Physics 215B: Particles and Fields
Winter 2019

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

Last updated: 2019/03/18, 17:04:49

Contents

0.1 Introductory remarks

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.

It's not surprising that QFT is so useful, since this situation happens all over the place. Some examples of 'stuff' are: the atoms in a solid, or the electrons in those atoms, or the spins of those electrons. A less obvious, but more visible, example is the electromagnetic field, even in vacuum. More examples are provided by other excitations of the vacuum, and it will be our job here to understand those very electrons and atoms that make up a solid in these terms. The vacuum has other less-long-lasting excitations which are described by the Standard Model of particle physics.

Some examples of QFT are Lorentz invariant ('relativistic'). That's a nice simplification when it happens. Indeed this seems to happen in particle physics. We're going to focus on this case for much of this quarter. Still I would like to emphasize: though some of the most successful applications of QFT are in the domain of high energy particle physics, this is not a class on that subject, and I will look for opportunities to emphasize the universality of QFT.

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. This is a crucial point of departure for QFT. It's a necessary consequence of Lorentz symmetry, but the converse is false: particle production can happen without relativity.

'Divergences'. Another intrinsic and famous feature of QFT discernible from the definition I gave above is its flirtation with infinity. I said that there is 'stuff at each point of space'; how much stuff is that? Well, there are two senses in which 'the number of points of space' is infinite: (1) space can go on forever (the infrared (IR)), and (2) in the continuum, in between any two points of space are more points (the ultraviolet (UV)). The former may be familiar from statistical mechanics, where it is associated with the thermodynamic limit, which is where interesting things happen. For our own safety, we'll begin our discussion in a padded room, protected on both sides from the terrors of the infinite.

Prof. Jenkins tells me that 215A ended just as Feynman diagrams were being drawn for the first time. I think therefore that it will be useful to retreat a bit and rederive the diagrammatic expansion from another (in many ways simpler) point of view, namely the path integral.

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* (Wiley)

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

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

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

David Tong's lecture notes

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

0.3 Conventions

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

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

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

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

D is the number of spacetime dimensions, d is the number of space dimensions.

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 The path integral makes some things easy

1.1 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:

$$\mathbf{H} = \sum_{n=1}^{N} \left(\frac{\mathbf{p}_n^2}{2m} + \frac{1}{2}\kappa \left(\mathbf{q}_n - \mathbf{q}_{n-1}\right)^2 \right) + \lambda \mathbf{q}^4. \tag{1.1}$$

Notice that this system is an ordinary QM 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 $V = k_{ab}\mathbf{q}_a\mathbf{q}_b$. 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 $\omega_k^2 = \frac{2\kappa}{m}\sin^2\frac{ka}{2}$. 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.

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

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

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:

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

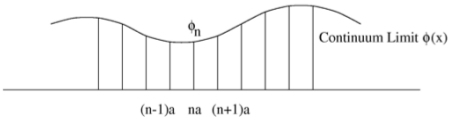
An integral expression for this can be obtained by inserting resolutions of the identity

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

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

Scalar field theory in one dimension. [Zee §1.3] The 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. So let's try to take the 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.]

We'll use

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

The path integral becomes:

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

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

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

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 $a^4 (\partial_x q)^4$ which 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 $\frac{\partial}{\partial q_n}$. [1]

From the phonon problem, we automatically found $r = u = 0$, and the equation of motion is just the wave equation $(\partial_t^2 - v_s^2 \partial_x^2) q = 0$, where we see that v_s is the sound speed. This happened because of the symmetry $q_n \to q_n + \epsilon$. This is the operation

[1] Functional derivatives will be very useful to us. The definition is

$$\frac{\delta \phi(x)}{\delta \phi(y)} = \delta(x - y) \tag{1.2}$$

plus the Liebniz properties (linearity, product rule). More prosaically, they are just partial derivatives, if we define a collection of values of the independent variable $\{x_n\}$ to regard as grid points, and let

$$\phi_n \equiv \phi(x_n)$$

so that (1.2) is just

$$\frac{\partial \phi_n}{\partial \phi_m} = \delta_{nm}.$$

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

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.)

We can construct a hamiltonian from this action by defining a canonical field-momentum density $\pi(x) = \frac{\partial \mathcal{L}}{\partial_t q} = \mu \partial_t q$ and doing the Legendre transformation:

$$H = \sum_n \left(p_n \dot{q}_n - L_n\right) = \int dx \left(\pi(x)\dot{q}(x) - \mathcal{L}\right) = \int dx \left(\frac{\pi(x)^2}{2\mu} + \mu v_s^2 \left(\partial_x q(x)\right)^2 + rq^2 + uq^4 + ...\right) . \tag{1.3}$$

Note that I suppress the dependence of all the fields on t just so it doesn't get ugly, not because it isn't there. Also, I emphasize that the position along the chain x here is just a *label* on the fields, not a degree of freedom or a quantum operator.

The field q is called a *scalar field* because it doesn't have any indices decorating it. This is to be distinguished from *e.g.* the Maxwell field, which is a vector field, and which we'll discuss soon. (Note that vibrations of a crystal in *three* dimensions actually do involve vector indices! We omit this complication.)

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 (in addition to the continuum limit) 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

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

Continuum (free) scalar field theory in $d+1$ dimensions. These continuum expressions are easy to generalize to scalar field theory in any number of dimensions. Let's do this directly in infinite volume and set $\mu = 1$ by rescaling fields. The action is

$$S[\phi] = \int d^d x dt \ \left(\frac{1}{2}\dot{\phi}^2 - \frac{1}{2}v_s^2 \vec{\nabla}\phi \cdot \vec{\nabla}\phi - V(\phi)\right). \tag{1.4}$$

This is almost what we would have found for the long-wavelength ($ka \ll 1$) description of a d-dimensional lattice of masses on springs, like a mattress (except that there would have been one ϕ for each direction in which the atoms can wiggle). The equation of motion is

$$0 = \frac{\delta S[\phi]}{\delta \phi(x)} = -\partial_t^2 \phi + v_s^2 \nabla^2 \phi - V'(\phi). \tag{1.5}$$

For the harmonic case $V(\phi) = \frac{1}{2} m^2 \phi^2$ we know what we're doing, and (1.5) is called the Klein-Gordon equation,

$$0 = \left(\partial_\mu \partial^\mu + m^2\right) \phi. \tag{1.6}$$

(Notice that I've set $v_s = c = 1$ here, and this is where we have committed to a choice of signature convention; take a look at the conventions page §0.3.). In relativistic notation, the Lagrangian density is just $\mathcal{L} = \frac{1}{2}\left(\partial_\mu \phi \partial^\mu \phi - m^2 \phi^2\right)$. This describes free continuum real massive relativistic scalar quantum field theory. (Match the adjectives to the associated features of the lagrangian; collect them all!)

The canonical momentum is $\pi = \frac{\partial \mathcal{L}}{\partial \dot{\phi}} = \dot{\phi}$ and the Hamiltonian (which we can instantly promote to a quantum operator by using boldface symbols) is then

$$\mathbf{H} = \int d^d x \left(\frac{\boldsymbol{\pi}(x)^2}{2} + \frac{1}{2} v_s^2 \left(\vec{\nabla} \boldsymbol{\phi} \cdot \vec{\nabla} \boldsymbol{\phi} \right) + \frac{1}{2} m^2 \boldsymbol{\phi}^2 \right).$$

Note that all these terms are positive.

A translation invariant linear problem is solved by Fourier transforms: $\boldsymbol{\phi}(x) = \int \mathrm{đ}^d k \; e^{-\mathbf{i}\vec{k}\cdot\vec{x}} \boldsymbol{\phi}_k$, and $\boldsymbol{\pi}(x) = \int \mathrm{đ}^d k \; e^{-\mathbf{i}\vec{k}\cdot\vec{x}} \boldsymbol{\pi}_k$, this is

$$\mathbf{H} = \int \mathrm{đ}^d k \left(\frac{1}{2} \boldsymbol{\pi}_k \boldsymbol{\pi}_{-k} + \frac{1}{2} \left(v_s^2 k^2 + m^2 \right) \boldsymbol{\phi}_k \boldsymbol{\phi}_{-k} \right)$$

where $k^2 = (-\mathbf{i}\vec{k}) \cdot (\mathbf{i}\vec{k}) = \vec{k} \cdot \vec{k}$. This is merely a sum of decoupled oscillators, except for the coupling between wavenumbers k and $-k$. We can read off the normal mode frequencies, aka the dispersion relation:

$$\omega_k^2 = v_s^2 k^2 + m^2.$$

Notice that this is also the condition for a Fourier mode $e^{\mathbf{i}\vec{k}\cdot\vec{x} - \mathbf{i}\omega t}$ to solve the Klein-Gordon equation (1.6).

We can decouple the modes with wavenumber k and $-k$ by introducing the ladder operators[2]

$$\boldsymbol{\phi}_k \equiv \sqrt{\frac{\hbar}{2\omega_k}} \left(\mathbf{a}_k + \mathbf{a}_{-k}^\dagger \right), \quad \boldsymbol{\pi}_k \equiv \frac{1}{\mathbf{i}} \sqrt{\frac{\hbar \omega_k}{2}} \left(\mathbf{a}_k - \mathbf{a}_{-k}^\dagger \right), \qquad [\mathbf{a}_k, \mathbf{a}_{k'}^\dagger] = (2\pi)^d \delta^{(d)}(k - k').$$

[2]Beware that the mode operators $\mathbf{a}_k$ defined here differ by powers of $2\pi/L$ from the finite-volume objects in the previous discussion. These agree with Peskin's conventions.

Their commutator follows from $[\phi(x), \pi(y)] = \mathbf{i}\delta^{(d)}(x-y)$. In terms of the ladder operators,

$$\mathbf{H} = \int \mathrm{đ}^d k \ \hbar\omega_k \left(\mathbf{a}_k^\dagger \mathbf{a}_k + \frac{1}{2}\right).$$

The field operators

$$\begin{aligned}
\boldsymbol{\phi}(\vec{x}) &= \int \mathrm{đ}^d k \ \sqrt{\frac{\hbar}{2\omega_k}} \left(e^{\mathbf{i}\vec{k}\cdot\vec{x}}\mathbf{a}_k + e^{-\mathbf{i}\vec{k}\cdot\vec{x}}\mathbf{a}_k^\dagger\right), \\
\boldsymbol{\pi}(\vec{x}) &= \frac{1}{\mathbf{i}} \int \mathrm{đ}^d k \ \sqrt{\frac{\hbar\omega_k}{2}} \left(e^{\mathbf{i}\vec{k}\cdot\vec{x}}\mathbf{a}_k - e^{-\mathbf{i}\vec{k}\cdot\vec{x}}\mathbf{a}_k^\dagger\right),
\end{aligned} \tag{1.7}$$

satisfy the canonical commutation relation

$$[\boldsymbol{\phi}(\vec{x}), \boldsymbol{\pi}(\vec{x}')] = \mathbf{i}\hbar\mathbb{1}\delta^d(\vec{x}-\vec{x}').$$

I emphasize that this is really the same equation as our starting point for each ball on springs:

$$[\mathbf{q}_n, \mathbf{p}_{n'}] = \mathbf{i}\hbar\mathbb{1}\delta_{nn'}.$$

The mode expansions (1.7) contain a great deal of information. First notice that $\boldsymbol{\phi}$ is manifestly hermitian. Next, notice that from $\boldsymbol{\phi}(\vec{x}) \equiv \boldsymbol{\phi}(\vec{x}, 0)$ by itself we cannot disentangle $\mathbf{a}_k$ and $\mathbf{a}_k^\dagger$, since only the combination $\mathbf{a}_k + \mathbf{a}_{-k}^\dagger$ multiplies $e^{\mathbf{i}\vec{k}\cdot\vec{x}}$. The momentum $\boldsymbol{\pi}$ contains the other linear combination. However, if we evolve the field operator in time using the Heisenberg equation we find

$$\phi(\vec{x}, t) \equiv e^{\mathbf{i}\mathbf{H}t}\phi(\vec{x})e^{-\mathbf{i}\mathbf{H}t} = \int \mathrm{đ}^d k \sqrt{\frac{\hbar}{2\omega_k}} \left(e^{\mathbf{i}\vec{k}\cdot\vec{x} - \mathbf{i}\omega_{\vec{k}}t}\mathbf{a}_k + e^{-\mathbf{i}\vec{k}\cdot\vec{x} + \mathbf{i}\omega_{\vec{k}}t}\mathbf{a}_k^\dagger\right). \tag{1.8}$$

Indeed we can check that the relation $\boldsymbol{\pi} = \dot{\boldsymbol{\phi}}$ holds.

Notice that the dependence on spacetime is via a sum of terms of the form:

$$e^{\mathbf{i}\vec{k}\cdot\vec{x} - \mathbf{i}\omega_{\vec{k}}t} = e^{\mathbf{i}k_\mu x^\mu}|_{k^0=\omega_{\vec{k}}}$$

and their complex conjugates. These are precisely all the solutions to the wave equation (1.6). For each $\vec{k}$, there are two solutions, one with positive frequency and one with negative frequency. You might have worried that solutions with both signs of the frequency mean that the world might explode or something (like it would if we tried to replace the Schrödinger equation for the wavefunction with a Klein-Gordon equation). This danger is evaded in a beautiful way: the coefficient of the positive frequency solution with wavenumber $\vec{k}$ is the destruction operator for the mode; the associated negative frequency term comes with the creation operator for the same mode, as a consequence of reality of the field.

1.2 Fields mediate forces

[Zee §1.3] 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:

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

As you see, this is a quantity for which we have a path integral representation. 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.

[End of Lecture 1]

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

$$\mathcal{L}(\phi) = \frac{1}{2}\left(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2\right) \overset{\text{IBP}}{=} -\frac{1}{2}\phi\left(\partial^2 + m^2\right)\phi + \text{total derivative}. \tag{1.9}$$

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

$$-\left(\partial^2 + m^2\right)D(x-y) = \delta(x-y). \tag{1.10}$$

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? On the homework you saw an integral of the form $\int_{\mathbb{R}} dq e^{-\frac{1}{2}qAq}$, which 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[3].

The equation (1.10) is translation-invariant and linear so you should not be surprised that it is solved by going to Fourier space (in space and time):

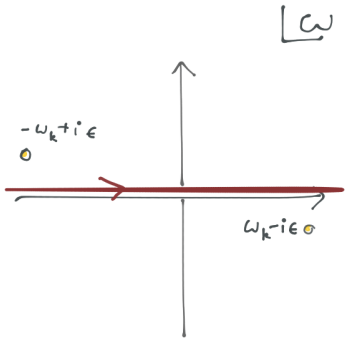
$$D(x) = \int \text{đ}^{d+1}k \; e^{\mathbf{i}k_\mu x^\mu} D_k, \quad \delta^{d+1}(x) = \int \text{đ}^{d+1}k \; e^{\mathbf{i}k_\mu x^\mu}.$$

in terms of which (1.10) becomes the algebraic equation $1 = (k^2 - m^2 + \mathbf{i}\epsilon)D_k$. Hence

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

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

$$D(x) = -\mathbf{i}\int \text{đ}^d k \left(\theta(t)\frac{e^{-\mathbf{i}(\omega_k t - \vec{k}\cdot\vec{x})}}{2\omega_k} + \theta(-t)\frac{e^{\mathbf{i}(\omega_k t - \vec{k}\cdot\vec{x})}}{2\omega_k}\right). \tag{1.11}$$

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

[3]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 the next subsection, we'll say more.

[4]We are using the Cauchy residue theorem $\oint_C dz f(z) = 2\pi\mathbf{i}\sum_{z_j} \text{Res}_{z=z_j} f$ 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 $\oint_{C_0} \frac{dz}{z} = \mathbf{i}\int_0^{2\pi} d\theta = 2\pi\mathbf{i}$.

The propagator. Who is $D(x)$, besides some quantity in terms of which we did a Gaussian integral? Recall from the homework that 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

$$D(x-y) \stackrel{?}{=} \langle 0|\phi(x)\phi(y)|0\rangle \tag{1.12}$$

– 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 (1.9) 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.)

Why the '?' in (1.12)? 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 (1.11) the thing to do is to *time-order* them:

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

To see this, plug in the mode expansion (1.7) to see *e.g.*

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

(where $k^0 = \omega_k, q^0 = \omega_q$ to satisfy the KG equation), which reproduces the first term in (1.11)[5][6].

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

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

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

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

[6]In comparing to (1.11), it helps to notice that we can redefine the $\vec{k}$ integration variable to reverse the sign of the exponent of the spatial part, $\int \mathrm{đ}^d k f(\vec{k}^2) e^{\mathbf{i}\vec{k}\cdot\vec{x}} = \int \mathrm{đ}^d k f(\vec{k}^2) e^{-\mathbf{i}\vec{k}\cdot\vec{x}}$. (Thanks to Hung-Hwa Lin for help during lecture.)

$\int dx^0 e^{-\mathbf{i}k^0x^0}\left(e^{\mathbf{i}\vec{k}\cdot\vec{x}_1}+e^{\mathbf{i}\vec{k}\cdot\vec{x}_2}\right)$. The interaction between the two lumps *mediated* by the mattress field ϕ will then be described by the J_1J_2 cross-terms in $W[J]$:

$$W[J] = -\frac{2}{2}\int dx^0 \int dy^0 \int \text{đ}k^0 e^{\mathbf{i}k^0(x^0-y^0)} \int \text{đ}^3k \frac{e^{\mathbf{i}\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{k^2-m^2+\mathbf{i}\epsilon} + ... \tag{1.13}$$

$$= -\int dx^0 \underbrace{\left(\int \text{đ}k^0 2\pi\delta(k^0)\right)}_{=1} \int \text{đ}^3k \frac{e^{\mathbf{i}\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{k^2-m^2+\mathbf{i}\epsilon} + ... \tag{1.14}$$

$$= +\int dx^0 \int \text{đ}^3k \frac{e^{\mathbf{i}\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}}{\vec{k}^2+m^2-\mathbf{i}\epsilon} + ... \tag{1.15}$$

(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 $e^{\mathbf{i}W} = \langle 0|\, e^{-\mathbf{i}HT}\,|0\rangle = e^{-\mathbf{i}E_{\text{gs}}(J)T}$, so $W = -E_{\text{gs}}(J)T$. We learn that

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

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[7]

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

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

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

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 7) is called the Yukawa potential.

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

$$\int \text{đ}^3k \frac{e^{\mathbf{i}\vec{k}\cdot\vec{x}}}{\vec{k}^2+M^2} \overset{y\equiv\cos\theta}{=} \frac{1}{(2\pi)^2}\int_0^\infty \frac{k^2dk}{k^2+M^2}\underbrace{\int_{-1}^1 dy e^{\mathbf{i}kyr}}_{=2\frac{\sin kr}{kr}} = \frac{1}{(2\pi)^2 r}\int_{-\infty}^\infty \frac{dk\,k\sin kr}{k^2+M^2}$$

$$= \frac{1}{(2\pi)^2 r}\left(\frac{1}{2\mathbf{i}}\int_{-\infty}^\infty dk \frac{ke^{\mathbf{i}kr}}{k^2+M^2} + h.c.\right) \quad \text{close contour in UHP for free}$$

$$= \frac{1}{(2\pi)^2 r}\frac{1}{2\mathbf{i}} 2\pi\mathbf{i}\frac{\mathbf{i}Me^{\mathbf{i}(\mathbf{i}M)r}}{2\mathbf{i}M}\cdot 2 = \frac{e^{-Mr}}{4\pi r}.$$

1.3 Euclidean path integral and Wick rotation

Here is a route to defining the path integral (actually the same as the replacement $m^2 \to m^2 - \mathbf{i}\epsilon$) which makes clearer what is going on with the initial and final states.

The whole point here can be made for a single mode of the field – a single harmonic oscillator – with action

$$S[q] = \frac{1}{2}\int dt \left((\partial_t q)^2 - \Omega^2 q^2\right) - \int Jq$$

(where $\Omega^2 = \vec{k}^2 + m^2$ if you like). Consider the replacement $\tau = \mathbf{i}t$ in the action:

$$S[q] = -\frac{1}{2}\mathbf{i}\int d\tau \left(-(\partial_\tau q)^2 - \Omega^2 q^2\right) + \mathbf{i}\int d\tau Jq = \mathbf{i}\int d\tau \left(\frac{1}{2}\left((\partial_\tau q)^2 + \Omega^2 q^2\right) + Jq\right).$$

With this replacement, the path integral becomes

$$\int [Dq] e^{-\int d\tau \left(\frac{1}{2}\left((\partial_\tau q)^2 + \Omega^2 q^2\right) + Jq\right)} \equiv \int [Dq] e^{-S_E[q]}.$$

This integrand suppresses configurations with large q, and large $\partial_\tau q$, and the integral is therefore totally well-defined. The euclidean action is[8]

$$S_E[q] = \int d\tau \left(\frac{1}{2}\left((\partial_\tau q)^2 + \Omega^2 q^2\right) + Jq\right) = \int d\tau \left(\frac{1}{2} q\left(-\partial_\tau^2 + \Omega^2\right) q + Jq\right)$$

where $(-\partial_\tau^2 + \Omega^2)$ is a positive operator (meaning all of its eigenvalues are positive). Call its inverse G, which then, by definition, satisfies

$$\left(-\partial_\sigma^2 + \Omega^2\right) G(\sigma, \tau) = \delta(\sigma - \tau)$$

The fact that our system is time-translation invariant means $G(\sigma, \tau) = G(\sigma - \tau)$. We can solve this equation in fourier space: $G(s) = \int đ\omega e^{i\omega\sigma} G_\omega$ makes it algebraic:

$$G_\omega = \frac{1}{\omega^2 + \Omega^2}$$

and we have

$$G(\tau) = \int đ\omega \frac{e^{\mathbf{i}\omega\tau}}{\omega^2 + \Omega^2} = e^{-\Omega|\tau|}\frac{1}{2\Omega}. \tag{1.16}$$

(Do it by residues: the integrand has poles at $\omega = \pm\mathbf{i}\Omega$ (see the figure 1 below). The absolute value of $|\tau|$ is crucial, and comes from the fact that the contour at infinity converges in the upper (lower) half plane for $\tau < 0$ ($\tau > 0$).)

[8] It is called euclidean because the $(\partial_\tau q)^2$ has the same sign as the spatial derivatives $(\partial_x q)^2$, so this is the action we get in euclidean spacetime with metric $\delta_{\mu\nu}$, rather than $\eta_{\mu\nu}$. Exercise: put back the spatial derivative terms and check that this is the case.

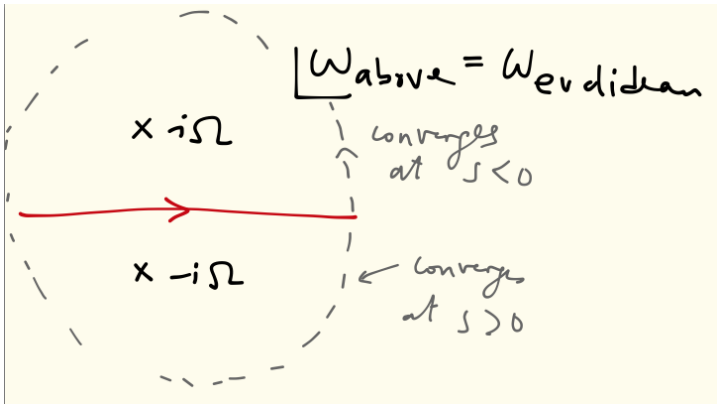
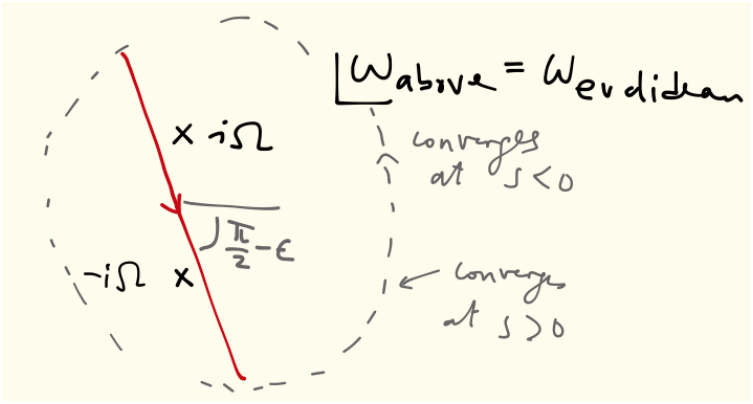


Figure 1: Poles of the integrand of the ω integral in (1.16).

I claim that the real-time calculation which keeps the oscillator in its groundstate is the *analytic continuation* of the one we did above, where we replace

$$\omega_{\rm Mink} = e^{-\mathbf{i}(\pi/2-\epsilon)}\omega_{\rm above} \tag{1.17}$$

where ϵ is (a familiar) infinitesimal. In the picture of the euclidean frequency plane in Fig. 1, this is a rotation by *nearly* 90 degrees. We don't want to go all the way to 90 degrees, because then we would hit the poles at $\pm\mathbf{i}\Omega$. The replacement (1.17) just means that if we integrate over *real* $\omega_{\rm Mink}$, we rotate the contour in the integral over ω as follows:



as a result we pick up the same poles at $\omega_{\rm above} = \pm\mathbf{i}\Omega$ as in the euclidean calculation. Notice that we had better also rotate the argument of the function, τ, at the same time to maintain convergence, that is:

$$\omega_{\rm eucl} = -\mathbf{i}\omega_{\rm Mink}, \quad \omega_{\rm eucl}t_{\rm eucl} = \omega_{\rm Mink}t_{\rm Mink}, \quad t_{\rm eucl} = +\mathbf{i}t_{\rm Mink}. \tag{1.18}$$

So this is giving us a contour prescription – a prescription for negotiating the poles – for the real-frequency integral. The result is the *Feynman* contour, and it is the same as

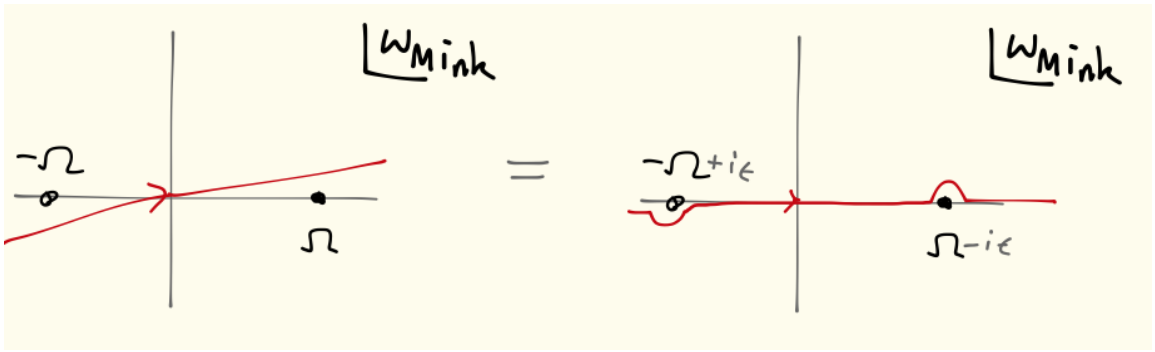


Figure 2: The Feynman contour in the $\omega_{\rm Mink}$ complex plane.

what we got from $m^2 \to m^2 - \mathbf{i}\epsilon$: depending on the sign of the (real) time separation of the two operators (recall that t is the difference), we close the contour around one pole or the other, giving the *time-ordered* propagator. For the case of a free scalar field, the replacement $m^2 \to m^2 - \mathbf{i}\epsilon$ had the same effect of rotating the real-frequency contour away from the poles. It is also the same as shifting the frequency by $\Omega \to \Omega - \mathbf{i}\epsilon$, as indicated in the right part of Fig. 2. This prescription works in a case where there is no m^2 term.

Notice for future reference that the euclidean action and real-time action are related by

$$S_{\rm eucl}[Q] = \int dt_{\rm eucl} \frac{1}{2}\left(\left(\frac{\partial Q}{\partial t_{\rm eucl}}\right)^2 + \Omega^2 Q^2\right) = -\mathbf{i} S_{\rm Mink}[Q] = -\mathbf{i}\int dt_{\rm Mink}\frac{1}{2}\left(\left(\frac{\partial Q}{\partial t_{\rm Mink}}\right)^2 - \Omega^2 Q^2\right) .$$

because of (1.18). This means the path integrand is $e^{-S_{\rm eucl}} = e^{\mathbf{i}S_{\rm Mink}}$.

[End of Lecture 2]

Euclidean evolution. Now, why does the contour coming from the euclidean path integral put the oscillator into its groundstate? The point in life of the euclidean time evolution to prepare the groundstate from an arbitrary state:

$$e^{-\beta \mathbf{H}}\,|\text{any}\rangle = \sum_n e^{-\beta E_n}|n\rangle\langle n|\text{any}\rangle \propto |\text{gs}\rangle + \mathcal{O}\left(e^{-\beta(E_1 - E_{\rm gs})}\right) \tag{1.19}$$

– the euclidean-time propagator $e^{-\beta \mathbf{H}}$ beats down the amplitude of any excited state relative to the groundstate, for large enough β.

And the euclidean path integral gives a formula for this euclidean propagation amplitude. Recall that the path integral representation for the real-time propagation

amplitude is

$$\langle f|\, e^{-\mathbf{i}\mathbf{H}t}\, |i\rangle = \int_{f\leftarrow i} [dq] e^{\mathbf{i}\int_0^t dtL}.$$

On the RHS here, we sum over all paths between i and f in time t (*i.e.* $q(0) = q_i, q(t) = q_f$), weighted by a phase $e^{\mathbf{i}\int dtL}$. But that means you also know a representation for

$$\sum_f \langle f|\, e^{-\beta\mathbf{H}}\, |f\rangle \equiv \mathrm{tr} e^{-\beta\mathbf{H}}$$

– namely, you sum over all *periodic* paths $q_i = q_f$ in *imaginary time* $t = -\mathbf{i}\beta$. So:

$$Z(\beta) = \mathrm{tr} e^{-\beta\mathbf{H}} = \oint [dq] e^{-\int_0^\beta d\tau L} = \oint [dq] e^{-S_{\mathrm{eucl}}[q]}$$

The LHS is the partition function in quantum statistical mechanics. The RHS is the euclidean functional integral we've been using. [For more on this, see Zee §V.2]

The period of imaginary time, $\beta \equiv 1/T$, is the inverse temperature. We've been studying the limit as $\beta \to \infty$. Taking $\beta \to \infty$ means $T \to 0$, and you'll agree that at $T = 0$ we project onto the groundstate (if there's more than one groundstate we have to think more).

Time-ordering. To summarize the previous discussion: in real time, we must choose a state, and this means that there are many Green's functions, not just one: $\langle\psi|\, q(t)q(s)\, |\psi\rangle$ depends on $|\psi\rangle$, unsurprisingly.

But we found a special one which arises by analytic continuation from the euclidean Green's function, which is unique[9]. It is

$$G(s,t) = \langle \mathcal{T}\,(q(s)q(t))\rangle \ ,$$

the time-ordered, or Feynman, Green's function, and I write the time-ordering symbol $\mathcal{T}$ to emphasize this. I emphasize that from our starting point above, the time ordering arose because we have to close the contour in the UHP (LHP) for $t < 0$ ($t > 0$).

[9] Another important perspective on the uniqueness of the euclidean Green's function and the non-uniqueness in real time: in euclidean time, we are inverting an operator of the form $-\partial_\tau^2 + \Omega^2$ which is *positive* (≡ all its eigenvalues are positive) – recall that $-\partial_\tau^2 = \hat{p}^2$ is the square of a hermitian operator. If all the eigenvalues are positive, the operator has no kernel, so it is completely and unambiguously invertible. This is why there are no poles on the axis of the (euclidean) ω integral in (1.16). In real time, in contrast, we are inverting something like $+\partial_t^2 + \Omega^2$ which annihilates modes with $\partial_t = \mathbf{i}\Omega$ (if we were doing QFT in $d > 0+1$ this equation would be the familiar $p^2 - m^2 = 0$). These are called 'on-shell states', they are actual states in the spectrum of the Hamiltonian of the system. So the operator we are trying to invert has a kernel and this is the source of the ambiguity. In frequency space, this is reflected in the presence of poles of the integrand *on* the contour of integration; the choice of how to negotiate them encodes the choice of Green's function.

Let's pursue this one more step. The same argument tells us that the generating functional for real-time, time-ordered correlation functions of q is

$$Z[J] = \left\langle \mathcal{T} e^{\mathbf{i}\int dt q(t) J(t)} \right\rangle = \langle 0| \mathcal{T} e^{\mathbf{i}\int Jq} |0\rangle \,, \tag{1.20}$$

in the sense that

$$\langle \mathcal{T} q(t_1) q(t_2) ... \rangle = \frac{1}{Z} \frac{\delta}{\delta J(t_1)} \frac{\delta}{\delta J(t_2)} ... Z[J]|_{J=0} \,.$$

In the second step of (1.20) I just emphasized that the real time expectation value here is really a *vacuum* expectation value, as long as we use the $\mathbf{i}\epsilon$ prescription above to define the integrals. This quantity has the picturesque interpretation as the *vacuum persistence amplitude*, in the presence of the source J.[10]

So we see that in general, the correlation functions that are computed by this "$\mathbf{i}\epsilon$ prescription" of Wick rotating from Euclidean spacetime are time-ordered vacuum expectation values:

$$\frac{1}{Z} \int [D\phi] e^{\mathbf{i} S_{m^2 \to m^2 - \mathbf{i}\epsilon}} f(\phi) = \langle 0 | \mathcal{T} f(\phi) | 0 \rangle \,.$$

Causality. In other treatments of this subject, you will see the Feynman contour motivated by ideas about causality. This was not the logic of our discussion here but it is reassuring that we end up in the same place. Note that even in 0+1 dimensions there is a useful notion of causality: effects should come after their causes.

1.4 Feynman diagrams from the path integral

Subsection §1.3 was a sophisticated discussion of QFT in $0+1$ dimensions (*i.e.*ordinary quantum mechanics of a single particle), since we focussed on a single mode. To attempt to demystify some more of the structure we'll discover in QFT, let's regress even further, and consider the case of QFT in $0+0$ dimensions. By the path-integral representation, this means ordinary integrals. If everything is positive, this is probability theory.

Suppose we want to do the integral

$$Z(J) = \int_{-\infty}^{\infty} dq \; e^{-\frac{1}{2}m^2 q^2 - \frac{g}{4!}q^4 + Jq} \equiv \int dq \; e^{-S(q)} \,. \tag{1.21}$$

[10] Actually, more useful is the generating function of *connected* correlation functions:

$$\langle \mathcal{T} q(t_1) q(t_2) ... \rangle_c = \frac{\delta}{\delta J(t_1)} \frac{\delta}{\delta J(t_2)} ... log Z[J]|_{J=0} \,.$$

where $\langle q_1 q_2 \rangle_c \equiv \langle q_1 q_2 \rangle - \langle q_1 \rangle \langle q_2 \rangle$. Higher-point connected correlation functions are defined by subtracting the gaussian answer. Connected correlation functions are well-named because they are computed by connected Feynman diagrams, as we'll discuss more next.

It is the path integral for ϕ^4 theory with fewer labels. For $g = 0$, this is a gaussian integral which we know how to do. For $g \neq 0$ it's not an elementary function of its arguments. We can develop a (non-convergent!) series expansion in g by writing it as

$$Z(J) = \int_{-\infty}^{\infty} dq\ e^{-\frac{1}{2}m^2q^2+Jq}\left(1 - \frac{g}{4!}q^4 + \frac{1}{2!}\left(-\frac{g}{4!}q^4\right)^2 + \cdots\right)$$

and integrating term by term. And the term with q^{4n} (that is, the coefficient of $\frac{1}{n!}\left(\frac{-g}{4!}\right)^n$) is

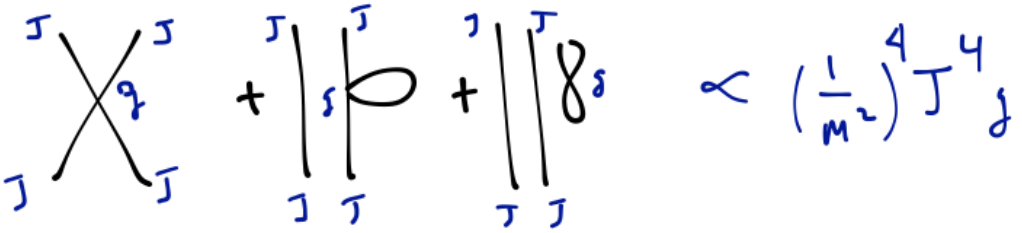
$$\int_{-\infty}^{\infty} dq\ e^{-\frac{1}{2}m^2q^2+Jq}q^{4n} = \left(\frac{\partial}{\partial J}\right)^{4n}\int_{-\infty}^{\infty} dq\ e^{-\frac{1}{2}m^2q^2+Jq} = \left(\frac{\partial}{\partial J}\right)^{4n} e^{\frac{1}{2}J\frac{1}{m^2}J}\sqrt{\frac{2\pi}{m^2}}\ .$$

So:

$$Z(J) = \sqrt{\frac{2\pi}{m^2}}e^{-\frac{g}{4!}\left(\frac{\partial}{\partial J}\right)^4}e^{\frac{1}{2}J\frac{1}{m^2}J}\ .$$

This is a double expansion in powers of J and powers of g. The process of computing the coefficient of $J^n g^m$ can be described usefully in terms of diagrams. There is a factor of $1/m^2$ for each line (the propagator), and a factor of $(-g)$ for each 4-point vertex (the coupling), and a factor of J for each external line (the source). For example, the coefficient of gJ^4 comes from:

$$\sim \left(\frac{1}{m^2}\right)^4 gJ^4.$$

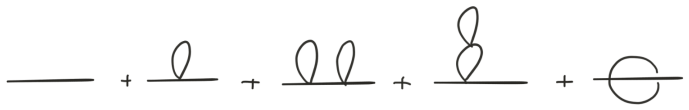


There is a symmetry factor which comes from expanding the exponential: if the diagram has some symmetry preserving the external labels, the multiplicity of diagrams does not completely cancel the $1/n!$.

As another example, consider the analog of the two-point function:

$$G \equiv \langle q^2 \rangle\,|_{J=0} = \frac{\int dq\ q^2\ e^{-S(q)}}{\int dq\ e^{-S(q)}} = -2\frac{\partial}{\partial m^2}\log Z(J=0). \tag{1.22}$$

In perturbation theory this is:

$G \simeq$  $+\ \mathcal{O}(g^3)$

$$= m^{-2}\left(1 \quad -\frac{1}{2}gm^{-4} \quad + \quad \frac{2}{3}g^2m^{-8} \quad + \mathcal{O}(g^3)\right) \tag{1.23}$$

To get the numerical coefficients note that Wick's theorem for this simple case is

$$\langle q^k\rangle_0 = \begin{cases} 0, & k \text{ odd} \\ (k-1)!!, & k\text{even} \end{cases} \tag{1.24}$$

– the number of ways of pairing k objects. Here $(k-1)!! \equiv (k-1)(k-3)(k-5)...\cdot 3\cdot 1$. This is because there are $k-1$ choices of partner for the first q, after which there are $k-3$ choices of partner for the next one, etc.

Some important structural comments: A diagram contributing to G which has any part not connected to the external legs is cancelled by the expansion of the denominator $Z = \int dq e^{-S(q)}$. The contributions to Z are called 'vacuum diagrams' (since they have no external lines, so they are like an amplitude for nothing to turn back into nothing). Z is a sum over *all* diagrams with no external lines, including disconnected ones. As you saw in 215A, this sum exponentiates: $Z = e^{\sum(\text{ connected diagrams })}$.

Some labels. Some of these points are clearer if we put back some of the labels. So consider the slightly more complicated case

$$Z = \int \prod_{i=1}^N dq_i e^{-S(q)}, \quad S(q) = \frac{1}{2}q_i A_{ij} q_j + \frac{g}{4!}\sum_i q_i^4 \equiv S_0 + \frac{g}{4!}\sum_i q_i^4.$$

(Think of i as like a position index, and A as a difference operator, so this is a discretization of ϕ^4 theory.) Then we can develop a perturbative expansion by writing

$$\langle q_1\cdots q_k\rangle = \frac{\int \prod_{i=1}^N dq_i e^{-S(q)} q_1\cdots q_k}{Z} \tag{1.25}$$

$$= \frac{\int \prod_{i=1}^N dq_i e^{-S_0(q)} e^{-\frac{g}{4!}\sum_i q_i^4} q_1\cdots q_k}{\int \prod_{i=1}^N dq_i e^{-S_0(q)} e^{-\frac{g}{4!}\sum_i q_i^4}} \tag{1.26}$$

$$\simeq \frac{\sum_{n=0}\left(-\frac{g}{4!}\right)^n/n! \int dq e^{-S_0}\sum_{i_1,\ldots,i_n}\prod_{i_a}^n q_{i_a}^4 q_1\cdots q_k}{\sum_{n=0}\left(-\frac{g}{4!}\right)^n/n! \int dq e^{-S_0}\sum_{i_1,\ldots,i_n}\prod_{i_a}^n q_{i_a}^4}. \tag{1.27}$$

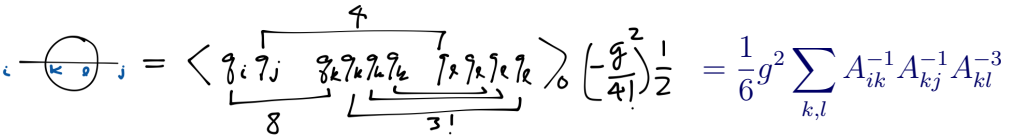
(Note the step with the $\simeq$ is where we exchange the sum over n with the integral over q.)

Now the general statement of Wick's theorem is:

$$\langle q_1\cdots q_k\rangle_0 = \begin{cases} 0, & k \text{ odd} \\ \sum_{\text{contractions}} A^{-1}_{i_1 i_2}\cdots A^{-1}_{i_{k/2-1} i_{k/2}}, & k\text{even} \end{cases}.$$

Here the sum is over all ways of pairing up the k fields. (Note that this reduces to (1.24) if we remove all the labels.)

Let's think about a particular term, *e.g.* an $\mathcal{O}(g^2)$ contribution to $\langle q_i q_j \rangle$:

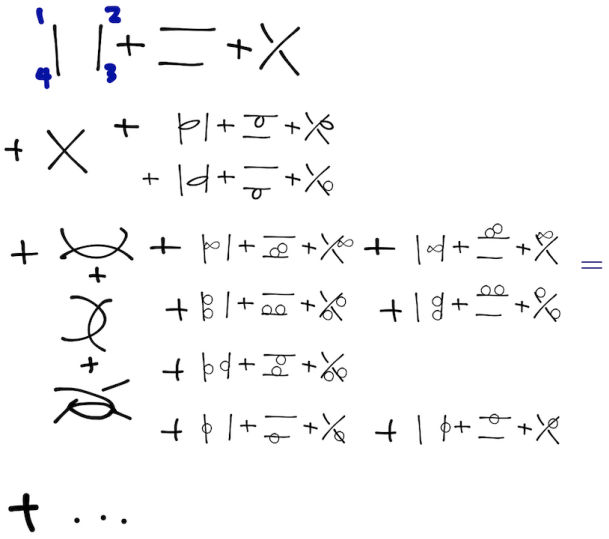


$$= \frac{1}{6} g^2 \sum_{k,l} A^{-1}_{ik} A^{-1}_{kj} A^{-3}_{kl}$$

The factor of $1/6$ is called a *symmetry factor*. Instead of by explicit combinatorics, we could have gotten this number by dividing by the order of the automorphism group of the diagram. An automorphism of the diagram is a map from the diagram to itself which preserves the external lines and the connectivity. In this diagram, we can permute the three internal lines amongst themselves, giving $|S_3| = 3! = 6$ in the denominator. *Don't get hung up on the symmetry factors.*

As a final example for now, here is the expansion of the four-point function:

$$\langle q_1 q_2 q_3 q_4 \rangle \simeq$$



$$= A^{-1}_{14} A^{-1}_{23} + A^{-1}_{12} A^{-1}_{34} + A^{-1}_{13} A^{-1}_{24}$$
$$- g \sum_i \left(A^{-1}_{1i} A^{-1}_{2i} A^{-1}_{3i} A^{-1}_{4i} + (A^{-1}_{1i} A^{-1}_{i4} A^{-1}_{23} + \cdots) \right)$$
$$+ \frac{g^2}{2!} \sum_{ij} \left((\tfrac{1}{2} A^{-1}_{1i} A^{-1}_{4i} A^{-2}_{ij} A^{-1}_{j2} A^{-1}_{j3} + \cdots) + (\tfrac{1}{2} A^{-1}_{1i} A^{-2}_{ij} A^{-1}_{jj} A^{-1}_{4i} A^{-1}_{32} + \cdots) \right)$$
$$+ \cdots$$

[End of Lecture 3]

All the labels. The Feynman diagrams we're going to draw all the time are the same but with more labels. Notice that each of the qs in our integral could come with a label, $q \to q_a$. Then each line in our diagram would be associated with a matrix $(m^{-2})_{ab}$ which is the inverse of the quadratic term $q_a m^2_{ab} q_b$ in the action. If our diagrams have loops we get free sums over the label. If that label is conserved by the interactions, the vertices will have some delta functions. In the case of translation-invariant field theories we can label lines by the conserved momentum k. Each comes with a factor of the free propagator $\frac{\mathbf{i}}{k^2+m^2+\mathbf{i}\epsilon}$, each vertex conserves momentum, so

comes with $\mathbf{i}g\delta^D\left(\sum k\right)(2\pi)^D$, and we must integrate over momenta on internal lines $\int \mathrm{d}^D k$.

Brief comments about large orders of perturbation theory.

- The perturbation series about $g = 0$ does *not* converge. How do I know? One way to see this is to notice that if I made g even infinitesimally negative, the integral itself would not converge (the potential would be unbounded below), and $Z_{g=-|\epsilon|}$ is not defined. Therefore Z_g as a function of g cannot be analytic in a neighborhood of $g = 0$. This argument is due to Dyson, and applies also in most QFTs. This means there is more to QFT than perturbation theory: the perturbation series does not define the field theory amplitudes.

- The expansion of the exponential in the integrand is clearly convergent for each q. The place where we went wrong is exchanging the order of integration over q and summation over n.

- In this case, the perturbation expansion can be given a closed form expression:

$$Z(0) \simeq \sqrt{\frac{2\pi}{m^2}} \sum_n \frac{(-1)^n}{n!} \frac{2^{2n+\frac{1}{2}}}{(4!)^n} \Gamma\left(2n+\frac{1}{2}\right) \left(\frac{g}{m^4}\right)^n . \tag{1.28}$$

- The expansion for G is of the form

$$G \simeq m^{-2} \sum_{n=0}^{\infty} c_n \left(\frac{g}{m^4}\right)^n .$$

  When n is large, the coefficients satisfy $c_{n+1} \overset{n\gg 1}{\simeq} -\frac{2}{3} n c_n$ (you can see this by looking at the coefficients in (1.28)) so that $|c_n| \sim n!$. This factorial growth of the number of diagrams is general in QFT and is another way to see that the series does not converge.

- The fact that the coefficients c_n grow means that there is a *best* number of orders to keep. The errors start getting bigger when $c_{n+1}\left(\frac{g}{m^4}\right) \sim c_n$, that is, at order $n \sim \frac{3m^4}{2g}$. So if you want to evaluate G at this value of the coupling, you should stop at that order of n. An estimate of the error by the next term left out gives something that goes like $e^{-\#/g}$.

- I said above that the fact that the perturbation series doesn't converge means that it doesn't define the field theory amplitudes. What does it miss? To answer

this, consider trying to do the integral (1.21) by saddle point (at $J = 0$ for simplicity):

$$0 = S'(q_\star) = m^2 q_\star + \frac{g}{3!} q_\star^3.$$

(Note the resemblance to the equations of motion.) This has *three* solutions:

$$q_\star = 0, \quad q_\star = \pm \mathbf{i} \sqrt{\frac{3! m^2}{g}}.$$

The expansion about the 'trivial' saddle at $q_\star$ (where the action is $S(q_\star = 0) = 0$) reproduces the perturbation series. At the other saddles,

$$S\left(q_\star = \pm \mathbf{i} \sqrt{\frac{3! m^2}{g}}\right) = -\frac{3m^4}{2g}, \tag{1.29}$$

which means their contribution would go like $e^{+\frac{3m^4}{2g}}$, which actually would blow up at weak coupling, $g \to 0$. These saddles are not on the contour and don't contribute for small positive g, but more generally (as for example when $m^2 < 0$), there will be effects that go like $e^{-\frac{a}{|g|}}$. This is a function whose series expansion in g at $g = 0$ is identically zero. You can never find it by doing perturbation theory in g about $g = 0$.

- A technique called *Borel resummation* can sometimes produce a well-defined function of g from an asymptotic series whose coefficients diverge like $n!$. The idea is to make a new series

$$B(z) \equiv \sum_{n=0} \frac{c_n}{n!} z^n$$

whose coefficients are ensmallened by $n!$. Then to get back $Z(g)$ we use the identity

$$1 = \frac{1}{n!} \int_0^\infty dz e^{-z} z^n$$

and do the Laplace transform of $B(z)$:

$$\int_0^\infty dz B(z) e^{-z/g} = \sum_{m=0} c_m \frac{\int_0^\infty dz e^{-z/g} z^m}{m!} = g \sum_{m=0}^\infty c_m g^m = g Z(g).$$

This procedure requires both that the series in $B(z)$ converges and that the Laplace transform can be done. In fact this procedure works in this case.

The fact that the number of diagrams at large order grows like $n!$ is correlated with the existence of saddle-point contributions to $Z(g)$ which go like $e^{-a/g}$.

This is because they are associated with singularities of $B(z)$ at $z = a$; such a singularity means the sum of $\frac{c_n}{n!}z^n$ must diverge at $z = a$. (More generally, non-perturbative effects which go like $e^{-a/g^{1/p}}$ (larger if $p > 1$) are associated with (faster) growth like $(pn)!$. In string theory, $p = 2$. See this classic work.)

- In fact in this case, we know the whole function. The integral actually does have a name – it's a Bessel function:

$$Z(J=0) = \frac{2}{\sqrt{m^2}}\sqrt{\rho}e^{\rho}K_{\frac{1}{4}}(\rho), \qquad \rho \equiv \frac{3m^4}{4g}$$

  (for $\mathrm{Re}\sqrt{\rho} > 0$), as `Mathematica` will tell you. Because we know about Bessel functions, in this case we can actually figure out what happens at strong coupling, when $g \gg m^4$, using the asymptotics of the Bessel function.

- The functions $G(g)$ and $Z(g)$ can be analytically continued in g away from the real axis, and can in fact be defined on the whole complex g plane. It has a branch cut on the negative real axis, across which its discontinuity is related to its imaginary part. The imaginary part goes like $e^{-\frac{a}{|g|}}$ near the origin and can be computed by a tunneling calculation like (1.29).

  How did we know Z has a branch cut? One way is from the asymptotics of the Bessel function. But, better, *why* does Z satisfy the Bessel differential equation as a function of the couplings? The answer, as you'll check on the homework, is that the Bessel equation is a Schwinger-Dyson equation,

$$0 = \int_{-\infty}^{\infty} \frac{\partial}{\partial q}\left(\text{something } e^{-S(q)}\right) \tag{1.30}$$

  which results from demanding that we can change integration variables in the path integral.

For a bit more about this, you might look at sections 3 and 4 of this recent paper from which I got some of the details here. See also the giant book by Zinn-Justin. There is a deep connection between the large-order behavior of the perturbation series about the trivial saddle point and the contributions of non-trivial saddle points. The keywords for this connection are *resurgence* and *trans-series* and a starting reference is here.

The class of equations (1.30) is very important: it shows that the equations of motion are true in Green's functions, up to *contact terms*, the contributions where the

$\frac{\partial}{\partial q}$ hits the 'something'. For example, in scalar field theory

$$0 = \int [D\phi] \frac{\delta}{\delta\phi(y)} \left(\phi(x) e^{\mathbf{i}S[\phi]} \right) = \left\langle \phi(x) \mathbf{i} \frac{\delta S}{\delta\phi(y)} \right\rangle + \delta^{d+1}(x-y).$$

In the special case where S is quadratic, $S = \int \phi A \phi$, this shows that the two-point function is a Green's function for the quadratic operator A.

1.5 Lagrangian field theory

[Here we fill in the bits of Peskin §2.2 that we missed above.] Let's consider a classical field theory in the Lagrangian description. This means that the degrees of freedom are a set of fields $\phi_r(x)$, where r is a discrete index (for maybe spin or polarization or flavor), and we specify the dynamics by the classical action. If the world is kind to us (in this class we assume this), the action is an integral over space and time of a Lagrangian density

$$S[\phi] \equiv \int d^{d+1}x \mathcal{L}(\phi, \partial^\mu \phi).$$

This important assumption is an implementation of *locality*.

This central object encodes the field equations, the canonical structure on the phase space, the Hamiltonian, the symmetries of the theory.

I've sneakily implied that we are going to assume Lorentz invariance, so that $\mathcal{L}$ depends on the 4-vector $\partial^\mu \phi$, and not its components separately. I am also going to assume that the action S is real.

Two examples to keep in mind are the Klein-Gordon Lagrangian:

$$\mathcal{L}_{KG} = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - \frac{1}{2}m^2\phi^2$$

and the Maxwell Lagrangian:

$$\mathcal{L}_{EM} = -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} = \frac{1}{4e^2}\left(E^2 - B^2\right)$$

with $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ and A_μ regarded as the independent degrees of freedom.

A word about units: in units with $\hbar = c = 1$, everything has units of mass to some power, called its mass dimension. Energy and momentum $p_\mu = \hbar k_\mu$ have mass dimension $+1$. The space and time coordinates x^μ have mass dimension -1. The action goes in the exponential of the path integral measure $\int [D\phi] e^{\frac{\mathbf{i}S}{\hbar}}$ and so must be dimensionless. So the Lagrangian density has mass dimension $d+1$. This means that the KG field has mass dimension $\frac{d-1}{2}$ (and the mass m has mass dimension 1 (yay!)). In $d+1 = 3+1$ dimensions, $E \sim \dot{A}, B \sim \vec{\nabla} A$ have mass dimension 2 and A has mass dimension one (and e is dimensionless). This is nice because then the covariant derivative $\partial_\mu + A_\mu$ has mass dimension one. Notice that $E^2 + B^2$ has dimension 4 which is good for an energy per unit volume.

object	mass dim.
p_μ	1
x^μ	-1
S	0
$\mathcal{L}$	$d+1$
ϕ	$\frac{d-1}{2}$
A_μ	1
$\vec{E}, \vec{B}, F_{\mu\nu}$	2

The equation of motion is

$$0 = \frac{\delta S}{\delta \phi_r(x)}.$$

Note the functional derivative. You can check that in the case when $\mathcal{L}$ depends only on ϕ and $\partial_\mu\phi$, this is the same as the Lagrange EOM

$$0 = \frac{\partial \mathcal{L}}{\partial \phi_r} - \partial^\mu \frac{\partial \mathcal{L}}{\partial(\partial^\mu \phi_r)}$$

(for each r) which I can't remember. Note that since we are interested here in the bulk equations of motion, we ignore boundary terms unless we are interested in field theory on a space with boundary. That is a worthy subject but an unnecessary complication for now.

By redefining the field by *e.g.* $\phi \equiv \frac{1}{D}(\chi - B/C)$, we can make the KG theory uglier

$$\mathcal{L} = A + B\chi + \frac{1}{2}C\chi^2 + \frac{1}{2}D\partial^\mu\chi\partial_\mu\chi.$$

From the path integral point of view, the field is just an integration variable. Sometimes, its normalization is meaningful, like in the phonon example where it began its life as the displacement of the atoms from their equilibrium. So you see that relative to the most general possible Lagrange density for a scalar field, we are not losing generality except in our neglect of interactions, and in our neglect of terms with more derivatives. The former neglect we will repair little by little in this course, by doing perturbation theory. The latter is justified well by the renormalization group philosophy, which is a subject for later.

Canonical field momentum and Hamiltonian. The Hamiltonian viewpoint in field theory has the great virtue of bringing out the physical degrees of freedom. It has the great shortcoming that it picks out the time coordinate as special and obscures Lorentz symmetry.

