times N$ representation is called the *fundamental representation* of $\mathsf{SU}(N)$. Other representations of the group come from other sets of T^A_Rs that satisfy the same algebra (4.6), but can have other dimensions. For example, the structure constants themselves $\left(T^A_{\text{adj}}\right)_{BC} \equiv -\mathbf{i} f_{ABC}$ furnish the representation matrices for the adjoint representation.

Local invariance. The transformation above was *global* in the sense that the parameter λ was independent of spacetime. This is an actual symmetry of the physical system associated with (4.4). Let's consider how we might change the model in (4.4) to make it invariant under a *local* transformation, with $\lambda = \lambda(x)$. In the Abelian case, we have learned the recipe

$$\Phi \mapsto e^{\mathbf{i}\lambda(x)}\Phi(x), A_\mu \mapsto A_\mu + \partial_\mu \lambda, \quad \partial_\mu \Phi \rightsquigarrow D_\mu \Phi = (\partial_\mu - \mathbf{i}A_\mu)\Phi \mapsto e^{\mathbf{i}\lambda(x)} D_\mu \Phi.$$

In words: by replacing partial derivatives with covariant derivatives, we can make gauge-invariant Lagrangians. The same thing works in the non-abelian case:

$$\begin{gathered}(D_\mu \Phi)_\alpha \equiv \partial_\mu \Phi_\alpha - \mathbf{i} A^A_\mu T^A_{\alpha\beta}\Phi_\beta \\ \Phi \mapsto \Phi + \mathbf{i}\lambda^A(x) T^A \Phi, \quad A^A_\mu \mapsto A^A_\mu + \partial_\mu \lambda^A - f_{ABC}\lambda^B A^C_\mu(x).\end{gathered} \tag{4.8}$$

The difference is that there is a term depending on A in the shift of the gauge field A. The following Yang-Mills Lagrangian density is a natural generalization of Maxwell:

$$\mathcal{L}_{YM} = -\frac{1}{4g^2}\sum_A \left(\underbrace{\partial_\mu A^A_\nu - \partial_\nu A^A_\mu + f_{ABC} A^B_\mu A^C_\nu}_{=F^A_{\mu\nu} = -F^A_{\nu\mu}} \right)^2 = -\frac{1}{2g^2}\text{tr} F_{\mu\nu}F^{\mu\nu}. \tag{4.9}$$

[End of Lecture 14]

The field strength

$$F^A_{\mu\nu} \mapsto F^A_{\mu\nu} - f_{ABC}\lambda^B F^C_{\mu\nu} = F^A_{\mu\nu} + \mathbf{i}\lambda^B \left(T^B_{\rm adj}\right)_{AC} F^C_{\mu\nu} \tag{4.10}$$

is designed so that it transforms in the adjoint representation, and therefore S_{YM} is gauge-invariant. In the last step of (4.9), we regarded F as an $N\times N$ matrix $F = F^A T^A$, and chose a basis of the Lie algebra with $\mathrm{tr} T^A T^B = \frac{1}{2}\delta^{AB}$. In terms of this matrix F, the finite version of (4.10) is just $F \mapsto \Lambda F \Lambda^{-1}$ (compare (4.7)), which makes it manifest that $\mathrm{tr} F^2$ is invariant.

4.3 Interlude on differential forms

[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, which I would find it rather inhibiting not to be able to use in the next section. 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. Suppose x^μ are some local coordinates on X.

A p-form on X is made of 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 $(\mathrm{d}x)^2 = 0$. Under a coordinate transformation, they transform like $d\tilde{x}^\mu = \frac{\partial \tilde{x}^\mu}{\partial x^\nu} dx^\nu$. To understand the point of the antisymmetry of the product of forms, consider the case of two dimensions: $d\tilde{x}\wedge d\tilde{y} = J dx \wedge dy$, where $J = \det \frac{\partial(\tilde{x},\tilde{y})}{\partial(x,y)}$ is the Jacobian for the change of variables.

Familiar examples include the gauge potential $A = A_\mu \mathrm{d}x^\mu$, and its field strength $F = \frac{1}{2} F_{\mu\nu} \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$. The point in life of a p-form is to be integrated over a p-dimensional submanifold of spacetime. 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. For example, given a curve C in X parameterized as $x^\mu(s)$, we can make the scalar quantity

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

and this would be the same if we chose some other parameterization or some other local coordinates (by the chain rule). This is the phase acquired by a unit-charge particle when moving along the path C in the background gauge field configuration described by A.

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

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

[42] This product satisfies $A_p \wedge B_q = (-1)^{pq} B_q \wedge A_p$. The space of differentiable p-forms on a manifold X is sometimes denoted $\Omega^p(X)$; this a vector space (let's say over $\mathbb{R}$), since we can add p forms and multiply them by real numbers.

The *exterior derivative* d is a linear operator acting on forms as

$$\begin{aligned} \mathrm{d} : \Omega^p(X) &\to \quad \Omega^{p+1}(X) \\ A \quad &\mapsto \mathrm{d}A = dx^\nu \wedge \partial_\nu A \end{aligned}$$

by

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

You can check that

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

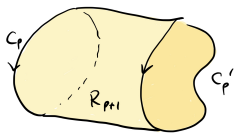
because derivatives commute when acting on smooth functions. Notice that $F = \mathrm{d}A$ in the example above. Denoting the boundary of a region D by ∂D, Stokes' theorem is

$$\int_{D_p} \mathrm{d}\alpha_{p-1} = \int_{\partial D_p} \alpha_{p-1}.$$

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: $\mathrm{d}\omega_p = 0$. ω_p closed means that $\int_{C_p} \omega_p$ depends only on the topology of C_p, in the sense that

$$\int_{C_p} \omega_p - \int_{C'_p} \omega_p = \int_{C_p - C'_p} \omega_p = \int_{\partial R_{p+1}} \omega_p \overset{\text{Stokes}}{=} \int_{R_{p+1}} d\omega_p = 0.$$



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

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

where [..] means sum over permutations with a -1 for odd permutations. Try not to get caught up in the numerical prefactors. In my expression below for the exterior derivative also there is an annoying combinatorial prefactor.

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. ω_p is exact means it is a total derivative, a boundary term, so $\int_{C_p} \omega_p \overset{\text{Stokes}}{=} \int_{\partial C_p} \alpha_{p-1}$ vanishes if C_p doesn't have a boundary.

To get some familiarity with the above language let's think about the case $\mathcal{M} = \mathbb{R}^3$ for a moment. Then $\Omega^0(\mathbb{R}^3)$ and $\Omega^3(\mathbb{R}^3)$ are both spanned by ordinary functions, while $\Omega^1(\mathbb{R}^3)$ and $\Omega^2(\mathbb{R}^3)$ are both spanned by vector fields – functions with a single index. On functions, $df = \partial_i f dx^i$. On 1-forms,

$$d(f_i dx^i) = (\partial_y f_z - \partial_z f_y)\, dy\wedge dz + (\partial_x f_y - \partial_y f_x)\, dx\wedge dy + (\partial_z f_x - \partial_x f_z)\, dz\wedge dx = \frac{1}{3!}\epsilon_{ijk}\partial_i f_j \epsilon_{klm} dx^l \wedge dx^m.$$

On 2-forms

$$d(f_x dy\wedge dz + f_y dz\wedge dx + f_z dx\wedge dy) = \partial_i f_i dx\wedge dy\wedge dz.$$

So this accounts for all the classic operations of vector calculus:

$$d(\text{0-form}) = \text{gradient}, \quad d(\text{1-form}) = \text{curl}, \quad d(\text{2-form}) = \text{divergence}.$$

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

A classical physics context where one encounters a cohomological question is in fluid dynamics: given a vector field, say describing the flow of a fluid on some space X, when is it the gradient of a well-defined function on X? Or in electrostatics on some space X, an allowed electric field configuration must be the gradient of a scalar potential on $X \backslash$ the locations of the charges.

Now suppose we have a metric on X, *i.e.* a way of measuring distances $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$. Then 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

$$\left(\star A^{(p)}\right)_{\mu_1 \dots \mu_{d-p}} \equiv \frac{\sqrt{g}}{p!} \epsilon_{\mu_1 \dots \mu_d} A^{(p)\ \mu_{d-p+1} \dots \mu_d}$$

where $\sqrt{g} \equiv \sqrt{\det g}$ and indices are raised with the inverse metric $g^{\mu\nu}$.

Here are some familiar statements written in the above language. The electromagnetic field is a 2-form on spacetime, $\mathbb{R}^4$ (I use ijk for spatial indices and $\mu\nu$ for spacetime indices):

$$F = dA = E_i dx^i \wedge dt + B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy = E_i dx^i \wedge dt + B_i dx^j \wedge dx^k \epsilon_{ijk}/2.$$

The dual field strength, in flat spacetime, is

$$\star F = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma} dx^\mu \wedge dx^\nu = -B_i dx^i \wedge dt + E_i dx^j \wedge dx^j \epsilon_{ijk}/2.$$

Maxwell's equations (away from charges) are $dF = 0, d \star F = 0$. (Including charges, they are $dF = \star j_m, d \star F = \star j_e$, which expression makes electromagnetic duality manifest.) The first is the Bianchi identity, which is automatic if A is well-defined, while the second is the equations of motion associated with the Maxwell action

$$S[A] = -\frac{1}{2e^2} \int F \wedge \star F = -\frac{1}{4e^2} \int d^D x \sqrt{g} F_{\mu\nu} F^{\mu\nu}.$$

Maxwell's equations say that both F and $\star F$ represent cohomology classes in spacetime (minus the locations of charges). Consider the simplest possible nontrivial example of a point charge at rest at the origin of coordinates. The field strength is $F = q\frac{dr\wedge dt}{r^2} = -qd\left(\frac{dt}{r}\right)$, well-defined in $\mathcal{M} \equiv \mathbb{R}^4 \setminus \mathbb{R}_t$, where we remove the origin at all times. F is clearly exact and hence represents the trivial class in $H^2(\mathcal{M})$. However, using $dr = x^i dx^i/r$ and *e.g.* $\star dx \wedge dt = dy \wedge dz$, we have

$$\star F = q\frac{xdy \wedge dz + ydz \wedge dx + zdx \wedge dy}{r}.$$

This is a nontrivial element of $H^2(\mathcal{M})$, as you can see by integrating it over the appropriate Gaussian surface, *i.e.* the 2-sphere surrounding the particle to get $\int_{S^2} \star F = 4\pi q$. So we can interpret the charge of the particle as an element of the cohomology group $H^2(\mathcal{M})$.

Abelian p-form gauge fields. The usual Maxwell field strength is $F_2 = dA_1$. It is invariant under gauge transformations $A_1 \to A_1 + d\lambda_0$ since $d^2 = 0$. A large family of useful generalizations of this is p-form abelian gauge fields:

$$F_{p+1} = dA_p, \quad \delta A_p = d\lambda_{p-1}.$$

Again the field strength is gauge invariant by $d^2 = 0$. An action is

$$S[A] = -\frac{1}{2g^2}\int F_{p+1} \wedge \star F_{p+1} \propto -\int d^D x \frac{\sqrt{g}}{(p+1)!} F_{\mu_1\cdots\mu_{p+1}} F^{\mu_1\cdots\mu_{p+1}}.$$

For $p = 0$, this is just $L = -\frac{1}{2g^2}(\partial\phi)^2$, a massless scalar. The equations of motion are

$$0 = \frac{\delta S}{\delta A(x)} \propto d \star F(x).$$

In flat spacetime of enough dimensions, we can fourier transform and see that this describes a massless excitation with a spin that depends on p and D.

The analog of minimal coupling for a p-form gauge field is to a p-dimensional worldvolume:

$$\Delta S = e\int_{X_p} A_p$$

– this is the worldvolume of a $(p-1)$-brane, an object with $p-1$ spatial dimensions.

In spacetime dimensions $D = 2 \bmod 4$, it is consistent with the equations of motion to impose a (anti-)self-duality equation: $F_{D/2} = \pm \star F_{D/2}$, which gets rid of half the degrees of freedom. In $D = 2$ this describes a chiral scalar. In $D = 4$, this describes only one circular polarization of the photon. The Maxwell-like action, however, vanishes when imposing this condition and there is no covariant action principle for such fields. They also enjoy various gravitational anomalies.

The duality operation $dA_p = \star dA^\vee_{D-p-2}$ exchanges the Bianchi identity and the equation of motion. We can learn something by giving a path-integral derivation of the duality. The partition function (euclidean) is

$$\int [dA] e^{-\frac{1}{2g^2}\int dA\wedge\star dA} = \int \frac{[dAdBdA^\vee]}{(\star)} e^{-\frac{1}{2g^2}\int (F-B)\wedge\star(F-B)+\mathbf{i}\int B\wedge dA^\vee} = \int [dA^\vee] e^{-\frac{g^2}{2}\int dA^\vee\wedge\star dA^\vee}.$$

In the first step, we introduce $A^\vee$ as a Lagrange multiplier to impose $dB = 0$ (and $\oint_X B \in 2\pi\mathbb{Z}$ for all 2-cycles X). The middle object has a new redundancy under

$$(\star) \qquad A \to A + \Lambda, B \to B + d\Lambda \tag{4.11}$$

for an arbitrary p-form Λ; when $dB = 0$ (and B has integral periods) this can be used to set $B = 0$, giving back the first expression. In the second step, we set $A = 0$ and do the gaussian integral over B, producing a nontrivial action for $A^\vee$. (The same manipulation works for other gauge-invariant observables.) Notice that the coupling constant gets inverted. A simple example of this is $p = 0, D = 2$ which relates a compact scalar of radius $R = \frac{1}{g}$ to one with radius R – this is called T-duality. The manipulation above is described in §2.2 here for $D = 4, p = 1$, and here for $D = 2, p = 0$.

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 this 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* $\Delta\tilde{\omega} = 0$.

An application of this is Poincare duality: $b^p(X) = b^{d-p}(X)$ if X has a volume form. This follows because the map $H^p \to H^{d-p}$ which takes $[\omega_p] \mapsto [\star\omega_p]$ 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.

For more in this direction, take a look at the notes here.

The machinery of differential forms is very useful.

4.4 Gauge fields as connections

The formulae in §4.2 are not too hard to verify, but where did they come from? Suppose we wanted to attach an N-dimensional complex vector space to each point in spacetime; on each vector space we have an action of $\mathsf{SU}(N)$, by $\Phi_\alpha(x) \mapsto \Lambda_{\alpha\beta}(x)\Phi_\beta(x)$. Suppose we would like to do physics in a way that is independent of the choice of basis for this space, at each point. We would like to be able to compare $\Phi(x)$ and $\Phi(y)$ (for example to make kinetic energy terms) in a way that respects these independent rotations. To do this, we need more structure: we need a *connection* (or *comparator*) W_{xy}, an object that transforms like (suppressing the $\alpha = 1..N$ indices) $W_{xy} \mapsto \Lambda(x)W_{xy}\Lambda^{-1}(y)$, so that $\Phi^\dagger(x)W_{xy}\Phi(y)$ is invariant. The connection between two points W_{xy} may depend on how we get from x to y. We demand that $W(x,x) = \mathbb{1}$, the composition law, $W(C_2 \circ C_1) = W(C_2)W(C_1)$ (this is matrix multiplication), and $W(-C) = W^{-1}(C)$, where $-C$ is the path C taken in the opposite direction.

But if we have a W_{xy} for any two points, you can't stop me from considering nearby points and defining (the covariant derivative)

$$D_\mu\Phi(x) \equiv \lim_{\Delta x\to 0} \frac{W(x,x+\Delta x)\Phi(x+\Delta x) - \Phi(x)}{\Delta x^\mu} \mapsto \Lambda(x)D_\mu\Phi(x) \ . \tag{4.12}$$

Expanding near $\Delta x \to 0$, we can let

$$W(x,x+\Delta x) = \mathbb{1} - \mathbf{i}e\Delta x^\mu A_\mu(x) + \mathcal{O}(\Delta x^2) \tag{4.13}$$

this defines the gauge field A_μ (sometimes also called the connection). To make the gauge transformation of the non-abelian connection field $A \mapsto A^\Lambda$ obvious, just remember that the covariant derivative of a field is designed to transform like the field: $D_\mu\Phi \mapsto D_\mu^{A^\Lambda}(\Lambda\Phi) \stackrel{!}{=} \Lambda\left(D_\mu^A\Phi\right)$ which means $A_\mu^\Lambda = \Lambda A_\mu\Lambda^{-1} - (\partial_\mu\Lambda)\,\Lambda^{-1}$. (This formula also works in the abelian case $\Lambda = e^{\mathbf{i}\lambda}$, and knows about the global structure of the group $\lambda \simeq \lambda + 2\pi$.)

