QFT III

Patrick Draper
2025

Contents

Introduction

Matter is made of atoms, and atoms are made of electrons and nuclei, and nucleons are made of quarks and gluons. Invaluably, however, it was unnecessary for Galileo, Huygens, Hooke, Newton, Lagrange, Hamilton, and the other developers of classical mechanics from the sixteenth to the nineteenth centuries to know anything of atomic theory, nuclear physics, or quantum chromodynamics. The basic tools that allow us to understand anything quantitative about the universe are dimensional analysis and decoupling of widely separated scales. You are familiar with dimensional analysis. Decoupling, roughly speaking, is the idea that to describe phenomena at one length (or energy, or velocity...) scale D, with finite loss of accuracy, we can set dimensionful parameters that are very large compared to D to infinity, and parameters that are very small compared to D to zero. You do not have to worry about the curvature of the universe, and you do not have to worry about the Bohr radius or the QCD scale, in order to accurately predict the motion of the planets. Subsequently, we usually have only a small number of dimensionful parameters left in the problem, and observables can be written in terms of them using dimensional analysis, up to some dimensionless coefficients that are often $\mathcal{O}(1)$. The price for this simplification is that no one theory, if it is to be useful, will describe all scales of observation. Chemistry has no explanation for the binding of nucleons in nuclei.

Effective field theory (EFT) is the business of applying dimensional analysis and decoupling to construct simple, predictive, systematically improvable theories describing the dynamics of fields at characteristic length or energy scales. The basic inputs to an EFT are a collection of degrees of freedom that we expect to matter at the scale of interest; a specification of any symmetries that the dynamics must exactly or approximately respect; and a finite collection of coupling parameters. The construction of an EFT usually proceeds by identifying the most general Lagrangian or Hamiltonian that could describe the dynamics, consistent with the symmetries, then truncating it on the basis of dimensional analysis.

This description is so broad that one can correctly guess that EFTs are useful in both classical and quantum mechanics, in mixed and pure states, in relativistic and nonrelativistic physics, in continuum and lattice models, and in all dimensions. In these notes we will focus on relativistic quantum field theory and high energy particle physics. Even here there are myriad uses of EFTs. For example, the notion of separation of scales can be applied in various ways in particle physics, and the simplest of these is in the hierarchies of particle masses. Here is a tower of a few of the elementary particle masses, by order of magnitude:

$$\begin{aligned} m_\nu &\sim 1 \text{ eV} \\ m_e, m_u &\sim 1 \text{ MeV} \\ m_\mu, m_\pi &\sim 100 \text{ MeV} \\ m_\tau, m_c &\sim 1 \text{ GeV} \\ m_W, m_Z, m_h, m_t &\sim 100 \text{ GeV} \\ m_{SUSY} &\gtrsim 10 \text{ TeV???} \\ m_{GUT} &\sim 10^{16} \text{ GeV???} \end{aligned}$$

In general, EFTs can simplify the description of particle physics at a particular energy scale by eliminating all particles that are too heavy to be produced in the interactions.

In quantum field theory and in statistical mechanics, the *why* and the *how* of decoupling is explained by the renormalization group (RG). Roughly, the renormalization group tells us that the influence of heavy degrees of freedom of mass M to a process involving light degrees of freedom with characteristic energy $E \ll M$ can be captured by $\mathcal{O}(1)$ changes in a finite number of physical constants, plus a power series in E/M. In fact RG provides both a conceptual framework to understand why EFT works and a technical tool for simplifying precision computations. For this reason these notes begin with a review of renormalization and RG. The goal is to expand on what is normally covered in introductory QFT courses and develop tools that are particularly useful for carrying out EFT analyses. For example, we will see how to generalize the "old" notions of renormalizability, broadening the space of useful quantum field theories. We will also see how RG flow gives rise to towers of EFTs, and how to match EFTs onto each other at their boundaries of validity.

There are two complementary ways of looking at EFTs that we will study:

- *Top down.* Given a theory ("UV," "microscopic," "short-distance," "high-energy"), EFT techniques allow the systematic elimination of the theory's heavy degrees of freedom in order to obtain an effective ("IR," "macroscopic," "long-distance," "low-energy") theory with the same dynamics for the light degrees of freedom. Eliminating the heavy DOF is called "integrating out" and fixing the Lagrangian coefficients in the EFT so that the same IR dynamics is obtained is called "matching."

- *Bottom up.* Given a set of light degrees of freedom, a symmetry structure, and a scale M, using EFT techniques we can construct the most general Lagrangian that can contribute to dynamical process at fixed order in the E/M expansion.

Much as a power series often has a finite radius of convergence, EFTs have natural ranges of validity. We would not expect the description of electron scattering to be well-described by QED for center of mass energies of order 100 MeV, where muon and pion pair production is possible. Furthermore if we are interested in photon-photon scattering at energies below 1 MeV, the electron can be removed and an EFT written just for the photon. One goal of this course is to understand how to view the Standard Model through the lens of EFT. It is convenient and computationally precise to treat the low energy effects of the weak interactions, including beta decay and flavor-changing processes in hadron physics, using EFT generated from the top-down.

Meanwhile bottom-up EFT is essential in the description of the strong interactions at low energies. We will also come to think of the SM itself as an EFT, generated by some unknown physics at even shorter distances, and we will see how the SM should be modified to incorporate testable E/M corrections.

After discussing RG, we will introduce top-down EFT. However, we will see that this approach is mostly not useful for the description of physics around the QCD scale. We will then take a step back and review the constraints global symmetries impose on QFTs, showing that they can guide the bottom-up construction of EFTs. With this machinery we will study chiral perturbation theory, the EFT describing pion physics, among other things.

There is an important feature of renormalization that we will gloss over somewhat in the first pass: the concept of "natural" magnitudes for renormalizable parameters. We will see that the SM has two severe "naturalness problems," and that curing these problems is a highly nontrivial exercise in model building.

This summarizes the first part of the course. The second part will be about topological objects in QFTs: instantons, monopoles, strings, domain walls, and the roles these objects play in the infrared dynamics. We will also see some other applications of EFT reasoning. However, I haven't written a proper introduction for this part of the course yet. Material is present in the notes but is still somewhat under construction.

Time permitting, there will be a third part of the course, introducing supersymmetry. This material is not yet incorporated into these notes.