The canonical field momentum is defined to be

$$\pi(x) = \frac{\partial \mathcal{L}}{\partial(\partial_t \phi(x))}.$$

Notice that this expression assumes a local Lagrangian density. π is actually a 'field momentum density' in the sense that the literal canonical momentum is $\frac{\partial}{\partial\dot\phi(x)}L = d^dx\pi(x)$ (as opposed to $\mathcal{L}$). I will often forget to say 'density' here.

The hamiltonian is then

$$H = \sum_n p_n q_n - L = \int d^dx \left(\pi(x)\dot\phi(x) - \mathcal{L}\right) \equiv \int d^dx\ \mathfrak{h}.$$

Noether's theorem and the Noether method. Yay, symmetries. Why do physicists love symmetries so much? One reason is that they offer possible resting

places along our never-ending chains of 'why?' questions. For example, one answer to the question "Why QFT?" is (certainly this is the one given in Weinberg's text, but just as certainly it is not the only one): quantum mechanics plus Poincaré symmetry.

They are also helpful for solving physical systems: Continuous symmetries are associated with conserved currents. Suppose the action is invariant under a continuous transformation of the fields ϕ, $\phi(x) \mapsto \phi'(x)$. (The invariance of the action is what makes the transformation a symmetry.) 'continuous' here means we can do the transformation just a little bit, so that $\phi(x) \mapsto \phi(x) + \epsilon\Delta\phi(x)$ where ϵ is an infinitesimal parameter.

If the transformation with constant ϵ (independent of space and time) is a symmetry, then the variation of the action with $\epsilon = \epsilon(x,t)$ must be proportional to $\partial_\mu\epsilon$ (at least assuming some smoothness properties of the action), and so that it vanishes $\forall\phi$ when ϵ is constant:

$$S[\phi + \epsilon(x)\Delta\phi] - S[\phi] = \int d^d x dt \partial_\mu \epsilon(x) j^\mu \overset{\text{IBP}}{=} -\int d^d x dt \epsilon(x) \partial_\mu j^\mu \ .$$

But if the equations of motion are obeyed, then the action is invariant under *any* variation of ϕ, including this one, for arbitrary $\epsilon(x)$. But this means that $\partial_\mu j^\mu = 0$, the current is conserved. These words are an accurate description of the equation because they mean that the charge

$$Q_R \equiv \int_R d^d x \ j^0$$

in some region of space R can only change by leaving the region (assume the definition of R is independent of time):

$$\partial_t Q_R = \int_R d^d x \ \partial_t j^0 = -\int_R d^d x \ \vec{\nabla}\cdot\vec{j} = -\int_{\partial R} d^{d-1} x \hat{n}\cdot\vec{j}$$

where in the last step we used Stokes' theorem.

This trick with pretending the parameter depends on space is called the *Noether method*. More prosaically, the condition that the action is invariant means that the Lagrangian density changes by a total derivative (we assume boundary terms in the action can be ignored):

$$\mathcal{L}(\phi', \partial_\mu\phi') \overset{\text{symmetry}}{=} \mathcal{L}(\phi, \partial_\mu\phi) + \epsilon\partial_\mu\mathcal{J}^\mu$$

but on the other hand, by Taylor expansion,

$$\mathcal{L}(\phi', \partial_\mu\phi') \overset{\text{calculus}}{=} \mathcal{L}(\phi, \partial_\mu\phi) + \epsilon\left(\frac{\partial\mathcal{L}}{\partial\phi}\Delta\phi + \frac{\partial\mathcal{L}}{\partial\left(\partial_\mu\phi\right)}\partial_\mu\Delta\phi\right)$$

$$\overset{\text{IBP}}{=} \mathcal{L}(\phi, \partial_\mu \phi) + \epsilon \left(\underbrace{\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)}}_{eom} \right) \Delta\phi + \epsilon \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \Delta\phi \right) \ .$$

By combining the previous two equations for $\mathcal{L}(\phi')$, we see that on configurations which satisfy the EOM, $0 = \partial_\mu j^\mu$ with

$$j^\mu = \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_r)} \Delta\phi_r - \mathcal{J}^\mu. \tag{1.31}$$

Notice that I stuck the index back in at the last step.

There is a converse to the Noether theorem, which is easier to discuss directly in quantum mechanics. Given a conserved charge Q, that is, a hermitian operator with $[H, Q] = 0$, we can make a symmetry transformation of the fields ϕ by

$$\delta\phi \equiv \mathbf{i}\epsilon[Q, \phi]. \tag{1.32}$$

We'll say that Q *generates* the symmetry, for the following reason. (1.32) is the infinitesimal version of the finite transformation

$$\phi \to \phi' \equiv e^{\mathbf{i}\epsilon Q} \phi e^{-\mathbf{i}\epsilon Q}.$$

The object $\mathbf{U} \equiv e^{\mathbf{i}\epsilon Q}$ is a unitary operator (since Q is hermitian) which represents the action of the symmetry on the Hilbert space of the QFT. It is a symmetry in the sense that it commutes with the time evolution operator $e^{-\mathbf{i}Ht}$.

Some examples will be useful:

- For example, suppose $S[\phi]$ only depends on ϕ through its derivatives, for example, $S[\phi] = \int \frac{1}{2} \partial_\mu \phi \partial^\mu \phi$. Then there is a shift symmetry $\phi \to \phi' \equiv \phi + \epsilon$. Letting ϵ depend on spacetime, the variation of the action is $S[\phi + \epsilon(x)] - S[\phi] = -\int \epsilon \partial_\mu \partial^\mu \phi$, so the current is $j_\mu = \partial_\mu \phi$. Let's check the converse: Indeed, the charge $Q = \int_{\text{space}} j_0$ generates the symmetry in the sense that for small ϵ, the variation in the field is

  $$\delta\phi \equiv \phi' - \phi = \epsilon = \mathbf{i}\epsilon[Q, \phi]$$

  (if we were doing classical mechanics, we should replace $\mathbf{i}[Q, \phi]$ with the Poisson bracket). Using our expression for the current this is

  $$\delta\phi = \mathbf{i}\epsilon \big[\int d^d y \underbrace{\dot{\phi}(y)}_{=\pi(y)}, \phi(x) \big] = \epsilon$$

  which is indeed true by the canonical commutation relations. In this case the finite transformation is again $\phi \to \phi + \epsilon$.

- Consider a complex scalar Φ, and suppose $S[\Phi, \Phi^\star]$ is invariant under $\Phi \to e^{\mathbf{i}\epsilon}\Phi = \Phi + \mathbf{i}\epsilon\Phi + \mathcal{O}(\epsilon^2)$, such as $S = \int (\partial\Phi^\star\partial\Phi - V(\Phi^\star\Phi))$. This $\mathsf{U}(1)$ phase transformation can be rewritten in terms of the real and imaginary parts as an $\mathsf{SO}(2)$ rotation. The charge can be written as

$$Q = \int d^d x j^0 = \int \mathrm{đ}^d p \left(\mathbf{a}_p^\dagger \mathbf{a}_p - \mathbf{b}_p^\dagger \mathbf{b}_p\right)$$

where the two sets of creation and annihilation operators are associated with excitations of Φ and $\Phi^\dagger$ respectively. (That is, quantize $\phi_{1,2}$ as we did for a single real scalar field, in terms of mode operators $\mathbf{a}_{1,2}$ respectively. Then let $\mathbf{a} \equiv \mathbf{a}_1 + \mathbf{i}\mathbf{a}_2, \mathbf{b} \equiv \mathbf{a}_1 - \mathbf{i}\mathbf{a}_2$, up to numerical prefactors.) So the particles created by $\mathbf{a}$ and $\mathbf{b}$ have opposite charge (this follows given the mode expansion $\Phi_k \sim \mathbf{a}_k + \mathbf{b}_{-k}^\dagger$) and can be interpreted as each others' antiparticles: there can be symmetry-respecting processes where an $\mathbf{a}$ particle and $\mathbf{b}$ particle take each other out.

[End of Lecture 4]

The previous two examples are related. Consider the case where $V(\Phi^\star\Phi) = \lambda(\Phi^\star\Phi - v^2)^2$ Changing variables to polar coordinates in field space, $\Phi = \rho e^{\mathbf{i}\phi}$, the Lagrangian is

$$\mathcal{L} = \rho^2(\partial\phi)^2 + (\partial\rho)^2 - \lambda(\rho^2 - v^2)^2.$$

If λ is big, the potential forces $\rho = v$, and its fluctuations are heavy, and we are left with $\mathcal{L} = v^2(\partial\phi)^2$, where $\phi \to \phi + \epsilon$ is a symmetry. Notice that $\phi \equiv \phi + 2\pi$ is periodic.

- Consider spacetime translations, $x^\mu \to x^\mu - a^\mu$. We can think of this as a transformation of the fields by

$$\phi(x) \mapsto \phi(x + a) = \phi(x) + a^\nu \underbrace{\partial_\nu \phi}_{\equiv \Delta_\nu \phi} + \mathcal{O}(a^2).$$

Our transformation parameter is now itself a four-vector, so we'll get a four-vector of currents T^μ_ν. This will be a symmetry as long as the lagrangian doesn't depend explicitly on space and time (so $\partial_\nu \mathcal{L} = 0$) but rather depends on space and time only via the fields (so $0 \neq \frac{d}{dx^\nu}\mathcal{L} \overset{\text{chain rule}}{=} \partial_\nu\phi\frac{\partial\mathcal{L}}{\partial\phi} + \partial_\mu\partial_\nu\phi\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}$). Let's use the prosaic method for this one: the shift in the Lagrangian density also can be found by Taylor expansion

$$\mathcal{L} \mapsto \mathcal{L} + a^\mu \frac{d}{dx^\mu}\mathcal{L} = \mathcal{L} + a^\nu \partial_\mu \left(\delta^\mu_\nu \mathcal{L}\right).$$

So the formula (1.31) gives

$$T^\mu_\nu = \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \phi\right)} \underbrace{\partial_\nu \phi}_{\Delta_\nu \phi} - \mathcal{L}\delta^\mu_\nu.$$

For the time translation, the conserved charge T^0_0 gives back the hamiltonian density $\mathfrak{h} = \pi\dot{\phi} - \mathcal{L}$ obtained by Legendre transformation. The conserved quantity from spatial translations is the momentum carried by the field, which for the KG field is

$$\mathbf{P}_i = \int d^d x \; T^0_i = -\int d^d x \; \pi \partial_i \phi.$$

For the Maxwell field, this gives the Poynting vector.

There is some ambiguity in the definition of the stress tensor (associated with the possibility of adding total derivatives to $\mathcal{L}$).

Let's check that the expression above for the conserved momentum agrees with our expectations. In particular, in free field theory the total momentum of the state $\left|\vec{k}_1, \cdots \vec{k}_n\right\rangle$ should be just the sum of the momenta of the particles, $\vec{P} = \sum_{\ell=1}^n \hbar \vec{k}_\ell$ (with interactions the story can be more complicated). Indeed

$$\mathbf{P}_i = -\int d^d x \; \pi \partial_i \phi = \int \mathrm{đ}^d k k_i \mathbf{a}^\dagger_{\vec{k}} \mathbf{a}_{\vec{k}}$$

agrees with this. (Notice that I used rotation invariance of the vacuum to not worry about a possible constant term.)

- I have the impression that you learned all about the rest of the Poincaré group already in 215A.

2 From correlation functions to the S matrix

We've derived an expression for correlation functions, such as (1.22), in terms of a sum of diagrams connected to the external lines, ordered by the number of powers of the coupling constant. Our next goal is to organize this sum.

First let's make contact with the way the diagrammatic expansion was (I think) introduced in 215A. A time-ordered real-time Green's function has path integral representation (with $\mathbf{i}\epsilon$ prescription implicit, and $\phi_i \equiv \phi(x_i)$)

$$G^n \equiv \langle\Omega|\mathcal{T}\phi_1\cdots\phi_n|\Omega\rangle = Z^{-1}\int[D\phi]\phi_1\cdots\phi_n e^{\mathbf{i}S[\phi]} \tag{2.1}$$

$$= \frac{\int[D\phi]\phi_1\cdots\phi_n e^{-\mathbf{i}\int V(\phi)}e^{\mathbf{i}S_0[\phi]}}{\int[D\phi]e^{\mathbf{i}S_0[\phi]}e^{-\mathbf{i}\int V(\phi)}} \tag{2.2}$$

$$= \frac{\langle 0|\mathcal{T}\phi_1\cdots\phi_n e^{-\mathbf{i}\int V(\phi)}|0\rangle}{\langle 0|\mathcal{T}e^{-\mathbf{i}\int V(\phi)}|0\rangle}. \tag{2.3}$$

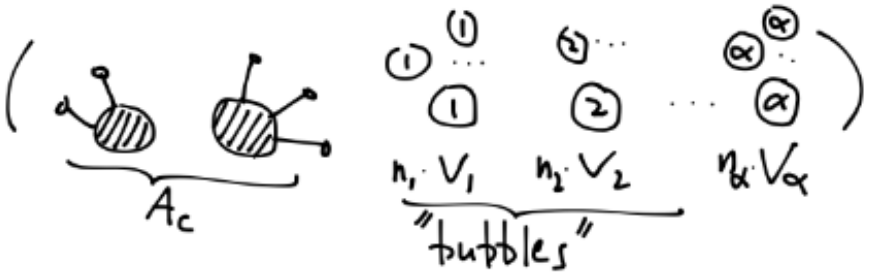
Here we've written $S = S_0 - \int V$ where S_0 is gaussian. The last object here is a time-ordered expectation value in the *free theory*, which we know how to compute by Wick contraction.

Two comments about this formula: (1) It must be admitted that in (2.1) the variable 'ϕ' is seriously overloaded: on the LHS it is used to represent a (Heisenberg-picture) operator, while on the RHS it is used to represent a (functional) integration variable. (2) This formula (or a related one) is sometimes called the Dyson formula for interaction-picture time evolution.

Taylor expanding the exponential $e^{-\mathbf{i}\int V}$ in (2.3) reproduces the diagrammatic expansion. (Notice that in real-time quantities, the interaction vertex comes with a factor of $-\mathbf{i}g$.) The denominator is the sum of bubble diagrams. In both numerator and denominator, the disconnected diagrams exponentiate, and therefore cancel. Here is a reminder of why this is true:

The exponentiation of the disconnected diagrams. [Peskin page 96] There are some patterns in these sums of diagrams to which it behooves us to attend. (The

following discussion transcends the ϕ^4 example.) The general diagram has the form:



Only some of the components are attached to the external legs; for a given diagram A, call the factor associated with these components A_c (note that A_c need not be *fully* connected). The rest of the diagram is made of a pile of 'bubbles' of various types V_i (each one internally connected, but disconnected from the external lines) and multiplicities n_i (*e.g.* V_1 could be a figure eight, and there could be $n_1 = 2$ of them. These bubbles (or 'vacuum bubbles') would be there even if we didn't have any external lines, and they would have the same value; they are describing the fluctuations intrinsic to the vacuum. The amplitude associated with the general diagram is then

$$\mathcal{M}_A = \mathcal{M}_{A_c} \cdot \frac{V_1^{n_1}}{n_1!} \cdot \frac{V_2^{n_2}}{n_2!} \cdots \frac{V_\alpha^{n_\alpha}}{n_\alpha!}$$

where the $n_i!$ factors are the most important appearance of symmetry factors: they count the number of ways to permute the identical copies of V_i amongst themselves.

The numerator of $G^{(n)}$ is then

$$\begin{aligned} G^{(n)}_{\text{numerator}} = \langle 0| \mathcal{T}\left(\phi_1 \cdots \phi_n e^{-\mathbf{i}\int V}\right)|0\rangle = \sum_A \mathcal{M}_A &= \sum_{A_c} \mathcal{M}_{A_c} \sum_{\{n_i=0\}} \frac{V_1^{n_1}}{n_1!} \cdot \frac{V_2^{n_2}}{n_2!} \cdots \frac{V_\alpha^{n_\alpha}}{n_\alpha!} \\ &= \sum_{A_c} \mathcal{M}_{A_c} \cdot e^{V_1} \cdot e^{V_2} \cdots e^{V_\alpha} \\ &= \sum_{A_c} \mathcal{M}_{A_c} e^{\sum_i V_i} \end{aligned} \tag{2.4}$$

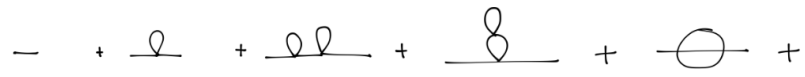
– the bubbles always exponentiate to give the same factor of $e^{\sum_i V_i}$, independent of the external data in G. In particular, consider the case of $n = 0$, where there are no external lines and hence no A_c:

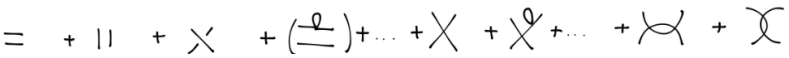
$$G^{(0)}_{\text{numerator}} = \langle 0| \mathcal{T} e^{-\mathbf{i}\int V}|0\rangle = 1 \cdot e^{\sum_i V_i}$$

But we care about this because it is the denominator of the actual Green's function:

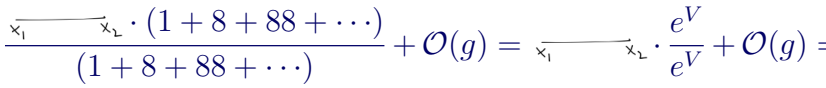
$$G^{(n)} = \frac{\langle 0| \mathcal{T}\left(\phi_1 \cdots \phi_n e^{-\mathbf{i}\int V}\right)|0\rangle}{\langle 0| \mathcal{T} e^{-\mathbf{i}\int V}|0\rangle} = \frac{G^{(n)}_{\text{numerator}}}{G^{(0)}_{\text{numerator}}} = \sum_{A_c} \mathcal{M}_{A_c} \,. \tag{2.5}$$

And with that we can forget all about the bubbles. So for example,

$G^{(2)} =$ 

$G^{(4)} =$ 

Notice that in this manipulation (2.5) we are adding terms of many orders in perturbation theory in the coupling g. If we want an answer to a fixed order in g, we can regard anything of higher order as zero, so for example, it makes perfect sense to write

$G^{(2)} =$ 

(I only drew one kind of bubble in the previous expression since that one was easy to type.)

Momentum space Green's functions from Feynman diagrams. In translation-invariant problems, things are usually a little nicer in momentum space. In ϕ^4 theory in $d+1$ dimensions, let's think about

$$\tilde{G}^{(n)}(p_1 \cdots p_n) \equiv \prod_{i=1}^{n} \int d^{d+1}x_i e^{-\mathbf{i}p_i x_i} G^{(n)}(x_1 \cdots x_n).$$

This an off-shell Green's function, a function of general p, not necessarily $p^2 = m^2$. It will, however, vanish unless $\sum_i p_i^\mu = 0$ by translation invariance. Consider a fully-connected contribution to it, at order g^N. (We'll get the others by multiplying these bits.)

In ϕ^4 theory, we need to make a diagram by connecting n external position vertices x_i to N 4-valent vertices z_a using Feynman propagators $\Delta_F(y_A - y_B) = \int \mathrm{d}^{d+1}q_r e^{-\mathbf{i}(y_A - y_B)q_r} \frac{\mathbf{i}}{q_r^2 - m^2 + \mathbf{i}\epsilon}$, where $\{y_A\} = \{z_a, x_i\}$. All of the position dependence is in these exponentials.

Since each propagator has two ends, the number of lines (by the fully-connected assumption) is

$$N_I = \frac{\#\text{ of ends of lines}}{2} = \frac{n + 4N}{2} = \frac{n}{2} + 2N.$$

This is the number of q integrals, a priori.

The integral over the external positions x_i (in the Fourier transform) gives $\int d^{d+1}x_i e^{\mathbf{i}x_i(q_i - p_i)} = \not{\delta}^{d+1}(q_i - p_i)$ and so we can label the external lines by p_i (and we lose n q integrals).

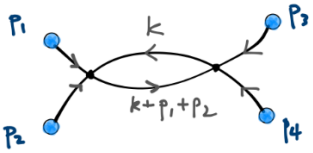
The integral over the position of each internal vertex is of the form $\int d^{d+1}z e^{\mathbf{i}z(\sum_r q_r)}$ where q_r are the momenta associated to the lines coming into the vertex. So each

internal vertex decreases the number of q integrals by 1. One combination of the momenta is fixed by overall momentum conservation so we have left

$$N_I - n - (N-1) = N - \frac{n}{2} + 1 \equiv N_L \tag{2.6}$$

momentum integrals. This number is ≥ 0 for fully connected diagrams, and it is the number of *loops* in the diagram. (This counting is the same as in a Kirchoff's law resistor network problem.)[11]

For example, consider a particular contribution to $G^{(4)}$ ($n = 4$ external legs) and $N = 2$ interaction vertices



which has $N_I = \dfrac{4 + 2\cdot 4}{2} = 6.$

In the example, $N_L = 2 - 2 + 1 = 1$ which agrees with one undetermined momentum integral. This gives the amplitude

$$\mathcal{M}_{FC}(p_1 \cdots p_n) = (-\mathbf{i}g)^N \cdot s(FC)\delta^{(d+1)}(\sum p_i) \int \prod_{\text{loops},\alpha=1}^{N_L} \mathrm{d}^{d+1}k_\alpha \prod_{\text{lines},r} \frac{\mathbf{i}}{q_r^2 - m^2 + \mathbf{i}\epsilon}$$

$$= \frac{(-\mathbf{i}g)^2}{2!}\delta^{d+1}(\sum_{i=1}^{4} p_i) \prod_{i=1}^{n=4} \frac{\mathbf{i}}{p_i^2 - m^2 - \mathbf{i}\epsilon} \int \mathrm{d}^{d+1}k \frac{\mathbf{i}}{k^2 - m^2 + \mathbf{i}\epsilon} \frac{\mathbf{i}}{(p_1 + p_2 + k)^2 - m^2 + \mathbf{i}\epsilon}$$

(You might notice that the integral over k is in fact formally infinite, since at large k it goes like $\int^\Lambda \frac{d^4k}{k^2} \sim \log(\Lambda)$. Try to postpone that worry.) The propagators for the external lines just factor out, and can be brought outside the momentum integrals. Notice that here p is general, and this function has poles when the external particles go on-shell, $p_i^2 = m^2$.

So here are the momentum space Feynman rules for Green's function in ϕ^4 theory:

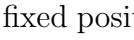
- Every line gives a factor of  $= \frac{\mathbf{i}}{p^2 - m^2 + \mathbf{i}\epsilon} = \tilde{\Delta}_F(p)$. Notice that since $\Delta_F(x-y) = \Delta_F(y-x)$, the choice of how we orient the momenta is not so fateful.

- An internal vertex gives  $\rightsquigarrow (-\mathbf{i}g)\int d^{d+1}z e^{-\mathbf{i}\sum_i p_i z} = (-\mathbf{i}g)\delta^{d+1}(\sum_i p_i)$, momentum conservation at each vertex. So, set $\sum_i p_i = 0$ at each vertex (I've assumed the arrows are all pointing toward the vertex). After imposing momentum

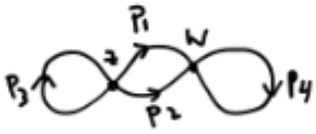
[11] Here's a proof that (2.6) is the number of loops in the diagram: place the $N + n$ internal and external vertices on the page. Add the propagators one at a time. You must add $N + n - 1$ just to make the diagram fully connected. After that, each line you add makes a new loop.

conservation, the remaining consequence of the vertex is

 $= -\mathbf{i}g.$

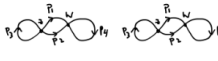
- Integrate over the loop momenta $\prod_{\alpha=1}^{N_L} \mathrm{đ}^{d+1} q_\alpha$ for each undetermined momentum variable. There is one for each loop in the diagram. You should think of these integrals as just like the Feynman path integral: if there is more than one way to get from here to there, we should sum over the amplitudes.
- Multiply by the wretched symmetry factor $s(A)$.
- For $\tilde{G}(p)$, multiply by an overall $\not{\delta}^{d+1}(\sum p)$ in each diagram.
- An external vertex at fixed position,  $= e^{-\mathbf{i}px}$. (Such vertices would arise if we wanted to compute $G(x)$ using momomentum-space feynman rules.) More generally, external vertices are associated with the wavefunctions of the states we are inserting; here they are plane waves.

Here is another perspective on the exponentiation of the vacuum bubbles. Consider the diagram:

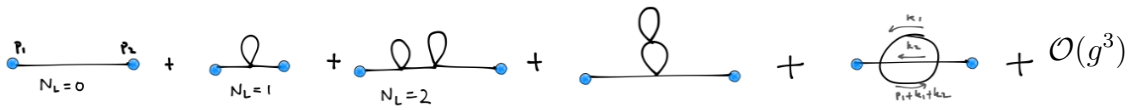
 $= (-\mathbf{i}g)^2 \prod_{i=1}^{4} \int \mathrm{đ}^{d+1} p_i \not{\delta}^{d+1}(p_1+p_2) \not{\delta}^{d+1}(p_1+p_2) \cdots$

The two delta functions come from the integrals over $z_{1,2}$, and we can restore sense by remembering this:

$$\left(\not{\delta}^{d+1}(p_1+p_2)\right)^2 = \not{\delta}^{d+1}(p_1+p_2) \int d^{d+1} z_2 = \not{\delta}^{d+1}(p_1+p_2) VT$$

where VT is the volume of spacetime. This factor arises because this process can happen anywhere, anytime. There is one such factor for each connected component of a collection of vacuum bubbles, so for example the diagram $\left($  $\right)$ is proportional to $(VT)^2$. But the free energy $\propto \log Z = \log G^{(0)}$ should be extensive, $\propto VT$. Therefore, the vacuum bubbles must exponentiate.

The whole two point function in momentum space is then (through order g^2) :

$\tilde{G}^{(2)} =$  $+ \mathcal{O}(g^3)$

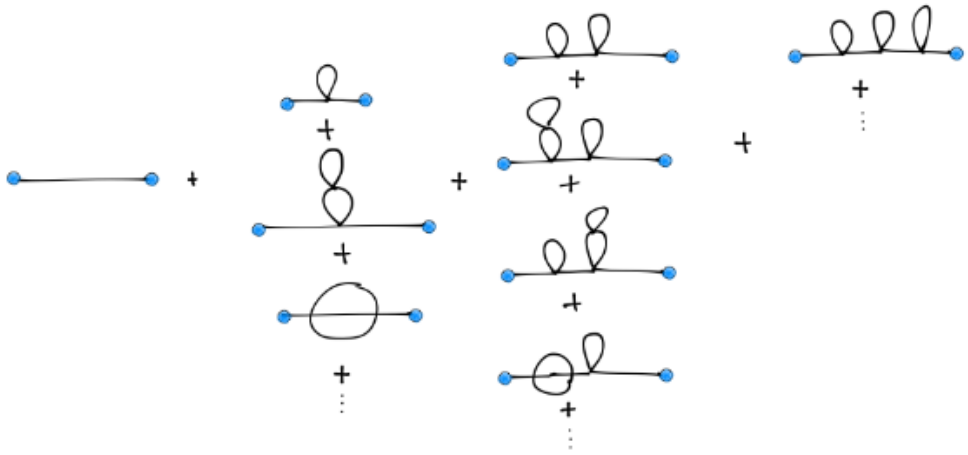
(2.7)

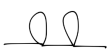
I draw the blue dots to emphasize the external propagators. Notice that for the two-point function, the number of loops is $N_L = N - \frac{n}{2} + 1 = N$, the same as the number of powers of g. More generally, for $n \neq 2$, there is an additive shift: N_L = constant plus number of powers of g.

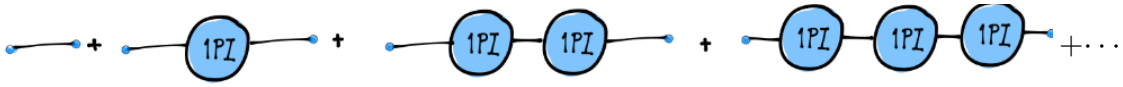
Organizing the propagator. We would like to unpack the physics contained in the correlation functions which we've learned to compute in perturbation theory. The first interesting one is the two-point function aka the propagator. Let's factor out the overall delta function by writing:

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

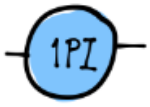
It will be useful to re-organize this sum, in the following way:



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

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

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

$$\begin{aligned}\tilde{G}^{(2)}(p) &= \frac{\mathbf{i}}{p^2-m_0^2} + \frac{\mathbf{i}}{p^2-m_0^2}(-\mathbf{i}\Sigma(p))\frac{\mathbf{i}}{p^2-m_0^2} + \frac{\mathbf{i}}{p^2-m_0^2}(-\mathbf{i}\Sigma(p))\frac{\mathbf{i}}{p^2-m_0^2}(-\mathbf{i}\Sigma(p))\frac{\mathbf{i}}{p^2-m_0^2} + \cdots \\ &= \frac{\mathbf{i}}{p^2-m_0^2}\left(1 + \frac{\Sigma}{p^2-m_0^2} + \left(\frac{\Sigma}{p^2-m_0^2}\right)^2 + \cdots\right) \\ &= \frac{\mathbf{i}}{p^2-m_0^2}\frac{1}{1-\frac{\Sigma}{p^2-m_0^2}} = \frac{\mathbf{i}}{p^2-m_0^2-\Sigma(p)}. \end{aligned} \tag{2.8}$$

[End of Lecture 5]

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, $m_0^2 + \Sigma(p)|_{\text{pole}}$ is the physical mass, 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

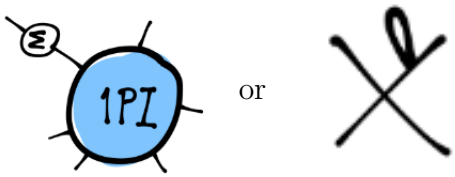
$$\tilde{G}^{(2)}(p) \equiv \left(\frac{\mathbf{i}Z}{p^2-m^2} + \text{regular bits}\right) \tag{2.9}$$

This equation defines the residue 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. Later we will see that unitarity requires $Z \leq 1$. Notice that if we know Σ only to some order in perturbation theory, then (2.8) is still true, up to corrections at higher order.

The notion of 1PI extends to diagrams for $\tilde{G}^{(n>2)}(p_1 \cdots p_n)$. Let

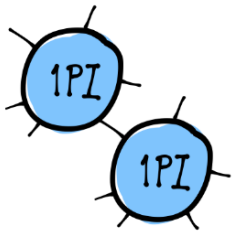
$\tilde{G}^{(n)}_{1PI}(p_1 \cdots p_n) \equiv$ 

where the blob indicates the sum over all 1PI diagrams with n external *nubbins* (notice that these do not have the blue circles that were present before). This means G_{1PI} does not include diagrams like:



or .

Notice that 1PI diagrams are *amputated* – their external limbs have been cut off.

LSZ reduction formula. This is almost what we need to make S-matrix elements. If we multiply the n-point function by $\prod_{i=1}^{n} \frac{p_i^2 - m^2}{\sqrt{Z}}$ we cancel out the propagators from the external legs, near the mass shell. This object is naturally called the *amputated* n-point function. (It differs from the 1PI n-point Green's function because of diagrams like this one 

which is amputated but not 1PI.) If we then take $p_i^2 \to m^2$, we keep only the part of $\tilde{G}$ which is singular on the mass-shell. And here's why we care about that:

Claim (the LSZ reduction formula):

$$S_{fi} \equiv \langle \vec{p}_1 \cdots \vec{p}_n | S | \vec{k}_1 \cdots \vec{k}_m \rangle = \prod_{a=1}^{n+m} \left(\lim_{P_a^0 \to E_{\vec{P}_a}} \frac{P_a^2 - m^2}{\mathbf{i}\sqrt{Z}} \right) \tilde{G}^{(n+m)} \left(k_1 \cdots k_m, -p_1 \cdots - p_n \right) \tag{2.10}$$

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 which are not fully connected.

This formula provides the bridge from time-ordered Green's functions (which we know how to compute in perturbation theory now) 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 has several steps. The field operators in this discussion are all in Heisenberg picture.

1. First, for a free field, the mode expansion implies that we can extract the ladder operators by:
$$\sqrt{2\omega_k} a_k = \mathbf{i} \int d^d x \; e^{\mathbf{i}kx} \left(-\mathbf{i}\omega_k + \partial_0 \right) \phi_{\text{free}}(x)$$

$$\sqrt{2\omega_k}a_k^\dagger = -\mathbf{i}\int d^dx\; e^{-\mathbf{i}kx}\left(+\mathbf{i}\omega_k + \partial_0\right)\phi_{\text{free}}(x) \tag{2.11}$$

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

2. Now let's pretend that we can turn the interactions off at $t = \pm\infty$, so that the asymptotic states we are scattering are free particles.[12] This allows us to write the field in terms of some pretend free fields of mass m (not m_0!)

$$\phi(x)\begin{cases} \overset{t\to-\infty}{\rightsquigarrow} & Z^{\frac{1}{2}}\phi_{\text{in}}(x) \\ \overset{t\to+\infty}{\rightsquigarrow} & Z^{\frac{1}{2}}\phi_{\text{out}}(x)\end{cases}.$$

The factors of $\sqrt{Z}$ are required to get the correct two point functions (2.9) 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 reference time for $\phi_{\text{in, out}}$ is $\pm\infty$ respectively. Since they are free fields, we can use (2.11) to write

$$\sqrt{2\omega_k}a^{(\text{in})\dagger} = -\mathbf{i}\int d^dx\, e^{-\mathbf{i}kx}\left(+\mathbf{i}\omega_k + \partial_0\right)\phi_{\text{in}}(x) = -\mathbf{i}Z^{-1/2}\int d^dx\, e^{-\mathbf{i}kx}\left(+\mathbf{i}\omega_k + \partial_0\right)\phi(x)\big|^{t\to-\infty}$$

where in the second step we used the independence on time in (2.11), 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 manifestly covariant using the fundamental theorem of calculus:

$$\begin{aligned}\sqrt{2\omega_k}\left(a^{(\text{in})\dagger} - a^{(\text{out})\dagger}\right) &= \mathbf{i}Z^{-1/2}\int_{-\infty}^{\infty} dt\partial_t\left(\int d^dx\; e^{-\mathbf{i}kx}\left(\mathbf{i}\omega_k + \partial_0\right)\phi(x)\right) \\ &\overset{\text{IBP in time}}{=} \mathbf{i}Z^{-1/2}\int d^{d+1}x\left(e^{-\mathbf{i}kx}\partial_0^2\phi - \phi\cdot\underbrace{\partial_0^2\; e^{-\mathbf{i}k_\mu x^\mu}}_{\left(\vec{\nabla}^2-m^2\right)e^{-\mathbf{i}kx}}\right) \\ &\overset{\text{IBP in space}}{=} \mathbf{i}Z^{-1/2}\int d^{d+1}x e^{-\mathbf{i}kx}\left(\Box + m^2\right)\phi(x)\end{aligned} \tag{2.12}$$

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 none of the incoming momenta k_i is the same as any outgoing momentum p_j.

$$\langle p_1\cdots p_n|\, S\,|k_1\cdots k_m\rangle$$

[12] 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.

$$
\begin{aligned}
&= \prod_{p,k} \sqrt{2\omega}\, \langle\Omega| \prod a_p^{\text{out}} S \prod a_k^{\text{in}\dagger} |\Omega\rangle \\
&= \prod_{p,k} \sqrt{2\omega}\, \langle\Omega| \mathcal{T}\left(\prod a_p^{\text{out}} S \prod a_k^{\text{in}\dagger}\right) |\Omega\rangle \qquad {}^{a^{\text{out}} \text{ lives at } t=+\infty} \\
&= \prod_{p,k} \sqrt{2\omega}\, \langle\Omega| \mathcal{T}\left(\prod a_p^{\text{out}} S \left(a_{k_1}^{\text{in}\dagger} - a_{k_1}^{\text{out}\dagger}\right) \prod_2^m a_k^{\text{in}\dagger}\right) |\Omega\rangle \qquad {}^{\text{since } p_i \neq k_j,\ \text{use } \langle 0| a^{\text{out}\dagger} = 0} \\
&\overset{(2.12)}{=} \mathbf{i} Z^{-1/2} \int d^{d+1}x_1 e^{-\mathbf{i}k_1 x_1} \langle\Omega| \mathcal{T}\left(\prod \sqrt{\omega_p} a_p^{\text{out}} S\left(\Box + m^2\right)\phi(x_1) \prod_2^m \sqrt{\omega_k} a_k^{\text{in}\dagger}\right) |\Omega\rangle \\
&= \mathbf{i} Z^{-1/2} \int d^{d+1}x_1 e^{-\mathbf{i}k_1 x_1} \left(\Box + m^2\right) \langle\Omega| \mathcal{T}\left(\prod \sqrt{\omega_p} a_p^{\text{out}} S\phi(x_1) \prod_2^m \sqrt{\omega_k} a_k^{\text{in}\dagger}\right) |\Omega\rangle + \text{X}
\end{aligned}
$$

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

5. Now do this for every particle to get

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

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 (2.10).

Comment: In our discussion of QFT, a special role has been played by 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 $\mathcal{O}$:

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

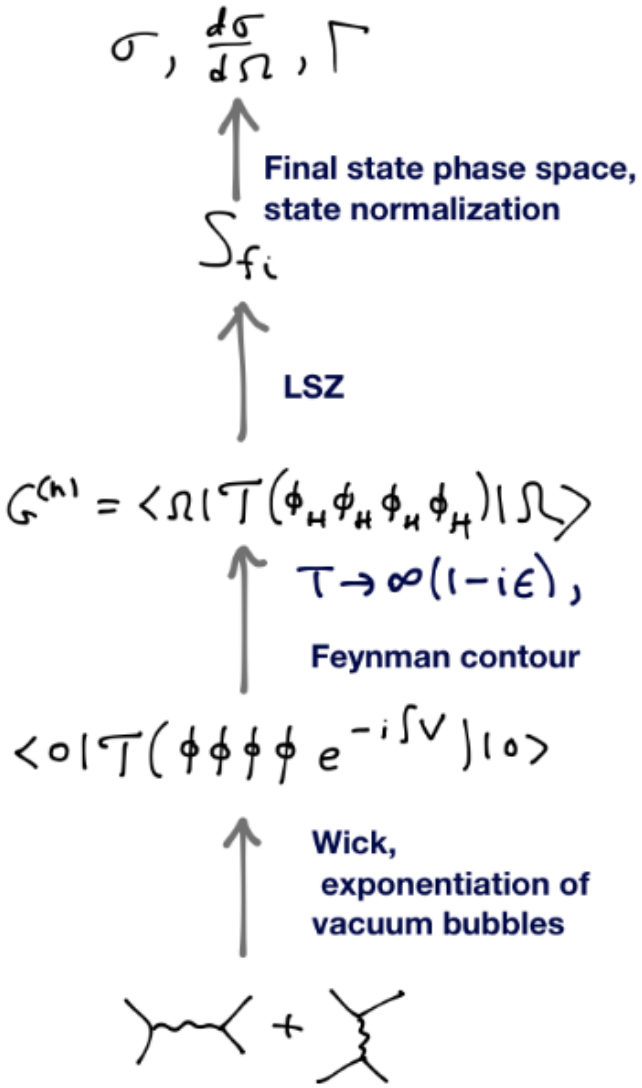
In this case, there is a more general statement of LSZ:

$$
\textstyle\prod_{a\in i} \left(Z_a^{-1/2} \mathbf{i} \int d^{d+1}x_a e^{-\mathbf{i}p_a x_a} \left(\Box_a + m_a^2\right)\right)
$$

$$\prod_{b\in f}\left(Z_b^{-1/2}\mathbf{i}\int d^{d+1}x_b e^{+\mathbf{i}p_b x_b}\left(\Box_b+m_b^2\right)\right)\; G_{\mathcal{O}}^{(n)}(1\cdots n) = \langle\{p_f\}|\,S\,|\{p_a\}\rangle \tag{2.13}$$

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

Here is a summary of the long logical route connecting Feynman diagrams to measurable quantities in particle physics:



The final step was covered in 215A.

S-matrix from Feynman diagrams. The end result of the previous discussion is a prescription to compute S-matrix elements from Feynman diagrams. In a translation-invariant system, the S matrix always has a delta function outside of it. Also we are not so interested in the diagonal elements of the S matrix where nothing happens. So more useful than the S matrix itself are the scattering amplitudes $\mathcal{M}$ defined by

$$\langle f| \left(S - \mathbb{1}\right) |i\rangle \equiv (2\pi)^{d+1}\delta^{(d+1)}\left(\sum_f p_f - \sum_i p_i\right) \mathbf{i}\mathcal{M}_{fi} \ . \tag{2.14}$$

(The object $\mathbf{i}\mathcal{M}\delta^{d+1}(\sum p)$ is sometimes called the transfer matrix. The $\mathbf{i}$ is a convention.)

The rules for the Feynman diagram calculation of $\mathcal{M}$ (for ϕ^4 theory, as a representative example) are:

1. Draw all amputated diagrams with appropriate external nubbins for the initial and final states. For a diagram with N_L loops think of N_L letters that are like k or q or p to call the undetermined loop momenta.
2. For each vertex, impose momentum conservation and multiply by the coupling $(-\mathbf{i}\lambda)$.
3. For each *internal* line, put a propagator.
4. For each *external* line, put a factor of $\sqrt{Z}$.
5. For each loop, integrate over the associated momentum $\int \mathrm{d}^{d+1}k$.

A comment about rule 1: For tree-level diagrams (diagrams with no loops), 'amputate' just means leave off the propagators for the external lines. More generally, it means leave off the resummed propagator (2.8). In particular, a diagram like  is already included by using the correct Z and the correct m.

I skipped the example in lecture. Please read through it and make sure you are happy about it. Please ask questions if you are not. We will do some examples in QED soon.

Example: snucleon scattering. [Here we follow Tong §3.5 very closely] Let's consider an example with a complex scalar field Φ interacting with a real scalar field ϕ with Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu\Phi^\star\partial^\mu\Phi - \frac{1}{2}m^2\Phi^\star\Phi + \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}M^2\phi^2 + \mathcal{L}_I \tag{2.15}$$

$$\text{with} \quad \mathcal{L}_I = -g\Phi^\star\Phi\phi.$$

In specifying initial states below, I will need names for the mode operators of the two fields:

$$\phi = \int \frac{\mathrm{đ}^d p}{\sqrt{2\omega_p}} \left(\mathbf{a}_p e^{-\mathbf{i}px} + \mathbf{a}_p^\dagger e^{\mathbf{i}px}\right)|_{p^0=\omega_p}$$

$$\Phi = \int \frac{\mathrm{đ}^d p}{\sqrt{2E_p}} \left(\mathbf{b}_p e^{-\mathbf{i}px} + \mathbf{c}_p^\dagger e^{\mathbf{i}px}\right)|_{p^0=E_p}$$

where I've written $\omega_p \equiv \sqrt{M^2+p^2}, E_q \equiv \sqrt{m^2+q^2}$. Notice that the $\Phi \to e^{-\mathbf{i}\alpha}\Phi$ symmetry is conserved; the charge is

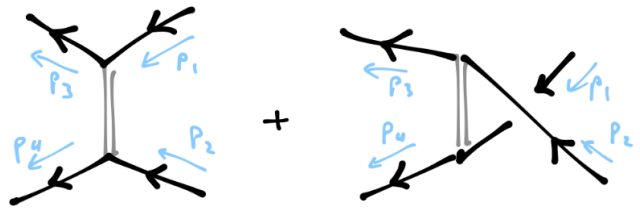
$$q = N_c - N_b.$$

But the ϕ particles are not conserved.[13] Relative to ϕ^4 theory, the differences in the Feynman rules are: we have two kinds of propagators, one of which is *oriented* (to keep track of the flow of the conserved Φ number through the diagrams), and instead of a 4-point vertex which costs $-\mathbf{i}g$, we have a 3-point vertex for $\phi\Phi^\star\Phi$ which costs $-\mathbf{i}g$.

Let's consider $2 \to 2$ scattering of Φ particles, so

$$|i\rangle = |\vec{p}_1, \vec{p}_2\rangle\,, |f\rangle = |\vec{p}_3, \vec{p}_4\rangle \quad \text{with} \quad |\vec{p}_i, \vec{p}_j\rangle \equiv \sqrt{2E_{\vec{p}_i}}\sqrt{2E_{\vec{p}_j}}\,\mathbf{b}^\dagger_{\vec{p}_i}\mathbf{b}^\dagger_{\vec{p}_j}\,|0\rangle\,.$$

(To appreciate some of the beauty of the diagram technique, see Tong §3.3.3 for the artisanal version of this calculation.) The Feynman rules above give, to leading nonzero order,

$\mathbf{i}\mathcal{M} =$ 

$$= (-\mathbf{i}g)^2 \left(\frac{\mathbf{i}}{(p_1-p_3)^2 - M^2 + \mathbf{i}\epsilon} + \frac{\mathbf{i}}{(p_1-p_4)^2 - M^2 + \mathbf{i}\epsilon}\right). \tag{2.16}$$

The diagrams depict two 'snucleons' Φ (solid lines with arrows indicating snucleons versus antisnucleons) exchanging a meson ϕ (double gray line, with no arrow) with momentum $k \equiv p_1 - p_3 = p_4 - p_2$ (first term) or $k \equiv p_4 - p_1 = p_2 - p_3$ (second term). Time goes to the left as always. Notice that here I am being careful about using arrows

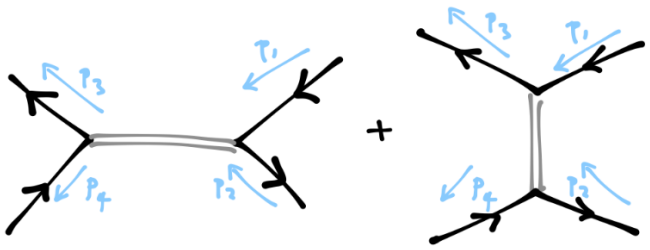
[13] You might notice a possible problem with this theory: what happens to the quadratic term for Φ when ϕ is very negative? Let's not take it too seriously.

on the lines to indicate flow of particle number through the diagram, while the extra (light blue) arrows indicate momentum flow.

The meson in these diagrams is virtual, or off-shell, in the sense that it does not satisfy its equation of motion $k^2 \neq M^2$. In fact, each of these diagrams is actually the sum of retarded and advanced exchange of real on-shell particles. (For more on this statement, see Schwartz chapter 4 or §4.5 of the notes here.) The two diagrams included in (2.16) make the amplitude symmetric under interchanging the two particles in the initial or final state, as it must be because they are indistinguishable bosons.

Two more examples with the same ingredients are useful for comparison. If we instead scatter a snucleon and an *anti-snucleon*, so $|i\rangle = \sqrt{2E_{\vec{p}_1}}\sqrt{2E_{\vec{p}_2}}\mathbf{b}^\dagger_{\vec{p}_1}\mathbf{c}^\dagger_{\vec{p}_2}|0\rangle$, then the leading diagrams are

$$\mathbf{i}\mathcal{M} = \quad + \quad$$

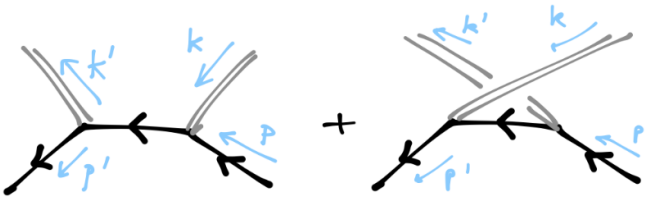


$$= (-\mathbf{i}g)^2 \left(\frac{\mathbf{i}}{(p_1+p_2)^2 - M^2 + \mathbf{i}\epsilon} + \frac{\mathbf{i}}{(p_1-p_3)^2 - M^2 + \mathbf{i}\epsilon} \right). \tag{2.17}$$

This one has a new ingredient: in the first diagram, the meson momentum is $k = p_1 + p_2$, which *can* be on-shell, and the $\mathbf{i}\epsilon$ matters. This will produce a big bump, a *resonance*, in the answer as a function of the incoming center-of-mass energy $\sqrt{s} \equiv \sqrt{(p_1+p_2)^2}$.

Finally, we can scatter a meson and a snucleon:

$$\mathbf{i}\mathcal{M} = \quad + \quad$$



$$= (-\mathbf{i}g)^2 \left(\frac{\mathbf{i}}{(p+k)^2 - m^2 + \mathbf{i}\epsilon} + \frac{\mathbf{i}}{(p-k')^2 - m^2 + \mathbf{i}\epsilon} \right). \tag{2.18}$$

Now the intermediate state is a snucleon.

There is a common notation for the Lorentz-invariant combinations of the momenta appearing in these various processes, called Mandelstam variables, of which s is one. A concise summary appears in §3.5.1 of Tong's notes.

3 QED

3.1 Quantum light: Photons

I skipped this subsection in lecture. Please read through it and make sure you are happy about it. Please ask questions if you are not.

The quantization of the Maxwell field is logically very similar to the case of a harmonic chain. There are just a few complications from its several polarizations, and from the fact that quantum mechanics means that the vector potential is real and necessary (whereas classically it is just a convenience). This is a quick-and-dirty version of the story. I mention it here to emphasize that the machinery we are developing applies to a system you have already thought a lot about!

Maxwell's equations (with $c = 1$) are:

$$\epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} = 0 \qquad \vec{\nabla}\cdot\vec{B} = 0, \qquad \vec{\nabla}\times\vec{E} = -\partial_t\vec{B}, \tag{3.1}$$

$$\partial^\mu F_{\mu\nu} = 4\pi j_\nu \qquad \vec{\nabla}\cdot\vec{E} = 4\pi\rho, \qquad \nabla\times\vec{B} = \partial_t\vec{E} + 4\pi\vec{j} \tag{3.2}$$

(where the familiar electric and magnetic fields are $E^i = -F^{0i}$ and $\epsilon^{ijk}B^k = -F^{ij}$). The first two equations (3.1) are constraints on $\vec{E}$ and $\vec{B}$ which mean that their components are not independent. This is annoying for trying to treat them quantumly. To get around this we introduce *potentials* $A_\mu = (\Phi, \vec{A})_\mu$ which determine the fields by taking derivatives and which automatically solve the constraints (3.1):

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \qquad \text{aka} \qquad \vec{E} = -\vec{\nabla}\Phi - \partial_t\vec{A}, \quad \vec{B} = \vec{\nabla}\times\vec{A}.$$

Potentials related by a *gauge transformation*

$$\vec{A} \to \vec{A}^\lambda = \vec{A} - \vec{\nabla}\lambda, \quad \Phi \to \Phi^\lambda = \Phi + \partial_t\lambda$$

for any function $\lambda(\vec{r}, t)$, give the same $\vec{E}, \vec{B}$. The Bohm-Aharonov effect is proof that (some of the information in) the potential is real and useful, despite this redundancy. We can partially remove this redundancy be choosing our potentials to satisfy *Coulomb gauge*

$$\vec{\nabla}\cdot\vec{A} = 0\ .$$

In the absence of sources $\rho = 0 = \vec{j}$, we can also set $\Phi = 0$. In this gauge, Ampere's law becomes

$$c^2\vec{\nabla}\times\left(\vec{\nabla}\times\vec{A}\right) = c^2\vec{\nabla}\cdot\left(\vec{\nabla}\cdot\vec{A}\right) - c^2\nabla^2\vec{A} = -\partial_t^2\vec{A} \quad i.e. \ \boxed{\partial_t^2\vec{A} - c^2\nabla^2\vec{A} = 0}.$$

This wave equation is different from the scalar wave equation $\Box\phi = 0$ in three ways:

- we're in three spatial dimensions,
- the speed of sound v_s has been replaced by the speed of light c,
- the field $\vec{A}$ is a vector field obeying the constraint $\vec{\nabla} \cdot \vec{A} = 0$. In fourier space $\vec{A}(x) = \sum_k e^{i\vec{k}\cdot\vec{x}} \vec{A}(k)$ this condition is

$$0 = \vec{k} \cdot \vec{A}(k)$$

  – the vector field is *transverse*.

An action which gives rise to Maxwell's equations is

$$S[A] = \int d^4x \left(-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \right) = \int d^4x \mathcal{L}_{\text{Maxwell}}. \quad \mathcal{L}_{\text{Maxwell}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} = \frac{1}{2} \left(E^2 - B^2 \right).$$

Note that we must regard A as the dynamical variable to obtain (3.2) by $0 = \frac{\delta S}{\delta A_\mu(x)}$. The canonical momentum of A is then $\Pi_{A_i} = \frac{\partial \mathcal{L}_{\text{Maxwell}}}{\partial \dot{A}_i} = E^i$. So the Hamiltonian is[14]:

$$\mathbf{H} = \frac{1}{2} \int d^3x \left(\vec{E}^2 + c^2 \vec{B}^2 \right) \quad . \tag{3.3}$$

Here $\vec{E} = -\partial_t \vec{A}$ plays the role of field momentum $\pi(x)$ in (1.3), and $\vec{B} = \vec{\nabla} \times \vec{A}$ plays the role of the spatial derivative $\partial_x q$. We immediately see that we can quantize this system just like for the scalar case, with the canonical commutator

$$[\phi(x), \pi(x')] = \mathbf{i}\hbar\delta(x - x') \qquad \rightsquigarrow \qquad [\mathbf{A}_i(\vec{r}), \mathbf{E}_j(\vec{r}')] = -\mathbf{i}\hbar\delta^3(\vec{r} - \vec{r}')\delta_{ij}$$

where $i, j = 1..3$ are spatial indices[15]. So we can immediately write down an expression for the quantum Maxwell field in terms of independent creation and annihilation operators:

$$\vec{\mathbf{A}}(\vec{r}) = \int \mathrm{d}^3k \frac{1}{\sqrt{2\omega_k}} \sum_{s=1,2} \left(\mathbf{a}_{\vec{k},s} \vec{e}_s(\hat{k}) e^{\mathbf{i}\vec{k}\cdot\vec{r}} + \mathbf{a}^\dagger_{\vec{k},s} \vec{e}^\star_s(\hat{k}) e^{-\mathbf{i}\vec{k}\cdot\vec{r}} \right)$$

[14] You may also recall that the energy density of a configuration of Maxwell fields is $u = \frac{1}{2}\left(\vec{E}^2 + \vec{B}^2\right)$. This result can be obtained either by Legendre transformation of $\mathcal{L}_{\text{Maxwell}}$, or from T^0_0, the energy momentum tensor.

[15] As a check, note that using this Hamiltonian and the canonical commutator, we can reproduce Maxwell's equations using Ehrenfest's theorem:

$$\left\langle \partial_t^2 \vec{A} \right\rangle = \partial_t \left\langle \vec{E} \right\rangle = -\frac{\mathbf{i}}{\hbar} \left\langle [\mathbf{H}, \vec{E}] \right\rangle = \left\langle c^2 \vec{\nabla}^2 \vec{A} \right\rangle.$$

The field momentum is $\vec{\mathbf{E}} = -\partial_t \vec{A}$:

$$\vec{\mathbf{E}}(\vec{r}) = \mathbf{i}\sqrt{\frac{\omega_k}{2}} \sum_{s=1,2} \left(\mathbf{a}_{\vec{k},s} \vec{e}_s(\hat{k}) e^{\mathbf{i}\vec{k}\cdot\vec{r}} - \mathbf{a}^\dagger_{\vec{k},s} \vec{e}^\star_s(\hat{k}) e^{-\mathbf{i}\vec{k}\cdot\vec{r}} \right)$$

Also, the magnetic field operator is

$$\vec{\mathbf{B}} = \vec{\nabla} \times \vec{\mathbf{A}} = \sum_{\vec{k}} \sum_s \sqrt{\frac{\hbar}{2\epsilon_0 \omega_k L^3}} \mathbf{i}\vec{k} \times \left(\mathbf{a}_{\vec{k},s} \vec{e}_s(\hat{k}) e^{\mathbf{i}\vec{k}\cdot\vec{r}} - \mathbf{a}^\dagger_{\vec{k},s} \vec{e}^\star_s(\hat{k}) e^{-\mathbf{i}\vec{k}\cdot\vec{r}} \right) \ ;$$

the magnetic field is analogous to $\vec{\nabla}\phi$ in the scalar field theory[16]. Plugging these expressions into the Hamiltonian (3.3), we can write it in terms of these oscillator modes (which create and annihilate *photons*). As for the scalar field, the definitions of these modes were designed to make this simple: It is:

$$\mathbf{H} = \sum_{\vec{k},s} \hbar\omega_k \left(\mathbf{a}^\dagger_{\vec{k},s} \mathbf{a}_{\vec{k},s} + \frac{1}{2} \right).$$

Notice that in this case we began our story in the continuum, rather than with microscopic particles connected by springs. (However, if you read Maxwell's papers you'll see that he had in mind a particular UV completion involving gears and cogs. I actually don't understand it; if you do please explain it to me.)