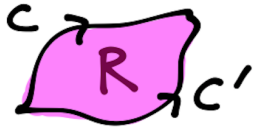
The composition law then implies that

$$W(x,y+\Delta x) = \left(\mathbb{1} - \mathbf{i}e\Delta x^\mu A_\mu(y) + \mathcal{O}(\Delta x^2)\right) W(x,y) \ . \tag{4.14}$$

This is a differential equation for W which can be integrated: $W_{xy} \stackrel{?}{=} e^{-\mathbf{i}e\int_{C_{xy}} A_\mu(\tilde{x})d\tilde{x}^\mu}$ where C_{xy} is a path in spacetime from x to y. What if G is not abelian? Then I need to tell you the ordering of the exponential. We know from Dyson's equation that the solution is

$$W_{xy} = \mathcal{P}e^{-\mathbf{i}e\int_{C_{xy}} A_\mu(\tilde{x})d\tilde{x}^\mu}$$

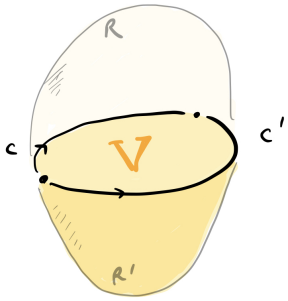
where $\mathcal{P}$ indicates *path-ordering* along the path C_{xy}, just like the time-ordered exponential we encountered in interaction-picture perturbation theory.



To what extent *does* W_{xy} depend on the path? In the abelian case,

$$W_C = W_{C'}e^{\mathbf{i}e\oint_{C-C'} A} \stackrel{\text{Stokes}}{=} W_{C'}e^{\mathbf{i}e\int_R F_{\mu\nu}dx^\mu dx^\nu}$$

where $\partial R = C - C'$ is a 2d surface whose boundary is the difference of paths.



But which 2d surface? (I continue to speak about the abelian case for the rest of this paragraph.) The difference in phase between two possible choices is $e^{\mathbf{i}e\int_{R-R'} F} \stackrel{\text{Stokes}}{=} e^{\mathbf{i}e\int_V dF}$ where $\partial V = R - R'$ is the 3-volume whose boundary is the difference of the two regions. The integrand vanishes by the Bianchi identity, which is actually an identity if $F = dA$ and A is smooth. You might think this prevents magnetic sources, which appear on the RHS of the Maxwell equation $dF = \star j_m$. But actually $\int_V dF$ only appears here in the combination $e^{\mathbf{i}e\int_V dF}$, so magnetic sources are perfectly consistent with independence of the choice of R, as long as their charge $q \equiv \int_V dF = \oint_{\partial V} F$ is quantized $ge \in 2\pi\mathbb{Z}$. This is Dirac quantization.

Imagine inserting an infinitesimal rectangle to the path which moves by dx^μ then by dx^ν and then back and back. The difference in the action on Φ is

$$dx^\mu dx^\nu [D_\mu, D_\nu]\Phi = -\mathbf{i}e dx^\mu dx^\nu F_{\mu\nu}\Phi. \quad \text{(no sum on } \mu,\nu) \tag{4.15}$$

The commutator of covariant derivatives is not a differential operator, but a function $[D_\mu, D_\nu] = -\mathbf{i}eF_{\mu\nu}$. (Note that this same maneuver defines the Riemann tensor in terms of derivatives covariant with respect to coordinate changes.) This same relation holds in the non-abelian case:

$$F_{\mu\nu} = \frac{\mathbf{i}}{e}[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu - \mathbf{i}e[A_\mu, A_\nu].$$

This object is Lie-algebra-valued, so can be expanded in a basis of the Lie algebra: $F_{\mu\nu} = F^A_{\mu\nu}T^A$, so more explicitly,

$$F^A_{\mu\nu} = \partial_\mu A^A_\nu - \partial_\nu A^A_\mu + ef_{ABC}A^B_\mu A^C_\nu.$$

Since it is made from products of covariant derivatives, $[D, D]\Phi \mapsto \Lambda[D, D]\Phi$, it must transform in the adjoint representation, $F \mapsto \Lambda F \Lambda^{-1}$, which in infinitesimal form returns us to (4.10)

$$F^A_{\mu\nu} \mapsto F^A_{\mu\nu} - f^{ABC}\lambda^B F^C_{\mu\nu}.$$

In this exposition, Maldacena uses currency exchange as an analogy to explain gauge theory. Suppose we had three countries arranged in a triangle, each with its own paper currency. Suppose there is a big lake in the middle so that there is no triple junction. There are exchange rates across each pair of borders. If these rates are chosen poorly, an enterprising person can generate wealth by going around in a closed loop (in the correct direction) exchanging currency at each border crossing.

The fact that the choice of currency is arbitrary (12,000 Zoobels = 5 Zinkys) is a gauge redundancy. The value one generates by going in a loop is like magnetic flux. The further point is that the value of a given unit of paper currency is arbitrary, so currency exchanges are like gauge transformations. If instead of using paper currency, the countries used something with intrinsic value (like say baked goods) it would be locally obvious if someone were doing something dumb. The intrinsic value of baked goods (at least if everyone agreed about it) plays the role of a Higgs field. The price of baked goods is not invariant under currency revaluation – this is the statement that the Higgs field is not gauge invariant. But the local obviousness of whether you are getting a good deal on donuts allows us to fix a gauge; this is like a mass for the gauge field. Maldacena has a nice appendix where he makes the discussion quite concrete.

4.5 Actions for gauge fields

The Yang-Mills (YM) action (4.9) is a gauge invariant and Lorentz invariant local functional of A. If the gauge field is to appear in $D = \partial + A$ it must have the same dimension as ∂, so $\mathcal{L}_{YM}$ has naive scaling dimension 4, like the Maxwell term, so $[g^2] = 4 - D$ – it is marginal in $D = 4$. Notice that unlike the Maxwell term, $\mathcal{L}_{YM}$ is not quadratic in A: it contains cubic and quartic terms in A, whose form is determined by the gauge algebra f_{ABC}. Non-abelian gauge fields interact with themselves in a very definite way. This interaction is relevant in $D < 4$ and irrelevant in $D > 4$.

In even spacetime dimensions, another gauge invariant, Lorentz invariant local functional of A is the total-derivative term $S_\theta = \theta \int \mathrm{tr} \frac{F}{2\pi} \wedge ... \wedge \frac{F}{2\pi}$ with $D/2$ factors of F. The D-form $\mathrm{tr} \frac{F}{2\pi} \wedge ... \wedge \frac{F}{2\pi} = d\omega(A)$ is exact (*e.g.* in $D = 4$, in the abelian case,

$F \wedge F = d(A \wedge F)$, or in components, $\epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} = 2\partial_\mu \left(\epsilon^{\mu\nu\rho\sigma} A_\nu F_{\rho\sigma}\right)$; more generally, ω is the Chern-Simons form, more below). Because it is exact, this doesn't affect the equations of motion or perturbation theory, but it does matter non-perturbatively. We'll see (when we study anomalies) that for smooth gauge field configurations in a closed spacetime, this functional is an integer. This means that θ is a periodic variable, since it appears in the partition function only as $e^{\mathbf{i}\theta n}$ with $n \in \mathbb{Z}$. This coupling θ violates CP symmetry (notice that $F \wedge F$ has one time derivative and three spatial derivatives) for $\theta \neq 0, \pi$. In QCD, this coupling of the gluons is constrained to be very small because it would give an electric dipole moment to the neutron, which the neutron doesn't seem to have; this mystery is called the strong CP problem.

In odd spacetime dimensions, we should consider the Chern-Simons (CS) term (the $D = 2 + 1$ version of which we just encountered) which in the abelian case looks like $S_{\text{CS}}[A] \overset{\text{abelian}}{=} \int A \wedge \frac{F}{2\pi} \wedge ... \wedge \frac{F}{2\pi}$ with $(D-1)/2$ factors of F. (In the non-Abelian case, there is an extra term to make its integral gauge invariant: in 3d, $S_{CS}[A] \propto \int \text{tr}\left(A \wedge F + \frac{2}{3} A \wedge A \wedge A\right)$.) This term *does* affect the equations of motion. It breaks parity symmetry. Notice that in $D = 2 + 1$ it is more relevant than the Maxwell or Yang-Mills term. Notice that the CS term is dangerous in the sense that the integrand is not gauge invariant; this suggests that something interesting will happen if we put such a thing on a space with boundary. It is a central ingredient in quantum Hall physics in $D = 2 + 1$, where, for one thing, it gives the gauge field fluctuations a mass (via a mechanism different from the Anderson-Higgs mechanism).

[End of Lecture 15]

In general dimension, we can make more couplings out of just F if we take more derivatives, but they will have higher dimension, *i.e.* their coefficients will be inverse powers of some energy scale, and so we can ignore them at low energy.

We can couple YM gauge fields to matter by returning to our starting point: *e.g.* if $\psi(x) \mapsto \Lambda_R \psi(x)$ is a Dirac field transforming in some representation R of the gauge group, then $D_\mu \psi = \left(\partial_\mu - \mathbf{i} T_R^A A_\mu^A\right) \psi$ also transforms in representation R, so

$$\bar{\psi} \mathbf{i} \gamma^\mu D_\mu \psi + V(\bar{\psi}\psi)$$

is a gauge-invariant lagrangian density. The lowest-dimension couplings of A to matter are determined by the representation matrices T_R^A, which generalize the electric charge.

I described the case where the gauge group is $\mathsf{SU}(N)$ above, but this construction gives a sensible QFT for any finite-dimensional compact Lie group. QCD is the case where $G = \mathsf{SU}(3)$ and the quarks are in the fundamental **3** representation, i.e. the T_R^A matrices are the ones I described above, just a collection of (eight) 3×3 hermitian and traceless matrices, satisfying the orthonormality condition $\text{tr} T^A T^B = \frac{1}{2}\delta^{AB}$. A

particular basis for them is called the Gell-Mann matrices:

$$\lambda_1 = X_{12} = \begin{pmatrix} & 1 & \\ 1 & & \\ & & \end{pmatrix}, \quad \lambda_2 = Y_{12} = \begin{pmatrix} & -\mathbf{i} & \\ \mathbf{i} & & \\ & & \end{pmatrix}, \quad \lambda_3 = Z_{12} = \begin{pmatrix} 1 & & \\ & -1 & \\ & & \end{pmatrix}, \tag{4.16}$$

$$\lambda_4 = X_{13} = \begin{pmatrix} & & 1 \\ & & \\ 1 & & \end{pmatrix}, \quad \lambda_5 = Y_{13} = \begin{pmatrix} & & -\mathbf{i} \\ & & \\ \mathbf{i} & & \end{pmatrix},$$

$$\lambda_6 = X_{23} = \begin{pmatrix} & & \\ & & 1 \\ & 1 & \end{pmatrix}, \quad \lambda_7 = Y_{23} = \begin{pmatrix} & & \\ & & -\mathbf{i} \\ & \mathbf{i} & \end{pmatrix}, \quad \lambda_8 = (Z_{13} + Z_{23})/\sqrt{3} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & -2 \end{pmatrix}/\sqrt{3} \ .$$

Notice that two of the eight (in blue) are diagonal in this basis. This is the most that can be simultaneously diagonalized, this number is called the *rank* of the Lie algebra.

You might expect that we would start doing perturbation theory in g now. There is lots of physics there, but it takes a little while to get there. Instead let's first think about how we might define the thing non-perturbatively and see what we learn from that.

4.6 Lattice gauge theory

We've seen that scalar field theory can emerge from a lattice model, such as the model of balls and springs with which we began this quarter. It can also emerge from a spin system, with a two-state system at each site of a lattice. Such a lattice model is then a non-perturbative unitary regulator of the QFT in question, and comes with many advantages. One is that it means the QFT can arise as a low-energy description of suitably-arranged condensed matter. Another is that the QFT can be simulated in various ways. Depending on the signs appearing in the action this may involve Monte Carlo sampling of the path integral.

The following beautiful construction was found by Wegner and Wilson and Polyakov; a good review is this one by Kogut.

Consider discretizing euclidean spacetime into a hypercubic lattice (for simplicity). On each *link* xy of the lattice we place a G-valued matrix $U_{xy}^{\alpha\beta}$. We define $U_{yx} = U_{xy}^{-1}$, as we did for the comparator in (4.13). Three good examples to keep in mind (in decreasing order of difficulty) are:

1. $\mathsf{G} = \mathsf{U}(N)$, in which case each U is a complex $N \times N$ matrix with $UU^\dagger = 1$. Here $\alpha, \beta = 1..N$.

2. $\mathsf{G} = \mathsf{U}(1)$, in which case U is just a phase (a 1×1 matrix) $U_{xy} = e^{\mathbf{i}\theta_{xy}}$, $\theta_{xy} \in [0, 2\pi)$.

3. $\mathsf{G} = \mathbb{Z}_n$, in which case $U = e^{2\pi\mathbf{i}\ell/n}, \ell = 1, \cdots n$, is a phase with $U^n = 1$. For $n = 2$, this is a classical spin living on the links of the lattice.

Please think of $U_{xy} = \mathcal{P}e^{\mathbf{i}\int_x^y A_\mu(r)dr^\mu}$ as the comparator (or Wilson line) along the link (except that there is no such degree of freedom as $A_\mu(r)$ at other values of r). As such, we impose the gauge equivalence relation $U_{xy} \mapsto g_x^\dagger U_{xy} g_y$, where $g_x \in \mathsf{G}$ for each x. We will accomplish this by two steps: by writing an action $S[U]$ that has this invariance, and by integrating over $\{U\}$ with an invariant measure:

$$Z = \int \prod_\ell dU_\ell e^{-S[U]}.$$

Here $\int dU$ is the G-invariant (Haar) measure on G, which can be defined by the desiderata

$$\int_{\mathsf{G}} dU = 1, \quad \int_{\mathsf{G}} dU f(U) = \int_{\mathsf{G}} dU f(VU) = \int_{\mathsf{G}} dU f(UV), \forall V \in \mathsf{G}\ .$$

For $\mathsf{G} = \mathsf{U}(1)$, the Haar measure is just $\int_0^{2\pi} đ\varphi$; for $\mathsf{G} = \mathbb{Z}_n$, it just $\frac{1}{n}\sum_{\ell=1}^n$. You can figure out what it is for $\mathsf{SU}(2)$ (locally, it's the round measure on S^3). Notice the following lovely advantage of these conditions: there is no need to gauge fix anything. The existence of the Haar measure is where the compactness of G is important.

This is a statistical mechanics problem of the equilibrium thermodynamics of a bunch of classical rotors (slightly fancy ones in the $\mathsf{SU}(N)$ case). The review by Kogut does a great job of highlighting the fact that this class of problems is susceptible to all the tools of statistical mechanics.

What action should we use? Here is a good way to make something invariant under the gauge group: Consider the comparator for a *closed* path C_{xx} which starts at x and ends at x:

$$W(C_{xx}) = \mathcal{P}e^{-\mathbf{i}\int_{C_{xx}} A}.$$

How does this transform? $W(C_{xx}) \mapsto g_x^{-1} W(C_{xx}) g_x$, but, for non-abelian G, it's still a matrix! A gauge-*in*variant object is

$$W(C) \equiv \mathrm{tr} W(C_{xx}) = \mathrm{tr}\mathcal{P}e^{-\mathbf{i}\int_{C_{xx}} A}$$

where the g_x and g_x^{-1} can eat each other by cylicity of the trace. The action should be local. We can make something gauge invariant and as local as possible by considering a path C that goes around a single plaquette of the lattice: $C = \partial\square$. This is Wilson's action:

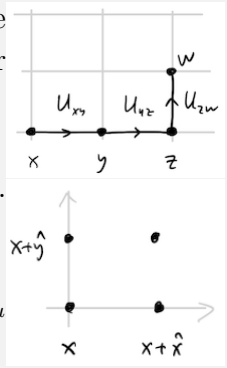
$$S[U] = \frac{1}{2f^2}\sum_\square \mathrm{Re}\, S_\square, \quad S_\square \equiv W(\partial\square) = \mathrm{tr}\prod_{\ell\in\partial\square} U_\ell = \mathrm{tr}\left(U_{x,x+dx}U_{x+dx,x+dx+dy}U_{x+dx+dy,x+dy}U_{x+dy,x}\right).$$

Now let's focus on the nonabelian case, such as $\mathsf{G} = \mathsf{SU}(N)$, and take seriously the idea that $U_{x,x+dx} = e^{-\mathbf{i}\int_x^{x+dx} A_\mu dx^\mu}$, where $A_\mu(x)$ is an element of the Lie algebra $\mathsf{g} = \mathsf{su}(N)$. An application of the CBH[43] formula $e^{sA}e^{sB} = e^{sA+sB+\frac{s^2}{2}[A,B]+\mathcal{O}(s^3)}$ shows that for a plaquette oriented in the $\mu\nu$ plane $\square_{\mu\nu}$, with lattice spacing a,

$$\begin{aligned}
\mathrm{Re}\, S_{\square_{\mu\nu}} &\overset{\text{CBH}}{=} \frac{1}{2f^2}\mathrm{Re}\,\mathrm{tr}\left(e^{-\mathbf{i}a^2 F_{\mu\nu}} + \mathcal{O}(a^3)\right) \\
&= \frac{1}{2f^2}\mathrm{Re}\,\mathrm{tr}\left(\mathbb{1} - \mathbf{i}a^2 F_{\mu\nu} - \frac{1}{2}a^4 F_{\mu\nu}F_{\mu\nu} + \mathcal{O}(a^5)\right) \qquad \text{(no sum on } \mu,\nu) \\
&= \frac{1}{2f^2}\left(\mathrm{tr}\mathbb{1} - \frac{a^4}{2}\mathrm{tr}F_{\mu\nu}F_{\mu\nu} + \ldots\right) + \mathcal{O}(a^5) \\
&= \mathcal{L}_{YM}(\square) + \text{const} + \mathcal{O}(a^5).
\end{aligned}$$

In the last step the term linear in $F_{\mu\nu}$ dropped out, either because $\mathrm{tr}F = 0$ (in the non-abelian case) or because it is pure imaginary and drops out of the real part. The coupling g is related to f in some way that can be figured out. So it is plausible that this model has a continuum limit governed by the Yang-Mills action. Realizing this possibility requires that the model defined by Z have a correlation length much larger than the lattice spacing, which is a physics question.

