Physics 215C: Particles and Fields
Spring 2019

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

Last updated: 2019/04/15, 14:44:11
## Contents

0.1 Introductory remarks for the third quarter ....................... 3  
0.2 Sources and acknowledgement .................................. 6  
0.3 Conventions ....................................................... 7  

1 Anomalies 8  

2 Effective field theory 19  
  2.1 A parable on integrating out degrees of freedom ............... 19  
  2.2 Introduction to effective field theory .......................... 24  
  2.3 The color of the sky ............................................ 29  
  2.4 Fermi theory of Weak Interactions .............................. 31  
  2.5 Loops in EFT .................................................. 32  
  2.6 The Standard Model as an EFT. ................................. 38  
  2.7 Superconductors ................................................ 41  
  2.8 Effective field theory of Fermi surfaces ....................... 45  
  2.9 Pions ........................................................... 56  

3 Topological terms in field theory actions 65  
  3.1 Coherent state path integral for fermions ...................... 65  
  3.2 Coherent state path integral for spins ........................ 65  
  3.3 Topological terms from integrating out fermions ............... 65
0.1 Introductory remarks for the third quarter

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

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

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

- There is more to QFT than the S-matrix. In a particle-physics QFT course (like 215 so far) you learn that the purpose in life of correlation functions or green’s functions or off-shell amplitudes is that they have poles (at $p^\mu p_\mu - m^2 = 0$) whose residues are the S-matrix elements, which are what you measure (or better, are the distribution you sample) when you scatter the particles which are the quanta of the fields of the QFT. I want to make two extended points about this:

  1. In many physical contexts where QFT is relevant, you can actually measure the off-shell stuff. This is yet another reason why including condensed matter in our field of view will deepen our understanding of QFT.

  2. This is good, because the Green’s functions don’t always have simple poles! There are lots of interesting field theories where the Green’s functions instead have power-law singularities, like $G(p) \sim \frac{1}{p^{2\Delta}}$. If you Fourier transform this, you don’t get an exponentially-localized packet. The elementary excitations created by a field whose two point function does this are not particles. (Any conformal field theory (CFT) is an example of this.) The theory of particles (and their dance of creation and annihilation and so on) is an important but proper subset of QFT.

- In addition to its importance in high energy physics, I want to emphasize that QFT is also quite central in many aspects of condensed matter physics, and we will learn about this. From the point of view of someone interested in QFT, high energy particle physics has the severe drawback that it offers only one example! (OK, for some purposes we can think about QCD and the electroweak theory separately...)

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

• A central theme this quarter is coarse-graining in quantum systems with extensive degrees of freedom, aka the RG in QFT.

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

More precisely, in a quantum mechanical system, we specify the degrees of freedom by their Hilbert space; by an extensive system, I’ll mean one in which the Hilbert space is of the form $\mathcal{H} = \otimes_{\text{patches of space}} \mathcal{H}_{\text{patch}}$ and the interactions are local $H = \sum_{\text{patches}} H(\text{nearby patches})$. (Actually, the Hilbert space of a gauge theory is not of this form; rather, it is a subspace of such a space which satisfies the Gauss law. This is a source of a lot of confusion, which I hope to dispel.)

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

To continue the sod example in 2+1 dimensions, a person laying the sod in the picture above cares that the sod doesn’t fall apart, and rolls nicely onto the ground (as long as we don’t do high-energy probes like bending it violently or trying to lay it down too quickly). These long-wavelength properties of rigidity and elasticity are collective, emergent properties of the microscopic constituents (sod molecules) – we can describe the dynamics involved in covering the Earth with sod (never mind whether this is a good idea in a desert climate) without knowing the microscopic theory of the sod molecules (‘grass’). Our job is to think about the relationship between the microscopic model (grassodynamics) and its macroscopic counterpart (in this case, suburban landscaping). In my experience, learning to do this is approximately synonymous with understanding.
• The crux of many problems in physics is the correct choice of variables with which to label the degrees of freedom. Often the best choice is very different from the obvious choice; a name for this phenomenon is ‘duality’. There are many examples of it and we will study some of them. This word is dangerous because it is about ambiguities in our (physics) language. I would like to reclaim it.

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

• We’ll be interested in models with a finite number of degrees of freedom per unit volume. This last is important, because we are going to be interested in the thermodynamic limit.

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

Topics which I hope to discuss this quarter include:

• effects of topology in QFT (this includes anomalies, topological solitons and defects, topological terms in the action)

• some illustrations of effective field theory (perhaps cleverly mixed in with the other subjects)

• the uses and limitations of path integrals of various kinds

• Large-\(N\) expansions

• more deep mysteries of gauge theory and its emergence in physical systems.

• duality.

I welcome your suggestions regarding which subjects in QFT we should study.
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)
- Banks, *Modern Quantum Field Theory: A Concise Introduction* (Cambridge)
- Schwartz, *Quantum field theory and the standard model* (Cambridge)
- Xiao-Gang Wen, *Quantum Field Theory of Many-Body Systems*
- Sidney Coleman, *Aspects of Symmetry*
- Alexander Polyakov, *Gauge Fields and Strings*
- Eduardo Fradkin, *Field Theories of Condensed Matter Systems*

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:

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

(spacelike is negative). We will also write $\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = (\partial_t, \vec{\nabla}_x)^\mu$, 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 $\hbar \equiv \frac{\hbar}{2\pi}$ is $dk \equiv \frac{dk}{2\pi}$. I will also write $\delta^d(q) \equiv (2\pi)^d \delta(d)(q)$. I will try to be consistent about writing Fourier transforms as

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

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


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. It means that the theory *cannot* be regulated in a way that preserves the symmetry.

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.)

You could say that we have already seen a dramatic example of an anomaly: the violation of classical scale invariance (*e.g.* in massless $\phi^4$ 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. Partly for this reason, the concept of ‘anomaly’ is not native to the condensed matter literature (but has recently been flourishing there).

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 $\pi^0 \rightarrow \gamma\gamma$, which would not happen if the chiral current were conserved. (The relationship between the chiral current and the pion is explained in §2.9.)

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.

1.0.1 Chiral anomaly

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

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

which means that $\gamma^5$ commutes with the Lorentz generators

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

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

Note that in $D = 4k$ dimensions, if $\psi_L$ is a left-handed spinor in representation $r$ of some group $G$, then its image under CPT, $\psi_L^{\text{CPT}}(t, \vec{x}) \equiv i\gamma^0 (\psi_L(-t, -\vec{x}))^*$, is right-handed and transforms in representation $\bar{r}$ of $G$. Therefore chiral symmetries arise when the Weyl fermions transform in *complex representations* of the symmetry group, where $\bar{r} \neq r$. (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 $\gamma^5$ is diagonal:

$$\gamma^\mu = \begin{pmatrix} 0 & \bar{\sigma}^\mu \\ \sigma^\mu & 0 \end{pmatrix}, \quad \sigma^\mu \equiv (\mathbb{1}, \vec{\sigma})^\mu, \quad \bar{\sigma}^\mu \equiv (\mathbb{1}, -\vec{\sigma})^\mu, \quad \gamma^5 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}.$$
This makes the reducibility of the Dirac representation of $SO(3,1)$ manifest, since the Lorentz generators are $\propto [\gamma^\mu,\gamma^\nu]$ block diagonal in this basis. The gammas are a map from the $(1,2^R)$ representation to the $(2^L,1)$ representation. It is sometimes useful to denote the $2^R$ indices by $\alpha,\beta = 1,2$, and the $2^L$ 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:

$$\bar{\chi} = \chi^\dagger \gamma^0, \quad \bar{\chi} \gamma^5 = \chi^\dagger \gamma^{05} = -\chi^\dagger \gamma^5 = -\bar{\chi}.$$

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 $\psi$ is in representation $r$ means that its transformation law is

$$\delta \psi_a = i \epsilon^A (t^A)_{ab} \psi_b$$

where $t^A, A = 1,..\dim G$ are generators of $G$ in representation $r$; 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 $\psi^{*T}$ transforms:

$$\delta \psi^{*T}_a = -i \epsilon^A (t^A)_{ab} \psi^{*T}_b.$$

So:

$$t^A_T = -(t^A)^T.$$

The condition for a complex representation is that this is different from $t^A_T$ (actually we have to allow for relabelling of the generators and the basis). The simplest case is $G = U(1)$, 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):

$$e^{iS_{\text{eff}}[A]} \equiv \int [D\psi \bar{D} \bar{\psi}] e^{iS[\psi,\bar{\psi},A]}.$$  

We must specify how the fermions coupled to the gauge field. The simplest example is if $A$ is a $U(1)$ gauge field and $\psi$ is minimally coupled:

$$S[\psi, \bar{\psi}, A] = \int d^D x \bar{\psi} i D \psi, \quad D \psi \equiv \gamma^\mu (\partial_\mu + iA_\mu) \psi.$$
We will focus on this example, but you could imagine instead that $A_\mu$ is a non-Abelian gauge field for the group $G$, and $\psi$ is in a representation $R$, with gauge generators $T^A(R) (A = 1...\dim G)$, so the coupling would be

$$\bar{\psi}i\slashed{D}\psi = \bar{\psi}_a \gamma^\mu \left( \partial_\mu \delta_{ab} + iA_\mu A^A(R)_{ab} \right) \psi_b.$$  

(1.1)

Much of the discussion below applies for any even $D$.

In the absence of a mass term, the action (in the Weyl basis) involves no coupling between $L$ and $R$:

$$S[\psi, \bar{\psi}, A] = \int d^Dx \left( \bar{\psi}_L^i \sigma^\mu D_\mu \psi_L + \bar{\psi}_R^i \sigma^\mu D_\mu \psi_R \right)$$

and therefore is invariant under the global chiral rotation

$$\psi \to e^{i\alpha \gamma^5} \psi, \quad \psi^\dagger \to \psi^{\dagger} e^{-i\alpha \gamma^5}, \quad \bar{\psi} \to \bar{\psi} e^{i\alpha \gamma^5}. \quad \text{That is:} \quad \psi_L \to e^{i\alpha} \psi_L, \quad \psi_R \to e^{-i\alpha} \psi_R.$$

(The mass term couples the two components

$$L_m = \bar{\psi} \left( \text{Re} m + \text{Im} m \gamma^5 \right) \psi = m \bar{\psi}_L^i \psi_R^i + h.c.;$$

notice that the mass parameter is complex.) The associated Noether current is $j_\mu^5 = \bar{\psi} \gamma^\mu \gamma_\mu \psi$, and it seems like we should have $\partial^\mu j_\mu^5 = 0$. This follows from the massless (classical) Dirac equation $0 = \gamma^\mu \partial_\mu \psi$. (With the mass term, we would have instead $\partial^\mu j_\mu^5 = 2i\bar{\psi} \left( \text{Re} \gamma^5 + \text{Im} m \right) \psi$.)

Notice that there is another current $j_\mu = \bar{\psi} \gamma^\mu \psi$. $j_\mu$ 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

$$S_{\text{eff}}[A_\mu] = S_{\text{eff}}[A_\mu + \partial_\mu \lambda] \sim \log \left( e^{\int A(x) \partial_\mu j_\mu^5} \right) + S_{\text{eff}}[A_\mu].$$

No matter what happens we can’t find an anomaly in $j_\mu$. 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 $\psi'(x) = e^{i\alpha(x) \gamma^5} \psi(x)$ into the action:

$$S_F[\psi'] = \int d^Dx \bar{\psi}' e^{i\alpha(x) \gamma^5} i\slashed{D} e^{i\alpha(x) \gamma^5} \psi = \int d^Dx \left( \bar{\psi} i\slashed{D} \psi + \bar{\psi}_i \gamma^5 \left( \partial_\mu \alpha \right) \psi \right) \overset{\text{IBP}}{=} S_F[\psi] - i \int \alpha(x) \partial^\mu \text{tr} \bar{\psi} \gamma^5 \gamma_\mu \psi.$$

Then we can completely get rid of $\alpha(x)$ if we can change integration variables, i.e. if $[D \psi'] = [D \psi]$. Usually this is true, but here we pick up an interesting Jacobian.
Claim:

\[ e^{i S_{\text{eff}}[A]} = \int [D\psi' D\bar{\psi}'] e^{i S_F[\psi']} = \int [D\psi D\bar{\psi}] e^{i S_F[\psi]} + \int d^D x \alpha(x) (\partial_\mu j^5_\mu - A(x)) \]

where \( A \) comes from the variation of the measure. That is,

\[ [D\psi' D\bar{\psi}'] = [D\psi D\bar{\psi}] \det \left( e^{i \alpha \gamma^5} \right) \equiv e^{-i \int \alpha A} \]

so

\[ e^{-i \int \alpha A} = e^{\text{Tr} \log e^{i \alpha \gamma^5}} = e^{\text{Tr}(i \alpha \gamma^5)} \]

or more explicitly

\[ A(x) = \sum_n \text{tr} \xi_n(x) \gamma^5 \xi_n(x) \quad (1.2) \]

where \( \xi_n \) are a basis of eigenspinors of the Dirac operator.