The vacuum energy is

$$E_0 = \frac{1}{2} \sum_{\vec{k},s} \hbar\omega_k = \frac{L^3}{(2\pi)^3} \int d^3k \hbar c k.$$

The fact that $\sum_k$ is no longer a finite sum might be something to worry about. This vacuum energy has physical consequences, since it can depend on boundary conditions placed on the field by conducting objects, as we'll discuss in §4.1.

3.2 More on vector fields

A few things we did not do yet for vector fields: study the propagator, figure out the data on external states, and understand the relation of between the masslessness of the photon and gauge invariance.

[16] I should say a little more about the *polarization vectors*, $\vec{e}_s$. They conspire to make it so that there are only two independent states for each $\vec{k}$ and they are transverse $\vec{k} \cdot \vec{e}_s(\hat{k}) = 0$, so $s = 1, 2$. The polarization vectors of a given $\vec{k}$ can be chosen to satisfy the following completeness relation:

$$\sum_s e_{si}(\hat{k}) e^\star_{sj}(\hat{k}) = \delta_{ij} - \hat{k}_i \hat{k}_j \ . \tag{3.4}$$

This says that they span the plane perpendicular to $\hat{k}$.

Consider the following Lagrangian for a vector field A_μ (which I claim is the most general quadratic Poincaré-invariant Lagrangian with at most two derivatives):

$$\mathcal{L} = -\frac{1}{2}\left(\partial_\mu A^\nu \partial^\mu A_\nu + a \underbrace{\partial_\mu A^\mu \partial_\nu A^\nu}_{=(\partial A)^2} + bA_\mu A^\mu + c\epsilon^{\mu\nu\rho\sigma}\partial_\mu A_\nu \partial_\rho A_\sigma \right).$$

The sign is chosen so that spatial derivatives are suppressed, and the normalization of the first term is fixed by rescaling A. (Another possible-seeming term, $\partial_\mu A^\nu \partial_\nu A^\mu$, is related to the second term by two IBPs.) The last term is a total derivative, $\epsilon^{\mu\nu\rho\sigma}\partial_\mu A_\nu \partial_\rho A_\sigma \propto \partial_\mu \left(\epsilon^{\mu\nu\rho\sigma} A_\nu \partial_\rho A_\sigma\right)$, and will not affect the EoM or anything at all in perturbation theory; it is called a θ term.

The EoM are

$$0 = \frac{\delta}{\delta A^\nu(x)} \int \mathcal{L} = -\partial^2 A_\nu - a\partial_\nu \left(\partial \cdot A\right) + bA_\nu \tag{3.5}$$

which (like any translation-invariant linear equation) is solved by Fourier transforms $A_\mu(x) = \epsilon_\mu e^{-\mathbf{i}kx}$, if

$$k^2\epsilon_\mu + ak_\mu \left(k\cdot\epsilon\right) + b\epsilon_\mu = 0.$$

There are two kinds of solutions: longitudinal ones with $\epsilon_\mu \propto k_\mu$ (for which the dispersion relation is $k^2 = -\frac{b}{1+a}$), and transverse solutions $\epsilon \cdot k = 0$ with dispersion $k^2 = -b$. The longitudinal mode may be removed by taking $b \neq 0$ and $a \to -1$, which we will do from now on. This gives the Proca Lagrangian:

$$\mathcal{L}_{a=-1, b\equiv -\mu^2} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\mu^2 A_\mu A^\mu,$$

where as usual $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$. Note that the EOM (Proca equation) $0 = \partial^\cdot F_{\cdot\nu} + \mu^2 A_\nu$ implies $0 = \partial^\nu A_\nu$ by $0 = \partial^\mu \partial^\nu F_{\mu\nu}$. So each component of A_μ satisfies (by (3.5)) the KG equation, $k^2 = \mu^2$, and the transverse condition $\epsilon \cdot k = 0$. In the rest frame, $k^\mu = (k^0, \vec{0})^\mu$, we can choose a basis of plane-wave transverse solutions which are eigenstates of the vector rotation generator

$$J^z = \mathbf{i}\begin{pmatrix} & +1 & \\ -1 & & \\ & & \end{pmatrix}, \text{ namely, } \epsilon^{(\pm)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ \mp\mathbf{i} \\ 0 \end{pmatrix}, \epsilon^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

They are normalized so that $\epsilon^{(r)} \cdot \epsilon^{(s)} = +\delta^{rs}$ and $\sum_{r=\pm 1,0} \epsilon^{(r)\star}_\mu \epsilon^{(r)}_\nu = -\eta_{\mu\nu} + \frac{k_\mu k_\nu}{\mu^2}$ so that they project out $\epsilon \propto k$. Notice that in the massless case, only two of these three polarization states will be transverse to k^μ. If $\vec{k} \propto \hat{z}$ (for example in the massless case with $k^\mu = (E, 0, 0, E)^\mu$) then these ϵ are also all helicity eigenstates: $h = \vec{J}\cdot\hat{k} = J^z$.

Canonical stuff: The canonical momenta are $\pi^i = \frac{\partial \mathcal{L}}{\partial \dot{A}_i} = -F^{0i} = E^i$ (as for electrodynamics in §3.1) and $\pi^0 = \frac{\partial \mathcal{L}}{\partial \dot{A}_0} = 0$. This last bit is a little awkward, but it just means we can solve the equations of motion for A_0 algebraically in terms of the other (real) dofs:

$$0 = \frac{\delta S}{\delta A_0} = \vec{\nabla}\cdot\vec{E} - \mu^2 A_0 = (-\nabla^2 + \mu^2)A_0 + \vec{\nabla}\cdot\dot{\vec{A}} \implies A_0(\vec{x}) = \int d^3y e^{-\mu|\vec{x}-\vec{y}|} \frac{\left(-\vec{\nabla}\cdot\dot{\vec{A}}\right)}{4\pi|\vec{x}-\vec{y}|}\,. \tag{3.6}$$

So at each moment A_0 is determined by A_i. (Notice that this is still true for $\mu \to 0$.) The hamiltonian density is (after using $\pi^i = F^{0i}$, integration by parts, and the equations of motion for A_0)

$$\mathfrak{h} = +\frac{1}{2}\left(F_{0i}^2 + \frac{1}{2}F_{ij}^2 + \mu^2 A_i^2 + \mu^2 A_0^2\right) = \frac{1}{2}\left(\vec{E}^2 + \vec{B}^2 + \mu^2\vec{A}^2 + \mu^2 A_0^2\right) \geq 0,$$

where positivity follows from the fact that it is a sum of squares of real things.

The canonical equal time commutators are then

$$[A_i(t,\vec{x}), F^{j0}(t,\vec{y})] = \mathbf{i}\delta_i^j \delta^{(3)}(\vec{x}-\vec{y})$$

which if we add up the plane wave solutions as

$$A_\mu(x) = \sum_{r=1,2,3}\int \frac{\text{đ}^3k}{\sqrt{2\omega_k}}\left(e^{-\mathbf{i}kx}\mathbf{a}_k^r \epsilon_\mu^{(r)} + e^{+\mathbf{i}kx}\mathbf{a}_k^{r\dagger}\epsilon_\mu^{(r)\star}\right)$$

give the bosonic ladder algebra for each mode

$$[\mathbf{a}_k^r, \mathbf{a}_p^{s\dagger}] = \not{\delta}^{(3)}(\vec{k}-\vec{p})\delta^{rs}.$$

The normal-ordered hamiltonian is

$$:H: = \sum_r \int \text{đ}^3k\ \omega_k \mathbf{a}_k^{r\dagger}\mathbf{a}_k^r.$$

Using the mode expansion above, the propagator for the $A_\mu(x)$ field is found to be

$$\langle 0|\mathcal{T}A_\mu(x)A_\nu(y)|0\rangle = \int \text{đ}^4k e^{-\mathbf{i}k(x-y)}\left[\frac{-\mathbf{i}(\eta_{\mu\nu} - k_\mu k_\nu/\mu^2)}{k^2 - \mu^2 + \mathbf{i}\epsilon}\right]. \tag{3.7}$$

Notice that like in the spinor case the polarization sum $\sum_r \epsilon_\mu^{r\star}\epsilon_\nu^r = -(\eta_{\mu\nu} - k_\mu k_\nu/\mu^2)$ appears in the numerator of the propagator. (Note that there are 3 orthonormal polarizations, so this is a rank-3 matrix; its kernel is the longitudinal direction, k^μ.) The quantity in square brackets is then the momentum-space propagator. Since $\langle 0|\, A_\mu(x)\, |k,r\rangle =$

$\epsilon^r_\mu(k)e^{-\mathbf{i}kx}$, a vector in the initial state produces a factor of $\epsilon^r_\mu(k)$, and in the final state gives $\epsilon^\star$.

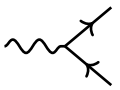
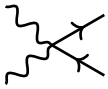
Massless case. In the limit $\mu \to 0$ some weird stuff happens. If we couple A_μ to some object j^μ made of other matter, by adding $\Delta\mathcal{L} = j^\mu A_\mu$, then we learn that $\partial_\mu A^\mu = \mu^{-2}\partial_\mu j^\mu$. This means that in order to take $\mu \to 0$, it will be best if the current is conserved $\partial_\mu j^\mu$.

One example is the QED coupling, $j^\mu = \bar{\Psi}\gamma^\mu\Psi$. Here j^μ is the Noether current for the symmetry $\Psi \to e^{\mathbf{i}\alpha}\Psi$ of the Dirac Lagrangian. This coupling $A_\mu j^\mu$ arises from the 'minimal coupling' prescription of replacing $\partial_\mu \to D_\mu = \partial_\mu + \mathbf{i}eqA_\mu$ in the Dirac Lagrangian. In this case, the model, with Lagrangian

$$\mathcal{L} = \bar{\Psi}\left(\mathbf{i}\not{D} - m\right)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{\mu^2}{2}A_\mu A^\mu|_{\mu^2=0},$$

has a local invariance under $A_\mu \to A_\mu + \partial_\mu\lambda(x)/e, \Psi(x) \to e^{\mathbf{i}q\lambda(x)}\Psi(x)$. For λ non-constant (and going to zero far away), this is a redundancy of our description rather than a symmetry (for example, they have the same configuration of $\vec{E}, \vec{B}, \oint A$). That is, configurations related by this gauge transformation should be regarded as equivalent.

[End of Lecture 6]

Another example can be obtained by taking a complex scalar and doing the same replacement: $\mathcal{L} = D_\mu\Phi^\star D^\mu\Phi + ...$ Notice that in this case the vertex involves a derivative, so it comes with a factor of  $= -\mathbf{i}eq(p_\Phi + p_{\Phi^\star})^\mu$. Also, there is a $A_\mu A_\nu\Phi^\star\Phi$ coupling, which gives a vertex  $= -\mathbf{i}e^2q^2\eta_{\mu\nu}$.

How do I know that configurations related by a gauge transformation should be regarded as equivalent? If not, the kinetic operator for the massless vector field $\left(\eta_{\mu\nu}\left(\partial^\rho\partial_\rho\right) - \partial_\mu\partial_\nu\right)A^\nu = 0$ is not invertible (even in Euclidean section!), since it annihilates $A_\nu = \partial_\nu\lambda$.

What's the propagator for a massless vector field, then? One strategy is to simply ignore the gauge equivalence and use the same propagator (3.7) that we found in the massive case with $\mu \to 0$. Since the dynamics are gauge invariant, it will never make gauge-variant stuff, and the longitudinal bits $\propto k_\mu k_\nu$ in (3.7) (which depend on μ) will just drop out, and we can take $\mu \to 0$ in the denominator at the end. This actually works. The guarantee that it works is the QED Ward identity: any amplitude with an

external vector $\epsilon(k)_\mu$ is of the form

$$= \quad \mathbf{i}\mathcal{M} = \mathbf{i}\mathcal{M}^\mu(k)\epsilon_\mu(k)$$

and if all external fermion lines are on-shell then

$$\mathcal{M}^\mu(k)k_\mu = 0.$$

There is a complicated diagrammatic proof of this statement in Peskin; Schwartz §8.4 argues that it is a necessary condition for Lorentz invariance of $\mathcal{M} = \epsilon_\mu \cdot \mathcal{M}^\mu$; and we will see some illustrations of it below (I also recommend Zee §II.7). But it is basically a statement of current conservation: such an amplitude is made (by LSZ and the photon Schwinger-Dyson equation) from a correlation function involving an insertion of the electromagnetic current $j^\mu(k) = \int d^4x\ e^{-\mathbf{i}kx} j^\mu(x)$, in the form, $\mathcal{M}^\mu \sim \ldots \langle\Omega|....j^\mu(k)...|\Omega\rangle$, and $k_\mu j^\mu(k) = 0$ is current conservation[17].

This property guarantees that we will not emit either of the unphysical polarizations of massless photons, since the amplitude to do so is either $\mathcal{A}(\text{ emit } \epsilon_\lambda \propto k_\lambda) = \epsilon_\mu\mathcal{M}^\mu \propto k_\mu\mathcal{M}^\mu \overset{\text{Ward}}{=} 0$, or the $\mu \to 0$ limit of

$$\mathcal{A}\left(\begin{matrix}\text{emit } \epsilon^L_\lambda = \frac{1}{\mu}(k,0,0,-\omega)_\lambda \\ \text{with } k^\lambda = (\omega,0,0,k)^\lambda\end{matrix}\right) \propto \epsilon^L_\mu\mathcal{M}^\mu = \frac{1}{\mu}\left(k\mathcal{M}^0 - \omega\mathcal{M}^3\right) = \frac{1}{\mu}\left(k\mathcal{M}^0 - \underbrace{\sqrt{k^2+\mu^2}}_{=k+\frac{\mu^2}{2k}+\ldots}\mathcal{M}^3\right)$$
$$= \frac{1}{\mu}\underbrace{k_\mu\mathcal{M}^\mu}_{=0,\text{by Ward}} - \underbrace{\frac{\mu}{2k}\mathcal{M}^3 + \mathcal{O}(\mu^3)}_{\to 0 \text{ as } \mu\to 0} \overset{\mu\to 0}{\to} 0.$$

For the same reason, in summing over photon final-states (for example in computing a cross section), we'll have

$$\sum_\epsilon |\mathcal{M}|^2 = \sum_\epsilon \epsilon(k)_\mu\epsilon(k)^\star_\nu\mathcal{M}^\mu(k)\mathcal{M}^\nu(k)^\star.$$

[17] Current conservation $\partial_\mu j^\mu$ is a statement which requires the equations of motion (recall the proof of Noether's theorem). Recall that equations of motion are true in correlation functions, up to contact terms, using the independence of the path integral on choice of integration variables. By contact terms, I mean terms which are only nonzero when two operators are at the same point. So you can worry about the contact terms in the argument for the Ward identity. The reason they do not contribute is that all the operators in the correlation function (using the LSZ formula) correspond to external states. A collision between the operators creating the external particles would lead to a disconnected amplitude, which could only contribute for degenerate kinematical configurations, and we can ignore them. If you would like to read more words about this, look at Schwartz §14.8, or §3.3.

This sum is only over the *two* physical polarizations of the massless photon. If we choose a frame with $k^\mu = (k, 0, 0, k)^\mu$, the Ward identity says $0 = k_\mu \mathcal{M}^\mu = k(\mathcal{M}^0 - \mathcal{M}^3)$, so

$$\sum_\epsilon \epsilon(k)_\mu \epsilon(k)^\star_\nu \mathcal{M}^\mu(k)\mathcal{M}^\nu(k)^\star = |\mathcal{M}^1|^2 + |\mathcal{M}^2|^2 = |\mathcal{M}^1|^2 + |\mathcal{M}^2|^2 + |\mathcal{M}^3|^2 - |\mathcal{M}^0|^2 = -\eta_{\mu\nu}\mathcal{M}^\mu(k)\mathcal{M}^\nu(k)^\star, \tag{3.8}$$

that is, just like in the numerator of the propagator, we can replace

$$\sum_\epsilon \epsilon(k)_\mu \epsilon(k)^\star_\nu \rightsquigarrow -\eta_{\mu\nu}$$

since they differ by stuff proportional to k^μ which vanishes when contracted with the rest of the amplitude. In (3.8) we see explicitly that the crazy timelike polarization (which looks like negative probability) cancels the longitudinal polarization $\vec{\epsilon} \propto \vec{k}$.

Gauge fixing. You might not be happy with the accounting procedure I've advocated above, where unphysical degrees of freedom are floating around in intermediate states and only drop out at the end by some formal trick. In that case, a whole zoo of formal tricks called *gauge fixing* has been prepared for you. Here's a brief summary to hold you over until we really need it for the non-Abelian case.

At the price of Lorentz invariance, we can make manifest the physical dofs, by choosing *Coulomb gauge*. That means we restrict $\partial_\mu A^\mu = 0$ (so far, so Lorentz invariant) *and also* $\vec{\nabla} \cdot \vec{A} = 0$. Looking at (3.6), we see that this kills off the bit of A_0 that depended on $\vec{A}$. We also lose the helicity-zero polarization $\vec{\nabla} \cdot \vec{A} \propto \epsilon^{(0)}$. But the Coulomb interaction is instantaneous action at a distance.

To keep Lorentz invariance, we can instead merely discourage configurations with $\partial \cdot A \neq 0$ by adding a term to the action

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} - \frac{1}{2\xi}(\partial \cdot A)^2$$

for some arbitrary number ξ. Physics should not depend on ξ, and this is a check on calculations. The propagator is

$$\langle \mathcal{T} A_\mu(x) A_\nu(y) \rangle_\alpha = \int \mathrm{d}^4 k \; e^{-\mathbf{i}k(x-y)} \left[\frac{-\mathbf{i}(\eta_{\mu\nu} - (1-\xi)k_\mu k_\nu/\mu^2)}{k^2 - \mu^2 + \mathbf{i}\epsilon} \right]$$

and again the bit with $k_\mu k_\nu$ must drop out. $\xi = 1$ is called Feynman gauge and makes this explicit. $\xi = 0$ is called Landau gauge and makes the propagator into a projector onto $k_\perp$.

It becomes much more important to be careful about this business in non-Abelian gauge theory.

3.3 On the non-perturbative proof of the Ward identity

[Schwartz §14.8] First, consider a Green's function from which we might make an S-matrix element by LSZ,

$$G \equiv \langle\Omega|\mathcal{T}\mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)|\Omega\rangle = \int [D\Psi] e^{\mathbf{i}S}\mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)$$

where the operators $\mathcal{O}_1(x_1) \mapsto e^{-\mathbf{i}Q_i\alpha}\mathcal{O}_1(x_1)$ have charge Q_i under a global $\mathsf{U}(1)$ symmetry. For example the $\mathcal{O}(x)$ could be just the elementary field $\Psi(x)$ [18].

Now change variables in the path integral so that $\mathcal{O}_i(x_i) \mapsto e^{-\mathbf{i}Q_i\alpha(x_i)}\mathcal{O}_i(x_i)$; the action will shift by $S \mapsto S - \int \alpha\partial_\mu j^\mu$ where j^μ is the Noether current. The path integral doesn't change at all, so its infinitesimal variation is

$$0 = \delta G = \int [D\Psi]\left(-\int \mathbf{i}\alpha\partial^\mu j_\mu e^{\mathbf{i}S}\mathcal{O}_1\cdots\mathcal{O}_n - \mathbf{i}\sum_i Q_i\alpha(x_i)e^{\mathbf{i}S}\mathcal{O}_1\cdots\mathcal{O}_n\right) \tag{3.9}$$

$$= \int d^Dx\alpha(x)\left[\mathbf{i}\partial_\mu\langle j^\mu(x)\mathcal{O}_1\cdots\mathcal{O}_n\rangle - \sum_i Q_i\delta^D(x-x_i)G\right]. \tag{3.10}$$

Since this is true for any $\alpha(x)$, we learn that the thing in square brackets is zero: $\partial_\mu j^\mu = 0$ up to contact terms. This is called the Ward-Takahashi identity.

Now suppose we do this same manipulation in a gauge theory, like QED. The additional terms in S are $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \mathbf{i}A_\mu\bar{\Psi}\gamma^\mu\Psi$, which are invariant under the transformation, so don't change these statements. Notice that the transformation we're doing here is *not* the gauge transformation, since A_μ doesn't transform – we're only doing the gauge transformation on the matter fields here, so their kinetic terms actually shift and produce the $\alpha\partial^\mu j_\mu$ term above. Photon field insertions in G don't contribute, since they have charge zero here.

Next, think about the LSZ formula for an S-matrix element with (say) two external photons:

$$\mathcal{M} = \langle\epsilon,...\epsilon_k...|S|...\rangle \overset{LSZ}{=} \epsilon^\mu\epsilon_k^\rho\mathbf{i}^n\int d^4x e^{\mathbf{i}px}\Box_{\mu\nu}\int d^4x_1 e^{\mathbf{i}p_kx_k}\Box^k_{\rho\sigma}\int ...\langle A^\nu(x)...A^\sigma(x_k)...\rangle \tag{3.11}$$

where $\Box_{\mu\nu}$ is shorthand for the photon kinetic operator $\Box_{\mu\nu} \equiv \Box\eta_{\mu\nu} - \partial_\mu\partial_\nu/\mu^2$, and I'm ignoring the wavefunction renormalization factors ($\sqrt{Z}$) for simplicity. The Schwinger-Dyson equation for A_μ then implies that

$$\Box^k_{\rho\sigma}\Box_{\mu\nu}\langle A^\nu(x)...A^\sigma(x_k)...\rangle = \Box^k_{\rho\sigma}\left(\langle j^\mu(x)...A^\sigma(x_k)...\rangle - \mathbf{i}\delta^4(x-x_k)\eta_{\mu\sigma}\langle...\rangle\right) \tag{3.12}$$

$$= \langle j_\mu(x)...j_\sigma(x_k)\rangle - \mathbf{i}\Box\delta(x-x_k)\eta_{\mu\sigma}\langle...\rangle \tag{3.13}$$

[18] You'll have to trust me for now that the path integral for fermionic fields exists. That's the only information about it we'll need here. Also I've absorbed the factor of Z^{-1} into $[D\Psi]$.

First of all, this is why I said we could get the S-matrix elements with photons from correlators with currents. But notice that this is only true up to the contact terms. But those are disconnected amplitudes which we can ignore.

Finally, set the polarization of one of the photons equal to its momentum $\epsilon = p$. Then

$$p^\mu \mathcal{M}_\mu = \epsilon_k^\rho \mathbf{i}^n \int d^4x e^{\mathbf{i}px} \int d^4x_1 e^{\mathbf{i}p_k x_k} \int dy e^{\mathbf{i}q_1 y} \left(\mathbf{i}\partial\!\!\!/_y + m_1\right) ... \langle -\mathbf{i}\partial_\mu j^\mu ... j_\rho(x_1) ... \Psi(y) .. \rangle \tag{3.14}$$

$$= \left(\not{q}_1 - m_1\right)\left(\not{q}_2 - m_2\right) \cdots \sum_j Q_j \tilde{G}(..., q_j \pm p, ...) \tag{3.15}$$

where the $\pm$ depends on whether particle j is incoming or outgoing. At the last step we used the Fourier transform of (3.10).

Now here's the punchline: The $\tilde{G}$ on the RHS of (3.15) has poles at $(q_j \pm p)^2 = m_j^2$, and *not* at $q_j^2 = m_j^2$. So when it's multiplied by $\not{q}_j - m_j = \frac{q_j^2 - m_j^2}{\not{q} + m}$ it will vanish. End of story. Notice that no use of perturbation theory was made here.

3.4 Feynman rules for QED

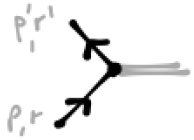
First, Feynman rules for Dirac fermion fields, more generally[19]. As always in these notes, time goes to the left, so I draw the initial state on the right (like the ket) and the final state on the left (like the bra).

1. An internal fermion line gives

$$= \frac{\mathbf{i}}{\not{k} - m_\Psi}$$

[19] Another good example of a QFT with interacting fermions is the Yukawa theory theory of a Dirac fermion field plus a scalar ϕ and an interaction

$$V = g\phi\bar{\Psi}\Psi \quad \Longrightarrow$$



$$= -\mathbf{i}g\delta^{rr'}. \tag{3.16}$$

Notice that in $3+1$ dimensions, $[g] = +4 - [\phi] - 2[\Psi] = 4 - 1 - 2\frac{3}{2} = 0$, the coupling is dimensionless. This describes more realistically the interactions between nucleons (which are fermions, as opposed to snucleons) and scalar pions, which hold together nuclei. It also is a crude sketch of the Higgs coupling to matter; notice that if ϕ is some nonzero constant $\langle\phi\rangle$, then there is a contribution to the mass of the fermions, $g\langle\phi\rangle$.

which is a matrix on the spinor indices.

There are four possibilities for an external fermion line of definite momentum. Here u, v are respectively the positive- and negative-energy solutions of the Dirac equation, $\left(\not{p} - m\right) u^r(k) = 0$, $\left(\not{p} + m\right) v^r(k) = 0$. (For a reminder, see *e.g.* §5.4 of my 215A lecture notes.)

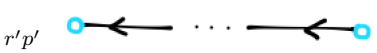
2.  $= \quad \overbrace{\Psi|k,r\rangle} = u^r(k)$

3.  $= \quad \langle k,r|\overbrace{\Psi} = \bar{u}^r(k)$

4.  $= \quad \overbrace{\bar{\Psi}|k,r\rangle} = \bar{v}^r(k)$

5.  $= \quad \langle k,r|\overbrace{\Psi} = v^r(k).$

6. Some advice: When evaluating a Feynman diagram with spinor particles, always begin at the head of the particle-number arrows on the fermion lines, and keep going along the fermion line until you can't anymore. This will keep the spinor indices in the form of matrix multiplication. Why: every Lagrangian you'll ever encounter has fermion parity symmetry, under which every fermionic field gets a minus sign; this means fermion lines cannot end, except on external legs. The result is always of the form of a scalar function (not a matrix or a spinor) made by sandwiching gamma matrices between external spinors:



$$= \sum_{a,b\ldots=1..4} \bar{u}^{r'}(p')_a \,(\text{pile of gamma matrices})_{ab}\, u^r(p)_b$$

Furthermore, in S-matrix elements the external spinors $u(p), v(p)$ satisfy the equations of motion $(\not{p} - m)u(p) = 0$, a fact which can be used to our advantage to shrink the pile of gammas.

There can also be fermion lines which form internal loops (though not at tree level, by definition). In this case, the spinor indices form a trace,

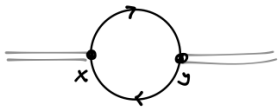
$$\sum_a (\text{pile of gamma matrices})_{aa} \equiv \text{tr}\,(\text{pile of gamma matrices})\,.$$

We'll learn to compute such traces below (around (3.18)); in fact, traces appear even in the case with external fermions if we do not measure the spins.

7. Diagrams related by exchanging external fermions have a relative minus sign.

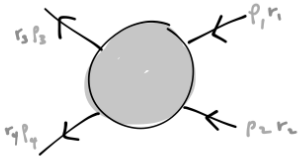
8. Diagrams with an odd number of fermion loops have an extra minus sign.

The last two rules are best understood by looking at an example in detail.

To understand rule 8 consider the following amplitude in the Yukawa theory with interaction (3.16): 

It is a contribution to the meson propagator. It is proportional to

$$\sum_{abcd} \overbrace{\bar{\Psi}_a(x)\Psi_b(x)\bar{\Psi}_c(y)\Psi_d(y)} = (-1)\mathrm{tr}\overbrace{\Psi(x)\bar{\Psi}(y)}\overbrace{\Psi(x)\bar{\Psi}(y)} = (-1)\mathrm{tr}S_F(x-y)S_F(x-y)$$

[Peskin page 119] To understand rule 7 consider $\Psi\Psi \to \Psi\Psi$ (nucleon) scattering in the Yukawa theory: 

The blob represents the matrix element

$$_0\langle p_3 r_3; p_4 r_4|\, \mathcal{T} e^{-\mathbf{i}\int V d^4 z}\, |p_1 r_1; p_2 r_2\rangle_0$$

where the initial state is

$$|p_1 r_1; p_2 r_2\rangle_0 \propto \mathbf{a}^{r_1\dagger}_{p_1}\mathbf{a}^{r_2\dagger}_{p_2}|0\rangle$$

and the final state is

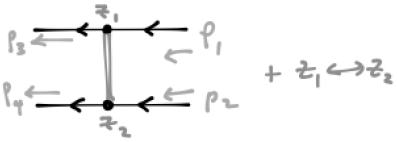
$$_0\langle p_3 r_3; p_4 r_4| = (|p_3 r_3; p_4 r_4\rangle_0)^\dagger \propto \langle 0|\, \mathbf{a}^{r_4}_{p_4}\mathbf{a}^{r_3}_{p_3} = -\langle 0|\, \mathbf{a}^{r_3}_{p_3}\mathbf{a}^{r_4}_{p_4}$$

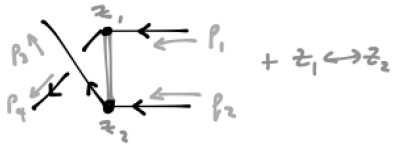
where note that the dagger reverses the order.

The leading contribution comes at second order in V:

$$_0\langle p_3 r_3; p_4 r_4|\, \mathcal{T}\left(\frac{1}{2!}(\mathbf{i}g)^2 \int d^4 z_1 \int d^4 z_2 \left(\bar{\Psi}\Psi\phi\right)_1 \left(\bar{\Psi}\Psi\phi\right)_2\right) |p_1 r_1; p_2 r_2\rangle_0$$

To get something nonzero we must contract the ϕs with each other. The diagrams at right indicate best the possible ways to contract the fermions. Exchanging the roles of z_1 and z_2 interchanges two pairs of fermions so costs no signs and cancels the $\frac{1}{2!}$.
The overall sign is annoying but can be fixed by demanding that the diagonal bit of the S-matrix give





$$\langle p_3 p_4|\, (\mathbb{1} + ...)\, |p_1 p_2\rangle = +\delta(p_1 - p_3)\delta(p_2 - p_4) + \cdots$$

The relative sign is what we're after, and it comes by comparing the locations of fermion

operators in the contractions in the two diagrams at right. In terms of the contractions, these $t-$ and $u-$ channel diagrams are related by leaving the annihilation operators alone and switching the contractions between the creation operators and the final state. Denoting by $\mathbf{a}^\dagger_{1,2}$ the fermion creation operators coming from the vertex at $z_{1,2}$,

$$\begin{aligned}\langle 0|\,\mathbf{a}_{p_4}\underbrace{\mathbf{a}_{p_3}\mathbf{a}_1^\dagger}\,\mathbf{a}_2^\dagger...\;+\;\langle 0|\,\mathbf{a}_{p_4}\,\underbrace{\mathbf{a}_{p_3}\mathbf{a}_1^\dagger\mathbf{a}_2^\dagger}...\\ =\langle 0|\,\mathbf{a}_{p_4}\underbrace{\mathbf{a}_{p_3}\mathbf{a}_1^\dagger}\,\mathbf{a}_2^\dagger...\;-\;\langle 0|\,\underbrace{\mathbf{a}_{p_4}\mathbf{a}_1^\dagger}\,\underbrace{\mathbf{a}_{p_3}\mathbf{a}_2^\dagger}...\end{aligned}$$

In the last expression the fermion operators to be contracted are all right next to each other and we see the relative minus sign.

While we're at it, let's evaluate this whole amplitude to check the Feynman rules I've claimed and get some physics out. It is

$$S_{fi}=-g^2\int dz_1dz_2\int \mathrm{d}\!\!\!^4q\frac{e^{-\mathbf{i}q(z_1-z_2)}\mathbf{i}}{q^2-m^2+\mathbf{i}\epsilon}\left(e^{-\imath z_2(p_1-p_3)}\bar{u}^{r_3}(p_3)u^{r_1}(p_1)\cdot e^{-\imath z_1(p_2-p_4)}\bar{u}^{r_4}(p_4)u^{r_2}(p_2)-(3\leftrightarrow 4)\right).$$

In the first (t-channel) term, the integrals over $z_{1,2}$ gives $\not{\delta}(p_1-p_3-q)\not{\delta}(p_2-p_4-q)$, and the q integral then gives $\delta(p_1+p_2-p_3-p_4)$, overall momentum conservation. In the second (u-channel) term, $q=p_1-p_4=p_3-p_2$. Altogether,

$$S_{fi}=\mathbb{1}+\not{\delta}^4(p_T)\mathbf{i}\mathcal{M}$$

with, to leading order,

$$\mathbf{i}\mathcal{M}=-\mathbf{i}g^2\left(\frac{1}{t-m^2}\left(\bar{u}_3u_1\right)\left(\bar{u}_4u_2\right)-\frac{1}{u-m^2}\left(\bar{u}_4u_1\right)\left(\bar{u}_3u_2\right)\right)\tag{3.17}$$

with $t\equiv(p_1-p_3)^2, u\equiv(p_1-p_4)^2$. This minus sign implements Fermi statistics.

Yukawa force revisited. In the non-relativistic limit, we can again relate this amplitude to the force between particles, this time with the actual spin and statistics of nucleons. In the COM frame, $p_1=(m,\vec{p}), p_2=(m,-\vec{p})$ and $p_3=(m,\vec{p}'), p_4=(m,-\vec{p}')$. In the non-relativistic limit, the spinors become $u^r_p=\begin{pmatrix}\sqrt{\sigma\cdot p}\xi^r\\ \sqrt{\bar{\sigma}\cdot p}\xi^r\end{pmatrix}\to\sqrt{m}\begin{pmatrix}\xi^r\\ \xi^r\end{pmatrix}$ so that $\bar{u}_3u_1\equiv\bar{u}(p_3)^{r_3}u(p_1)^{r_1}=2m\xi^\dagger_{r_3}\xi_{r_1}=2m\delta_{r_3r_1}$. Let's simplify our lives and take two *distinguishable* fermions (poetically, they could be proton and neutron, but let's just add a label to our fermion fields; they could have different masses, for example, or different couplings to ϕ, call them g_1,g_2). Then we only get the t-channel diagram. The intermediate scalar momentum is $q=p_1-p_3=(0,\vec{p}-\vec{p}')$ so $t=(p_1-p_3)^2=-\vec{q}^2=-\left(\vec{p}-\vec{p}'\right)^2$ and

$$\mathbf{i}\mathcal{M}_{NR,COM}=\mathbf{i}g_1g_2\frac{1}{\vec{q}^2+m_\phi^2}4m^2\delta^{r_1r_3}\delta^{r_2r_4}.$$

Compare this to the NR Born approximation matrix element

$$\begin{aligned} 2\pi\delta(E_p - E_{p'})\left(-\mathbf{i}\tilde{V}(\vec{q})\right) &= {}_{NR}\langle\vec{p}'|\, S\,|\vec{p}\rangle_{NR} \\ &= \sum_{r_4}\int \mathrm{d}\!\!^{-3}p_4 V \underbrace{\prod_{i=1}^{4}\frac{1}{\sqrt{2E_i}}}_{=\frac{1}{\sqrt{2m}^4}} S(34\leftarrow 12) \\ &= 2\pi\delta(E_p - E_{p'})\delta^{r_1 r_3}\frac{\mathbf{i}g_1 g_2}{\vec{q}^2 + m_\phi^2} \end{aligned}$$

where in the second line we summed over possible final states of the second (target) particle, and corrected the relativistic normalization, so that ${}_{NR}\langle\vec{p}'|\vec{p}\rangle_{NR} = \not{\delta}^3(p-p')$. This is completely independent of the properties of the second particle. We infer that the scalar mediates a force with potential $U(x) = -\frac{g_1 g_2 e^{-m_\phi r}}{4\pi r}$. It is attractive if $g_1 g_2 > 0$.

[End of Lecture 7]

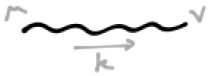
Back to QED. The new ingredients in QED are the propagating vectors, and the interaction hamiltonian $V = e\bar{\Psi}\gamma^\mu\Psi A_\mu$. The rest of the Feynman rules are

9. The interaction vertex gets a

$$= \quad -\mathbf{i}e\gamma^\mu$$

10. An external photon in the initial state gets a $\epsilon^\mu(p)$, and in the final state gets a $\epsilon^{\mu\star}(p)$.

11. An internal photon line gets a



$$= \quad \frac{\mathbf{i}}{k^2 - m_\gamma^2}\left(-\eta^{\mu\nu} + (1-\xi)k^\mu k^\nu/k^2\right)$$

where $m_\gamma = 0$ (it's sometimes useful to keep it in there for a while as an IR regulator) and the value of ξ is up to you (meaning that your answers for physical quantities should be independent of ξ).

Spinor trace ninjutsu.

The trace is cyclic: $\quad \mathrm{tr}\,(AB\cdots C) = \mathrm{tr}\,(CAB\cdots)\,. \qquad (3.18)$

Our gamma matrices are 4×4, so $\quad \mathrm{tr}\mathbb{1} = 4.$

$$\mathrm{tr}\gamma^\mu = \mathrm{tr}\left(\gamma^5\right)^2 \gamma^\mu \overset{(3.18)}{=} \mathrm{tr}\gamma^5\gamma^\mu\gamma^5 \overset{\{\gamma^5,\gamma^\mu\}=0}{=} -\mathrm{tr}\gamma^\mu = 0. \tag{3.19}$$

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.

$$\mathrm{tr}\gamma^\mu\gamma^\nu \overset{\text{clifford}}{=} -\mathrm{tr}\gamma^\nu\gamma^\mu + 2\eta^{\mu\nu}\mathrm{tr}\mathbb{1} \overset{(3.18)}{=} -\mathrm{tr}\gamma^\mu\gamma^\nu + 8\eta^{\mu\nu} \implies \mathrm{tr}\gamma^\mu\gamma^\nu = 4\eta^{\mu\nu}. \tag{3.20}$$

$$\mathrm{tr}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma = 4\left(\eta^{\mu\nu}\eta^{\rho\sigma} + \eta^{\sigma\mu}\eta^{\nu\rho} - \eta^{\mu\rho}\eta^{\nu\sigma}\right). \tag{3.21}$$

Why is this? The completely antisymmetric bit vanishes because it is proportional to γ^5 which is traceless (by the same argument as (3.19)). If any pair of indices is the same then the other two must be too by (3.20). 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.

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

3.5 QED processes at leading order

Now we are ready to do *lots* of examples, nearly all of which (when pushed to the end) predict cross sections which are verified by experiments to about one part in 137.[20] Here $\frac{1}{137} \approx \alpha \equiv \frac{e^2}{4\pi}$ is the small number by which the next order corrections are suppressed. [21]

Did I mention that the antiparticle of the electron, predicted by the quantum Dirac theory (*i.e.* by Dirac), is the positron? It has the same mass as the electron and the opposite electromagnetic charge, since the charge density is the 0 component of the electromagnetic current, $j^\mu = \bar{\Psi}\gamma^\mu\Psi$, so the charge is

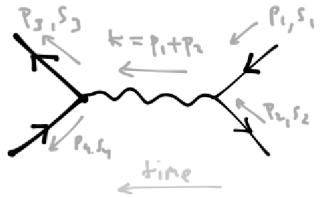
$$\int d^3x j^0(x) = \int \bar{\Psi}\gamma^0\Psi = \int \Psi^\dagger\Psi = \int \mathrm{d}^3p \sum_s \left(\mathbf{a}^\dagger_{p,s}\mathbf{a}_{ps} - \mathbf{b}^\dagger_{p,s}\mathbf{b}_{ps}\right).$$

[20] I guess it is this overabundance of scientific victory in this area that leads to the intrusion of so many names of physicists in the following discussion.

[21] This statement is true naively (in the sense that the next diagrams which are nonzero come with two more powers of e), and also true in fact, but in between naiveté and the truth is a long road of renormalization, which begins in the next section.

So $\mathbf{b}^\dagger$ creates a positron.

[Schwarz §13.3, Peskin §5.1] Perhaps the simplest to start with is scattering of electrons and positrons. We can make things even simpler (one diagram instead of two) by including also the muon, which is a heavy version of the electron[22], and asking about the process $\mu^+\mu^- \leftarrow e^+e^-$. At leading order in e, this comes from

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

$$= (-\mathbf{i}e\bar{u}^{s_3}(p_3)\gamma^\mu v^{s_4}(p_4))_{\text{muons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)k_\mu k_\nu}{k^2}\right)}{k^2} (-\mathbf{i}e\bar{v}^{s_2}(p_2)\gamma^\nu u^{s_1}(p_1))_{\text{electrons}} \quad (3.22)$$

with $k \equiv p_1 + p_2 = p_3 + p_4$ by momentum conservation at each vertex. I've labelled the spinors according to the particle types, since they depend on the mass.

Ward identity in action. What about the $k_\mu k_\nu$ term in the photon propagator? The spinors satisfy their equations of motion, $\not{p}_1 u_1 = m_e u_1$ (where $u_1 \equiv u_{p_1}^{s_1}$ for short) and $\bar{v}_2 \not{p}_2 = -m_e \bar{v}_2$. The k_ν appears in

$$k_\nu \bar{v}_2 \gamma^\nu u_1 = \bar{v}_2 \left(\not{p}_1 + \not{p}_2\right) u_1 = \bar{v}_2 \not{p}_1 u_1 + \bar{v}_2 \not{p}_2 u_1 = (m - m)\bar{v}u = 0.$$

(The other factor is also zero, but one factor of zero is enough.) Therefore

$$\mathcal{M} = \frac{e^2}{s} \bar{u}_3 \gamma_\mu v_4 \cdot \bar{v}_2 \gamma^\mu u_1$$

where $s \equiv k^2 = (p_1 + p_2)^2 = E_{CoM}^2$ is the Mandelstam variable. And I am relying on you to remember which spinors refer to muons (3,4) and which to electrons (1,2).

Squaring the amplitude. We need to find $\mathcal{M}^\dagger$ (the dagger here really just means complex conjugate, but let's put dagger to remind ourselves to transpose and reverse the order of all the matrices). Recall the special role of γ^0 here:

$$\gamma_\mu^\dagger \gamma_0 = \gamma_0 \gamma_\mu, \quad \gamma_0^\dagger = \gamma_0.$$

This means that for any two Dirac spinors,

$$\left(\bar{\Psi}_1 \gamma^\mu \Psi_2\right)^\dagger = \bar{\Psi}_2 \gamma^\mu \Psi_1.$$

[22] Who ordered *that*? (I. I. Rabi's reaction to learning about the muon.) I hope you don't find it too jarring that the number of 'elementary' particles in our discussion increased by three in the last two paragraphs. People used to get really disgruntled about this kind of thing. But here we have, at last, uncovered the true purpose of the muon, which is to halve the number of Feynman diagrams in this calculation (compare (3.28)).

(This is the same manipulation that showed that the Dirac Lagrangian was hermitian.) So

$$\mathcal{M}^\dagger = \frac{e^2}{s}\left(\bar{v}_4\gamma^\mu u_3\right)\left(\bar{u}_1\gamma_\mu v_2\right).$$

and therefore

$$|\mathcal{M}_{\mu^+\mu^-\leftarrow e^+e^-}|^2 = \frac{e^4}{s^2}\underbrace{\left(\bar{v}_4\gamma^\mu u_3\right)\left(\bar{u}_3\gamma^\nu v_4\right)}_{\text{out}}\cdot\underbrace{\left(\bar{u}_1\gamma_\mu v_2\right)\left(\bar{v}_2\gamma_\nu u_1\right)}_{\text{in}}. \tag{3.23}$$

These objects in parentheses are just c-numbers, so we can move them around, no problem. I've grouped them into a bit depending only on the initial state (the electron stuff 1, 2) and a bit depending only on the final state (the muon stuff 3,4).

Average over initial, sum over final. In the amplitude above, we have fixed the spin states of all the particles. Only very sophisticated experiments are able to discern this information. So suppose we wish to predict the outcome of an experiment which does not measure the spins of the fermions involved. We must sum over the final-state spins using

$$\sum_{s_4} v_a^{s_4}(p_4)\bar{v}_b^{s_4}(p_4) = \left(\not{p}_4 - m_\mu\right)_{ab} = \sum_{s_4}\bar{v}_b^{s_4}(p_4)v_a^{s_4}(p_4)$$

(where I wrote the last expression to emphasize that these are just c-numbers) and

$$\sum_{s_3} u_a^{s_3}(p_3)\bar{u}_b^{s_3}(p_3) = \left(\not{p}_3 + m_\mu\right)_{ab}.$$

Looking at just the 'out' factor of $|\mathcal{M}|^2$ in (3.23), we see that putting these together produces a spinor trace, as promised:

$$\begin{aligned}
& \sum_{s_3,s_4}\left(\bar{u}(p_3)_a^{s_3}\gamma^\mu_{ab}\underbrace{v(p_4)_b^{s_4}\left(\bar{v}(p_4)_c^{s_4}\right.}_{(\not{p}_4-m_\mu)_{bc}}\gamma^\nu_{cd}u^{s_3}(p_3)_d\right) \\
= \quad & \gamma^\mu_{ab}(\not{p}_4 - m_\mu)_{bc}\gamma^\nu_{cd}(\not{p}_3 + m_\mu)_{da} \\
= \quad & \mathrm{tr}\left(\gamma^\mu\left(\not{p}_4 - m_\mu\right)\gamma^\nu\left(\not{p}_3 + m_\mu\right)\right) \\
= \quad & p_{4\rho}p_{3\sigma}\mathrm{tr}\gamma^\mu\gamma^\rho\gamma^\nu\gamma^\sigma - m_\mu^2\mathrm{tr}\gamma^\mu\gamma^\nu \\
\overset{(3.20),(3.21)}{=} \quad & 4\left(p_4^\mu p_3^\nu + p_3^\nu p_4^\mu - \underbrace{p_3\cdot p_4}_{\equiv p_{34}}\eta^{\mu\nu} - m_\mu^2\eta^{\mu\nu}\right)
\end{aligned} \tag{3.24}$$

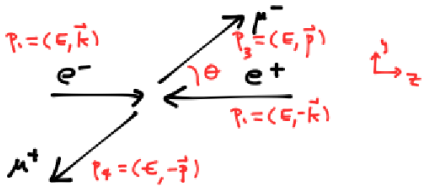
If also we don't know the initial (electron) spins, then the outcome of our experiment is the *average* over the initial spins, of which there are four possibilities. Therefore, the relevant probability for unpolarized scattering is

$$\frac{1}{4}\sum_{s_{1,2,3,4}}|\mathcal{M}|^2 \quad = \quad \frac{e^4}{4s^2}\mathrm{tr}\left(\gamma^\mu\left(\not{p}_4 - m_\mu\right)\gamma^\nu\left(\not{p}_3 + m_\mu\right)\right)\mathrm{tr}\left(\gamma_\nu\left(\not{p}_2 - m_e\right)\gamma_\mu\left(\not{p}_1 + m_e\right)\right)$$

$$\overset{(3.24)\text{ twice}}{=} \frac{8e^4}{s^2}\left(p_{13}p_{24}+p_{14}p_{23}+m_\mu^2 p_{12}+m_e^2 p_{34}+2m_e^2 m_\mu^2\right)$$
$$\overset{\text{algebra}}{=} \frac{2e^4}{s^2}\left(t^2+u^2+4s(m_e^2+m_\mu^2)-2(m_e^2+m_\mu^2)^2\right) \tag{3.25}$$

In the second step of (3.25) the $p_{12}p_{34}$ terms cancel. In the last step of (3.25) we used all the Mandelstam variables:

$$\begin{aligned} s &\equiv (p_1+p_2)^2 = (p_3+p_4)^2 = \quad E_{CoM}^2 = 4E^2 \\ t &\equiv (p_1-p_3)^2 = (p_2-p_4)^2 = m_e^2+m_\mu^2-2E^2+2\vec{k}\cdot\vec{p} \\ u &\equiv (p_1-p_4)^2 = (p_2-p_3)^2 = m_e^2+m_\mu^2-2E^2-2\vec{k}\cdot\vec{p} \end{aligned}$$



where the particular kinematic variables (in the rightmost equalities) are special to this problem, in the center of mass frame (CoM), and are defined in the figure at right. Really there are only two independent Lorentz-invariant kinematical variables, since $s+t+u=\sum_i m_i^2$.

Now we can use the formula for a differential cross section with a two-body final state, in the CoM frame (for the derivation, see these notes, §4.7):

$$\begin{aligned} \left(\frac{d\sigma}{d\Omega}\right)_{CoM} &= \frac{1}{64\pi^2 E_{CoM}^2}\frac{|\vec{p}|}{|\vec{k}|}\left(\frac{1}{4}\sum_{\text{spins}}|\mathcal{M}|^2\right) \\ &= \frac{\alpha^2}{16E^6}\frac{|\vec{p}|}{|\vec{k}|}\left(E^4+|\vec{k}|^2|\vec{p}|^2\cos^2\theta+E^2(m_e^2+m_\mu^2)\right) \end{aligned} \tag{3.26}$$

where $\alpha \equiv \frac{e^2}{4\pi}$ is the fine structure constant. This can be boiled a bit with kinematical relations $|\vec{k}| = \sqrt{E^2-m_e^2}, |\vec{p}| = \sqrt{E^2-m_\mu^2}$ to make manifest that it depends only on two independent kinematical variables, which we can take to be the CoM energy E and the scattering angle θ in $\vec{k}\cdot\vec{p} = |\vec{k}||\vec{p}|\cos\theta$ (best understood from the figure). It simplifies a bit if we take $E \gg m_e$, and more if we take $E \gg m_\mu \sim 200 m_e$, to

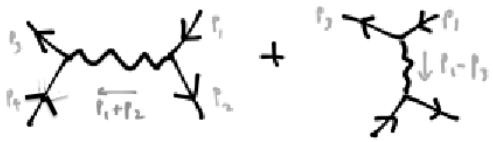
$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4E_{CoM}^2}\left(1+\cos^2\theta\right). \tag{3.27}$$

In fact, the two terms here come respectively from spins transverse to the scattering plane and in the scattering plane; see Schwartz §5.3 for an explanation.

There is a lot more to say about what happens when we scatter an electron and a positron! Another thing that can happen is that the final state could be an electron and positron again (Bhabha scattering[23]).

[23] See figure 3 here. Now remember that a person doesn't have much control over their name. By the way, I totally believe the bit about *non-perturbative strings = lint*.

They are not necessarily the same e^- and e^+, though (except in the sense that they are all the same), because another way to get there at tree level is the second, t-channel, diagram, at right. The intermediate photon in that diagram has $k_t = (p_1 - p_3)$, so that the denominator of the propagator is $t = k_t^2 = (p_1 - p_3)^2$ instead of s.



Squaring this amplitude gives

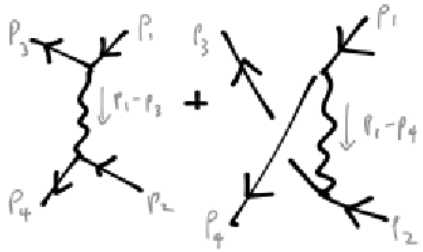
$$|\mathcal{M}_s + \mathcal{M}_t|^2 = |\mathcal{M}_s|^2 + |\mathcal{M}_t|^2 + 2\mathrm{Re}\,(\mathcal{M}_s \mathcal{M}_t^\star), \tag{3.28}$$

interference terms. Interference terms mean that you have to be careful about the overall sign or phase of the amplitudes.

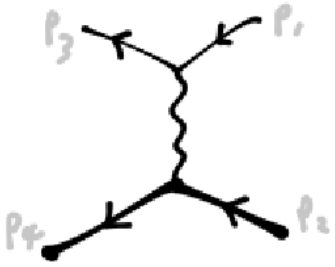
You may be surprised that the cross section (3.27) decreases with energy. Mechanically this comes mainly from the $1/s^2$ from the photon propagator: as s grows, the intermediate photon is more and more off-shell. But more deeply, it's because above we've studied an *exclusive* cross-section, in the sense that we fixed the final state to be exactly a muon and an antimuon. At higher energies, nothing new happens, because the final state is fixed.

It has also been very valuable to think about *inclusive* cross-sections for e^+e^- scattering, because in this way you can make anything that the s-channel photon couples to, if you put enough energy into it. The inclusive cross section for (e^+e^- goes to anything) does grow with energy, and jumps at energies which are thresholds for new particles in the final state. In this way, for example, we can also make *quarks* (more specifically quark-antiquark pairs) since they also carry electric charge. See Peskin pp 139-140 (and our later discussion in §5.3) for a bit more about that, and in particular how this observable gives evidence that there are three colors of quarks.

$\boxed{e^-e^- \leftarrow e^-e^-}$. What happens if instead we scatter two electrons (Möller scattering)? In that case, the leading order diagrams are the ones at right. Now the intermediate photons have $k_t = (p_1 - p_3)$ and $k_u = (p_1 - p_4)$ respectively, so that the denominator of the propagator is t and u in the two diagrams. The evaluation of these diagrams has a lot in common with the ones for $e^+e^- \to e^+e^-$, namely you just switch some of the legs between initial and final state.



The relation between such amplitudes is called *crossing symmetry*. Let's illustrate it instead for $e^-\mu^- \leftarrow e^-\mu^-$, where again there is only one diagram, related by crossing to (3.30). The diagram is the one at right. (The muon is the thicker fermion line.)



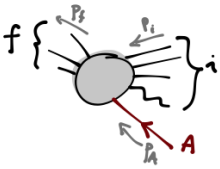
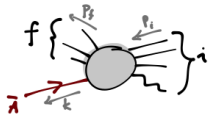
$$\mathbf{i}\mathcal{M} = \quad = (-\mathbf{i}e\bar{u}_3\gamma^\mu u_1))_{\text{electrons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)k_\mu k_\nu}{k^2}\right)}{k^2} (-\mathbf{i}e\bar{u}_2\gamma^\nu u_4)_{\text{muons}} \tag{3.29}$$

with $k \equiv p_1 - p_3 = p_2 - p_4$. It differs from (3.30) by replacing the relevant vs with us for the initial/final antiparticles that were moved into final/initial particles, and relabelling the momenta. After the spin sum,

$$\frac{1}{4}\sum_{s_{1,2,3,4}} |\mathcal{M}|^2 = \frac{e^4}{4t^2}\mathrm{tr}\left(\gamma^\mu\left(\not{p}_4 + m_\mu\right)\gamma^\nu\left(\not{p}_2 + m_\mu\right)\right)\mathrm{tr}\left(\gamma_\nu\left(\not{p}_3 + m_e\right)\gamma_\mu\left(\not{p}_1 + m_e\right)\right)$$

this amounts to the replacement $(p_1, p_2, p_3, p_4) \to (p_1, -p_3, p_4, -p_2)$; on the Mandelstam variables, this is just the permutation $(s, t, u) \to (t, u, s)$.

Crossing symmetry more generally. 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:

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

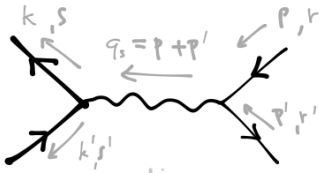
(If you must, note that this is another sense in which an antiparticle is a particle going backwards in time.) If A is a spinor particle, the sum relations for particles and antiparticles are different:

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

– 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. We'll study a well-motivated example in more detail next.

Mott formula. By studying scattering of an electron from a heavy charged fermion (a muon is convenient) we can reconstruct the cross section for scattering off a Coulomb potential (named after Mott). This example will be important later in §4, where we'll figure out how it is corrected by other QED processes.