Chapter 1

Renormalization and RG

1.1 Dimensional analysis at tree level

In relativistic QFT, we can use

$$\begin{aligned} &[c] = L/T \\ &[\hbar c] = E \cdot T \quad \left(10^{-11} \text{ MeV} \cdot \text{cm}\right) \end{aligned} \tag{1.1}$$

to express all dimensionful quantities as energies:

$$1/L, 1/T, M \to E \tag{1.2}$$

The action is dimensionless and spacetime coordinates have dimension −1. From this we may infer the classical scaling dimensions, or "engineering dimensions," of fields and couplings:

$$\begin{aligned}
&[S] = 0 \quad [x^\mu] = -1 \\
&\left[\int d^d x \, (\partial^\mu \phi)^2\right] = 0 \quad \Rightarrow \quad [\phi] = \frac{d}{2} - 1 \ \ (= 1 \text{ in } d = 4) \\
&\left[\int d^d x \, \bar{\psi} i \partial\!\!\!/ \psi\right] = 0 \quad \Rightarrow \quad [\psi] = \frac{d-1}{2} \ \ (= 3/2 \text{ in } d = 4) \\
&\int d^d x \, m^2 \phi^2 \quad \Rightarrow \quad \left[m^2\right] = 2 \ \ (\text{all } d) \\
&\int d^d x \, \lambda_n \phi^n \quad \Rightarrow \quad [\lambda] = d\left(1 - \frac{n}{2}\right) + n \ \ (= 4 - n \text{ in } d = 4)
\end{aligned} \tag{1.3}$$

At tree level, we can use dimensional analysis to make a powerful observation about correlation functions. In $d = 4$, for example, consider the n-point momentum space correlator of scalar fields:

$$\langle \phi(p_i) \ldots \phi(p_n) \rangle \equiv \delta^4\left(\sum_i p_i\right) \times G_n\left(p_i; \lambda_i\right). \tag{1.4}$$

The Fourier-space field $\phi(p)$ have classical scaling dimensions -3 in four dimensions. Let all momenta and energies be of a similar order E. Then

$$G_n\left(p_i;\lambda_i\right) \simeq E^{4-3n} f\Bigg(\underbrace{\frac{m^2}{E^2}, \lambda_4, \lambda_6 E^2, \lambda_8 E^4, \ldots}_{\text{dimensionless ratios}}\Bigg). \tag{1.5}$$

As $E \to 0$,

$$G_n\left(p_i;\lambda_i\right) \approx E^{4-3n}\left[f\left(\frac{m^2}{E^2}, \lambda_4, 0, 0, \cdots\right) + \mathcal{O}\left(\lambda_6 E^2\right)\right]. \tag{1.6}$$

We learn:

- The mass becomes very important, or "strongly coupled," at low E. We say that ϕ^2, and any operators with couplings of dimension $[\lambda_n] > 0$, are "relevant operators."
- The importance of λ_4 is the same for all E. ϕ^4, and in general any operators with couplings of vanishing dimension, are "marginal operators."
- All "higher-dimension operators" (HDOs), or those with couplings of dimension $[\lambda_n] < 0$, become weakly coupled at low energies, with importance vanishing as $E^{-[\lambda_n]}\lambda_n$. These are "irrelevant operators."

The dimension in which an operator becomes marginal is called the critical dimension of the operator.

In the early days, irrelevant operators were thought to be problematic, indicating a loss of perturbative renormalizability. We will see they are not a big deal, and in perturbation theory the scalings above are still true after including quantum corrections, up to small calculable corrections. Instead, it is the relevant operators that are sometimes problematic, for a different reason.

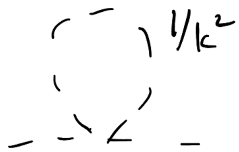
1.2 Quantum Dimensional Analysis 1

The previous discussion must be modified in quantum theory. The basic issue is that quantum corrections introduce another scale.

Correlation functions receive quantum corrections. In perturbation theory, these may be computed from loop diagrams. The corresponding Feynman integrals may be divergent: the loop integrals sample arbitrarily high momenta, or arbitrarily short distances, causing the integrals to diverge.

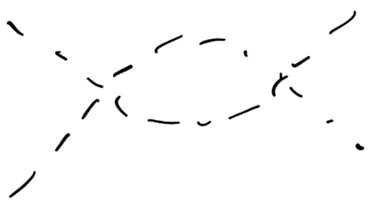
A first step in dealing with these divergences is to identify them in theories with couplings of only positive-semidefinite engineering dimension: in this case there are only a finite number of "primitive" (independent) divergences in one-particle irreducible (1PI) graphs. Historically, this is also how the theory of renormalization was developed. We will relax the restriction on the coupling dimensions later.

Example: ϕ^4 in 4D : $\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{4!}\lambda\phi^4$. At one loop, divergences appear in the two-point function



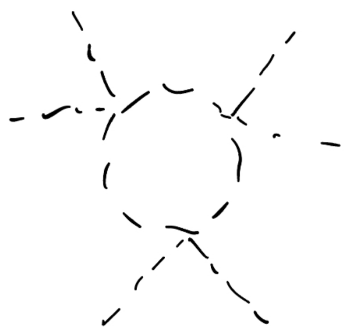
$$\sim \int^\Lambda \frac{d^4k}{k^2+m^2} \sim \Lambda^2 + m^2 \log \Lambda/m \tag{1.7}$$

and in the four-point function



$$\sim \int^\Lambda \frac{d^4k}{(k^2+m^2)^2} \sim \log(\Lambda/m) \tag{1.8}$$

However, after that the divergent behavior terminates:



$$\sim \int^\Lambda \frac{d^4k}{(k^2+m^2)^3} \sim 1/m^2 + 1/\Lambda^2 \tag{1.9}$$

Note: to extract the leading UV divergences precisely, we can set $m^2, p_i \to 0$. This makes the extraction much easier. m^2 dependence can be restored on dimensional grounds.

At first sight, divergences suggest a disturbing sensitivity of the low-momentum behavior to high-momentum degrees of freedom, and moreover a failure of perturbation theory. However, the renormalization procedure demonstrates that this is not such a big deal. The idea is that (1) there is a one-to-one correspondence between divergences and local operators in $\mathcal{L}$, and (2) Lagrangian parameters are not observables, so we may absorb divergences into redefined couplings: "bare" $\to$ "renormalized." In terms of the renormalized couplings, the correlation functions are finite. To carry out the procedure we need:

- a way to isolate the divergences. "regularization"
- a prescription for absorption. "renormalization"

There will be some freedom in both of these steps.

Regardless of how we regulate and renormalize the theory, a consequence of the renormalization procedure is the appearance of a "renormalization scale" or a "subtraction scale" μ in the correlation functions:

$$G_n\left(p_i;\lambda_i,\mu\right) = E^{4-3n} f\left(\frac{m^2}{E^2},\lambda_4,\lambda_6 E^2,\ldots,\mu/E\right) \tag{1.10}$$

Compare with Eq. (1.5). Furthermore, the μ dependence is not arbitrary. Exploiting this, we will see how quantum corrections modify classical scaling.