The expression above for \( S_{\text{eff}} \) is actually independent of \( \alpha \), since the path integral is invariant under a change of variables. For a conserved current, \( \alpha \) would multiply the divergence of the current and this demand would imply current conservation. Here this implies that instead of current conservation we have a specific violation of the current:

\[ \partial_\mu j^5_\mu = A(x). \]

**What is the anomaly \( A \)?** [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

\[ \langle \partial_\mu j^5_\mu \rangle = \partial_\mu \langle \bar{\psi}(x) \gamma^\mu \gamma^5 \psi(x) \rangle \]

– the coincident operators on the RHS need to be regulated.

The classical (massless) Dirac equation immediately implies that the axial current is conserved (up to contact terms)

\[ \partial_\mu (\bar{\psi} \gamma^\mu \gamma^5 \psi) \equiv 0. \]

Consider, on the other hand, the (Euclidean vacuum) expectation value

\[ j^5_\mu \equiv \langle \bar{\psi}(x) \gamma^\mu \gamma^5 \psi(x) \rangle \equiv Z^{-1}[A] \int [D\psi D\bar{\psi}] e^{-S_F[\psi]} j^5_\mu(x) \]

\[ = -\text{Tr} \, \gamma^\mu \gamma^5 G^{(A)}(x,x) \quad (1.3) \]
where $G$ is the Green’s function of the Dirac operator in the gauge field background (and the figure is from Polyakov’s book). We can construct it out of eigenfunctions of $i\partial$:

$$i\partial \xi_n(x) = \epsilon_n \xi_n(x), \quad \bar{\xi}_n(x) i\gamma^\mu \left( -\partial_\mu + iA_\mu \right) = \epsilon_n \bar{\xi}_n \quad (1.4)$$

in terms of which

$$G(x, x') = \sum_n \frac{1}{\epsilon_n} \xi_n(x) \bar{\xi}_n(x'). \quad (1.5)$$

(I am suppressing spinor indices all over the place, note that here we are taking the outer product of the spinors.)

We want to define the coincidence limit, as $x' \to x$. The problem with this limit arises from the large $|\epsilon_n|$ eigenvalues; the contributions of such short-wavelength modes are local and most of them can be absorbed in renormalization of couplings. It should not (and does not) matter how we regulate them, but we must pick a regulator. A convenient choice here is heat-kernel regulator:

$$G_s(x, x') \equiv \sum_n e^{-s\epsilon_n^2} \frac{1}{\epsilon_n} \xi_n(x) \bar{\xi}_n(x')$$

and

$$J_5^\mu (x) = \sum_n e^{-s\epsilon_n^2} \frac{1}{\epsilon_n} \xi_n(x) \gamma^5 \gamma_\mu \xi_n(x).$$

The anomaly is

$$\partial^\mu J_5^\mu = \partial^\mu \langle j_5^\mu \rangle = \sum_n i\partial^\mu \left( \bar{\xi}_n \gamma_\mu \gamma^5 \xi_n \right) \frac{e^{-s\epsilon_n^2}}{\epsilon_n}.$$ 

The definition (1.4) says

$$i\partial^\mu \left( \bar{\xi}_n \gamma^5 \gamma_\mu \xi_n \right) = 2\epsilon_n \bar{\xi}_n \gamma_5 \xi_n$$

using $\{\gamma^5, \gamma^\mu\} = 0$. (Notice that the story would deviate dramatically here if we were studying the vector current which lacks the $\gamma^5$.) This gives

$$\partial^\mu J_5^\mu (x) = 2\text{tr} \gamma \langle x | \gamma^5 e^{-s(\partial)^2} | x \rangle$$

with

$$(i\partial)^2 = -\left( \gamma_\mu \left( \partial_\mu + iA_\mu \right) \right)^2 = -\left( \partial_\mu + iA_\mu \right)^2 - \frac{i}{2} \Sigma_{\mu\nu} F^{\mu\nu}$$

where $\Sigma_{\mu\nu} \equiv \frac{1}{2} [\gamma_\mu, \gamma_\nu]$ is the spin Lorentz generator. This is (1.2), now better defined by the heat kernel regulator.

[End of Lecture 1]

---

1 Actually, this step is full of danger. (Polyakov has done it to me again. Thanks to Sridip Pal for discussions of this point.) See §1.0.2 below.
We’ve shown that in any even dimension,
\[ \partial_{\mu} \langle j_{5}^{\mu}(x) \rangle = 2 \text{tr}_{\gamma} \langle x | \gamma^{5} e^{sp^{2}} | x \rangle \]  
(1.6)

This can now be expanded in small \( s \), which amounts to an expansion in powers of \( A, F \). If there is no background field, \( A = 0 \), we get

\[ \langle x | e^{-s (iD)^{2}} | x \rangle = \int d^{D} p \ e^{-sp^{2}} = \frac{K_{D}}{s^{D/2}} \frac{1}{\Omega_{D}} \frac{D=4}{16\pi^{2}s^{2}}. \]  
(1.7)

This term will renormalize the charge density
\[ \rho(x) = \langle \psi^\dagger \psi(x) \rangle = \text{tr}_{\gamma} \gamma^{0} G(x, x), \]
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
\[ \text{tr} (\gamma^{5} G(x, x)) |_{A=0} \propto \text{tr}_{\gamma^{5}} = 0. \]

Similarly, bringing down more powers of \( (\partial + A)^{2} \) doesn’t give something nonzero since the \( \gamma^{5} \) 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:

\[ \text{tr} \left( \gamma^{5} e^{-s (iD)^{2}} \right) = \frac{s^{2} (i^{2})}{16\pi^{2}s^{2}} \cdot \text{tr}_{\gamma} \langle x | e^{-s (iD)^{2}} | x \rangle \cdot \text{tr}_{\gamma^{5}} \langle F_{\mu\nu} F_{\rho\lambda} \rangle + O(s^{-1}). \]

In the abelian case, just ignore the trace over color indices, \( \text{tr}_{c} \). The terms that go like positive powers of \( s \) go away in the continuum limit. Therefore

\[ \partial_{\mu} J_{5}^{\mu} = -2 \cdot \frac{1}{16\pi^{2}s^{2}} \cdot \frac{s^{2}}{8} \cdot 4\epsilon^{\mu\nu\rho\lambda} \text{tr}_{c} F_{\mu\nu} F_{\rho\lambda} + O(s^{1}) = -\frac{1}{8\pi^{2}} \text{tr} F_{\mu\nu} (\ast F)^{\mu\nu}. \]  
(1.8)

(Here \( (\ast F)^{\mu\nu} \equiv \frac{1}{8} \epsilon^{\mu\nu\rho\lambda} F_{\rho\lambda} \).) This is the chiral anomaly formula. It can also be usefully written as:

\[ \partial_{\mu} J_{5}^{\mu} = -\frac{1}{8\pi^{2}} \text{tr} F \wedge F = -\frac{1}{32\pi^{2}} \vec{E} \cdot \vec{B}. \]

• This object on the RHS is a total derivative. In the abelian case it is

\[ F \wedge F = d \left( A \wedge F \right). \]
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:

$$\Delta Q_A \equiv \Delta (N_L - N_R) = \int dt \partial_t J_5^0 = \int_{M_4} \partial^\mu J_5^\mu = 2 \int_{M_4} \frac{F \wedge F}{16\pi^2}$$

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 (1.3). Which term in that expansion gave the nonzero contribution to the axial current violation? In $D = 4$ it is the diagram with three current insertions, the ABJ triangle diagram. So in fact we did end up computing the triangle diagram. But this calculation also shows that nothing else contributes, even non-perturbatively.

• 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 $\gamma^5$ – 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 (1.1) through the whole calculation, the only difference is that the trace in (1.8) would have included a trace over representations of the gauge group. And we could have considered also a non-abelian flavor transformation in the chiral symmetry

$$\psi_I \rightarrow \left(e^{i\gamma_5 g^a \tau^a}\right)_{I,I} \psi_J$$

for some flavor rotation generator $\tau^a$. Then we would have found:

$$\partial^\mu j_5^{a\mu} = \frac{1}{16\pi^2} \epsilon^{\mu\nu\rho\lambda} F_{\mu\nu}^A F_{\rho\lambda}^B \text{tr}_{c,a} (T^A T^B \tau^a).$$

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.
1.0.2 Zeromodes of the Dirac operator

Do you see 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 zermodes, eigenspinors with eigenvalue \( \epsilon_n = 0 \). In that case, \( i\slashed{D} \) is not invertible, and the expression (1.5) for \( G \) is ambiguous. This factor of \( \epsilon_n \) is about to be cancelled when we compute the divergence of the current and arrive at (1.2). Usually this kind of thing is not a problem because we can lift the zermodes a little and put them back at the end. But here it is actually hiding something important. The zermodes cannot just be lifted. This is true because nonzero modes of \( i\slashed{D} \) must come in left-right pairs: this is because \( \{ \gamma^5, i\slashed{D} \} = 0 \), so \( i\slashed{D} \) and \( \gamma^5 \) cannot be simultaneously diagonalized in general. That is: if \( i\slashed{D}\xi = \epsilon\xi \) then \( (\gamma^5\xi) \) is also an eigenvector of \( i\slashed{D}\xi \), with eigenvalue \(-\epsilon\). Only for \( \epsilon = 0 \) does this fail, so zermodes can come by themselves. So you can’t just smoothly change the eigenvalue of some \( \xi_0 \) 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 \((i\slashed{D})^2\) (with eigenvalue \( \epsilon_n^2 \)) and of \( \gamma^5 \) with \( \gamma^5 = \pm \) respectively.

This leads us to a deep fact, called the (Atiyah-Singer) index theorem: only zermodes can contribute to the anomaly. Any mode \( \xi_n \) with nonzero eigenvalue has a partner (with the same eigenvalue of \((i\slashed{D})^2\)) with the opposite sign of \( \gamma^5 \); hence they cancel exactly in

\[
\text{tr}\gamma^5 e^{-s(i\slashed{D})^2} = \sum_{n,L/R} \chi_n^{L/R} \gamma^5 \chi_n^{L/R} e^{-s\epsilon_n^2} + \text{zeromodes}.
\]

So the anomaly equation tells us that the number of zermodes of the Dirac operator, weighted by handedness (i.e. with a + for \( L \) and - for \( R \)) is equal to

\[
n_L - n_R = \int d^Dx A(x) = \int \frac{1}{16\pi^2} F \wedge F.
\]

A practical consequence for us is that it makes manifest that the result is independent of the regulator \( s \).

Another consequence is that in the background of a gauge field configuration with nonzero \( n_L - n_R = q \), the vacuum to vacuum amplitude

\[
Z[A^0] = \int [D\psi D\bar{\psi}] e^{i \int d^Dx \bar{\psi} A\psi} = \det i\slashed{D} = 0
\]

vanishes, since it is the determinant of an operator with a kernel. Rather, only amplitudes for transitions which change the chiral charge by \( q \) are allowed.
1.0.3 The physics of the anomaly

[Polyakov, page 102; Kaplan 0912.2560 §2.1; Alvarez-Gaumé] Consider non-relativistic free (i.e. no 4-fermion interactions) fermions in 1+1 dimensions, e.g. with 1-particle dispersion \( \omega_k = \frac{1}{2m} \vec{k}^2 \). The groundstate of \( N \) such fermions is described by filling the \( N \) lowest-energy single particle levels, up the Fermi momentum: \( |k| \leq k_F \) are filled. We must introduce an infrared regulator so that the levels are discrete – put them in a box of length \( L \), so that \( k_n = \frac{2\pi n}{L} \). (In Figure 1, the red circles are possible 1-particle states, and the green ones are the occupied ones.) The lowest-energy excitations of this groundstate come from taking a fermion just below the Fermi level \( k = |k_F - k_1| \) and putting it just above, \( k = |k_F + k_2| \); the energy cost is

\[
E_{k_1-k_2} = \frac{1}{2m} (k_F + k_1)^2 - \frac{1}{2m} (k_F - k_2)^2 \approx \frac{k_F}{m} (k_1 - k_2)
\]

– 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:

\[
(\omega - v_F \delta k) \psi_L = 0, \quad (\omega + v_F \delta k) \psi_R = 0
\]

where \( \delta k \equiv k - k_F \).

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 \( \psi_L \) and \( \psi_R \) 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

\[
H = -t \sum_n c_n^\dagger c_{n+1} + \text{h.c.}
\]

where the dispersion would be \( \omega_k = -2t (\cos ka - 1) \sim \frac{1}{2m} k^2 + O(k^4) \) with \( \frac{1}{2m} = ta^2 \).

\[\text{Figure 1: Green dots represent occupied 1-particle states. Top: In the groundstate. Bottom: After applying } E_x(t).\]
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

\[
\Delta p = e \int dt E_x(t).
\]

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 \( \Delta p \), 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

\[
S[\psi] = \int dx dt \bar{\psi} (i\gamma^\mu \partial_\mu) \psi,
\]

where \( \gamma^\mu \) are \( 2 \times 2 \) and the upper/lower component of \( \psi \) 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

\[
\Delta Q_A = \Delta(N_L - N_R) = 2 \frac{\Delta p}{2\pi/L} = \frac{L}{\pi} \Delta p = \frac{L}{\pi} e \int dt E_x(t) = \frac{e}{\pi} \int dt dx E_x = \frac{e}{2\pi} \int \epsilon_{\mu\nu} F^{\mu\nu}
\]

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

\[
\partial^\mu J_A^\mu = \frac{e}{2\pi} \epsilon_{\mu\nu} F^{\mu\nu}.
\]

This agrees exactly with the anomaly equation in \( D = 1+1 \) produced by the calculation above in (1.6) (see the homework).
2 Effective field theory

2.1 A parable on integrating out degrees of freedom

Here’s another parable from QM which gives some useful perspective on renormalization in QFT and on the notion of effective field theory.