$\boxed{\mu^+\mu^- \leftarrow e^+e^-}$**.** Consider again 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. We found the amplitude

$$\mathbf{i}\mathcal{M}_{\mu^+\mu^-\leftarrow e^+e^-} = \quad \text{  } \quad = \left(-\mathbf{i}e\bar{u}^s(k)\gamma^\mu v^{s'}(k')\right)_{\text{muons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu q_\nu}{q^2}\right)}{q^2} \left(-\mathbf{i}e\bar{v}^{r'}(p')\gamma^\nu u^r(p)\right)_{\text{electrons}} \tag{3.30}$$

(with $q \equiv p + p' = k + k'$)[24] and the (unpolarized) scattering probability density

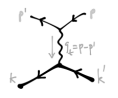
$$\frac{1}{4}\sum_{\text{spins}} |\mathcal{M}|^2 \overset{\text{spinor traces}}{=} \frac{1}{4}\frac{e^4}{s^2} E^{\mu\nu} M_{\mu\nu},$$

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 $s = q^2 = (p+p')^2$.

$\boxed{e^-\mu^- \leftarrow e^-\mu^-}$**.** To get from this 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 $\mu^+\mu^- \leftarrow e^+e^-$ (3.30), we must replace the relevant vs with us for the initial/final antiparticles that were moved into final/initial particles, and we must replace $p' \to -p', k' \to -k'$:

$$\mathbf{i}\mathcal{M} = \quad \text{  } \quad = \left(-\mathbf{i}e\bar{u}(p')\gamma^\mu u(p)\right)_{\text{electrons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu^t q_\nu^t}{q_t^2}\right)}{q_t^2} \left(-\mathbf{i}e\bar{u}(k)\gamma^\nu u(k')\right)_{\text{muons}} \tag{3.31}$$

[24] Relative to the notation I used earlier, $p_1 = p, p_2 = p', p_3 = k, p_4 = k'$.

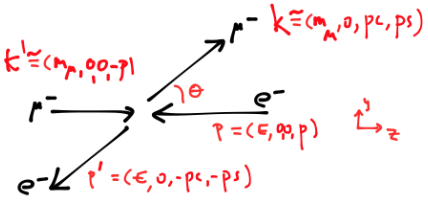
with $q_t \equiv p - p' = k - k'$. After the spin sum,

$$\frac{1}{4}\sum_{s,s',r,r'} |\mathcal{M}|^2 = 4\frac{e^4}{t^2}\ \left(-p_\mu p'_\nu - p'_\mu p_\nu - \eta_{\mu\nu}(-p\cdot p' + m_e^2)\right) \cdot \left(-k_\mu k'_\nu - k'_\mu k_\nu - \eta_{\mu\nu}(-k\cdot k' + m_\mu^2)\right) \tag{3.32}$$

On the Mandelstam variables, this is just the permutation $(s,t,u) \to (t,u,s)$.

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. '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, $E' \simeq E$. In the diagram of the kinematics at right, annoyingly, $s \equiv \cos\theta, c \equiv \sin\theta$. (Sorry.)



This means that the muon-line tensor factor $M_{\mu\nu}$ in (4.11) simplifies dramatically:

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

In the electron line, we'll need the ingredient

$$-p\cdot p' + m_e^2 = -E^2 + \vec{p}^2\cos\theta + m_e^2 = -\vec{p}^2(1-\cos\theta). \tag{3.33}$$

So

$$\begin{aligned} E^{\mu\nu}M_{\mu\nu} = 32m_\mu^2 E^{00} &= 32m_\mu^2(2E^2 + \eta^{00}(-p\cdot p' + m_e^2)) \\ &\overset{(3.33)}{=} 32m_\mu^2(2E^2 - \vec{p}^2(1-\cos\theta)) \\ &\overset{\text{trig}}{=} 32m_\mu^2 2(E^2 - \vec{p}^2\sin^2\theta/2) \overset{\beta^2 \equiv \vec{p}^2/E^2}{=} 64m_\mu^2 E^2(1-\beta^2\sin^2\theta/2)\ . \end{aligned}$$

From the two-body phase space, the cross section is

$$d\sigma = \underbrace{\frac{1}{v_{\rm rel}}}_{=\beta}\frac{1}{2E}\frac{1}{2m_\mu}\frac{4z^2e^4}{t^2}64m_\mu^2E^2(1-\beta^2\sin^2\theta/2)\frac{d\Omega}{16\pi^2}\frac{p}{E_{\rm total}}$$
$$\overset{E_{\rm total}\sim m_\mu}{=} \frac{4E}{\beta}\frac{z^2e^4(1-\beta^2\sin^2\theta/2)}{t^2}\frac{d\Omega}{16\pi^2}.$$

Noting that $t=(p-p')^2=-2\vec{p}^2(1-\cos\theta)$, we get

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

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.

Electron-proton scattering. The answer is basically the same if we think of the heavy particle in (4.10) as a proton (we have to flip the sign of the charge but this gets squared away since there is no interference in this case). $ep\to ep$ is called Rutherford scattering, for good reason[25]. More generally, the Mott formula applies to scattering electrons off of heavy *pointlike* charged particles. For ep collisions at high enough energies, this formula fails because the proton has structure. At even higher energies it works again because the electron scatters off pointlike, approximately free quarks.

Electron-photon scattering. In the case of the process $e^-\gamma\leftarrow e^-\gamma$, [26] we meet a new ingredient, namely external photons:

$$\mathbf{i}\mathcal{M} = \quad + \quad \equiv \mathbf{i}\mathcal{M}_s + \mathbf{i}\mathcal{M}_t$$
$$= (-\mathbf{i}e)^2\epsilon_1^\mu\epsilon_4^{\star\nu}\bar{u}_3\left(\gamma_\nu\frac{\mathbf{i}\not{k}_s+m}{s-m^2}\gamma_\mu+\gamma_\mu\frac{\mathbf{i}\not{k}_t+m}{t-m^2}\gamma_\nu\right)u_2\ . \qquad (3.34)$$

The two amplitudes have a relative plus since we only mucked with the photon contractions, they just differ by how the gamma matrices are attached. If you don't believe me, draw the contractions on this:

$$\langle\gamma e|\,(\bar\Psi\not{A}\Psi)_1(\bar\Psi\not{A}\Psi)_2\,|\gamma e\rangle$$

[25] If you don't know why, you should go read *Inward Bound*, by Abraham Pais, as soon as possible.

[26] which at high energy is called Compton scattering and at low energies is called Thomson scattering. Despite my previous curmudgeonly footnote chastising the innocent reader for an imagined incomplete knowledge of the history of science, I do have a hard time remembering which name goes where. Moreover, as much as I revere the contributions of many of these folks, I find that using their names makes me think about the people instead of the physics. No one owns the physics! It's the same physics for lots of space aliens, too.

(I'm not going to TeX it, thank you).

Now, if we don't measure the photon polarizations, we need

$$P = \frac{1}{4} \sum_{\text{polarizations, spins}} |\mathcal{M}|^2.$$

The key ingredient is the completeness relation

$$\sum_{i=1,2} \epsilon_\mu^{i\star}(k)\epsilon_\nu^i(k) = -\eta_{\mu\nu} + \text{something proportional to } k^\mu k^\nu.$$

We can do various incantations to find a definite coefficient of $k^\mu k^\nu$, but *it will not matter* because of the Ward identity: anytime there is an external photon $\epsilon(k)_\mu$, the amplitude is $\mathcal{M} = \mathcal{M}_\mu \epsilon^\mu(k)$ and satisfies $k^\mu \mathcal{M}_\mu = 0$. Therefore, we can ignore the term about which I was vague and we have

$$\begin{aligned}\sum_{\text{polarizations}} |\mathcal{M}|^2 &= \sum_i \epsilon_\mu^{i\star} \mathcal{M}^{\mu\star} \mathcal{M}^\nu \epsilon_\nu^i = -\eta_{\mu\nu} \mathcal{M}^{\mu\star} \mathcal{M}^\nu + (\text{terms with } \mathcal{M}_\mu k^\mu) \\ &= -\mathcal{M}_\mu^\star \mathcal{M}^\mu.\end{aligned}$$

Don't be scared of the minus sign, it's because of the mostly minus signature, and makes the thing positive. But notice the opportunity to get negative probabilities if the gauge bosons don't behave!

A dramatic process related by crossing to Compton scattering is pair annihilation, $\mathcal{M}_{\gamma\gamma \leftarrow e^+e^-}$. See the end of Peskin §5, where he has a nice plot comparing to experimental data the result for $\frac{d\sigma}{d\Omega}$ as a function of scattering angle.

[End of Lecture 8]

4 To infinity and beyond

At this point we are capable of successfully computing the amplitudes and cross-sections for many processes using QED. More precisely, we can do a good job of the leading-order-in-α amplitudes, using Feynman diagrams which are trees – no 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. When we do that we're going to encounter some confusing stuff. A place we've already encountered this stuff is in the additive constant in the Hamiltonian; this has physical consequences as we'll see in thinking about Casimir forces in §4.1.

We don't encounter these short-distance issues in studying tree-level diagrams because in a tree-level diagram, the quantum numbers (and in particular the momenta) of the intermediate states are fixed by the external states. 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, about which surely we have no information yet.

In order to put ourselves in the right frame of mind to think about that stuff, we'll make a brief retreat in §4.2 to systems with finitely many degrees of freedom.

Then we'll apply some of these lessons to a toy field theory example (scalar field theory). Then we'll come back to perturbation theory in QED. Reading assignment for this chapter: Zee §III.

4.1 Casimir effect: vacuum energy is real

[A. Zee, *Quantum Field Theory in a Nutshell*, §I.9] This subsection has two purposes. One is to show that the $\frac{1}{2}\hbar\omega$ energy of the vacuum of the quantum harmonic oscillator (which appeared in our discussion of quantum sound and light) 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 in the case of many oscillators as for the radiation field) and compare the difference.

The other purpose is to give an object lesson in asking the right questions. In physics, the right question is often a question which 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.

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

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

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[27]. Replace

$$\begin{aligned} f(d) \rightsquigarrow f(a,d) &= \hbar\frac{\pi}{2d}\sum_{j=1}^{\infty} e^{-a\omega_j/\pi} j \\ &= -\frac{\pi\hbar}{2}\partial_a \underbrace{\left(\sum_{j=1}^{\infty} e^{-aj/d}\right)}_{=\frac{1}{1-e^{-a/d}}-1} \\ &= +\frac{\pi\hbar}{2d}\frac{e^{a/d}}{\left(e^{a/d}-1\right)^2} \\ &\stackrel{a\ll d}{\simeq} \hbar\left(\underbrace{\frac{\pi d}{2a^2}}_{\to\infty \text{ as } a\to 0} - \frac{\pi}{24d} + \frac{\pi a^2}{480 d^3} + \ldots\right) \end{aligned} \tag{4.1}$$

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

$$F = -\partial_d E_0 = -\left(f'(d) - f'(L-d)\right)$$

[27] 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.

$$
\begin{aligned}
&= -\hbar\left(\left(\frac{\pi}{2a^2}+\frac{\pi}{24d^2}+\mathcal{O}(a^2)\right)-\left(\frac{\pi}{2a^2}+\frac{\pi}{24\left(L-d\right)^2}+\mathcal{O}(a^2)\right)\right)\\
&\overset{a\to 0}{=} -\frac{\pi\hbar}{24}\left(\frac{1}{d^2}-\frac{1}{(L-d)^2}\right)\\
&\overset{L\gg d}{=} -\frac{\pi\hbar c}{24d^2}\left(1+\mathcal{O}(d/L)\right)\ . \qquad (4.2)
\end{aligned}
$$

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

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 which 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 (available electronically here!).

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 ($\lambda = 0$ in (1.1)).

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

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

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.

4.2 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.

Here we will study a simple quantum mechanical example (that is: an example with a finite number of degrees of freedom)[28] 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 which we'll need in examples where perturbating is our only option.

Consider the following ('bare') action:

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

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 $\hbar = c = 1$, all dimensionful quantites are some power of a length. Let $-[X]$ denote the number of powers of length in the units of the quantity X; that is, if $X \sim (\text{length})^{\nu(X)}$ then we have $[X] = -\nu(X)$, a number. We have:

$$[t] = [\text{length}/c] = -1 \implies [dt] = -1.$$

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 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.

[28] I learned this example from 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:

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

The delta function is

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

So the Schrödinger equation says

$$\begin{aligned}\left(-\frac{1}{2}\nabla^2 - E\right)\psi(q) &= -V(q)\psi(q)\\ \int \text{đ}^2p e^{ip\cdot q}\left(\frac{p^2}{2} - E\right)\varphi(p) &= +g_0\delta^2(q)\psi(0)\\ &= +g_0\left(\int \text{đ}^2p e^{\mathbf{i}p\cdot q}\right)\psi(0) \qquad (4.3)\end{aligned}$$

which (integrating the both-hand side of (4.3) over q: $\int d^2q e^{-\mathbf{i}p\cdot q}$ ((4.3))) 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:

- $\psi(\vec{q}=0) = \int \text{đ}^2p\varphi(\vec{p}) = 0$. 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 undergrad 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 (rather: logarithmic dependence on the cutoff). It's the only kind 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 $|p| > \Lambda$, or maybe we could say that the potential is not really a delta function if we look more closely. The choice of narrative here shouldn't 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 $V = -g_0\delta$).

[End of Lecture 9]

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

$$\epsilon_B = \frac{\Lambda^2}{2} \frac{1}{e^{\frac{2\pi\hbar^2}{g_0}} - 1} \ . \tag{4.4}$$

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[29]. 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 which 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.*

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

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

[This d.t. phenomenon was maybe first seen in a perturbative field theory in S. Coleman, E. Weinberg, *Phys Rev* **D7** (1973) 1898. We'll come back to their example.]

There are more lessons in this example. Go back to (4.4):

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

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.

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 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:

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

[30] 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 (fixed points, limit cycles...)

Def: A fixed point $g_0^\star$ of a flow is a point 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.)

Where are the fixed points in our example?

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

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 this (high-energy) definition of the direction of flow, $g_0 = 0$ is an attractive 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 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.

[30] 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 UV, toward high energy. This is the high-energy particle physics perspective, where we learn more physics by going to higher energies. As we will see, there is a strong argument to be made for the other perspective, 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.

On the other hand, the d.t. 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 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 d.t. 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.

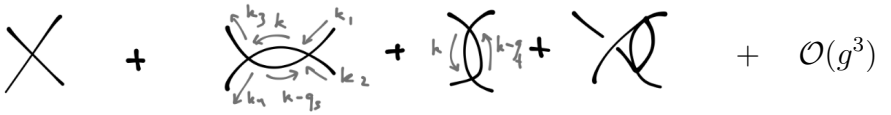
4.3 A simple example of perturbative renormalization in QFT

[Zee §III.1, Schwartz §15.4] Now let's consider an actual field theory but a simple one, namely the theory of a real scalar field in four dimensions, with

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - m^2\phi^2 - \frac{g}{4!}\phi^4. \tag{4.5}$$

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

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

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

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

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

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

Parametrizing ignorance. Recall our discovery of the scalar field at the beginning of the quarter by starting with a chain of springs, and looking at the long-wavelength (small-wavenumber) modes. 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 stops.

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 from §4.1: in that calculation, many different choices of regulators for the mode sum corresponded to different material properties of the conducting plates. The leading 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 ($-\pi/a \leq p_i < \pi/a$ for the cubic lattice). Instead, for simplicity, let's keep just impose a hard cutoff on the euclidean momentum $\sum_{i=0}^d p^2 \leq \Lambda^2$.

Parametrizing ignorance is another way to say 'doing 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 which 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, $C = \frac{1}{16\pi^2}$.

Observables can be predicted from other observables. Again, the boldface statement might sound like some content-free tweet from some 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'.[31] *This* we can relate to our theory letters:

$$-\mathbf{i}g_P = \mathbf{i}\mathcal{M}(s_0,t_0,u_0) = -\mathbf{i}g + \mathbf{i}Cg^2L_0 + \mathcal{O}(g^3) \tag{4.6}$$

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.)

Renormalization. Now here comes the big gestalt shift: Solve this equation (4.6) for the stupid letter g

$$\begin{aligned} -\mathbf{i}g &= -\mathbf{i}g_P - \mathbf{i}Cg^2L_0 + \mathcal{O}(g^3) \\ &= -\mathbf{i}g_P - \mathbf{i}Cg_P^2L_0 + \mathcal{O}(g_P^3). \end{aligned} \tag{4.7}$$

and eliminate g from the discussion:

$$\mathbf{i}\mathcal{M}(s,t,u) \;=\; -\mathbf{i}g + \mathbf{i}Cg^2L + \mathcal{O}(g^3)$$

[31] 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. I'll say more later, but if you are impatient see the beginning of Schwartz, chapter 18.

$$\begin{aligned} &\overset{(4.7)}{=} -\mathbf{i}g_P - \mathbf{i}Cg_P^2 L_0 + \mathbf{i}Cg_P^2 L + \mathcal{O}(g_P^3) \\ &= -\mathbf{i}g_P + \mathbf{i}Cg_P^2 \left(\log \frac{s_0}{s} + \log \frac{t_0}{t} + \log \frac{u_0}{u} \right) + \mathcal{O}(g_P^3). \end{aligned} \tag{4.8}$$

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 §4.2, it was eliminated by letting the coupling vary with it, $g = g(\Lambda)$, according to (4.7). 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 (4.5), with $m = 0$ for simplicity):

$$\mathcal{L} = -\frac{1}{2}\phi\Box\phi - \frac{g_P}{4!}\phi^4 - \frac{\delta_g}{4!}\phi^4 \tag{4.9}$$

but written in terms of the measured coupling g_P, and some as-yet-undetermined 'counterterm' δ_g. Then

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

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

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

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.

4.4 Radiative corrections to the Mott formula

Recall that by studying scattering of an electron from a heavy charged fermion (a muon is convenient) we reconstructed the cross section for scattering off a Coulomb potential (named after Mott). Our next goal is to figure out how this cross section is corrected by other QED processes.

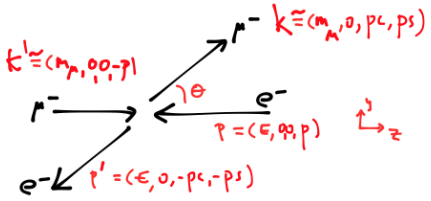
Recall that

$$\mathbf{i}\mathcal{M} = \quad = (-\mathbf{i}e\bar{u}(p')\gamma^\mu u(p)))_{\text{electrons}} \frac{-\mathbf{i}\left(\eta_{\mu\nu} - \frac{(1-\xi)q_\mu^t q_\nu^t}{q_t^2}\right)}{q_t^2} (-\mathbf{i}e\bar{u}(k)\gamma^\nu u(k'))_{\text{muons}} \quad (4.10)$$

with $q_t \equiv p - p' = k - k'$. After the spin sum,

$$\frac{1}{4}\sum_{s,s',r,r'} |\mathcal{M}|^2 = 4\frac{e^4}{t^2} \left(-p_\mu p'_\nu - p'_\mu p_\nu - \eta_{\mu\nu}(-p\cdot p' + m_e^2)\right) \cdot \left(-k_\mu k'_\nu - k'_\mu k_\nu - \eta_{\mu\nu}(-k\cdot k' + m_\mu^2)\right) \quad (4.11)$$

Consider the limit where the target μ particle is much heavier than the electron. '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, $c \equiv \cos\theta, s \equiv \sin\theta$.

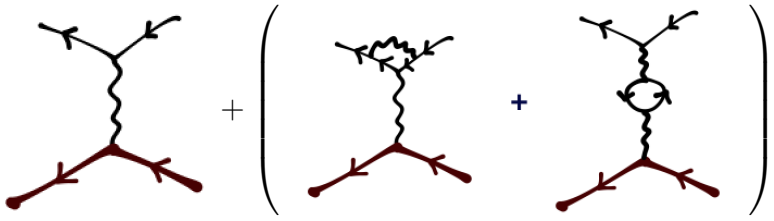


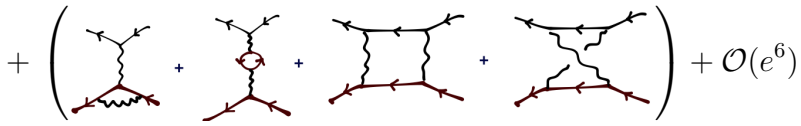
The answer we found after some boiling was:

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

If we take $\beta \ll 1$ in this formula we get the Rutherford formula.

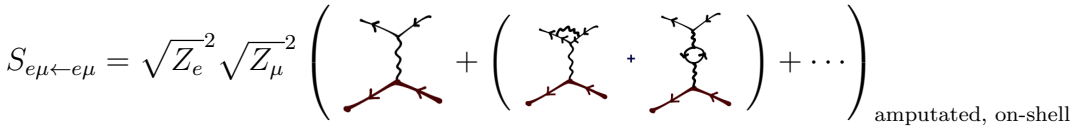
Radiative corrections. Now it's time to think about perturbative corrections to this cross section. Given that the leading-order calculation reproduced the classical physics of the Coulomb potential, you can think of what we are doing as effectively discovering (high-energy or short-distance) quantum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

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



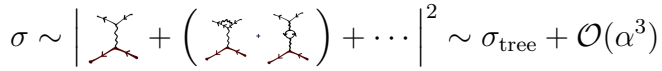
• What do the one-loop diagrams in the second line have in common? They have an internal muon line. Why does this matter? When the energy going through the line is much smaller than the muon mass, then the propagator is $\frac{i(\not{k}+m_\mu)}{k^2-m_\mu^2} \sim \frac{1}{m_\mu}$ and its relative contribution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



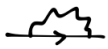
and in determining the locations of the poles whose residues are the S-matrix elements. We'll take care of these when we talk about the electron self-energy.

• 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

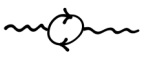


from the cross term between the tree and one-loop amplitudes.

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

• electron self-energy: 

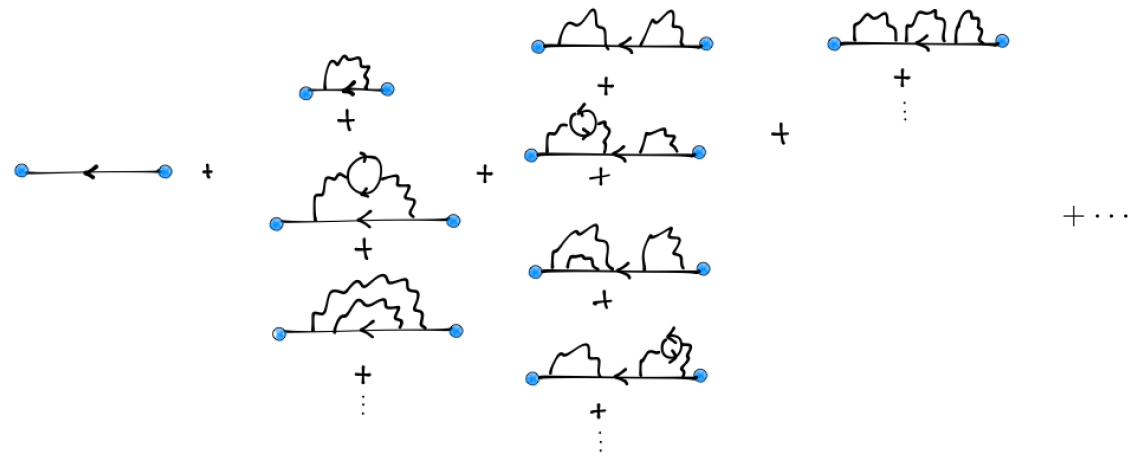
• vertex correction: 

• vacuum polarization (photon self-energy): 

[End of Lecture 10]

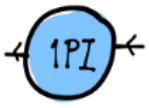
4.5 Electron self-energy in QED

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

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

(4.12)

As we did for the scalar field theory in §3, 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

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

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

The full two point function is then:

$$\begin{aligned}
\tilde{G}^{(2)}(p) &= \mathbf{i}S + \mathbf{i}S\,(-\mathbf{i}\Sigma(p))\,\mathbf{i}S + \mathbf{i}S\,(-\mathbf{i}\Sigma(p))\,\mathbf{i}S\,(-\mathbf{i}\Sigma(p))\,\mathbf{i}S + \cdots \\
&= \mathbf{i}S\,(1 + \Sigma S + \Sigma S \Sigma S + \cdots) = \mathbf{i}S\frac{1}{1 - \Sigma S} \\
&= \frac{\mathbf{i}}{\not{p} - m_0}\frac{1}{1 - \Sigma\frac{1}{\not{p} - m_0}} = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(p)}.
\end{aligned} \tag{4.13}$$

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 $\not{p}$, and therefore they commute;

we could do these manipulations in the eigenbasis of $\not{p}$. This fully corrected propagator has a pole at

$$\not{p} = m \equiv m_0 + \Sigma(m) \ . \tag{4.14}$$

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 (4.12), the leading (one-loop) contribution is

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

Notice that I am relying on the Ward identity to enforce the fact that only the traverse 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!

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

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

This allows us to combine the denominators into one:

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

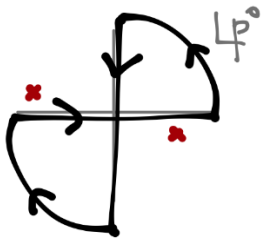
Step 2: Now we can complete the square

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

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 p^0 integral don't occur in the first and third octants of the complex p^0 plane. (And the integrand decays at large $|p^0|$.) 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

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

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

$$\begin{aligned}\Sigma_2(p) &= e^2 \int_0^1 dx \int \frac{\ell^2 d\ell^2}{2} \frac{(2\pi^2)}{(2\pi)^4} \frac{2(2m_0 - x\not{p})}{(\ell^2 + \Delta)^2} \\ &= \frac{e^2}{8\pi^2} \int_0^1 dx (2m_0 - x\not{p}) \mathcal{J} \end{aligned} \tag{4.15}$$

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

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

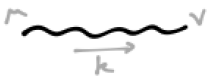
You knew this UV divergence was coming. To be more precise, let's add zero:

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

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



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

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 isn't manifest. This 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. You shouldn't take the regulated theory too seriously: for example, the wrong-sign propagator means wrong-sign kinetic terms for the PV fields. This 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. A nice thing about PV is that it is Lorentz invariant. A class of regulators which 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.

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

$$\begin{aligned}\mathcal{J} \rightsquigarrow\ & \mathcal{J}_{\Delta(\mu^2)} - \mathcal{J}_{\Delta(\Lambda^2)} \\ =\ & \left[\left(\ln\left(\ell^2+\Delta(\mu^2)\right)-1\right) - \left(\ln\left(\ell^2+\Delta(\Lambda^2)\right)-1\right)\right]\Big|_0^\infty \\ =\ & \ln\frac{\ell^2+\Delta(\mu^2)}{\ell^2+\Delta(\Lambda^2)}\Big|_0^\infty \\ =\ & \ln 1/1 - \ln\frac{\Delta(\mu^2)}{\Delta(\Lambda^2)} = \ln\frac{\Delta(\Lambda^2)}{\Delta(\mu^2)}.\end{aligned}$$

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

$$\Sigma_2(p)_{PV} = \frac{\alpha}{2\pi}\int_0^1 dx(2m_0 - x\not{p}) \ln \frac{x\Lambda^2}{x\mu^2 + (1-x)m_0^2 - x(1-x)p^2}. \tag{4.16}$$

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. We return to (4.14): the shift in the mass as a result of this one-loop self-energy is

$$\begin{aligned}
\delta m \equiv m - m_0 &= \Sigma_2(\not{p} = m) + \mathcal{O}(e^4) = \Sigma_2(\not{p} = m_0) + \mathcal{O}(e^4) \\
&= \frac{\alpha}{2\pi}\int_0^1 dx\ (2-x)m_0 \ln \frac{x\Lambda^2}{\underbrace{x\mu^2 + (1-x)m_0^2 + x(1-x)m_0^2}_{= x\mu^2 + (1-x^2)m_0^2 \equiv f(x,m_0,\mu)}} \\
&= \frac{\alpha}{2\pi}\int_0^1 dx\ (2-x)m_0 \left(\underbrace{\ln \frac{\Lambda^2}{m_0^2}}_{\text{divergent}} + \underbrace{\ln \frac{xm_0^2}{f(x,m_0,\mu)}}_{\text{relatively small}} \right) \\
&\approx \frac{\alpha}{2\pi}\left(2 - \frac{1}{2}\right) m_0 \ln \frac{\Lambda^2}{m_0^2} = \frac{3\alpha}{4\pi} m_0 \ln \frac{\Lambda^2}{m_0^2}.
\end{aligned} \tag{4.17}$$

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.

Mass renormalization. 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 magnetic field and measure the radius of curvature of its orbit and set $m_e v^2/r = evB/c$), so

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

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 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

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{\not{p} - m_0 - \Sigma(p)} \overset{p\sim m}{\simeq} \frac{\mathbf{i}Z}{\not{p} - m} + \text{regular terms.} \tag{4.18}$$

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

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

So then (4.18) becomes

$$\tilde{G}^{(2)}(p) \overset{p\sim m}{\sim} \frac{\mathbf{i}}{\not{p} - m - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}(\not{p} - m)} = \frac{\mathbf{i}}{(\not{p} - m)\left(1 - \frac{\partial\Sigma}{\partial\not{p}}|_{m_0}\right)} \tag{4.19}$$

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

$$\begin{aligned}\delta Z = \frac{\partial\Sigma_2}{\partial\not{p}}|_{m_0} &\overset{(4.16)}{=} \frac{\alpha}{2\pi}\int_0^1 dx \left(-x\ln\frac{x\Lambda^2}{f(x,m_0,\mu)} + (2m_0 - xm_0)\frac{-2x(1-x)}{f(x,m_0,\mu)}\right) \\ &= -\frac{\alpha}{4\pi}\left(\ln\frac{\Lambda^2}{m_0^2} + \text{finite}\right).\end{aligned} \tag{4.20}$$

Here $f = f(x, m_0, \mu)$ is the same quantity defined in the second line of (4.17). We'll see below 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.

4.6 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.

4.6.1 Self-energy in ϕ^4 theory

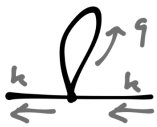
[Zee §III.3] Let's return to the ϕ^4 theory in $D = 3 + 1$ for a moment. The $\mathcal{M}_{\phi\phi\leftarrow\phi\phi}$ amplitude is not the only place where the cutoff appears.

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:

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

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

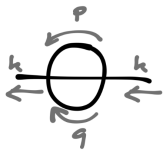
$$\delta\Sigma_1(k) = \quad = -\mathbf{i}g\int^\Lambda \mathrm{đ}^4 q \frac{\mathbf{i}}{q^2 - m^2 + \mathbf{i}\epsilon} = \delta\Sigma_1(k=0) \sim g\Lambda^2$$



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 (4.19)), 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:

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



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.

Putting this together, the inverse propagator is

$$D^{-1}(k) = D_0^{-1}(k) - \Sigma(k) = k^2 - m^2 - \underbrace{(\delta\Sigma_1(0) + A_0)}_{\equiv a\sim\Lambda^2} - k^2 A_1 - k^4 A_2 + \cdots$$

The $\cdots$ here includes both higher orders in g ($\mathcal{O}(g^3)$) and higher powers of k^2, *i.e.* higher derivative terms. If instead the physical pole were at a nonzero value of the mass, we should Taylor expand about $k^2=m_P^2$ instead:

$$D^{-1}(k) = D_0^{-1}(k) - \Sigma(k) = k^2 - m_0^2 - \underbrace{(\delta\Sigma_1(0) + A_0)}_{\equiv a\sim\Lambda^2} - (k^2-m_P^2)A_1 - (k^2-m_P^2)^2 A_2 + \cdots$$

where now $A_n \equiv \frac{1}{n!}\left(\frac{\partial}{\partial k^2}\right)^n \Sigma_2(k^2)|_{k^2=m_P^2}$.

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

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

Some points to notice: • $\delta Z = A_1$.

• 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 was only log divergent. This UV sensitivity of scalar fields is ubiquitous[32] (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

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

the height of a wave in the 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 $\phi_R \equiv Z^{-1/2}\phi$. 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

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

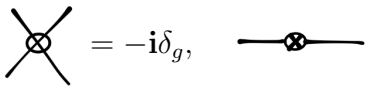
with

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

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

(4.21)

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 (4.21) 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)?

4.6.2 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 $\mathcal{A}$ (order by order in the loop expansion). Define the 'superficial' degree of divergence of $\mathcal{A}$ to be $D_\mathcal{A}$ if

$\mathcal{A} \sim \Lambda^{D_\mathcal{A}}$ (in the limit that $\Lambda \ll$ all other scales – this is an asymptotic statement). 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 $D_\mathcal{A} = 4 - B_E$. [End of Lecture 11]

Why didn't it 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,

$$L = B_I - (V - 1). \tag{4.22}$$

We already discussed a version of this statement around (2.6). Math proof[33]: 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 which we must integrate: $\prod_{\alpha=1}^{B_I} \int \mathrm{đ}^D q_\alpha$. We then stick a $\delta^{(D)}(\sum q)$ for each vertex, but one of these gives the overall momentum conservation $\delta^{(D)}(k_T)$, so we have $V - 1$ fewer momentum integrals. For the example above, (4.22) 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 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

$$B_I = 2V - B_E/2. \tag{4.23}$$

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

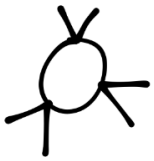
Now we count powers of momenta:

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

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:

$$\begin{aligned} D_{\mathcal{A}} = [\mathcal{A}] &= DL - 2B_I \\ &\overset{(4.22)}{=} B_I(D-2) - D(V-1) \\ &\overset{(4.23)}{=} D + \frac{2-D}{2} B_E + V(D-4). \end{aligned}$$

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 $B_E = 6$:



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

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. [34]

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:

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

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

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

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. (I bet it is possible to come up with more pathological examples.) 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.

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 $2V_y = 2F_E + F_I$ 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

$$D_{\mathcal{A}} = D + (D-4)\left(V_g + \frac{1}{2}V_y\right) + B_E\left(\frac{2-D}{2}\right) + F_E\left(\frac{1-D}{2}\right). \tag{4.24}$$

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 Yukawa interactions) has

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

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.

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).

4.6.3 Naive scale invariance in 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 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:

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

Notice that the $D = 1$ case agrees with our quantum mechanics counting from §4.2. 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 $ds^2 = g_{ij}d\phi^i d\phi^j$.

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

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

– 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

$$[g] = D \cdot \frac{p_D - p}{p_D} \begin{cases} < 0 \text{ when } p > p_D, & \text{non-renormalizable or irrelevant} \\ = 0 \text{ when } p = p_D, & \text{renormalizable or marginal} \\ > 0 \text{ when } p < p_D, & \text{super-renormalizable or relevant.} \end{cases} \tag{4.25}$$

Consider the 'non-renormalizable' case. Suppose we calculate in QFT some quantity f 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

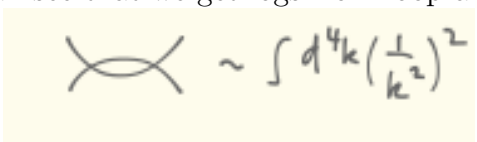
$$[f] = n[g] + [c_n] \implies [c_n] = [f] - n[g]$$

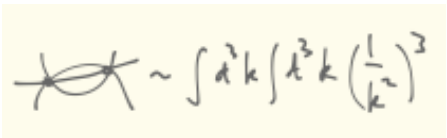
So if $[g] < 0$, c_n must have more and more powers of some mass (inverse length) as n increases. What dimensionful quantity makes up the difference? Sometimes it is masses or external momenta. But generically, it gets made up by UV divergences (if everything is infinite, dimensional analysis can fail, 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: $c_n = \tilde{c}_n \left(\Lambda_{UV}\right)^{-n[g]}$, where $\tilde{c}_n$ is dimensionless, and $n[g] < 0$ – it's higher and higher powers of the cutoff.

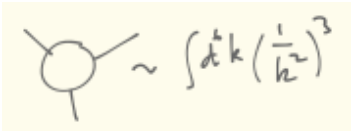
Consider the renormalizable (classically scale invariant) case: $[c_n] = [f]$, since $[g] = 0$. But in fact, what you'll get is something like

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

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 (4.25) 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.

4.7 Vertex correction in QED

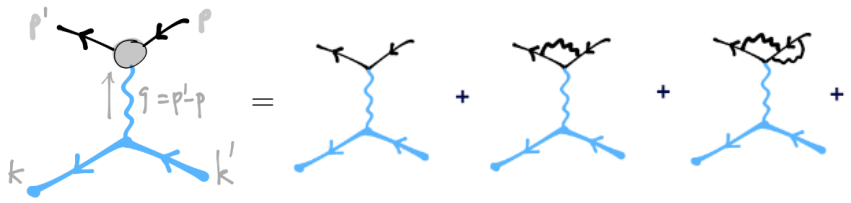
[Peskin chapter 6, Schwartz chapter 17, Zee chapter III.6] Back to work on QED. The vertex correction  has some great physics payoffs:

- 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![35] 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 you're overstepping your bounds (by taking too seriously your Lagrangian parameters or your knowledge of short distances), IR divergences mean you are *asking the wrong question*.

[35] 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.

To get started, consider the following class of diagrams.



$$\equiv \mathbf{i}\mathcal{M} \qquad = \mathbf{i}e^2\left(\bar{u}(p')\Gamma^\mu(p,p')u(p)\right)\frac{1}{q^2}\bar{u}(K')\gamma_\mu u(K) \tag{4.26}$$

The shaded blob is the *vertex function* Γ. The role of the light blue factors is just to make and propagate the photon which 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 $q^2 = 2m^2 - 2p'\cdot p$.

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, p', γ^μ 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

$$\Gamma^\mu(p,p') = A\gamma^\mu + B(p+p')^\mu + C(p-p')^\mu \tag{4.27}$$

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:

$$0 \overset{\text{Ward}}{=} q_\mu\bar{u}(p')\Gamma^\mu u(p) \overset{(4.27)}{=} \bar{u}(p')\left(A\underbrace{\not{q}}_{=\not{p}'-\not{p}\overset{\bar{u}(p')\ldots u(p)}{=}m-m=0} + B\underbrace{(p+p')\cdot(p-p')}_{=m^2-m^2=0} + Cq^2\right)u(p)$$

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[36]. The Gordon identity

[36] 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 $m_1 = m_2$.

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* $F_{1,2}$:

$$\Gamma^\mu(p,p') = \gamma^\mu F_1(q^2) + \frac{\mathbf{i}\sigma^{\mu\nu}q_\nu}{2m}F_2(q^2). \tag{4.28}$$

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 fact of gauge invariance.

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 (4.28) 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 $\bar{u}(p')\Gamma^i u(p)A_i(q) = -\vec{\mu}\cdot\vec{B}(q) + ...$, (similarly to the homework problem with the γ^5 interaction) 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 $2F_2(q^2=0)$.

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:

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

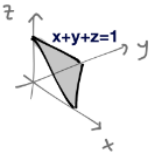
with $k' \equiv k+q$.

(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[37]:

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



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

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

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

$$\Delta = -xyq^2 + (1-z)^2m^2 + zm_\gamma^2.$$

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

[37] 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

$$\frac{1}{A} = \int_0^\infty ds\ e^{-sA}. \tag{4.29}$$

Applying this identity to each factor gives

$$\frac{1}{A_1A_2\cdots A_n} = \int_0^\infty ds_1 \cdots \int_0^\infty ds_n\ e^{-\sum_{i=1}^n s_iA_i}.$$

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

$$\frac{1}{A_1A_2\cdots A_n} = \int_0^\infty d\tau\tau^{n-1}\prod_{i=1}^n\int_0^1 dx_i\delta\left(\sum_{i=1}^n x_i - 1\right)\ e^{-\tau\sum_i x_iA_i}.$$

Now do the integral over τ, using $\int_0^\infty d\tau\tau^{n-1}e^{-\tau X} = \frac{(n-1)!}{X^n}$ (differentiate (4.29) wrt A), to arrive at

$$\frac{1}{A_1A_2\cdots A_n} = \prod_{i=1}^n\int_0^1 dx_i\delta\left(\sum_{i=1}^n x_i - 1\right)\frac{(n-1)!}{(\sum_i x_iA_i)^n}.$$

integrals we need are just

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

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

$$\int \frac{\mathrm{đ}^D \ell}{(\ell^2 - \Delta)^m} = (-1)^m \frac{\mathbf{i}}{(4\pi)^{D/2}} \frac{\Gamma\left(m - \frac{D}{2}\right)}{\Gamma(m)} \left(\frac{1}{\Delta}\right)^{m - \frac{D}{2}}. \tag{4.30}$$

$$\int \frac{\mathrm{đ}^D \ell \; \ell^2}{(\ell^2 - \Delta)^m} = (-1)^{m-1} \frac{D}{2} \frac{\mathbf{i}}{(4\pi)^{D/2}} \frac{\Gamma\left(m - \frac{D}{2} - 1\right)}{\Gamma(m)} \left(\frac{1}{\Delta}\right)^{m - \frac{D}{2} - 1}. \tag{4.31}$$

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

(0) But for now let's persist in using the Pauli Villars regulator. (I call this step (0) instead of (3) because it should have been there all along.) Here this means we subtract from the amplitude the same quantity with m_γ 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

$$\begin{aligned} N^\mu &= \bar{u}(p')\gamma^\nu \left(\not{k} + \not{q} + m_e\right) \gamma^\mu \left(\not{k} + m_e\right) \gamma_\nu u(p) \\ &= -2 \left(\mathcal{A} \bar{u}(p')\gamma^\mu u(p) + \mathcal{B} \bar{u}(p') \sigma^{\mu\nu} q_\nu u(p) + \mathcal{C} \bar{u}(p') q^\mu u(p)\right) \end{aligned} \tag{4.32}$$

where

$$\begin{aligned} \mathcal{A} &= -\frac{1}{2}\ell^2 + (1-x)(1-y)q^2 + (1 - 4z + z^2)m^2 \\ \mathcal{B} &= \mathbf{i} m z (1 - z) \\ \mathcal{C} &= m(z-2)(y - x) \; . \end{aligned} \tag{4.33}$$

The blood of many men was spilled 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. [End of Lecture 12]

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:

4.7.1 Anomalous magnetic moment

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

$$\begin{aligned} F_2(q^2) &= \frac{2m}{e} \cdot (\text{the term with } \mathcal{B}\ ) \\ &= \frac{2m}{e} 4e^3\,(\mathbf{i}m) \int dxdydz\delta(x+y+z-1) z(1-z) \underbrace{\int \frac{\text{đ}^4\ell}{(\ell^2-\Delta)^3}}_{=\frac{-\mathbf{i}}{32\pi^2\Delta}} \\ &= \frac{\alpha}{\pi} m^2 \int dxdydz\delta(x+y+z-1) \frac{z(1-z)}{(1-z)^2m^2 - xyq^2}. \end{aligned} \tag{4.34}$$

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: $g = 2.00232$.

4.7.2 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 $m_\gamma = 0$)

$$\begin{aligned} \int dxdydz\delta(x+y+z-1) \frac{m^2(1-4z+z^2)}{\Delta} &= m^2 \int_0^1 dz \int_0^{1-z} dy \frac{-2+2(1-z)+(1-z)^2}{(1-z)^2m^2} \\ &= \int_0^1 dz \frac{-2}{(1-z)} + \text{finite}, \end{aligned} \tag{4.35}$$

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

$$\Gamma^\mu = \gamma^\mu \left(1 - \frac{\alpha}{2\pi} f_{IR}(q^2) \ln\left(\frac{-q^2}{{m_\gamma}^2}\right)\right) + \text{ stuff which is finite as } m_\gamma \to 0. \tag{4.36}$$

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 (4.36) 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,

$$F_1(q^2) = 1 + f(q^2) + \delta_1 + \mathcal{O}(\alpha^2)$$

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_1 here is a counterterm 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_1 so that

$$1 = F_1(0) \implies \delta_1 = -f(0) \overset{m_e/q \to 0}{\to} -\frac{e^2}{8\pi^2} \frac{1}{2} \ln \frac{\Lambda^2}{m_\gamma^2}.$$

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

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

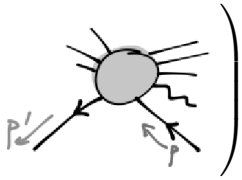
$$F_1(q^2)|_{m_e=0} = 1 - \frac{e^2}{16\pi^2} \left(\ln^2 \frac{-q^2}{m_\gamma^2} + 3 \ln \frac{-q^2}{m_\gamma^2}\right) + \text{finite}.$$

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 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 for those particles to do 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.

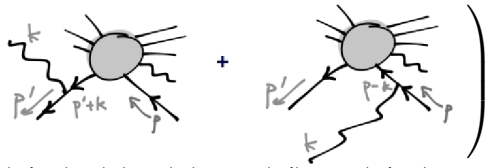
Being more inclusive, then, we cannot distinguish amplitudes of the form

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



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 photon is real (k^2=0) and it is *soft*, in the sense that $k^0 < E_c$, the detector cutoff. So we can approximate the numerator of the second term as

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

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 and $k \ll p$. Therefore

$$\mathcal{M}\left(e\mu + \text{one soft } \gamma \leftarrow e\mu\right) = e\bar{u}(p')\mathcal{M}_0(p',p)u(p)\left(\frac{p'\cdot\epsilon^\star}{p'\cdot k + \mathbf{i}\epsilon} - \frac{p\cdot\epsilon^\star}{p\cdot k - \mathbf{i}\epsilon}\right) \qquad (4.37)$$

This is bremsstrahlung. Before we continue this calculation to find the inclusive amplitude which a real detector actually measures, let's pause to relate the previous expression to some physics we know. 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 [38]

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

This is a good approximation to the motion a free particle which 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 $f_{IR}(q^2)=\frac{\alpha}{\pi}\ln\left(\frac{-q^2}{m^2}\right)$ (which entered our lives in (4.36)) 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 $\frac{1}{2}\int d^3x\,(E^2+B^2)=\int \text{đ}^3k\hbar\omega_k n_k$. (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.

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

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

[38] 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.

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[39]

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

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 regard as an ugly detail has saved the day.

I didn't show explicitly that the coefficient of the log is the same function $f_{IR}(q^2)$. In fact this function is $f_{IR}(q^2) = \frac{1}{2}\log(-q^2/m^2)$, so the product $f_{IR} \ln q^2 \sim \ln^2 q^2$ is the Sudakov double logarithm. A benefit of the calculation which 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

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

[39] 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 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.

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] Consider a diagram with n soft external photons, summed over ways of distributing them on an initial and final electron line:

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

Here the difference in each factor is just as in (4.37), one term from initial and one from final-state emission; expanding the product gives the sum over $n_f = 1 - n_i$, the number coming from the final-state line. 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 $k_\alpha = -k_\beta \equiv k$ and tying them together with a propagator and summing over k:

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

The factor of $\frac{1}{2}$ accounts for the symmetry under exchange of $\alpha \leftrightarrow \beta$. For the case of $n = 2$, this is the whole story, and this is

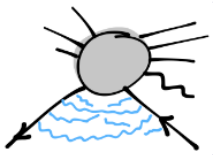
$$\bar{u}\mathbf{i}\mathcal{M}_0 u \cdot \mathbf{X} = \left(\quad \right) \cdot \left(\quad \right)_{\text{soft part}}$$

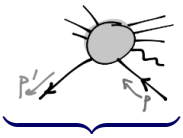
(where here 'soft part' means the part which is singular in m_γ) from which we conclude that

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

Taking the most IR-divergent bit with m virtual soft photons (order α^m) for each m gives

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





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

$$\begin{aligned}
d\sigma_{1\gamma} &= \int d\Pi \underbrace{\sum_{\text{pols}} \epsilon^\mu \epsilon^{\star\nu}}_{=-\eta^{\mu\nu}} \mathcal{M}_\mu \mathcal{M}_\nu^\star \\
&= |\bar{u}(p')\mathcal{M}_0 u(p)|^2 \int \frac{\mathrm{d}^3 k}{2E_k} (-\eta_{\mu\nu})\, e^2 \left(\frac{p'}{p'\cdot k} - \frac{p}{p\cdot k} \right)^\mu \left(\frac{p'}{-p'\cdot k} - \frac{p}{-p\cdot k} \right)^\nu \\
&\equiv d\sigma_0 \mathbf{Y}, \\
\mathbf{Y} &= \frac{\alpha}{\pi} f_{IR}(q^2) \ln\left(\frac{E_c^2}{m_\gamma^2} \right).
\end{aligned}$$

(The integral is done in Peskin, page 201.) 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.

[End of Lecture 13]

4.7.3 Some magic from gauge invariance of QED

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

$$Z_2 = 1 + \frac{\partial \Sigma}{\partial \not{p}}|_{\not{p}=m_0} + \mathcal{O}(e^4) = 1 - \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{m^2} + \text{finite} + \mathcal{O}(\alpha^2).$$

We care about this because there is a factor of Z_2 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

$$\begin{aligned} S_{e\mu \leftarrow e\mu} &= \left(\sqrt{Z_2(e)}\right)^2 \left(\quad + \left(\quad \cdot \quad \right) + \cdots \right) \\ &= \left(1 - \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{m^2} + \cdots\right) e^2 \bar{u}(p') \left(\gamma^\mu \left(1 + \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{m^2} + \cdots\right) + \alpha \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} \right) u(p) \end{aligned}$$

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_1 for the vertex. But how many counterterms do we get here? Is there a point of view which 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.

4.8 Vacuum polarization

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

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

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:

$$\mathcal{L} = \bar{\psi}\left(\partial\!\!\!/ + \mathbf{i}A\!\!\!/ - m\right)\psi - \frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu}.$$

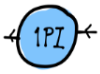
With this normalization, instead of measuring the coupling between electrons and photons, the coupling constant e measures the difficulty a photon has propagating through space:

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

None of the physics is different, since each internal photon line still has two ends on a $\bar{\psi}A\!\!\!/\psi$ 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 $\bar{\psi}\partial\!\!\!/\psi$ and $\bar{\psi}A\!\!\!/\psi$ terms[40]. Therefore, any counterterm we need for the $\bar{\psi}\partial\!\!\!/\psi$ term (which comes from the electron self-energy correction and is traditionally called δZ_2) must be the *same* as the counterterm for the $\bar{\psi}A\!\!\!/\psi$ 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.

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

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

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

(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

[40]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 $\tilde{A}_\mu \to \tilde{A}_\mu - \partial_\mu\lambda(x)/e$.

photon propagator is $\frac{-\mathbf{i}\eta_{\mu\nu}}{q^2}$.) 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

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

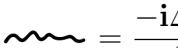
Let $A \equiv \Pi q^2$ so that

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

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

$$\Delta_{T\rho}^{\mu} \Delta_{T\ \nu}^{\rho} = \Delta_{T\nu}^{\mu} \tag{4.38}$$

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



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

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

$$\begin{aligned}
\tilde{G}^{(2)}(q) &= \\
&\overset{(4.38)}{=} \frac{-\mathbf{i}\Delta_T}{q^2}\left(1 + \mathbf{i}\Pi q^2 \Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right) + \mathbf{i}\Pi q^2 \Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right)\mathbf{i}\Pi q^2 \Delta_T\left(\frac{-\mathbf{i}\Delta_T}{q^2}\right) + \cdots\right) \\
&\overset{\Delta_T^2=\Delta_T}{=} \frac{-\mathbf{i}\Delta_T}{q^2}\left(1 + \Pi\Delta_T + \Pi^2\Delta_T + \cdots\right) = \frac{-\mathbf{i}\Delta_T}{q^2}\frac{1}{1-\Pi(q^2)}.
\end{aligned} \tag{4.39}$$

Does the photon get a mass? 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 $q^2 = 0$. If $\Pi(q^2)$ is regular at $q^2 = 0$, then the photon remains massless. 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?

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

But here is an identity:

$$\frac{1}{\not p + \not q - m}\not q \frac{1}{\not p - m} = \frac{1}{\not p - m} - \frac{1}{\not p + \not q - m}. \tag{4.40}$$

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

Why do I say 'if'? If the integral depends on the UV limit, this shift is not innocuous. So we have to address the cutoff dependence.

In addition to the (lack of) mass renormalization, we've figured out that the electromagnetic field strength renormalization is

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

We need Z_γ for example for the S-matrix for processes with external photons, like Compton scattering.

Claim: If we do it right[41], the cutoff dependence looks like[42]:

$$\Pi_2(q^2) = \frac{\alpha_0}{4\pi}\left(-\frac{2}{3}\ln\Lambda^2 + \underbrace{2D(q^2)}_{\text{finite}}\right)$$

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

$$\frac{e_0^2\Delta_T}{q^2} \rightsquigarrow \frac{Z_3 e_0^2 \Delta_T}{q^2},$$

and $Z_3 = \frac{1}{1-\Pi(0)}$ 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

[41] What I mean here is: if we do it in a way which 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 page 202-204, does. Alternatively, we could use dimensional regularization everywhere.

[42] The factor in front of the $\ln\Lambda$ can be made to look like it does in other textbooks using $\alpha = \frac{e^2}{4\pi}$, so that

$$\frac{\alpha_0}{4\pi}\left(\frac{2}{3}\ln\Lambda^2\right) = \frac{e_0^2}{12\pi^2}\ln\Lambda.$$

in terms of $e \equiv \sqrt{Z_3} e_0$ where $\frac{e^2}{4\pi} = \frac{1}{137}$ is the measured fine structure constant (at low energy). To this order, then, we write

$$e_0^2 = e^2 \left(1 + \frac{\alpha_0}{4\pi} \frac{2}{3} \ln \Lambda^2\right) + \mathcal{O}(\alpha^2). \tag{4.41}$$

$$m_0 = m + \mathcal{O}(\alpha_0) = m + \mathcal{O}(\alpha). \tag{4.42}$$

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

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

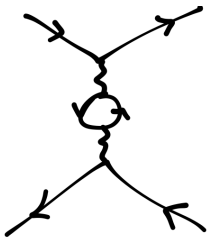
– in QED, the quantum fluctuations *reduce* the charge, as you might expect from the interpretation of this phenomenon as dielectric screening by virtual e^+e^- pairs.

In the example case of $e\mu \leftarrow e\mu$ scattering, the full one-loop UV cutoff dependence then looks like

$$\begin{aligned} S_{e\mu \leftarrow e\mu} &= \sqrt{Z_e^2} \left(1 - \frac{\alpha_0}{4\pi} \ln \Lambda^2 + \frac{\alpha_0}{2\pi} A(m_0)\right) e_0^2 \\ &\quad L_\mu \bar{u}(p') \left[\gamma^\mu \left(1 + \frac{\alpha_0}{4\pi} \ln \Lambda^2 + \frac{\alpha_0}{2\pi} (B + D) + \frac{\alpha_0}{4\pi} \left(-\frac{2}{3} \ln \Lambda^2\right)\right) + \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} \frac{\alpha_0}{2\pi} C(q^2, m_0)\right] u(p) \\ &= e^2 L_\mu \bar{u}(p') \left[\gamma^\mu \left(1 + \frac{\alpha}{2\pi}(A + B + D)\right) + \frac{\mathbf{i}\sigma^{\mu\nu} q_\nu}{2m} \frac{\alpha}{2\pi} C\right] u(p) + \mathcal{O}(\alpha^2) \end{aligned} \tag{4.43}$$

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 (4.41) and (4.42) 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

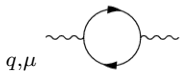


$$\propto e_0 \frac{1}{1 - \Pi(0)} e_0 = e^2.$$

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.

4.8.1 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 is

$$N^{\mu\nu} = 2\ell^\mu\ell^\nu - \eta^{\mu\nu}\ell^2 - 2x(1-x)q^\mu q^\nu + \eta^{\mu\nu}\left(m^2 + x(1-x)q^2\right) + \text{terms linear in } \ell^\mu \quad .$$

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

$$\mathbf{i}\Pi_2^{\mu\nu} \propto e^2 \int^\Lambda d^4\ell_E \frac{\ell_E^2 \eta^{\mu\nu}}{\left(\ell_E^2-\Delta\right)^2} + ... \propto e^2\Lambda^2\eta^{\mu\nu}$$

This is *NOT* of the form $\Pi^{\mu\nu} = \Delta_T^{\mu\nu}\Pi(p^2)$; rather it produces a correction to the photon mass proportional to the cutoff. What happened? Our cutoff was not gauge invariant. Oops.[43]

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

[43] 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

$$-\sum_a c_a \int \text{đ}^D k \text{tr}\left((\mathbf{i}e\gamma^\mu)\frac{\mathbf{i}}{\not q+\not k-m_a}(\mathbf{i}e\gamma^\nu)\frac{\mathbf{i}}{\not q-m_a}\right) \sim \int^\Lambda \text{đ}^4 k\left(\frac{\sum_a c_a}{k^2}+\frac{\sum_a c_a m_a^2}{k^4}+\cdots\right).$$

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, the answer is

$$\Pi_2(q^2) = \frac{1}{2\pi^2}\int dx x(1-x)\ln\frac{M^2}{m^2-x(1-x)q^2}$$

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.