1.3 Perturbative renormalization example: φ^3 in 6D

Bare lagrangian:

$$L_0 = \frac{1}{2}\left(\partial_\mu\varphi_0\right)^2 - \frac{1}{2}m_0^2\varphi_0^2 - \frac{1}{6}g_0\varphi_0^3$$

Standard procedure: introduce counterterms. Split $\mathcal{L}$:

$$\begin{aligned}
&L = \frac{1}{2}z_\varphi\left(\partial_\mu\varphi\right)^2 - \frac{1}{2}z_m m^2\varphi^2 - \frac{1}{6}z_g g\varphi^3 \\
&\varphi_0 = \sqrt{z_\varphi}\varphi \\
&m_0^2 = z_m z_\varphi^{-1} m^2 \\
&g_0 = z_g z_\varphi^{-3/2} g.
\end{aligned} \tag{1.11}$$

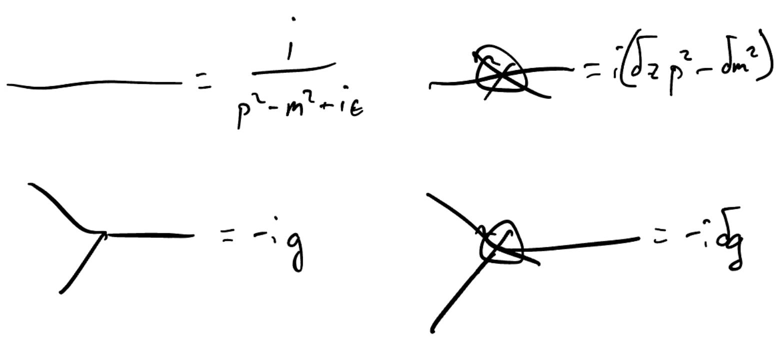
(There are many equivalent ways to introduce z's. All a matter of convention, no physics here.) Then split z's:

$$z_\varphi \equiv 1 + \delta z \tag{1.12}$$

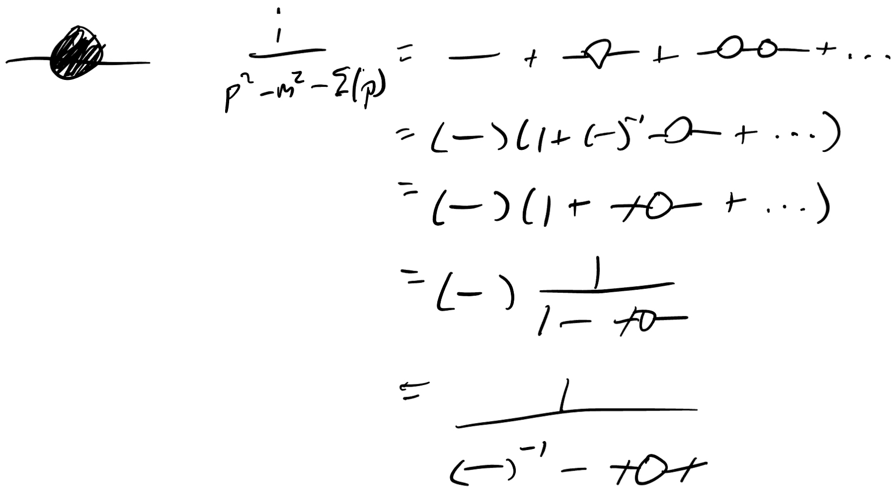
$$z_m \equiv 1 + \delta m^2/m^2 \tag{1.13}$$

$$z_g \equiv 1 + \delta g/g \tag{1.14}$$

Now $\mathcal{L}$ takes the form of a renormalized Lagrangian plus a counterterm Lagrangian, $L = L_r + L_{ct}$. Feynman rules:



We can focus on IPI diagrams, since connected Green functions are built from them. For example, the "bubble chain" of 1-loop 1PI self energy corrections (denoted $\not{-}O\not{-}$) builds the connected self-energy:



This generalizes to higher loop 1P1 and n-point Green functions.

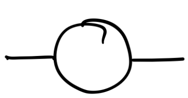
We will do two regularization and two subtraction prescriptions. One is useful for intuition, and the other is useful for practical computation.

1.3.1 Euclidean momentum cutoff regularization

We compute the Feynman integrals in Euclidean signature, following Wick rotation $p^0 = ip^0_E$. This modifies the Feynman rule for the quadratic counterterm to

$$i\left(\delta z p^2 - \delta m^2\right) \to -i\left(\delta z p_E^2 + \delta m^2\right). \tag{1.15}$$

The self-energy diagram is

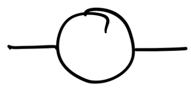


$$= \frac{1}{2}(-ig)^2 \int \frac{id^6 k_E}{(2\pi)^6} \frac{i}{k_E^2 + m^2} \frac{i}{\left((k_E + p_E)^2 + m^2\right)}. \tag{1.16}$$

We drop the subscript E for notational convenience, $k_E \to k$. In the large k regime of the integral we can expand the integrand as a series in p/k and m/k. So doing, we see that the integral is of the form

$$\begin{aligned} &\int \frac{d^6 k}{k^4} + \frac{d^6 k}{k^6}\left(p^2, m^2\right) + \cdots \\ &\sim (\text{quadratic divg}) + (\text{log divg}) + (\text{finite}) \end{aligned} \tag{1.17}$$

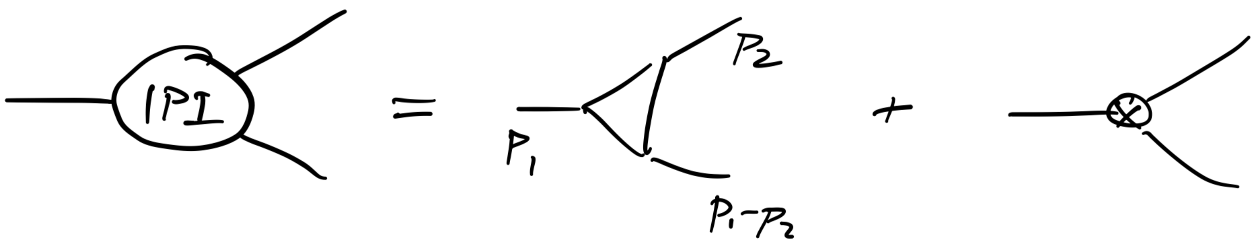
We regulate the divergences with a hard cutoff on the magnitude of the Euclidean loop momentum. After introducing Feynman parameters, shifting the integration variable, and performing the angular integrals, we have



$$
\begin{aligned}
&= \frac{ig^2}{128\pi^3} \int_0^1 dx \int_0^\Lambda \frac{l^5 dl}{(l^2+\Delta_x)^2} \\
&= \frac{ig^2}{128\pi^3} \int_0^1 dx \left[\frac{\Lambda^2}{2} + \Delta_x \log \frac{m^2}{\Lambda^2} + (\Lambda - \text{indep}) + \mathcal{O}\left(\Delta_x/\Lambda\right) \right] \\
&= \frac{ig^2}{128\pi^3} \left(\frac{\Lambda^2}{2} + \left(m^2 + \frac{p^2}{6} \right) \log \frac{m^2}{\Lambda^2} + \dots \right) \qquad (1.18)
\end{aligned}
$$

where $\Delta_x = m^2 + p^2 x(1-x)$.