[Banks p. 138] Consider a system of two coupled harmonic oscillators. We will assume one of the springs is much stiffer than the other: let’s call their natural frequencies $\omega_0$, $\Omega$, with $\omega_0 \ll \Omega$. The euclidean-time action is

$$S[Q,q] = \int dt \left[ \frac{1}{2} (\dot{q}^2 + \omega_0^2 q^2) + \frac{1}{2} (\dot{Q}^2 + \Omega^2 Q^2) + g Q q^2 \right] \equiv S_{\omega_0}[q] + S_\Omega[Q] + S_{\text{int}}[Q,q].$$

(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:

$$Z = \int [dQ dq] e^{-S[Q,q]}.$$

Since I put a minus sign rather than an $i$ in the exponent (and the potential terms in the action have + signs), this is a euclidean path integral.

Let’s consider what happens if we do the path integral over the heavy mode $Q$, and postpone doing the path integral over $q$. This step, naturally, is called integrating out $Q$, and we will see below why this is a good idea. The result just depends on $q$; we can think of it as an effective action for $q$:

$$e^{-S_{\text{int}}[q]} := \int [dQ] e^{-S[Q,q]} = e^{-S_{\omega_0}[q]} \langle e^{-S_{\text{int}}[Q,q]} \rangle_Q.$$

Here $\langle \ldots \rangle_Q$ 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 $S_{\text{int}}$):

$$\langle e^{-S_{\text{int}}[Q,q]} \rangle_Q = \int [dQ] e^{-S_\Omega[Q] - \int ds J(s) Q(s)} = \mathcal{N} e^{\frac{1}{2} \int ds dt J(s) G(s,t) J(t)}.$$

This last equality is an application of the ‘fundamental theorem of path integrals,’ i.e. the gaussian integral. Here $J(s) \equiv gq(s)^2$. The normalization factor $\mathcal{N}$ is independent of $J$ and hence of $q$. And $G(s,t)$ is the inverse of the linear operator appearing in $S_\Omega$, the euclidean Green’s function:

$$S_\Omega[Q] = \int ds dt Q(s) G^{-1}(s,t) Q(t).$$
More usefully, $G$ satisfies
\[
(-\partial_s^2 + \Omega^2) G(s, t) = \delta(s - t).
\]
The fact that our system is time-translation invariant means $G(s, t) = G(s - t)$. We can solve this equation in fourier space: $G(s) = \int d\omega e^{-i\omega s} G_\omega$ makes it algebraic:
\[
G_\omega = \frac{1}{\omega^2 + \Omega^2}
\]
and we have
\[
G(s) = \int d\omega e^{-i\omega s} \frac{1}{\omega^2 + \Omega^2}.
\] (2.1)

So we have:
\[
e^{-S_{\text{eff}}[q]} = e^{-S_{\omega_0}[q]} e^{-\int dt ds \frac{g^2}{2}q(s)^2G(s,t)q(t)^2}.
\]
or taking logs
\[
S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dt ds \frac{g^2}{2}q(s)^2G(s,t)q(t)^2.
\] (2.2)

$Q$ mediates an interaction of four $q$s, 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 $q$s what can it do besides turn back into two $q$s? 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 (2.1):
\[
G(s) \approx \int d\omega e^{-i\omega s} \frac{1}{\omega^2 + \Omega^2} \frac{1}{1 + \frac{\omega^2}{\Omega^2}} \approx \frac{1}{\Omega^2} \delta(s) + \frac{1}{\Omega^4} \partial_s^2 \delta(s) + ...
\]

Plug this back into (2.2):
\[
S_{\text{eff}}[q] = S_{\omega_0}[q] + \int dt \frac{g^2}{2\Omega^2} q(t)^4 + \int dt \frac{g^2}{2\Omega^4} \dot{q}^2 q^2 + ...
\]
The effects of the heavy mode $Q$ are now organized in a \textit{derivative expansion}, with terms involving more derivatives suppressed by more powers of the high energy scale $\Omega$.

\[ A \text{ useful mnemonic for integrating out the effects of the heavy field in terms of Feynman diagrams: to picture } Q \text{ as propagating for only a short time (compared to the external time } t - s), \text{ we can contract its propagator to a point. The first term on the RHS shifts the } q^4 \text{ term, the second shifts the kinetic term, the third involves four factors of } \dot{q}... \]

On the RHS of this equation, we have various interactions involving four $q$s, 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 \textit{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 \textit{more than two derivatives} of $q$. This is OK. We’ve just derived the right way to think about such terms: we treat them as a perturbation, and they are part of a never-ending series of terms which become less and less important for low-energy questions. If we want to ask questions about $q$ at energies of order $\omega$, we can get answers that are correct up to effects of order $(\frac{\omega}{\Omega})^{2n}$ by keeping the $n$th term in this expansion.

Conversely if we are doing an experiment with precision $\Delta$ at energy $\omega$, we can measure the effects of up to the $n$th term, with

\[ (\frac{\omega}{\Omega})^{2n} \sim \Delta. \]

Another important lesson: $S_{\text{eff}}[q]$ contains couplings with negative dimensions of energy

\[ \sum_n c_n (\partial^n q)^2 q^2, \quad \text{with } c_n \sim \frac{1}{\Omega^{2n}}, \]

exactly the situation where the $S$-matrix grows too fast at high energies that we discussed in 215B. 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.

[End of Lecture 2]
2.1.1 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 \(k\), \(\Omega = \sqrt{k^2 + m^2}\). 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 \(q \rightarrow -q\), so there are no odd powers of \(q\)). For more on this, see Physics 217.

The result of that calculation was that fluctuations of \(Q\) mediate various \(q^4\) interactions. It adds to the action for \(q\) the following: \(\Delta S_{\text{eff}}[q] \sim \int dt ds q^2(t)G(t - s)q^2(s)\), as in Fig. 2.3.

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 \(\omega_0\), 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 \(\frac{1}{\Omega}\) (or more accurately, powers of \(\frac{\partial}{\partial \Omega}\)).

(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:

![Graph showing the kernel G(s-t)](image)

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 $M_{\text{Planck}}$, whose role in the toy model is played by $\Omega$. And the natural energy scale of your sandwich is much less than $M_{\text{Planck}}$.

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:

$$G(s) = \int d\omega \frac{e^{-i\omega s}}{\omega^2 + \Omega^2} = e^{-\Omega|s|} \frac{1}{2\Omega}. \quad (2.4)$$

Do it by residues: the integrand has poles at $\omega = \pm i\Omega$. 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 $Q$:

$$Z[J] \equiv \left\langle e^{\int dt Q(t)J(t)} \right\rangle_Q, \quad \langle Q(t_1)Q(t_2)\ldots \rangle_Q = \frac{\delta}{\delta J(t_1)} \frac{\delta}{\delta J(t_2)} \ldots \log Z[J]$$
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 $-\partial_t^2 + \Omega^2$, 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.

### 2.2 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. A. Manohar]

Taking the example of the previous subsection to its logical conclusion, we are led to the idea of an *effective field theory* (EFT), or, *how to do physics without a theory of everything*. (You may notice that all the physics that has been done has been done without a theory of everything.) It is a weaponized version of selective inattention.

The basic idea is that the Hamiltonian (or the action) should contain all terms consistent with symmetries, organized according to an expansion in decreasing relevance to low energy physics. This is an implementation of the *totalitarian principle of physics*, that anything that can happen must happen. (This principle is best understood from the perspective of Wilson’s work on the renormalization group, which is the subject of Physics 217.)

**Diatribe about ‘renormalizability’.** There is no reason to demand that a field theory that we have found to describe physics in some regime should be a valid descrip-
tion 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 $\omega/\Omega$.)

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 $e^{+\frac{\alpha}{137}}$ where $\alpha \sim 1$ 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 $E_{\text{new}}$ 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 case 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 (some of which we will discuss in more detail below) and where you can learn about them:

- Hydrodynamics [Kovtun]
- 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.
2.3 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 $\Delta E$ that it takes to excite atomic energy levels we’ve left out of the discussion. We allow no inelastic scattering. This means we require

$$E_\gamma \ll \Delta E \sim \frac{\alpha}{a_0} \ll a_0^{-1} \ll M_{\text{atom}}$$

(2.5)

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 $v \phi_v$. $v^\mu v_\mu = 1$ and $v_\mu = (1, 0, 0, 0)_\mu$ in the atom’s rest frame. The (Lorentz-singlet) Lagrangian can depend on $v^\mu$. We can write a Lagrangian for the free atoms as

$$L_{\text{atom}} = \phi_v^\dagger v^\mu \partial_\mu \phi_v .$$

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 com is just $\partial_t \phi_{v=(1,0)} = 0$. (If we didn’t define the zero of energy to be at the rest mass, there would be an additional term $M_{\text{atom}} \phi_v^\dagger \phi_v$.) Notice that the kinetic term $\phi_v^\dagger \frac{v^2}{2M_{\text{atom}}} \phi_v$ is a very small correction given our hierarchy of scales (2.5).

So the Lagrangian density is

$$L_{\text{Maxwell}}[A] + L_{\text{atom}}[\phi_v] + L_{\text{int}}[A, \phi_v]$$

and we must determine $L_{\text{int}}$. 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 $F_{\mu \nu} =$
\[ \partial_\mu A_\nu - \partial_\nu A_\mu, \text{ and not } A_\mu \text{ directly, by gauge invariance, because the atom, and hence } \phi_v, \text{ is neutral.} \] 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 \( \phi_v \).) Therefore

\[
L_{\text{int}} = c_1 \phi_v^\dagger \phi_v F_{\mu\nu} F^{\mu\nu} + c_2 \phi_v^\dagger \phi_v v^\sigma F_{\sigma\mu} \phi_v \phi_v (v^\lambda \partial_\lambda) F_{\mu\nu} F^{\mu\nu} + \ldots
\]

\ldots indicates terms with more derivatives and more powers of velocity (\textit{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 \( a_0 \sim (\alpha m_e)^{-1} \) 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 \( \sigma \sim c_1^2 \sim a_0^6 \); 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 \( E_\gamma^2 \) 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 \( \Delta E \). 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.
2.4 Fermi theory of Weak Interactions

[from §5 of A. Manohar’s EFT lectures] As another example of EFT, let’s think about part of the Standard Model.

\[ L_{EW} \supset -\frac{1}{2} \left( \partial_\mu W^+_\nu - \partial_\nu W^+_\mu \right) \left( \partial^\mu W^{-\nu} - \partial^\nu W^{-\mu} \right) + M_W W^+ W^- \]

\[ - \frac{ig}{\sqrt{2}} \bar{\psi}_i \gamma^\mu P_L \psi_j W^+_\mu V_{ij} \quad + \text{terms involving } Z \text{ bosons} \]

Some things intermediate, off-shell W bosons can do: \( \mu \) 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:

\[ \delta S_{\text{eff}} \sim \left( \frac{ig}{\sqrt{2}} \right)^2 V_{ij} V^*_{k\ell} \int \! d^4p \frac{-ig_{\mu\nu}}{p^2 - M_W^2} \left( \bar{\psi}_i \gamma^\mu P_L \psi_j \right) (p) \left( \bar{\psi}_k \gamma^\nu P_L \psi_\ell \right) (-p) \]

(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, the \( W \) carries electric charge, so the charges of \( \bar{\psi}_i \) and \( \psi_j \) in (2.6) must differ by one.) This is non-local at scales \( p \gg M_W \) (recall the discussion of the subsection §2.1). But for \( p^2 \ll M_W^2 \),

\[ \frac{1}{p^2 - M_W^2} \left. \right|_{p^2 \ll M_W^2} \approx - \frac{1}{M_W^2} \left( 1 + \frac{p^2}{M_W^2} + \frac{p^4}{M_W^4} + \ldots \right) \]

\[ S_F = -\frac{4G_F}{\sqrt{2}} V_{ij} V^*_{k\ell} \int \! d^4x \left( \bar{\psi}_i \gamma^\mu P_L \psi_j \right) (x) \left( \bar{\psi}_k \gamma^\nu P_L \psi_\ell \right) (x) + \mathcal{O} \left( \frac{1}{M_W^2} \right) + \text{kinetic terms for fermions} \]

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 (2.8) 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 ...

### 2.5 Loops in EFT

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 arises by asking about radiative corrections in the 4-Fermi theory to the coupling between the fermions and the photon (or the Z boson).