Before examining the partition sum, you might ask how would we add charged matter?
If we place fundamentals $q_x \mapsto g_x q_x$ at each site, we can make gauge invariants of the form (for example) $q_x^\dagger U_{xy}U_{yz}U_{zw}q_w$, or most simply, we can make a kinetic term for q by



$$S_{\text{kin}}[q,U] = \frac{1}{a^\#}\sum_{x,\hat{\ell}} q_x^\dagger U_{x,x+\hat{\ell}} q_{x+\hat{\ell}} \simeq \int d^D x\, \bar{q}(x)\left(\not{D} - m\right) q(x) + \ldots$$

where $\hat{\ell}$ runs over the lattice generators, and where $D_\mu = \partial_\mu - \mathbf{i}A_\mu$ is the covariant derivative, and we used its definition (4.12).
As you can see from the last expression, the expression I've written is for a fermionic field; in the euclidean path integral this is accomplished by making q_x a *Grassmann variable*. I'll explain what this is in the next subsection. For bosonic fields, where q_x is an ordinary number, the second-order terms are the leading terms which aren't a total derivative. There is some drama about the number of components of the spinor field one gets. It is not hard to get a massive Dirac fermion charged under a $\mathsf{U}(1)$ gauge field, like in QED. It is impossible to get a chiral spectrum, like a single Weyl fermion, from a gaussian, local lattice action; this is called the Nielsen-Ninomiya theorem. You might think 'oh that's

[43] Charlie-Baker-Hotel? Campbell-Baker-Hausdorff.

not a problem, because in the Standard Model there is the same number of L and R Weyl fermions,' but it is still a problem because they carry different representations under the electroweak gauge group. The word 'gaussian' is a real loophole, but not an easy one.

How do we get physics from the lattice gauge theory path integral Z? We need to find some gauge-invariant observables (since anything we stick in the integrand that isn't gauge-invariant will average to zero – this is sometimes called 'Elitzur's theorem'). In the pure YM theory, a good one is our friend the Wilson loop $W(C) = \mathrm{tr}\left(\prod_{\ell\in C} U_\ell\right) \simeq \mathrm{tr}\mathcal{P}e^{\mathbf{i}\oint_C A}$. What physics does its expectation values encode? Recall what happened in abelian gauge theory when we added an external source to measure the force mediated by various fields, for example in the Maxwell theory:

$$\lim_{T\to\infty} Z^{-1}\int DA\; e^{\mathbf{i}S_{\mathrm{Maxwell}}[A]+\mathbf{i}\int A_\mu J^\mu} = e^{-\mathbf{i}V(R)T}.$$

Here we took $J^\mu(x) = \eta^{\mu 0}\left(\delta^d(\vec{x}) - \delta^d(\vec{x}-(R,0,0))\right)$ for t in an interval of duration T, and zero before and after, two charges are held at distance R for a time T. $V(R)$ is the energy of the resulting configuration of (here, electromagnetic) fields, *i.e.* the Coulomb potential. If instead we let the charge and anticharge annihilate at $t = 0$ and $t = T$, this is a single charge moving along a rectangular loop $C_{R\times T}$ in spacetime, with sides R and T, and the result is just the expectation value of the associated Wilson loop. Going back to Euclidean spacetime, this is

$$\langle W(C_{R\times T})\rangle = Z^{-1}\int\prod dU\; e^{-\frac{1}{2f^2}\sum_\square \mathrm{Re}\,S_\square} W(C_{R\times T}) \overset{T\gg R}{\simeq} e^{-V(R)T},$$

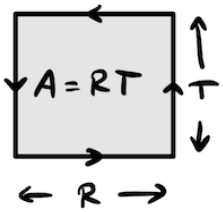
where the LHS is the expectation value of a gauge invariant operator. There can be some funny business associated with the corners and the spacelike segments, and this is the reason that we look for the bit of the free energy that is extensive in T.

In the case of the Maxwell theory in the continuum, this is a gaussian integral, which we can do (see the homework), and $\log\left\langle e^{\mathbf{i}\oint_{C_{R\times T}} A}\right\rangle \simeq -E(R)T - f(T)R$ with $E(R)\sim\frac{1}{R}$, goes something like the *perimeter* of the loop C. In the case of a short-ranged interaction, such as from a massive gauge field, the perimeter law would be more literally satisfied.

In contrast, a confining force between the charges would obtain if $\langle W(C_{R\times T})\rangle \overset{T\gg R}{\simeq} e^{-V(R)T}$ with instead

$$V(R) = \sigma R \quad\Longrightarrow\; F = -\frac{\partial V}{\partial R} = -\sigma\ .$$

This is a distance-independent attractive force between the charges. In this case $\log\langle W\rangle \sim RT$ goes like the *area* of the (inside of the) loop, so confinement is associated with an area law for Wilson loops. A constant force means a linear potential, so it is as if the charges are connected by a string of constant tension (energy per unit length) σ. This means only neutral ('colorless') objects have finite energy, confinement.

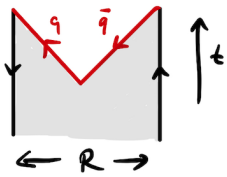


A small warning about the area law: with charged matter, the existence of an area law may depend on the representation in which we put the external charges:

$$W(C,R) = \mathrm{tr}_R \mathcal{P} e^{\mathbf{i}\oint_C A^A T_R^A}$$

where T_R^A are the generators of G in some representation R; this is the phase associated with a (very heavy and hence non-dynamical) particle in representation R. For some choices of R, it might be possible and energetically favorable for the vacuum to pop out *dynamical* charges which then screen the force between the two external charges (by forming singlets with them).

This happens if the energy stored in the string $\sim \sigma R \gtrsim 2M_q$, the threshold for producing the dynamical charges. That is, the string connecting the external sources may be able to break. $\mathsf{G} = \mathsf{SU}(N)$ has a center $\mathbb{Z}_N \subset \mathsf{SU}(N)$ under which the adjoint is neutral, so a Wilson loop in a representation carrying $\mathbb{Z}_N$ charge (such as the fundamental, in which it acts by $\mathbb{Z}_N$ phases times the identity) cannot be screened by pure glue. QCD, which has dynamical fundamentals, is more subtle.



This point, however, motivates the study of the dynamics of lattice gauge theories to address the present question: Where might such an area law come from? I'll give some hints for how to think about it.

Hint 0: Monte Carlo simulations. By now, Monte Carlo simulation of lattice gauge theory, and in particular actual QCD, is a giant industry. Its practitioners are able to calculate the spectrum of hadrons to a few percent, with good agreement with experiments. I'll explain below how masses of hadrons are extracted from such a theory.

[End of Lecture 16]

Hint 1: monopole condensation and dual Meissner effect. [Banks' book has a very nice discussion of this. I actually tricked myself into explaining most of this story earlier when we were talking about superconductors.] Recall that a single magnetic monopole is not a finite energy situation inside an infinite superconductor, because it has a tensionful Abrikosov flux string attached to it. A monopole and an antimonopole are linearly confined, with a constant force equal to the string tension.

On the other hand, electric-magnetic duality is a familiar invariance of Maxwell's equations:

$$\partial^\mu F_{\mu\nu} = J^{(e)}_\nu, \partial^\mu \tilde{F}_{\mu\nu} = J^{(m)}_\nu \tag{4.17}$$

is invariant under the replacements

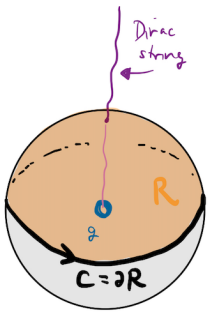
$$F_{\mu\nu} \to \tilde{F}_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}, \quad J^{(e)}_\nu \to J^{(m)}_\nu.$$

In doing a weak-coupling expansion (*e.g.* as we did in QED), we make a choice (having not seen magnetic charges, they must be heavy) to *solve* the second equation of (4.17) by introducing a smooth vector potential A_μ via

$$F_{\mu\nu}(x) = \partial_\mu A_\nu - \partial_\nu A_\mu + \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\int d^4y J^{(m)}(y)^\sigma f^\rho(x-y)$$

with $\partial_\rho f^\rho(x) = \delta^4(x)$. Here we are treating the magnetic sources as fixed, *e.g.* because they are heavy. The support of the function f^ρ is called the *Dirac string*.

In this description, a monopole is placed at the end of a long and infinitely thin solenoid, which carries away its magnetic flux $\int_{\text{sphere around monopole}} B = \int_{\text{cross-section of solenoid}} B = g$, and is invisible classically. Quantumly, it could be detected by Aharonov-Bohm effect of a charged particle going around it $e^{\mathbf{i}e\oint A} = e^{\mathbf{i}e\int B} = e^{\mathbf{i}eg}$ unless $eg \in 2\pi\mathbb{Z}$, Dirac quantization again. (For particles with both electric and magnetic charge (they are called dyons), the condition is $q_1m_2 - q_2m_1 \in 2\pi\mathbb{Z}$.)



So, the duality interchanges electric and magnetic things. If condensation of electric charge (meaning $\langle|\Phi|\rangle = v$ for some electrically charged field Φ) means that A_μ is massive (Anderson-Higgs effect) and that monopoles are confined by tensionful magnetic flux tubes, then we can just replace the relevant words to learn the following: Condensation of magnetic charge $\langle|\Phi_m|\rangle \neq 0$ means that some dual photon ($\tilde{A}_\mu$ with $d\tilde{A} = \tilde{F}$) is massive, and that *electric* charges are linearly confined by tensionful *electric* flux tubes.

This was pointed out by Mandelstam and 't Hooft in 1974. In 1994 Seiberg and Witten (hep-th/9407087) showed in detail that confinement actually happens by condensation of magnetic charge in a highly supersymmetric example. In abelian lattice models, we can actually implement the duality transformation explicitly by various path integral tricks. One path through this story (found in 1978 by Banks, Myerson, Kogut and also Peskin) is described in Banks' book. Along the way, one encounters dualities with many familiar statistical mechanical models, such as the XY model. I hope we will come back to this next quarter.

Hint 2: Strong coupling expansion. In thinking about an integral of the form

$$\int DU\ e^{\beta \sum_\square S_\square} W(C)$$

it is hard to resist trying to expand the exponential in β.

Unlike the perturbation series we've been talking about for months, this series has a finite radius of convergence. To understand this, it is useful to recognize that this expansion is structurally identical to the high-temperature expansion of a thermal partition function. (This is why I've used the symbol β for the coupling constant in front of the action.) For each configuration C, the function $e^{-\beta h(C)}$ is analytic in β about $\beta = 0$ (notice that $e^{-\frac{1}{T}}$ is analytic about $T = \infty$!). The only way to get a singularity at $\beta = 0$ would be if the sum over configurations (in the thermodynamic limit) did it; this would be a phase transition at $T = \infty$; that doesn't happen because the correlation length inevitably goes to zero at $T = \infty$: every site is so busy being buffeted by thermal fluctuations that it doesn't care about the other sites at all.[44]

In the non-abelian case, we get to do all kinds of fun stuff with characters of the group. For simplicity, let's focus on an abelian example, which will have a similar structure (though different large-β (weak coupling) physics). So take $U_\ell = e^{\mathbf{i}\theta_\ell} \in \mathsf{U}(1)$, in which case

$$S_{\square_{\mu\nu}}[U] = -\left(1 - \cos\theta_{\mu\nu}\right),\ \ \theta_{\mu\nu}(x) = \theta_\mu(x+\nu) - \theta_\mu(x) - \theta_\nu(x+\mu) + \theta_\nu(x) \equiv \Delta_\nu\theta_\mu - \Delta_\mu\theta_\nu(x).$$

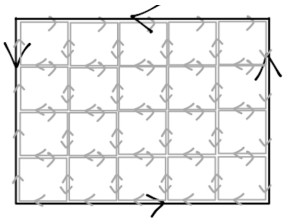
First let's consider the case where the world is a single plaquette. Then, using the identity $\int_0^{2\pi} đ\theta\ e^{\mathbf{i}n\theta} = \delta_{n,0}$,

$$\begin{aligned}\langle W(\square)\rangle &= \int \prod_\ell dU_\ell\ \underbrace{U_1U_2U_3U_4}_{\equiv S_\square}\left(1 + \beta(S_\square + S_\square^\dagger) + \frac{1}{2}\beta^2\left(S + S^\dagger\right)^2 + \frac{1}{3!}\beta^3\left(S + S^\dagger\right)^3 + \cdots\right) \\ &= \beta\underbrace{\langle S_\square S_{-\square}\rangle}_{=1} + \frac{\beta^3}{2}\langle S_{2\square}S_{-2\square}\rangle + \mathcal{O}(\beta^5) = \beta^{A(\square)}\left(1 + \mathcal{O}(\beta^2)\right) = e^{-\sigma(\beta)\text{Area}} \end{aligned} \tag{4.18}$$

with $\sigma(\beta) = |\ln\beta|$ is the string tension in this crude approximation. Here the area of the loop was just 1. I've written $S_{-\square} \equiv S_\square^\dagger$ and $S_{2\square} = S_\square^2$, the latter of which is only true in abelian cases.

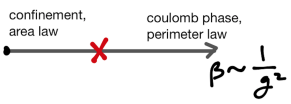
[44] For a much more formal and, I think, less illuminating proof, see for example J-M Drouffe and J-B Zuber, Physics Reports **102** (1983) section 3.1.2. Thanks to Tarun Grover for framing the above argument.

If instead we consider a model on a larger lattice (say $D = 2$ for now) and a loop that encloses many plaquettes, we must pull down at least one factor of $\beta S_\square^\dagger$ for each plaquette inside the loop, in order to cancel the link factors in the integrand. We can get more factors of beta if we pull down more cancelling pairs of $\beta^n S_\square^n S_{-\square}^n$, but these terms are subleading at small β. The leading contribution is $\langle W(C)\rangle = e^{-\sigma(\beta)\text{Area}}(1 + \mathcal{O}(\beta^2))$, where 'Area' denotes the area bounded by the curve C: an area law.



This area-law behavior of the Wilson loop in the strong-coupling expansion is the analog of the fact that in a spin system like the Ising model at high temperature, the correlation functions are short-ranged, $\langle s_x s_0\rangle \sim e^{-|x|/\xi}$ – the correlations are short-ranged. Since the series converges, this conclusion can be made completely rigorous, and only charge-neutral excitations have finite energy in the thermodynamic limit. In what sense is confinement a mystery then? Well, a hint is that our argument applies equally well (and in fact the calculation we did was) for abelian gauge theory! But QED doesn't confine – we calculated the Wilson loop at weak coupling and found a perimeter law – what gives?

The answer is that, in Abelian gauge theory, there is a phase transition in between weak and strong coupling, so the weak coupling regime is not an analytic continuation of the strong coupling series answer. Ruling out this possibility in Yang-Mills theory would be lucrative.



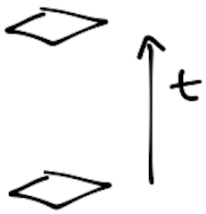
In fact, though, the Wilson loop expectation itself can exhibit a phase transition, even if other observables don't. I've drawn the pictures above as if the world were two-dimensional, in which case we just cover every plaquette inside the loop. In $D > 2$, we have to choose a surface whose boundary is the loop. Rather, $\langle W\rangle$ is a statistical sum over such surfaces, weighted by β^{area}:

$$\langle W(C)\rangle = \sum_{S|\partial S = C} \beta^{\text{Area}(S)}. \tag{4.19}$$

Such random-surface models often exhibit a *roughening transition* as β becomes larger, to a regime where floppy surfaces are not suppressed.

The same technology can be used to study the spectrum of excitations of the gauge theory, by considering correlations like

$$\left\langle S_R(t) S_R^\dagger(0) \right\rangle_c = \sum_\alpha |c_\alpha^R|^2 e^{-m_\alpha(R) t} \tag{4.20}$$



where S_R is the trace of a Wilson loop in representation R, around a single plaquette, and the two loops in question are separated only in time and are parallel. The subscript c means connected.
The right hand side is a sum over intermediate, gauge invariant states with the right quantum numbers, and $m_\alpha(R)$ are their masses. This is obtained by inserting a complete set of energy eigenstates.