Now let's look at the 3-point correlator at one loop:



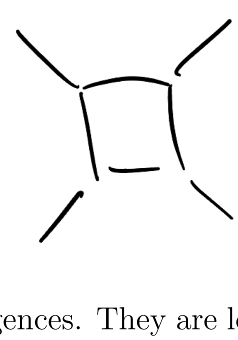
The large k behavior is

$$
\int \frac{d^6k}{k^6} + \int \frac{d^6k}{k^8} \left(m^2, p^2\right) + \dots \quad \sim \log\Lambda \ + \ \text{finite}. \qquad (1.19)
$$

So to isolate the UV-sensitive part, we can set $p \to 0$. (This makes it easy - no Feynman parameters for $p = 0$.) We'll keep m around for a convenient IR cutoff. In this limit the first one-loop diagram corresponds to the Feynman integral

$$
\begin{aligned}
(-ig)^3 \int \frac{id^6k}{(2\pi)^6} \frac{i}{(k^2+m^2)^3} &= -\frac{ig^3}{64\pi^3} \int_0^\Lambda \frac{k^5 dk}{(k^2+m^2)^3} \\
&= -\frac{ig^3}{64\pi^3} \log(\Lambda/m) + \text{finite}. \qquad (1.20)
\end{aligned}
$$

All higher-point IPI diagrams are finite. The four-point box diagram, for example, behaves as

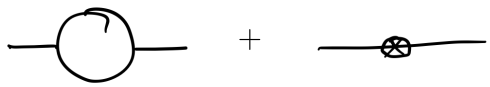


$$
\sim \int \frac{d^6k}{\left(k^2+m^2\right)^4} \sim \frac{1}{m^2}. \qquad (1.21)
$$

We have isolated the divergences. They are local,[1] so we can absorb them into redefinitions of the

[1]polynomials in p : $\int e^{ipx} p^n \sim \partial^\mu \delta(x)$.

local couplings in $\mathcal{L}$ - or equivalently, the counterterms. For example,



$$= \frac{ig^2}{256\pi^3}\left(\Lambda^2 - \left(2m^2 + \frac{p^2}{3}\right)\log\frac{\Lambda^2}{m^2} + \text{ finite }\right) - i\left(\delta z\, p^2 + \delta m^2\right) \tag{1.22}$$

which shows that the quadratic counterterms have the right momentum structure to absorb divergences in the self-energy diagram. We will carry out this renormalization below, but first, let us introduce another useful regulator.

1.3.2 Dimensional regularization

Physics is regularization and renormalization scheme independent. But sometimes a good scheme choice can make computations easier. We don't typically use hard momentum cutoffs in QED, because it breaks the gauge symmetry and Poincare symmetries. This can be repaired by tuning local counterterms, but it is tedious, and computations are easier if we use covariant schemes like Pauli-Villars or dimensional regularization. There are other reasons dimensional regularization is useful as well, as we will see.

In dimensional regularization, Feynman integrals are regulated by replacing

$$\mu^{2\epsilon}\int \frac{d^{d-2\epsilon}k}{k^{d,d-2,d-4\ldots}}. \qquad \text{for } \epsilon > 0, 1, 2, \ldots \text{ these are UV - finite }. \tag{1.23}$$

μ is a new scale needed to counteract the change in the engineering dimension of the integration measure. "Analytic continuation in spacetime dimensions."

Some more detailed examples of carrying out Feynman integrals in dim reg are reviewed at the end of this chapter. Here we proceed to dimensionally regularize the primitive one-loop divergences in the example of 6D φ^3 theory.

The one-loop self-energy diagram gives

$$\mathcal{I}_2 = ig^2\int_0^1 dx \int \frac{d^6 l}{(2\pi)^6}\frac{1}{(l^2+\Delta_x)^2} \qquad \left(\Delta_x = m^2 + p^2(1-x)x\right). \tag{1.24}$$

In $6 \to d$ dimensions,

$$\begin{aligned}
\mathcal{I}_2 &\Rightarrow \frac{ig^2}{2}\int_0^1 dx \frac{2(\sqrt{\pi})^d}{\Gamma(d/2)}\frac{1}{(2\pi)^d}\mu^{6-d}\left[\left(\frac{1}{\Delta_x}\right)^{2-\frac{d}{2}}\frac{\Gamma(d/2)\Gamma(2-d/2)}{2\Gamma(2)}\right] \qquad \text{set } d = 6-2\epsilon \\
&\Rightarrow \frac{ig^2}{2}\int_0^1 dx \frac{1}{(4\pi)^3}\left(\frac{4\mu^2\pi}{\Delta_x}\right)^{\epsilon}\Delta_x \underbrace{\Gamma(-1+\epsilon)}_{-\frac{1}{\epsilon}+\gamma-1} \\
&\Rightarrow \frac{ig^2}{128\pi^3}\int_0^1 dx \Delta_x\left[-\frac{1}{\epsilon} - \left(1 + \log\frac{\tilde{\mu}^2}{\Delta_x}\right)\right] \\
&= \frac{ig^2}{128\pi^3}\left(-\frac{1}{\epsilon}\right)\times\left(m^2 + \frac{p^2}{6}\right) + \cdots
\end{aligned} \tag{1.25}$$

At $p = 0$, the one loop vertex correction gives

$$(-ig)^3 \int \frac{id^6k}{(2\pi)^6} \frac{(-1)(i^3)}{(k^2+m^2)^3} \underset{d=6-2\epsilon}{\longrightarrow} -\frac{ig^3}{(4\pi)^3} \frac{1}{\Gamma(3)} \Gamma(\epsilon) \left(\frac{4\pi\mu}{m^2}\right)^{\epsilon}$$
$$= -\frac{ig^3}{128\pi^3} \left(\frac{1}{\epsilon} + \mathcal{O}(\log(\tilde{\mu}))\right). \tag{1.26}$$

The Euclidean momentum cutoff and dim reg are just two common examples of regulators. Finite spacetime lattices, Pauli-Villars, and smooth versions of the momentum cutoff all achieve the same end. Let us now turn to step two, the renormalization prescription.