We are just trying to estimate the magnitude of this correction, so don’t worry about the factors and the gamma matrices:

$$\sim I \equiv \frac{1}{M_W^2} \int_0^\Lambda \frac{d^4k}{k^2} \frac{1}{k^2} \text{tr}(\gamma\ldots) \sim \mathcal{O}(1).$$

Even worse, consider what happens if we use the vertex coming from the $(p^2/M_W^2)\ell$ correction in (2.7)

$$\sim I_\ell \equiv \frac{1}{M_W^2} \int_0^\Lambda \frac{d^4k}{k^2} \frac{1}{k^2} \left(\frac{k^2}{M_W^2}\right)^\ell \sim \mathcal{O}(1)$$

– 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 (MS). The crucial feature is that the dimensionful cutoff parameter appears only inside logarithms ($\log \mu$), and not as free-standing powers ($\mu^2$).
With such a scheme, we’d get instead

\[ I \sim \frac{m^2}{M_W^2} \log \mu \quad I_\ell \sim \left( \frac{m^2}{M_W^2} \right)^{\ell+1} \log \mu \]

where \( m \) is some mass scale other than the RG scale \( \mu \) (like a fermion mass parameter, or an external momentum, or a dynamical scale like \( \Lambda_{QCD} \)).

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 \( 2\ell \) 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.)

### 2.5.1 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
\[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \bar{\psi} (i\not{D} - m) \psi\]
with $D_\mu = \partial_\mu - ieA_\mu$. 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 \textit{either} in the vacuum polarization, the correction to the photon propagator:
\[\cdots\]. I will call this diagram $i\Pi_{\mu\nu}$.

So the information about the running of the coupling is encoded in the gauge field two-point function:
\[
\Pi_{\mu\nu} \equiv \langle A_\mu(p) A_\nu(q) \rangle = \left( p_\mu p_\nu - p^2 g_{\mu\nu} \right) \delta(p + q) \Pi(p^2) .
\]
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 $p^\mu$ which satisfies this. This determines the correlator up to a function of $p^2$, which we have called $\Pi(p^2)$.

The choice of scheme shows up in our choice of renormalization condition to impose on $\Pi(p^2)$:

**Mass-dependent scheme:** subtract the value of the graph at $p^2 = -M^2$ (a very off-shell, euclidean, momentum). That is, we impose a renormalization condition which says
\[
\Pi(p^2 = -M^2) \overset{!}{=} 1 \quad (2.9)
\]
(which is the tree-level answer with the normalization above).

The contribution of a fermion of mass $m$ and charge $e$ is (factoring out the momentum-conserving delta function):
\[
p_{\mu} \circ p_{\nu} = - \int d^Dk \text{tr} \left( (-ie\gamma^\mu) \frac{-i(k + m)}{k^2 - m^2} (-ie\gamma^\nu) \frac{-i(p + k + m)}{(p + k)^2 - m^2} \right)
\]
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

$$i\Pi^{\mu\nu} = -e^2 \int d^D \ell \int_0^1 dx \frac{N^{\mu\nu}}{(\ell^2 - \Delta)^2}$$

with $\ell = k + xp$ is a new integration variable, $\Delta \equiv m^2 - x(1-x)p^2$, and the numerator is

$$N^{\mu\nu} = 2\ell^{\mu} \ell^{\nu} - g^{\mu\nu} \ell^2 - 2x(1-x)p^{\mu} p^{\nu} + g^{\mu\nu} \left( m^2 + x(1-x)p^2 \right) + \text{terms linear in } \ell^{\mu} .$$

In dim reg, the one-loop vacuum polarization correction satisfies the gauge invariance 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:

$$\delta \Pi_2(p^2)^{\text{Peskin p. 252}} = -\frac{8e^2}{(4\pi)^{D/2}} \int_0^1 dx x(1-x) \frac{\Gamma(2 - D/2)}{\Delta^{2-D/2}} \frac{\ell^2}{\bar{\mu}^2}$$

$$\left[ \frac{D}{\rightarrow 4} \right] = -\frac{e^2}{2\pi^2} \int_0^1 dx (1-x) \left( \frac{2}{\epsilon} - \log \left( \frac{\Delta}{\mu^2} \right) \right) \quad (2.10)$$

where we have introduced the heralded $\mu$:

$$\mu^2 \equiv 4\pi \bar{\mu}^2 e^{-\gamma_E}$$

where $\gamma_E$ is the Euler-Mascheroni constant. In the second line of (2.10), we expanded the $\Gamma$-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:

$$0 \overset{(2.9)!}{=} \Pi_2^{(M)}(p^2 = -M^2) = \underbrace{\delta^{(M)}_{F^2}}_{\text{counterterm coefficient for } \frac{1}{4} F_{\mu\nu} F^{\mu\nu}} + \delta \Pi_2$$

$$\implies \Pi_2^{(M)}(p^2) = \frac{e^2}{2\pi^2} \int dx x(1-x) \log \left( \frac{m^2 - x(1-x)p^2}{m^2 + x(1-x)M^2} \right).$$

Notice that the $\mu$s 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 $\Pi$ is defined by the rule that we subtract the $1/\epsilon$ pole. This means that the counterterm is

$$\delta^{(\overline{\text{MS}})}_{F^2} = -\frac{e^2}{2\pi^2} 2 \int_0^1 dx x(1-x).$$

$$=1/6$$
(Confession: I don’t know how to state this in terms of a simple renormalization condition on $\Pi_2$. Also: the bar in $\bar{\text{MS}}$ refers to the (not so important) distinction between $\bar{\mu}$ and $\mu$. The resulting vacuum polarization function is

$$\Pi_2^{(\bar{\text{MS}})}(p^2) = \frac{e^2}{2\pi^2} \int_0^1 dx x (1-x) \log \left( \frac{m^2 - x(1-x)p^2}{\mu^2} \right).$$

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.** First in the mass-dependent scheme. The fermion contribution to the beta function for the EM coupling is

$$\beta_e^{(M)} = \frac{e}{2} M \partial_M \Pi_2(p^2) = -\frac{1}{2} \left( \frac{e^3}{2\pi} \right) \int_0^1 dx x (1-x) \left( \frac{-2M^2 x (1-x)}{m^2 + M^2 x (1-x)} \right) + \mathcal{O}(e^5)
\begin{cases}
\text{if } m \ll M, \quad \frac{e^3}{2\pi} \int_0^1 dx x (1-x) = \frac{e^3}{12\pi^2}, \\
\text{if } m \gg M, \quad \frac{e^3}{2\pi} \int_0^1 dx x (1-x) \frac{M^2 x (1-x)}{m^2} = \frac{e^3}{60\pi^2} \frac{M^2}{m^2}.
\end{cases}
$$

$$\beta_e^{(\bar{\text{MS}})} = \frac{e}{2} \mu \partial_\mu \Pi_2(p^2) = -\frac{1}{2} \frac{e^3}{2\pi^2} \int_0^1 dx x (1-x) \mu \partial_\mu \log \left( \frac{m^2 - p^2 x (1-x)}{\mu^2} \right)
\begin{cases}
= 1/6, \\
= -2
\end{cases}
= \frac{e^3}{12\pi^2}. \quad (2.12)

Also, the $\bar{\text{MS}}$ vacuum polarization behaves for small external momenta like

$$\Pi_2(p^2 \ll m^2) \simeq -\frac{e^3}{2\pi^2} \int_0^1 dx x (1-x) \log \frac{m^2}{\mu^2}
\begin{cases}
\gg 1, \text{for } \mu \ll m! \text{ bad!}
\end{cases}$$
Figure 2: 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.

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.
2.6 The Standard Model as an EFT.

The Standard Model. [Schwartz, §29]

<table>
<thead>
<tr>
<th></th>
<th>$L = (\nu_L, e_L)$</th>
<th>$e_R$</th>
<th>$\nu_R$</th>
<th>$Q = (u_L, d_L)$</th>
<th>$u_R$</th>
<th>$d_R$</th>
<th>$H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SU(3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>-</td>
</tr>
<tr>
<td>SU(2)</td>
<td>☐</td>
<td>-</td>
<td>-</td>
<td>☐</td>
<td>-</td>
<td>-</td>
<td>☐</td>
</tr>
<tr>
<td>U(1)$_Y$</td>
<td>$-\frac{1}{2}$</td>
<td>-1</td>
<td>0</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{2}{3}$</td>
<td>$-\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
</tr>
</tbody>
</table>

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, $\nu_R$.)

Whence the values of the charges under the 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 $p = \epsilon_{ijk}u_iu_jd_k$ have exactly opposite charges of the same magnitude.

The Lagrangian is just all the terms which are invariant under the gauge group $SU(3) \times SU(2) \times U(1)$ 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

$$SU(2) \times U(1)_Y \xrightarrow{\langle H \rangle} U(1)_{EM}.$$

The broken gauge bosons get masses from the Higgs kinetic term

$$|D_\mu H|^2 \left|_{H=\left(\begin{array}{c} 0 \\ v/\sqrt{2} \end{array}\right)}\right. \quad \text{with} \quad D_\mu H = \left(\partial_\mu - igW^a_\mu \tau^a - \frac{1}{2}igY_\mu\right)H$$

where $Y_\mu$ is the hypercharge gauge boson, and $W^a, a = 1, 2, 3$ are the SU(2) gauge bosons. There are two massive $W$-bosons with electric charge $\pm 1$ (as described in §2.4), with $M_W = \frac{v_0}{\sqrt{2}}$. The photon and $Z$ boson are the linear combinations of $Y$ and $W^3$ which diagonalize the remaining mass terms:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_w & \sin \theta_w \\ -\sin \theta_w & \cos \theta_w \end{pmatrix} \begin{pmatrix} W^3_\mu \\ Y_\mu \end{pmatrix}.$$
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

\[
L_{\text{Yukawa}} = -Y_{ij}^L \bar{L}_i H_R^j - Y_{ij}^u \bar{Q}_i H_d^j - Y_{ij}^d \bar{Q}_i (i\tau^2 H^*) u_R^j + h.c.
\]

The contortion with the \( \tau^2 \) 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 \( m_e = yev/\sqrt{2} \). There’s lots of drama about the matrices \( Y \) which can mix the generations. The mass for the \( \nu_R \) (which maybe could not exist – it doesn’t have any charges at all) you can figure out on the homework later.

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:

\[
\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)_Y \subset \text{SU}(5) \subset \text{SO}(10)
\]

\[
\text{one generation} = 10 \oplus 5 \oplus 1 = 16
\]

The singlet of \( \text{SU}(5) \) is the right-handed neutrino, but if we include it, one generation is an irreducible (spinor) representation of \( \text{SO}(10) \). 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 \( M_{\text{GUT}} \sim 10^{16}\text{GeV} \).) 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

\[
\delta L = \frac{1}{M^2} (\bar{\psi} A \psi) \cdot (\bar{\psi} B \psi)
\]

(where \( \psi \) represent various SM fermion fields and \( A, B \) can be various gamma and flavor matrices) with \( M \lesssim 10 \text{ TeV} \). Notice that I am talking now about interactions other than the electroweak interactions, which as we’ve just discussed, for energies above \( M_W \sim 80\text{GeV} \) cannot be treated as contact interactions – you can see the W’s 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 \( \leq 4 \)) 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 \( \theta \) 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 \( \Lambda_{\text{new}} \), so we can integrate it out and include its effects by corrections to the Standard Model:

\[
L = L_0 + \frac{1}{\Lambda_{\text{new}}} \mathcal{O}^{(5)} + \frac{1}{\Lambda_{\text{new}}^2} \sum_i c_i \mathcal{O}_i^{(6)} + \cdots
\]

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 SU(2)$_{\text{EW}}$ Higgs doublet and \( L^i = (\nu_L, e_L)^i \) is an SU(2)$_{\text{EW}}$ doublet of left-handed leptons, and \( L^c \equiv L^TC \) 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 \( L \rightarrow e^{i\alpha}L, Q \rightarrow Q, H \rightarrow H \).

At dimension 6, there are operators that directly violate baryon number, such as

\[
\epsilon_{\alpha\beta\gamma} (\bar{u}_R)^\alpha (u_R)^\beta (\bar{u}_R)^\gamma e_R.
\]

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 \( eqqq \) 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 \( \frac{1}{M_{\text{Planck}}^2} p\mathcal{O} \) would cause proton decay (into whatever \( \mathcal{O} \) makes). This predicts \( \Gamma_p \sim \frac{m_p^3}{M_{\text{Planck}}^2} \sim 10^{-13} \text{s}^{-1} \) 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 \((L \rightarrow L, Q \rightarrow e^{\imath\alpha} Q, H \rightarrow H)\). This is an emergent symmetry, expected to be violated by the UV completion.

• Surely nothing can prevent \( \Delta L \sim \left(\frac{1}{M_{\text{Planck}}} \right)^2 qqq\ell \). Happily, this is consistent with the observed proton lifetime.