In strong coupling expansion, we get a sum over discretized tubes of plaquettes, with one boundary at each loop (the connected condition prevents disconnected surfaces, and removes the contribution of the vacuum from the sum in (4.20)), the minimal number of plaquettes for a hypercubic lattice is $4t$, giving

$$\left\langle S_R(t) S_R^\dagger(0) \right\rangle_c \sim A\beta^{4t} \left(1 + \mathcal{O}(\beta^2)\right)$$

and the smallest glueball mass becomes $m_0 \sim 4|\ln\beta|$, similar to the scale of the string tension. Actually, the corrections exponentiate to give something of the form $m_0(R) = -4\ln\beta + \sum_k m_k(R)\beta^k$. (For more on that, see *e.g.* the monograph by Montvay and Münster called *Quantum Fields on a Lattice*.)

4.7 Fermion path integrals

We'll need these for our discussion of anomalies, and they are extremely useful for putting QCD (which differs from Yang-Mills theory by the addition of fermionic quarks) on the lattice, and for doing perturbative for QCD in the continuum, and even, perhaps surprisingly, for pure Yang-Mills theory.

I'm not going to derive the fermion path integral here. We can do that (by inserting a well-chosen resolution of the identity over and over) next quarter. Here I'm just going to motivate it, define it, and show that it does what we want.

[*e.g.* Schwartz §14.6] Canonical fermion operators satisfy anticommutation relations like $\{\psi(x), \bar{\psi}(y)\} = \mathbf{i}\hbar\delta^d(x-y)$. If we formally take $\hbar \to 0$, the Fermi fields are a bunch of objects that anticommute and square to zero. Such things are called Grassmann numbers

$$\theta_i\theta_j = -\theta_j\theta_i \ , \ i = 1..n$$

and the set of objects we get by multiplying and adding them (with coefficients in $\mathbb{C}$) is a Grassmann algebra. For $n = 1$, the most general element of the algebra is $g(\theta) = a + b\theta$. For $n = 2$, it is

$$g(\theta_1, \theta_2) = a + b\theta_1 + c\theta_2 + d\theta_1\theta_2. \tag{4.21}$$

A Grassmann algebra has an even part (made of products of even numbers of thetas, which therefore commute) and an odd part. I've named the object in (4.21) $g(\theta_1, \theta_2)$ as if it is a function of the Grassmann variables. This doesn't really mean anything, but if we go along with it, then (4.21) is actually Taylor's theorem for (two) Grassmann variables. The general case is very simple, there are only two terms in the expansion for each variable, 2^n terms altogether. A realization of Grassmann algebra that we've already seen is differential forms.

Integration is just as easy and in fact is 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:

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

So

$$\int d\psi_1 \cdots d\psi_n X = \partial_{\psi_1} \cdots \partial_{\psi_n} X.$$

Notice that there are no limits of integration. All Grassmann integrals are like the analog of

$$\int_{-\infty}^{\infty} dx f(x) = \int_{-\infty}^{\infty} dx f(x+a), \text{ if } \partial_x a = 0.$$

In fact the analogous condition is true:

$$\int (A + B\theta) d\theta = \int d\theta (A + B(\theta + \alpha)) \text{ if } \partial_\theta \alpha = 0.$$

The only integral, really, is the gaussian integral:

$$\int \underbrace{e^{-a\bar{\psi}\psi}}_{=1-a\bar{\psi}\psi} d\bar{\psi} d\psi = a.$$

[End of Lecture 17]

Many of these give

$$\int e^{-\bar{\psi}_i A_{ij} \cdot \psi_j} \prod_{i=1}^{M} d\bar{\psi}_i \prod_{i=1}^{M} d\psi_i = \int \prod_{i=1}^{M} d\bar{\psi}_i \prod_{i=1}^{M} d\psi_i \left(1 - \bar{\psi} A \psi + \frac{1}{2} \bar{\psi} A \psi \bar{\psi} A \psi + \cdots \right) \tag{4.22}$$

$$= \frac{1}{M!} \sum_{\text{perms},\sigma} (-1)^\sigma A_{1\sigma_1} A_{2\sigma_2} \cdots A_{M\sigma_M} \tag{4.23}$$

$$= \det A. \tag{4.24}$$

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

$$\int e^{-\bar{\psi} \cdot A \cdot \psi} d\bar{\psi} d\psi = \det A = e^{+\text{tr} \log A}$$

involves a sign in the exponent relative to the bosonic answer

$$\int e^{-\phi^\star \cdot A \cdot \phi} d\phi^\star d\phi = \frac{1}{\det A} = e^{-\text{tr} \log A}.$$

This is the same sign as the minus sign associated to fermion loops.

This inverse relation between boson integrals and fermion integrals is important in several places. Supersymmetry is a symmetry that relates bosons and fermions and in such theories the corresponding determinants often cancel; since the determinants are loop effects, this is an avatar of 'nonrenormalization theorems' of supersymmetric theories. The cancellation between bosonic and fermionic integrals is also used in the context of systems with disorder, where one wants

to average physical observables over some ensemble of space-dependent coupling constants. The factor of $1/Z$ in the normalization of correlation functions gets in the way of this (since Z depends on the couplings). Sometimes people use the *replica trick*, which involves studying n copies of the system and analytically continuing to $n \to 0$, which is as dicey as it sounds. Cancelling Z instead by introducing a combination of bosonic and fermionic modes is a method that is sometimes available and more reliable.
Here's the idea (due to Efetov). Imagine you are interested in computing the following Green's function of bosonic variables:

$$G^J_{ij}(t,t') = -\mathbf{i}\frac{1}{Z(J)}\int D\phi D\bar{\phi}\phi_i(t)\phi^\star_j(t')e^{\mathbf{i}S_J[\phi,\phi^\star]}, \tag{4.25}$$

with

$$S_J[\phi,\phi^\star] = \int dt\phi^\star_i D_{ij}\phi_j \equiv \int dt\phi^\star_i\left(\left(\mathbf{i}\partial_t+\mu\right)\delta_{ij} - J_{ij}\right)\phi_j \tag{4.26}$$

and

$$Z(J) = \int D\phi D\bar{\phi}e^{\mathbf{i}S_J[\phi,\phi^\star]} = \det D^{-1}. \tag{4.27}$$

And suppose we want to average G over some ensemble of couplings J, say with some gaussian distribution $P(J) \propto e^{-\sum_{ij}J^2_{ij}/(2\mathcal{J}^2)}$. The dependence of Z on J messes this all up. But now here comes the clever trick: rewrite

$$\frac{1}{Z(J)} = \det D = \int D\chi D\chi^\star e^{\mathbf{i}\int dt\chi^\star_i D_{ij}\chi_j} \tag{4.28}$$

where χ_i is a complex *grassmann* field. Now the whole integral over J in G is gaussian. See *e.g.* this paper for more discussion and references.

Correlation functions look like:

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

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

$$\left\langle\bar{\psi}_i\psi_j\right\rangle = \frac{\delta_{ij}}{a_i} \equiv \langle\bar{i}j\rangle \tag{4.29}$$

or, in a general basis,

$$\left\langle\bar{\psi}_i\psi_j\right\rangle = A^{-1}_{ij}.$$

Wick's theorem here is

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

With sources, the general gaussian integral is

$$\int e^{-\bar{\psi}_i A_{ij}\cdot\psi_j+\bar{\eta}_i\psi_i+\bar{\psi}_i\eta_i}\prod_{i=1}^{M}d\bar{\psi}_i\prod_{i=1}^{M}d\psi_i = e^{\bar{\eta}A^{-1}\eta}\int\prod d\bar{\psi}d\psi e^{-(\bar{\psi}-\bar{\eta}A^{-1})A(\theta-A^{-1}\eta)} = e^{\bar{\eta}A^{-1}\eta}\det A.$$

Now we can take a continuum limit: $\psi_i \rightsquigarrow \psi(x), f(\psi_i) \rightsquigarrow f[\psi]$. The (real-time) partition function for a free fermion field is

$$Z[\bar{\eta},\eta] = \int [D\bar{\psi}D\psi]e^{\mathbf{i}\int d^Dx\left(\bar{\psi}\left(\mathbf{i}\not{\partial}-m\right)\psi+\bar{\eta}\psi+\bar{\psi}\eta+\mathbf{i}\epsilon\bar{\psi}\psi\right)} \tag{4.30}$$

$$= \det\left(\mathbf{i}\not{\partial}-m\right)e^{\mathbf{i}\int d^Dx\int d^Dy\bar{\eta}(x)\left(\mathbf{i}\not{\partial}-m+\mathbf{i}\epsilon\right)^{-1}\eta(y)}. \tag{4.31}$$

You can check from this expression that we reproduce the Feynman propagator:

$$\left\langle\mathcal{T}\bar{\psi}(x)\psi(0)\right\rangle = \frac{\delta}{\delta\eta(x)}\frac{\delta}{\delta\bar{\eta}(0)}\log Z = \frac{1}{\mathbf{i}\not{\partial}-m+\mathbf{i}\epsilon}. \tag{4.32}$$

If we couple ψ minimally to a gauge field, the determinant (which here is an irrelevant constant) becomes an effective potential for the gauge field,

$$\det\left(\mathbf{i}\not{D}_A-m\right) = e^{\mathbf{i}S_{\text{eff}}[A]}. \tag{4.33}$$

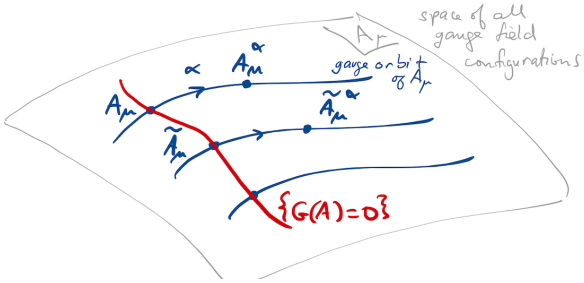
5 Non-abelian gauge fields in perturbation theory

5.1 Gauge fixing and Feynman rules

Gauge fixing. [Peskin §16.2, Schwartz §25] Consider the partition function for pure Yang-Mills theory (it will be easy to add quarks later):

$$Z \equiv \frac{1}{\mathrm{vol}(\mathcal{G})} \int [DA] e^{\mathbf{i}S[A]}.$$

We assume that $S[A]$ is some gauge-invariant functional of A, such as the Yang-Mills action. The integral over $[DA]$ goes over all configurations of the gauge field A. Here $\mathrm{vol}(\mathcal{G})$ is the volume of the gauge group – a copy of G for each point in space. We divide by it to cancel out the contributions from gauge-equivalent configurations of A. So the path integral is $\frac{\infty}{\infty}$. We would like to make this cancellation more explicit by fixing a gauge, $G(A(x)) = 0$. A cartoon of the space of all gauge field configurations, stratified into gauge orbits, is sketched at right; we want to restrict the integration to a gauge-fixed slice $G(A(x)) = 0$ (indicated in red) which intersects each orbit at one representative configuration. Perhaps surprisingly, this will be an application of fermion path integrals.



Note that the gauge-fixing function $G(A(x))$ must be a function of $A(x)$ (we assume it is a local condition) that is *not* gauge invariant, such as $\partial^\mu A_\mu(x)$. To implement this, we will insert the following form of the number 1:

$$1 = \Delta[A] \int [D\alpha] \delta[G(A^\alpha)] \tag{5.1}$$

where A^α denotes the gauge transformation of A by gauge parameter $\alpha(x)$. Here $\int [D\alpha]$ is the Haar measure on G and

$$A^\alpha_\mu = (A^\alpha_\mu)^a T^a = e^{\mathbf{i}\alpha\cdot T}\left(A_\mu + \frac{\mathbf{i}}{g}\partial_\mu\right) e^{-\mathbf{i}\alpha\cdot T} \tag{5.2}$$

$$= \left(A^a_\mu + \frac{1}{g}\partial_\mu\alpha^a + f^{abc}A^b_\mu\alpha^c\right) T^a + \mathcal{O}(\alpha^2) \tag{5.3}$$

$$= \left(A^a_\mu + \frac{1}{g}D_\mu\alpha^a\right) T^a + \mathcal{O}(\alpha^2) \tag{5.4}$$

is the image of A under a gauge transformation with parameter $\alpha^a(x)$ – at the last step D is the covariant derivative on a field in the adjoint representation. Here the indices $a, b, c = 1..\dim G$ run over the adjoint representation of G, ie the generators of the Lie algebra. The (Fadeev-Popov) determinant Δ defined by (5.1) is a Jacobian

$$\Delta[A] = \det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right) = \det_{xa,yb}\left(\frac{\delta G^a(A^\alpha(x))}{\delta\alpha^b(y)}\right). \tag{5.5}$$

A crucial property of Δ is that it is gauge invariant:

$$\Delta[A^{\alpha_1}]^{-1} \overset{(5.1)}{=} \int [D\alpha]\delta[G(A^{\alpha+\alpha_1})] = \int [D\left(\alpha+\alpha_1\right)]\delta[G(A^{\alpha+\alpha_1})] = \Delta[A]^{-1}. \tag{5.6}$$

(Here we used the fact that the measure on the group is invariant, $[D\alpha] = [D\left(\alpha+\alpha_1\right)]$. Forgive me for switching to a more-convenient additive notation for the group law here.)

Inserting this form of 1 (5.1) into Z (and changing the order of integration) gives:

$$Z = \frac{1}{\mathrm{vol}(\mathcal{G})}\int [D\alpha]\int [DA]\delta[G(A^\alpha)]\Delta[A]e^{\mathbf{i}S[A]} \tag{5.7}$$

$$= \frac{1}{\mathrm{vol}(\mathcal{G})}\int [D\alpha]\int [DA^\alpha]\delta[G(A^\alpha)]\Delta[A^\alpha]e^{\mathbf{i}S[A^\alpha]} \tag{5.8}$$

$$\overset{\tilde{A}\equiv A^\alpha}{=} \underbrace{\left(\frac{1}{\mathrm{vol}(\mathcal{G})}\int [D\alpha]\right)}_{=1}\int [D\tilde{A}]\delta[G(\tilde{A})]\Delta[\tilde{A}]e^{\mathbf{i}S[\tilde{A}]}\ . \tag{5.9}$$

In the first step we use the fact that $\int[DA] = \int[DA^\alpha], S[A] = S[A^\alpha]$ and (5.6). In the second step we change integration variables to $A^\alpha \equiv \tilde{A}$, and promptly drop the tilde. So we've cancelled the offending volume of the gauge group, and inserted a gauge-fixing delta function in the path integral.

The only price is the FP determinant Δ that we've acquired. What is it? It depends on the choice of gauge fixing function. Let's choose

$$G^a(A(x)) = \partial^\mu A^a_\mu(x) - \omega^a(x).$$

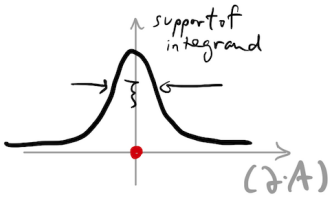
Rather than picking a particular ω, let's average over all possibilities with gaussian measure:

$$1 = N(\xi)\int [D\omega]e^{-\mathbf{i}\int d^Dx\frac{\omega^2(x)}{2\xi}}.$$

The normalization factor is just a constant which we can forget. Therefore

$$Z = N(\xi)\int [DA]\underbrace{\int [D\omega]\delta[\partial\cdot A-\omega]}_{=1} e^{-\mathbf{i}\int \frac{\omega^2}{2\xi}}\Delta[A]e^{\mathbf{i}S[A]} \tag{5.10}$$

$$= N(\xi)\int [DA]\Delta[A]e^{\mathbf{i}\left(S[A]-\int \frac{(\partial\cdot A)^2}{2\xi}\right)}. \tag{5.11}$$



Finally we must figure out what is $\Delta[A]$. Comparing to (5.4) (and remembering that there is a factor of $\delta[G]$ multiplying everything), Δ is the determinant of the operator

$$\frac{\delta G^a(A^\alpha(x))}{\delta\alpha(y)} = \frac{1}{g}\partial^\mu D_\mu \delta^{ab}\delta^D(x-y).$$

Notice that in the abelian case, this is independent of A (the covariant derivative D acting on the adjoint representation of $\mathsf{U}(1)$ is just ∂) and we can forget about it; that's why we didn't bother doing this for QED. In the nonabelian case,

$$\Delta = \det\left(\frac{1}{g}\partial^\mu D_\mu\right) = \int [DcD\bar{c}]e^{\mathbf{i}\int d^D x\bar{c}(-\partial^\mu D_\mu)c}.$$

At the last step we used the integration formula for gaussian grassmann integrals (and absorbed a factor of g into the definition of c, and ignored a constant factor). c here is a new complex scalar field in the theory ($\bar{c}\equiv c^\dagger$). Since D is the covariant derivative in the adjoint, it's a field in the adjoint of the gauge group. There's just one weird thing about it – it's a *fermionic* field with second-order kinetic terms, a ghost!

There are all kinds of bad things about fermions with second-order kinetic terms (see the discussion of the connection between spin and statistics around p. 138 here, or Schwartz §12). But those bad things only happen if the particles occur in external states. One purpose of a lot of the fancy stuff on this subject (such as BRST symmetry) is guaranteeing that we'll never make ghost particles while scattering the real particles. The loops of the ghosts, though, are crucial for getting correct and unitary answers. In particular, the optical theorem relates scattering states to particles appearing in loops. The contributions to the imaginary part of loops from the ghosts are required to cancel the unitarity-violating contributions from the unphysical polarization states of the gluons. (For the details of what is being cancelled see Peskin pp. 508-511, and for the cancellation itself, see 515-516, and a bit below.)