1.3.3 A mass-dependent renormalization scheme

Our first renormalization scheme is given by the following choices of counterterms in the cutoff case:

$$\delta z = -\frac{g^2}{256\pi^3} \frac{1}{3} \log \frac{\Lambda^2}{\mu^2}$$
$$\delta m^2 = \frac{g^2}{256\pi^3} \left[\left(\Lambda^2 - \mu^2\right) - 2m^2 \log \frac{\Lambda^2}{\mu^2} \right]$$
$$\delta g = -\frac{g^3}{128\pi^3} \log \frac{\Lambda^2}{\mu^2}. \tag{1.27}$$

The sum of the self energy diagram and counterterms, Eq. (1.22), becomes

$$\frac{ig^2}{256\pi^3} \left(\Lambda^2 - \left(2m^2 + \frac{p^2}{3}\right) \log \frac{\Lambda^2}{m^2} + \text{ finite}\right) - i\left(\delta z\, p^2 + \delta m^2\right)$$
$$= \frac{ig^2}{256\pi^3} \left(\mu^2 - \left(2m^2 + \frac{1}{3}p^2\right) \log \frac{\mu^2}{m^2} + \text{finite}\right). \tag{1.28}$$

Likewise the sum of the vertex correction and vertex counterterm becomes

$$-\frac{ig^3}{128\pi^3} \log(\Lambda/m) + \text{finite} - i\delta g$$
$$= -\frac{ig^3}{128\pi^3} \log(\mu/m) + \text{finite}. \tag{1.29}$$

Thus our choice of counterterms completely removes explicit Λ dependence at one loop, so that the $\Lambda \to \infty$ limit may be taken with fixed renormalized parameters.

In introductory QFT courses, the first renormalization schemes introduced are usually defined by the requirement that, after removing the cutoff, the loop diagrams and counterterms exactly cancel at some particular external momentum p_0. This is a physically intuitive scheme because it means that the renormalized parameters are, to very good approximation, the masses and interaction strengths that you would measure in scattering amplitudes around that momentum. The scheme above is a simplified version, but analogous. Note that the introduction of μ is necessary on dimensional grounds (because the argument of the logarithm must be dimensionless.) We see that in this scheme μ appears both inside logarithms and in polynomials (in the mass squared counterterm.) Such schemes are called "mass dependent."

1.3.4 A mass-independent renormalization scheme

It is also useful to consider "mass-independent" schemes. **In these schemes, μ only appears in logarithms, not polynomially.** The significance of this property will be discussed later.

We'll just consider mass-independent schemes in conjunction with dimensional regularization, since that is the most common setting. Some examples:

- Minimal subtraction (MS): choose counterterms to remove $1/\epsilon$ poles only. In effect, one just drops $1/\epsilon$ wherever is appears.
- Modified MS ($\overline{MS}$): choose counterterms to remove $1/\epsilon$ poles and $\log(\tilde{\mu}^2/\mu^2)$. In effect, one just drops $1/\epsilon$ wherever it appears, and replaces $\log(\tilde{\mu})$ by $\log(\mu)$.

The $d = 6$ φ^3 counterterms in dim reg with minimal subtraction are:

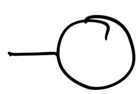
$$\begin{aligned}\delta z &= \frac{-g^2}{256\pi^3}\frac{1}{3}\frac{1}{\epsilon} \\ \delta m^2 &= -\frac{g^2}{128\pi^3}m^2\frac{1}{\epsilon} \\ \delta g &= -g^3\frac{1}{128\pi^3}\frac{1}{\epsilon}. \end{aligned} \tag{1.30}$$

Comparing with Eqs. (1.25),(1.26) we see that these counterterms are just throwing away the $1/\epsilon$ poles. Note that in this scheme the counterterms do not depend explicitly on μ. We can also see that any renormalized amplitude will carry explicit dependence on $\log\mu$. Thus, this scale, introduced in going from $6 \to 6 - 2\epsilon$ dimensions, may be treated like the renormalization scale in previous analysis. Masses and couplings depend on μ, so that observables are μ-independent.

1.3.5 The tadpole

An important technicality: we missed a lower-point diagram, the tadpole.

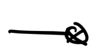
We could have added a term in the bare Lagrangian, $y_0\varphi_0$. This doesn't break any symmetries of the φ^3 theory. The one-loop tadpole diagram is



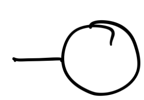
$$= \frac{-ig}{(2\pi)^6}\int \frac{id^6k}{k^2+m^2}$$

$$\sim \Lambda^4 + \Lambda^2 m^2 + m^4 \log\Lambda + \ldots \tag{1.31}$$

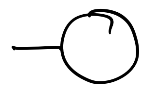
There is a quartic divergence. $\langle\varphi\rangle$ indicates we were not doing our perturbative expansion around the minimum of the potential. To do so, we choose our renormalization condition to be



$$= -$$



Then, if we take $y = 0$,

$\langle\varphi\rangle =$  $+$  $= 0$

Once we choose this renormalization condition, all tadpole subgraphs vanish and we can forget about it. We'll come back to the freedom to adjust linear sources (tadpoles) for the fields later when we discuss effective actions.

1.3.6 Outstanding issues

We have now done the basic steps to obtain all correlation functions of elementary fields in perturbation theory, or renormalize, φ^3 theory. This procedure is general:

- Identify basic divergent IPI graphs to fixed loop order
- regulate integrals
- absorb regulator dependence into local counterterms

There are some clear outstanding issues, however:

1. How do we choose μ?
2. What are the pros and cons of mass dependent and mass independent schemes?
3. How should we handle interactions of negative mass dimension?
4. What about the scaling of correlators of composite operators?
5. Can we say anything beyond perturbation theory?

We will address these issues and identify others in the following sections.

1.4 The renormalization scale and β-functions

Bare couplings + cutoff Λ are a complete description of a QFT. No renormalization scale μ needed. What μ describes is a family of ways to absorb large (divergent) parts of loops into renormalized, or effective, couplings.

Different choices of μ change how much of the loop integrals are swept up into the effective couplings (see Sec. 1.7), but physical predictions should be independent of μ. We can enforce this, and exploit it, via

$$\frac{d(\text{bare param})}{d\mu} = 0. \tag{1.32}$$

For example, in scalar field theory we have

$$\frac{dm_0^2}{d\log\mu} = 0 = \frac{dz_{m^2} z_\varphi^{-1} m^2}{d\log\mu}$$
$$\Rightarrow \frac{d\log m^2}{d\log\mu} = -\frac{d\log\left(z_{m^2} z_\varphi^{-1}\right)}{d\log\mu} \tag{1.33}$$

So the physically equivalent choices of renormalized mass-squared parameter, and different scales μ, are related by an ordinary differential equation.