There are \( \sim 10^2 \) dimension 6 operators that preserve baryon number, and therefore are not as tightly constrained\(^3\). (Those that induce flavor-changing processes in the SM are more highly constrained and must have \( \Lambda_{\text{new}} > 10^4 \text{ TeV} \).) 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\(^4\). [End of Lecture 4]

2.7 Superconductors

Recall from 215B our effective (Landau-Ginzburg) description of superconductors which reproduces the Meissner effect, the Abelian Higgs model:

\[
\mathcal{F} = \frac{1}{4} F_{ij} F_{ij} + |D_i \Phi|^2 + a |\Phi|^2 + \frac{1}{2} b |\Phi|^4 + \ldots
\]  

with \( D_i \Phi \equiv (\partial_i - 2\imath e A_i) \Phi \).

I want to make two more comments about this:

\(^3\)Recently, humans have gotten better at counting these operators. See this paper.
\(^4\)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”.

41
Symmetry breaking by fluctuations (Coleman-Weinberg). [Zee problem IV.6.9.] What happens near the transition, when $a = 0$ in (2.13)? 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 $\psi_\alpha$, $\alpha = \uparrow, \downarrow$.

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

$$\Phi \sim \psi_\alpha \psi_\beta \epsilon^{\alpha\beta}; \quad (2.14)$$

$\Phi$ 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)

$$S[\psi] = S_2[\psi] + \int dt d^d x \, u \psi^\dagger \psi \psi^\dagger \psi + h.c. \quad (2.15)$$

where

$$S_2 = \int dt \int d^d k \psi^\dagger_k \left( i \partial_t - \epsilon(k) \right) \psi_k.$$

Spin is important here so that $\psi_{\uparrow}^\dagger \psi_{\uparrow}^\dagger \psi_{\uparrow} \psi_{\uparrow}$ is nonzero. A mean field theory description of the condensation of Cooper pairs (2.14) is obtained by replacing the quartic term in (2.15) by expectation values:

$$S_{MFT}[\psi] = S_2[\psi] - \int dt d^d x \, u \langle \psi \psi \rangle \psi^\dagger \psi^\dagger + h.c.$$

$$= S_2[\psi] - \int dt d^d x \, u \Phi \psi^\dagger \psi^\dagger + h.c. \quad (2.16)$$

So an expectation value for $\Phi$ is a mass for the fermions. It is a funny kind of symmetry-breaking mass, but if you diagonalize the quadratic operator in (2.16) (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.

### 2.7.1 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 (2.15) 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:

$$e^{-iux^4} = \frac{1}{\sqrt{i\pi u}} \int_{-\infty}^{\infty} d\sigma \ e^{-\frac{1}{iu} \sigma^2 - 2iux^2 \sigma}.$$  \hspace{1cm} (2.17)

At the cost of introducing an extra field $\sigma$, we turn a quartic term in $x$ into a quadratic term in $x$. The RHS of (2.17) is gaussian in $x$ and we know how to integrate it over $x$. (The version with $i$ is relevant for the real-time integral.) Notice the weird extra factor of $i$ lurking in (2.17). This can be understood as arising because we are trying to use a scalar field $\sigma$, 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:

$$e^{-iux^2x^2} = \frac{1}{1\pi u} \int_{\mathbb{C}} d^2\sigma \ e^{-\frac{1}{iu} |\sigma|^2 - iux^2 \sigma - i\bar{x}^2 \sigma}, \hspace{1cm} (2.18)$$

where $\int_{\mathbb{C}} d^2\sigma ... \equiv \int_{-\infty}^{\infty} d\text{Re} \sigma \int_{-\infty}^{\infty} d\text{Im} \sigma ...$ (The field-independent prefactor is, as usual, not important for path integrals.)

We can use a field theory generalization of (2.18) to ‘decouple’ the 4-fermion interaction in (2.15):

$$Z = \int [D\psi D\psi^\dagger] e^{iS[\psi]} = \int [D\psi D\psi^\dagger D\sigma D\sigma^\dagger] e^{iS_2[\psi] + i \int d^Dx (\sigma \psi \psi^\dagger + h.c.) - \int d^Dx |\sigma|^2(x) \frac{1}{iu}} \hspace{1cm} (2.19)$$

The point of this is that now the fermion integral is gaussian. At the saddle point of the $\sigma$ integral (which is exact because it is gaussian), $\sigma$ is the Cooper pair field, $\sigma_{\text{saddle}} = u\psi\psi^\dagger$.

Notice that we made a choice here about in which ‘channel’ to make the decoupling – we could have instead introduced a different auxiliary field $\rho$ 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 $\psi^2$). 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 (2.19)? Let’s study the case where the single-fermion dispersion is $\epsilon(k) = \frac{k^2}{2m} - \mu$.

$$I_\psi[\sigma] \equiv \int [D\psi D\psi^\dagger] e^{\int dt dx \left( \bar{\psi}(i\partial_t - \frac{\nabla^2}{2m} - \mu)\psi + \bar{\psi}\sigma\psi + \bar{\psi}\psi^\dagger\sigma \right)}$$

The action here can be written as the integral of

$$L = \left( \bar{\psi} \psi \right) \begin{pmatrix} i\partial_t - \epsilon(-i\nabla) & \sigma \\ \sigma & -i\partial_t - \epsilon(-i\nabla) \end{pmatrix} \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix} \equiv \left( \bar{\psi} \psi \right) M \left( \psi \bar{\psi} \right)$$

so the functional integral is

$$I_\psi[\sigma] = \det M = e^{\text{tr} \log M(\sigma)}.$$

If $\sigma$ is constant (which will lower the energy), the matrix $M$ is diagonal in momentum space, and the integral remaining to be done is

$$\int [D\sigma D\sigma^\dagger] e^{-\int d^D x \frac{[\epsilon(\sigma)]^2}{2m} + \int d^D k \log \left( \omega^2 - \epsilon_k^2 - |\sigma|^2 \right)}.$$

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 $\sigma$, which as we saw earlier determines the fermion gap.

$$0 = \delta\text{exponent} = \frac{\delta\sigma}{2u} + \int d\omega d^D k \left( \frac{2\sigma}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + i\epsilon} \right).$$

We can do the frequency integral by residues:

$$\int d\omega \frac{1}{\omega^2 - \epsilon_k^2 - |\sigma|^2 + i\epsilon} = \frac{1}{2\pi} \frac{2\pi i}{2\sqrt{\epsilon_k^2 + |\sigma|^2}}.$$

The resulting equation is naturally called the gap equation:

$$1 = -2u \int d^D k' \frac{1}{\sqrt{\epsilon(k')^2 + |\sigma|^2}}$$

which you can imagine solving self-consistently for $\sigma$. Plugging back into the action (2.19) says that $\sigma$ determines the energy cost to have electrons around; more precisely, $\sigma$ is the energy required to break a Cooper pair.

Comments:
• Notice that a solution of (2.20) 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

\[
\sigma(p) = -\frac{1}{2} \int d^d p' \frac{u(p, p')\sigma(p')}{\sqrt{c(p')^2 + |\sigma(p')|^2}}.
\]

• 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 \( \sigma \) 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 \( \psi \psi^\dagger \psi^\dagger \psi \) into *e.g.* \( \psi \psi \sigma \) or \( \psi^\dagger \psi \rho \). BCS wins. This is explained beautifully in Polchinski, lecture 2, and R. Shankar. I will summarize the EFT framework for understanding this in §2.8.

• 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 \( \sigma \) 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 \( N \).

2.8 Effective field theory of Fermi surfaces

[Polchinski, lecture 2 (I recommend these notes very strongly), 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

\[ E_0 = \frac{1}{2} \frac{e^4 m}{\hbar^2} = \frac{1}{2} \frac{e^2}{a_0} \approx 13 \text{eV} \]

(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 \( \infty \) to first approximation. \( v_F/c \) suppresses spin orbit couplings (though large atomic numbers enhance them: \( \lambda_{\text{SO}} \propto Z v_F/c \)).

Let us attempt to construct a Wilsonian-natural effective field theory of this phenomenon. The answer is called Landau Fermi Liquid Theory. What are the right low-energy degrees of freedom? Let’s make a guess that they are like electrons – fermions with spin and electric charge. They will not have exactly the properties of free electrons, since they must incorporate the effects of interactions with all their friends. The ‘dressed’ electrons are called quasielectrons, or more generally quasiparticles.

Given the strong interactions between so many particles, why should the dofs have anything at all to do with electrons? Landau’s motivation for this description (which is not always correct) is that we can imagine starting from the free theory and adiabatically turning up the interactions. If we don’t encounter any phase transition along the way, we can follow each state of the free theory, and use the same labels in the interacting theory.

We will show that there is a nearly-RG-stable fixed point describing gapless quasielectrons. 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:

\[
S_{\text{free}}[\psi] = \int dt \, d^d p \left( i \psi_\sigma^\dagger(p) \partial_t \psi_\sigma(p) - (\epsilon(p) - \epsilon_F) \psi_\sigma^\dagger(p) \psi_\sigma(p) \right)
\]

where \( \sigma \) is a spin index, \( \epsilon_F \) 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 $p$. \(^5\) \(^6\)

The groundstate of the free theory is the filled Fermi sea:

$$|gs\rangle = \prod_{p|\epsilon(p) < \epsilon_F} \psi_p^\dagger |0\rangle, \quad \psi_p |0\rangle = 0, \quad \forall p.$$ 

(If you don’t like continuous products, put the system in a box so that $p$ is a discrete label.) The Fermi surface is the set of points in momentum space at the boundary of the filled states:

$$FS \equiv \{ p|\epsilon(p) = \epsilon_F \}.$$ 

The low-lying excitations are made by adding an electron just above the FS or removing an electron (creating a hole) just below.

In order to define the power-counting rules for our EFT, we would like to define a scaling transformation which focuses on the low-energy excitations. We scale energies by a factor $E \rightarrow bE, b < 1$. In relativistic QFT, $\vec{p}$ scales like $E$, toward zero, $\vec{p} \rightarrow b\vec{p}$, 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.

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 $\ell$. (Clearly there are some exceptional cases if the FS gets too wiggly. Ignore these for now.)

$$\epsilon(p) - \epsilon_F = \ell v_F(\vec{k}) + \mathcal{O}(\ell^2), \quad v_F \equiv \partial_\ell \epsilon|_{p=\vec{k}}.$$ 

\(^5\)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. 

\(^6\)We have chosen the normalization of $\psi$ to fix the coefficient of the $\partial_t$ term (this rescaling may depend on $p$).
So a scaling rule which accomplishes our goal of focusing on the FS is 
\[ E \to bE, \quad \vec{k} \to \vec{k}, \quad \vec{l} \to b\vec{l}. \]
This implies 
\[ dt \to b^{-1}dt, \quad d^{d-1}\vec{k} \to d^{d-1}\vec{k}, \quad d\vec{l} \to bd\vec{l}, \quad \partial_t \to b\partial_t. \]
\[ S_{\text{free}} = \int \left( i\psi^\dagger(p) \overleftrightarrow{\partial_p} \psi(p) - \ell v_F(k) \psi^\dagger(p)\psi(p) \right) \sim b^0 \sim b^{-\frac{1}{2}} \sim b^{-\frac{1}{2}} \]
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, \( \psi \to e^{i\theta}\psi \)

2. Spatial symmetries: time-translation invariance, and 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, \( SU(n) \) 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 \( \psi \) (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 \( \psi \)) and consistent with the symmetries; we can order them by the number of fermion operators involved. Particle number symmetry means every \( \psi \) comes with a \( \psi^\dagger \). The possible quadratic terms are:
\[ \int \left( \mu(k) \psi^\dagger_\sigma(p)\psi_\sigma(p) \right) \sim b^{-1} \]
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 $\partial_t$ or factor of $\ell$ costs a $b^1$ and makes the operator marginal; those terms are already present in $S_{\text{free}}$. Adding more than one makes it irrelevant.

**Quartic terms:**

$$S_4 = \int dt \prod_{i=1}^4 d^{d-1}k_i d\vec{\ell}_i \left( u(4 \cdot 1) \psi_\sigma(p_1) \psi_\sigma(p_3) \psi_{\sigma'}(p_2) \psi_{\sigma'}(p_4) \delta^d(\vec{p}_1 + \vec{p}_2 - \vec{p}_3 - \vec{p}_4) \right)$$

The minus signs on $p_{3,4}$ is because $\psi(p)$ removes a particle with momentum $p$. We assume $u$ depends only on $k, \sigma$, so does not scale – this will give the most relevant piece. How does the delta function scale?

$$\delta^d(\vec{p}_1 + \vec{p}_2 - \vec{p}_3 - \vec{p}_4) = \delta^d(k_1 + k_2 - k_3 - k_4 + \ell_1 + \ell_2 - \ell_3 - \ell_4) \approx \delta^d(k_1 + k_2 - k_3 - k_4)$$

In the last (questioned) step, we used the fact that $\ell \ll k$ to ignore the contributions of the $\ell$s. If this is correct then the delta function does not scale (since $k$s 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. [End of Lecture 5]

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), $\delta k_1$ and $\delta k_2$ 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 \( \epsilon \) 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 \( \ell \)s 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 (nesting) 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 $S_{\text{free}}$, is

$$ G(\epsilon,p = k + l) = \frac{i}{\epsilon(1 + i\eta) - v_F(k)\ell + \mathcal{O}(\ell)^2} $$

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

$$ G(t,p) = \langle \epsilon_F | c_p^\dagger(t)c_p(0) | \epsilon_F \rangle, \quad c_p^\dagger(t) \equiv e^{-iHt}c_p^\dagger e^{iHt} $$

and use the free-fermion fact $[H, c_p^\dagger] = \epsilon_p c_p^\dagger$.)