Altogether,

$$Z = \int [DADcD\bar{c}]e^{\mathbf{i}\left(S[A]-\int \frac{(\partial\cdot A)^2}{2\xi}+\int \bar{c}(-\partial^\mu D_\mu)c\right)}. \tag{5.12}$$

More generally, the ghosts are *negative degrees of freedom* whose role in life is to *cancel* the unphysical contributions of the timelike and longitudinal components of the gluon field. For example, in the free ($g = 0$) path integral in Feynman gauge, we have

$$Z(g=0) = \left(\det\left(-\partial^2\right)^{-D} \det\left(-\partial^2\right)^{+2}\right)^{\dim(\mathsf{G})/2} .$$

The contribution of the ghosts cancels two components' worth of the contribution from the gluons, leaving behind just the contributions of the $D-2$ transverse polarizations as we would hope. A physical application is getting the correct coefficient from two polarizations in the Stefan-Boltzmann law $F(T) = cL^3T^4$ when we compute this ratio of determinants on a thermal circle of radius $1/T$.

Feynman rules. More explicitly, the ghost action is (after some IBP)

$$\mathcal{L}_{\text{ghost}} = \bar{c}^a \left(-\partial^2 \delta^{ab} + g \overleftarrow{\partial}^\mu f^{abc} A^c_\mu\right) c^b.$$

The ghost propagator is then:

$$\left\langle c^a(x) \bar{c}^b(y) \right\rangle = \int \mathrm{đ}^4 k \; e^{-\mathbf{i}k(x-y)} \frac{\mathbf{i}}{k^2 + \mathbf{i}\epsilon} \delta^{ab}.$$

Let us use the lovely Yang-Mills action, $S[A] = \int \mathcal{L}_{YM}$

$$\mathcal{L}_{YM} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu}, \qquad F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu.$$

The resulting gluon propagator is

$$\left\langle A^a_\mu(x) A^b_\nu(y) \right\rangle = \int \mathrm{đ}^4 k \; e^{-\mathbf{i}k(x-y)} \frac{-\mathbf{i}}{k^2 + \mathbf{i}\epsilon} \delta^{ab} \left(\eta_{\mu\nu} - (1-\xi)\frac{k_\mu k_\nu}{k^2}\right)$$

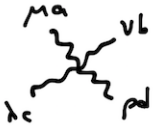
The gluon propagator is just like the photon one, times a δ^{ab} which conserves the color. $\xi = 1$ is Feynman gauge, which I'll use everywhere below.

The new Feynman rules are

$= \dfrac{\mathbf{i}}{k^2},$

$= -g f^{abc} k^\mu,$


$= g f^{abc} \left(\eta_{\mu\nu}(k_1 - k_2)_\lambda + \eta_{\nu\lambda}(k_2 - k_3)_\mu + \eta_{\lambda\mu}(k_3 - k_1)_\nu\right),$


$= -\mathbf{i}g^2 \left(f^{abe} f^{cde} \left(\eta_{\mu\lambda}\eta_{\nu\rho} - \eta_{\mu\rho}\eta_{\nu\lambda}\right) + f^{ade} f^{cbe} \left(\eta_{\mu\lambda}\eta_{\nu\rho} - \eta_{\mu\nu}\eta_{\rho\lambda}\right) + f^{ace} f^{bde} \left(\eta_{\mu\nu}\eta_{\lambda\rho} - \eta_{\mu\rho}\eta_{\nu\lambda}\right)\right).$

(Patterns: in the cubic coupling, the three terms cyclically permute the color and kinematic indices. In the quartic coupling, the second term is obtained from the first by the interchange $(b,\nu) \leftrightarrow (d,\rho)$, and the third is obtained from the first by $(b,\nu) \leftrightarrow (c,\lambda)$.)

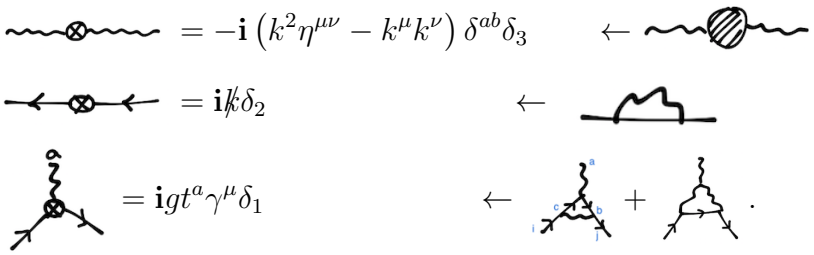
Including quarks doesn't mess with the gauge-fixing stuff. We'll take

$$\mathcal{L}_{\text{quarks}} = \bar{q}\left(\mathbf{i}\not{D} - m\right) q = \bar{q}_i \left(\gamma^\mu \left(\mathbf{i}\partial_\mu \delta_{ij} + g A^a_\mu t^a_{ij}\right) - m\delta_{ij}\right) q_j \tag{5.13}$$

Here i, j are color indices. For QCD, the color indices are $i, j = 1..3$ and and t^a_{ij} are the generators of $\mathsf{SU}(3)$ in the fundamental representation (a good basis of which are the Gell-Mann matrices (4.16)). In (5.13) I've suppressed the *flavor* indices – quarks come in various flavors, just like a muon is a different flavor of electron. I've also suppressed the coupling to electromagnetism, which is of the form $\bar{q} Q_q \gamma^\mu A_\mu q$, where $Q_q = 2/3$ or $-1/3$ for up- and down-type quarks respectively. Their fractional electric charge is one way in which we know we haven't seen free quarks.

With this addition, there's also a quark propagator that looks the same as an electron propagator, and the qqg vertex is  $= \mathbf{i} g \gamma^\mu t^a$. [End of Lecture 18]

We'll also need to add some counterterms

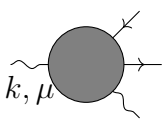


On the right, I've indicated which one-loop diagrams require us to add these respective counterterms.

5.2 Ghosts and unitarity

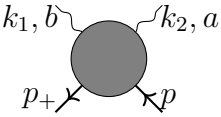
[Peskin §15.1,16.1, 16.3] A practical goal of maintaining gauge invariance everywhere above is to try to make a field theory with massless vector particles (if only in perturbation theory). Such particles have two transverse polarization states. It's a nontrivial thing to decouple the longitudinal modes of the gluons from physical processes.

Does it work? One concrete meaning of 'work' is that the Ward identity should hold. That is, in any amplitude with an external gauge boson,

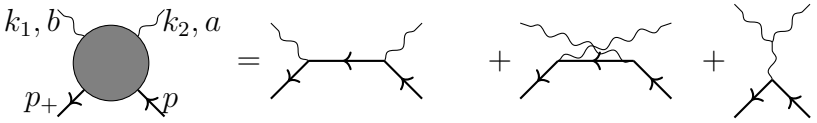
$$\mathbf{i}\epsilon^\mu(k)\mathcal{M}_\mu(k) = \tag{5.14}$$ 

then when all other *charged* particles are on-shell (notice that now this includes other gluons), $k^\mu \mathcal{M}_\mu(k) \stackrel{!}{=} 0$. For example, consider the process $q\bar{q} \to gg$, even at tree level (time is going up in these diagrams, and all momenta are pointed toward the future):

$$\mathbf{i}\mathcal{M} = \quad + \text{loops} \tag{5.15}$$



with



(5.16)

$$I + II = \mathbf{i}\epsilon^\star_\mu(k_1)\epsilon^\star_\nu(k_2)\mathcal{M}^{\mu\nu}_{12} \tag{5.17}$$

with

$$\mathbf{i}\mathcal{M}^{\mu\nu}_{12} = (\mathbf{i}g)^2 \bar{v}(p_+)\left(\gamma^\nu t^a \frac{\mathbf{i}}{\not{p} - \not{k}_2 - m}\gamma^\nu t^b + \gamma^\nu t^b \frac{\mathbf{i}}{\not{k}_2 - \not{p}_+ - m}\gamma^\mu t^a\right) u(p) \tag{5.18}$$

So let's see what happens if we replace $\epsilon^\star_\nu(k_2)$ with k_2^ν. As in the case of QED we'll use the EoM of the spinors: $(\not{p} - m)u(p) = 0, \bar{v}(p_+)\left(-\not{p}_+ - m\right) = 0$. The result is

$$\mathbf{i}\mathcal{M}^{\mu\nu}_{12}(k_2)_\nu = -g^2\bar{v}(p_+)\left(-\mathbf{i}\gamma^\mu[t^a, t^b]\right)u(p) = -g^2\bar{v}(p_+)\gamma^\mu u(p) f^{abc}t^c. \tag{5.19}$$

A nice check is that this vanishes in the abelian case where the diagram III is absent. More specifically, the result involves f^{abc}, which accounts for the color structure of the 3-gluon vertex in diagram III. That diagram is

$$III = \mathbf{i}\epsilon^\star_\mu(k_1)\epsilon^\star_\nu(k_2)\mathcal{M}^{\mu\nu}_3 \tag{5.20}$$

with

$$\mathbf{i}\mathcal{M}^{\mu\nu}_3 = \underbrace{\mathbf{i}g\bar{v}(p_+)\gamma_\rho t^c u(p)\frac{-\mathbf{i}}{k_3^2}gf^{abc}}_{\equiv \text{☼}_\rho}\left(\eta^{\mu\nu}(k_2 - k_1)^\rho + \eta^{\nu\rho}(k_3 - k_2)^\mu + \eta^{\rho\mu}(k_1 - k_3)^\nu\right). \tag{5.21}$$

where $k_3 \equiv -k_1 - k_2$. So replacing $\epsilon(k_2)$ with k_2 and eliminating $k_2 = -k_1 - k_3$ gives

$$(k_2)_\nu\, \mathbf{i}\mathcal{M}^{\mu\nu}_3 = \text{☼}_\rho\left(\eta^{\rho\mu}k_3^2 - k_3^\rho k_3^\mu - \eta^{\rho\mu}k_1^2 + k_1^\rho k_1^\mu\right). \tag{5.22}$$

The four terms here each have their own story. The third term is zero by the EoM for the spinors using $k_3 = p + p_+$. The second term vanishes because $k_1^2 = 0$ since the

other gluon is assumed to be on-shell. *IF* the other gluon were transverse, then the last term would vanish when this is contracted with $\epsilon(k_1)_\mu$. This would leave the first term, in which the k_3^2 cancels the denominator of the gluon propagator, leaving behind exactly the negative of the contribution from $I + II$.

We conclude that the coupling g in the 3-gluon vertex must be the same as the one in the coupling between quarks and gluons in order for this cancellation of the longitudinal bits to work. A similar argument for tree-level $gg \to gg$ implies that the coupling in the 4-gluon vertex must also be the same. (From our starting point, where g began its life in front of the Yang-Mills term, this equality is obvious, but we could wonder what happens if we violate it.)

But why should the other gluon be transverse?? Our goal here is to show that the non-tranvserse bits don't contribute to physics.

Let's back up a bit and think about a good basis of polarization vectors. Fix some lightlike k^μ, $k^2 = 0$. We can construct a basis satisfying

$$\epsilon_i^T \cdot k = 0,\ \epsilon_i^T \cdot \epsilon_j^{T\star} = \delta_{ij},\ \epsilon^\pm \cdot \epsilon_i^{T\star} = 0, \epsilon^+ \cdot \epsilon^- = 1,\ \epsilon^\pm \cdot \epsilon^\pm = 0 \tag{5.23}$$

where $\epsilon^T_{i=1,2}$ are transverse (physical) polarizations, and $\epsilon^\pm$ are longitudinal (unphysical) polarizations (which I choose to be real). If $k^\mu = (k^0, \vec{k})^\mu$, then we can take $\epsilon^\pm_\mu = \frac{1}{\sqrt{2}|\vec{k}|}\left(k^0, \pm\vec{k}\right)_\mu$. $\epsilon^\pm$ are called "forward" and "backward" longitudinal polarizations. We can resolve the Minkowski metric by

$$\eta_{\mu\nu} = -\sum_{i=1,2} \epsilon^T_{i\mu}\epsilon^{T\star}_{i\nu} + \epsilon^-_\mu\epsilon^+_\nu + \epsilon^+_\mu\epsilon^-_\nu. \tag{5.24}$$

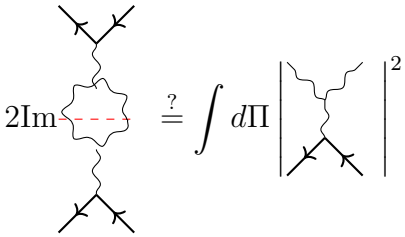
Here the first bit is the contribution from the transverse polarizations.

So let's revisit what happens if the other gluon is not transverse. In fact because of the last term in (5.22),

$$\begin{aligned} 0 \neq \mathbf{i}\mathcal{M}^{\mu\nu}_{12+3}\epsilon^-_\mu(k_1)\ \underbrace{\epsilon^+_\nu(k_2)}_{=\frac{1}{\sqrt{2}|\vec{k}_2|}k_{2\nu}} &\overset{(5.22)}{=} \mathbf{i}g\bar{v}(p_+)\gamma_\rho t^c u(p)\frac{-\mathbf{i}}{k_3^2}\epsilon^-_\mu(k_1)\frac{1}{\sqrt{2}|\vec{k}_2|}gf^{abc}\left(-k_1^\rho k_1^\mu\right) \\ &= -\mathbf{i}g^2\bar{v}(p_+)\gamma_\rho u(p)t^c f^{abc}\frac{-\mathbf{i}}{k_3^2}k_1^\rho\frac{|\vec{k}_1|}{|\vec{k}_2|}. \end{aligned} \tag{5.25}$$

What the heck? Maybe we can just ignore this final state? A place where we can't ignore it is in loops. Let's consider the contribution to the imaginary part of the amplitude for $gg \to gg$ coming from cutting an intermediate gluon bubble. The optical

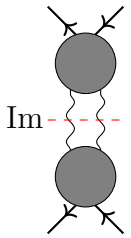
theorem would seem to say something like



(5.26)

where $d\Pi$ denotes the final-state phase space of two gluons.

More precisely, let's ask: what is the imaginary part of the contribution where there are two gluons in the intermediate state that go on-shell:



(5.27)

where the blobs are as above, *i.e.* one denotes $\mathbf{i}\mathcal{M}^{\mu\nu}_{12+3}$ above and the other denotes its channel dual that interchanges initial and final states, $\mathbf{i}\mathcal{M}'^{\mu\nu}_{12+3}$. The imaginary part from cutting the gluon propagators comes from replacing each one by

$$\frac{\mathbf{i}\eta_{\mu\rho}}{k^2} \rightsquigarrow -\mathbf{i}\eta_{\mu\rho}(-2\pi\mathbf{i})\delta(k^2)\ . \tag{5.28}$$

This gives

$$X = \int d\Pi \frac{1}{2}\mathbf{i}\mathcal{M}^{\mu\nu}\eta_{\mu\rho}\eta_{\nu\sigma}\mathbf{i}\mathcal{M}'^{\rho\sigma} \tag{5.29}$$

where the $\frac{1}{2}$ is in recognition of the identical nature of the two bosons, and $\mathcal{M}'$ is the channel-dual amplitude with the gluons in the initial state and the quarks in the final state. Now use (5.24) for both of the ηs in this expression. Using (5.23), all the bits mixing the transverse and longitudinal polarizations vanish (and $\mathcal{M}^{\mu\nu}\epsilon^{\pm}\epsilon^{\pm} = 0$, too), leaving

$$X = X_{\text{physical}} + \int d\Pi \frac{1}{2}\left(\mathbf{i}\mathcal{M}^{\mu\nu}\epsilon^-_\mu\epsilon^+_\nu\right)\left(\mathbf{i}\mathcal{M}'^{\mu\nu}\epsilon^+_\rho\epsilon^-_\sigma\right) + (+\leftrightarrow -)\,. \tag{5.30}$$

Now we use (5.25) (and a similar expression for $\mathcal{M}'$) to find

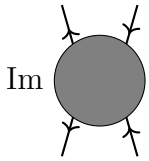
$$X = X_{\text{physical}} + \int d\Pi\ Y \tag{5.31}$$

with

$$\begin{aligned} Y &= \frac{1}{2}\left(\mathbf{i}g\bar{v}(p_+)\gamma_\mu t^c u(p)\frac{-\mathbf{i}}{(k_1+k_2)^2}(-gf^{abc}k_1^\mu)\right)\left(\mathbf{i}g\bar{v}(p')\gamma_\rho t^d v(p'_+)\frac{-\mathbf{i}}{(k_1+k_2)^2}(-gf^{abc}(-k_2)^\mu)\right) + (k_1\leftrightarrow k_2) \\ &\equiv AB. \end{aligned} \tag{5.32}$$

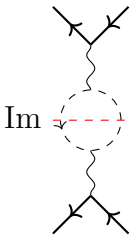
The $(k_1 \leftrightarrow k_2)$ term is equal to the first term by $k_1 + k_2 = p + p_+$ and the spinor EoM. The bad news is that extra term is not zero.