In φ^3 theory we computed the z's in a power series in the coupling. For example, with a cutoff regulator and a mass-dependent subtraction scheme, we obtained

$$z_\varphi = 1 + \delta z = 1 - \frac{g^2}{256\pi^3}\frac{1}{3}\log\Lambda^2/m^2 + \mathcal{O}(g^3) \tag{1.34}$$
$$z_{m^2} = 1 + \frac{\delta m^2}{m^2} = 1 + \frac{1}{m^2}\frac{g^2}{256\pi^3}\left(\Lambda^2 - \mu^2 - 2m^2\log\Lambda^2/\mu^2\right) + \mathcal{O}(g^3) \tag{1.35}$$
$$z_g = 1 + \frac{\delta g}{g} = 1 - \frac{g^2}{128\pi^3}\log\Lambda^2/\mu^2 + \mathcal{O}(g^3) \tag{1.36}$$

We see that the RHS of Eq. (1.33) is nonvanishing, so $m^2 = m^2(\mu)$. In general, renormalized parameters can be defined to "run" with the scale introduced during regularization and renormalization, according to first order, nonlinear, coupled ODEs called renormalization group equations (RGEs). It is convenient to define

$$\beta_{m^2} \equiv \frac{dm^2}{d\log\mu} = -m^2\frac{d\log\left(z_{m^2} z_\varphi^{-1}\right)}{d\log\mu} \tag{1.37}$$
$$\beta_g \equiv \frac{dg}{d\log\mu} = -g\frac{d\log\left(z_g z_\varphi^{-3/2}\right)}{d\log\mu} \tag{1.38}$$
$$\gamma_\varphi \equiv -\frac{d\log\varphi}{d\log\mu} = +\frac{d\log\left(z_\varphi^{1/2}\right)}{d\log\mu} \tag{1.39}$$

Let's compute $\beta_g, \beta_{m^2}, \gamma_\varphi$ in the mass-dependent scheme to one-loop order (leading order in powers of g.) First we use the chain rule to expand the total derivative $d/d\log\mu$:

$$\frac{d}{d\log\mu} \to \frac{\partial}{\partial\log\mu} + \underbrace{\frac{dg}{d\log\mu}}_{\beta_g}\frac{\partial}{\partial g} + \underbrace{\frac{dm^2}{d\log\mu}}_{\beta_m^2}\frac{\partial}{\partial m^2} \tag{1.40}$$

Then

$$\begin{aligned}\beta_g =& -g\left[\frac{\partial}{\partial\log\mu}+\beta_g\frac{\partial}{\partial g}+\beta_{m^2}\frac{\partial}{\partial m^2}\right]\log\left(z_g z_\varphi^{-3/2}\right)\\ =& -\frac{\partial\delta_g}{\partial\log\mu}+\frac{3}{2}g\frac{\partial\delta z}{\partial\log\mu}\qquad\text{(to leading order in } g)\\ =& -\frac{2g^3}{128\pi^3}+\frac{g^3}{256\pi^3}\\ =& -\frac{3g^3}{256\pi^3}.\end{aligned}\tag{1.41}$$

Similarly,

$$\begin{aligned}\beta_m^2 =& m^2\frac{\partial\delta z}{\partial\log\mu}-\frac{\partial\delta m^2}{\partial\log\mu}\qquad\text{(to leading order in } g)\\ =& \frac{g^2}{256\pi^3}\left(-\frac{10}{3}m^2+\mu^2\right).\end{aligned}\tag{1.42}$$

Note that the mass-squared β-function is explicitly μ-dependent.

In order for renormalization to work, all Λ-dependence has to drop out of these β-functions, so that they have smooth limits as the cutoff is removed. Renormalized perturbation theory accomplishes this order by order in weak coupling. At leading order, we get it for free. At higher order, it implies nontrivial cancellations that can be explicitly checked.

Now let's re-derive the 1-loop β-functions using $\overline{MS}$. In $6-2\epsilon$ dimensions, the bare action is

$$S=\int d^{6-2\epsilon}x\left[\frac{1}{2}(\partial_\mu\varphi_0)^2-m_0^2\varphi_0^2-\frac{g_0}{6}\varphi_0^3\right]\tag{1.43}$$

so the coupling is dimensionful,

$$[g_0]=\epsilon.\tag{1.44}$$

We require $\frac{dg_0}{d\log\mu}=0$. We'll compute β function for a *dimensionless* renormalized coupling, related to the bare coupling as $g_0=g\mu^\epsilon z_g z_\varphi{}^{-3/2}$. So our previous formulas (1.39) relating the beta functions to the z factors will require a small revision.

Again we expand the total derivative in partials. Note that because the counterterms are μ-independent in this scheme, the only explicit μ dependence arises from $\frac{\partial}{\partial\log\mu}\log(\mu^\epsilon)=\epsilon$. Then

$$0=\frac{d}{d\log\mu}\log(g\mu^\epsilon z_g z_\varphi^{-3/2})=\epsilon+\frac{1}{g}\beta_g+(\beta_g\frac{\partial}{\partial g}+\beta_{m^2}\frac{\partial}{\partial m^2})\log(z_g z_\varphi^{-3/2}).\tag{1.45}$$

There is no m^2 dependence, and we may solve this equation for β_g. We have to be a bit careful because there are two small numbers (g and ϵ). Keep the regulator finite until the very end, since we only expect our renormalization prescription to work to fixed order in g (we only computed one-loop divergences.) The solution has an expansion of the form $\beta_g \sim \frac{\epsilon}{1+\frac{1}{\epsilon}g^3} \sim \epsilon + g^3 + \dots$. Working it out, in the $d = 6$ limit we find

$$\beta_g = -\frac{3g^3}{256\pi^3} \tag{1.46}$$

after setting $\epsilon \to 0$ at the end. This is the same as (1.41): the β function for the coupling is scheme-independent at least through this loop order.

The mass squared beta function is obtained similarly. We don't have to add an epsilonic power of μ because the dimension of m^2 is still two. We write:

$$\begin{aligned} 0 &= \frac{d}{d\log\mu}\log(m^2 z_{m^2} z_\varphi^{-1}) \\ &= \frac{1}{m^2}\beta_{m^2} + (\beta_g\frac{\partial}{\partial g} + \beta_{m^2}\frac{\partial}{\partial m^2})\log(z_{m^2} z_\varphi^{-1}) \end{aligned} \tag{1.47}$$

and we find, in $d = 6$,

$$\beta_{m^2} = -\frac{5g^2 m^2}{384\pi^3} \tag{1.48}$$

which is the same as what we got in the mass-dependent scheme, without the μ^2 term. This is a general feature which may be taken as the definition of mass independent schemes:

- **in mass-independent schemes, the beta functions contain no explicit dependence on the renormalization scale.**

1.5 Quantum Dimensional Analysis 2

Now let us return to the scaling of correlation functions, including quantum corrections, and see the roles of the anomalous dimensions and beta functions. Consider again the momentum space correlator

$$\langle\varphi(p_1)\dots\varphi(p_n)\rangle \equiv G_n\left(p_i; m^2(\mu), g(\mu), \mu\right)\delta^d\Big(\sum_i p_i\Big). \tag{1.49}$$

Here we have in mind a general theory in d dimensions, where m^2 represents some mass parameter, g is a marginal coupling, and we omit other couplings for brevity.