Let’s assume rotation invariance. Then $V(\theta_3, \theta_1) = V(\theta_3 - \theta_1)$, $V_l = \int d\theta e^{i\theta} V(\theta)$. 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:

$$ -i\delta^{(1)} V = t \sigma \cdot \epsilon = \int_{\text{doIntegration}} (-iV)^2 \int_{b_0}^{\epsilon_0} \int_0^{d} \frac{d\epsilon d^{d-1}k' d\ell'}{(2\pi)^{d+1}} \frac{i^2}{(\epsilon + \epsilon' - v_F(k')\ell') (\epsilon - \epsilon' - v_F(k')\ell')} $$

By residues

$$ = V^2 \int_{b_0}^{\epsilon_0} \frac{d\epsilon' d^{d-1}k'}{(2\pi)^{d+1}} \frac{2\pi i}{v_F(k')} \left( \frac{\epsilon - \epsilon' - (\epsilon + \epsilon')}{-2\epsilon'} \right)^{-1} $$

$$ = +i \frac{V^2}{2} \int_{b_0}^{\epsilon_0} \frac{d\epsilon'}{\epsilon'} \int_{\text{dos at FS}}^{d} \frac{d^{d-1}k'}{(2\pi)^{d+1} v_F(k')} $$

Between the first and second lines, we did the $\ell'$ 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 $\ell$ 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

$$ b \frac{d}{db} V(b) = \beta_V = NV^2(b) + \mathcal{O}(V^3) $$

51
and the solution of the flow equation at $E = bE_1$ is

$$V(E) = \frac{V_1}{1 + NV_1 \log(E_1/E)} \begin{cases} 
\to 0 & \text{in IR for } V_1 > 0 \text{ (repulsive)} \\
\to -\infty & \text{in IR for } V_1 < 0 \text{ (attractive)}
\end{cases} \quad (2.22)$$

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 $\sigma$ (‘the BCS channel’) wins over the decoupling with $\rho$ (‘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 §2.7.1.

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\(^7\). The Goldstone theorem is satisfied by including a field

$$\vec{D} \propto \text{(local) displacement } \delta\vec{r} \text{ of ions from their equilibrium positions}$$

Most microscopically we have a bunch of coupled springs:

$$L_{\text{ions}} \sim \frac{1}{2} M \left( \dot{\delta}\vec{r} \right)^2 - k_{ij} \delta r^i \delta r^j + \ldots$$

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

$$S[\vec{D}] = (M)^{1/2} \delta\vec{r} = \frac{1}{2} \int dt d^d q \left( \partial_t D_i(q) \partial_t D_i(-q) - \omega^2_{ij}(q) D_i(q) D_j(-q) \right).$$

Here $\omega^2 \propto M^{-1}$. Their status as Goldstons means that the eigenvalues of $\omega^2_{ij}(q) \sim |q|^2$ 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.

\(^7\)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.
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
\[ dtq^4 (\partial_t D)^2 \sim b^{+1+2|D|} \implies D \sim b^{-\frac{1}{2}} \]
and the restoring force $dtqD^2\omega^2(q) \sim b^{-2}$ is relevant, and dominates over the $\partial_t^2$ term for
\[ E \leq E_D = \sqrt{\frac{m}{M}} E_0 \quad \text{the Debye energy.} \]
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?
\[
S_{\text{int}}[D, \psi] = \int dtq^3q^2k_1d\ell_1d^2k_2d\ell_2 M^{-\frac{1}{2}} g_i(q, k_1, k_2)D_i(q)\psi_\sigma^+(p_1)\psi_\sigma(p_2)\delta^3(p_1 - p_2 - q) \\
\sim b^{-1+1+1-3/2} = b^{-1/2} \tag{2.23}
\]
– 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 $-g^2$: \[ \xrightarrow{\text{phonons}} \]
Notice that the scale at which the coupling $V$ becomes strong ($V(E_{\text{BCS}}) \equiv 1$ in (2.22)) is
\[ E_{\text{BCS}} \sim E_D e^{-\frac{1}{NVD}}. \]
Two comments about this: First, it is non-perturbative in the interaction $V_D$. Second, it provides some verification of the role of phonons, since $E_D \sim M^{-1/2}$ can be varied by studying the same material with different isotopes and studying how the critical superconducting temperature ($\sim E_{\text{BCS}}$) scales with the nuclear mass.

Here’s the narrative, proceeding as a function of decreasing energy scale, beginning at $E_0$, the Planck scale of solids: (1) Electrons
repel each other by the Coulomb interaction. However, in a metal, this interaction is screened by processes like this: 

(the intermediate state is an electron-hole pair) and is short-ranged. It is still repulsive, however. As we coarse-grain more and more, we see more and more electron-hole pairs and the force weakens. (2) While this is happening, the electron-phonon interaction is relevant and growing. This adds an attractive bit to $V$. This lasts until $E_D$. (3) At $E_D$ the restoring force term in the phonon lagrangian dominates (for the purposes of their interactions with the electrons) and we can integrate them out. (4) What happens next depends on the sign of $V(E_D)$. If it’s positive, $V$ flows harmlessly to zero. If it’s negative, it becomes moreso until we exit the perturbative analysis at $E_{BCS}$, and vindicate our choice of Hubbard-Stratonovich channel above.

Further brief comments, for which I refer you to Shankar:

1. Putting back the possible angular dependence of the BCS interaction, the result at one loop is

$$\frac{dV(\theta_1 - \theta_3)}{d\ell} = -\frac{1}{8\pi^2} \int_0^{2\pi} d\theta V(\theta_1 - \theta)V(\theta - \theta_3)$$

or in terms of angular momentum components,

$$\frac{dV_l}{d\ell} = -\frac{V_l^2}{4\pi}.$$

2. This example is interesting and novel in that it is a (family of) fixed point(s) characterized by a dimensionful quantity, namely $k_F$. This leads to a phenomenon called *hyperscaling violation* where thermodynamic quantities need not have their naive scaling with temperature.

3. The one loop analysis gives the right answer to all loops in the limit that $N \equiv k_F/\Lambda \gg 1$, where $\Lambda$ 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 $\phi^4$ 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 $m^*$.
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 $i\eta$ prescription. After
doing the frequency integrals by residues, we get

$$\Sigma(k, \epsilon) = \int dq dk' \frac{|u_q|^2}{D - i\eta}$$

$$D \equiv \epsilon_k(1 + i\eta) + \epsilon_{k'}(1 - i\eta) - \epsilon_{k' + q}(1 + i\eta) - \epsilon_{k - q}(1 + i\eta)$$

(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 $k$:

$$\tau^{-1}(k) = \text{Im} \Sigma(k, \epsilon) = \pi \int dq dk' \delta(D)|u_q|^2 f(-\epsilon_{k'}) f(\epsilon_{k' + q}) f(\epsilon_{k - q})$$

where

$$f(\epsilon) = \lim_{T\to 0} \frac{1}{e^{\frac{-\epsilon}{T}} + 1} = \theta(\epsilon < \epsilon_F)$$

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
$\epsilon$ 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.
3 Topological terms in field theory actions

3.1 Coherent state path integral for fermions

[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:

\[
\{ c, c \} = 0, \quad \{ c^\dagger, c^\dagger \} = 0, \quad \{ c, c^\dagger \} = 1.
\]

With a single mode, the general Hamiltonian is

\[
H = c^\dagger c (\omega_0 - \mu)
\]

(\(\omega_0\) and \(\mu\) are (redundant when there is only one mode) constants). This algebra is represented on a two-state system |1\rangle = c^\dagger |0\rangle. We might be interested in its thermal partition function

\[
Z = \text{tr} e^{-\frac{H}{T}}.
\]

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

Let’s define coherent states for fermionic operators:

\[
c |\psi\rangle = \psi |\psi\rangle.
\]

(3.1)

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

\[
\psi_1 \psi_2 = -\psi_2 \psi_1, \quad c \psi = -\psi c
\]

– 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

\footnote{\text{For many modes,}}

\[
\{ c_i, c_j \} = 0, \quad \{ c_i^\dagger, c_j^\dagger \} = 0, \quad \{ c_i, c_j^\dagger \} = \delta_{ij}.
\]
consider multiple Grassmann numbers when we have more than one fermion mode, where \( \{c_1, c_2\} = 0 \) will require that they anticommute \( \{\psi_1, \psi_2\} = 0 \) (as in the definition (??)); note that we will be simultaneously diagonalizing operators which anticommute.

The solution to equation (??) is very simple:

\[
|\psi\rangle = |0\rangle - \psi |1\rangle
\]

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

Similarly, the left-eigenvector of the creation operator is

\[
\langle \bar{\psi} | = \langle \bar{\psi} | c, \quad \langle \bar{\psi} | = \langle 0| - \langle 1| \bar{\psi} = \langle 0| + \bar{\psi} \langle 1|.
\]

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

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

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

\[
f(\psi) = f_0 + f_1 \psi.
\]

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

\[
\int \psi d\psi = 1, \quad \int 1 d\psi = 0.
\]

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

\[
1 = \int \bar{\psi} \psi d\psi d\bar{\psi} = - \int \bar{\psi} \psi d\bar{\psi} d\psi.
\]

The only integral, really, is the Gaussian integral:

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

Many of these give

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

$$\langle \bar{\psi} \psi \rangle \equiv \int \bar{\psi}_1 \psi_1 e^{-\bar{\psi}_1 \psi_1} d\bar{\psi}_1 d\psi_1 = -\frac{1}{a} = -\langle \bar{\psi} \psi \rangle.$$

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

$$\langle \bar{\psi}_i \psi_j \rangle = \frac{\delta_{ij}}{a_i} \equiv \langle i j \rangle \quad (3.3)$$

and Wick’s theorem here is

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

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

$$\mathbb{I} = \int d\bar{\psi} d\psi \ e^{-\bar{\psi} \psi} |\psi\rangle \langle \bar{\psi}| \quad (3.4)$$

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

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

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

$$Z = \int d\bar{\psi}_0 d\psi_0 \ e^{-\bar{\psi}_0 \psi_0} \langle -\bar{\psi}_0 | \ e^{-\frac{H}{\mathcal{M}}} | (1 - \Delta \tau H) \cdots (1 - \Delta \tau H) \rangle$$

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

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

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

$$\bar{\psi}_M = -\bar{\psi}_0, \quad \psi_M = -\psi_0 \quad (3.5)$$

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:

\[
\langle \bar{\psi}_{l+1} | (1 - \Delta \tau H(c^\dagger, c)) | \psi_l \rangle = \langle \bar{\psi}_{l+1} | (1 - \Delta \tau H(\bar{\psi}_{l+1}, \psi_l)) | \psi_l \rangle \xrightarrow{\Delta \tau \to 0} e^{\bar{\psi}_{l+1}\psi_l} e^{-\Delta \tau H(\bar{\psi}_{l+1}, \psi_l)}.
\]

It was important that in \( H \) all \( c \)'s were to the right of all \( c^\dagger \)'s, i.e. that \( H \) was normal ordered.

So we have

\[
Z = \int \prod_{l=0}^{M-1} d\bar{\psi}_l d\psi_l e^{-\bar{\psi}_l \psi_l} e^{-\Delta \tau H(\bar{\psi}_{l+1}, \psi_l)}
\]

\[
= \int \prod_{l=0}^{M-1} d\bar{\psi}_l d\psi_l \exp \left( \Delta \tau \left( \frac{\bar{\psi}_{l+1} - \bar{\psi}_l}{\Delta \tau} \psi_l - H(\bar{\psi}_{l+1}, \psi_l) \right) = \partial_{\tau} \psi \right)
\]

\[
\simeq \int [D\bar{\psi} D\psi] \exp \left( \int_0^{1/T} d\tau \bar{\psi}(\tau) (-\partial_{\tau} - \omega_0 + \mu) \psi(\tau) \right) = \int [D\bar{\psi} D\psi] e^{-S[\bar{\psi}, \psi]}.
\] (3.6)

Points to note:

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

\[
\psi(\tau_l = \Delta \tau l) \equiv \psi_l, \quad \bar{\psi}(\tau_l = \Delta \tau l) \equiv \bar{\psi}_l.
\]

- We elided the difference \( H(\bar{\psi}_{l+1}, \psi_l) = H(\bar{\psi}_l, \psi_l) + O(\Delta \tau) \) 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 (??) on \( \psi(\tau + \frac{1}{T}) = -\psi(\tau) \) mean that in its fourier representation\(^9\)

\[
\psi(\tau) = T \sum_n \psi(\omega) e^{-i\omega_n \tau}, \quad \bar{\psi}(\tau) = T \sum_n \bar{\psi}(\omega) e^{i\omega_n \tau}
\] (3.7)

the Matsubara frequencies

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

are half-integer multiples of \( \pi T \).