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 which 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 which 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 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 $\int \text{đ}^D p f(p+q) = \int \text{đ}^D p f(p)$ [44]

[44] 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 (4.40).

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:

$$\left(\frac{\pi}{a}\right)^{D/2} = \int d^D x e^{-a\vec{x}^2} = \Omega_{D-1} \int_0^\infty x^{D-1} dx e^{-ax^2} = \frac{1}{2} a^{-\frac{D}{2}} \Gamma\left(\frac{D}{2}\right) \Omega_{D-1} \,. \tag{4.44}$$

This defines Ω_{D-1} for general D.

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:

$$\begin{aligned} \Pi_2(q^2) &\overset{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2-D/2)}{\Delta^{2-D/2}} \bar{\mu}^\epsilon \\ &\overset{D\to 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left(\frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right)\right) \end{aligned} \tag{4.45}$$

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 (9.9), 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 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 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 $\overline{\rm MS}$ scheme, in which Π is defined by the rule that we *subtract the* $1/\epsilon$ *pole*. This means that the counterterm is

$$\delta_{F^2}^{(\overline{\rm MS})} = -\frac{e^2}{2\pi^2}\frac{2}{\epsilon}\underbrace{\int_0^1 dx x(1-x)}_{=1/6}.$$

(Confession: I don't know how to state this in terms of a simple renormalization condition on Π_2. Also: the bar in $\overline{\rm MS}$ refers to the (not so important) distinction between $\bar{\mu}$ and μ.) The resulting vacuum polarization function is

$$\delta\Pi_2^{(\overline{\rm MS})}(q^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\log\left(\frac{m^2 - x(1-x)q^2}{\mu^2}\right).$$

[End of Lecture 14]

4.8.2 Physics from vacuum polarization

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

$$\tilde{V}(q) = \frac{1}{q^2}\frac{e^2}{1-\Pi(q^2)} \equiv \frac{e^2_{\rm eff}(q)}{q^2}. \tag{4.46}$$

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

$$\delta\Pi_2(q) = \frac{e^2}{2\pi^2}\int dx x(1-x)\ln\left(1 - \frac{q^2}{m^2}x(1-x))\right) \overset{q\ll m}{\simeq} \frac{e^2}{2\pi^2}\int dx x(1-x)\left(-\frac{q^2}{m^2}x(1-x))\right) = -\frac{q^2}{60\pi^2 m^2}$$

which means

$$\tilde{V}(q) \overset{q\ll m}{\simeq} \frac{e^2}{q^2} + \frac{e^2}{q^2}\left(-\frac{q^2}{30m^2}\right) + \cdots$$

and hence

$$V(r) = -\frac{e^2}{4\pi r^2} - \frac{e^4}{60\pi^2 m^2}\delta(r) + \cdots \equiv V + \Delta V.$$

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 (the Lamb shift).

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[45]

$$\Pi_2(q^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\ln\left(1 - \frac{q^2}{m^2}x(1-x)\right) \simeq \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\ln\left(-\frac{q^2}{m^2}\right) = \frac{e^2}{12\pi^2}\ln\left(-\frac{q^2}{m^2}\right).$$

Therefore, the effective charge in (4.46) at high momentum exchange is

$$e^2_{\text{eff}}(q^2) \overset{q^2 \gg m_e^2}{\simeq} \frac{e^2}{1 - \frac{e^2}{12\pi^2}\ln\left(-\frac{q^2}{m^2}\right)}. \tag{4.47}$$

(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*. 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 $q \ll m_e$.

[45] 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.

5 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 the 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'.

5.1 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. 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 [46]

$$\partial_\mu \mathcal{O}(x) = \mathbf{i}[\mathbf{P}_\mu, \mathcal{O}(x)] .$$

[46] 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:

$$-\mathbf{i}\mathcal{D}(x) = \theta(t)\,\langle 0|\, e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)e^{-\mathbf{i}\mathbf{P}x}\mathcal{O}^\dagger(0)\,|0\rangle + \theta(-t)\,\langle 0|\,\mathcal{O}^\dagger(0)e^{\mathbf{i}\mathbf{P}x}\mathcal{O}(0)e^{-\mathbf{i}\mathbf{P}x}\,|0\rangle\,. \quad (5.1)$$

Now we need a resolution of the identity operator on the *entire* QFT $\mathcal{H}$:

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

This innocent-looking n summation variable 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 (5.1); sticking the $\mathbb{1}$ 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 our 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):

$$-\mathbf{i}\mathcal{D}(q) = \int d^D x e^{\mathbf{i}qx}\mathbf{i}\mathcal{D}(x) = \mathbf{i}(2\pi)^{D-1}\sum_n \|\,\mathcal{O}_{0n}\,\|^2\left(\frac{\delta^{(D-1)}(\vec{q}-\vec{p}_n)}{q^0 - p_n^0 + \mathbf{i}\epsilon} - \frac{\delta^{(D-1)}(\vec{q}+\vec{p}_n)}{q^0 + p_n^0 - \mathbf{i}\epsilon}\right)\,. \quad (5.2)$$

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: $\left|\vec{k}\right\rangle$ (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

$$\left\langle \vec{k}\right| \mathcal{O}(0)\left|0\right\rangle = \frac{Z^{\frac{1}{2}}}{\sqrt{(2\pi)^{D-1}\, 2\omega_{\vec{k}}}} \tag{5.3}$$

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. [47] What is Z? From (5.3) 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:

$$\mathbb{1} = \mathbb{1}_1 + \cdots .$$

[47] 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 (5.3), 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 (5.2) which 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. (This result was promised when we discussed LSZ.)[48]

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

$$\mathrm{Im}\,\frac{1}{Q\mp\mathbf{i}\epsilon} = \pm\pi\delta(Q), \qquad (\text{for } Q \text{ real}). \tag{5.4}$$

we have

$$\mathrm{Im}\,\mathcal{D}(q) = \pi\,(2\pi)^{D-1}\sum_n \|\,\mathcal{O}_{0n}\,\|^2\left(\delta^D(q-p_n)+\delta^D(q+p_n)\right).$$

More explicitly (for real operators):

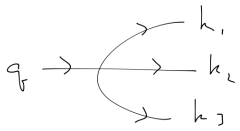
$$\mathrm{Im}\,\mathbf{i}\int d^Dx\,e^{\mathbf{i}qx}\,\langle 0|\,\mathcal{T}\mathcal{O}(x)\mathcal{O}(0)\,|0\rangle = \pi\,(2\pi)^{D-1}\sum_n \|\,\mathcal{O}_{0n}\,\|^2\left(\delta^D(q-p_n)+\underbrace{\delta^D(q+p_n)}_{=0\text{ for }q^0>0\text{ since }p_n^0>0}\right).$$

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, the contribution of a 1-particle state to the spectral density is:

$$\mathrm{Im}\,\mathcal{D}(q) = \ldots + \pi Z\delta(q^2-m^2).$$

[48] 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.

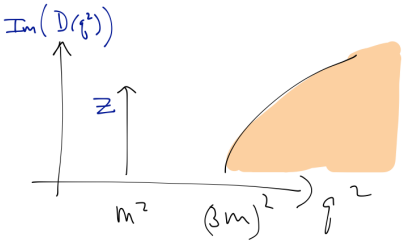
This quantity $\mathrm{Im}\,\mathcal{D}(q)$ (the spectral density of $\mathcal{O}$) is positive because it is the *number* of states (with D-momentum in an infinitesimal neighborhood of q), weighted by the modulus of their overlap with the state engendered by the operator on the groundstate.

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 can 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 (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 branch cut would start at $q^2=(2m)^2$.

Now we recall some complex analysis, in the form of the *Kramers-Kronig (or dispersion) relations*:

$$\mathrm{Re}\,G(z) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega \frac{\mathrm{Im}\,G(\omega)}{\omega - z}$$

(valid if $\mathrm{Im}\,G(\omega)$ 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, $q^2 = (3m)^2$.
- A sometimes useful fact which we've basically already shown (for real operators):

  $$\mathrm{Im}\mathcal{D}(q) = (2\pi)^D \sum_n \| \mathcal{O}_{0n} \|^2 \left(\delta^D(q-p_n) + \delta^D(q+p_n)\right) = \frac{1}{2}\int d^D x e^{\mathbf{i}qx} \langle 0| \left[\mathcal{O}(x), \mathcal{O}(0)\right] |0\rangle \ .$$

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 $s > 0$. 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

$$\mathcal{D}(k^2) = \int ds\ \rho(s) \frac{1}{k^2 - s + \mathbf{i}\epsilon}.$$

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. 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 §5.2.

Taking into account our assumption about single-particle states, this 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!)

- **Sum rule.** Finally, suppose that the field ϕ in question is a canonical 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[49]

$$1 = \int_0^{\infty} ds \rho(s). \tag{5.6}$$

[49] Here's how to see this. For free fields (chapter 2) 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):

$$\langle \Omega|[\phi(x), \phi(y)]|\Omega\rangle = \int d\mu^2 \rho(\mu^2) \left(\Delta_+(x - y, \mu^2) - \Delta_+(y - x, \mu^2)\right), \tag{5.5}$$

where ρ is the same spectral density as above. Now take $\partial_{x^0}|_{x^0 = y^0}$ of the BHS of (5.5) and use $\partial_t \Delta_+(x - y; \mu^2)|_{x^0 = y^0} = -\frac{\mathbf{i}}{2}\delta^{(d)}(\vec{x} - \vec{y})$.

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.

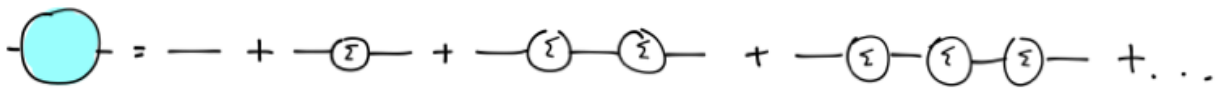
The idea of spectral representation and spectral density is more general than the Lorentz-invariant case. In particular, the spectral density of a Green's function is an important concept in the study of condensed matter. For example, the spectral density for the electron 2-point function is the thing that actually gets measured in angle-resolved photoemission experiments (ARPES).

5.2 Cutting rules and optical theorem

[Zee §III.8] So, that may have seemed like some math. What does this mean when we have in our hands a perturbative QFT? Consider the two point function of a relativistic scalar field ϕ which has a perturbative cubic interaction:

$$S = \int d^D x \left(\frac{1}{2} \left((\partial\phi)^2 + m^2\phi^2 \right) - \frac{g}{3!}\phi^3 \right).$$

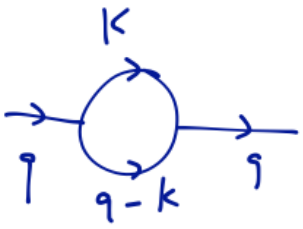
Sum the geometric series of 1PI insertions to get



$$\mathbf{i}\mathcal{D}_\phi(q) = \frac{\mathbf{i}}{q^2 - m^2 - \Sigma(q) + \mathbf{i}\epsilon}$$

where $\Sigma(q)$ is the 1PI two point vertex.

The leading contribution to Σ comes from the one loop diagram at right 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*.

[End of Lecture 15]

So: what is $\mathrm{Im}\,\Sigma_{1\text{ loop}}$ in this example?

We will use

$$\frac{1}{k^2 - m^2 + \mathbf{i}\epsilon} = \mathcal{P}\frac{1}{k^2 - m^2} - \mathbf{i}\pi\delta(k^2 - m^2) \equiv \mathcal{P} - \mathbf{i}\Delta$$

where $\mathcal{P}$ denotes 'principal part'. Then

$$\mathrm{Im}\,\Sigma_{1\text{ loop}}(q) = -\frac{1}{2}g^2 \int d\Phi\,(\mathcal{P}_1\mathcal{P}_2 - \Delta_1\Delta_2)$$

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). Altogether:

$$\begin{aligned}\mathbf{i}\Sigma(q) &= \frac{1}{2}\int d^dx\; e^{\mathbf{i}qx}\,(-\mathbf{i}g)^2\,\mathbf{i}\mathcal{D}(x)\mathbf{i}\mathcal{D}(x) \\ &= \frac{1}{2}g^2\int d\Phi \frac{1}{k_1^2-m_1^2+\mathbf{i}\epsilon}\frac{1}{k_2^2-m_2^2+\mathbf{i}\epsilon}\end{aligned} \tag{5.7}$$

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

$$\begin{aligned}0 &= \frac{1}{2}\int d^dx\; e^{\mathbf{i}qx}\,(\mathbf{i}g)^2\,\mathbf{i}\mathcal{D}_{\text{adv}}(x)\mathbf{i}\mathcal{D}_{\text{ret}}(x) \\ &= \frac{1}{2}g^2\int d\Phi \frac{1}{k_1^2-m_1^2-\sigma_1\mathbf{i}\epsilon}\frac{1}{k_2^2-m_2^2+\sigma_2\mathbf{i}\epsilon}\end{aligned} \tag{5.8}$$

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

$$\text{Im}\,(0) = \frac{1}{2}g^2\int d\Phi\,(\mathcal{P}_1\mathcal{P}_2 + \sigma_1\sigma_2\Delta_1\Delta_2)$$

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 which 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, which satisfy all the momentum conservation constraints:

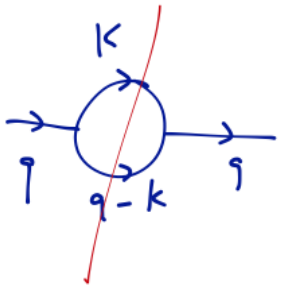
$$\mathrm{Im}\,\Sigma = \frac{1}{2}g^2 \int d\Phi 2\theta(k_1^0)\theta(k_2^0)\Delta_1\Delta_2 = \frac{1}{2}g^2 2 \int d\Phi\theta(k_1^0)\theta(k_2^0)\pi\delta(k_1^2-m^2)\pi\delta(k_2^2-m^2)$$

$$= \frac{g^2}{2}\frac{1}{2}\int \frac{\mathrm{đ}^{D-1}\vec{k}_1}{2\omega_{\vec{k}_1}}\frac{\mathrm{đ}^{D-1}\vec{k}_2}{2\omega_{\vec{k}_2}}(2\pi)^D\delta^D(k_1+k_2-q)\ .$$

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:

$$\mathrm{Im}\,\Sigma = \frac{1}{2}\sum_{\substack{\text{actual states } n \text{ of 2 particles}\\ \text{into which } \phi \text{ can decay}}} \|\,\mathcal{A}_{\phi\to n}\,\|^2 = m\Gamma. \tag{5.9}$$

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. 94 of my 215A notes).



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 which separates initial and final states (as at right) will 'cut' through some number of internal propagators; 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.

There is a small but important problem with the preceding discussion (pointed out by Brian Campbell-Deem): 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. But we can make the example viable (without changing the calculation at all) by thinking about a theory of two scalar fields, one light ϕ, one heavy Φ with lagrangian

$$\mathcal{L} = \frac{1}{2}\left((\partial\Phi)^2 - M^2\Phi^2 + (\partial\phi)^2 - m^2\phi^2 - g\phi^2\Phi\right)$$

and thinking about the self-energy for the (unstable) heavy particle.

The general form of (5.9) is a general consequence of unitarity. Recall that the S-matrix is

$$\mathcal{S}_{fi} = \langle f|\, e^{-\mathbf{i}\mathbf{H}T}\,|i\rangle \equiv (\mathbb{1} + \mathbf{i}\mathcal{T})_{fi}\,.$$

$$\mathbf{H} = \mathbf{H}^\dagger \implies \mathbb{1} = \mathcal{S}\mathcal{S}^\dagger \implies 2\mathrm{Im}\,\mathcal{T} \equiv \mathbf{i}\left(\mathcal{T}^\dagger - \mathcal{T}\right) \overset{\mathbb{1}=\mathcal{S}\mathcal{S}^\dagger}{=} \mathcal{T}^\dagger\mathcal{T}.$$

This is called the optical theorem and it is the same as the one taught in some QM classes. In terms of matrix elements:

$$2\mathrm{Im}\,\mathcal{T}_{fi} = \sum_n \mathcal{T}^\dagger_{fn}\mathcal{T}_{ni}$$

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 which can contribute to $\sum_n$ are two-particle states, so that $\sum_n$ will turn into $\int \frac{\mathrm{đ}\vec{k}_1}{2\omega_{k_1}} \frac{\mathrm{đ}\vec{k}_2}{2\omega_{k_2}}$, the two-particle density of states.

A bit more explicitly, introducing a basis of scattering states

$$\langle f|\,\mathcal{T}\,|i\rangle = \mathcal{T}_{fi} = \not{\delta}^4(p_f - p_i)\mathcal{M}_{fi}, \quad \mathcal{T}^\dagger_{fi} = \not{\delta}^4(p_f - p_i)\mathcal{M}^\star_{if}, \quad \text{(recall that } \not{\delta}^d \equiv (2\pi)^d\delta^d\text{)}$$

we have

$$\begin{aligned}\langle F|\,\mathcal{T}^\dagger\mathbb{1}\mathcal{T}\,|I\rangle &= \sum_n \langle F|\,\mathcal{T}^\dagger \sum_n \prod_{f=1}^n \int \frac{\mathrm{đ}^3 q_f}{2E_f}\,|\{q_f\}\rangle\,\langle\{q_f\}|\,\mathcal{T}\,|I\rangle \\ &= \sum_n \prod_{f=1}^n \int \frac{\mathrm{đ}^3 q_f}{2E_f} \not{\delta}^4(p_F - \sum_f q_f)\mathcal{M}^\star_{\{q_f\}F}\not{\delta}^4(p_I - \sum_f q_f)\mathcal{M}_{\{q_f\}I}\end{aligned}$$

Now notice that we have a $\not{\delta}^4(p_F - p_I)$ on both sides, and

$$\prod_{f=1}^n \int \frac{\mathrm{đ}^3 q_f}{2E_f} \not{\delta}^4(p_F - \sum_f q_f) = \int d\Pi_n$$

is the final-state phase space of the n particles. Therefore, the optical theorem says

$$\mathbf{i}\left(\mathcal{M}^\star_{IF} - \mathcal{M}_{FI}\right) = \sum_n \int d\Pi_n \mathcal{M}^\star_{\{q_f\}F}\mathcal{M}_{\{q_f\}I}.$$

Now consider *forward scattering*, $I = F$ (notice that here it is crucial that $\mathcal{M}$ is the transition matrix, $S = \mathbb{1} + \mathbf{i}\mathcal{T} = \mathbb{1} + \mathbf{i}\not{\delta}(p_T)\mathcal{M}$):

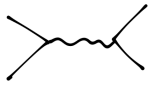
$$2\mathrm{Im}\,\mathcal{M}_{II} = \sum_n \int d\Pi_n |\mathcal{M}_{\{q_f\}I}|^2.$$

For the special case of 2-particle scattering, we can relate the RHS to the total cross section for $2 \to$ anything:

$$\mathrm{Im}\,\mathcal{M}(k_1, k_2 \leftarrow k_1, k_2) = 2E_{cm}p_{cm}\sigma(\text{anything} \leftarrow k_1, k_2).$$

Recall that for real x the imaginary part of a function of one variable with a branch cut, (like $\operatorname{Im}(x+\mathbf{i}\epsilon)^\nu = \frac{1}{2}\left((x+\mathbf{i}\epsilon)^\nu - (x-\mathbf{i}\epsilon)^\nu\right)$) is equal to (half) the *discontinuity* of the function $((x)^\nu)$ across the branch cut.

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: $\operatorname{Im}\Big($  $\Big)$.

So, consider the case $\mathbf{i}\mathcal{M} = \langle F|\, \mathbf{i}\mathcal{T}\, |I\rangle$ where both I and F are one-particle states. A special case of the LSZ formula says

$$\mathcal{M} = -\left(\sqrt{Z}\right)^2 \Sigma = -Z\Sigma \tag{5.10}$$

where $-\mathbf{i}\Sigma$ is the amputated 1-1 amplitude, that is, the self-energy, sum of all connected and amputated 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

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}}{p^2 - m_0^2 - \Sigma(p)} \simeq \frac{\mathbf{i}}{(p^2-m^2)\underbrace{\left(1 - \partial_{p^2} A|_{m^2}\right)}_{=Z^{-1}} - \mathbf{i}B} = \frac{\mathbf{i}Z}{(p^2-m^2) - \mathbf{i}BZ}. \tag{5.11}$$

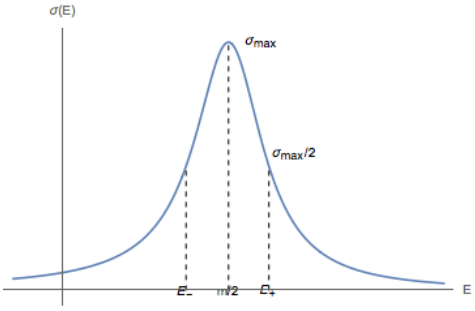
In terms of the *particle width* $\Gamma_w \equiv -ZB(m^2)/m$, this is

$$\tilde{G}^{(2)}(p) = \frac{\mathbf{i}Z}{(p^2-m^2) + \mathbf{i}m\Gamma_w}.$$

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 $E = E_\pm = \sqrt{\frac{m(m\pm\Gamma_w)}{4}} \simeq \frac{m}{2} \pm \frac{\Gamma}{4}$.



This width is the same as the decay rate, because of the optical theorem:

$$\Gamma_w = -\frac{BZ}{m} \overset{(5.10)}{=} -\frac{1}{m}\left(-\mathrm{Im}\,\mathcal{M}_{1\to1}\right) \overset{\text{optical}}{=} \frac{1}{m}\frac{1}{2}\sum_n \int_f d\Pi_n |\mathcal{M}_{\{q_f\}1}|^2 = \Gamma$$

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 (5.11) may not be such a good idea.

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 will 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:

$$\sigma(E \gg ...) \sim G^2 E^{-2-2k}. \tag{5.12}$$

This means that if $k \leq -1$, the cross section grows at high energy. 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.

5.3 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 $s = (k + k_+)^2 \gg m_e^2$:

$$\sigma^{\text{anything}\leftarrow e^+e^-} \overset{\text{optical thm}}{=} \frac{1}{2s}\text{Im}\,\mathcal{M}(e^+e^- \leftarrow e^+e^-) \tag{5.13}$$

where on the RHS, $\mathcal{M}$ is the forward scattering amplitude (meaning that the initial and final electrons have the same momenta). 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 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 (5.13):

$$\sigma^{\text{hadrons}\leftarrow e^+e^-} = \frac{1}{4}\sum_{\text{spins}}\frac{\text{Im}\,\mathcal{M}_h}{2s} = -\frac{4\pi\alpha}{s}\text{Im}\,\Pi_h(s). \tag{5.14}$$

(The initial and final spins are equal and we average over initial spins. We can ignore the longitudinal term $q^\mu q^\nu$ by the Ward identity. The spinor trace is $\sum_{\text{spins}}\bar{u}(k)\gamma_\mu v(k_+)\bar{v}(k_+)\gamma^\mu u(k) = -2k\cdot k_+ = -s$.) 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

$$\sigma^{L^+L^-\leftarrow e^+e^-} = \frac{4\pi}{3}\frac{\alpha^2}{s}F(M^2/s)$$

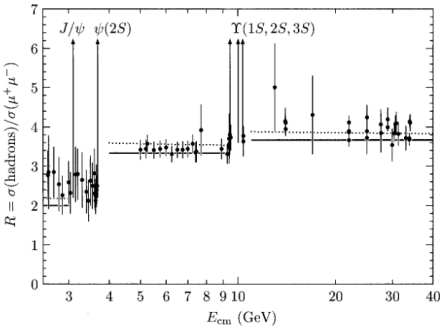
with

$$F(M^2/s) = \begin{cases} 0, & s < (2M)^2 \\ \sqrt{1-\frac{4M^2}{s}}\left(1+\frac{2M^2}{s}\right) = 1 + \mathcal{O}\left(M^2/s\right), & s > (2M)^2 \end{cases}.$$

In perturbative QCD, the leading order result is the same from each quark with small enough mass:

$$\sigma_0^{\text{quarks}\leftarrow e^+e^-} = \underbrace{3}_{\text{colors}} \sum_{\text{flavors, f}} Q_f^2 \frac{4\pi}{3}\frac{\alpha^2}{s} F(m_f^2/s).$$

This 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 $E = 2m_f$ for 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.



[End of Lecture 16]

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:

$$\mathbf{i}\Pi_h^{\mu\nu}(q) = -e^2 \int d^4x\; e^{-\mathbf{i}q\cdot x} \left\langle \Omega \right| \mathcal{T} J^\mu(x) J^\nu(0) \left| \Omega \right\rangle, \quad J^\mu(x) \equiv \sum_f Q_f \bar{q}_f(x) \gamma^\mu q_f(x).$$

(Here, the quark fields q_f are Dirac spinors, with Lagrangian $L_q = \sum_f \bar{q}_f \left(\mathbf{i}\not{\partial} - m_f\right) q_f$, $D_\mu = \partial_\mu - \mathbf{i}Q_f A_\mu + ...$, where the ... is the coupling to the gluon field which we'll discuss next chapter. They have a color index which runs from 1 to 3 which I've suppressed.) 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. (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 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

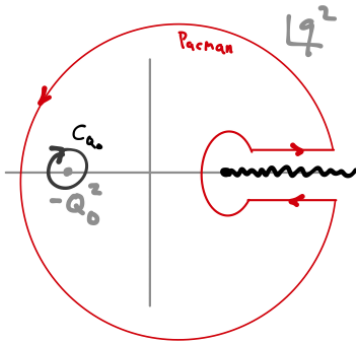
answer. One way to encode the information we know is to package it into moments:

$$I_n \equiv -4\pi\alpha \oint_{C_{Q_0}} \frac{dq^2}{2\pi\mathbf{i}} \frac{\Pi_h(q^2)}{(q^2+Q_0^2)^{n+1}} = -\frac{4\pi\alpha}{n!} \left(\partial_{q^2}\right)^n \Pi_h|_{q^2=-Q_0^2}.$$

The idea here is that the RHS can be computed by perturbative QCD. On the other hand, we know from the (appropriate generalization to currents of the) spectral representation sum rule (5.6) 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

$$\begin{aligned} I_n &= -4\pi\alpha \oint_{\text{Pacman}} \frac{dq^2}{2\pi\mathbf{i}} \frac{\Pi_h(q^2)}{(q^2+Q_0^2)^{n+1}} \\ &= -4\pi\alpha \int \frac{dq^2}{4\pi\mathbf{i}} \frac{\text{Disc}\Pi_h}{(q^2+Q_0^2)^{n+1}} \\ &\overset{(5.14)}{=} \frac{1}{\pi} \int_{s_{\text{threshhold}}}^{\infty} ds \frac{s}{(s+Q_0^2)^{n+1}} \sigma^{\text{hadrons}\leftarrow e^+e^-}(s). \end{aligned}$$



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.

6 Gauge theory

6.1 Massive vector fields as gauge fields

Consider a massive vector field B_μ with Lagrangian density

$$\mathcal{L}_B = -\frac{1}{4e^2}(dB)_{\mu\nu}(dB)^{\mu\nu} + \frac{1}{2}m^2 B_\mu B^\mu$$

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

$$\int d^D x \; j_\mu B^\mu = \int d^D x j_\mu A^\mu$$

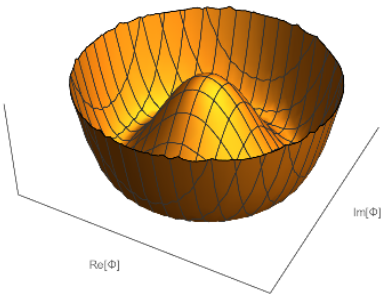
it is the same as coupling to A_μ.

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 $\theta(x) = 0$. This is called *unitary gauge*, and gives us back the Proca theory of $B = A$. Alternatively, consider the following slightly bigger (more dofs) theory:

$$\mathcal{L}_h \equiv -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}|D_\mu \Phi|^2 - V(|\Phi|)$$

where Φ is a complex, charged scalar field whose covariant derivative is $D_\mu \Phi = (\partial_\mu - \mathbf{i} A_\mu)\,\Phi$, and let's take

$$V(|\Phi|) = \kappa(|\Phi|^2 - v^2)^2$$



for some couplings κ, v. This is called an *Abelian Higgs model*. This potential has a $\mathsf{U}(1)$ symmetry $\Phi \to e^{\mathbf{i}\alpha}\Phi$, and a circle of minima at $|\Phi|^2 = v^2$ (if $v^2 > 0$, which we'll assume).

In polar coordinates in field space, $\Phi \equiv \rho e^{\mathbf{i}\theta}$, the Lagrangian is

$$\mathcal{L}_h = -\frac{1}{4e^2} F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\rho^2 (A_\mu - \partial_\mu \theta)^2 + (\partial\rho)^2 - V(\rho).$$

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

$$\partial_\rho^2 V|_{\rho=v} = m_\rho^2 = 8\kappa v^2 \overset{\kappa \gg 1}{\gg} m_A^2 = \langle \rho \rangle^2 = v^2.$$

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[50]. 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 renormalizible, 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, $0 = \partial_\mu F^{\mu\nu} - m^2 A^\nu \implies B(x) = Be^{-mx}$. 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, as you'll show on 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

[50] 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 (4.39).

the Cooper pair field (which has electric charge two). On the homework, you'll see a consequence of the charge of Φ for the flux quantization of vortices. I hope to say more about its origins in terms of electrons later in §9.8.1.

I mention here the Meissner effect and the resulting collimation of flux lines partly because it will be helpful for developing a picture of confinement. In particular: think about the energetics of a magnetic monopole (suppose we had one available[51]) 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.

6.2 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

$$\mathcal{L} = \sum_{\alpha=1}^{N} \partial_\mu \Phi_\alpha^\star \partial^\mu \Phi_\alpha - V(\Phi_\alpha^\star \Phi_\alpha) \tag{6.1}$$

(or $\mathcal{L} = \bar{\Psi}_\alpha \partial_\mu \Psi_\alpha$). The model actually has an $\mathsf{O}(2N)$ symmetry except that for kicks I grouped the scalars into pairs, and made the potential out of the combination $\sum_{\alpha=1}^{N} \Phi_\alpha^\star \Phi_\alpha$.

Lighting review of Lie groups and Lie algebras. (6.1) is invariant under the $\mathsf{U}(N)$ transformation

$$\Phi_\alpha \mapsto \Lambda_{\alpha\beta} \Phi_\beta, \quad \Lambda^\dagger \Lambda = \mathbb{1}. \tag{6.2}$$

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 which 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 non-abelian part: the T^A are the generators of $\mathsf{SU}(N)$, and are traceless, so $\mathsf{SU}(N) \ni$

[51] Here is the paper about the only one that's been detected by humans so far.

$\Lambda(\lambda^0 = 0)$ 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

$$[T^A, T^B] = \mathbf{i} f_{ABC} T^C \tag{6.3}$$

where f_{ABC} are the structure constants of the algebra. 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 (6.2), with Λ close to the identity, is

$$\Phi_\alpha \mapsto \Phi_\alpha + \mathbf{i}\lambda^A T^A_{\alpha\beta}\Phi_\beta. \tag{6.4}$$

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 which satisfy the same algebra (6.3), but can have other dimensions. For example, the structure constants themselves $\left(T^B_{\text{adj}}\right)_{AC} \equiv -\mathbf{i} f_{ABC}$ 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 (6.1). Let's consider how we might change the model in (6.1) 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:

$$\begin{gathered}(D_\mu \Phi)_\alpha \equiv \partial_\mu \Phi_\alpha - \mathbf{i} A^A_\mu T^A_{\alpha\beta}\Phi_\beta \\ \Phi \mapsto \Phi + \mathbf{i}\lambda^A(x) T^A \Phi, \quad A^A_\mu \mapsto A^A_\mu + \partial_\mu \lambda^A - f_{ABC}\lambda^B A^C_\mu(x).\end{gathered} \tag{6.5}$$

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:

$$\mathcal{L}_{YM} = -\frac{1}{4g^2}\sum_A \left(\underbrace{\partial_\mu A^A_\nu - \partial_\nu A^A_\mu + f_{ABC} A^B_\mu A^C_\nu}_{=F^A_{\mu\nu} = -F^A_{\nu\mu}} \right)^2 = -\frac{1}{4g^2}\text{tr} F_{\mu\nu}F^{\mu\nu}. \tag{6.6}$$

The field strength

$$F^A_{\mu\nu} \mapsto F^A_{\mu\nu} + f_{ABC}\lambda^B F^C_{\mu\nu} = F^A_{\mu\nu} + \mathbf{i}\lambda^B \left(T^B_{\text{adj}}\right)_{AC} F^C_{\mu\nu} \tag{6.7}$$

is designed so that it transforms in the adjoint representation, and therefore S_{YM} is gauge-invariant. (Regarding F as an $N \times N$ matrix $F = F^A T^A$, the finite version of (6.7) is $F \mapsto \Lambda F \Lambda^{-1}$ (compare (6.4)), which makes it manifest that $\text{tr} F^2$ is invariant.)

6.3 Interlude on differential forms (and algebraic topology)

[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 a completely antisymmetric p-index tensor,

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

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

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

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

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

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

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

[52] The space of p-forms on a manifold X is sometimes denoted $\Omega^p(X)$, especially when it is to be regarded as a vector space (let's say over $\mathbb{R}$).

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

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

where [..] means sum over permutations with a -1 for odd permutations. Try not to get caught up in the numerical prefactors. In my expression below for the exterior derivative also there is an annoying combinatorial prefactor.

The *exterior derivative* d acts on forms as

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

by

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

You can check that

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

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

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

[End of Lecture 17]

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

$$\int_{C_p} \omega_p - \int_{C'_p} \omega_p = \int_{C_p - C'_p} \omega_p = \int_{\partial R_{p+1}} \omega_p \overset{\text{Stokes}}{=} \int_{R_{p+1}} d\omega_p = 0.$$

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.

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

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

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 $b^0(S^1) = b^1(S^1) = 1$.

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

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

by

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

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

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

The machinery of differential forms is very useful.

6.4 Gauge fields as connections

The formulae in §6.2 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 $\Phi_\alpha(x) \mapsto \Lambda_{\alpha\beta}(x)\Phi(x)$. Suppose we would like to do physics in a way which 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 which respects these independent rotations. To do this, we need more structure: we need a *connection* (or *comparator*) W_{xy}, an object which transforms like $W_{xy} \mapsto \Lambda(x) W_{xy} \Lambda^{-1}(y)$, 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 $W(\emptyset) = \mathbb{1}$, $W(C_2 \circ C_1) = W(C_2)W(C_1)$ 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

$$D_\mu \Phi(x) \equiv \lim_{\Delta x \to 0} \frac{W(x, x+\Delta x)\Phi(x+\Delta x) - \Phi(x)}{\Delta x^\mu} \mapsto \Lambda(x) D_\mu \Phi(x) \ . \tag{6.8}$$

Expanding near $\Delta x \to 0$, we can let

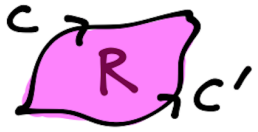
$$W(x, x+\Delta x) = \mathbb{1} - \mathbf{i} e \Delta x^\mu A_\mu(x) + \mathcal{O}(\Delta x^2) \tag{6.9}$$

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 equation (6.9) 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 in the exponent. 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.

[53]

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

$$dx^\mu dx^\nu [D_\mu, D_\nu]\Phi = -\mathbf{i}e dx^\mu dx^\nu F_{\mu\nu}\Phi.$$

The commutator of covariant derivatives is not an operator, but a function $[D_\mu, D_\nu] =$

[53] Which 2d surface? Let me speak about the abelian case for the rest of this footnote. 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 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.

$-\mathbf{i}eF_{\mu\nu}$. (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: $F_{\mu\nu} = F^A_{\mu\nu}T^A$, so more explicitly,

$$F^A_{\mu\nu} = \partial_\mu A^A_\nu - \partial_\nu A^A_\mu - \mathbf{i}e f_{ABC} A^B_\mu A^C_\nu.$$

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 (6.7)

$$F^A_{\mu\nu} \mapsto F^A_{\mu\nu} - f^{ABC}\lambda^B F^C_{\mu\nu}.$$

6.5 Actions for gauge fields

The Yang-Mills (YM) action (6.6) 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 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.

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. Because it is exact, this doesn't affect the equations of motion or perturbation theory (*e.g.* in $D = 4$, in the abelian case, $F \wedge F = d(A \wedge F)$, or in components, $\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} = 2\partial_\mu(\epsilon^{\mu\nu\rho\sigma}A_\nu F_{\rho\sigma})$) 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 coupling θ violates CP symmetry (notice that $F \wedge F$ has one time derivative and three spatial derivatives). 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 term (the $D = 2 + 1$ version of which we just encountered) which in the abelian case looks like $S_{\mathrm{CS}}[A] \overset{\text{abelian}}{=} \int A \wedge \frac{F}{2\pi} \wedge ... \wedge \frac{F}{2\pi}$ with $(D-1)/2$ factors of F. (In the non-Abelian case, there is an extra term: in 3d, $S_{CS}[A] \propto \int \mathrm{tr}\left(A \wedge F + \frac{2}{3}A \wedge A \wedge A\right)$.) 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. It is important in quantum Hall physics in $D = 2 + 1$, where it gives the gauge field fluctuations a mass.

In general dimension, we can make more couplings out of just A if we take more derivatives, but they will have higher dimension.

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

$$\bar{\psi} \gamma^\mu D_\mu \psi + V(\bar{\psi}\psi)$$

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.

You might expect that we would starting doing perturbation theory in g now. There is lots of physics there, but it takes a little while to get there. Given how limited our time is this quarter, we will instead think about how we might define the thing non-perturbatively and see what we learn from that. [End of Lecture 18]

6.6 Fermion path integrals

We'll need these for our discussion of anomalies, and they are extremely useful for doing perturbative gauge theory for QCD (which differs from Yang-Mills theory by the addition of fermionic quarks), and even for pure Yang-Mills theory.

[*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 consider $\hbar \to 0$, the fermi fields are a bunch of objects which 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

$$g(\theta_1, \theta_2) = a + b\theta_1 + c\theta_2 + d\theta_1\theta_2. \tag{6.10}$$

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 (6.10) $g(\theta_1, \theta_2)$ as if it is a function of the Grassmann variables. This doesn't really mean anything, but if we go along with it, then (6.10) is actually Taylor's theorem for Grassmann variables.

It's 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:

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

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:

$$\int (A + B\theta) d\theta = \int d\theta (A + B(\theta + \alpha)) \text{ if } \partial_x \alpha = 0.$$

The only integral, really, is the gaussian integral:

$$\int \underbrace{e^{-a\bar{\psi}\psi}}_{=1-a\bar{\psi}\psi} d\bar{\psi} d\psi = a.$$

Many of these give

$$\int e^{-\bar{\psi}_i A_{ij} \cdot \psi_j} \prod_{i=1}^{M} d\bar{\psi}_i \prod_{i=1}^{M} d\psi_i = \int \prod_{i=1}^{M} d\bar{\psi}_i \prod_{i=1}^{M} d\psi_i \left(1 - \bar{\psi} A \psi + \frac{1}{2} \bar{\psi} A \psi \bar{\psi} A \psi + \cdots\right) \tag{6.11}$$

$$= \frac{1}{n!} \sum_{\text{perms},\sigma} (-1)^{\sigma} A_{1\sigma_1} A_{2\sigma_2} \cdots A_{M\sigma_M} \tag{6.12}$$

$$= \det A. \tag{6.13}$$

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

$$\int e^{-\bar{\psi} \cdot A \cdot \psi} d\bar{\psi} d\psi = \det A = e^{+\operatorname{tr} \log A}$$

involves a sign in the exponent relative to the bosonic answer

$$\int e^{-\phi^\star \cdot A \cdot \phi} d\phi^\star d\phi = \frac{1}{\det A} = e^{-\operatorname{tr}\log A}.$$

This is the same sign as the minus sign associated to fermion loops.

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 grassman variables we use the action $S = \sum_i a_i \bar\psi_i \psi_i$ (diagonalize A above) then

$$\left\langle \bar\psi_i \psi_j \right\rangle = \frac{\delta_{ij}}{a_i} \equiv \langle \bar i j \rangle \tag{6.14}$$

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

$$\int e^{-\bar\psi_i A_{ij} \cdot \psi_j + \bar\eta_i \psi_i + \bar\psi_i \eta_i} \prod_{i=1}^M d\bar\psi_i \prod_{i=1}^M d\psi_i = e^{\bar\eta A^{-1} \eta} \int \prod d\bar\psi d\psi e^{-(\theta - \eta A^{-1}) A (\theta - A^{-1}\eta)} = e^{\bar\eta A^{-1}\eta} \det A.$$

Now we can take a continuum limit: $\psi_i \rightsquigarrow \psi(x), f(\theta) \rightsquigarrow f[\psi]$. The partition function for a free fermion field is

$$Z[\bar\eta, \eta] = \int [D\bar\psi D\psi] e^{\mathbf{i}\int d^D x \left(\bar\psi \left(\mathbf{i}\not\partial - m\right)\psi + \bar\eta\psi + \bar\psi\eta + \mathbf{i}\epsilon\bar\psi\psi\right)} \tag{6.15}$$

$$= \det\left(\mathbf{i}\not\partial - m\right) e^{\mathbf{i}\int d^D x \int d^D y \bar\eta(y) \left(\mathbf{i}\not\partial - m + \mathbf{i}\epsilon\right)^{-1} \eta(x)}. \tag{6.16}$$

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.

6.7 Lattice gauge theory

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^{ab}_{xy}. We demand that $U_{yx} = U^{-1}_{xy}$, as we did for the comparator in (6.9). 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 $a, b = 1..N$.

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.

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 thing 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 $S[U]$ which 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 $\mathsf{G} = \mathsf{U}(1)$, it is just $\int_0^{2\pi} đ\varphi$; for $\mathsf{G} = \mathbb{Z}_n$, it just $\sum_{\ell=1}^n$. 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.

This is a statistical mechanics problem of the 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. We can make something gauge invariant and as local as possible by considering a path C which goes around a single plaquette of the lattice: $C = \partial\square$. This is Wilson's action:

$$S[U] = \frac{1}{2f^2}\sum_\square S_\square, \quad S_\square \equiv \mathrm{Re}W(\partial\square) = \mathrm{Re}\,\mathrm{tr}\prod_{\ell\in\partial\square} U_\ell = \mathrm{Re}\,\mathrm{tr}\left(U_{x,x+dx}U_{x+dx,x+dx+dy}U_{x+dx+dy,x+dy}U_{x+dy,x}\right).$$

Now let's focus on he $\mathsf{G} = \mathsf{SU}(N)$ case, 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 $\mathsf{su}(N)$. An application of the CBH[54] 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,

$$\begin{aligned} S_{\square_{\mu\nu}} &\overset{\text{CBH}}{=} \frac{1}{2f^2}\text{Re}\,\text{tr}\left(e^{-\mathbf{i}a^2 F_{\mu\nu}} + \mathcal{O}(a^3)\right) \\ &= \frac{1}{2f^2}\text{Re}\,\text{tr}\left(\mathbb{1} - \mathbf{i}a^2 F_{\mu\nu} - \frac{1}{2}a^4 F_{\mu\nu}F_{\mu\nu} + \mathcal{O}(a^5)\right) \\ &= \frac{1}{2f^2}\left(\text{tr}\mathbb{1} - \frac{a^4}{2}\text{tr}F_{\mu\nu}F_{\mu\nu} + ...\right) = \mathcal{L}_{YM}(\square) + \text{const.} \end{aligned}$$

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, 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 $q_x^\dagger U_{xy}U_{yz}U_{zw}q_w$, or most simply, we can make a kinetic term for q by

$$S_q = \frac{1}{a^\#}\sum_{x,\ell} q_x^\dagger U_{x,x+\ell} q_{x+\ell} \simeq \int d^D x\; q^\dagger(x)\left(\not{D} - m\right)q(x) + ...$$

where $D_\mu = \partial_\mu - \mathbf{i}A_\mu$ is the covariant derivative, and we used its definition (6.8). The expression I've written is for a grassmann, spinor field; for bosonic fields 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 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). In the pure YM theory, a good one is our friend the Wilson loop $W(C) = \text{tr}\left(\prod_{\ell\in C} U_\ell\right) \simeq \text{tr}\mathcal{P}e^{\mathbf{i}\oint_C A}$. What physics does it encode? Recall what happened when we added an external source to measure the force

[54] Charlie-Baker-Hotel? Campbell-Baker-Hausdorff.

mediated by various fields, for example in the Maxwell theory:

$$\lim_{T\to\infty} Z^{-1} \int DA\; e^{\mathbf{i}S_{\text{Maxwell}}[A]+\mathbf{i}\int A_\mu J^\mu} = e^{-\mathbf{i}V(R)T}.$$

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 which is extensive in T. [End of Lecture 19]

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, 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.

A small warning about the area law: in general, 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). $\mathsf{G} = \mathsf{SU}(N)$ 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 two hints for how to think about it.

Hint 1: 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. 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.[55]

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

$$S_{\square_{\mu\nu}}[U] = -\left(1 - \cos\theta_{\mu\nu}\right),\ \ \theta_{\mu\nu}(x) = \theta_\mu(x+\nu) - \theta_\mu(x) - \theta_\nu(x+\mu) + \theta_\nu(x) \equiv \Delta_\nu \theta_\mu - \Delta_\mu \theta_\nu(x).$$

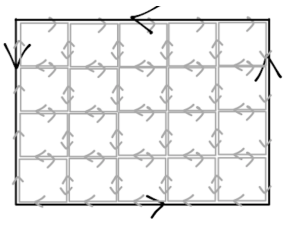
First let's consider the case where the world is a single plaquette. Then, using the identity $\int_0^{2\pi} \text{đ}\theta\ e^{\mathbf{i}n\theta} = \delta_{n,0}$,

$$\begin{aligned}\langle W(\square)\rangle &= \int \prod_\ell dU_\ell\ U_1 U_2 U_3 U_4 \left(1 + \beta(S_\square + S_\square^\dagger) + \frac{1}{2}\beta^2 \left(S + S^\dagger\right)^2 + \frac{1}{3!}\beta^3 \left(S + S^\dagger\right)^3 + \cdots\right) \\ &= \beta \underbrace{\langle S_\square S_{-\square}\rangle}_{=1} + \frac{\beta^3}{2}\langle S_{2\square} S_{-2\square}\rangle + \mathcal{O}(\beta^5) = \beta^{A(\square)}\left(1 + \mathcal{O}(\beta^2)\right) = e^{-f(\beta)\text{Area}} \end{aligned} \tag{6.17}$$

[55] 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.

with $f(\beta) = |\ln \beta|$ in this crude approximation. Here the area of the loop was just 1. I've written $S_{2\square} = S_\square^2$, which is only true in abelian cases.

If instead we consider a loop which encloses many plaquettes, we must pull down at least one factor of $\beta S_\square^\dagger$ for each plaquette, 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 $\langle W(C)\rangle = e^{-f(\beta)\text{Area}}\,(1 + \mathcal{O}(\beta^2))$, an area law.



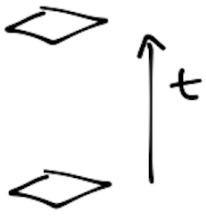
Since the series converges, this conclusion can be made completely rigorous. In what sense is confinement a mystery then? Well, a hint is that our argument applies equally well (and in fact the calculation we did was) for abelian gauge theory! But QED doesn't confine – we calculated the Wilson loop at weak coupling and found a perimeter law – what gives?

The answer is that there is a phase transition in between weak and strong coupling, so weak coupling 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}. Such surface models often exhibit a *roughening transition* as β becomes larger and floppy surfaces are not suppressed.

By the way, the same technology can be used to study the spectrum of excitations of the gauge theory, by considering correlations like

$$\left\langle S_R(t) S_R^\dagger(0)\right\rangle_c = \sum_\alpha |c_\alpha^R|^2 e^{-m_\alpha(R)t}$$



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), 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 $m_0(R) = -4\ln\beta + \sum_k m_k(R)\beta^k$.

Hint 2: monopole condensation and dual Meissner effect.

[Banks' book has a very nice discussion of this.] 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:

$$\partial^\mu F_{\mu\nu} = J_\nu^{(e)}, \partial^\mu \tilde{F}_{\mu\nu} = J_\nu^{(m)} \tag{6.18}$$

is invariant under the replacements

$$F_{\mu\nu} \to \tilde{F}_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}, \quad J_\nu^{(e)} \to J_\nu^{(m)}.$$

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 (6.18) 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*. 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. So, 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 that: 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 this happens 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.

7 Non-abelian gauge fields in perturbation theory

7.1 Gauge fixing and Feynman rules

Gauge fixing. [Peskin §16.2, Scwhartz §25] Consider the partition function for Yang-Mills theory (it will be easy to add the 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. We would like to make this cancellation (which is ∞/∞) more explicit by fixing a gauge, $G(A) = 0$. Perhaps surprisingly, this will be an application of fermion path integrals.

Note that the gauge-fixing function $G(A)$ must be a function of A which is *not* gauge invariant, such as $\partial^\mu A_\mu$. To do so, we will insert the following form of the number 1:

$$1 = \Delta[A] \int [D\alpha] \delta[G(A^\alpha)]. \tag{7.1}$$

Here

$$A^\alpha_\mu = (A^\alpha_\mu)^a T^a = e^{\mathbf{i}\alpha\cdot T} \left(A_\mu + \frac{\mathbf{i}}{g}\partial_\mu \right) e^{-\mathbf{i}\alpha\cdot T} \tag{7.2}$$

$$= \left(A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a + f^{abc} A^b_\mu \alpha^c \right) T^a + \mathcal{O}(\alpha^2) \tag{7.3}$$

$$= \left(A^a_\mu + \frac{1}{g} D_\mu \alpha^a \right) T^a + \mathcal{O}(\alpha^2) \tag{7.4}$$

is the gauge-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. The (Fadeev-Popov) determinant Δ defined by (7.1) is a Jacobian $\Delta[A] = \det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right)$. A crucial property of Δ is that it is gauge invariant:

$$\Delta[A^{\alpha_1}]^{-1} \overset{(7.1)}{=} \int [D\alpha]\delta[G(A^{\alpha+\alpha_1})] = \int [D\left(\alpha+\alpha_1\right)]\delta[G(A^{\alpha+\alpha_1})] = \Delta[A]^{-1}. \tag{7.5}$$

(Here we used the fact that the measure on the group is invariant, $[D\alpha] = [D\left(\alpha+\alpha_1\right)]$.)

Inserting this form of 1 (7.1) into Z (and changing the order of integration) gives:

$$Z = \frac{1}{\mathrm{vol}(\mathcal{G})} \int [D\alpha] \int [DA]\delta[G(A^\alpha)]\Delta[A]e^{\mathbf{i}S[A]} \tag{7.6}$$

$$= \frac{1}{\mathrm{vol}(\mathcal{G})} \int [D\alpha] \int [DA^\alpha]\delta[G(A^\alpha)]\Delta[A^\alpha]e^{\mathbf{i}S[A^\alpha]} \tag{7.7}$$

$$\overset{\tilde{A}\equiv A^\alpha}{=} \underbrace{\left(\frac{1}{\mathrm{vol}(\mathcal{G})} \int [D\alpha]\right)}_{=1} \int [D\tilde{A}]\delta[G(\tilde{A})]\Delta[\tilde{A}]e^{\mathbf{i}S[\tilde{A}]} \ . \tag{7.8}$$

In the first step we use the fact that $\int [DA] = \int [DA^\alpha], S[A] = S[A^\alpha]$ and (7.5). 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

$$G[A] = \partial^\mu A^A_\mu(x) - \omega^A(x).$$

Rather than picking a particular ω, let's average over all possibilities with gaussian measure:

$$1 = N(\xi) \int [D\omega] e^{-\mathbf{i}\int d^D x \frac{\omega^2(x)}{2\xi}}.$$

The normalization factor is just a constant which we can forget. Therefore

$$Z = N(\xi) \int [DA] \underbrace{\int [D\omega]\delta[\partial \cdot A - \omega]}_{=1} e^{-\mathbf{i}\int \frac{\omega^2}{2\xi}} \Delta[A]e^{\mathbf{i}S[A]} \tag{7.9}$$

$$= N(\xi) \int [DA]\Delta[A]e^{\mathbf{i}\left(S[A] - \int \frac{(\partial\cdot A)^2}{2\xi}\right)}. \tag{7.10}$$

Finally we must figure out what is $\Delta[A]$. Comparing to (7.4) (and remembering that there is a factor of $\delta[G]$ multiplying everything), Δ is the determinant of the operator

$$\frac{\delta G[A^\alpha]}{\delta\alpha} = \frac{1}{g}\partial^\mu D_\mu.$$

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.

$$\Delta = \det\left(\frac{1}{g}\partial^\mu D_\mu\right) = \int [DcD\bar{c}]e^{\mathbf{i}\int d^D x \bar{c}(-\partial^\mu D_\mu)c}.$$

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!

We saw all kinds of bad things about fermions with second-order kinetic terms in our discussion of spin-statistics. 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.)

Altogether,

$$Z = \int [DADcD\bar{c}] e^{\mathbf{i}\left(S[A] - \int \frac{(\partial \cdot A)^2}{2\xi} + \int \bar{c}(-\partial^\mu D_\mu)c\right)}. \tag{7.11}$$

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

$$Z = \left(\det\left(-\partial^2\right)^{-D/2} \det\left(-\partial^2\right)^{+1}\right)^{\dim(\mathsf{G})}.$$

The contribution of the ghosts cancels two components' worth of the contribution from the gluons.

Feynman rules. More explicitly, the ghost action is

$$\mathcal{L}_{\text{ghost}} = \bar{c}^a \left(-\partial^2 \delta^{ab} - g\vec{\partial}^\mu f^{abc} A^c_\mu\right) c^b.$$

The ghost propagator is then:

$$\left\langle c^a(x)\bar{c}^b(y)\right\rangle = \int \mathrm{d}^4k \; e^{-\mathbf{i}k(x-y)} \frac{-\mathbf{i}}{k^2 + \mathbf{i}\epsilon} \delta^{ab}.$$

Let us use the lovely Yang-Mills action, $S[A] = \int \mathcal{L}_{YM}$

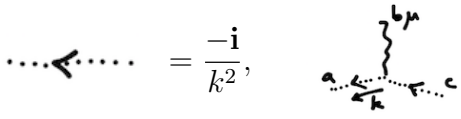
$$\mathcal{L}_{YM} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu}, \qquad F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu.$$

The resulting gluon propagator is

$$\left\langle A^a_\mu(x) A^b_\nu(y) \right\rangle = \int \mathrm{d}^4k \; e^{-\mathbf{i}k(x-y)} \frac{-\mathbf{i}}{k^2 + \mathbf{i}\epsilon} \delta^{ab} \left(\eta_{\mu\nu} - (1-\xi) \frac{k_\mu k_\nu}{k^2} \right)$$

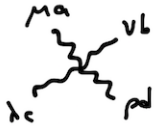
The gluon propagator is just like the photon one, times a δ^{ab} which conserves the color. $\xi = 1$ is Feynman gauge, which I'll use everywhere below.

The new Feynman rules are



$$= \frac{-\mathbf{i}}{k^2}, \qquad = -g f^{abc} k^\mu,$$

$$= g f^{abc} \left(\eta_{\mu\nu}(k_1 - k_2)_\lambda + \eta_{\nu\lambda}(k_2 - k_3)_\mu + \eta_{\lambda\mu}(k_3 - k_1)_\nu \right),$$



$$= -\mathbf{i}g^2 \left(f^{abe} f^{cde} \left(\eta_{\mu\lambda}\eta_{\nu\rho} - \eta_{\mu\rho}\eta_{\nu\lambda} \right) + f^{ade} f^{cbe} \left(\eta_{\mu\lambda}\eta_{\nu\rho} - \eta_{\mu\nu}\eta_{\rho\lambda} \right) + f^{ace} f^{bde} \left(\eta_{\mu\nu}\eta_{\lambda\rho} - \eta_{\mu\rho}\eta_{\nu\lambda} \right) \right).$$

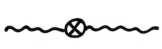
(Patterns: in the cubic coupling, the three terms cyclically permute the color and kinematic indices. In the quartic coupling, the second term is obtained from the first by the interchange $(b, \nu) \leftrightarrow (d, \rho)$, and the third is obtained from the first by $(b, \nu) \leftrightarrow (c, \lambda)$.)