Dimensional analysis tells us

$$G_n\left(xp_i; m^2(\mu), g(\mu), \mu\right) = x^A G_n\left(p_i; m^2(\mu)/x^2, g(\mu), \mu/x\right) \tag{1.50}$$

where $A = n(-1-d/2)+d$ counts up the classical scaling dimensions of the momentum space fields minus the scaling dimension of the momentum-conserving delta function. Also, using $\varphi \equiv z_\varphi^{-1/2}\varphi_0$ and $\gamma_\varphi \equiv \frac{1}{2}\frac{d\delta z}{d\log\mu}$, we have

$$\frac{dG_n}{d\log\mu} = -\frac{n}{2}\gamma_\varphi G_n. \tag{1.51}$$

Integrating this RGE,

$$G_n\left(p_i; m^2(\mu_0), g(\mu_0), \mu_0\right) = e^{\frac{n}{2}\int_{\mu_0}^{\mu_1}\gamma_\varphi d\log\mu} G_n\left(p_i; m^2(\mu_1), g(\mu_1), \mu_1\right). \tag{1.52}$$

Thus

$$G_n\left(xp_i; m^2(\mu_0), g(\mu_0), \mu_0\right) = e^{\frac{n}{2}\int_{\mu_0}^{\mu_1}\gamma_\phi d\log\mu} G_n\left(xp_i; m^2(\mu_1), g(\mu_1), \mu_1\right) \tag{1.53}$$

$$= e^{\frac{n}{2}\int_{\mu_0}^{\mu_1}\gamma_\phi d\log\mu} x^A G_n\left(p_i; \frac{m^2(\mu_1)}{x^2}, g(\mu_1), \frac{\mu_1}{x}\right). \tag{1.54}$$

Taking $\mu_1/x = \mu_0$, we find

$$G_n(xp_i; m^2(\mu_0), g(\mu_0), \mu_0) = e^{\frac{n}{2}\int_{\mu_0}^{x\mu_0}\gamma_\varphi d\log\mu} x^A G_n\left(p_i; \frac{m^2(x\mu_0)}{x^2}, g(x\mu_0), \mu_0\right). \tag{1.55}$$

Now suppose that we are in a regime where $\gamma_\varphi \approx \gamma_\varphi^*$ is approximately constant over the integral in the exponent. Then

$$e^{\frac{n}{2}\int_{\mu_0}^{x\mu_0}\gamma_\varphi d\log\mu} \approx x^{\frac{n}{2}\gamma_\varphi^*}. \tag{1.56}$$

Eq. (1.55) and its generalizations are our main result thus far. We conclude that the correlation function data of a QFT scales classically, with two modifications:

- The engineering dimension A appearing in Eq. (1.50) is modified by an "anomalous dimension" factor. In the case where γ is approximately constant,

  $$A_{\text{classical}} = n(-1 - d/2) + d \qquad \Rightarrow \qquad A_{\text{quantum}} = n(-1 - d/2 + \Delta) + d, \tag{1.57}$$

  where

  $$\Delta \equiv \gamma_\varphi^*/2. \tag{1.58}$$

- The dimensionless interaction strength is altered to $g(\mu_0) \to g(x\mu_0)$, where $\frac{\partial g}{\partial\log\mu} = \beta_g$. The positive-dimension coupling m^2 has a classical scaling piece (the $1/x^2$) and a quantum scaling $m^2(\mu_0) \to m^2(x\mu_0)$, where $\frac{dm^2}{d\log\mu} = \beta_{m^2}$.

This discussion is exact. In practice, the beta functions and anomalous dimensions are computed in perturbation theory, and the renormalization scale is chosen to be $\mu_0 \sim |p_i|$. Then (cf. Sec. 1.7) there are no large logs of the form $\log(p/\mu_0)$ in the perturbative expansion of $G_n(p; g, \mu_0)$.

The correlation functions are computed as functions of couplings, external momenta, and μ. So we can also expand the total derivative in Eq. (1.51) in partial derivatives:

$$\frac{dG_n}{d\log\mu} = \frac{\partial G_n}{\partial\log\mu} + \beta_{g_i}\frac{\partial G_n}{\partial g_i} + \beta_{m^2}\frac{\partial G_n}{\partial m^2} = -\frac{n}{2}\gamma_\varphi G \tag{1.59}$$

This is the Callan-Symanzik equation. It is sometimes used to derive the beta functions, as an alternative to the method used above. We will come back to the CS equation later when we discuss the anomalous dimensions of composite operators.

Below we will discuss consequences of scaling and interpretations of RG evolution from several perspectives. First, however, let us address item 3 from the list of outstanding issues above, interactions of negative mass dimension.

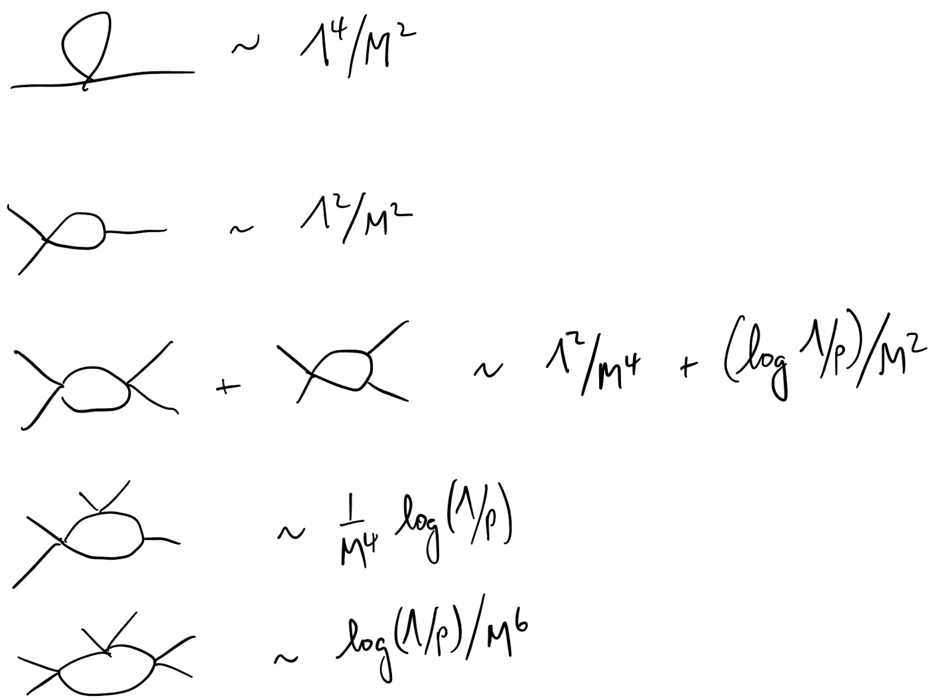
1.6 Renormalizing nonrenormalizable couplings

In our tree level analysis of scaling, it was easy to see why higher-dimension operators were "irrelevant" in the infrared. What about quantum corrections?