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

\(^9\bar{\psi} \) is still not the complex conjugate of \( \psi \) 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

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

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

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

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 (??)). The Berry phase term is actually

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

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

\[
(1 - e^{i\omega_n \tau}) \to i\omega_n \tau.
\]

Clearly this replacement is just fine if

\[
\omega_n \tau \ll 1
\]

for all \( \omega_n \) which matter. Which \( \omega_n \) contribute? I claim that if we use a reasonable \( H = H_{\text{quadratic}} + H_{\text{int}} \), reasonable quantities like \( Z, \langle O^\dagger O \rangle \), 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 \( T = 0 \):

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

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):

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

The partition function is

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

Notice that in the zero-temperature limit

$$T \sum_{\omega_n} \xrightarrow{T \to 0} \int \frac{d\omega}{2\pi} \equiv \int \bar{d}\omega.$$

(This is the same fact as $V \sum_k \xrightarrow{V \to \infty} \int d^d k$ in the thermodynamic limit.) So the zero-temperature partition function is

$$Z \xrightarrow{T \to 0} \int D\bar{\psi}(\omega) D\psi(\omega) \exp \left( \int_{-\infty}^{\infty} d\omega \tilde{\psi}(\omega) (i\omega - \omega_0 + \mu) \psi(\omega) \right).$$

Using the Gaussian-integral formula (??) you can see that the propagator for $\psi$ is

$$\langle \bar{\psi}(\omega_1) \psi(\omega_2) \rangle = \frac{\delta_{\omega_1,\omega_2}}{i\omega_1 - \omega_0 + \mu} \xrightarrow{T \to 0} \frac{2\pi}{i\omega_1 - \omega_0 + \mu}.$$

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

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

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

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

$$\langle N \rangle = \frac{1}{Z} \int D\bar{\psi} D\psi \ e^{-S[\bar{\psi},\psi]} \bar{\psi}(\tau_N + \Delta \tau) \psi(\tau_N)$$

$$= \int_{-\infty}^{\infty} d\omega \ e^{i\omega \Delta \tau} \frac{e^{i\omega (\mu - \omega_0 + \mu)}}{i\omega - \omega_0 + \mu} = \theta(\mu - \omega_0).$$

(3.9)
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, $\int_0^\Lambda \frac{d\omega}{\omega} \sim \log \Lambda$, 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.  

3.2 Topological terms from integrating out fermions

[Abanov ch 7] Here is a quick application of fermionic path integrals. Consider a 0+1
dimensional model of spinful fermions $c_\alpha, \alpha = \uparrow, \downarrow$ coupled to a single spin $s, \vec{S}$. Let’s
couple them in an SU(2)-invariant way:

$$H_K = M (c_\alpha^\dagger \vec{\sigma} c_\beta) \cdot \vec{S}$$

by coupling the spin of the fermion $c_\alpha^\dagger \vec{\sigma} c_\beta$ to the spin. ‘$K$’ is for ‘Kondo’. Notice
that $M$ is an energy scale. (Ex: find the spectrum of $H_K$.)

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

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

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

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

$$\langle c_\alpha^\dagger(t) c_\beta(0) \rangle \equiv \langle \bar{\psi}_\alpha(t) \psi_\beta(0) \rangle$$

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

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

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

\[10]\text{The calculation between the first and second lines of (??) is familiar to us – it is a single Wick}
\text{contraction, and can be described as a feynman diagram with one line between the two insertions.}
\text{More prosaically, it is}

$$\langle \bar{\psi}(\tau_N + \Delta \tau) \psi(\tau_N) \rangle \equiv T^2 \sum_{n,m} e^{i(\omega_n - \omega_m)\tau + i\omega_n \Delta \tau} \langle \bar{\psi}(\omega_n) \psi(\omega_m) \rangle \equiv T \sum_m \frac{e^{i\omega_m \Delta \tau}}{\omega_n - \omega_m + \mu} \mid_{\tau \to 0} \int d\omega \frac{e^{i\omega \Delta \tau}}{\omega - \omega + \mu}.$$
with
\[ S_{\text{eff}}[\vec{n}] = S_0[\vec{n}] - \log \det (\partial_t - M \vec{n} \cdot \vec{\sigma}) \equiv - \log \det D. \]

The variation of the effective action under a variation of \( \vec{n} \) is:
\[ \delta S_{\text{eff}} = - \text{tr} \left( \delta D \delta D^{-1} \right) = - \text{tr} \left( \delta D \dagger \left( D D \dagger \right)^{-1} \right) \]
where \( D \dagger = - \partial_t + M \vec{n} \cdot \vec{\sigma} \). This is
\[ \delta S_{\text{eff}} = M \text{tr} \left( \delta \vec{n} \cdot \vec{\sigma} (\partial_t + M \vec{n} \cdot \vec{\sigma}) \left( - \partial_t^2 + M^2 - M \vec{n} \cdot \vec{\sigma} \right)^{-1} \right). \quad (3.10) \]

We can expand the denominator in \( \dot{\vec{n}}/M \) (and use \( n^2 = 1 \)) to get
\[ \delta S_{\text{eff}} = \int dt \left( - \frac{M}{|M|} \frac{1}{2} \delta \vec{n} \cdot \left( \vec{n} \times \dot{\vec{n}} \right) + \frac{1}{4M} \delta \vec{n} \cdot \vec{n} + \ldots \right) \]
where ... is higher order in the expansion and we ignore it. But we know this is the variation of
\[ S_{\text{eff}} = -2\pi \frac{M}{|M|} W_0 + \int_0^T dt \left( \frac{1}{8M \vec{n}^2} \right) + O \left( \frac{\dot{\vec{n}}^2}{M} \right) \]
where \( W_0 \) is the WZW term. Integrating out the fermions has shifted the coefficient of the WZW term from \( s \to s \mp \frac{1}{2} \) depending on the sign of \( M \). This is satisfying: we are adding angular momenta, \( s \otimes \frac{1}{2} = (s - \frac{1}{2}) \oplus (s + \frac{1}{2}) \). If \( M > 0 \), it is an antiferromagnetic interaction whose groundstates will be the ones with smaller eigenvalue of \( \vec{S}^2 \). If \( M < 0 \), it is ferromagnetic, and the low-energy manifold grows.

The second term in \( S_{\text{eff}} \) is a shift of \( K \). Higher-order terms are suppressed by more powers of \( \frac{\dot{\vec{n}}}{M} \), so for \( \dot{\vec{n}} \ll M \), this is a local action. That means that the coupling to \( n \) must have gapped out the fermions. That the term proportional to \( M \) is a funny mass term for the fermions is clear from the expression for \( DD\dagger \) in (??): when \( n \) is static, \( DD\dagger = - \partial_t^2 + M^2 \), so that the fermion propagator is
\[ \langle \bar{\psi}_\alpha(t) \psi_\beta(0) \rangle = \left( \frac{D}{D\dagger} \right)_t = \int \frac{d\omega}{2\pi} e^{i\omega t} \frac{\omega + iM \vec{n} \cdot \vec{\sigma}_{\alpha\beta}}{\omega^2 + M^2} \sim e^{-Mt} \]
which is short-ranged in time. So indeed the fermions are fast modes in the presence of the coupling to the \( n \)-field.

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

The above calculation generalizes well to higher dimensions. For many examples of its application, see this paper. (The context for this paper will become clearer in §??).
3.3 Coherent state path integral for spins

3.4 Pions

[Schwartz §28.1] Below the scale of electroweak symmetry breaking, we can forget the \( W \) and \( Z \) bosons. Besides the 4-Fermi interactions, the remaining drama is QCD and electromagnetism:

\[
L_{QCD_2} = -\frac{1}{4} F_{\mu\nu}^2 + i \sum_{\alpha=L,R} \sum_f \bar{q}_{\alpha f} \gamma^\mu p_{\alpha f} - \bar{q} M q.
\]

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

\[
\begin{pmatrix} u \\ d \end{pmatrix} \rightarrow U \begin{pmatrix} u \\ d \end{pmatrix}, \quad U \in SU(N_f = 2).
\]

If, further, there were no masses \( m = 0 \), then \( L \) and \( R \) decouple and we also have chiral symmetry, \( q \rightarrow e^{i\gamma_5 \alpha} q \), i.e.

\[
q_L \rightarrow V q_L, q_R \rightarrow V^{-1} q_R, \quad V \in SU(N_f = 2).
\]

Why do I restrict to \( SU(2) \) and not \( U(2) \)? The central bit of the axial symmetry \( U(1)_A \) 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 \( \eta' \). The central bit of the vectorlike transformation \( q \rightarrow e^{\beta \gamma_5} q \) 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

\[
\langle \bar{q} f q_f \rangle = V^3
\]

independent of flavor \( f \). This condensate spontaneously breaks

\[
SU(2)_L \times SU(2)_R \rightarrow SU(2)_{\text{isospin}},
\]

the diagonal combination. \( \begin{pmatrix} u \\ d \end{pmatrix} \) is a doublet. Since \( p = u_\alpha d_\gamma d_\beta \epsilon_{\alpha\beta\gamma}, n = u_\gamma d_\beta d_\gamma \epsilon_{\alpha\beta\gamma}, \) 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 (2.25) (‘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 (2.25). One way to do this is to introduce a field $\Sigma$ which transforms like

$$\text{SU}(2)_L \times \text{SU}(2)_R : \Sigma \rightarrow g_L \Sigma g_R^\dagger, \quad \Sigma^\dagger \rightarrow g_R \Sigma^\dagger g_L^\dagger,$$

(this will be called a linear sigma model, because $\Sigma$ transforms linearly) – we have in mind $\bar{q}_\alpha q_\beta \sim \Sigma_{\alpha\beta}$. We can make singlets (hence an action) out of $\Sigma_{ij} \Sigma_{ji} = \text{tr} \Sigma \Sigma^\dagger \equiv |\Sigma|^2$:

$$\mathcal{L} = |\partial_\mu \Sigma|^2 + m^2 |\Sigma|^2 - \frac{\lambda}{4} |\Sigma|^4 + \cdots \quad \text{(3.13)}$$

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 (2.24)), which preserves $\text{SU}(2)_{\text{isospin}}$ (under which $\Sigma \rightarrow g \Sigma g^\dagger$). We can parametrize the fluctuations about this configuration as

$$\Sigma(x) = \frac{V + \sigma(x)}{\sqrt{2}} e^{\frac{2\imath a}{F_\pi} x^a},$$

where $F_\pi = V = \frac{2m}{\sqrt{\lambda}}$ is be chosen to give $\pi^a(x)$ canonical kinetic terms. The $\pi^a$ parametrize the directions of field space in which the potential is flat (like the field $\theta$ in the discussion of the Mexican hat in §???). Under $g_{L/R} = e^{\imath \theta_{L/R} x^a}$, the pion field transforms as

$$\pi^a \rightarrow \pi^a + \frac{F_\pi}{2} (\theta_L^a - \theta_R^a) \quad - \quad \frac{1}{2} f^{abc} (\theta_L^a + \theta_R^a) \pi^c.$$

The fields $\pi^\pm, \pi^0$ create pions, they transform in the adjoint representation of the diagonal $\text{SU}(2)_{\text{isospin}}$, and they shift under the broken symmetry. This shift symmetry forbids mass terms $\pi^2$. The radial excitation $\sigma$, on the other hand, is a fiction which we’ve introduced in (2.26), and which has no excuse to stick around at low energies (and does not). We can put it out of its misery by taking $m \rightarrow \infty, \lambda \rightarrow \infty$ fixing $F_\pi$. 65
In the limit, the useful field to use is
\[ U(x) \equiv \frac{\sqrt{2}}{V} \Sigma(x)|_{\sigma=0} = e^{\frac{2i\sigma_{\pi}x}{F_{\pi}}} \]

which is unitary \(UU^\dagger = U^\dagger U = 1\). This last identity means that all terms in an action for \(U\) require derivatives, so (again) no mass for \(\pi\). The most general Lagrangian for \(U\) can be written as an expansion in derivatives, and is called the \textit{chiral Lagrangian}:

\[ \mathcal{L}_\chi = \frac{F_{\pi}^2}{4} \text{tr} D_\mu U D^\mu U^\dagger + L_1 \text{tr} (D_\mu U D^\mu U^\dagger)^2 + L_2 \text{tr} D_\mu U D_\nu U^\dagger \text{tr} D^\nu U^\dagger D_\mu U + L_3 \text{tr} D_\mu U D^\mu U^\dagger D_\nu U D^\nu U^\dagger + \cdots \]

(3.14)

In terms of \(\pi\), the leading term expands into

\[ L_\chi = \frac{1}{2} \partial_\mu \pi^a \partial^\mu \pi^a + \frac{1}{F_{\pi}^2} \left( -\frac{1}{3} \pi^0 \pi^0 D_\mu \pi^+ D^\mu \pi^- + \cdots \right) + \frac{1}{F_{\pi}^4} \left( \frac{1}{18} (\pi^- \pi^+)^2 D_\mu \pi^0 D^\mu \pi^0 + \cdots \right) \]

This fixes the relative coefficients of many irrelevant interactions, all with two derivatives, suppressed by powers of \(F_{\pi}\). The expansion of the \(L_i\) terms have four derivatives, and are therefore suppressed by further powers of \(E/F_{\pi}\), the promised small parameter of this EFT.