Including quarks doesn't mess with the gauge-fixing stuff. We'll take

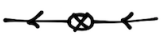
$$\mathcal{L}_{\text{quarks}} = \bar{q} \left(\mathbf{i}\not{D} - m \right) q = \bar{q}_i \left(\gamma^\mu \left(\mathbf{i}\partial_\mu \delta_{ij} + g A^a_\mu t^a_{ij} \right) - m\delta_{ij} \right) q_j$$

Here i, j are color indices. For QCD, $i, j = 1..3$ and and t^a_{ij} are the generators of $\mathsf{SU}(3)$ in the fundamental representation (a good basis of which are called Gell-Mann matrices). Then there's also a quark propagator, and the qqg vertex,  $= \mathbf{i}g\gamma^\mu t^a$.

We'll also need to add some counterterms

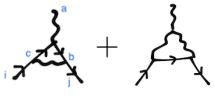


$$= -\mathbf{i}\left(k^2 \eta^{\mu\nu} - k^\mu k^\nu \right) \delta^{ab} \delta_3 \qquad \leftarrow$$



$$= \mathbf{i}\not{k}\delta_2 \qquad \leftarrow$$

$$= \mathbf{i}g t^a \gamma^\mu \delta_1 \qquad \leftarrow$$



.

On the right, I've indicated which one-loop diagrams require us to add these respective counterterms.

7.2 QCD beta function

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 is

$$\mathcal{L} = -\frac{1}{4}Z_3(\partial A)^2 + Z_2\bar{q}\left(\mathbf{i}\partial\!\!\!/ - Z_m m_R\right)q - Z_{3c}\bar{c}^a\Box c^a \tag{7.12}$$

$$- \mu^{\epsilon/2}g_R Z_{A^3} f^{abc}(\partial_\mu A^a_\nu)A^{\mu b}A^{\nu c} - \frac{1}{4}\mu^\epsilon g_R^2 Z_{A^4}\left(f^{abc}A^b_\mu A^c_\nu\right)\left(f^{ade}A^{\mu d}A^{\nu e}\right) \tag{7.13}$$

$$+ \mu^{\epsilon/2}g_R Z_1 A^a_\mu \bar{q}\gamma^\mu t^a q + \mu^{\epsilon/2}g_R Z_{1c} f^{abc}\partial_\mu\bar{c}^a A^b_\mu c^c \tag{7.14}$$

Here I've written the counterterms in terms of $Z = 1 - \delta$. 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.[56]

The *bare fields* are the ones whose quadratic terms are just $(\partial A^0)^2$ and $\overline{q^0}\mathbf{i}\partial\!\!\!/ q^0$, *i.e.* $A^0_\mu = \sqrt{Z_3}A_\mu, q^0 = \sqrt{Z_2}q, c^0 = \sqrt{Z_{3c}}c$. 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 of μ.[57] Therefore

$$0 = \mu\partial_\mu g_0 = \mu\partial_\mu\left(\mu^{\frac{\epsilon}{2}}g_R Z_1 Z_3^{-1/2}Z_2^{-1}Z_1\right) = g_0\left(\frac{\epsilon}{2} + \frac{1}{g_R}\underbrace{\mu\partial_\mu g_R}_{=\beta(g_R)} + \mu\partial_\mu\left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2\right)\right)$$

Now the counterterms δ will depend on μ through $g_R(\mu)$, so we can use the chain rule:

$$\mu\partial_\mu\delta = \mu\frac{dg_R}{d\mu}\frac{\partial}{\partial g_R}\delta = \beta(g_R)\frac{\partial\delta}{\partial g_R}.$$

[56] 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 $g \to \mu^{\epsilon/2}g$.

[57] 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 §4.2 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.

So we have

$$\beta(g_R) = -\frac{\epsilon}{2} g_R - g_R \mu \partial_\mu \left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2 \right) \tag{7.15}$$

$$= -\frac{\epsilon}{2} g_R - g_R \underbrace{\beta(g_R)}_{= -\frac{\epsilon}{2} g_R + \mathcal{O}(g_R^2)} \partial_{g_R} \left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2 \right) \tag{7.16}$$

$$= -\frac{\epsilon}{2} g_R + \frac{\epsilon}{2} g_R^2 \partial_{g_R} \left(\delta_1 - \frac{1}{2}\delta_3 - \delta_2 \right) + \mathcal{O}(g_R^4). \tag{7.17}$$

At the last step we solved the equation perturbatively. 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.

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

$$\mathbf{i}\mathcal{M}_q^{\mu\nu ab}(q) = \quad = -\mathrm{tr}_F(t^a t^b)\,(\mathbf{i}g)^2\,\bar{\mu}^{4-D}\int \mathrm{đ}^D k \frac{\mathbf{i}}{(q-k)^2 - m^2}\frac{\mathbf{i}}{k^2-m^2}\mathrm{tr}\left[\gamma^\mu\left(\not{k} - \not{q} + m\right)\gamma^\nu\left(\not{k} + m\right)\right].$$

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, 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 $D = 4 - \epsilon$,

$$\mathbf{i}\mathcal{M}_q^{\mu\nu ab}(q) = N_f T_F\left(q^2\eta^{\mu\nu} - q^\mu q^\nu\right)\delta^{ab}\frac{g^2}{16\pi^2}\left(-\frac{8}{3}\frac{1}{\epsilon} - \frac{20}{9} - \frac{4}{3}\ln\frac{\mu^2}{-q^2} + \mathcal{O}(\epsilon)\right) \tag{7.18}$$

where N_f is the number of flavors of quarks (*e.g.* up, down...), counting Dirac multiplets,

More novel are the gluon and ghost loops:

$$\mathbf{i}\mathcal{M}_{\text{ghost}}^{\mu\nu ab} = \quad = (-1)(-g)^2\int \mathrm{đ}^D k \frac{\mathbf{i}}{(k-q)^2}\frac{\mathbf{i}}{k^2} f^{cad} k^\mu f^{dbc}(k-q)^\nu$$

$$\overset{(7.19)}{=} g^2\frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}}\delta^{ab}C_2(\mathsf{G})\int_0^1 dx \left(\frac{1}{\Delta}\right)^{2-D/2}\left(\eta^{\mu\nu}\left(\frac{1}{2}\Gamma\left(2-\frac{D}{2}\right)\Delta\right) + q^\mu q^\nu\left(x(1-x)\Gamma\left(2-\frac{D}{2}\right)\right)\right.$$

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 $T_r^a T_r^a = C_2(r)\mathbb{1}_{d(r)\times d(r)}$. In particular, for the adjoint rep,

$$T^a_{\text{adj}} T^a_{\text{adj}} = (f^a)_{bc}(f^a)_{cd} = f^{abc}f^{acd} \equiv C_2(\mathsf{G})\delta^{bd}. \tag{7.19}$$

$C_2(r)$ is related to the normalization of the generators: $\text{tr} T_r^a T_r^b = c(r)\delta^{ab}$ (remember, we chose $c(\text{fundamental}) = \frac{1}{2}$). Contracting with δ^{ab} gives $d(r)C_2(r) = d(\mathsf{G})c(r)$. For the $\mathsf{SU}(N)$, $c_2(\mathsf{G}) = N$. See Peskin page 502 for a derivation.

$$\mathbf{i}\mathcal{M}_3^{\mu\nu ab} = \quad = \frac{g^2}{2}\bar{\mu}^{4-D}\int \mathrm{đ}^D k \frac{-\mathbf{i}}{k^2}\frac{-\mathbf{i}}{(k-q)^2} f^{acd}f^{bcd}N^{\mu\nu} \tag{7.20}$$

$$= -\frac{g^2}{2}\frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}}\delta^{ab}C_2(\mathsf{G})\int_0^1 dx \left(\frac{1}{\Delta}\right)^{2-D/2}\left(\eta^{\mu\nu}A + q^\mu q^\nu B + \eta^{\mu\nu}q^2 C\right). \tag{7.21}$$

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 $\mathcal{M}_4 \sim \int \frac{\mathrm{đ}^4 k}{k^2} \sim \Lambda^2$:

$$\mathbf{i}\mathcal{M}_4^{\mu\nu ab} = \quad = \frac{\mathbf{i}g^2}{2}\bar{\mu}^{4-D}\int \mathrm{đ}^D k \frac{-\mathbf{i}}{k^2}\eta^{\rho\lambda}\delta^{cd}\left(f^{abe}f^{cde}\left(\delta^\mu_\lambda\delta^\nu_\rho - \delta^\mu_\rho\delta^\nu_\lambda\right)\right.$$
$$\left. + f^{ade}f^{cbe}\left(\delta^\mu_\lambda\delta^\nu_\rho - \eta^{\mu\nu}\eta_{\rho\lambda}\right) + f^{ace}f^{bde}\left(\eta^{\mu\nu}\eta_{\lambda\rho} - \delta^\mu_\rho\delta^\nu_\lambda\right)\right)$$
$$= -g^2\delta^{ab}\eta^{\mu\nu}C_2(\mathsf{G})(D-1)\bar{\mu}^{4-D}\int \frac{\mathrm{đ}^D k}{k^2}\frac{(q-k)^2}{(q-k)^2}$$
$$= -g^2\delta^{ab}\eta^{\mu\nu}C_2(\mathsf{G})(D-1)\bar{\mu}^{4-D}\int_0^1 dx\left(\frac{1}{\Delta}\right)^{2-D/2}\left(-\frac{D}{2}\Gamma\left(1-\frac{D}{2}\right)\Delta + (1-x)^2q^2\Gamma\left(2-\frac{D}{2}\right)\right). \tag{7.22}$$

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. [58]

The glue contributions to the gluon vacuum polarization (not including quarks yet, since those are optional) are then

$$\mathcal{M}^{\mu\nu ab}_{\text{glue}}(q) = (\mathcal{M}_3 + \mathcal{M}_4 + \mathcal{M}_{\text{ghost}})^{\mu\nu ab} = \delta^{ab} C_2(\mathsf{G}) g^2 \frac{\bar{\mu}^{4-D}}{(4\pi)^{D/2}} \int_0^1 dx \left(\frac{1}{\Delta}\right)^{2-D/2} \cdot \left[\eta^{\mu\nu}\Gamma\left(1-\frac{D}{2}\right)\Delta\left(\underbrace{-\frac{1}{2}+\frac{3(D-1)}{2}-\frac{D(D-1)}{2}}_{=-\frac{1}{2}(D-2)^2}\right) + q^\mu q^\nu \Gamma\left(2-\frac{D}{2}\right)\mathfrak{a} + \eta^{\mu\nu} q^2 \Gamma\left(2-\frac{D}{2}\right)\mathfrak{b}\right]$$

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 $D=2$. 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 (7.18), 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 we must take

$$\delta_3 = \frac{1}{\epsilon}\frac{g^2}{16\pi^2}\left(\frac{10}{3}C_2(\mathsf{G}) - \frac{8}{3}N_f T_F\right).$$

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

$$\mathbf{i}\Sigma_2^{ij}(\not{p}) = \int \mathrm{d}^D k t^a_{ik} \gamma^\mu \frac{\mathbf{i}\not{k}\delta^{kl}}{k^2 + \mathbf{i}\epsilon} t^b_{lj} \gamma_\mu \frac{-\mathbf{i}\delta^{ab}}{(k-p)^2 + \mathbf{i}\epsilon}$$

[58] 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.

The color factors are

$$t^a_{ik} t^b_{lj} \delta^{ab} \delta^{kl} = \sum_a (t^a t^a)_{ij} = C_2(F) \delta_{ij}$$

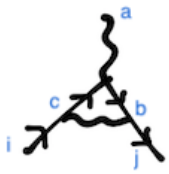
where $C_2(F)$ is the quadratic Casimir in the fundamental representation, which for $\mathsf{SU}(N)$ is $C_2(F) = \frac{N^2-1}{2N}$. The momentum integral is the same as in QED and we find

$$\Sigma_2^{ij}(\not{p}) = \frac{g^2}{8\pi^2} \delta^{ij} C_2(F) \frac{1}{\epsilon} \not{p} + \text{finite} \,. \tag{7.23}$$

The quark wavefunction renormalization counterterm δ_2 contributes as $\Sigma^{ij} = \ldots + \delta_2 \not{p}$, so we must set

$$\delta_2 = \frac{1}{\epsilon} \frac{g^2}{16\pi^2} \left(-2C_2(F)\right).$$

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 (notice that it's useful to keep the quark mass around for a bit here) and the color factors are

$$t^c t^a t^b \delta^{bc} = t^b t^a t^b = t^b t^b t^a + t^b [t^a, t^b] = C_2(F) T^a + \mathbf{i} f^{abc} t^b t^c.$$

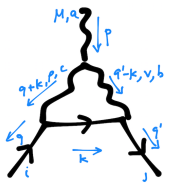
By antisymmetry of f^{abc}, the second term is

$$\mathbf{i} f^{abc} t^b t^c = \mathbf{i} f^{abc} \frac{1}{2} [t^b, t^c] = -\frac{1}{2} f^{abc} f^{bcd} t^d = -\frac{1}{2} C_2(\mathsf{G}) t^a.$$

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

$$-\mathbf{i}g\Gamma^\mu_{\text{new}}(p^2) = (\mathbf{i}g)^2 g\bar{\mu}^{4-D}\int \text{đ}^D k \gamma^\rho \frac{\mathbf{i}}{\not{k}}\gamma^\nu \frac{-\mathbf{i}}{(q+k)^2}\frac{-\mathbf{i}}{(q'-k)^2}\cdot \\ \left(\eta^{\mu\nu}(2q+q'+k)^\rho + \eta^{\nu\rho}(-q+q'-2k)^\mu + \eta^{\rho\mu}(k-2q'-q)^\nu\right)$$

(I find the opposite sign from Schwartz here. This sign cancels against the one in (7.24).) 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

$$\begin{aligned}
-\Gamma^\mu_{\text{new}}(0) &= g^2\bar{\mu}^{4-D}\int \text{đ}^D k \frac{\gamma_\rho \not{k}\gamma_\nu}{k^6}\left(\eta^{\mu\nu}k^\rho - 2\eta^{\nu\rho}k^\mu + \eta^{\rho\mu}k^\nu\right) \\
&= g^2\bar{\mu}^{4-D}\int \frac{\text{đ}^D k}{k^6}\left(2k^2\gamma^\mu - 2\gamma_\rho \not{k}\gamma^\rho k^\mu\right) \\
&= g^2\left(4-\frac{4}{D}\right)\gamma^\mu\bar{\mu}^{4-D}\int\frac{\text{đ}^D k}{k^4} \qquad \gamma_\rho\gamma^\nu\gamma^\rho = (2-D)\gamma^\nu,\ \int k^\mu k^\nu \ldots = \int \frac{k^2}{D}\eta^{\mu\nu}\ldots \\
&= g^2\mathbf{i}\gamma^\mu\frac{g^2}{16\pi^2}\left(\frac{6}{\epsilon} + 3\log\frac{\mu^2}{-p^2} + \text{finite}\right)
\end{aligned}$$

where at the last step we put back the gluon momentum to make up the dimensions. Finally the color factor is

$$f^{abc}(t^c t^b) = \frac{1}{2}\mathbf{i}f^{abc}f^{cbd}t^d = -\mathbf{i}\frac{1}{2}C_2(\mathsf{G})t^a. \tag{7.24}$$

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 (7.17), 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 SM), β is negative, in which case such a non-Abelian gauge theory is asymptotically free, as

promised many times. Defining a running coupling as in (4.47), we find a crucial plus sign relative to (4.47)

$$g^2_{\text{eff}}(q^2) = \frac{g^2}{1 + \frac{g^2}{16\pi^2} C \log\left(\frac{|q^2|}{M^2}\right)}, \quad C \equiv \frac{11}{3}N - \frac{2}{3}N_f$$

and the coupling grows as q decreases, and shrinks at large q. 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 $E^{ia} \equiv F^{0ia}$):

$$g\rho^a = D_i E^{ia} = \partial_i E^{ia} + g f^{abc} A^b_i E^{ic},$$

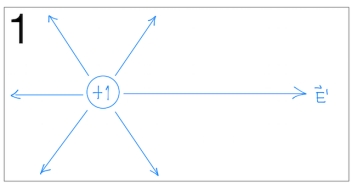
where ρ is the charge density (*e.g.* the number density of quarks if we included them in the theory), and for $\mathsf{SU}(2)$, the structure constants are $f^{abc} = \epsilon^{abc}$. Instead of dynamical quarks, let's consider a static color source particle: $\rho^a(x) = \delta^{(3)}(\vec{x})\delta^{a1}$ pointing in a particular color direction, so the equation we wish to solve is

$$\partial_i E^{ia} = g\delta^{(3)}(\vec{x})\delta^{a1} + g f^{abc} A^{bi} E^{ic}. \tag{7.25}$$

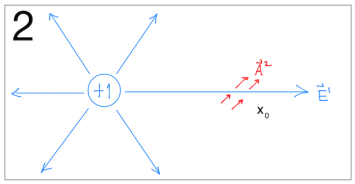
Let's solve this perturbatively in g in several steps.

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.

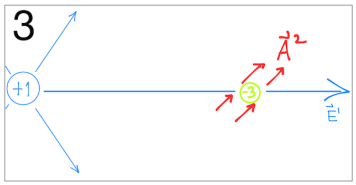


2. The quantum mechanics comes in here:
   consider a fluctuation of the vector potential in the 2d color direction $A^{b=2,i}(\vec{x})$, with support localized somewhere, call it x_0, away from the origin. Suppose it points in some direction, somewhat aligned with $\vec{x}$, its displacement from the source.

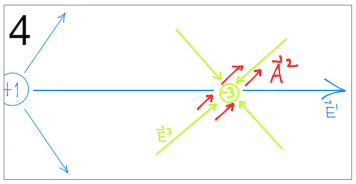


3. Here comes the iteration.

   The second term on the RHS of (7.25) 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.



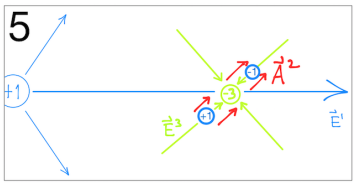
4. This produces a new Coulomb field $\vec{E}^3(\vec{x}) \sim -\frac{\vec{x}-\vec{x}_0}{|x-x_0|^3}$ pointing towards x_0.



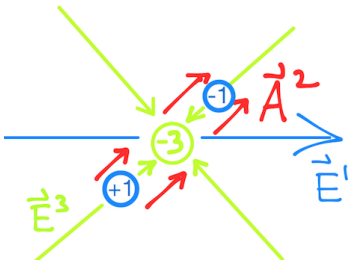
5. Now look at the second term on the RHS of (7.25) again:

$$\vec{\nabla} \cdot \vec{E}^1 = \ldots + g\epsilon^{123}\vec{A}^2 \cdot \vec{E}^3$$

   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.

8 Anomalies and fermion path integrals

8.1 Coherent state path integrals for fermions

[Peskin, §95, Shankar, *Principles of QM*, path integrals revisited. In this chapter of his great QM textbook, Shankar sneaks in lots of insights useful for modern condensed matter physics.]

Consider the algebra of a single fermion mode operator[59]:

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

With a single mode, the general Hamiltonian is

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

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

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

(In this example, it happens to equal $Z = 1 + e^{-\frac{\omega_0 - \mu}{T}}$, as you can see by computing the trace in the eigenbasis of $\mathbf{n} = \mathbf{c}^\dagger \mathbf{c}$. But never mind that; the one mode is a proxy for many, where it's not quite so easy to sum.) How do we write this trace as a path integral? We can do this by insertion lots of resolutions of the identity (this is sometimes called 'Trotterizing'), using any resolution of the identity on $\mathcal{H}$, so there can be many very-different-looking answers to this question.

Let's define coherent states for fermionic operators:

$$\mathbf{c} \, |\psi\rangle = \psi \, |\psi\rangle \, . \tag{8.1}$$

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

$$\psi_1 \psi_2 = -\psi_2 \psi_1, \psi \mathbf{c} = -\mathbf{c} \psi \tag{8.2}$$

– they anticommute with each other and with fermionic operators, and commute with ordinary numbers and bosons. They seem weird but they are easy. We'll need to consider multiple grassmann numbers when we have more than one fermion mode,

[59] For many modes,

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

where $\{\mathbf{c}_1, \mathbf{c}_2\} = 0$ will require that they anticommute $\{\psi_1, \psi_2\} = 0$ (as in the definition (8.2)); note that we will be simultaneously diagonalizing operators which *anti*commute.

The solution to equation (8.1) is very simple:

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

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

Similarly, the left-eigenvector of the creation operator is

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

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

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

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

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

Integration is just as easy and 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:

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

The only integral, really, is the gaussian integral:

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

Many of these give

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

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

$$\int e^{-\bar\psi\cdot A\cdot\psi} d\bar\psi d\psi = \det A = e^{+\mathrm{tr}\log A}$$

involves a sign in the exponent relative to the bosonic answer

$$\int e^{-\phi^\star \cdot A \cdot \phi} d\phi^\star d\phi = \overset{-1}{\det} A = e^{-\operatorname{tr}\log A}.$$

This is the same sign as the minus sign associated to fermion loops.

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

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

$$\left\langle \bar\psi_i \psi_j \right\rangle = \frac{\delta_{ij}}{a_i} \equiv \langle \bar i j \rangle \tag{8.3}$$

and Wick's theorem here is

$$\left\langle \bar\psi_i \bar\psi_j \psi_k \psi_l \right\rangle = \langle \bar i l \rangle \langle \bar j k \rangle - \langle \bar i k \rangle \langle \bar j l \rangle .$$

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

$$\mathbb{1} = \int d\bar\psi d\psi \; e^{-\bar\psi\psi} \left|\psi\right\rangle \left\langle\bar\psi\right| \tag{8.4}$$

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

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

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

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

Now insert (8.4) in between each pair of Trotter factors to get

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

Because of the $-\bar\psi$ in (8.4), to get this nice expression we had to define an extra letter

$$\bar\psi_M = -\bar\psi_0, \quad \psi_M = -\psi_0 \tag{8.5}$$

so we could replace $\langle -\bar\psi_0| = \langle \bar\psi_M|$.

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

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

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

So we have

$$\begin{aligned} Z &= \int \prod_{l=0}^{M-1} d\bar\psi_l d\psi_l e^{-\bar\psi_l\psi_l} e^{\bar\psi_{l+1}\psi_l} e^{-\Delta\tau H(\bar\psi_{l+1},\psi_l)} \\ &= \int \prod_{l=0}^{M-1} d\bar\psi_l d\psi_l \exp\left(\Delta\tau \left(\underbrace{\frac{\bar\psi_{l+1}-\bar\psi_l}{\Delta\tau}}_{=\partial_\tau\bar\psi} \psi_l - H(\bar\psi_{l+1},\psi_l)\right)\right) \\ &\simeq \int [D\bar\psi D\psi] \exp\left(\int_0^{1/T} d\tau\ \bar\psi(\tau)\left(-\partial_\tau - \omega_0 + \mu\right)\psi(\tau)\right) = \int [D\bar\psi D\psi] e^{-S[\bar\psi,\psi]}. \end{aligned} \tag{8.6}$$

Points to note:

- In the penultimate step we defined, as usual, continuum fields
$$\psi(\tau_l = \Delta\tau l) \equiv \psi_l, \qquad \bar\psi(\tau_l = \Delta\tau l) \equiv \bar\psi_l.$$
- We elided the difference $H(\bar\psi_{l+1},\psi_l) = H(\bar\psi_l,\psi_l) + \mathcal{O}(\Delta\tau)$ in the last expression. This difference is usually negligible and sometimes helpful (an example where it's helpful is the discussion of the number density below).
- The APBCs (8.5) on $\psi(\tau + \frac{1}{T}) = -\psi(\tau)$ mean that in its fourier representation[60]
$$\psi(\tau) = T\sum_n \psi(\omega) e^{-\mathbf{i}\omega_n\tau}, \qquad \bar\psi(\tau) = T\sum_n \bar\psi(\omega) e^{\mathbf{i}\omega_n\tau} \tag{8.7}$$
the *Matsubara frequencies*
$$\omega_n = (2n+1)\pi T, \quad n \in \mathbb{Z}$$
are odd multiples of πT.
- The measure $[D\bar\psi D\psi]$ is defined by this equation, just as in the bosonic path integral.

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

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

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

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

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

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

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

Continuum limit warning (about the red $\simeq$ in (8.6)). The Berry phase term is *actually*

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

and in (8.6) we have kept only the leading nonzero term:

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

Clearly this replacement is just fine if

$$\omega_n\tau \ll 1$$

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

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

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

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

Frequency space, $T \to 0$.

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

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

The partition function is

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

Notice that in the zero-temperature limit

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

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

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

Using the gaussian-integral formula (8.3) you can see that the *propagator* for ψ is

$$\left\langle\bar{\psi}(\omega_1)\psi(\omega_2)\right\rangle = \underbrace{\frac{\delta_{\omega_1,\omega_2}}{T}}_{\overset{T\to 0}{\to}\delta(\omega_1-\omega_2)}\ \frac{2\pi}{\mathbf{i}\omega_1 - \omega_0 + \mu}. \tag{8.8}$$

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

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

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

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

$$\begin{aligned}\langle\mathbf{N}\rangle &= \frac{1}{Z}\int D\bar{\psi}D\psi\ e^{-S[\bar{\psi},\psi]}\bar{\psi}(\tau_N+\Delta\tau)\psi(\tau_N)\\ &= \int_{-\infty}^{\infty}\text{đ}\omega\frac{e^{\mathbf{i}\omega\Delta\tau}}{\mathbf{i}\omega-\omega_0+\mu} = \theta(\mu-\omega_0).\end{aligned} \tag{8.9}$$

Which is the right answer: the mode is occupied in the groundstate only if $\omega_0 < \mu$. In the last step we used the fact that $\Delta\tau > 0$ to close the contour in the UHP; so we only pick up the pole if it is in the UHP. Notice that this quantity is very *UV sensitive*: if we put a frequency cutoff on the integral, $\int^\Lambda \frac{d\omega}{\omega} \sim \log\Lambda$, the integral diverges logarithmically. For most calculations the $\Delta\tau$ can be ignored, but here it told us the right way to treat the divergence. [61]

8.2 Anomalies

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

Topology means the study of quantities which can't vary smoothly, but can only jump. Like quantities which must be integers. *Anomalies* are an example of a topological phenomenon in QFT, which is therefore robust against any change in the QFT which can be made continuously (like varying masses or couplings, or the cutoff or the resolution of our description, *i.e.* a renormalization group transformation).

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

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

[61] The calculation between the first and second lines of (8.9) is familiar to us – it is a single Wick contraction, and can be described as a feynman diagram with one line between the two insertions. More prosaically, it is

$$\left\langle \bar\psi(\tau_N+\Delta\tau)\psi(\tau_N)\right\rangle \overset{(8.7)}{=} T^2\sum_{nm} e^{\mathbf{i}(\omega_n-\omega_m)\tau+\mathbf{i}\omega_n\Delta\tau}\left\langle\bar\psi(\omega_n)\psi(\omega_m)\right\rangle \overset{(8.8)}{=} T\sum_m \frac{e^{\mathbf{i}\omega_n\Delta\tau}}{\mathbf{i}\omega_n-\omega_0+\mu} \overset{T\to 0}{\to} \int \text{đ}\omega \frac{e^{\mathbf{i}\omega\Delta\tau}}{\mathbf{i}\omega-\omega_0+\mu}.$$

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

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

The example we will focus on here is the *chiral anomaly*. This is an equation for the violation of the chiral (aka axial) current for fermions coupled to a background gauge field. The chiral anomaly was first discovered in perturbation theory, by computing a certain Feynman diagram with a triangle; the calculation was motivated by the experimental observation of the process $\pi^0 \to \gamma\gamma$, which would not happen if the chiral current were conserved. (The relationship between the chiral current and the pion is explained in §9.7.)

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

8.2.1 Chiral anomaly

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

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

which means that γ^5 commutes with the Lorentz generators

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

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

Note that in $D = 4k$ dimensions, if ψ_L is a left-handed spinor in representation $\mathbf{r}$ of some group G, then its image under CPT, $\psi_L^{CPT}(t,\vec{x}) \equiv \mathbf{i}\gamma^0\,(\psi_L(-t,-\vec{x}))^\star$, is right-handed and transforms in representation $\bar{\mathbf{r}}$ of G. Therefore chiral symmetries arise

when the Weyl fermions transform in *complex representations* of the symmetry group, where $\bar{\mathbf{r}} \neq \mathbf{r}$. (In $D = 4k+2$, CPT maps left-handed fields to left-handed fields. For more detail on discrete symmetries and Dirac fields, see Peskin §3.6.)

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

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

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

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

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

For a continuous symmetry G, we can be more explicit about the meaning of a complex representation. The statement that ψ is in representation $\mathbf{r}$ means that its transformation law is

$$\delta\psi_a = \mathbf{i}\epsilon^A \left(t^A_{\mathbf{r}}\right)_{ab} \psi_b$$

where t^A, $A = 1..\dim G$ are generators of G in representation $\mathbf{r}$; for a compact lie group G, we may take the t^A to be Hermitian. The conjugate representation, by definition, is one with which you can make a singlet of G – it's the way $\psi^{\star T}$ transforms:

$$\delta\psi^{\star T}_a = -\mathbf{i}\epsilon^A \left(t^A_{\mathbf{r}}\right)^T_{ab} \psi^{\star T}_b \ .$$

So:

$$t^A_{\bar{\mathbf{r}}} = -\left(t^A_{\mathbf{r}}\right)^T.$$

The condition for a complex representation is that this is different from $t^A_{\mathbf{r}}$ (actually we have to allow for relabelling of the generators and the basis). The simplest case is $G = \mathrm{U}(1)$, where t is just a number indicating the charge. In that case, any nonzero charge gives a complex representation.

Consider the effective action produced by integrating out Dirac fermions coupled to a *background* gauge field (the gauge field is just going to sit there for this whole calculation):

$$e^{\mathbf{i}S_{\text{eff}}[A]} \equiv \int [D\psi D\bar{\psi}]\; e^{\mathbf{i}S[\psi,\bar{\psi},A]} \; .$$

We must specify how the fermions coupled to the gauge field. The simplest example is if A is a $U(1)$ gauge field and ψ is minimally coupled:

$$S[\psi,\bar{\psi},A] = \int \mathrm{d}^D x \bar{\psi}\mathbf{i}\not{D}\psi, \qquad \not{D}\psi \equiv \gamma^\mu \left(\partial_\mu + \mathbf{i}A_\mu\right)\psi.$$

We will focus on this example, but you could imagine instead that A_μ is a non-Abelian gauge field for the group G, and ψ is in a representation R, with gauge generators $T^A(R)$ ($A = 1...\dim G$), so the coupling would be

$$\bar{\psi}\mathbf{i}\not{D}\psi = \bar{\psi}_a\gamma^\mu\left(\partial_\mu\delta_{ab} + \mathbf{i}A^A_\mu T^A(R)_{ab}\right)\psi_b \; . \tag{8.10}$$

Much of the discussion below applies for any even D.

In the absence of a mass term, the action (in the Weyl basis) involves no coupling between L and R:

$$S[\psi,\bar{\psi},A] = \int \mathrm{d}^D x \left(\psi_L^\dagger \mathbf{i}\sigma^\mu D_\mu \psi_L + \psi_R^\dagger \mathbf{i}\bar{\sigma}^\mu D_\mu \psi_R\right)$$

and therefore is invariant under the global chiral rotation

$$\psi \to e^{\mathbf{i}\alpha\gamma^5}\psi, \quad \psi^\dagger \to \psi^\dagger e^{-\mathbf{i}\alpha\gamma^5}, \quad \bar{\psi} \to \bar{\psi}e^{+\mathbf{i}\alpha\gamma^5}. \quad \text{That is:} \;\; \psi_L \to e^{\mathbf{i}\alpha}\psi_L, \quad \psi_R \to e^{-\mathbf{i}\alpha}\psi_R.$$

(The mass term couples the two components

$$L_m = \bar{\psi}\left(\mathrm{Re}\, m + \mathrm{Im}\, m\gamma^5\right)\psi = m\psi_L^\dagger\psi_R + h.c.;$$

notice that the mass parameter is complex.) The associated Noether current is $j^5_\mu = \bar{\psi}\gamma^5\gamma_\mu\psi$, and it seems like we should have $\partial^\mu j^5_\mu \stackrel{?}{=} 0$. This follows from the massless (classical) Dirac equation $0 = \gamma^\mu\partial_\mu\psi$. (With the mass term, we would have instead $\partial^\mu j^5_\mu \stackrel{?}{=} 2\mathbf{i}\bar{\psi}\left(\mathrm{Re}\, m\gamma^5 + \mathrm{Im}\, m\right)\psi$.)

Notice that there is another current $j^\mu = \bar{\psi}\gamma^\mu\psi$. j^μ is the current which is coupled to the gauge field, $L \ni A_\mu j^\mu$. The conservation of this current is required for gauge invariance of the effective action

$$S_{\text{eff}}[A_\mu] \stackrel{!}{=} S_{\text{eff}}[A_\mu + \partial_\mu\lambda] \sim \log\left\langle e^{\mathbf{i}\int\lambda(x)\partial_\mu j^\mu}\right\rangle + S_{\text{eff}}[A_\mu].$$

No matter what happens we can't find an anomaly in j^μ. The anomalous one is the other one, the *axial current*.

To derive the conservation law we can use the Noether method. This amounts to substituting $\psi'(x) \equiv e^{\mathbf{i}\alpha(x)\gamma^5}\psi(x)$ into the action:

$$S_F[\psi'] = \int \mathrm{d}^D x \bar{\psi} e^{+\mathbf{i}\alpha\gamma^5} \mathbf{i}\not{D} e^{\mathbf{i}\alpha\gamma^5}\psi = \int \mathrm{d}^D x \left(\bar{\psi}\mathbf{i}\not{D}\psi + \bar{\psi}\mathbf{i}\gamma^5\left(\not{\partial}\alpha\right)\psi\right) \overset{\text{IBP}}{=} S_F[\psi] - \mathbf{i}\int \alpha(x)\partial^\mu \mathrm{tr}\bar{\psi}\gamma^5\gamma_\mu\psi.$$

Then we can completely get rid of $\alpha(x)$ if we can change integration variables, *i.e.* if $[D\psi'] \overset{?}{=} [D\psi]$. Usually this is true, but here we pick up an interesting Jacobian.

Claim:

$$e^{\mathbf{i}S_{\text{eff}}[A]} = \int [D\psi' D\bar{\psi}'] e^{\mathbf{i}S_F[\psi']} = \int [D\psi D\bar{\psi}] e^{\mathbf{i}S_F[\psi] + \int \mathrm{d}^D x \alpha(x)\left(\partial_\mu j_5^\mu - \mathcal{A}(x)\right)}$$

where

$$\mathcal{A}(x) = \sum_n \mathrm{tr}\bar{\xi}_n\gamma^5\xi_n \tag{8.11}$$

where ξ_n are a basis of eigenspinors of the Dirac operator. The contribution to $\mathcal{A}$ can be attributed to zeromodes of the Dirac operator.

The expression above is actually independent of α, since the path integral is invariant under a change of variables. For a conserved current, α would multiply the divergence of the current and this demand would imply current conservation. Here this implies that instead of current conservation we have a specific violation of the current:

$$\partial^\mu j_\mu^5 = \mathcal{A}(x).$$

What is the anomaly $\mathcal{A}$? [Polyakov §6.3] An alternative useful (perhaps more efficient) perspective is that the anomaly arises from trying to define the axial current operator, which after all is a composite operator. Thus we should try to compute

$$\langle \partial_\mu j_5^\mu \rangle = \partial_\mu \left\langle \bar{\psi}(x)\gamma^\mu\gamma^5\psi(x)\right\rangle$$

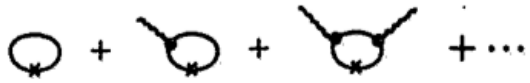
– the coincident operators on the RHS need to be regulated.

The classical (massless) Dirac equation immediately implies that the axial current is conserved

$$\partial_\mu\left(\bar{\psi}\gamma^\mu\gamma^5\psi\right) \overset{?}{=} 0.$$

Consider, on the other hand, the (Euclidean vacuum) expectation value

$$J_\mu^5 \equiv \left\langle \bar{\psi}(x)\gamma_\mu\gamma^5\psi(x)\right\rangle \equiv Z^{-1}[A]\int [D\psi D\bar{\psi}] e^{-S_F[\psi]} j_\mu^5(x)$$

$$= \quad \text{  }$$

$$= -\mathrm{Tr}\ \gamma\gamma_\mu\gamma^5 G^{[A]}(x,x) \tag{8.12}$$

where G is the Green's function of the Dirac operator in the gauge field background (and the figure is from Polyakov's book). We can construct it out of eigenfunctions of $\mathbf{i}\not{D}$:

$$\mathbf{i}\not{D}\xi_n(x) = \epsilon_n\xi_n(x), \qquad \bar{\xi}_n(x)\mathbf{i}\gamma^\mu\left(-\overleftarrow{\partial}_\mu + \mathbf{i}A_\mu\right) = \epsilon_n\bar{\xi}_n \tag{8.13}$$

in terms of which[62]

$$G(x,x') = \sum_n \frac{1}{\epsilon_n}\xi_n(x)\bar{\xi}_n(x'). \tag{8.14}$$

(I am suppressing spinor indices all over the place, note that here we are taking the outer product of the spinors.)

We want to define the coincidence limit, as $x' \to x$. The problem with this limit arises from the large $|\epsilon_n|$ eigenvalues; the contributions of such short-wavelength modes are local and most of them can be absorbed in renormalization of couplings. It should not (and does not) matter how we regulate them, but we must pick a regulator. A convenient choice here is heat-kernel regulator:

$$G_s(x,x') \equiv \sum_n e^{-s\epsilon_n^2}\frac{1}{\epsilon_n}\xi_n(x)\bar{\xi}_n(x')$$

and

$$J^5_\mu(x) = \sum_n e^{-s\epsilon_n^2}\frac{1}{\epsilon_n}\bar{\xi}_n(x)\gamma^5\gamma_\mu\xi_n(x)\ .$$

The anomaly is

$$\partial^\mu J^5_\mu = \partial^\mu\left\langle j^5_\mu\right\rangle = \sum_n \mathbf{i}\partial^\mu\left(\bar{\xi}_n\gamma_\mu\gamma^5\xi_n\right)\frac{e^{-s\epsilon_n^2}}{\epsilon_n}.$$

The definition (8.13) says

$$\mathbf{i}\partial^\mu\left(\bar{\xi}_n\gamma^5\gamma_\mu\xi_n\right) = 2\epsilon_n\bar{\xi}_n\gamma_5\xi_n$$

using $\{\gamma^5,\gamma^\mu\} = 0$. (Notice that the story would deviate dramatically here if we were studying the vector current which lacks the γ^5.) This gives

$$\partial^\mu J^5_\mu = 2\mathrm{Tr}\ \gamma^5 e^{-s\left(\mathbf{i}\not{D}\right)^2}$$

[62] Actually, this step is full of danger. (Polyakov has done it to me again. Thanks to Sridip Pal for discussions of this point.) See §8.2.2 below.

with

$$(\mathbf{i}\not{D})^2 = -\left(\gamma_\mu\left(\partial_\mu + \mathbf{i}A_\mu\right)\right)^2 = -\left(\partial_\mu + A_\mu\right)^2 - \frac{\mathbf{i}}{2}\Sigma_{\mu\nu}F^{\mu\nu}$$

where $\Sigma_{\mu\nu} \equiv \frac{1}{2}[\gamma_\mu, \gamma_\nu]$ is the spin Lorentz generator. This is (8.11), now better defined by the heat kernel regulator. We've shown that in any even dimension,

$$\partial^\mu \left\langle j^5_\mu(x)\right\rangle = 2\mathrm{Tr}\ {}_\alpha\gamma^5 e^{s\not{D}^2} \tag{8.15}$$

This can now be expanded in small s, which amounts to an expansion in powers of A, F. If there is no background field, $A = 0$, we get

$$\left\langle x|e^{-s\left(\mathbf{i}\not{\partial}\right)^2}|x\right\rangle = \int \mathrm{d}\!\!^{\text{-}D}p\ e^{-sp^2} = \underbrace{K_D}_{=\frac{\Omega_{D-1}}{(2\pi)^D}\text{ as before}} \frac{1}{s^{D/2}} \overset{D=4}{=} \frac{1}{16\pi^2 s^2}. \tag{8.16}$$

This term will renormalize the charge density

$$\rho(x) = \left\langle \psi^\dagger\psi(x)\right\rangle = \mathrm{tr}\gamma^0 G(x,x),$$

for which we must add a counterterm (in fact, it is accounted for by the counterterm for the gauge field kinetic term, *i.e.* the running of the gauge coupling). But it will not affect the axial current conservation which is proportional to

$$\mathrm{tr}\left(\gamma^5 G(x,x)\right)|_{A=0} \propto \mathrm{tr}\gamma^5 = 0.$$

Similarly, bringing down more powers of $(\partial + A)^2$ doesn't give something nonzero since the γ^5 remains.

In $D = 4$, the first term from expanding $\Sigma_{\mu\nu}F^{\mu\nu}$ is still zero from the spinor trace. (Not so in $D = 2$.) The first nonzero term comes from the next term:

$$\mathrm{tr}\left(\gamma_5 e^{-s\left(\mathbf{i}\not{D}\right)^2}\right)_{xx} = \underbrace{\left\langle x|e^{-s(\mathbf{i}D)^2}|x\right\rangle}_{\overset{(8.16)}{=}\frac{1}{16\pi^2 s^2}+\mathcal{O}(s^{-1})} \cdot\frac{s^2}{8}\cdot(\mathbf{i}^2)\underbrace{\mathrm{tr}\left(\gamma^5\Sigma^{\mu\nu}\Sigma^{\rho\lambda}\right)}_{=4\epsilon^{\mu\nu\rho\lambda}}\cdot\underbrace{\mathrm{tr}_c}_{\text{color}}(F_{\mu\nu}F_{\rho\lambda}) + \mathcal{O}(s^1)\ .$$

In the abelian case, just ignore the trace over color indices, tr_c. The terms that go like positive powers of s go away in the continuum limit. Therefore

$$\partial_\mu J^\mu_5 = -2\cdot\frac{1}{16\pi^2 s^2}\cdot\frac{s^2}{8}\cdot 4\epsilon^{\mu\nu\rho\lambda}\mathrm{tr}_c F_{\mu\nu}F_{\rho\lambda} + \mathcal{O}(s^1) = -\frac{1}{8\pi^2}\mathrm{tr}F_{\mu\nu}\left(\star F\right)^{\mu\nu}. \tag{8.17}$$

(Here $(\star F)^{\mu\nu} \equiv \frac{1}{8}\epsilon^{\mu\nu\rho\lambda}F_{\rho\lambda}$.) This is the chiral anomaly formula. It can also be usefully written as:

$$\partial_\mu J^\mu_5 = -\frac{1}{8\pi^2}\mathrm{tr}F\wedge F = -\frac{1}{32\pi^2}\vec{E}\cdot\vec{B}.$$

- This object on the RHS is a total derivative. In the abelian case it is

$$F \wedge F = \mathrm{d}\left(A \wedge F\right) \ .$$

  Its integral over spacetime is a topological (in fact $16\pi^2$ times an integer) characterizing the gauge field configuration. How do I know it is an integer? The anomaly formula! The change in the number of left-handed fermions minus the number of right-handed fermions during some time interval is:

$$\Delta Q_A \equiv \Delta\left(N_L - N_R\right) = \int \mathrm{d}t \partial_t J_0^5 = \int_{M_4} \partial^\mu J_\mu^5 = 2\int_{M_4} \frac{F \wedge F}{16\pi^2}$$

  where M_4 is the spacetime region under consideration. If nothing is going on at the boundaries of this spacetime region (*i.e.* the fields go to the vacuum, or there is no boundary, so that no fermions are entering or leaving), we can conclude that the RHS is an integer.

- Look back at the diagrams in (8.12). Which term in that expansion gave the nonzero contribution to the axial current violation? In $D = 4$ it is the diagram with three current insertions, the ABJ triangle diagram. So in fact we did end up computing the triangle diagram. But this calculation also shows that nothing else contributes, even non-perturbatively.

- We chose a particular regulator above. The answer we got did not depend on the cutoff; in fact, whatever regulator we used, we would get this answer.

- Consider what happens if we redo this calculation in other dimensions. We only consider even dimensions because in odd dimensions there is no analog of γ^5 – the Dirac spinor representation is irreducible. In $2n$ dimensions, we need n powers of F to soak up the indices on the epsilon tensor. Actually there is an analogous phenomenon in odd dimensions (sometimes called parity anomaly) of an effect that is independent of the masses of the fields which you'll study on the homework. Instead of F^n, the thing that appears is the Chern-Simons term.

- If we had kept the non-abelian structure in (8.10) through the whole calculation, the only difference is that the trace in (8.17) would have included a trace over representations of the gauge group; and we could have considered also a non-abelian flavor transformation

$$\psi_I \to \left(e^{\mathbf{i}\gamma^5 g^a \tau^a}\right)_{IJ} \psi_J$$

  for some flavor rotation generator τ^a. Then we would have found:

$$\partial^\mu j_\mu^{5a} = \frac{1}{16\pi^2}\epsilon^{\mu\nu\rho\lambda} F_{\mu\nu}^A F_{\rho\lambda}^B \mathrm{tr}_{c,a}\left(T^A T^B \tau^a\right) .$$

A similar statement applies to the case of multiple species of fermion fields: their contributions to the anomaly add. Sometimes they can cancel; the Electroweak gauge interactions are an example of this.

8.2.2 Zeromodes of the Dirac operator

Do you see now why I said that the step involving the fermion Green's function was full of danger? The danger arises because the Dirac operator (whose inverse is the Green's function) can have *zeromodes*, eigenspinors with eigenvalue $\epsilon_n = 0$. In that case, $\mathbf{i}\not{D}$ is *not* invertible, and the expression (8.14) for G is ambiguous. This factor of ϵ_n is about to be cancelled when we compute the divergence of the current and arrive at (8.11). Usually this kind of thing is not a problem because we can lift the zeromodes a little and put them back at the end. But here it is actually hiding something important. The zeromodes cannot just be lifted. This is true because nonzero modes of $\mathbf{i}\not{D}$ must come in left-right pairs: this is because $\{\gamma^5, \mathbf{i}\not{D}\} = 0$, so $\mathbf{i}\not{D}$ and γ^5 cannot be simultaneously diagonalized in general. That is: if $\mathbf{i}\not{D}\xi = \epsilon\xi$ then $(\gamma^5\xi)$ is also an eigenvector of $\mathbf{i}\not{D}\xi$, with eigenvalue $-\epsilon$. Only for $\epsilon = 0$ does this fail, so zeromodes can come by themselves. So you can't just smoothly change the eigenvalue of some ξ_0 from zero unless it has a partner with whom to pair up. By taking linear combinations

$$\chi_n^{L/R} = \frac{1}{2}\left(1 \pm \gamma^5\right)\xi_n$$

these two partners can be arranged into a pair of simultaneous eigenvectors of $(\mathbf{i}\not{D})^2$ (with eigenvalue ϵ_n^2) and of γ^5 with $\gamma^5 = \pm$ respectively.

This leads us to a deep fact, called the (Atiyah-Singer) *index theorem*: *only* zeromodes can contribute to the anomaly. Any mode ξ_n with nonzero eigenvalue has a partner (with the same eigenvalue of $(\mathbf{i}\not{D})^2$) with the opposite sign of γ^5; hence they cancel exactly in

$$\mathrm{tr}\gamma^5 e^{-s(\mathbf{i}\not{D})^2} = \sum_{n,L/R} \bar{\chi}_n^{L/R}\gamma^5\chi_n^{L/R} e^{-s\epsilon_n^2} + \text{zeromodes} \quad .$$

So the anomaly equation tells us that the number of zeromodes of the Dirac operator, weighted by handedness (*i.e.* with a + for L and - for R) is equal to

$$N_L - N_R = \int d^D x \mathcal{A}(x) = \int \frac{1}{16\pi^2} F \wedge F.$$

A practical consequence for us is that it makes manifest that the result is independent of the regulator s.

8.2.3 The physics of the anomaly

[Polyakov, page 102; Kaplan 0912.2560 §2.1; Alvarez-Gaumé] Consider non-relativistic free (*i.e.* no 4-fermion interactions) fermions in 1+1 dimensions, *e.g.* with 1-particle dispersion $\omega_k = \frac{1}{2m}\vec{k}^2$. The groundstate of N such fermions is described by filling the N lowest-energy single particle levels, up the Fermi momentum: $|k| \leq k_F$ are filled. We must introduce an infrared regulator so that the levels are discrete – put them in a box of length L, so that $k_n = \frac{2\pi n}{L}$. (In Figure 3, the red circles are possible 1-particle states, and the green ones are the occupied ones.) The lowest-energy excitations of this groundstate come from taking a fermion just below the Fermi level $k = |k_F - k_1|$ and putting it just above, $k = |k_F + k_2|$; the energy cost is

$$E_{k_1-k_2} = \frac{1}{2m}(k_F + k_1)^2 - \frac{1}{2m}(k_F - k_2)^2 \simeq \frac{k_F}{m}(k_1 - k_2)$$

The energy – we get relativistic dispersion with velocity $v_F = \frac{k_F}{m}$. The fields near these Fermi points in k-space satisfy the Dirac equation[63]:

$$(\omega - v_F\delta k)\,\psi_L = 0, \quad (\omega + v_F\delta k)\,\psi_R = 0$$

where $\delta k \equiv k - k_F$.

[63] This example is worthwhile for us also because we see the relativistic Dirac equation is emerging from a non-relativistic model; in fact we could have started from an even more distant starting point – *e.g.* from a lattice model, like

$$H = -t\sum_n c_n^\dagger c_{n+1} + h.c.$$

where the dispersion would be $\omega_k = -2t(\cos ka - 1) \sim \frac{1}{2m}k^2 + \mathcal{O}(k^4)$ with $\frac{1}{2m} = ta^2$.

It would therefore seem to imply a conserved axial current – the number of left moving fermions minus the number of right moving fermions. But the fields ψ_L and ψ_R are not independent; with high-enough energy excitations, you reach the bottom of the band (near $k = 0$ here) and you can't tell the difference. This means that the numbers are *not* separately conserved.

We can do better in this 1+1d example and show that the amount by which the axial current is violated is given by the anomaly formula. Consider subjecting our poor 1+1d free fermions to an electric field $E_x(t)$ which is constant in space and slowly varies in time. Suppose we gradually turn it on and then turn it off; here gradually means slowly enough that the process is adiabatic. Then each particle experiences a force $\partial_t p = eE_x$ and its net change in momentum is

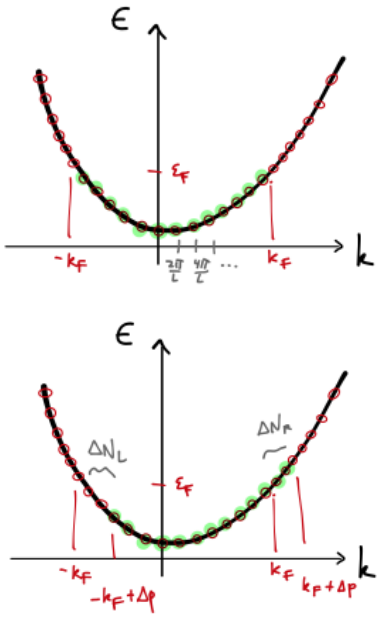


Figure 3: Green dots represent occupied 1-particle states. Top: In the groundstate. Bottom: After applying $E_x(t)$.

$$\Delta p = e \int \mathrm{d}t E_x(t).$$

This means that the electric field puts the fermions in a state where the Fermi surface $k = k_F$ has shifted to the right by Δp, as in the figure. Notice that the total number of fermions is of course the same – charge is conserved.

Now consider the point of view of the low-energy theory at the Fermi points. This theory has the action

$$S[\psi] = \int \mathrm{d}x\mathrm{d}t\bar{\psi}\left(\mathbf{i}\gamma^\mu\partial_\mu\right)\psi \ ,$$

where γ^μ are 2×2 and the upper/lower component of ψ creates fermions near the left/right Fermi point. In the process above, we have added N_R right-moving particles and taken away N_L left-moving particles, that is *added* N_L left-moving holes (aka anti-particles). The axial charge of the state has changed by

$$\Delta Q_A = \Delta(N_L - N_R) = 2\frac{\Delta p}{2\pi/L} = \frac{L}{\pi}\Delta p = \frac{L}{\pi}e\int \mathrm{d}t E_x(t) = \frac{e}{\pi}\int \mathrm{d}t\mathrm{d}x E_x = \frac{e}{2\pi}\int \epsilon_{\mu\nu}F^{\mu\nu}$$

On the other hand, the LHS is $\Delta Q_A = \int \partial^\mu J^A_\mu$. We can infer a local version of this equation by letting E vary slowly in space as well, and we conclude that

$$\partial_\mu J^\mu_A = \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu}.$$

This agrees exactly with the anomaly equation in $D = 1+1$ produced by the calculation above in (8.15) (see the homework).

9 Effective field theory

9.1 A parable on integrating out degrees of freedom

Here's a second 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 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:

$$\begin{aligned} e^{-S_{\rm eff}[q]} &:= \int [dQ] e^{-S[q,Q]} \\ &= e^{-S_{\omega_0}[q]} \left\langle e^{-S_{\rm int}[Q,q]} \right\rangle_Q \end{aligned}$$

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 $S_{\rm int}$:

$$\left\langle e^{-S_{\rm int}[Q,q]} \right\rangle_Q = \int [dQ] e^{-S_\Omega[Q] - \int dsJ(s)Q(s)} = \mathcal{N} e^{\frac{1}{4}\int dsdtJ(s)G(s,t)J(t)} .$$

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:

$$S_\Omega[Q] = \int dsdtQ(s)G^{-1}(s,t)Q(t).$$

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

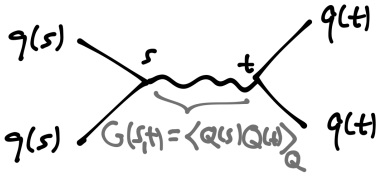
$$G(s) = \int đ\omega e^{-\mathbf{i}\omega s} \frac{1}{\omega^2 + \Omega^2}. \tag{9.1}$$

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

$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dtds \frac{g^2}{2} q(s)^2 G(s,t) q(t)^2 \ . \tag{9.2}$$



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.

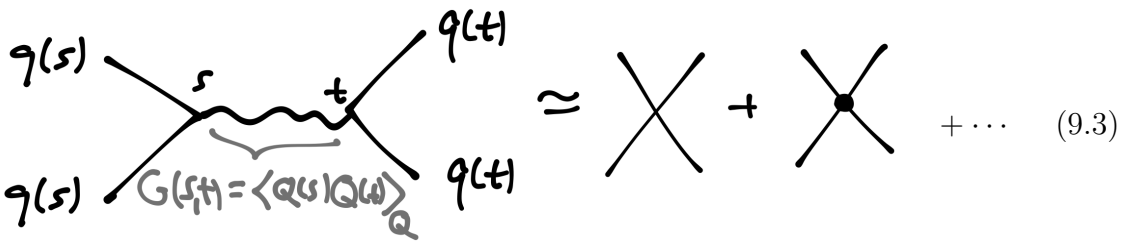
But now suppose we are interested in times much longer than $1/\Omega$, say times comparable to the period of oscillation of the less-stiff spring $2\pi/\omega$. We can accomplish this by Taylor expanding under the integrand in (9.1):

$$G(s) \overset{s \gg 1/\Omega}{\simeq} \int đ\omega e^{-\mathbf{i}\omega s} \frac{1}{\Omega^2} \underbrace{\frac{1}{1 + \frac{\omega^2}{\Omega^2}}}_{= \sum_n (-1)^n \left(\frac{\omega^2}{\Omega^2}\right)^n} \simeq \frac{1}{\Omega^2}\delta(s) + \frac{1}{\Omega^4}\partial_s^2 \delta(s) + \ldots$$

Plug this back into (9.2):

$$S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dt \frac{g^2}{2\Omega^2} q(t)^4 + \int dt \frac{g^2}{2\Omega^4} \dot{q}^2 q^2 + \ldots$$

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.