The good news is that there is another contribution to



(5.33)

namely the process with an intermediate ghost loop, whose cut is exactly the tree diagrams we studied above:



$$= - \int d\Pi \mathcal{M}_{q\bar{q}\to c\bar{c}} \mathcal{M}_{c\bar{c}\to q\bar{q}} \tag{5.34}$$

The $-$ is the fermion loop minus sign. But above we computed

$$\mathbf{i}\mathcal{M}_{q\bar{q}\to c\bar{c}} = \mathbf{i} g \bar{v}(p_+)\gamma_\mu t^c u(p) \frac{-\mathbf{i}}{(k_1+k_2)^2}\left(-g f^{abc} k_1^\mu\right) = A \tag{5.35}$$

$$\mathbf{i}\mathcal{M}_{c\bar{c}\to q\bar{q}} = \mathbf{i} g \bar{u}(p')\gamma_\rho t^d v(p'_+) \frac{-\mathbf{i}}{(k_1+k_2)^2}\left(-g f^{abd} (-k_2)^\mu\right) = B \ . \tag{5.36}$$

This exactly cancels the contributions of the unphysical polarizations, and the optical theorem holds with only the physical polarizations contributing.

BRST invariance. Let's revisit the gauge-fixed QED Lagrangian:

$$\mathcal{L}_{\rm FP}^{\rm abelian} = \mathcal{L}_0[A_\mu, \psi, \phi] - \frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2 - \bar{c}\Box c \tag{5.37}$$

where the first term is the kinetic terms for the photon and whatever charged or uncharged matter there is; its crucial property is that it is gauge invariant. The second term fixes the gauge invariance $A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha$ up to $\alpha(x)$ with

$$\Box \alpha = 0. \tag{5.38}$$

Now consider (don't ask why) a gauge transformation where the gauge parameter is

$$\alpha(x) = \theta c(x) \tag{5.39}$$

where $c(x)$ is the ghost field configuration and θ is a fixed grassmann variable, so that $\alpha(x)$ is a commuting object. Note that the EoM for the ghosts is $\Box c(x) = 0$, so this $\alpha(x)$ satisfies the condition (5.38) on-shell. $\mathcal{L}_0[A_\mu, \psi, \phi]$ is invariant under this, but

$$\left(\partial_\mu A^\mu\right)^2 \to \left(\partial_\mu A^\mu\right)^2 + \frac{2}{e}(\partial_\mu A^\mu)\theta\Box c + \frac{1}{e^2}(\theta\Box c)\,(\theta\Box c)\,. \tag{5.40}$$

The last term vanishes by $\theta^2 = 0$. We can cancel this variation by varying

$$\bar{c}(x) \to \bar{c}(x) - \frac{1}{e}\theta\frac{1}{\xi}\partial_\mu A^\mu(x). \tag{5.41}$$

Note that despite their names c and $\bar{c}$ are independent grassmann variables and need not transform the same way. This is called a BRST transformation. It is a residual redundancy of the gauge-fixed action.

Now let's see the analog for the nonabelian case. The reason to care about this is that the cancellation we've described above seems kind of delicate, and the FP ghosts are weird and make people nervous. Wouldn't it be great if there were some giant mathematical machine we could use that would guarantee that this works at all orders in every process?

$$\mathcal{L}_{\text{FP}} = \mathcal{L}_0[A_\mu, \psi, \phi] - \frac{1}{2\xi}\left(\partial^\mu A^a_\mu\right)^2 - \partial_\mu \bar{c}^a D_\mu c^a \tag{5.42}$$

where $D_\mu c^a = \partial_\mu c^a + g f^{abc} A^b_\mu c^c$. Consider the following transformation, again a gauge transformation parametrized by the c-ghost field:

$$\begin{aligned}
\psi_i &\to \psi_i + \mathbf{i}\theta c^a t^a_{ij}\psi_j \\
A^a_\mu &\to A^a_\mu + \frac{1}{g}\theta D_\mu c^a \\
\bar{c}^a &\to \bar{c}^a - \frac{1}{g}\theta\frac{1}{\xi}\partial_\mu A^a_\mu,\ c^a \to c^a + ?\ .
\end{aligned} \tag{5.43}$$

As above, L_0 is invariant since this is a gauge transformation. As above, the transformation of c cancels the variation of the gauge-fixing term $(\partial_\mu A^\mu)^2$. But now because of the extra variation of A, $D_\mu c^a$ gets a term added to it of the form $\theta f^{abc} D_\mu c^b c^c$. So if we take

$$c^a \to c^a - \frac{1}{2}\theta f^{abc} c^b c^c \tag{5.44}$$

then $\mathcal{L}_{\text{FP}}$ is invariant. Some comments: this is a supersymmetry transformation, in the sense that the parameter of the transformation, θ, is a grassmann variable, so it relates bosonic fields and fermionic fields. If we write the BRST transformation of a given field $\mathcal{O}$ as $\mathcal{O} \to \mathcal{O} + \theta\Delta\mathcal{O}$, then the linear operator Δ satisfies $\Delta^2 = 0$, just like the exterior derivative d or the boundary operator ∂. Because the BRST transformation is a gauge redundancy, states related by $|\psi'\rangle = |\psi\rangle + \Delta|\phi\rangle$ are physically equivalent. Moreover, demanding that physical states are gauge invariant means that they should satisfy $\Delta|\text{phys}\rangle = 0$. Thus,

$$\text{physical states} = \frac{\text{states with } \Delta|\psi\rangle = 0}{(|\psi\rangle \simeq |\psi\rangle + \Delta|\phi\rangle)}. \tag{5.45}$$

This is called the cohomology of the BRST operator Δ.

The point of this is that it constrains the renormalization of nonabelian gauge theory, specifically where the ghosts can appear, and is preserved by loops. It is a crucial ingredient in the demonstration that a finite number of counterterms are required to renormalize the theory. Let's talk about those counterterms now.

5.3 QCD beta function

[Schwartz §26.6] We're going to calculate the beta function for the QCD coupling g. We'll use dim reg, so the beta function is defined as $\beta(g_R) = \mu\partial_\mu g_R$, where μ is the scale that appears when we replace 4-dimensional integrals with D-dimensional integrals, and where g_R is the renormalized coupling. Here is a good device for working out the beta function in dim reg. Very explicitly, the whole Lagrangian in $D = 4 - \epsilon$ dimensions is

$$\mathcal{L} = -\frac{1}{4}Z_3(\partial A)^2 + Z_2\bar{q}\left(\mathbf{i}\partial\!\!\!/ - Z_m m_R\right)q - Z_{1c}\bar{c}^a\Box c^a \tag{5.46}$$

$$- \mu^{\epsilon/2}g_R Z_{A^3} f^{abc}(\partial_\mu A^a_\nu)A^{\mu b}A^{\nu c} - \frac{1}{4}\mu^\epsilon g_R^2 Z_{A^4}\left(f^{abc}A^b_\mu A^c_\nu\right)\left(f^{ade}A^{\mu d}A^{\nu e}\right) \tag{5.47}$$

$$+ \mu^{\epsilon/2}g_R Z_1 A^a_\mu \bar{q}\gamma^\mu t^a q + \mu^{\epsilon/2}g_R Z_{3c} f^{abc}\partial_\mu\bar{c}^a A^b_\mu c^c \tag{5.48}$$

The first line is the quadratic piece about which we're expanding, and the next two lines are interactions, all proportional to the (assumed-small) coupling g_R. Here I've written the counterterms in terms of $Z_X = 1 - \delta_X$ for each term. Notice that there are *four* counterterms $(Z_1, Z_{A^3}, Z_{A^4}, Z_{3c})$ all of which describe corrections to g – they are related by gauge invariance, just like how in QED the vertex correction and the electron self-energy were related.[45]

The *bare fields* are the ones whose quadratic terms are just $(\partial A^0)^2$ and $\overline{q^0}\mathbf{i}\partial\!\!\!/ q^0$ and $\bar{c}^0\Box c^0$, *i.e.* $A^0_\mu = \sqrt{Z_3}A_\mu, q^0 = \sqrt{Z_2}q, c^0 = \sqrt{Z_{1c}}c$. The bare coupling is the coefficient of the interaction written in terms of the bare fields, *e.g.*

$$\mathcal{L}_{qqg} = \underbrace{\mu^{\frac{4-D}{2}}g_R Z_1 Z_3^{-1/2}Z_2^{-1}}_{=g_0} A^{0a}_\mu \bar{q}^0\gamma^\mu t^a q^0.$$

Now here comes the trick: the bare coupling doesn't know anything about our choice

[45] Note that I also include the dimensional-analysis-restoring dim reg factor of $\mu^{\epsilon/2}$ explicitly in $\mathcal{L}$. We can see that this is the right thing to do by rescaling $\tilde{A} = gA$ so that the coupling appears only in $-\frac{1}{4g^2}\mathrm{tr}F^2$. In that case, the action and the coupling g are both dimensionless in $D = 4-\epsilon$ dimensions if we write $S = -\int d^Dx \frac{\mu^{D-4}}{4g^2}\mathrm{tr}F^2$. This is the same as replacing $g \to \mu^{\epsilon/2}g$.

of μ.[46] Therefore

$$0 = \mu\partial_\mu g_0 = \mu\partial_\mu \left(\mu^{\frac{\epsilon}{2}} g_R Z_1 Z_3^{-1/2} Z_2^{-1} Z_1\right) = g_0 \left(\frac{\epsilon}{2} + \frac{1}{g_R} \underbrace{\mu\partial_\mu g_R}_{=\beta(g_R)} + \mu\partial_\mu \left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right)\right) + \mathcal{O}(g^3)\,. \tag{5.49}$$

Now in $\overline{\text{MS}}$, the counterterms δ will depend on μ through $g_R(\mu)$ (at one loop they'll just be $\frac{g_R^2}{\epsilon}$ times numbers), so we can use the chain rule:

$$\mu\partial_\mu \delta = \mu \frac{dg_R}{d\mu}\frac{\partial}{\partial g_R}\delta = \beta(g_R)\frac{\partial \delta}{\partial g_R}.$$

Solving (5.49) perturbatively for β, we have

$$\beta(g_R) = -\frac{\epsilon}{2} g_R - g_R \mu\partial_\mu \left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right) + \mathcal{O}(g^4) \tag{5.50}$$

$$= -\frac{\epsilon}{2} g_R - g_R \underbrace{\beta(g_R)}_{=-\frac{\epsilon}{2}g_R + \mathcal{O}(g_R^2)} \partial_{g_R}\left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right) + \mathcal{O}(g^4) \tag{5.51}$$

$$= -\frac{\epsilon}{2} g_R + \frac{\epsilon}{2} g_R^2 \partial_{g_R}\left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right) + \mathcal{O}(g_R^4). \tag{5.52}$$

So we need to know how the counterterms $\delta_{1,2,3}$ depend on the coupling. We could have chosen a different term to focus on, which would have required knowing a different set of the counterterms; we'd have to get the same answer by gauge invariance.

Gluon vacuum polarization. The Ward identity in QCD still implies that the gluon self-energy is transverse:

$$\mathbf{i}\Pi_{ab}^{\mu\nu}(q) = -\mathbf{i}\Pi_{ab}\left(q^2\eta^{\mu\nu} - q^\mu q^\nu\right)$$

$$= \quad + \quad + \quad + \quad + \quad$$

$$\equiv \mathcal{M}_q + \mathcal{M}_3 + \mathcal{M}_4 + \mathcal{M}_{\text{ghost}} - \mathbf{i}\left(k^2\eta^{\mu\nu} - k^\mu k^\nu\right)\delta^{ab}\delta_3$$

through one loop. In Feynman gauge, we have

$$\mathbf{i}\mathcal{M}_q^{\mu\nu ab}(q) = \quad = -\text{tr}_F(t^a t^b)\left(\mathbf{i}g\right)^2 \bar{\mu}^{4-D}\int \mathrm{d}^D k \frac{\mathbf{i}}{(q-k)^2 - m^2}\frac{\mathbf{i}}{k^2 - m^2}\text{tr}\left[\gamma^\mu\left(\not{k} - \not{q} + m\right)\gamma^\nu\left(\not{k} + m\right)\right]$$

$$= \text{tr}_F(t^a t^b)\mathbf{i}\mathcal{M}_{QED}(e \to g). \tag{5.53}$$

[46] This is a different perspective than we have when the scale introduced in the renormalization scheme is a UV cutoff. There, if we hold fixed the physical coupling we *must* vary the bare coupling with the UV cutoff, and in fact its variation defines the beta function, as in §1.1. The two perspectives are related to each other like active and passive transformations; the object under study is the transformation itself which here is encoded in the beta function.

There are no surprises here – it looks just like the electron loop contribution to the photon vacuum polarization. The color trace is $\mathrm{tr}_F(t^a t^b) = T_F \delta^{ab} = \frac{1}{2}\delta^{ab}$ for the fundamental representation. Since we're interested in the UV-singular structure (recall that we're trying to compute the counterterms $\delta_{1,2,3}$), we can simplify our lives by setting the quark masses to zero. Using exactly the same tricks as for QED, the answer is then, near $D = 4 - \epsilon$,

$$\mathbf{i}\mathcal{M}_q^{\mu\nu ab}(q) = N_f T_F \left(q^2\eta^{\mu\nu} - q^\mu q^\nu\right)\delta^{ab}\frac{g^2}{16\pi^2}\left(-\frac{8}{3}\frac{1}{\epsilon} - \frac{20}{9} - \frac{4}{3}\ln\frac{\mu^2}{-q^2} + \mathcal{O}(\epsilon)\right) \tag{5.54}$$

where N_f is the number of flavors of quarks (*e.g.* up, down ... 6 altogether, so far), counting Dirac multiplets.

More novel are the gluon and ghost loops:

$$\mathbf{i}\mathcal{M}^{\mu\nu ab}_{\text{ghost}} = \quad = (-1)(-g)^2\int \mathrm{d}\!\!^{\,D}k \frac{\mathbf{i}}{(k-q)^2}\frac{\mathbf{i}}{k^2} f^{cad}k^\mu f^{dbc}(k-q)^\nu$$

$$\overset{(5.55)}{=} g^2\frac{\bar\mu^{4-D}}{(4\pi)^{D/2}}\delta^{ab}C_2(\mathsf{G})\int_0^1 dx\left(\frac{1}{\Delta}\right)^{2-D/2}\left(\eta^{\mu\nu}\left(\frac{1}{2}\Gamma\left(2-\frac{D}{2}\right)\Delta\right) + q^\mu q^\nu\left(x(1-x)\Gamma\left(2-\frac{D}{2}\right)\right)\right.$$

The big (-1) is because the ghosts are fermionic. To get to the second line, we used Feynman parameters and completed the square and did the integral over $\ell \equiv k + xq$. $\Delta \equiv x(x-1)q^2$. The new ingredient is the color stuff.

Quadratic Casimir. Recall that the total angular momentum $\mathbf{J}^2 = j(j+1)\mathbb{1}$ has $[\mathbf{J}^2, \vec{J}] = 0$ – it's a Casimir for $\mathsf{SU}(2)$, proportional to the identity on each irrep. This works for any Lie algebra:

$$T^2 \equiv T^a T^a \quad \text{satisfies} \quad [T^b, T^2] = 0, \forall b.$$

In any representation r then we have $T_r^a T_r^a = C_2(r)\mathbb{1}_{d(r)\times d(r)}$ (by Schur's lemma). In particular, for the adjoint rep,

$$-\left(T^a_{\text{adj}}T^a_{\text{adj}}\right)_{bd} = (f^a)_{bc}(f^a)_{cd} = f^{abc}f^{acd} \equiv C_2(\mathsf{G})\delta^{bd}. \tag{5.55}$$

$C_2(r)$ is related to the normalization of the generators: $\mathrm{tr}T_r^a T_r^b = c(r)\delta^{ab}$ (remember, we chose $c(\text{fundamental}) = \frac{1}{2}$). Contracting with δ^{ab} gives $\dim(r)C_2(r) = \dim(\mathsf{G})c(r)$. For $\mathsf{SU}(N)$, $C_2(\mathsf{G}) = N$. See Peskin page 502 for a derivation.

$$\mathbf{i}\mathcal{M}_3^{\mu\nu ab} = \quad = \frac{g^2}{2}\bar\mu^{4-D}\int \mathrm{d}\!\!^{\,D}k\frac{-\mathbf{i}}{k^2}\frac{-\mathbf{i}}{(k-q)^2}f^{acd}f^{bcd}N^{\mu\nu} \tag{5.56}$$

$$= -\frac{g^2}{2}\frac{\bar\mu^{4-D}}{(4\pi)^{D/2}}\delta^{ab}C_2(\mathsf{G})\int_0^1 dx\left(\frac{1}{\Delta}\right)^{2-D/2}\left(\eta^{\mu\nu}A + q^\mu q^\nu B + \eta^{\mu\nu}q^2 C\right). \tag{5.57}$$

The $\frac{1}{2}$ is a symmetry factor, since gluons are real, the two internal gluon lines can be exchanged. $\Delta = x(x-1)q^2$ is the same as before.

$$A = 3(D-1)\Gamma\left(1-\frac{D}{2}\right)\Delta, B = (6(x^2-x+1)-D(1-2x)^2)\Gamma\left(2-\frac{D}{2}\right), C = (-2x^2+2x-5)\Gamma\left(2-\frac{D}{2}\right).$$

The term with A represents a would-be-quadratic divergence. In dim reg this shows up as a pole at $D = 2$.