**Pion masses.** The pions aren’t actually massless: \(m_{\pi^\pm} \sim 140\text{MeV}\). In terms of quarks, one source for such a thing is the quark mass term \(\mathcal{L}_{\text{QCD}} \ni \bar{q}Mq\). This explicitly breaks the isospin symmetry if the eigenvalues of \(M\) aren’t equal. But an \textit{invariance} of \(\mathcal{L}_{\text{QCD}}\) is

\[ q_{L/R} \rightarrow g_{L/R} \bar{q}_{L/R}, \quad M \rightarrow g_L M g_R^\dagger. \]

(3.15)

Think of \(M\) as a background field (such a thing is sometimes called a \textit{spurion}). If \(M\) were an actual dynamical field, then (2.28) 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 (2.28), 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 (2.28) should be an invariance. This determines

\[ \Delta \mathcal{L}_\chi = \frac{V^3}{2} \text{tr} (MU + M^\dagger U^\dagger) + \cdots = V^3(m_u + m_d) - \frac{V^3}{2F_{\pi}^2} (m_u + m_d) \sum_a \pi^2_a + \mathcal{O}(\pi^2). \]

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

\[ m_{\pi^2} \simeq \frac{V^3}{F_{\pi}^2} (m_u + m_d) \]

66
which is called the Gell-Mann Oakes Renner relation.

**Electroweak interactions.** You may have noticed that I used covariant-looking $D$s in (2.27). That’s because the $\text{SU}(2)_L$ symmetry we’ve been speaking about is actually gauged by $W^a_{\mu}$. (The electroweak gauge boson kinetic terms are in the $\cdots$ of (2.27).) Recall that

$$\mathcal{L}_{\text{Weak}} \ni gW^a_{\mu} \left( J^a_{\mu} - J^5_{\mu} \right) = gW^a_{\mu} \left( V_{ij} \bar{Q}_i \gamma^\mu \frac{1 - \gamma^5}{2} \tau^a Q_j + \bar{L}_i \gamma^\mu \tau^a \frac{1 - \gamma^5}{2} \bar{L}_i \right)$$

where $Q_1 = \begin{pmatrix} u \\ d \end{pmatrix}$, $L_1 = \begin{pmatrix} e \\ \nu_e \end{pmatrix}$ are doublets of $\text{SU}(2)_L$.

Now, in equations, the statement “a pion is a Goldstone boson for the axial $\text{SU}(2)$” is:

$$\langle 0 \mid J^5_{\mu}(x) \mid \pi^b(p) \rangle = ip_\mu F_\pi e^{-ip \cdot x} \delta^{ab}$$

where the state $\mid \pi^b(p) \rangle$ is a one-pion state of momentum $p$. If the vacuum were invariant under the symmetry transformation generated by $J_\mu$, 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 $p_\mu$ and using current conservation (ignoring the explicit breaking just mentioned) would give $0 = p^2 F^2_\pi = m^2_\pi F^2_\pi$, a massless dispersion for the pions.

Combining the previous two paragraphs, we see that the following process can happen

$$\pi^{\text{Goldstone}} \rightarrow J^5_{\mu} \text{ electroweak interaction} \rightarrow \text{leptons}$$

(3.16)

and in fact is responsible for the dominant decay channel of charged pions. (Time goes from left to right in these diagrams, sorry.)

$$\mathcal{M}(\pi^+ \rightarrow \mu^+\nu_\mu) = \frac{G_F}{\sqrt{2}} F_\pi p^\mu \bar{\nu}_\mu \gamma^\mu (1 - \gamma^5) u_\mu$$

where the Fermi constant $G_F \sim 10^{-5} GeV^{-2}$ (known from e.g. $\mu \rightarrow e^{-\bar{\nu}_e}\nu_\mu$) is a good way to parametrize the Weak interaction amplitude. Squaring this and integrating
over two-body phase space gives the decay rate

$$\Gamma(\pi^+ \to \mu^+ \nu_\mu) = \frac{G_F^2 F_\pi^2 m_\pi m_\mu}{4\pi} \left( 1 - \frac{m_\mu^2}{m_\pi^2} \right)^2.$$ 

(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 \(\pi^-\) into back-to-back spin-half particles by the weak interaction (which only produces \(L\) particles and \(R\) antiparticles) can’t happen if helicity is conserved – the mass term is required to flip the \(e_L\) into an \(e_R\).) This contributes most of \(\tau_{\pi^+} = \Gamma^{-1} = 2.6 \cdot 10^{-8}\) s.

Knowing further the mass of the muon \(m_\mu = 106\) MeV then determines \(F_\pi = 92\) MeV which fixes the leading terms in the chiral Lagrangian. This is why \(F_\pi\) 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:

$$q_\mu \langle p_1, p_2 | J_\mu^{5,a=3} \ | 0 \rangle = -\frac{e^2}{4\pi^2} \epsilon^{\nu\lambda\alpha\beta} p_1^\nu \epsilon_1^\lambda p_2^\alpha \epsilon_2^\beta$$

where \(\langle p_1, p_2 \rangle \) is a state with two photons of polarizations \(\epsilon_{1,2}\). I know this because it is a matrix element of the \(J_\epsilon J_\epsilon J_{SU(2)}\)-axial anomaly,

$$\partial_\mu J^{\mu5a} = -\frac{e^2}{16\pi^2} \epsilon^{\nu\lambda\alpha\beta} F_{\nu\lambda} F_{\alpha\beta} \text{tr} (\tau^a Q^2)$$

where \(Q = \begin{pmatrix} 2/3 & 0 \\ 0 & -1/3 \end{pmatrix} \) is the quark charge matrix. Comments: (1) this symmetry acts by \(u \to e^{i\theta} u, d \to e^{-i\theta} \), and is not the same as the anomalous \(U(1)_A\) (which does \(q_i \to e^{i\theta} q_i\) for every flavor), and it’s also not the same as isospin \(u \to e^{i\theta} u, d \to e^{-i\theta}\), which is not chiral, and not spontaneously broken. Confusing! (2) The rate of \(\pi^0\) 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 \(SU(3)_L \times SU(3)_R\) chiral symmetry (see Schwartz §30.6).

Wait – what \(SU(3)\)?
**SU(3) and baryons.** The strange quark mass is also pretty small $m_s \sim 95$ MeV, and $\langle \bar{s}s \rangle \sim V^3$. This means the approximate invariance and symmetry breaking pattern is actually $SU(3)_L \times SU(3)_R \rightarrow SU(3)_{\text{diag}}$, meaning that there are $16 - 8 = 8$ pseudo NGBs. Besides $\pi^{\pm,0}$, the others are the kaons $K^{\pm,0}$ and $\eta$. It’s still only the $SU(2)_L$ 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 $SU(3)$, the baryons are in the representation

$$3 \otimes 3 \otimes 3 = (6 \oplus \overline{3}) \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1$$

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 $SU(2)$ approximation, we can include the nucleons $N_{L/R} = (p,n)_{L/R}$ 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 $\Sigma$ 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 $\pi \rightarrow -\pi$ which is not a symmetry of QCD).

The chiral Lagrangian governs a non-linear sigma model (NL$\sigma$M)– 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 $(SU(N_f)_L \times SU(N_f)_R)$ and $H$ is the unbroken subgroup $SU(N_f)_{\text{diagonal}}$. We can parametrize this space by $U = e^{i\pi^a T^a/2}$ where the $T^a$ includes only generators of the broken part of the group, so the $\pi^a$ are coordinates on $G/H$.

A WZW term is a term which we can sometimes add to a NL$\sigma$M 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 \( \hat{n} \) takes values on the unit sphere \( S^2 \), \( 1 = \sum_{a=1,2,3} \hat{n}^2_a \). This is a special case of a coset space \( G/H = SU(2)/U(1) \).

In order to write the WZW term in a manifestly symmetric way (under the \( SO(3) \) 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 \( u = 1 \):

\[
\hat{n}(t, u = 1) \equiv \hat{n}(t), \quad \hat{n}(t, u = 0) \equiv (0, 0, 1)
\]

it goes to the north pole at the other end of the extra dimension for all \( t \). Consider periodic boundary conditions in time \( \hat{n}(2\pi) = \hat{n}(0) \). 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 \( \hat{n}^{1,2,3} \) as

\[
\mathcal{W}_1[\hat{n}] = \frac{2\pi}{\Omega_2} \int_{B_2} \hat{n}^a d\hat{n}^b \wedge d\hat{n}^c \epsilon_{abc} = \frac{1}{4\pi} \int_{\mathcal{M}} dt \left( 1 - \cos \theta \right) \partial_t \phi.
\]

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 \( 2\pi \) (the integer is the winding number of the map).

The coefficient \( k \) of \( \mathcal{W}_1 \) in the action \( \Delta S[\hat{n}] = k \mathcal{W}_1[\hat{n}] \) 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

\[
\mathcal{W}_D = c \int_{B_{D+1}} \tr U^{-1} dU \wedge U^{-1} dU \wedge \cdots \wedge U^{-1} dU.
\]

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 $W_D$ with respect to $U$ is (for even $D$):\footnote{Why do I restrict to even $D$?}
\[
\delta W_D = (D + 1)c \int_{B_{D+1}} \text{tr} \left\{ (U^{-1}dU)^D \frac{\delta (U^{-1}dU)}{=U^{-1}(\deltaUU^{-1})U} \right\}
\]
\[
= (D + 1)c \int_{B_{D+1}} \text{tr} \left\{ (dUU^{-1})^D d(\deltaUU^{-1}) \right\}
\]
\[
= (D + 1)c \int_{B_{D+1}} d\text{tr} \left\{ (U^{-1}dU)^D U^{-1} \delta U \right\}
\]
\[
\overset{\text{Stokes}}{=} (D + 1)c \int_{M} \text{tr} \left\{ (U^{-1}dU)^D U^{-1} \delta U \right\}
\]
which only depends on the field configuration on $M$, 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 $c \int_{S^{D+1}} \text{tr}(U^{-1}dU)^D \in \mathbb{Z}$ is the winding number), and the coefficient must be an integer. (In $D = 4$, we have $c = -\frac{1}{24\pi^2}$.)

This WZW term is less topological than the theta term we discussed above, in the sense that it affects the equations of motion for $\hat{n}(t)$. The variation of $W$ 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 $M_D$:

\[
\text{tr} \left\{ (U^{-1}dU)^D \frac{\delta UU^{-1}}{} \right\} = \epsilon_{\mu_1 \cdots \mu_{D+1}} \text{tr} \left\{ U^{-1} \partial_{\mu_1} U \cdots U^{-1} \partial_{\mu_{D+1}} \right\}
\]

but $\epsilon_{\mu_1 \cdots \mu_{D+1}} = (-1)^{D+1} \epsilon^{\mu_1 \cdots \mu_{D+1}}$ so $W_D = (-1)^D W_D$ vanishes in odd dimensions. The step from (2.33) to (2.34) also relies on this fact. Using $1 = U^{-1}U$ and hence $0 = \delta(U^{-1}U) = d(U^{-1}U)$, so that
\[
dU^{-1} = -U^{-1}dUU^{-1},
\]
the term by which (2.33) and (2.34) differ is
\[
\text{tr} \left\{ (d(U^{-1}dU)^D) \delta UU^{-1} \right\}
\]

\[
\overset{\text{product rule}}{=} \text{tr} \left\{ (dUU^{-1} \wedge (U^{-1}dU)^{D-1} - (U^{-1}dU \wedge dUU^{-1} \wedge (U^{-1}dU)^{D-2} + \cdots) \delta UU^{-1} \right\}
\]

\[
= \text{tr} \left\{ (U^{-1}dUU^{-1} \wedge dUU^{-1} \wedge (U^{-1}dU)^{D-1} - U^{-1}dUU^{-1} \wedge dUU^{-1} \wedge (U^{-1}dU)^{D-2} + \cdots) \delta UU^{-1} \right\}
\]

\[
= \text{tr} \left\{ (1 - 1 + 1 - 1 \cdots) (U^{-1}dU)^{D-1} \delta UU^{-1} \right\} D-1 \text{ even}.
\]

See Weinberg, vol 2, §23.4 for more.

71
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 $\pi^0 \rightarrow \gamma \gamma$. 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 $U : (\text{space} \cup \infty \sim S^4) \rightarrow G/H$, 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, $G/H = SU(3) \times SU(3)/SU(3)_{\text{preserved}} \simeq SU(3)_{\text{broken}}$). The configuration is topological in the sense that as a map from $S^3 \rightarrow G/H$, 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 $B_\mu = \frac{e_{\mu\nu\alpha\beta}}{24\pi^2} \text{tr} U^{-1} \partial_\nu U U^{-1} \partial_\alpha U U^{-1} \partial_\beta U$ whose conserved charge $\int_{\text{space}} B_0$ is 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 $\pi$ on a spacetime trajectory where the soliton makes a $2\pi$ 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 of 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).