(9.3)

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 x 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 at (5.12). 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.

9.1.1 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 paradigm 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). Alas, this is all I'll say about it until Physics 217 in Fall 2018.

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. 9.3.

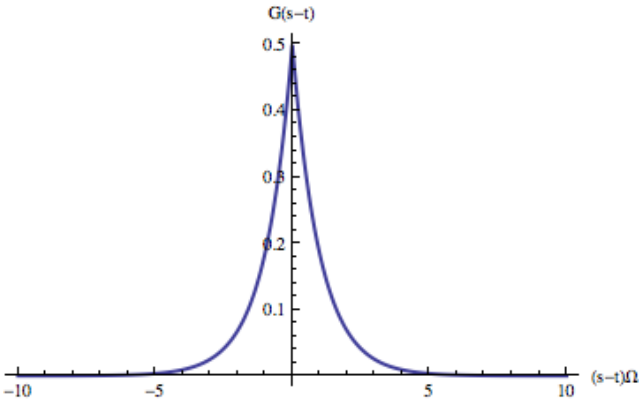
If we have the hubris to care about the exact answer, it's nonlocal in time. But if we want exact answers then we'll have to do the integral over q, too. On the other hand, the hierarchy of scales $\omega_0 \ll \Omega$ is useful if we ask questions about energies of order ω_0, e.g.

$$\langle q(t)q(0)\rangle \text{ with } t \sim \frac{1}{\omega_0} \gg \Omega$$

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 that 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 the-

ory of quantum gravity which 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:

$$G(s) = \int đ\omega \frac{e^{-\mathbf{i}\omega s}}{\omega^2+\Omega^2} = e^{-\Omega|s|}\frac{1}{2\Omega}. \tag{9.4}$$

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/introduce 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:

$$Z[J] \equiv \left\langle e^{\int dt Q(t) J(t)} \right\rangle_Q \ , \quad \langle Q(t_1)Q(t_2)...\rangle_Q = \frac{\delta}{\delta J(t_1)}\frac{\delta}{\delta J(t_2)}... \log Z[J]$$

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:

$$G(s,t) = \langle Q(s)Q(t)\rangle_Q = \frac{\delta}{\delta J(t)}\frac{\delta}{\delta J(s)}\log Z[J].$$

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.

9.2 Introduction to effective field theory

[Some nice lecture notes on effective field theory can be found here: J. Polchinski, A. Manohar, I. Rothstein, D. B. Kaplan, H. Georgi. Aneesh Manohar has written an excellent and provocative new set of lectures from last year's Les Houches school which should appear on the arXiv sometime soon.]

Diatribe about 'renormalizability'. Taking the example of the previous subsection to its logical conclusion, we are led to the idea of an *effective field theory* (EFT). (The Wilsonian perspective on renormalization – namely that we should include all possible operators consistent with symmetries and let dimensional analysis and the dynamics decide which are important at low energies – makes this idea even more inevitable.) There is no reason to demand that a field theory that we have found to describe physics in some regime should be a valid description of the world to arbitrarily short (or long!) distances. This is a happy statement: there can always be new physics that has been so far hidden from us. Rather, an EFT comes with a regime of validity, and with necessary cutoffs. As we will discuss, in a useful implementation of an EFT,

the cutoff implies a small parameter in which we can expand (and hence compute). (In the example of $S_{\text{eff}}[q]$ of the previous subsection, the small parameter is ω/Ω.)

Caring about renormalizibility is pretending to know about physics at arbitrarily short distances. Which you don't.

Even when theories are renormalizable, this apparent victory is often false. For example, QED requires only two independent counterterms (for the mass and for the fine structure constant), and is therefore by the old-fashioned definition renormalizable, but it is superseded by the electroweak theory above 80GeV. Also: the coupling in QED actually increases logarithmically at shorter distances, and ultimately reaches a *Landau pole* at SOME RIDICULOUSLY HIGH ENERGY (of order $e^{+\frac{c}{\alpha}}$ where $\alpha \sim \frac{1}{137}$ is the fine structure constant (*e.g.* at the scale of atomic physics) and c is some numerical number. Plugging in numbers gives something like 10^{330} GeV, which is quite a bit larger than the Planck scale). This is of course completely irrelevant for physics and even in principle because of the previous remark about electroweak unification. And if not because of that, because of the Planck scale. A heartbreaking historical fact is that Landau and many other smart people gave up on QFT as a whole because of this silly fantasy about QED in an unphysical regime.

We will see below that even in QFTs which are non-renormalizable in the strict sense, there is a more useful notion of renormalizability: effective field theories come with a small parameter (often some ratio of mass scales), in which we may expand the action. A useful EFT requires a finite number of counterterms *at each order in the expansion*.

Furthermore, I claim that this is *always* the definition of renormalizability that we are using, even if we are using a theory which is renormalizable in the traditional sense, which allows us to pretend that there is no cutoff. That is, there could always be corrections of order $\left(\frac{E}{E_{\text{new}}}\right)^n$ where E is some energy scale of physics that we are doing and E_{new} is some UV scale where new physics might come in; for large enough n, this is too small for us to have seen. The property of renormalizibility that actually matters is that we need a finite number of counterterms at each order in the expansion in $\frac{E}{E_{\text{new}}}$.

Renormalizable QFTs are in some sense *less* powerful than non-renormalizable ones – the latter have the decency to tell us when they are giving the wrong answer! That is, they tell us at what energy new physics *must* come in; with a renormalizable theory we may blithely pretend that it is valid in some ridiculously inappropriate regime like 10^{330} GeV.

Notions of EFT. There is a dichotomy in the way EFTs are used. Sometimes one knows a lot about the UV theory (*e.g.*

- electroweak gauge theory,
- QCD,
- electrons in a solid,
- water molecules

...) but it is complicated and unwieldy for the questions one wants to answer, so instead one develops an effective field theory involving just the appropriate and important dofs (*e.g.*, respectively,

- Fermi theory of weak interactions,
- chiral lagrangian (or HQET or SCET or ...),
- Landau Fermi liquid theory (or the Hubbard model or a topological field theory or ...),
- hydrodynamics (or some theory of phonons in ice or ...)

...). As you can see from the preceding lists of examples, even a single UV theory can have many different IR EFTs depending on what phase it is in, and depending on what question one wants to ask. The relationship between the pairs of theories above is always coarse-graining from the UV to the IR, though exactly what plays the role of the RG parameter can vary wildly. For example, in the example of the Fermi liquid theory, the scaling is $\omega \to 0$, and momenta scale towards the Fermi surface, not $\vec{k} = 0$.

A second situation is when one knows a description of some low-energy physics up to some UV scale, and wants to try to infer what the UV theory might be. This is a common situation in physics! Prominent examples include: the Standard Model, and quantized Einstein gravity. Occasionally we (humans) actually learn some physics and an example of an EFT from the second category moves to the first category.

Instructions for EFT. Answer the following questions:

1. what are the dofs?
2. what are the symmetries?
3. where is the cutoff on its validity?

Then write down all interactions between the dofs which preserve the symmetry in an expansion in derivatives, with higher-dimension operators suppressed by more powers of the UV scale.

I must also emphasize two distinct usages of the term 'effective field theory' which are common, and which the discussion above is guilty of conflating (this (often slippery) distinction is emphasized in the review article by Georgi linked at the beginning of this subsection). The Wilsonian perspective advocated above produces a low-energy description of the physics which is really just a way of solving (if you can) the original model; very reductively, it's just a physically well-motivated order for doing the integrals. If you really integrate out the high energy modes exactly, you will get a non-local action for the low energy modes. This is to be contrasted with the local actions one uses in practice, by truncating the derivative expansion. It is the latter which is really the action of the effective field theory, as opposed to the *full theory*, with some of the integrals done already. The latter will give correct answers for physics below the cutoff scale, and it will give them much more easily.

Some interesting and/or important examples where EFT has been useful (most of which we will not discuss explicitly) and where you can learn about them:

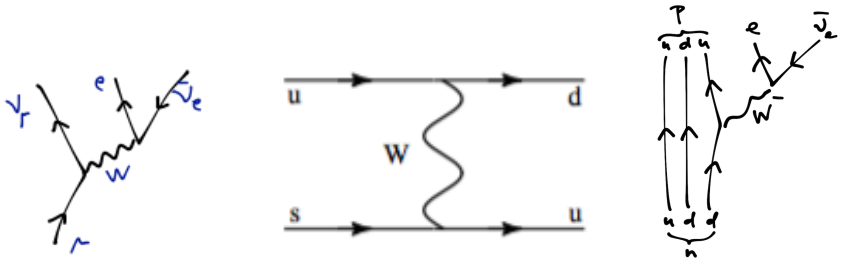
- Hydrodynamics [Kovtun]
- Fermi liquid theory [J. Polchinski, R. Shankar, *Rev. Mod. Phys.* **66** (1994) 129]
- chiral perturbation theory [D. B. Kaplan, §4]
- heavy quark effective theory [D. B. Kaplan, §1.3, Manohar and Wise, *Heavy Quark Physics*]
- random surface growth (KPZ) [Zee, chapter VI]
- color superconductors [D. B. Kaplan, §5]
- gravitational radiation from binary mergers [Goldberger, Rothstein, Porto]
- soft collinear effective theory [Becher, Stewart]
- magnets [Zee, chapter VI.5, hep-ph/9311264v1]
- effective field theory of cosmological inflation [Senatore et al, Cheung et al, Porto]
- effective field theory of dark matter direct detection [Fitzpatrick et al]

There are many others, the length of this list was limited by how long I was willing to spend digging up references.

9.3 Fermi theory of Weak Interactions

[from §5 of A. Manohar's EFT lectures] As a first example, let's think about part of the Standard Model.

$$L_{EW} \ni -\frac{1}{2}\left(\partial_\mu W^+_\nu - \partial_\nu W^+_\mu\right)\left(\partial^\mu W^{-\nu} - \partial^\nu W^{-\mu}\right) + M_W W^+_\mu W^{-\mu} \qquad (9.5)$$
$$-\frac{\mathbf{i}g}{\sqrt{2}}\bar{\psi}_i\gamma^\mu P_L\psi_j W^+_\mu V_{ij} \quad + \text{terms involving } Z \text{ bosons}$$



Some things intermediate, off-shell W bosons can do: μ decay, $\Delta S = 1$ processes, neutron decay

If we are asking questions with external momenta less than M_W, we can integrate out W and make our lives simpler:

$$\delta S_{eff} \sim \left(\frac{\mathbf{i}g}{\sqrt{2}}\right)^2 V_{ij}V^\star_{k\ell}\int \mathrm{d}^D p \frac{-\mathbf{i}g_{\mu\nu}}{p^2 - M_W^2}\left(\bar{\psi}_i\gamma^\mu P_L\psi_j\right)(p)\left(\bar{\psi}_k\gamma^\nu P_L\psi_\ell\right)(-p)$$

(I am lying a little bit about the W propagator in that I am not explicitly projecting out the fourth polarization with the negative residue. Also hidden in my notation is the fact that the W carries electric charge, so the charges of $\bar{\psi}_i$ and ψ_j in (9.5) must differ by one.) This is non-local at scales $p \gtrsim M_W$ (recall the discussion of the previous subsection). But for $p^2 \ll M_W^2$,

$$\frac{1}{p^2 - M_W^2} \overset{p^2 \ll M_W^2}{\simeq} -\frac{1}{M_W^2}\left(1 + \underbrace{\frac{p^2}{M_W^2} + \frac{p^4}{M_W^4} + \ldots}_{\text{derivative couplings}}\right) \qquad (9.6)$$

$$S_F = -\frac{4G_F}{\sqrt{2}}V_{ij}V^\star_{kl}\int d^4x \left(\bar{\psi}_i\gamma^\mu P_L\psi_j\right)(x)\left(\bar{\psi}_k\gamma_\mu P_L\psi_\ell\right)(x) + \mathcal{O}\left(\frac{1}{M_W^2}\right) + \text{kinetic terms for fermions} \qquad (9.7)$$

where $G_F/\sqrt{2} \equiv \frac{g^2}{8M_W^2}$ is the Fermi coupling. We can use this (Fermi's) theory to compute the amplitudes above, and it is much simpler than the full electroweak theory

(for example I don't have to lie about the form of the propagator of the W-boson like I did above). It was discovered first and used quite effectively long before the existence of Ws was suspected.

On the other hand, this theory is *not the same* as the electroweak theory; for example it is not renormalizable, while the EW theory is. Its point in life is to help facilitate the expansion in $1/M_W$. There is something about the expression (9.7) that should make you nervous, namely the big red 1 in the $1/M_W^2$ corrections: what makes up the dimensions? This becomes an issue when we ask about ...

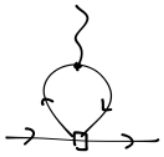
9.4 Loops in EFT

I skipped this subsection in lecture. Skip to §9.5. Suppose we try to define the Fermi theory S_F with a euclidean momentum cutoff $|k_E| < \Lambda$. We expect that we'll have to set $\Lambda \sim M_W$. A simple example which shows that this is problematic is to ask about radiative corrections in the 4-Fermi theory to the coupling between the fermions and the Z (or the photon).

We are just trying to estimate the magnitude of this correction, so don't worry about the factors and the gamma matrices:

$$\sim I \equiv \underbrace{\frac{1}{M_W^2}}_{\propto G_F} \underbrace{\int^\Lambda \text{đ}^4k \frac{1}{k}\frac{1}{k} \text{tr}\,(\gamma...)}_{\sim \int^\Lambda k\text{d}k \sim \Lambda^2 \sim M_W^2} \sim \mathcal{O}(1).$$

Even worse, consider what happens if we use the vertex coming from the $\left(\frac{p^2}{M_W^2}\right)^\ell$ correction in (9.6)



$$\sim I_\ell \equiv \frac{1}{M_W^2} \int^\Lambda \text{đ}^4k \frac{1}{k^2} \left(\frac{k^2}{M_W^2}\right)^\ell \sim \mathcal{O}(1)$$

– it's also unsuppressed by powers of ... well, anything. This is a problem.

Fix: A way to fix this is to use a "mass-independent subtraction scheme", such as dimensional regularization and minimal subtraction ($\overline{\text{MS}}$). The crucial feature is that the dimensionful cutoff parameter appears only inside logarithms ($\log \mu$), and not as free-standing powers (μ^2).

With such a scheme, we'd get instead

$$I \sim \frac{m^2}{M_W^2} \log \mu \qquad I_\ell \sim \left(\frac{m^2}{M_W^2}\right)^{\ell+1} \log \mu$$

where m is some mass scale *other* than the RG scale μ (like a fermion mass parameter, or an external momentum, or a dynamical scale like Λ_{QCD}).

We will give a more detailed example next. The point is that in a mass-independent scheme, the regulator doesn't produce new dimensionful things that can cancel out the factors of M_W in the denominator. It respects the 'power counting': if you see 2ℓ powers of $1/M_W$ in the coefficient of some term in the action, that's how many powers will suppress its contributions to amplitudes. This means that the EFT is like a renormalizable theory *at each order in the expansion* (here in $1/M_W$), in that there is only a finite number of allowed vertices that contribute at each order (counterterms for which need to be fixed by a renormalization condition). The insatiable appetite for counterterms is still insatiable, but it eats only a finite number at each order in the expansion. Eventually you'll get to an order in the expansion that's too small to care about, at which point the EFT will have eaten only a finite number of counterterms.

There is a price for these wonderful features of mass-independent schemes, which has two aspects:

- Heavy particles (of mass m) don't decouple when $\mu < m$. For example, in a mass-independent scheme for a gauge theory, heavy charged particles contribute to the beta function for the gauge coupling even at $\mu \ll m$.
- Perturbation theory will break down at *low* energies, when $\mu < m$; in the example just mentioned this happens because the coupling keeps running.

We will show both these properties very explicitly in the next subsection. The solution of both these problems is to integrate out the heavy particles by hand at $\mu = m$, and make a new EFT for $\mu < m$ which simply omits that field. Processes for which we should set $\mu < m$ don't have enough energy to make the heavy particles in external states anyway. (For some situations where you should still worry about them, see Aneesh Manohar's notes linked above.)

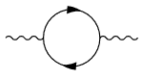
9.4.1 Comparison of schemes, case study

The case study we will make is the contribution of a charged fermion of mass m to the running of the QED gauge coupling.

Recall that the QED Lagrangian is

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \bar{\psi}\left(\mathbf{i}\not{D} - m\right)\psi$$

with $D_\mu = \partial_\mu - \mathbf{i}eA_\mu$. By redefining the field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ by a constant factor we can move around where the e appears, *i.e.* by writing $\tilde{A} = eA$, we can make the gauge kinetic term look like $\frac{1}{4e^2}\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu}$. This means that the charge renormalization can be seen *either* in the vacuum polarization, the correction to the photon propagator:



. I will call this diagram $\mathbf{i}\Pi_{\mu\nu}$.

So the information about the running of the coupling is encoded in the gauge field two-point function:

$$\Pi_{\mu\nu} \equiv \langle A_\mu(p)A_\nu(q)\rangle = \left(p_\mu p_\nu - p^2 g_{\mu\nu}\right)\not{\delta}(p+q)\Pi(p^2) \ .$$

The factor $P_{\mu\nu} \equiv p_\mu p_\nu - p^2 g_{\mu\nu}$ is guaranteed to be the polarization structure by the gauge invariance Ward identity: $p^\mu \langle A_\mu(p)A_\nu(q)\rangle = 0$. That is: $p^\mu P_{\mu\nu} = 0$, and there is no other symmetric tensor made from p^μ which satisfies this. This determines the correlator up to a function of p^2, which we have called $\Pi(p^2)$.

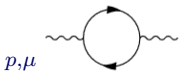
The choice of scheme shows up in our choice of renormalization condition to impose on $\Pi(p^2)$:

Mass-dependent scheme: subtract the value of the graph at $p^2 = -M^2$ (a very off-shell, euclidean, momentum). That is, we impose a renormalization condition which says

$$\Pi(p^2 = -M^2) \stackrel{!}{=} 1 \tag{9.8}$$

(which is the tree-level answer with the normalization above).

The contribution of a fermion of mass m and charge e is (factoring out the momentum-conserving delta function):



$$= -\int \mathrm{d}^D k \,\mathrm{tr}\left((-\mathbf{i}e\gamma^\mu)\frac{-\mathbf{i}\left(\not{k}+m\right)}{k^2-m^2}(-\mathbf{i}e\gamma^\nu)\frac{-\mathbf{i}\left(\not{p}+\not{k}+m\right)}{(p+k)^2-m^2}\right)$$

The minus sign out front is from the fermion loop. Some boiling, which you can find in Peskin (page 247) or Zee (§III.7), reduces this to something manageable. The steps involved are: (1) a trick to combine the denominators, like the Feynman trick $\frac{1}{AB} = \int_0^1 dx \left(\frac{1}{(1-x)A+xB}\right)^2$. (2) some Dirac algebra, to turn the numerator into a polynomial

in k, p. As Zee says, our job in this course is not to train to be professional integrators. The result of this boiling can be written

$$\mathbf{i}\Pi^{\mu\nu} = -e^2 \int \mathrm{đ}^D \ell \int_0^1 dx \frac{N^{\mu\nu}}{\left(\ell^2 - \Delta\right)^2}$$

with $\ell = k + xp$ is a new integration variable, $\Delta \equiv m^2 - x(1-x)p^2$, and the numerator is

$$N^{\mu\nu} = 2\ell^\mu \ell^\nu - g^{\mu\nu}\ell^2 - 2x(1-x)p^\mu p^\nu + g^{\mu\nu}\left(m^2 + x(1-x)p^2\right) + \text{terms linear in } \ell^\mu \quad .$$

In dim reg, the one-loop vacuum polarization correction satisfies the gauge invaraince Ward identity $\Pi^{\mu\nu} = P^{\mu\nu}\delta\Pi_2$ (unlike the euclidean momentum cutoff which is not gauge invariant). A peek at the tables of dim reg integrals shows that $\delta\Pi_2$ is:

$$\begin{aligned}
\delta\Pi_2(p^2) &\overset{\text{Peskin p. 252}}{=} -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2 - D/2)}{\Delta^{2-D/2}} \bar{\mu}^\epsilon \\
&\overset{D\to 4}{=} -\frac{e^2}{2\pi^2} \int_0^1 dx x(1-x) \left(\frac{2}{\epsilon} - \log\left(\frac{\Delta}{\mu^2}\right)\right)
\end{aligned} \tag{9.9}$$

where we have introduced the heralded μ:

$$\mu^2 \equiv 4\pi\bar{\mu}^2 e^{-\gamma_E}$$

where γ_E is the Euler-Mascheroni constant. In the second line of (9.9), we expanded the Γ-function about $D = 4$; there are other singularities at other integer dimensions.

Mass-dependent scheme: Now back to our discussion of schemes. I remind you that in a mass-independent scheme, we demand that the counterterm cancels $\delta\Pi_2$ when we set the external momentum to $p^2 = -M^2$, so that the whole contribution at order e^2 is :

$$0 \overset{(9.8)!}{=} \Pi_2^{(M)}(p^2 = -M^2) = \underbrace{\delta_{F^2}^{(M)}}_{\text{counterterm coefficient for } \frac{1}{4}F_{\mu\nu}F^{\mu\nu}} + \delta\Pi_2$$

$$\implies \Pi_2^{(M)}(p^2) = \frac{e^2}{2\pi^2} \int dx x(1-x) \log\left(\frac{m^2 - x(1-x)p^2}{m^2 + x(1-x)M^2}\right).$$

Notice that the μs go away in this scheme.

Mass-Independent scheme: This is to be contrasted with what we get in a mass-independent scheme, such as $\overline{\text{MS}}$, in which Π is defined by the rule that we *subtract the* $1/\epsilon$ *pole*. This means that the counterterm is

$$\delta_{F^2}^{(\overline{\text{MS}})} = -\frac{e^2}{2\pi^2}\frac{2}{\epsilon} \underbrace{\int_0^1 dx x(1-x)}_{=1/6}.$$

(Confession: I don't know how to state this in terms of a simple renormalization condition on Π_2. Also: the bar in $\overline{\text{MS}}$ refers to the (not so important) distinction between $\bar{\mu}$ and μ.) The resulting vacuum polarization function is

$$\Pi_2^{(\overline{\text{MS}})}(p^2) = \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\log\left(\frac{m^2 - x(1-x)p^2}{\mu^2}\right).$$

Next we will talk about beta functions, and verify the claim above about the failure of decoupling. First let me say some words about what is failing. What is failing – the price we are paying for our power counting – is the basic principle of the RG, namely that physics at low energies shouldn't care about physics at high energies, except for small corrections to couplings. An informal version of this statement is: you don't need to know about nuclear physics to make toast. A more formal version is the *Appelquist-Carazzone Decoupling Theorem*, which I will not state (Phys. Rev. D11, 28565 (1975)). So it's something we must and will fix.

Beta functions. $\boxed{M}$: First in the mass-dependent scheme. Demanding that physics is independent of our made-up RG scale, we find

$$0 = M\frac{d}{dM}\Pi_2^{(M)}(p^2) = \left(M\frac{\partial}{\partial M} + \beta_e^{(M)} e\frac{\partial}{\partial e}\right)\Pi_2^{(M)}(p^2) = \left(M\frac{\partial}{\partial M} + \beta_e^{(M)} \underbrace{\cdot 2}_{\text{to this order}}\right)\Pi_2^{(M)}(p^2)$$

where I made the high-energy physics definition of the beta function[64]:

$$\beta_e^{(M)} \equiv \frac{1}{e}\left(M\partial_M e\right) = -\frac{\partial_\ell e}{e}, \quad M \equiv e^{-\ell} M_0 \ .$$

Here ℓ is the RG time again, it grows toward the IR. So we find

$$\beta_e^{(M)} = -\frac{1}{2}\left(\frac{e^2}{2\pi}\right)\int_0^1 dx x(1-x)\left(\frac{-2M^2 x(1-x)}{m^2 + M^2 x(1-x)}\right) + \mathcal{O}(e^3)$$
$$\begin{cases} \overset{m\ll M}{\simeq} \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x) = \frac{e^2}{12\pi^2} \\ \\ \overset{m\gg M}{\simeq} \frac{e^2}{2\pi^2}\int_0^1 dx x(1-x)\frac{M^2 x(1-x)}{m^2} = \frac{e^2}{60\pi^2}\frac{M^2}{m^2} \end{cases} . \tag{9.10}$$

$$\boxed{\overline{\text{MS}}}: \quad 0 = \mu\frac{d}{d\mu}\Pi_2^{(\overline{\text{MS}})}(p^2) = \left(\mu\frac{\partial}{\partial\mu} + \beta_e^{(\overline{\text{MS}})} e\frac{\partial}{\partial e}\right)\Pi_2^{(\overline{\text{MS}})}(p^2) = \left(\mu\frac{\partial}{\partial\mu} + \beta_e^{(\overline{\text{MS}})} \underbrace{\cdot 2}_{\text{to this order}}\right)\Pi_2^{(\overline{\text{MS}})}(p^2)$$

[64] I've defined these beta functions to be dimensionless, *i.e.* they are $\partial_{\log M}\log(g)$; this convention is not universally used.

$$\implies \beta_e^{(\overline{\text{MS}})} = -\frac{1}{2}\frac{e^2}{2\pi^2}\underbrace{\int_0^1 dx x(1-x)}_{=1/6}\underbrace{\mu\partial_\mu \log\frac{m^2 - p^2x(1-x)}{\mu^2}}_{=-2}$$
$$= \frac{e^2}{12\pi^2}. \tag{9.11}$$

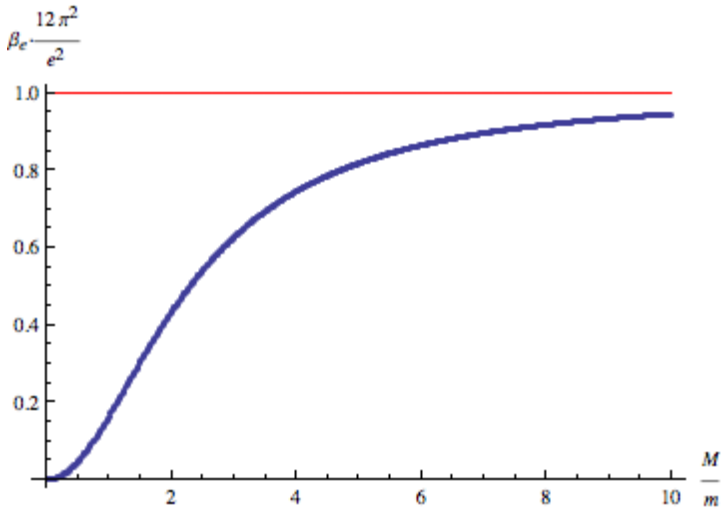
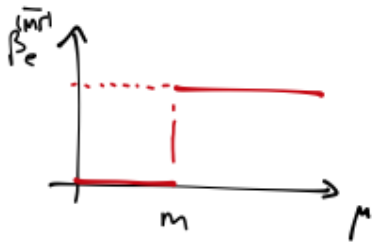


Figure 4: The blue curve is the mass-dependent-scheme beta function; at scales $M \ll m$, the mass of the heavy fermion, the fermion sensibly stops screening the charge. The red line is the $\overline{\text{MS}}$ beta function, which is just a constant, pinned at the UV value.

Also, the $\overline{\text{MS}}$ vacuum polarization behaves for small external momenta like

$$\Pi_2(p^2 \ll m^2) \simeq -\frac{e^2}{2\pi^2}\int_0^1 dx x(1-x) \underbrace{\log\frac{m^2}{\mu^2}}_{\gg 1, \text{for } \mu \ll m! \text{ bad!}}$$

As I mentioned, the resolution of both these problems is simply to define a new EFT for $\mu < m$ which omits the heavy field. Then the strong coupling problem goes away and the heavy fields do decouple. The price is that we have to do this by hand, and the beta function jumps at $\mu = m$; the coupling is continuous, though.



9.5 The Standard Model as an EFT.

The Standard Model. [Schwartz, §29]

	$L = \begin{pmatrix}\nu_L \\ e_L\end{pmatrix}$	e_R	ν_R	$Q = \begin{pmatrix}u_L \\ d_L\end{pmatrix}$	u_R	d_R	H
$\mathsf{SU}(3)$	-	-	-	□	□	□	-
$\mathsf{SU}(2)$	□	-	-	□	-	-	□
$\mathsf{U}(1)_Y$	$-\frac{1}{2}$	-1	0	$\frac{1}{6}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{2}$

Table 1: The Standard Model fields and their quantum numbers under the gauge group. □ indicates fundamental representation, - indicates singlet. Except for the Higgs, each column is copied three times; each copy is called a *generation*. Except for the Higgs all the matter fields are Weyl fermions of the indicated handedness. Gauge fields as implied by the gauge groups. (Some people might leave out the right-handed neutrino, ν_R.)

Whence the values of the charges under the $\mathsf{U}(1)$ ("hypercharge")? The condition $Y_L + 3Y_Q = 0$ (where Y is the hypercharge) is required by anomaly cancellation. This implies that electrons and protons $p = \epsilon_{ijk} u_i u_j d_k$ have exactly opposite charges of the same magnitude.

The Lagrangian is just all the terms which are invariant under the gauge group $\mathsf{SU}(3) \times \mathsf{SU}(2) \times \mathsf{U}(1)$ with dimension less than or equal to four – all renormalizable terms. This includes a potential for the Higgs, $V(|H|) = m_H^2 |H|^2 + \lambda |H|^4$, where it turns out that $m_H^2 \leq 0$. The resulting Higgs vacuum expectation value breaks the Electroweak part of the gauge group

$$\mathsf{SU}(2) \times \mathsf{U}(1)_Y \overset{\langle H \rangle}{\rightsquigarrow} \mathsf{U}(1)_{EM}.$$

The broken gauge bosons get masses from the Higgs kinetic term

$$|D_\mu H|^2| \Big|_{H=\begin{pmatrix}0 \\ v/\sqrt{2}\end{pmatrix}} \quad \text{with} \quad D_\mu H = \left(\partial_\mu - \mathbf{i} g W_\mu^a \tau^a - \frac{1}{2}\mathbf{i} g' Y_\mu\right) H$$

where Y_μ is the hypercharge gauge boson, and $W^a, a = 1, 2, 3$ are the $\mathsf{SU}(2)$ gauge bosons. There are two massive W-bosons with electric charge ± 1 (as described in §9.3), with $M_W = \frac{vg}{2}$. The photon and Z boson are the linear combinations of Y and W^3 which diagonalize the remaining mass terms:

$$\begin{pmatrix}A_\mu \\ Z_\mu\end{pmatrix} = \begin{pmatrix}\cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w\end{pmatrix} \begin{pmatrix}W_\mu^3 \\ Y_\mu\end{pmatrix}.$$

Here $\tan\theta_w \equiv \frac{g'}{g}$ defines the Weinberg angle. The masses are $M_\gamma = 0$ and $M_Z = \frac{M_W}{\cos\theta_w} < M_W$.

Fermion masses come from (dimension-four) Yukawa couplings

$$\mathcal{L}_{\text{Yukawa}} = -Y_{ij}^\ell \bar{L}_i H e_R^j - Y_{ij}^u \bar{Q}^i H d_R^j - Y_{ij}^d \bar{Q}^i \left(\mathbf{i}\tau^2 H^\star\right) u_R^j + h.c.$$

The contortion with the τ^2 is required to make a hypercharge invariant. Plugging in the Higgs vev to *e.g.* the lepton terms gives $-m_e \bar{e}_L e_R + h.c.$ with $m_e = y_e v/\sqrt{2}$. There's lots of drama about the matrices Y which can mix the generations. the mass for the ν_R (which maybe could not exist – it doesn't have any charges at all) you'll figure out on the homework.

Here is a useful mnemonic for remembering the table of quantum numbers (possibly it is more than that): There are larger simple Lie groups that contain the SM gauge group as subgroups:

$$\begin{array}{ccccc} \mathsf{SU}(3)\times\mathsf{SU}(2)\times\mathsf{U}(1)_Y & \subset & \mathsf{SU}(5) & \subset & \mathsf{SO}(10) \\ \text{one generation} & = & 10\oplus\bar{5}\oplus 1 & = & 16 \end{array}$$

The singlet of $\mathsf{SU}(5)$ is the right-handed neutrino, but if we include it, one generation is an irreducible (spinor) representation of $\mathsf{SO}(10)$. This idea is called grand unification. It is easy to imagine that the gauge group is actually the larger groups on the right, and another instance of the Higgs mechanism accomplishes the breaking down to the Standard Model. (The running of the respective gauge couplings go in the right direction with approximately the right rate to unify to a single value at $M_{GUT} \sim 10^{16} GeV$.) Notice that this idea means leptons and quarks are in the same representations – they can turn into each other. This predicts that the proton should not be perfectly stable. Next we'll say more about this.

Beyond the Standard Model with EFT. At what energy does the Standard Model stop working? Because of the annoying feature of renormalizibility, it doesn't tell us. However, we have experimental evidence against a cutoff on the Standard Model (SM) at energies less than something like 10 TeV. The evidence I have in mind is the absence of interactions of the form

$$\delta L = \frac{1}{M^2}\left(\bar{\psi} A \psi\right)\cdot\left(\bar{\psi} B \psi\right)$$

(where ψ represent various SM fermion fields and A, B can be various gamma and flavor matrices) with $M \lesssim 10$ TeV. Notice that I am talking now about interactions *other* than the electroweak interactions, which as we've just discussed, for energies above $M_W \sim 80$GeV cannot be treated as contact interactions – you can see the Ws propagate!

If such operators were present, we would have found different answers for experiments at LEP. But such operators would be present if we consider new physics in addition to the Standard Model (in most ways of doing it) at energies less than 10 TeV. For example, many interesting ways of coupling in new particles with masses that make them accessible at the LHC would have generated such operators.

A little more explicitly: the Standard Model Lagrangian L_0 contains all the renormalizable (*i.e.* engineering dimension ≤ 4) operators that you can make from its fields (though the coefficients of the dimension 4 operators do vary through quite a large range, and the coefficients of the two relevant operators – namely the identity operator which has dimension zero, and the Higgs mass, which has engineering dimension two, are strangely small, and so is the QCD θ angle).

To understand what lies beyond the Standard Model, we can use our knowledge that whatever it is, it is probably heavy (it could also just be very weakly coupled, which is a different story), with some intrinsic scale $\Lambda_{\rm new}$, so we can integrate it out and include its effects by corrections to the Standard Model:

$$L = L_0 + \frac{1}{\Lambda_{\rm new}}\mathcal{O}^{(5)} + \frac{1}{\Lambda_{\rm new}^2}\sum_i c_i \mathcal{O}_i^{(6)} + \cdots$$

where the $\mathcal{O}$s are made of SM fields, and have the indicated engineering dimensions, and preserve the necessary symmetries of the SM (Lorentz symmetry and gauge invariance).

In fact there is only one kind of operator of dimension 5 meeting these demands:

$$\mathcal{O}^{(5)} = c_5 \epsilon_{ij} \left(\bar{L}^c\right)^i H^j \epsilon_{kl} L^k H^l$$

where $H^i = (h^+, h^0)^i$ is the $\mathrm{SU}(2)_{EW}$ Higgs doublet and $L^i = (\nu_L, e_L)^i$ is an $\mathrm{SU}(2)_{EW}$ doublet of left-handed leptons, and $\bar{L}^c \equiv L^T C$ where C is the charge conjugation matrix. (I say 'kind of operator' because we can have various flavor matrices in here.) On the problem set you get to see from whence such an operator might arise, and what it does if you plug in the higgs vev $\langle H \rangle = (0, v)$. This term violates lepton number symmetry ($L \to e^{\mathbf{i}\alpha_L} L, Q \to Q, H \to H$).

At dimension 6, there are operators that directly violate baryon number, such as

$$\epsilon_{\alpha\beta\gamma} (\bar{u}_R)^c_\alpha (u_R)_\beta \, (\bar{u}_R)^c_\gamma \, e_R.$$

You should read the above tangle of symbols as '$qqq\ell$' – it turns three quarks into a lepton. The epsilon tensor makes a color SU(3) singlet; this thing ϵqqq has the quantum numbers of a baryon, such as the proton and neutron. The long lifetime of the proton (you can feel it in your bones – see Zee p. 413) then directly constrains the scale of new physics appearing in front of this operator.

Two more comments about this:

- If we didn't know about the Standard Model, (but after we knew about QM and GR and EFT (the last of which people didn't know before the SM for some reason)) we should have made the estimate that dimension-5 Planck-scale-suppressed

operators like $\frac{1}{M_{\rm Planck}} p\mathcal{O}$ would cause proton decay (into whatever $\mathcal{O}$ makes). This predicts $\Gamma_p \sim \frac{m_p^3}{M_{\rm Planck}^2} \sim 10^{-13} s^{-1}$ which is *not* consistent with our bodies not glowing. Actually it is a remarkable fact that there are no gauge-invariant operators made of SM fields of dimension less than 6 that violate baryon number symmetry ($L \to L, Q \to e^{i\alpha_B} Q, H \to H$). This is an *emergent* symmetry, expected to be violated by the UV completion.

- Surely nothing can prevent $\Delta L \sim \left(\frac{1}{M_{\rm Planck}}\right)^2 qqq\ell$. Happily, this is consistent with the observed proton lifetime.

There are $\sim 10^2$ dimension 6 operators that preserve baryon number, and therefore are not as tightly constrained[65]. (Those that induce flavor-changing processes in the SM are more highly constrained and must have $\Lambda_{\rm new} > 10^4$ TeV.) Two such operators are considered equivalent if they differ by something which vanishes by the tree-level SM equations of motion. This is the right thing to do, even for off-shell calculations (like green's functions and for fields running in loops). You know this from a previous problem set: the EOM are true as operator equations – Ward identities resulting from being free to change integration variables in the path integral[66].

[65] Recently, humans have gotten better at counting these operators. See this paper.

[66] There are a few meaningful subtleties here, as you might expect if you recall that the Ward identity is only true up to contact terms. The measure in the path integral can produce a Jacobian which renormalizes some of the couplings; the changes in source terms will drop out of S-matrix elements (recall our discussion of changing field variables in the Consequences of Unitarity section.) but can change the form of Green's functions. For more information on the use of eom to eliminate redundant operators in EFT, see Arzt, hep-ph/9304230 and Georgi, "On-Shell EFT".

9.6 The color of the sky

[from hep-ph/9606222 and nucl-th/0510023] Why is the sky blue? Basically, it's because the blue light from the sun scatters in the atmosphere more than the red light, and you (I hope) only look at the scattered light.

Here is an understanding of this fact using the EFT logic. Consider the scattering of photons off atoms (in a gas) at low energies. Low energy means that the photon does not have enough energy to probe the substructure of the atom – it can't excite the electrons or the nuclei. This means that the atom is just a particle, with some mass M.

The dofs are just the photon field and the field that creates an atom.

The symmetries are Lorentz invariance and charge conjugation invariance and parity. We'll use the usual redundant description of the photon which has also gauge invariance.

The cutoff is the energy ΔE that it takes to excite atomic energy levels we've left out of the discussion. We allow no inelastic scattering. This means we require

$$E_\gamma \ll \Delta E \sim \frac{\alpha}{a_0} \ll a_0^{-1} \ll M_{\rm atom} \tag{9.12}$$

Because of this separation of scales, we can also ignore the recoil of the atom, and treat it as infinitely heavy.

Since there are no charged objects in sight – atoms are neutral – gauge invariance means the Lagrangian can depend on the field strength $F_{\mu\nu}$. Let's call the field which destroys an atom with velocity v ϕ_v. $v^\mu v_\mu = 1$ and $v_\mu = (1,0,0,0)_\mu$ in the atom's rest frame. The (Lorentz-singlet) Lagrangian can depend on v^μ. We can write a Lagrangian for the free atoms as

$$L_{\rm atom} = \phi_v^\dagger \mathbf{i} v^\mu \partial_\mu \phi_v \quad .$$

This action is related by a boost to the statement that the atom at rest has zero energy – in the rest frame of the atom, the eom is just $\partial_t \phi_{v=(1,\vec{0})} = 0$. (If we didn't define the zero of energy to be at the rest mass, there would be an additional term $M_{\rm atom}\phi_v^\dagger \phi_v$.) Notice that the kinetic term $\phi_v^\dagger \frac{\vec{\nabla}^2}{2M_{\rm atom}} \phi_v$ is a very small correction given our hierarchy of scales (9.12).

So the Lagrangian density is

$$L_{\rm Maxwell}[A] + L_{\rm atom}[\phi_v] + L_{\rm int}[A, \phi_v]$$

and we must determine $L_{\rm int}$. It is made from local, Hermitian, gauge-invariant, Lorentz invariant operators we can construct out of $\phi_v, F_{\mu\nu}, v_\mu, \partial_\mu$ (it can only depend on $F_{\mu\nu} =$

$\partial_\mu A_\nu - \partial_\nu A_\mu$, and not A_μ directly, by gauge invariance, because the atom, and hence ϕ_v, is neutral.). It should actually only depend on the combination $\phi_v^\dagger \phi_v$ since we will not create and destroy atoms. (Notice that we didn't have to specify the statistics of the atoms or ϕ_v.) Therefore

$$L_{\text{int}} = c_1 \phi_v^\dagger \phi_v F_{\mu\nu} F^{\mu\nu} + c_2 \phi_v^\dagger \phi_v v^\sigma F_{\sigma\mu} v_\lambda F^{\lambda\mu} + c_3 \phi_v^\dagger \phi_v \left(v^\lambda \partial_\lambda\right) F_{\mu\nu} F^{\mu\nu} + \dots$$

$\dots$ indicates terms with more derivatives and more powers of velocity (*i.e.* an expansion in $\partial \cdot v$). Which are the most important terms at low energies? Demanding that the Maxwell term dominate, we get the power counting rules (so time and space should scale the same way):

$$[\partial_\mu] = 1, \quad [F_{\mu\nu}] = 2$$

This then implies $[\phi_v] = 3/2, [v] = 0$ and therefore

$$[c_1] = [c_2] = -3, [c_3] = -4 \ .$$

Terms with more partials are more irrelevant.

What makes up these dimensions? They must come from the length scales that we have integrated out to get this description – the size of the atom $a_0 \sim \alpha m_e$ and the energy gap between the ground state and the electronic excited states $\Delta E \sim \alpha^2 m_e$. For $E_\gamma \ll \Delta E, a_0^{-1}$, we can just keep the two leading terms.

In the rest frame of the atom, these two leading terms $c_{1,2}$ represent just the scattering of E and B respectively. To determine their coefficients one would have to do a matching calculation to a more complete theory (compute transition rates in a theory that does include extra energy levels of the atom). But a reasonable guess is just that the scale of new physics (in this case atomic physics) makes up the dimensions: $c_1 \simeq c_2 \simeq a_0^3$. (In fact the magnetic term c_2 comes with extra factor of v/c which suppresses it.) The scattering cross section then goes like $\sigma \sim c_i^2 \sim a_0^6$; dimensional analysis ($[\sigma] = -2$ is an area, $[a_0^6] = -6$) then tells us that we have to make up four powers with the only other scale around:

$$\sigma \propto E_\gamma^4 a_0^6.$$

(The factor of E_γ^2 in the amplitude arises from $\vec{E} \propto \partial_t \vec{A}$.) Blue light, which has about twice the energy of red light, is therefore scattered 16 times as much.

The leading term that we left out is the one with coefficient c_3. The size of this coefficient determines when our approximations break down. We might expect this to come from the next smallest of our neglected scales, namely ΔE. That is, we expect

$$\sigma \propto E_\gamma^4 a_0^6 \left(1 + \mathcal{O}\left(\frac{E_\gamma}{\Delta E}\right)\right).$$

The ratio in the correction terms is appreciable for UV light.

9.7 Pions

[Schwartz §28.1] Below the scale of electroweak symmetry breaking, we can forget the W and Z bosons. Besides the 4-Fermi interactions, the remaining drama is QCD and electromagnetism:

$$\mathcal{L}_{QCD_2} = -\frac{1}{4}F_{\mu\nu}^2 + \mathbf{i}\sum_{\alpha=L,R}\sum_f \bar{q}_{\alpha f}\not{D}q_{\alpha f} - \bar{q}Mq.$$

Here f is a sum over quark flavors, which includes the electroweak doublets, u and d. Let's focus on just these two lightest flavors, u and d. We can diagonalize the mass matrix by a field redefinition (this is what makes the CKM matrix meaningful): $M = \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix}$. If it were the case that $m_u = m_d$, we would have *isospin* symmetry

$$\begin{pmatrix} u \\ d \end{pmatrix} \to U \begin{pmatrix} u \\ d \end{pmatrix}, \quad U \in \mathsf{SU}(N_f = 2).$$

If, further, there were no masses $m = 0$, then L and R decouple and we also have chiral symmetry, $q \to e^{\mathbf{i}\gamma_5\alpha}q$, *i.e.*

$$q_L \to Vq_L, q_R \to V^{-1}q_R, \; V \in \mathsf{SU}(N_f = 2).$$

Why do I restrict to $\mathsf{SU}(2)$ and not $\mathsf{U}(2)$? The central bit of the axial symmetry $\mathsf{U}(1)_A$ is anomalous – its divergence is proportional to the *gluon* theta term operator $F \wedge F$, which has all kinds of nonzero matrix elements. It's not a symmetry (see Peskin page 673 for more detail). The missing non-Goldstone boson is called the η'. The central bit of the vectorlike transformation $q \to e^{\mathbf{i}\alpha}q$ is *baryon number*, B. (Actually this is anomalous under the full electroweak symmetry, but $B - L$ is not).

The groundstate of QCD is mysterious, because of infrared slavery. Here's one piece of input from experiment and numerical simulation. Apparently it is the case that in the groundstate

$$\langle \bar{q}_f q_f \rangle = V^3 \tag{9.13}$$

independent of flavor f. This condensate spontaneously breaks

$$\mathsf{SU}(2)_L \times \mathsf{SU}(2)_R \to \mathsf{SU}(2)_{\text{isospin}}, \tag{9.14}$$

the diagonal combination. $\begin{pmatrix} u \\ d \end{pmatrix}$ is a doublet. Since $p = u_\alpha u_\beta d_\gamma \epsilon_{\alpha\beta\gamma}, n = u_\alpha d_\beta d_\gamma \epsilon_{\alpha\beta\gamma}$, this means that $\begin{pmatrix} p \\ n \end{pmatrix}$ is also a doublet. This symmetry is (explicitly) weakly broken by

the difference of the masses $m_d = 4.7\text{MeV} \neq m_u = 2.15\text{MeV}$ and by the electromagnetic interactions, since $q_d = -1/3 \neq q_u = 2/3$.

This symmetry-breaking structure enormously constrains the dynamics of the color singlets which are the low-energy excitations above the QCD vacuum (hadrons). Let us use the EFT strategy. We know that the degrees of freedom must include (pseudo-)Goldstone bosons for the symmetry breaking (9.14) ('pseudo' because of the weak explicit breaking).

Effective field theory. Since QCD is strongly coupled in this regime, let's use the knowing-the-answer trick: the low energy theory must include some fields which represent the breaking of the symmetry (9.14). One way to do this is to introduce a field Σ which transforms like

$$\mathsf{SU}(2)_L \times \mathsf{SU}(2)_R : \Sigma \to g_L \Sigma g_R^\dagger, \quad \Sigma^\dagger \to g_R \Sigma^\dagger g_L^\dagger$$

(this will be called a *linear* sigma model, because Σ transforms linearly) – we have in mind $\bar{q}_\alpha q_\beta \sim \Sigma_{\alpha\beta}$. We can make singlets (hence an action) out of $\Sigma_{ij}\Sigma^\dagger_{ji} = \text{tr}\Sigma\Sigma^\dagger \equiv |\Sigma|^2$:

$$\mathcal{L} = |\partial_\mu \Sigma|^2 + m^2 |\Sigma|^2 - \frac{\lambda}{4}|\Sigma|^4 + \cdots \tag{9.15}$$

which is designed to have a minimum at $\langle\Sigma\rangle = \frac{V}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$, with $V = 2m/\sqrt{\lambda}$ (here V is from (9.13)), which preserves $\mathsf{SU}(2)_{\text{isospin}}$ (under which $\Sigma \to g\Sigma g^\dagger$). We can parametrize the fluctuations about this configuration as

$$\Sigma(x) = \frac{V + \sigma(x)}{\sqrt{2}} e^{\frac{2\mathbf{i}\pi^a(x)\tau^a}{F_\pi}}$$

where $F_\pi = V = \frac{2m}{\sqrt{\lambda}}$ is be chosen to give $\pi^a(x)$ canonical kinetic terms. The π^a parametrize the directions of field space in which the potential is flat (like the field θ in the discussion of the Mexican hat in §6.1). Under $g_{L/R} = e^{\mathbf{i}\theta_{L/R}\tau^a}$, the pion field transforms as

$$\pi^a \to \pi^a + \underbrace{\frac{F_\pi}{2}\left(\theta_L^a - \theta_R^a\right)}_{\text{nonlinear realization of } \mathsf{SU}(2)_{\text{axial}}} - \underbrace{\frac{1}{2} f^{abc}\left(\theta_L^a + \theta_R^a\right)\pi^c}_{\text{linear realiz'n (adj rep) of } \mathsf{SU}(2)_{\text{isospin}}} .$$

The fields $\pi^\pm, \pi^0$ create pions, they transform in the adjoint representation of the diagonal $\mathsf{SU}(2)_{\text{isospin}}$, and they *shift* under the broken symmetry. This shift symmetry forbids mass terms π^2. The radial excitation σ, on the other hand, is a fiction which we've introduced in (9.15), and which has no excuse to stick around at low energies (and does not). We can put it out of its misery by taking $m \to \infty, \lambda \to \infty$ fixing F_π.

In the limit, the useful field to use is

$$U(x) \equiv \frac{\sqrt{2}}{V}\Sigma(x)|_{\sigma=0} = e^{\frac{2\mathrm{i}\pi^a \tau^a}{F_\pi}}$$

which is unitary $UU^\dagger = U^\dagger U = \mathbb{1}$. This last identity means that all terms in an action for U require derivatives, so (again) no mass for π. The most general Lagrangian for U can be written as an expansion in derivatives, and is called the *chiral Lagrangian*:

$$\mathcal{L}_\chi = \frac{F_\pi^2}{4}\mathrm{tr} D_\mu U\, D^\mu U^\dagger + L_1 \mathrm{tr}\left(D_\mu U\, D^\mu U^\dagger\right)^2 + L_2 \mathrm{tr} D_\mu U\, D_\nu U^\dagger \mathrm{tr} D^\nu U^\dagger D_\mu U + L_3 \mathrm{tr} D_\mu U\, D^\mu U^\dagger D_\nu U\, D^\nu U^\dagger + \cdots \tag{9.16}$$

In terms of π, the leading term expands into

$$L_\chi = \frac{1}{2}\partial_\mu \pi^a \partial^\mu \pi^a + \frac{1}{F_\pi^2}\left(-\frac{1}{3}\pi^0\pi^0 D_\mu \pi^+ D^\mu \pi^- + \cdots\right) + \frac{1}{F_\pi^4}\left(\frac{1}{18}\left(\pi^-\pi^+\right)^2 D_\mu \pi^0 D^\mu \pi^0 + \cdots\right)$$

This fixes the relative coefficients of many irrelevant interactions, all with two derivatives, suppressed by powers of F_π. The expansion of the L_i terms have four derivatives, and are therefore suppressed by further powers of E/F_π, the promised small parameter of this EFT.

Pion masses. The pions aren't actually massless: $m_{\pi^\pm} \sim 140\mathrm{MeV}$. In terms of quarks, one source for such a thing is the quark mass term $\mathcal{L}_{QCD} \ni \bar{q}Mq$. This explicitly breaks the isospin symmetry if the eigenvalues of M aren't equal. But an *invariance* of $\mathcal{L}_{QCD}$ is

$$q_{L/R} \to g_{L/R} q_{L/R},\ M \to g_L M g_R^\dagger. \tag{9.17}$$

Think of M as a background field (such a thing is sometimes called a *spurion*). If M were an actual dynamical field, then (9.17) would be a symmetry. In the effective action which summarizes all the drama of strong-coupling QCD in terms of pions, the field M must still be there, and if we transform it as in (9.17), it should still be an invariance. Maybe we're going to do the path integral over M later. (This 'spurion' trick has applications all over physics.)

So the chiral lagrangian $\mathcal{L}_\chi$ should depend on M and (9.17) should be an invariance. This determines

$$\Delta\mathcal{L}_\chi = \frac{V^3}{2}\mathrm{tr}\left(MU + M^\dagger U^\dagger\right) + \cdots = V^3(m_u + m_d) - \frac{V^3}{2F_\pi^2}(m_u + m_d)\sum_a \pi_a^2 + \mathcal{O}(\pi^2).$$

The coefficient V^3 is chosen so that the first term matches $\langle \bar{q}Mq\rangle = V^3(m_u + m_d)$. The second term then gives

$$m_\pi^2 \simeq \frac{V^3}{F_\pi^2}(m_u + m_d)$$

which is called the Gell-Mann Oakes Renner relation.

Electroweak interactions. You may have noticed that I used covariant-looking Ds in (9.16). That's because the $\mathsf{SU}(2)_L$ symmetry we've been speaking about is actually gauged by W^a_μ. (The electroweak gauge boson kinetic terms are in the $\cdots$ of (9.16).) Recall that

$$\mathcal{L}_{\text{Weak}} \ni gW^a_\mu \left(\underbrace{J^a_\mu - J^{5a}_\mu}_{\text{'V' - 'A'}} \right) = gW^a_\mu \left(V_{ij} \bar{Q}_i \gamma^\mu \frac{1-\gamma^5}{2} \tau^a Q_j + \bar{L}_i \gamma^\mu \tau^a \frac{1-\gamma^5}{2} L_i \right)$$

where $Q_1 = \begin{pmatrix} u \\ d \end{pmatrix}, L_1 = \begin{pmatrix} e \\ \nu_e \end{pmatrix}$ are doublets of $\mathsf{SU}(2)_L$.

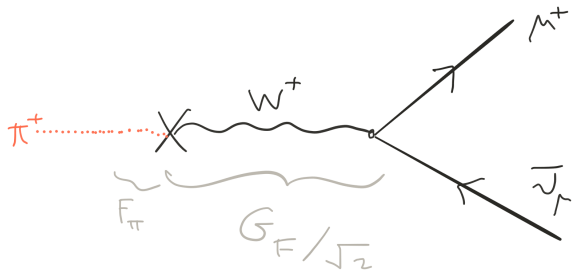
Now, in equations, the statement "a pion is a Goldstone boson for the axial $\mathsf{SU}(2)$" is:

$$\langle 0 | \, J^{5a}_\mu(x) \, | \pi^b(p) \rangle = \mathbf{i} p_\mu F_\pi e^{-\mathbf{i} p \cdot x} \delta^{ab}$$

where the state $|\pi^b(p)\rangle$ is a one-pion state of momentum p. If the vacuum were invariant under the symmetry transformation generated by J_μ, the BHS would vanish. The momentum dependence implements the fact that a global rotation ($p_\mu = 0$) does not change the energy. Contracting the BHS with p^μ and using current conservation (ignoring the explicit breaking just mentioned) would give $0 = p^2 F_\pi^2 = m_\pi^2 F_\pi^2$, a massless dispersion for the pions.