In the diagram which uses the quartic coupling, too, we find a quadratic divergence $\mathcal{M}_4 \sim \int \frac{\mathrm{đ}^4 k}{k^2} \sim \Lambda^2$:

$$\begin{aligned}
\mathbf{i}\mathcal{M}_4^{\mu\nu ab} = \quad &\text{  } \quad = \frac{\mathbf{i}g^2}{2}\bar{\mu}^{4-D}\int \mathrm{đ}^D k \frac{-\mathbf{i}}{k^2}\eta^{\rho\lambda}\delta^{cd}\left(f^{abe}f^{cde}\left(\delta^\mu_\lambda\delta^\nu_\rho - \delta^\mu_\rho\delta^\nu_\lambda\right)\right. \\
&\left. + f^{ade}f^{cbe}\left(\delta^\mu_\lambda\delta^\nu_\rho - \eta^{\mu\nu}\eta_{\rho\lambda}\right) + f^{ace}f^{bde}\left(\eta^{\mu\nu}\eta_{\lambda\rho} - \delta^\mu_\rho\delta^\nu_\lambda\right)\right) \\
&= -g^2\delta^{ab}\eta^{\mu\nu}C_2(\mathsf{G})(D-1)\bar{\mu}^{4-D}\int \frac{\mathrm{đ}^D k}{k^2}\frac{(q-k)^2}{(q-k)^2} \\
&= -g^2\delta^{ab}\eta^{\mu\nu}C_2(\mathsf{G})(D-1)\bar{\mu}^{4-D}\int_0^1 dx \left(\frac{1}{\Delta}\right)^{2-D/2}\left(-\frac{D}{2}\Gamma\left(1-\frac{D}{2}\right)\Delta + (1-x)^2q^2\Gamma\left(2-\frac{D}{2}\right)\right).
\end{aligned} \tag{5.58}$$

The monstrosity in the first line is just the quartic vertex. The first term vanishes by antisymmetry $\delta^{cd}f^{cde} = 0$. At the second line we multipled by $1 = \frac{(q-k)^2}{(q-k)^2}$ in order to put the integral into the same form as the other terms. [47]

The glue contributions to the gluon vacuum polarization (not including quarks yet, since those are optional) are then

$$\begin{aligned}
\mathcal{M}^{\mu\nu ab}_{\text{glue}}(q) = (\mathcal{M}_3 + \mathcal{M}_4 + \mathcal{M}_{\text{ghost}})^{\mu\nu ab} = \delta^{ab}C_2(\mathsf{G})g^2\frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}}\int_0^1 dx\left(\frac{1}{\Delta}\right)^{2-D/2}\cdot \\
\left[\eta^{\mu\nu}\Gamma\left(1-\frac{D}{2}\right)\Delta\left(\underbrace{-\frac{1}{2}+\frac{3(D-1)}{2}-\frac{D(D-1)}{2}}_{=-\frac{1}{2}(D-2)^2}\right) + q^\mu q^\nu\Gamma\left(2-\frac{D}{2}\right)\mathfrak{a} + \eta^{\mu\nu}q^2\Gamma\left(2-\frac{D}{2}\right)\mathfrak{b}\right]
\end{aligned}$$

Here $\mathfrak{a} = -3(x^2-x+1) + \frac{D}{2}(1-2x)^2 + x(1-x)$ and $\mathfrak{b} = x^2 - x + \frac{5}{2} - (1-x)^2(D-1)$. The coefficient of $\Gamma\left(1-\frac{D}{2}\right)$ has a factor of $D-2$, which cancels the pole at $D = 2$.

[47] Actually, there is a sense in which this contribution is zero in dim reg. After the rewriting, it's still zero, but only after doing the x integral. How can it affect anything then? The difference after the rewriting is merely that the cancellation of the pole at $D = 2$ happens in the integrand of the x integral, rather than only after integration.

Then using $\Gamma\left(1-\frac{D}{2}\right)(D-2) = -2\Gamma\left(2-\frac{D}{2}\right)$, this term combines with the other two. After some boiling using the $x \leftrightarrow 1-x$ symmetry, this is

$$\mathcal{M}^{\mu\nu ab}_{\text{glue}}(q) = \delta^{ab} C_2(\mathsf{G}) g^2 \frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}} \int_0^1 dx \left(\frac{1}{\Delta}\right)^{2-D/2} (\eta^{\mu\nu} - q^\mu q^\nu)\, \Gamma\left(2-\frac{D}{2}\right) \left(\left(1-\frac{D}{2}\right)(1-2x)^2 + 2\right)$$
$$\overset{D=4-\epsilon}{=} C_2(\mathsf{G})\,(\eta^{\mu\nu} - q^\mu q^\nu)\,\delta^{ab} \frac{g^2}{(4\pi)^2} \left(\frac{10}{3}\frac{1}{\epsilon} + \frac{31}{9} + \frac{5}{3}\ln\frac{\mu^2}{-q^2} + \mathcal{O}(\epsilon)\right).$$

Notice that compared to (5.54), the coefficient of the log (and of the pole in ϵ) has the opposite sign. From this we conclude that to cancel the ϵ^{-1} pole in the vacuum polarization (this is the $\overline{\text{MS}}$ scheme) we must take

$$\delta_3 = \frac{1}{\epsilon}\frac{g^2}{16\pi^2}\left(\frac{10}{3}C_2(\mathsf{G}) - \frac{8}{3}N_f T_F\right) , \tag{5.59}$$

where the first term is from pure glue effects and the second term is from the quark loop.

We're almost there. To get the beta function we also need δ_1 and δ_2.

Quark self-energy. This determines δ_2 and δ_m (the latter we won't need). The UV bit doesn't care about the mass, so let's set $m = 0$. Again it is just like QED except for the color factors. The one-loop correction to the quark self-energy is

$$\mathbf{i}\Sigma_2^{ij}(\not{p}) = \int \mathrm{d}\!\!^{\text{-}D} k\, t^a_{ik}\gamma^\mu \frac{\mathbf{i}\not{k}\delta^{kl}}{k^2 + \mathbf{i}\epsilon} t^b_{lj}\gamma_\mu \frac{-\mathbf{i}\delta^{ab}}{(k-p)^2 + \mathbf{i}\epsilon}$$

The color factors are

$$t^a_{ik}t^b_{lj}\delta^{ab}\delta^{kl} = \sum_a (t^a t^a)_{ij} = C_2(F)\delta_{ij}$$

where $C_2(F)$ is the quadratic Casimir in the fundamental representation, which[48] for $\mathsf{SU}(N)$ is $C_2(F) = \frac{N^2-1}{2N}$. The momentum integral is the same as in QED and we find (for $m_q = 0$)

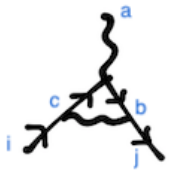
$$\Sigma_2^{ij}(\not{p}) = \frac{g^2}{8\pi^2}\delta^{ij}C_2(F)\frac{1}{\epsilon}\not{p} + \text{finite} . \tag{5.60}$$

The quark wavefunction renormalization counterterm δ_2 contributes as $\Sigma^{ij} = ... + \delta_2\delta^{ij}\not{p}$, so in $\overline{\text{MS}}$ we must set

$$\delta_2 = \frac{1}{\epsilon}\frac{g^2}{16\pi^2}\left(-2C_2(F)\right).$$

[48] To check this take the trace of the BHS and compare with $\text{tr}\, t^a t^b = \frac{1}{2}\delta^{ab}$.

Vertex correction. The vertex correction gets two contributions at one loop.



$$= \mathbf{i}g(t^c t^a t^b)_{ij}\delta^{bc}\Gamma^\mu_{\text{QED}}$$

where

$$\Gamma^\mu_{\text{QED}} = F_1(p^2)\gamma^\mu + \frac{\mathbf{i}\sigma^{\mu\nu}}{2m}p_\nu F_2(p^2)$$

is identical to the QED answer with e replaced with g (notice that it's useful to keep the quark mass around for a bit here). The color factors are

$$t^c t^a t^b \delta^{bc} = t^b t^a t^b = t^b t^b t^a + t^b[t^a, t^b] = C_2(F)t^a + \mathbf{i}f^{abc}t^b t^c.$$

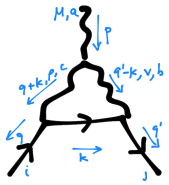
By antisymmetry of f^{abc}, the second term is

$$\mathbf{i}f^{abc}t^b t^c = \mathbf{i}f^{abc}\frac{1}{2}[t^b, t^c] = -\frac{1}{2}f^{abc}f^{bcd}t^d = -\frac{1}{2}C_2(\mathsf{G})t^a.$$

Altogether, the divergent bit of this diagram is

$$= \mathbf{i}g\left(C_2(F) - \frac{1}{2}C_2(\mathsf{G})\right)t^a_{ij}\gamma^\mu\frac{g^2}{16\pi^2}\left(\frac{2}{\epsilon} + \ln\frac{\mu^2}{-p^2} + \text{finite}\right).$$

The other diagram is new:



$$= \mathbf{i}g f^{abc}(t^c t^b)_{ij}\Gamma^\mu_{\text{new}}$$

with

$$-\mathbf{i}g\Gamma^\mu_{\text{new}}(p^2) = (\mathbf{i}g)^2 g\bar{\mu}^{4-D}\int \mathrm{d}^D k\gamma_\rho\frac{\mathbf{i}}{\not{k}}\gamma_\nu\frac{-\mathbf{i}}{(q+k)^2}\frac{-\mathbf{i}}{(q'-k)^2}\cdot$$
$$\left(\eta^{\mu\nu}(2q+q'+k)^\rho + \eta^{\nu\rho}(-q+q'-2k)^\mu + \eta^{\rho\mu}(k-2q'-q)^\nu\right)$$

(I find the opposite sign from Schwartz here. This sign cancels against the one in (5.61).) The horrible numerator comes from the 3-gluon vertex, but in computing the

UV divergence we can set the external momenta to zero. This gives

$$\begin{aligned}
-\Gamma^{\mu}_{\text{new}}(p^2 \to 0) &= g^2 \bar{\mu}^{4-D} \int \text{đ}^D k \frac{\gamma_\rho \not{k} \gamma_\nu}{k^6} \left(\eta^{\mu\nu} k^\rho - 2\eta^{\nu\rho} k^\mu + \eta^{\rho\mu} k^\nu\right) \\
&= g^2 \bar{\mu}^{4-D} \int \frac{\text{đ}^D k}{k^6} \left(2k^2 \gamma^\mu - 2\gamma_\rho \not{k} \gamma^\rho k^\mu\right) \\
&= g^2 \left(4 - \frac{4}{D}\right) \gamma^\mu \bar{\mu}^{4-D} \int \frac{\text{đ}^D k}{k^4} \qquad \gamma_\rho \gamma^\nu \gamma^\rho = (2-D)\gamma^\nu, \int k^\mu k^\nu \ldots = \int \frac{k^2}{D} \eta^{\mu\nu} \ldots \\
&= \mathbf{i} \gamma^\mu \frac{g^2}{16\pi^2} \left(\frac{6}{\epsilon} + 3 \log \frac{\mu^2}{-p^2} + \text{finite}\right)
\end{aligned}$$

where at the last step we put back the gluon momentum to make up the dimensions. Finally the color factor is

$$f^{abc}(t^c t^b) = \frac{1}{2} \mathbf{i} f^{abc} f^{cbd} t^d = -\mathbf{i} \frac{1}{2} C_2(\mathsf{G}) t^a. \tag{5.61}$$

Altogether, the divergent part of the qqg vertex at one loop is then

$$\frac{1}{\epsilon} \mathbf{i} g t^a_{ij} \gamma^\mu \left(\left(2\left(C_2(F) - \frac{1}{2} C_2(\mathsf{G})\right) + 3C_2(\mathsf{G})\right) \frac{g^2}{16\pi^2} + \delta_1 \epsilon\right) \implies \delta_1 = \frac{1}{\epsilon} \frac{g^2}{16\pi^2} \left(-2C_2(F) - 2C_2(\mathsf{G})\right).$$

Combining all of this information using (5.52), the QCD beta function is (dropping the R subscripts on g_R)

$$\begin{aligned}
\beta(g) &= -\frac{\epsilon}{2} g + \frac{\epsilon}{2} g^2 \partial_g \left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right) + \mathcal{O}(g^4) \\
&\overset{D \to 4}{=} \frac{g^3}{16\pi^2} \left(-2C_2(F) - 2C_2(\mathsf{G}) - \frac{1}{2}\left(\frac{10}{3} C_2(\mathsf{G}) - \frac{8}{3} N_f T_F\right) - \left(-2C_2(F)\right)\right) + \mathcal{O}(g^4) \\
&= -\frac{g^3}{16\pi^2} \left(\frac{11}{3} C_2(\mathsf{G}) - \frac{4}{3} N_f T_F\right) \\
&\overset{\mathsf{SU}(N) \text{ with fundamental quarks}}{=} -\frac{g^3}{16\pi^2} \left(\frac{11}{3} N - \frac{2}{3} N_f\right).
\end{aligned}$$

If there are not too many species of quarks ($N_f < 6N = 18$, which is true in the Standard Model, where $N_F = 6$), β is negative, in which case such a non-Abelian gauge theory is asymptotically free, as I've promised many times in the preceding pages.

Defining a running coupling as in (2.72), and solving the resulting RG equation $\frac{d}{d \log M^2} g = \beta(g)$ using the above beta function (integrating from M^2 to $|q^2|$) we find a crucial plus sign relative to (2.72)

$$g^2_{\text{eff}}(q^2) = \frac{g^2}{1 + \frac{g^2}{16\pi^2} C \log\left(\frac{|q^2|}{M^2}\right)}, \quad C \equiv \frac{11}{3} N - \frac{2}{3} N_f \tag{5.62}$$

and the coupling grows as q decreases, and shrinks at large q. The former is consistent with the conjecture that the strong coupling lattice calculation that shows confinement is adiabatically connected to the continuum physics, and the latter means that we can still calculate some things. Actually it is a bit tricky to define the effective coupling in QCD, but (a more precise version of) this curve has been measured (see Peskin fig. 17.23 and Schwartz §26.3).

Qualitative picture of asymptotic freedom. [Peskin §16.7] The sign of the beta function in QED can be understood as charge screening by the vacuum – electron-positron pairs fluctuate into existence, and respond to the presence of a source in such a way as to decrease its field at long distance.

How does non-Abelian gauge theory manage to produce antiscreening? There is certainly still screening from the quarks, and since the gluons are charged, they will also produce screening. So it makes sense that too many quarks will spoil the soup. But whence the terms of the opposite sign in the beta function?

Following Peskin §16.7, consider pure (no quarks) $\mathsf{SU}(2)$ gauge theory, in Coulomb gauge $\partial_i A^{ia} = 0$. In this gauge, we sacrifice Lorentz covariance for more manifest unitarity – no ghosts, and no longitudinal and timelike polarization states. The equation of motion for A^{0a} is the Gauss law (in terms of $E^{ia} \equiv F^{0ia}$):

$$g\rho^a = D_i E^{ia} = \partial_i E^{ia} + g f^{abc} A_i^b E^{ic},$$

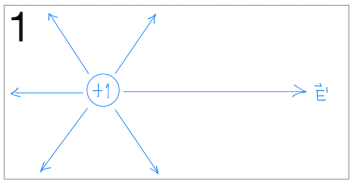
where ρ is the charge density (*e.g.* the number density of quarks if we included them in the theory), and for $\mathsf{SU}(2)$, the structure constants are $f^{abc} = \epsilon^{abc}$. Instead of dynamical quarks, let's consider a static color source particle: $\rho^a(x) = \delta^{(3)}(\vec{x})\delta^{a1}$ pointing in a particular color direction, so the equation we wish to solve is

$$\partial_i E^{ia} = g\delta^{(3)}(\vec{x})\delta^{a1} + g f^{abc} A^{bi} E^{ic}. \tag{5.63}$$

Let's solve this perturbatively in g in several steps.

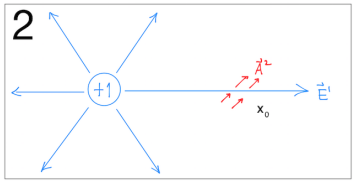
1. At leading order, the source produces a Coulomb field:

   $\vec{E}^a(x) = \frac{g\delta^{1a}\check{x}}{x^2}$. So far, this is just classical physics.



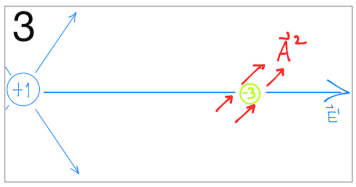
2. The quantum mechanics comes in here:

consider a fluctuation of the vector potential in the 2d color direction $A^{b=2,i}(\vec{x})$, with support localized somewhere, call it x_0, away from the origin. Suppose it points in some direction, somewhat aligned with $\vec{x}$, its displacement from the source.