Combining the previous two paragraphs, we see that the following process can happen

$$\pi \overset{\text{Goldstone}}{\to} J^5_\mu \overset{\text{electroweak interaction}}{\to} \text{leptons}$$



(9.18)

and in fact is responsible for the dominant decay channel of charged pions. (Time goes from left to right in these diagrams, sorry.)

$$\mathcal{M}(\pi^+ \to \mu^+ \nu_\mu) = \frac{G_F}{\sqrt{2}} F_\pi p^\mu \bar{v}_{\nu_\mu} \gamma^\mu (1-\gamma^5) u_\mu$$

where the Fermi constant $G_F \sim 10^{-5} GeV^{-2}$ (known from *e.g.* $\mu^- \to e^- \bar{\nu}_e \nu_\mu$) is a good way to parametrize the Weak interaction amplitude. Squaring this and integrating

over two-body phase space gives the decay rate

$$\Gamma(\pi^+ \to \mu^+ \nu_\mu) = \frac{G_F^2 F_\pi^2}{4\pi} m_\pi m_\mu^2 \left(1 - \frac{m_\mu^2}{m_\pi^2}\right)^2 .$$

(You can see from the answer why the decay to muons is more important than the decay to electrons, since $m_\mu/m_e \sim 200$. This is called *helicity suppression* – the decay of the helicity-zero π^- into back-to-back spin-half particles by the weak interaction (which only produces L particles and R antiparticles) can't happen if helicity is conserved – the mass term is required to flip the e_L into an e_R.) This contributes most of $\tau_{\pi^+} = \Gamma^{-1} = 2.6 \cdot 10^{-8} s$.

Knowing further the mass of the muon $m_\mu = 106\text{MeV}$ then determines $F_\pi = 92\text{MeV}$ which fixes the leading terms in the chiral Lagrangian. This is why F_π is called the *pion decay constant*. This gives a huge set of predictions for *e.g.* pion scattering $\pi^0\pi^0 \to \pi^+\pi^-$ cross sections.

Note that the neutral pion can decay by an anomaly into two photons:

$$q_\mu \langle p_1, p_2| J_\mu^{5,a=3}(q) |0\rangle = -c \frac{e^2}{4\pi^2} \epsilon^{\nu\lambda\alpha\beta} p_1^\nu \epsilon_1^\lambda p_2^\alpha \epsilon_2^\beta$$

where $\langle p_1, p_2|$ is a state with two photons of polarizations $\epsilon_{1,2}$. I know this because it is a matrix element of the $J_e J_e J_{\mathsf{SU}(2)-\text{axial}}$ anomaly,

$$\partial_\mu J^{\mu 5 a} = -\frac{e^2}{16\pi^2} \epsilon^{\nu\lambda\alpha\beta} F_{\nu\lambda} F_{\alpha\beta} \text{tr}\left(\tau^a Q^2\right)$$

where $Q = \begin{pmatrix} 2/3 & 0 \\ 0 & -1/3 \end{pmatrix}$ is the quark charge matrix. Comments: (1) this symmetry acts by $u \to e^{\mathbf{i}\theta\gamma^5} u, d \to e^{-\mathbf{i}\theta\gamma^5}$, and is *not* the same as the anomalous $\mathsf{U}(1)_A$ (which does $q_i \to e^{\mathbf{i}\theta\gamma^5} q_i$ for every flavor), and it's also not the same as isospin $u \to e^{\mathbf{i}\theta} u, d \to e^{-\mathbf{i}\theta}$, which is not chiral, and not spontaneously broken. Confusing! (2) The rate of π^0 decay (known since the 1940s) gives a measurement of the number of colors of QCD! (3) This effect can be encoded in the Lagrangian for the pions by a term

$$L \ni N_c \frac{e^2}{16\pi^2} \pi^0 \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma},$$

where $N_c = 3$ is the number of colors. The effective field theory consistently realizes the anomalies of the microscopic theory. This is an example of *'t Hooft anomaly matching*, a principle which can be used, for example, to prove that QCD *must* spontaneously break the $\mathsf{SU}(3)_L \times \mathsf{SU}(3)_R$ chiral symmetry (see Schwartz §30.6).

Wait – what $\mathsf{SU}(3)$?

SU(3) and baryons. The strange quark mass is also pretty small $m_s \sim 95\text{MeV}$, and $\langle \bar{s}s \rangle \sim V^3$. This means the approximate invariance and symmetry breaking pattern is actually $\mathsf{SU}(3)_L \times \mathsf{SU}(3)_R \to \mathsf{SU}(3)_{\text{diag}}$, meaning that there are $16 - 8 = 8$ pseudo NGBs. Besides $\pi^{\pm,0}$, the others are the kaons $K^{\pm,0}$ and η. It's still only the $\mathsf{SU}(2)_L$ that's gauged.

We can also include baryons $B = \epsilon_{\alpha\beta\gamma} q_\alpha q_\beta q_\gamma$. Since $q = (u, d, s) \in 3$ of the flavor $\mathsf{SU}(3)$, the baryons are in the representation

$$\begin{aligned} 3 \otimes 3 \otimes 3 &= (6 \oplus \bar{3}) \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1 \\ \square \otimes \square \otimes \square &= (\square\square \oplus \begin{smallmatrix}\square\\\square\end{smallmatrix}) \otimes \square = \square\square\square \oplus \begin{smallmatrix}\square\square\\\square\end{smallmatrix} \oplus \begin{smallmatrix}\square\square\\\square\end{smallmatrix} \oplus \begin{smallmatrix}\square\\\square\\\square\end{smallmatrix} \end{aligned} \tag{9.19}$$

The proton and neutron are in one of the octets. This point of view brought some order (and some predictions) to the otherwise-bewildering zoo of hadrons.

Returning to the two-flavor $\mathsf{SU}(2)$ approximation, we can include the nucleons $N_{L/R} = \begin{pmatrix} p \\ n \end{pmatrix}_{L/R}$ and couple them to pions by the symmetric coupling

$$\mathcal{L} \ni \lambda_{NN\pi} \bar{N}_L \Sigma N_R.$$

The expectation value for Σ gives a nucleon mass: $m_N = \lambda_{NN\pi} F_\pi$, where $\lambda_{NN\pi}$ can be measured by scattering. This is a cheap version of the *Goldberger-Treiman relation*; for a better one see Peskin pp. 670-672.

WZW terms in the chiral Lagrangian. Finally, I would be remiss not to mention that the chiral Lagrangian must be supplemented by WZW terms to have the correct realization of symmetries (in order to encode all the effects of anomalies, and in order to violate $\pi \to -\pi$ which is not a symmetry of QCD).

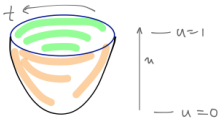
The chiral Lagrangian governs a *non-linear sigma model* (NLσM)– a QFT whose fields are maps from spacetime into some *target space*. In this case the target space is the coset space G/H, where G is the full symmetry group ($\mathsf{SU}(N_f)_L \times \mathsf{SU}(N_f)_R$) and H is the *unbroken* subgroup $\mathsf{SU}(N_f)_{\text{diagonal}}$. We can parametrize this space by $U = e^{\mathbf{i}\pi^a T^a \frac{2}{F_\pi}}$ where the T^a includes only generators of the broken part of the group, so the π^a are coordinates on G/H.

A WZW term is a term which we can sometimes add to a NLσM action; it is defined by the fact that it is symmetric under some group G, but isn't the integral of a symmetric local Lagrangian density in D dimensions. Making it manifestly symmetric requires the introduction of a fictitious extra dimension. This has the dramatic and surprising consequence that its coefficient is quantized.

To get the idea, consider a model in $D = 0 + 1$ where the field variable $\check{n}$ takes values on the unit sphere S^2, $1 = \sum_{a=1,2,3} \check{n}_a^2$. This is a special case of a coset space $\mathsf{G}/\mathsf{H} = \mathsf{SU}(2)/\mathsf{U}(1)$.

In order to write the WZW term in a manifestly symmetric way (under the $\mathsf{SO}(3)$ of rotations of the sphere, we have to extend the field into a (possibly fictitious) extra dimension whose coordinate is u.
We do this in such a way that the real system lives at $u = 1$:

$$\check{n}(t, u = 1) \equiv \check{n}(t), \quad \check{n}(t, u = 0) \equiv (0, 0, 1)$$



it goes to the north pole at the other end of the extra dimension for all t. Consider periodic boundary conditions in time $\check{n}(2\pi) = \check{n}(0)$. Then this means that the full space is really a disk with the origin at $u = 0$, and the boundary at $u = 1$. Call this disk B, its boundary $\partial B = \mathcal{M}$ is the real spacetime (here a circle).

We can write the WZW term in terms of the S^2-valued field $\check{n}^{1,2,3}$ as

$$\mathcal{W}_1[\check{n}] = \frac{2\pi}{\Omega_2} \int_{B_2} \check{n}^a \mathrm{d}\check{n}^b \wedge \mathrm{d}\check{n}^c \epsilon_{abc} = \frac{1}{4\pi} \int_{\mathcal{M}} \mathrm{d}t \, (1 - \cos\theta) \, \partial_t \phi.$$

The integrand here is the volume element of the image of a chunk of spacetime in the target S^2. If we integrate over the union of two balls with cancelling boundaries $B_2 \cup \bar{B}_2$, we get an integer multiple of 2π (the integer is the winding number of the map).

The coefficient k of $\mathcal{W}_1$ in the action $\Delta S[\check{n}] = k\mathcal{W}_1[\check{n}]$ must be an integer since B_1 and $\bar{B}_1$ give equally good definitions of $\mathcal{W}_2$, which differ by $2\pi k$. So this ambiguity will not affect the path integral if $k \in \mathbb{Z}$.

The generalization to a group-valued variable U in any dimension is of the form

$$\mathcal{W}_D = c \int_{B_{D+1}} \mathrm{tr} \underbrace{U^{-1}dU \wedge U^{-1}dU \wedge \cdots \wedge U^{-1}dU}_{D+1 \text{ of these}}.$$

Such terms are interesting when $\pi_{D+1}(\mathcal{M})$ is nontrivial, where $\mathcal{M}$ is the space where the fields live (the target space), that is, there are maps from S^{D+1} to $\mathcal{M}$ which cannot be smoothly deformed to the trivial map where every point in the base space goes to

the same point in the target. The variation of $\mathcal{W}_D$ with respect to U is (for even D)[67]:

$$\delta\mathcal{W}_D = (D+1)c\int_{B_{D+1}} \mathrm{tr}\left\{\left(U^{-1}dU\right)^D \underbrace{\delta\left(U^{-1}dU\right)}_{=U^{-1}d(\delta UU^{-1})U}\right\} \tag{9.21}$$

$$= (D+1)c\int_{B_{D+1}} \mathrm{tr}\left\{\left(dUU^{-1}\right)^D d(\delta UU^{-1})\right\} \tag{9.22}$$

$$= (D+1)c\int_{B_{D+1}} d\mathrm{tr}\left\{\left(U^{-1}dU\right)^D U^{-1}\delta U\right\} \tag{9.23}$$

$$\overset{\text{Stokes}}{=} (D+1)c\int_{\mathcal{M}} \mathrm{tr}\left\{\left(U^{-1}dU\right)^D U^{-1}\delta U\right\}$$

which only depends on the field configuration on $\mathcal{M}$, not on the extension to B_{D+1}. Again there can be topologically distinct ways to make the extension; demanding that they always give the same answer determines c in terms of volumes of spheres (so that $c\int_{S^{D+1}}\mathrm{tr}(U^{-1}dU)^{D+1}\in\mathbb{Z}$ is the winding number), and the coefficient must be an integer. (In $D=4$, we have $c=\frac{\mathbf{i}}{240\pi^2}$.)

This WZW term is *less* topological than the theta term we discussed above, in the sense that it affects the equations of motion for $\check{n}(t)$. The variation of $\mathcal{W}$ *is* local in D dimensions. The following table gives a comparison between theta terms and WZW terms for a field theory in D spacetime dimensions, on a spacetime M_D:

[67] Why do I restrict to even D?

$$\mathrm{tr}\left(U^{-1}dU\right)^{D+1} = \epsilon^{\mu_1\cdots\mu_{D+1}}\mathrm{tr}\left(U^{-1}\partial_{\mu_1}U\cdots U^{-1}\partial_{\mu_{D+1}}\right)$$

but $\epsilon^{\mu_1\cdots\mu_{D+1}} = -(-1)^{D+1}\epsilon^{\mu_{D+1}\mu_1\cdots\mu_D}$ so $\mathcal{W}_D = (-1)^D\mathcal{W}_D$ vanishes in odd dimensions. The step from (9.22) to (9.23) also relies on this fact. Using $1 = U^{-1}U$ and hence $0 = \delta(U^{-1}U) = d(U^{-1}U)$, so that

$$dU^{-1} = -U^{-1}dUU^{-1}, \tag{9.20}$$

the term by which (9.22) and (9.23) differ is

$$\mathrm{tr}\left\{\left(d\left(U^{-1}dU\right)^D\right)\delta UU^{-1}\right\}$$

$$\overset{\text{product rule}}{=} \mathrm{tr}\left\{\left(dU^{-1}\wedge dU\wedge\left(U^{-1}dU\right)^{D-1} - \left(U^{-1}dU\wedge dU^{-1}\wedge dU\wedge\left(U^{-1}dU\right)^{D-2}+\cdots\right)\delta UU^{-1}\right\}$$

$$\overset{(9.20)}{=} -\mathrm{tr}\left\{\left(U^{-1}dU\wedge U^{-1}dU\wedge\left(U^{-1}dU\right)^{D-1} - U^{-1}dUU^{-1}\wedge dUU^{-1}\wedge dU\wedge\left(U^{-1}dU\right)^{D-2}+\cdots\right)\delta UU^{-1}\right\}$$

$$= \mathrm{tr}\left\{\underbrace{(1-1+1-1...)}_{D-1\text{ of these}}\left(U^{-1}dU\right)^{D-1}\delta UU^{-1}\right\} \overset{D-1\text{ even}}{=} 0.$$

See Weinberg, vol 2, §23.4 for more.

theta term	WZW term
$\mathcal{H} = \int_{M_D} h$	$\mathcal{W}_D = \int_{B_{D+1}} w, \quad \partial B_{D+1} = M_D$
$h = \mathrm{d}q$	$w = \mathrm{d}v$
Doesn't affect EOM	Affects EOM
Invisible in perturbation theory	Appears in perturbation theory, *e.g.* in beta functions
$\mathcal{H} \in \mathbb{Z}$ for M_D closed	Coefficient of $\mathcal{W} \in \mathbb{Z}$ in order for path integral to be well-defined.

Pion physics is the context where these terms were first discovered, and where it was realized that their coefficients are quantized. In particular the coefficient of the WZW term $W_4[U]$ here is N_c, the number of colors, as Witten shows by explicitly coupling to electromagnetism, and finding the term that encodes $\pi^0 \to \gamma\gamma$. One dramatic consequence here is that the chiral Lagrangian (with some higher-derivative terms) has a topological soliton solution (the skyrmion) which is a *fermion* if the number of colors of QCD is odd. The field configuration $U(x,t)$ is constant in time and approaches the vacuum at infinity, so we can regard it as a map $U : (\text{space} \cup \infty \sim S^d) \to G/H$, where G is the full symmetry group and H is the unbroken subgroup, so G/H is the space of Goldstones (in the chiral Lagrangian, $G/H = \mathsf{SU}(3) \times \mathsf{SU}(3)/\mathsf{SU}(3)_{\text{preserved}} \simeq \mathsf{SU}(3)_{\text{broken}}$). The configuration is topological in the sense that as a map from $S^3 \to G/H$, it cannot be smoothly deformed to the trivial map – it represents a nontrivial element of $\pi_3(G/H)$. Its nontriviality is witnessed by a winding number, which can be written as the integral of a local density. In fact, the baryon number of this configuration comes from the anomalous (WZW) contribution to the baryon number current $B_\mu = \frac{\epsilon_{\mu\nu\alpha\beta}}{24\pi^2} \mathrm{tr} U^{-1}\partial_\nu U U^{-1}\partial_\alpha U U^{-1}\partial_\beta U$ whose conserved charge $\int_{\text{space}} B_0$ is exactly the winding number of the map from space (plus the point at infinity) to the space of goldstones. And finally this object a fermion because the WZW term evaluates to π on a spacetime trajectory where the soliton makes a 2π rotation. So this object is a fermionic particle which carries baryon number. It also carries isospin. It's a nucleon! Above we added nucleon fields to the chiral Lagrangian, but we actually didn't need to – they were already there as solitonic excitations. Note that the *size* of the soliton (the region of space over which the fields vary) is determined by the higher-derivative terms in the chiral lagrangian, so we shouldn't take too seriously the substructure of the proton predicted by this picture. But it doesn't do too badly.

I should also mention that WZW terms are important in the study of interacting

spin systems, for example in our understanding the dependence on the s of Heisenberg spin-s chains, and in phase transitions beyond the Landau-Ginzburg (symmetry-breaking) paradigm (*i.e.* deconfined quantum criticality). This is a subject for Physics 215C.

9.8 Superconductors

Recall from §6.1 our effective (Landau-Ginzburg) description of superconductors which reproduces the Meissner effect, the Abelian Higgs model:

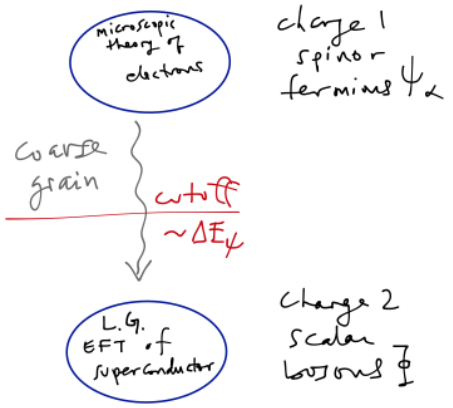
$$\mathcal{F} = \frac{1}{4}F_{ij}F_{ij} + |D_i\Phi|^2 + a|\Phi|^2 + \frac{1}{2}b|\Phi|^4 + ... \tag{9.24}$$

with $D_i\Phi \equiv (\partial_i - 2e\mathbf{i}A_i)\,\Phi$.

I want to make two more comments about this:

Symmetry breaking by fluctuations (Coleman-Weinberg). [Zee problem IV.6.9.] What happens near the transition, when $a = 0$ in (9.24)? Quantum fluctuations can lead to symmetry breaking.

New IR dofs. A feature of this example that I want you to notice: the microscopic description of real superconductor involves electrons – charge $1e$ spinor fermions, created by some fermionic operator ψ_α, $\alpha = \uparrow, \downarrow$.



We are describing the low-energy physics of a system of electrons in terms of a bosonic field, which (in simple 's-wave' superconductors) is roughly related to the electron field by

$$\Phi \sim \psi_\alpha \psi_\beta \epsilon^{\alpha\beta} \; ; \tag{9.25}$$

Φ is called a Cooper pair field. At least, the charges and the spins and the statistics work out. The details of this relationship are not the important point I wanted to emphasize. Rather I wanted to emphasize the dramatic difference in the correct choice of variables between the UV description (spinor fermions) and the IR description (scalar bosons). One reason that this is possible is that it costs a large energy to make a fermionic excitation of the superconductor. This can be understood roughly as follows: The microscopic theory of the electrons looks something like (ignoring the coupling to electromagnetism for now)

$$S[\psi] = S_2[\psi] + \int \mathrm{d}t\mathrm{d}^d x \; u\psi^\dagger\psi\psi^\dagger\psi + h.c. \tag{9.26}$$

where

$$S_2 = \int \mathrm{d}t \int \text{đ}^d k \psi_k^\dagger \left(\mathbf{i}\partial_t - \epsilon(k) \right) \psi_k .$$

Spin is important here so that $\psi_\uparrow^\dagger \psi_\uparrow \psi_\downarrow^\dagger \psi_\downarrow$ is nonzero. A mean field theory description of the condensation of Cooper pairs (9.25) is obtained by replacing the quartic term in (9.26) by expectation values:

$$\begin{aligned} S_{MFT}[\psi] &= S_2[\psi] - \int \mathrm{d}t \mathrm{d}^d x \; u \langle \psi\psi \rangle \, \psi^\dagger \psi^\dagger + h.c. \\ &= S_2[\psi] - \int \mathrm{d}t \mathrm{d}^d x \; u \Phi \psi^\dagger \psi^\dagger + h.c. \end{aligned} \tag{9.27}$$

So an expectation value for Φ is a mass for the fermions. It is a funny kind of symmetry-breaking mass, but if you diagonalize the quadratic operator in (9.27) (actually it is done below) you will find that it costs an energy of order $\Delta E_\psi = u \langle \Phi \rangle$ to excite a fermion. That's the cutoff on the LG EFT.

A general lesson from this example is: the useful degrees of freedom at low energies can be very different from the microscopic dofs.

9.8.1 Lightning discussion of BCS.

I am sure that some of you are nervous about the step from $S[\psi]$ to $S_{MFT}[\psi]$ above. To make ourselves feel better about it, I will say a few more words about the steps from the microscopic model of electrons (9.26) to the LG theory of Cooper pairs (these steps were taken by Bardeen, Cooper and Schreiffer (BCS)).

First recall the *Hubbard-Stratonovich transformation* aka *completing the square*. In 0+0 dimensional field theory:

$$e^{-\mathbf{i}u x^4} = \frac{1}{\sqrt{\mathbf{i}\pi u}} \int_{-\infty}^{\infty} \mathrm{d}\sigma \; e^{-\frac{1}{\mathbf{i}u}\sigma^2 - 2\mathbf{i}x^2\sigma} \; . \tag{9.28}$$

At the cost of introducing an extra field σ, we turn a quartic term in x into a quadratic term in x. The RHS of (9.28) is gaussian in x and we know how to integrate it over x. (The version with $\mathbf{i}$ is relevant for the real-time integral.) Notice the weird extra factor of $\mathbf{i}$ lurking in (9.28). This can be understood as arising because we are trying to use a scalar field σ, to mediate a repulsive interaction (which it is, for positive u) (see Zee p. 193, 2nd Ed).

Actually, we'll need a complex H-S field:

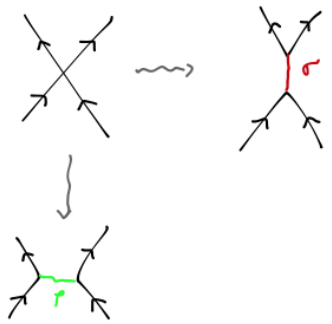
$$e^{-\mathbf{i}u x^2 \bar{x}^2} = \frac{1}{\mathbf{i}\pi u} \int_{\mathbb{C}} \mathrm{d}^2\sigma \; e^{-\frac{1}{\mathbf{i}u}|\sigma|^2 - \mathbf{i}x^2\bar{\sigma} - \mathbf{i}\bar{x}^2\sigma} \; , \tag{9.29}$$

where $\int_{\mathbb{C}} \mathrm{d}^2\sigma ... \equiv \int_{-\infty}^{\infty} d\mathrm{Re}\,\sigma \int_{-\infty}^{\infty} d\mathrm{Im}\,\sigma ...$ (The field-independent prefactor is, as usual, not important for path integrals.)

We can use a field theory generalization of (9.29) to 'decouple' the 4-fermion interaction in (9.26):

$$Z = \int [D\psi D\psi^\dagger] e^{\mathbf{i}S[\psi]} = \int [D\psi D\psi^\dagger D\sigma D\sigma^\dagger] e^{\mathbf{i}S_2[\psi] + \mathbf{i}\int d^D x (\bar\sigma\psi\psi + h.c.) - \int \mathrm{d}^D x \frac{|\sigma|^2(x)}{\mathbf{i}u}} \quad . \tag{9.30}$$

The point of this is that now the fermion integral is gaussian. At the saddle point of the σ integral (which is exact because it is gaussian), σ is the Cooper pair field, $\sigma_{\mathrm{saddle}} = u\psi\psi$.



Notice that we made a choice here about in which 'channel' to make the decoupling – we could have instead introduces a different auxiliary field ρ and written $S[\rho,\psi] = \int \rho\psi^\dagger\psi + \int \frac{\rho^2}{2u}$, which would break up the 4-fermion interaction in the t-channel (as an interaction of the fermion density $\psi^\dagger\psi$) instead of the s (BCS) channel (as an interaction of Cooper pairs ψ^2). At this stage both are correct, but they lead to different mean-field approximations below. That the BCS mean field theory wins is a consequence of the RG.

How can you resist doing the fermion integral in (9.30)? Let's study the case where the single-fermion dispersion is $\epsilon(k) = \frac{\vec{k}^2}{2m} - \mu$.

$$I_\psi[\sigma] \equiv \int [D\psi D\psi^\dagger] e^{\mathbf{i}\int \mathrm{d}t\mathrm{d}^d x \left(\psi^\dagger\left(\mathbf{i}\partial_t - \frac{\nabla^2}{2m} - \mu\right)\psi + \psi\bar\sigma\psi + \bar\psi\bar\psi\sigma\right)}$$

The action here can be written as the integral of

$$L = \begin{pmatrix}\bar\psi & \psi\end{pmatrix} \begin{pmatrix} \mathbf{i}\partial_t - \epsilon(-\mathbf{i}\nabla) & \sigma \\ \bar\sigma & -(\mathbf{i}\partial_t - \epsilon(-\mathbf{i}\nabla))\end{pmatrix} \begin{pmatrix}\psi \\ \bar\psi\end{pmatrix} \equiv \begin{pmatrix}\bar\psi & \psi\end{pmatrix} M \begin{pmatrix}\psi \\ \bar\psi\end{pmatrix}$$

so the integral is

$$I_\psi[\sigma] = \det M = e^{\mathrm{tr}\log M(\sigma)}.$$

If σ is constant (which will lower the energy), the matrix M is diagonal in momentum space, and the integral remaining to be done is

$$\int [D\sigma D\sigma^\dagger] e^{-\int \mathrm{d}^D x \frac{|\sigma(x)|^2}{2\mathbf{i}u} + \int \mathrm{d}^D k \log\left(\omega^2 - \epsilon_k^2 - |\sigma|^2\right)}.$$

It is often possible to do this integral by saddle point. This can justified, for example, by the largeness of the volume of the Fermi surface, $\{k|\epsilon(k) = \mu\}$, or by large N number

of species of fermions. The result is an equation which determines σ, which as we saw earlier determines the fermion gap.

$$0 = \frac{\delta \text{exponent}}{\delta \bar{\sigma}} = \mathbf{i}\frac{\sigma}{2u} + \int đ\omega đ^d k \frac{2\sigma}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + \mathbf{i}\epsilon} \ .$$

We can do the frequency integral by residues:

$$\int đ\omega \frac{1}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + \mathbf{i}\epsilon} = \frac{1}{2\pi} 2\pi \mathbf{i} \frac{1}{2\sqrt{\epsilon_k^2 + |\sigma|^2}}.$$

The resulting equation is naturally called the *gap equation*:

$$1 = -2u \int đ^d p' \frac{1}{\sqrt{\epsilon(p')^2 + |\sigma|^2}} \tag{9.31}$$

which you can imagine solving self-consistently for σ. Plugging back into the action (9.30) says that σ determines the energy cost to have electrons around; more precisely, σ is the energy required to break a Cooper pair.

Comments:

- Notice that a solution of (9.31) requires $u < 0$, an *attractive* interaction. Superconductivity happens because the u that appears here is not the bare interaction between electrons, which is certainly repulsive (and long-ranged). This is where the phonons come in in the BCS discussion.

- If we hadn't restricted to a delta-function 4-fermion interaction $u(p, p') = u_0$ at the outset, we would have found a more general equation like

$$\sigma(\vec{p}) = -\frac{1}{2} \int đ^d p' \frac{u(p, p')\sigma(\vec{p}')}{\sqrt{\epsilon(p')^2 + |\sigma(p')|^2}} \ .$$

- A conservative perspective on the preceding calculation is that we have made a variational ansatz for the groundstate wavefunction, and the equation we solve for σ is minimizing the variational energy – finding the best wavefunction within the ansatz.

- I haven't included here effects of the fluctuations of the fermions. In fact, they make the four-fermion interaction which leads to Cooper pairing marginally *relevant*. This breaks the degeneracy in deciding how to split up the $\psi\psi\psi^\dagger\psi^\dagger$ into *e.g.* $\psi\psi\sigma$ or $\psi^\dagger\psi\rho$. BCS wins. This is explained beautifully in Polchinski, lecture 2, and R. Shankar. If there were time, I would summarize the EFT framework for understanding this in §9.9.

- I've tried to give the most efficient introduction I could here. I left out any possibility of k-dependence or spin dependence of the interactions or the pair field, and I've conflated the pair field with the gap. In particular, I've been sloppy about the dependence on k of σ above.
- You can study a very closely related manipulation on the problem set, in an example (the Gross-Neveu model) where the saddle point is justified by large N.

9.9 Effective field theory of Fermi surfaces

[Polchinski, lecture 2, and R. Shankar] Electrically conducting solids are a remarkable phenomenon. An arbitrarily small electric field $\vec{E}$ leads to a nonzero current $\vec{j} = \sigma\vec{E}$. This means that there must be gapless modes with energies much less than the natural cutoff scale in the problem.

Scales involved: The Planck scale of solid state physics (made by the logic by which Planck made his quantum gravity energy scale, namely by making a quantity with dimensions of energy out of the available constants) is

$$E_0 = \frac{1}{2}\frac{e^4 m}{\hbar^2} = \frac{1}{2}\frac{e^2}{a_0} \sim 13\text{eV}$$

(where $m \equiv m_e$ is the electron mass and the factor of 2 is an abuse of outside information) which is the energy scale of *chemistry*. Chemistry is to solids as the melting of spacetime is to particle physics. There are other scales involved however. In particular a solid involves a lattice of nuclei, each with $M \gg m$ (approximately the proton mass). So m/M is a useful small parameter which controls the coupling between the electrons and the lattice vibrations. Also, the actual speed of light $c \gg v_F$ can generally also be treated as ∞ to first approximation. v_F/c suppresses spin orbit couplings (though large atomic numbers enhance them: $\lambda_{\text{SO}} \propto Zv_F/c$).

Let us attempt to construct a Wilsonian-natural effective field theory of this phenomenon. The answer is called Landau Fermi Liquid Theory. What are the right low-energy degrees of freedom? Let's make a guess that they are like electrons – fermions with spin and electric charge. They will not have exactly the properties of free electrons, since they must incorporate the effects of interactions with all their friends. The 'dressed' electrons are called quasielectrons, or more generally quasiparticles.

Given the strong interactions between so many particles, why should the dofs have anything at all to do with electrons? Landau's motivation for this description (which

is not always correct) is that we can imagine starting from the free theory and adiabatically turning up the interactions. If we don't encounter any phase transition along the way, we can follow each state of the free theory, and use the same labels in the interacting theory.

We will show that there is a nearly-RG-stable fixed point describing gapless quasi-electrons. Notice that we are not trying to match this description directly to some microscopic lattice model of a solid; rather we will do bottom-up effective field theory.

Having guessed the necessary dofs, let's try to write an action for them consistent with the symmetries. A good starting point is the free theory:

$$S_{\text{free}}[\psi] = \int dt\, \mathrm{d}^d p \left(\mathbf{i}\psi^\dagger_\sigma(p)\partial_t \psi_\sigma(p) - (\epsilon(p) - \epsilon_F)\, \psi^\dagger_\sigma(p)\psi_\sigma(p) \right)$$

where σ is a spin index, ϵ_F is the Fermi energy (zero-temperature chemical potential), and $\epsilon(p)$ is the single-particle dispersion relation. For non-interacting non-relativistic electrons in free space, we have $\epsilon(p) = \frac{p^2}{2m}$. It will be useful to leave this as a general function of p. [68] [69]

The groundstate is the filled Fermi sea:

$$|\text{gs}\rangle = \prod_{p|\epsilon(p)<\epsilon_F} \psi^\dagger_p |0\rangle\,, \quad \psi_p |0\rangle = 0, \quad \forall p.$$

(If you don't like continuous products, put the system in a box so that p is a discrete label.) The Fermi surface is the set of points in momentum space at the boundary of the filled states:

$$\text{FS} \equiv \{p|\epsilon(p) = \epsilon_F\}.$$

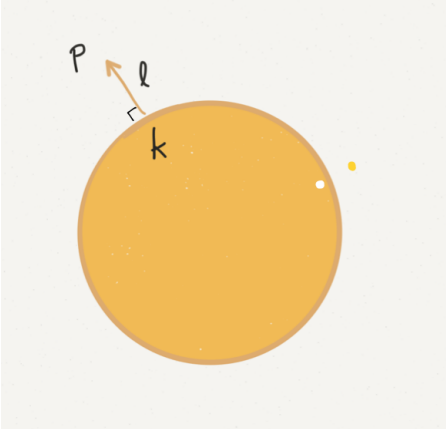
The low-lying excitations are made by adding an electron just above the FS or removing an electron (creating a hole) just below.

We would like to define a scaling transformation which focuses on the low-energy excitations. We scale energies by a factor $E \to bE, b < 1$. In relativistic QFT, $\vec{p}$ scales like E, toward zero, $\vec{p} \to b\vec{p}$, since all the low-energy stuff is near the single special point $\vec{p} = 0$. Here the situation is much more interesting because there is a whole surface of low-energy stuff on the FS. This will lead to what's called *hyperscaling violation* – we can't just count powers of momentum.

[68] Notice that we are assuming translation invariance. I am not saying anything at the moment about whether translation invariance is discrete (the ions make a periodic potential) or continuous.

[69] We have chosen the normalization of ψ to fix the coefficient of the ∂_t term (this rescaling may depend on p).

One way to implement this is to introduce a hierarchical labeling of points in momentum space, by breaking the momentum space into *patches* around the FS. (An analogous strategy of labeling is also used in heavy quark EFT and in SCET.)



We'll use a slightly different strategy, following Polchinski. To specify a point $\vec{p}$, we pick the nearest point $\vec{k}$ on the FS, $\epsilon(\vec{k}) = \epsilon_F$ (draw a line perpendicular to the FS from $\vec{p}$), and let

$$\vec{p} = \vec{k} + \vec{\ell}.$$

So $d-1$ of the components are determined by $\vec{k}$ and one is determined by ℓ. (Clearly there are some exceptional cases if the FS gets too wiggly. Ignore these for now.)

$$\epsilon(p) - \epsilon_F = \ell v_F(\vec{k}) + \mathcal{O}(\ell^2), \qquad v_F \equiv \partial_p \epsilon|_{p=k}.$$

So a scaling rule which accomplishes our goal of focusing on the FS is

$$E \to bE, \quad \vec{k} \to \vec{k}, \quad \vec{l} \to b\vec{\ell}.$$

This implies

$$dt \to b^{-1}dt, \quad d^{d-1}\vec{k} \to d^{d-1}\vec{k}, \quad d\vec{\ell} \to bd\vec{\ell}, \quad \partial_t \to b\partial_t$$

$$S_{\text{free}} = \int \underbrace{dt\ d^{d-1}\vec{k}\ d\vec{\ell}}_{\sim b^0} \left(\mathbf{i}\psi^\dagger(p) \underbrace{\partial_t}_{\sim b^1} \psi(p) - \underbrace{\ell v_F(k)}_{\sim b^1} \psi^\dagger(p)\psi(p) \right)$$

In order to make this go like b^0 we require $\psi \to b^{-\frac{1}{2}}\psi$ near the free fixed point.

Next we will play the EFT game. To do so we must enumerate the symmetries we demand of our EFT:

1. Particle number, $\psi \to e^{\mathbf{i}\theta}\psi$

2. Spatial symmetries: either (a) continuous translation invariance and rotation invariance (as for *e.g.* liquid ^{3}He) or (b) lattice symmetries. This means that momentum space is periodically identified, roughly $p \simeq p + 2\pi/a$ where a is the lattice spacing (the set of independent momenta is called the Brillouin zone (BZ)) and p is only conserved modulo an inverse lattice vector $2\pi/a$; the momentum There can also be some remnant of rotation invariance preserved by the lattice. Case (b) reduces to case (a) if the Fermi surface does not go near the edges of the BZ.

3. Spin rotation symmetry, $\mathsf{SU}(n)$ if $\sigma = 1..n$. In the limit with $c \to \infty$, this is an internal symmetry, independent of rotations.

4. Let's assume that $\epsilon(p) = \epsilon(-p)$, which is a consequence of *e.g.* parity invariance.

Now we enumerate all terms analytic in ψ (since we are assuming that there are no other low-energy dofs integrating out which is the only way to get non-analytic terms in ψ) and consistent with the symmetries; we can order them by the number of fermion operators involved. Particle number symmetry means every ψ comes with a $\psi^\dagger$. The possible **quadratic terms** are:

$$\int \underbrace{dt\ d^{d-1}\vec{k}\ d\vec{\ell}}_{\sim b^0}\ \mu(k)\ \underbrace{\psi^\dagger_\sigma(p)\psi_\sigma(p)}_{\sim b^{-1}} \sim b^{-1}$$

is relevant. This is like a mass term. But don't panic: it just shifts the FS around. The *existence* of a Fermi surface is Wilson-natural (*i.e.* a stable assumption given generic coefficients of all possible terms in the action); any precise location or shape (modulo something enforced by symmetries, like roundness) is not.

Adding one extra ∂_t or factor of ℓ costs a b^1 and makes the operator marginal; those terms are already present in S_{free}. Adding more than one makes it irrelevant.

Quartic terms:

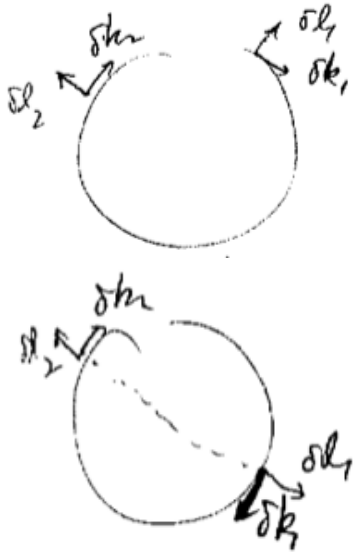
$$S_4 = \int \underbrace{dt\ \prod_{i=1}^{4} d^{d-1}\vec{k}_i d\vec{\ell}_i}_{\sim b^{-1+4-4/2}}\ u(4\cdots 1)\psi^\dagger_\sigma(p_1)\psi_\sigma(p_3)\psi^\dagger_{\sigma'}(p_2)\psi_{\sigma'}(p_4)\delta^d(\vec{p}_1+\vec{p}_2-\vec{p}_3-\vec{p}_4)$$

Note the similarity with the discussion of the XY model in §**??**. The minus signs on $p_{3,4}$ is because $\psi(p)$ *removes* a particle with momentum p. We assume u depends only on k, σ, so does not scale – this will give the most relevant piece. How does the delta function scale?

$$\delta^d\left(\vec{p}_1+\vec{p}_2-\vec{p}_3-\vec{p}_4\right) = \delta^d\left(k_1+k_2-k_3-k_4+\ell_1+\ell_2-\ell_3-\ell_4\right) \overset{?}{\simeq} \delta^d\left(k_1+k_2-k_3-k_4\right)$$

In the last (questioned) step, we used the fact that $\ell \ll k$ to ignore the contributions of the ℓs. If this is correct then the delta function does not scale (since ks do not), and $S_4 \sim b^1$ is irrelevant (and quartic interactions with derivatives are moreso). If this were correct, the free-fixed point would be exactly stable.

There are two important subtleties: (1) there exist phonons. (2) the questioned equality above is questionable because of kinematics of the Fermi surface. We will address these two issues in reverse order.



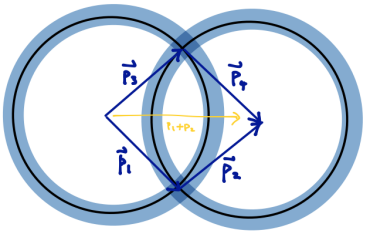
The **kinematic subtlety** in the treatment of the scaling of $\delta(p_1 + p_2 - p_3 - p_4)$ arises because of the geometry of the Fermi surface. Consider scattering between two points on the FS, where (in the labeling convention above)

$$p_3 = p_1 + \delta k_1 + \delta \ell_1, \quad p_4 = p_2 + \delta k_2 + \delta \ell_2,$$

in which case the momentum delta function is

$$\delta^d(p_1 + p_2 - p_3 - p_4) = \delta^d(\delta k_1 + \delta \ell_1 + \delta k_2 + \delta \ell_2).$$

For generic choices of the two points $p_{1,2}$ (top figure at left), δk_1 and δk_2 are linearly independent and the $\delta\ell$s can indeed be ignored as we did above. However, for two points with $p_1 = -p_2$ (they are called *nested*, as depicted in the bottom figure at left), then one component of $\delta k_1 + \delta k_2$ is automatically zero, revealing the tiny $\delta\ell$s to the force of (one component of) the delta function. In this case, $\delta(\ell)$ scales like b^{-1}, and for this particular kinematic configuration the four-fermion interaction is (classically) marginal. Classically marginal means quantum mechanics has a chance to make a big difference.



A useful visualization is at right ($d = 2$ with a round FS is shown; this is what's depicted on the cover of the famous book by Abrikosov-Gorkov-Dzyaloshinski): the blue circles have radius k_F; the yellow vector is the sum of the two initial momenta $p_1 + p_2$, both of which are on the FS; the condition that $p_3 + p_4$, each also on the FS, add up to the same vector means that p_3 must lie on the intersection of the two circles (spheres in $d > 2$). But when $p_1 + p_2 = 0$, the two circles are on top of each other so they intersect everywhere! Comments:

1. We assumed that both p_1 and $-p_2$ were actually *on* the FS. This is automatic if $\epsilon(p) = \epsilon(-p)$, *i.e.* if ϵ is only a function of p^2.

2. This discussion works for any $d > 1$.

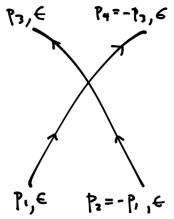
3. **Forward scattering.** There is a similar phenomenon for the case where $p_1 = p_3$ (and hence $p_2 = p_4$). This is called *forward scattering* because the final momenta are the same as the initial momenta. (We could just as well take $p_1 = p_4$ (and hence $p_2 = p_3$).) In this case too the delta function will constrain the ℓs and will therefore scale.

The tree-level-marginal 4-Fermi interactions at special kinematics leads to a *family* of fixed points labelled by 'Landau parameters'. In fact there is whole *function*'s worth of fixed points. In 2d, the fixed point manifold is parametrized by the forward-scattering function

$$F(\theta_1, \theta_2) \equiv u(\theta_4 = \theta_2, \theta_3 = \theta_1, \theta_2, \theta_1)$$

(Fermi statistics implies that $u(\theta_4 = \theta_1, \theta_3 = \theta_2, \theta_2, \theta_1) = -F(\theta_1, \theta_2)$.) and the BCS-channel interaction:

$$V(\theta_1, \theta_3) = u(\theta_4 = -\theta_3, \theta_3, \theta_2 = -\theta_1, \theta_1).$$



Now let's think about what decision the fluctuations make about the fate of the nested interactions. The first claim, which I will not justify here, is that F is not renormalized at one loop. The interesting bit is the renormalization of the BCS interaction:
The electron propagator, obtained by inverting the kinetic operator S_{free}, is

$$G(\epsilon, p = k + l) = \frac{1}{\epsilon(1 + \mathbf{i}\eta) - v_F(k)\ell + \mathcal{O}(\ell)^2}$$

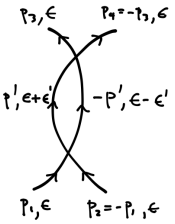
where I used $\eta \equiv 0^+$ for the infinitesimal specifying the contour prescription. (To understand the contour prescription for the *hole* propagator, it is useful to begin with

$$G(t, p) = \langle \epsilon_F | \, c_p^\dagger(t) c_p(0) \, | \epsilon_F \rangle \,, \quad c_p^\dagger(t) \equiv e^{-\mathbf{i}\mathbf{H}t} c_p^\dagger e^{\mathbf{i}\mathbf{H}t}$$

and use the free-fermion fact $[\mathbf{H}, c_p^\dagger] = \epsilon_p c_p^\dagger$.)

Let's assume rotation invariance. Then $V(\theta_3, \theta_1) = V(\theta_3 - \theta_1)$, $V_l = \int \text{đ}\theta e^{\mathbf{i}l\theta} V(\theta)$. Different angular momentum sectors decouple from each other at one loop.

We will focus on the s-wave bit of the interaction, so V is independent of momentum. We will integrate out just a shell in energy (depicted by the blue shaded shell in the Fermi surface figures). The interesting contribution comes from the following diagram:

$\delta^{(1)}V =$ 

$$= \mathbf{i}V^2 \int_{b\epsilon_0}^{\epsilon_0} \frac{d\epsilon' d^{d-1}k' d\ell'}{(2\pi)^{d+1}} \frac{1}{(\epsilon + \epsilon' - v_F(k')\ell')\,(\epsilon - \epsilon' - v_F(k')\ell')}$$

do $\int d\ell'$ by residues

$$= \mathbf{i}V^2 \int \frac{d\epsilon' d^{d-1}k'}{(2\pi)^{d+1}} \frac{1}{v_F(k')} \left(\underbrace{\epsilon - \epsilon' - (\epsilon + \epsilon')}_{=-2\epsilon'} \right)^{-1}$$

$$= -V^2 \underbrace{\int_{b\epsilon_0}^{\epsilon_0} \frac{d\epsilon'}{\epsilon'}}_{=\log(1/b)} \underbrace{\int \frac{d^{d-1}k'}{(2\pi)^d v_F(k')}}_{\text{dos at FS}} \tag{9.32}$$

Between the first and second lines, we did the ℓ' integral by residues. The crucial point is that we are interested in external energies $\epsilon \sim 0$, but we are integrating out a shell near the cutoff, so $|\epsilon'| > |\epsilon|$ and the sign of $\epsilon + \epsilon'$ is opposite that of $\epsilon - \epsilon'$; therefore there is a pole on either side of the real ℓ axis and we get the same answer by closing the contour either way. On one side the pole is at $\ell' = \frac{1}{v_F(k')}(\epsilon + \epsilon')$. (In the t-channel diagram (what Shankar calls ZS), the poles are on the *same* side and it therefore does not renormalize the four-fermion interaction.)

The result to one-loop is then

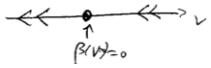
$$V(b) = V - V^2 N \log(1/b) + \mathcal{O}(V^3)$$

with $N \equiv \int \frac{d^{d-1}k'}{(2\pi)^d v_F(k')}$ is the density of states at the Fermi surface. From this we derive the beta function

$$b\frac{d}{db}V(b) = \beta_V = NV^2(b) + \mathcal{O}(V^3)$$

and the solution of the flow equation at $E = bE_1$ is

$$V(E) = \frac{V_1}{1 + NV_1 \log(E_1/E)} \begin{cases} \to 0 & \text{in IR for } V_1 > 0 \text{ (repulsive)} \\ \to -\infty & \text{in IR for } V_1 < 0 \text{ (attractive)} \end{cases} \tag{9.33}$$

There is therefore a very significant dichotomy depending on the sign of the coupling at the microscopic scale E_1, as in this phase diagram: 

The conclusion is that if the interaction starts attractive at some scale it flows to large attractive values. The thing that is decided by our perturbative analysis is that (if $V(E_1) > 0$) the decoupling we did with σ ('the BCS channel') wins over the decoupling with ρ ('the particle-hole channel'). What happens at $V \to -\infty$? Here we need non-perturbative physics.

The non-perturbative physics is in general hard, but we've already done what we can in §9.8.1.

The remaining question is: Who is V_1 and why would it be attractive (given that Coulomb interactions between electrons, while screened and therefore short-ranged, are repulsive)? The answer is:

Phonons. The lattice of positions taken by the ions making up a crystalline solid spontaneously break many spacetime symmetries of their governing Hamiltonian. This

implies a collection of gapless Goldstone modes in any low-energy effective theory of such a solid[70]. The Goldstone theorem is satisfied by including a field

$$\vec{D} \propto \text{(local) displacement } \delta\vec{r} \text{ of ions from their equilibrium positions}$$

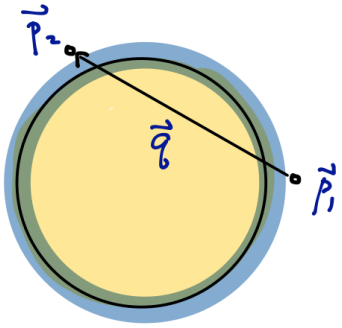
Most microscopically we have a bunch of coupled springs:

$$L_{\text{ions}} \sim \frac{1}{2} M \left(\delta\dot{\vec{r}}\right)^2 - k_{ij}\delta r^i \delta r^j + ...$$

with spring constants k independent of the nuclear mass M. It is useful to introduce a canonically normalized field in terms of which the action is

$$S[\vec{D} = (M)^{1/2}\,\delta\vec{r}] = \frac{1}{2}\int dt d^d q \left(\partial_t D_i(q)\partial_t D_i(-q) - \omega_{ij}^2(q) D_i(q) D_j(-q)\right).$$

Here $\omega^2 \propto M^{-1}$. Their status as Goldstones means that the eigenvalues of $\omega_{ij}^2(q) \sim |q|^2$ at small q: moving everyone by the same amount does not change the energy. This also constrains the coupling of these modes to the electrons: they can only couple through derivative interactions.



For purposes of their interactions with the electrons, a nonzero q which keeps the e^- on the FS must scale like $q \sim b^0$. Therefore

$$dt d^d q \left(\partial_t D\right)^2 \sim b^{+1+2[D]} \quad \Longrightarrow \quad D \sim b^{-\frac{1}{2}}$$

and the restoring force $dtdq D^2 \omega^2(q) \sim b^{-2}$ is relevant, and dominates over the ∂_t^2 term for

$$E < E_D = \sqrt{\frac{m}{M}} E_0 \quad \text{the Debye energy.}$$

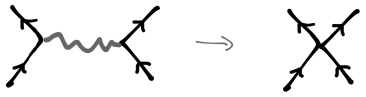
This means that phonons mediate *static* interactions below E_D – we can ignore retardation effects, and their effects on the electrons can be fully incorporated by the four-fermion interaction we used above (with some $\vec{k}$ dependence). How do they couple to the electrons?

$$S_{\text{int}}[D, \psi] = \int dt q^3 q d^2 k_1 d\ell_1 d^2 k_2 d\ell_2 \; M^{-\frac{1}{2}} g_i(q, k_1, k_2) D_i(q) \psi_\sigma^\dagger(p_1) \psi_\sigma(p_2) \delta^3(p_1 - p_2 - q)$$

[70] Note that there is a subtlety in counting Goldstone modes from spontaneously broken spacetime symmetries: there are more symmetry generators than Goldstones. Basically it's because the associated currents differ only by functions of spacetime; but a localized Goldstone particle is anyway made by a current times a function of spacetime, so you can't sharply distinguish the resulting particles. Some useful references on this subject are Low-Manohar and most recently Watanabe-Murayama.

$$\sim b^{-1+1+1-3/2} = b^{-1/2} \tag{9.34}$$

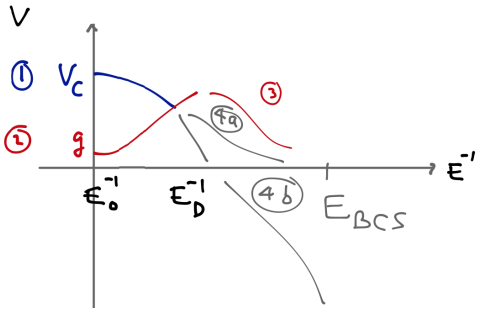
– here we took the delta function to scale like b^0 as above. This is relevant when we use the $\dot{D}^2$ scaling for the phonons; when the restoring force dominates we should scale D differently and this is irrelevant for generic kinematics. This is consistent with our previous analysis of the four-fermion interaction.

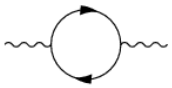
The summary of this discussion is: phonons do not destroy the Fermi surface, but they *do* produce an attractive contribution to the 4-fermion interaction, which is relevant in some range of scales (above the Debye energy). Below the Debye energy, it amounts to an addition to V that goes like $-g^2$: 

Notice that the scale at which the coupling V becomes strong ($V(E_{\rm BCS}) \equiv 1$ in (9.33)) is

$$E_{\rm BCS} \sim E_D e^{-\frac{1}{NV_D}}.$$

Two comments about this: First, it is non-perturbative in the interaction V_D. Second, it provides some verification of the role of phonons, since $E_D \sim M^{-1/2}$ can be varied by studying the same material with different isotopes and studying how the critical superconducting temperature ($\sim E_{\rm BCS}$) scales with the nuclear mass.



Here's the narrative, proceeding as a function of decreasing energy scale, beginning at E_0, the Planck scale of solids: (1) Electrons repel each other by the Coulomb interaction. However, in a metal, this interaction is *screened* by processes like this: 

(the intermediate state is an electron-hole pair) and is short-ranged. It is still repulsive, however. As we coarse-grain more and more, we see more and more electron-hole pairs and the force weakens. (2) While this is happening, the electron-phonon interaction is relevant and growing. This adds an attractive bit to V. This lasts until E_D. (3) At E_D the restoring force term in the phonon lagrangian dominates (for the purposes of their interactions with the electrons) and we can integrate them out. (4) What happens next depends on the sign of $V(E_D)$. If it's positive, V flows harmlessly to zero. If it's negative, it becomes moreso until we exit the perturbative analysis at $E_{\rm BCS}$, and vindicate our choice of Hubbard-Stratonovich channel above.

Further brief comments, for which I refer you to Shankar:

1. Putting back the possible angular dependence of the BCS interaction, the result

at one loop is

$$\frac{dV(\theta_1-\theta_3)}{d\ell} = -\frac{1}{8\pi^2}\int_0^{2\pi} đ\theta V(\theta_1-\theta)V(\theta-\theta_3)$$

or in terms of angular momentum components,

$$\frac{dV_l}{d\ell} = -\frac{V_l^2}{4\pi}.$$

2. This example is interesting and novel in that it is a (family of) fixed point(s) characterized by a dimensionful quantity, namely k_F. This leads to a phenomenon called *hyperscaling violation* where thermodynamic quantities need not have their naive scaling with temperature.

3. The one loop analysis gives the right answer to all loops in the limit that $N \equiv k_F/\Lambda \gg 1$, where Λ is the UV cutoff on the momentum.

4. The forward scattering interaction (for any choice of function $F(\theta_{13})$) is not renormalized at one loop. This means it is exactly marginal at leading order in N.

5. Like in ϕ^4 theory, the sunrise diagram at two loops is the first appearance of wavefunction renormalization. In the context of the Fermi liquid theory, this leads to the renormalization of the effective mass which is called $m^\star$.

Another consequence of the FS kinematics which I should emphasize more: it allows the quasiparticle to be stable. The leading contribution to the decay rate of a one-quasiparticle state with momentum k can be obtained applying the optical theorem to the following process.

The intermediate state is two electrons with momenta $k'+q$ and $k-q$, and one *hole* with momentum k'. The hole propagator has the opposite $\mathbf{i}\eta$ prescription. After doing the frequency integrals by residues, we get

$$\Sigma(k,\epsilon) = \int đq\, đk' \frac{|u_q|^2}{D-\mathbf{i}\eta}$$

$$D \equiv \epsilon_k(1+\mathbf{i}\eta) + \epsilon_{k'}(1-\mathbf{i}\eta) - \epsilon_{k'+q}(1+\mathbf{i}\eta) - \epsilon_{k-q}(1+\mathbf{i}\eta)$$

(Notice that this is the eyeball diagram which gives the lowest-order contribution to the wavefunction renormalization of a field with quartic interactions.) By the optical theorem, its imaginary part is the (leading contribution to the) inverse-lifetime of the quasiparticle state with fixed k:

$$\tau^{-1}(k) = \mathrm{Im}\,\Sigma(k,\epsilon) = \pi \int đq\, đk' \delta(D)|u_q|^2 f(-\epsilon_{k'}) f(\epsilon_{k'+q}) f(\epsilon_{k-q})$$

where

$$f(\epsilon) = \lim_{T\to 0} \frac{1}{e^{\frac{\epsilon-\epsilon_F}{T}}+1} = \theta(\epsilon < \epsilon_F)$$

is the Fermi function. This is just the demand that a particle can only scatter into an empty state and a hole can only scatter into a filled state. These constraints imply that all the energies are near the Fermi energy: both $\epsilon_{k'+q}$ and $\epsilon_{k'}$ lie in a shell of radius ϵ about the FS; the answer is proportional to the density of possible final states, which is thus

$$\tau^{-1} \propto \left(\frac{\epsilon}{\epsilon_F}\right)^2 .$$

So the width of the quasiparticle resonance is

$$\tau^{-1} \propto \epsilon^2 \ll \epsilon$$

much smaller than its frequency – it is a sharp resonance, a well-defined particle.