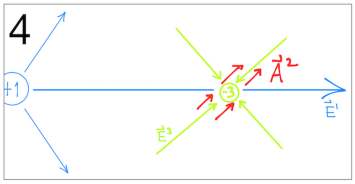


3. Here comes the iteration.

   The second term on the RHS of (5.63) is then $g\epsilon^{abc}\vec{A}^b \cdot \vec{E}^c \propto -\delta^{c3}\vec{A}^2 \cdot \vec{E}^1$, a sink for the color-electric field in the 3rd color direction.



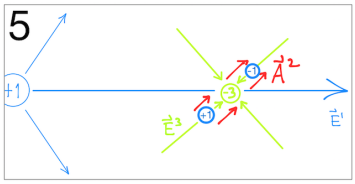
4. This produces a new Coulomb field $\vec{E}^3(\vec{x}) \sim -\frac{\vec{x}-\vec{x}_0}{|x-x_0|^3}$ pointing towards x_0.



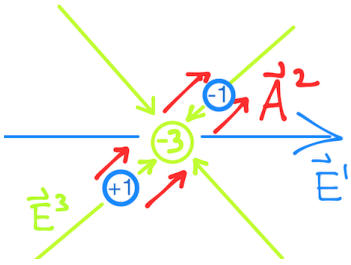
5. Now look at the second term on the RHS of (5.63) again:

$$\vec{\nabla} \cdot \vec{E}^1 = \ldots + g\epsilon^{123}\vec{A}^2 \cdot \vec{E}^3$$

   it is a *source* (*sink*) for the color field in direction 1 where $\vec{A}^2$ and $\vec{E}^3$ are parallel (antiparallel).



But if the fluctuation $\vec{A}^2$ points away from the source, then in the region closer to the source, $\vec{A}^2 \cdot \vec{E}^3 > 0$, and farther from the source they are anti-aligned. This produces a dipole source for $\vec{E}^1$ which points *toward* the original charge, and therefore anti-screens its field.



Warning: on the other hand, if the fluctuation $\vec{A}^2$ points toward the original source then this process produces a dipole pointing away from the original source, which contributes to screening. I'm not sure if this picture can be made quantitative.

Disclaimer. This discussion just scratches the surface of the physics of QCD! Many measurable phenomena can be calculated using the machinery we've set up. Please see Peskin §17 and Schwartz §32.

6 A parable on integrating out degrees of freedom

Here's another parable from QM which gives some useful perspective on renormalization in QFT and on the notion of effective field theory.

[Banks p. 138] Consider a system of two coupled harmonic oscillators. We will assume one of the springs is much stiffer than the other: let's call their natural frequencies ω_0, Ω, with $\omega_0 \ll \Omega$. The euclidean-time action is

$$S[Q,q] = \int dt \left[\frac{1}{2}\left(\dot{q}^2 + \omega_0^2 q^2\right) + \frac{1}{2}\left(\dot{Q}^2 + \Omega^2 Q^2\right) + gQq^2 \right] \equiv S_{\omega_0}[q] + S_\Omega[Q] + S_{\rm int}[Q,q].$$

(The particular form of the q^2Q coupling is chosen for convenience. Don't take too seriously the physics at large negative Q.) We can construct physical observables in this model by studying the path integral:

$$Z = \int [dQdq] e^{-S[Q,q]}.$$

Since I put a minus sign rather than an i in the exponent (and the potential terms in the action have + signs), this is a euclidean path integral.

Let's consider what happens if we do the path integral over the heavy mode Q, and postpone doing the path integral over q. This step, naturally, is called *integrating out* Q, and we will see below why this is a good idea. The result just depends on q; we can think of it as an *effective action* for q:

$$e^{-S_{\rm eff}[q]} := \int [dQ] e^{-S[q,Q]}$$

$$= e^{-S_{\omega_0}[q]} \left\langle e^{-S_{\rm int}[Q,q]} \right\rangle_Q$$

Here $\langle ... \rangle_Q$ indicates the expectation value of ... in the (free) theory of Q, with the action $S_\Omega[Q]$. It is a gaussian integral (because of our choice of $S_{\rm int}$):

$$\left\langle e^{-S_{\rm int}[Q,q]} \right\rangle_Q = \int [dQ] e^{-S_\Omega[Q] - \int ds J(s) Q(s)} = \mathcal{N} e^{\frac{1}{4}\int ds dt J(s) G(s,t) J(t)} \ .$$

This last equality is an application of the 'fundamental theorem of path integrals,' *i.e.* the gaussian integral. Here $J(s) \equiv gq(s)^2$. The normalization factor $\mathcal{N}$ is independent of J and hence of q. And $G(s,t)$ is the inverse of the linear operator appearing in S_Ω, the euclidean Green's function:

$$S_\Omega[Q] = \int ds dt Q(s) G^{-1}(s,t) Q(t).$$

More usefully, G satisfies

$$\left(-\partial_s^2 + \Omega^2\right) G(s,t) = \delta(s-t) \ .$$

The fact that our system is time-translation invariant means $G(s,t) = G(s-t)$. We can solve this equation in fourier space: $G(s) = \int đ\omega e^{-i\omega s} G_\omega$ makes it algebraic:

$$G_\omega = \frac{1}{\omega^2 + \Omega^2}$$

and we have

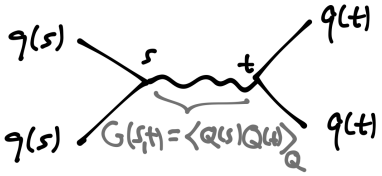
$$G(s) = \int đ\omega e^{-\mathbf{i}\omega s} \frac{1}{\omega^2 + \Omega^2}. \tag{6.1}$$

So we have:

$$e^{-S_{\text{eff}}[q]} = e^{-S_{\omega_0}[q]} e^{-\int dtds \frac{g^2}{2} q(s)^2 G(s,t) q(t)^2}$$

or taking logs

$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dtds \frac{g^2}{2} q(s)^2 G(s,t) q(t)^2 \ . \tag{6.2}$$



Q mediates an interaction of four qs, an anharmonic term, a self-interaction of q. In Feynman diagrams, the leading interaction between q's mediated by Q comes from the diagram at left.

And the whole thing comes from exponentiating disconnected copies of this diagram. There are no other diagrams: once we make a Q from two qs what can it do besides turn back into two qs? Nothing. And no internal q lines are allowed, they are just sources, for the purposes of the Q integral.

But it is *non-local*: we have two integrals over the time in the new quartic term. This is unfamiliar, and *bad*: *e.g.* classically we don't know how to pose an initial value problem using this action.

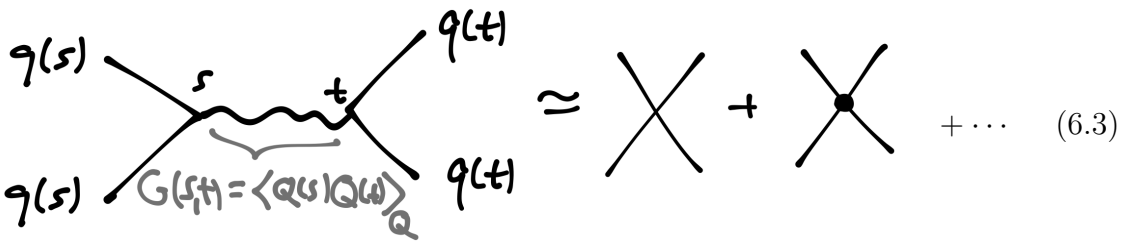
But now suppose we are interested in times much longer than $1/\Omega$, say times comparable to the period of oscillation of the less-stiff spring $2\pi/\omega_0$. We can accomplish this by Taylor expanding under the integrand in (6.1):

$$G(s) = \int đ\omega e^{-\mathbf{i}\omega s} \frac{1}{\Omega^2} \underbrace{\frac{1}{1+\frac{\omega^2}{\Omega^2}}}_{=\sum_n (-1)^n \left(\frac{\omega^2}{\Omega^2}\right)^n} \overset{s \gg 1/\Omega}{\simeq} \frac{1}{\Omega^2}\delta(s) + \frac{1}{\Omega^4}\partial_s^2 \delta(s) + ...$$

Plug this back into (6.2):

$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dt \frac{g^2}{2\Omega^2} q(t)^4 + \int dt \frac{g^2}{2\Omega^4} \dot{q}^2 q^2 + ...$$

The effects of the heavy mode Q are now organized in a *derivative expansion*, with terms involving more derivatives suppressed by more powers of the high energy scale Ω.



(6.3)

A useful mnemonic for integrating out the effects of the heavy field in terms of Feynman diagrams: to picture Q as propagating for only a short time (compared to the external time $t-s$), we can contract its propagator to a point. The first term on the RHS shifts the q^4 term, the second shifts the kinetic term, the third involves four factors of $\dot{q}$...

On the RHS of this equation, we have various interactions involving four qs, which involve increasingly many derivatives. The first term is a quartic potential term for q: $\Delta V = \frac{g}{\Omega^2} q^4$; the leading effect of the fluctuations of Q is to *shift* the quartic self-coupling of q by a finite amount (note that we could have included a bare $\lambda_0 q^4$ potential term).

Notice that if we keep going in this expansion, we get terms with *more than two derivatives* of q. This is OK. We've just derived the right way to think about such terms: we treat them as a perturbation, and they are part of a never-ending series of terms which become less and less important for low-energy questions. If we want to ask questions about q at energies of order ω, we can get answers that are correct up to effects of order $\left(\frac{\omega}{\Omega}\right)^{2n}$ by keeping the nth term in this expansion.

Conversely if we are doing an experiment with precision Δ at energy ω, we can measure the effects of up to the nth term, with

$$\left(\frac{\omega}{\Omega}\right)^{2n} \sim \Delta.$$

Another important lesson: $S_{\text{eff}}[q]$ contains couplings with negative dimensions of energy

$$\sum_n c_n \left(\partial_t^n q\right)^2 q^2, \quad \text{with } c_n \sim \frac{1}{\Omega^{2n}},$$

exactly the situation where the S-matrix grows too fast at high energies that we discussed above in (3.17). In this case we know exactly where the probability is going: if we have enough energy to see the problem ($E \sim \Omega$), we have enough energy to kick the heavy mode Q out of its groundstate.

6.0.1 Attempt to consolidate understanding

We've just done some coarse graining: focusing on the dofs we care about (q), and actively ignoring the dofs we don't care about (Q), except to the extent that they affect those we do (e.g. the self-interactions of q).

Above, we did a calculation in a QM model with two SHOs. This is a microcosm of QFT in many ways. For one thing, free quantum fields are bunches of harmonic oscillators with natural frequency depending on k, $\Omega = \sqrt{\vec{k}^2 + m^2}$. Here we kept just two of these modes (one with large k, one with small k) for clarity. Perhaps more importantly, QM is just QFT in 0+1d. The more general QFT path integral just involves more integration variables. The idea of the Wilsonian RG (for continuum field theory) is essentially to do the integrals over the modes in descending order of wavenumber, and at each stage make the expansion described above to get a local action. And notice that basically all possible terms are generated, consistent with the symmetries (here for example, there is a $\mathbb{Z}_2$ symmetry under which $q \to -q$, so there are no odd powers of q).

The result of that calculation was that fluctuations of Q mediate various q^4 interactions. It adds to the action for q the following: $\Delta S_{\text{eff}}[q] \sim \int dtds q^2(t) G(t-s) q^2(s)$, as in Fig. 6.3.

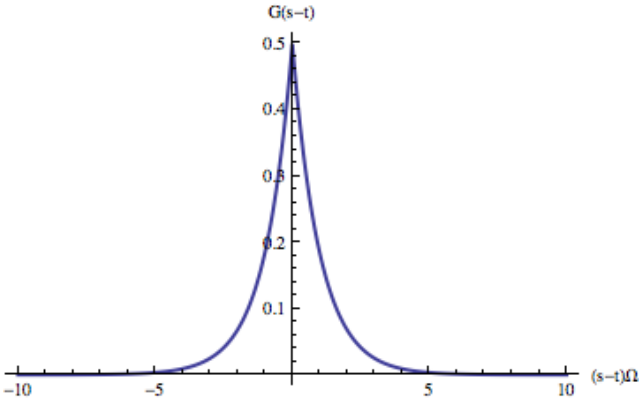
If we have the hubris to care about the exact answer, it's nonlocal in time. But if we want exact answers then we'll have to do the integral over q, too. On the other hand, the hierarchy of scales $\omega_0 \ll \Omega$ is useful if we ask questions about energies of order ω_0, e.g.

$$\langle q(t)q(0)\rangle \text{ with } t \sim \frac{1}{\omega_0} \gg \frac{1}{\Omega}.$$

Then we can Taylor expand the function $G(t-s)$, and we find a series of corrections in powers of $\frac{1}{t\Omega}$ (or more accurately, powers of $\frac{\partial_t}{\Omega}$).

(Notice an important asymmetry: it's not so useful to integrate out light degrees of freedom to get an action for the heavy degrees of freedom; that would necessarily be nonlocal and stay nonlocal and we wouldn't be able to treat it using ordinary techniques.)

The crucial point is that the scary non-locality of the effective action that we saw only extends a distance of order $\frac{1}{\Omega}$; the kernel $G(s-t)$ looks like this:



The mechanism we've just discussed is an essential ingredient in getting *any* physics done at all. *Why* can we do physics despite

the fact that we do not understand the theory of quantum gravity that governs Planckian distances? We happily do lots of physics without worrying about this! This is because the effect of those Planckian quantum gravity fluctuations – whatever they are, call them Q – on the degrees of freedom we do care about (*e.g.* the Standard Model, or an atom, or the sandwich you made this morning, call them collectively q) are encoded in terms in the effective action of q which are suppressed by powers of the high energy scale $M_{\rm Planck}$, whose role in the toy model is played by Ω. And the natural energy scale of your sandwich is much less than $M_{\rm Planck}$.

I picked the Planck scale as the scale to ignore here for rhetorical drama, and because we really are ignorant of what physics goes on there. But this idea is equally relevant for *e.g.* being able to describe water waves by hydrodynamics (a classical field theory) without worrying about atomic physics, or to understand the physics of atoms without needing to understand nuclear physics, or to understand the nuclear interactions without knowing about the Higgs boson, and so on deeper into the onion of physics.

This wonderful situation, which makes physics possible, has a price: since physics at low energies is so insensitive to high energy physics, it makes it hard to learn about high energy physics! People have been very clever and have learned a lot in spite of this vexing property of the RG called decoupling. We can hope that will continue. (Cosmological inflation plays a similar role in hiding the physics of the early universe. It's like whoever designed this game is *trying* to hide this stuff from us.)

The explicit functional form of $G(s)$ (the inverse of the (euclidean) kinetic operator for Q) is:

$$G(s) = \int đ\omega \frac{e^{-\mathbf{i}\omega s}}{\omega^2 + \Omega^2} = e^{-\Omega|s|}\frac{1}{2\Omega}. \tag{6.4}$$

Do it by residues: the integrand has poles at $\omega = \pm\mathbf{i}\Omega$. The absolute value of $|s|$ is crucial, and comes from the fact that the contour at infinity converges in the upper (lower) half plane for $s < 0$ ($s > 0$).

Next, some comments about ingredients in this discussion, which provide a useful opportunity to review some important QFT technology:

- Please don't be confused by the formal similarity of the above manipulations with

the construction of the generating functional of correlation functions of Q:

$$Z[J] \equiv \left\langle e^{\int dt Q(t) J(t)} \right\rangle_Q \ , \quad \langle Q(t_1)Q(t_2)...\rangle_Q^{\text{connected}} = \frac{\delta}{\delta J(t_1)} \frac{\delta}{\delta J(t_2)} ... \log Z[J]$$

It's true that what we did above amounts precisely to constructing $Z[J]$, and plugging in $J = g_0 q^2$. But the motivation is different: in the above q is also a dynamical variable, so we don't get to pick q and differentiate with respect to it; we are merely postponing doing the path integral over q until later.

- Having said that, what is this quantity $G(s)$ above? It is the (euclidean) two-point function of Q:

$$G(s,t) = \langle Q(s)Q(t)\rangle_Q^{\text{connected}} = \frac{\delta}{\delta J(t)} \frac{\delta}{\delta J(s)} \log Z[J]|_{J=0}.$$

The middle expression makes it clearer that $G(s,t) = G(s-t)$ since nobody has chosen the origin of the time axis in this problem. This euclidean Green's function, the inverse of $-\partial_\tau^2 + \Omega^2$, is unique, once we demand that it falls off at large separation (unlike the real-time Green's function).

- **Adding more labels.** Quantum mechanics is quantum field theory in 0+1 dimensions. Except for our ability to do all the integrals, everything we are doing here generalizes to quantum field theory in more dimensions: quantum field theory *is* quantum mechanics (with infinitely many degrees of freedom). With more spatial dimensions, we'll want to use the variable x for the spatial coordinates (which are just labels on the fields!) and it was in anticipation of this step that I used q instead of x for my oscillator position variables.