Let's add an irrelevant φ^4 operator to our 6D φ^3 theory:

$$L \supset -\frac{1}{6} g\varphi^3 - \frac{1}{24}\frac{\lambda}{M^2}\varphi^4 \tag{1.60}$$

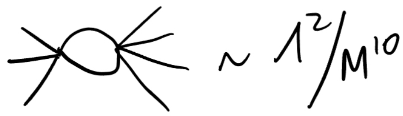
Here M is some large mass scale and $[\lambda_4] = 0$. Now there are new divergences already at one loop:

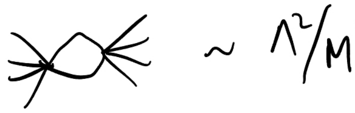


If we follow our nose, we would try to absorb into the existing counterterms $\delta z, \delta m^2, \delta g$, as well as a new counterterm $\delta\lambda_4$, and various other couplings and counterterms, including (but not limited to)

$$\frac{\lambda_5}{M^4}\varphi^5, \frac{\lambda_6}{M^6}\varphi^6 \Leftrightarrow \delta\lambda_5, \delta\lambda_6 \tag{1.61}$$

If we include φ^5 and φ^6 in the theory from the start, then we find additional divergences





which implies that we need to include an operator $\frac{\lambda_7}{M_8^8}\varphi^7$ with counterterm $\delta\lambda_7$, and an operator $\frac{\lambda_8}{M^{10}}\varphi^8$ with counterterm $\delta\lambda_8$. And so far we have not even considered higher dimension operators involving derivatives.

This seems to be spiraling out of control. However, all is not lost. Scaling gives us a way out. By dimensional analysis,

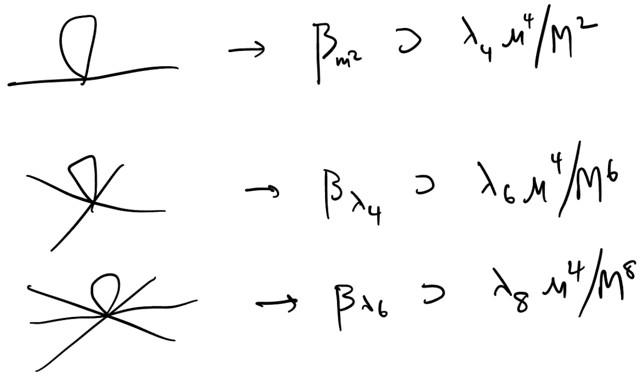
$$\begin{aligned} &G_n\left(xp_i; g(\mu_0), m^2(\mu_0), \frac{\lambda_4(\mu_0)}{M^2}, \frac{\lambda_6(\mu_0)}{M^4}, \ldots, \mu_0\right) \\ &= x^A e^{\frac{n}{2}\int_{\mu_0}^{\mu_1}\gamma_\varphi d\log\mu} G_n\left(p_i; g(\mu_1), \frac{m^2(\mu_1)}{x^2}, \frac{\lambda_4(\mu_1)}{M^2}x^2, \frac{\lambda_6(\mu_1)}{M^4}x^4, \ldots, \frac{\mu_1}{x}\right) \\ &= x^A e^{\frac{n}{2}\int_{\mu_0}^{x\mu_0}\gamma_\varphi d\log\mu} G_n\left(p_i; g(\mu_1), \frac{m^2(\mu_1)}{x^2}, \frac{\lambda_4(\mu_1)}{M^2}x^2, \frac{\lambda_6(\mu_1)}{M^4}x^4, \ldots, \mu_0\right) \end{aligned} \tag{1.62}$$

where we set $\mu_1 = x\mu_0$ in the last line. Now suppose that $x \ll 1$ and $m \ll M$, and further suppose that μ_0 and p_i are not much larger than M (although they might be of order M.). Then we see that the low-frequency, long-wavelength correlation functions of the full theory (the left hand side of (1.62)) are equal to the (scaled) correlation functions of the original renormalizable theory (right-hand side of (1.62) with $\vec{\lambda} = 0$), with a modified coupling $g \to g(x\mu_0)$, and with small corrections from higher-dimension operators. The effects of the HDOs are suppressed by energy ratios x and m/M: contributions from $\lambda_4(\mu_1)$ are of $\mathcal{O}(x^2)$ or $\mathcal{O}(m^2/M^2)$; contributions from $\lambda_6(\mu_1)$ are of $\mathcal{O}(x^4)$ or $\mathcal{O}(m^4/M^4)$, and so on. This is almost the same as our classical scaling analysis, just with some RG for the dimensionless couplings and the mass squared parameter, and the anomalous dimension factor.

The conclusions is this: **if we compute correlation functions using perturbation theory in both dimensionless couplings and in dimensionless ratios of p_{ext}/M and m/M, then at fixed order in these ratios, we only need to specify a finite number of renormalized couplings – up to a fixed, maximal operator dimension.** Practically speaking, we can just throw out divergences that are proportional to higher orders in p_{ext}/M and m/M. In principle we could keep them, and absorb them into the renormalized couplings of higher dimension operators, but afterward, what Eq. (1.62) tells us is that these operators do not contribute to the low-energy correlation functions at the prescribed order in energy ratios.

This is the generalization of "renormalizable" to "effective" QFTs. We also see why, at extremely low energies, the theory becomes arbitrarily well-approximated by a renormalizable QFT.

With a mass-dependent renormalization scheme, we are actually restricted to $\mu_0 \ll M$. Otherwise, in running the couplings down to μ_1, we would require all λ_i in the β functions:



This can be truncated for $\mu \ll M$, but not for $\mu \sim M$. If $\mu_0 \sim M$ then in a typical mass-dependent scheme, operators of all orders mix with each other strongly at the start of the flow down to μ_1. Thus even if we plan on working at low energies at some fixed operator dimension D, we cannot solve the RGEs for the values of these couplings at low scales, because they receive unsuppressed contributions from operators of all dimensions at the start of the flow.

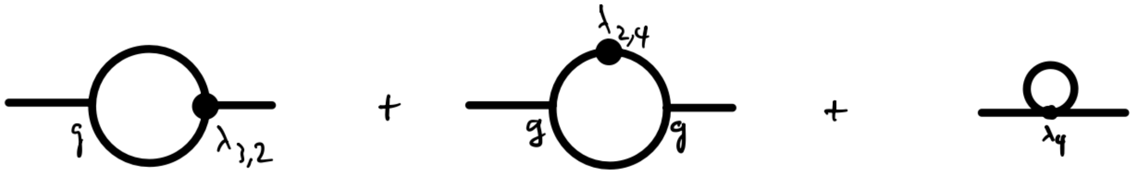
We can do better by using a mass-independent subtraction scheme. (There is a price for this, related to decoupling, discussed later.)

As an example, let us investigate the φ^3 theory including operators up to dimension 8 (corresponding to p_{ext}^2/M^2), using $DR + \overline{MS}$ to regulate and renormalize. Up to total derivatives, the possible dimension 8 operators include:

$$\lambda_4\phi^4, \quad \lambda_{3,2}\phi^2\partial^2\phi, \quad \lambda_{2,4}(\partial^2\phi)^2 \tag{1.63}$$

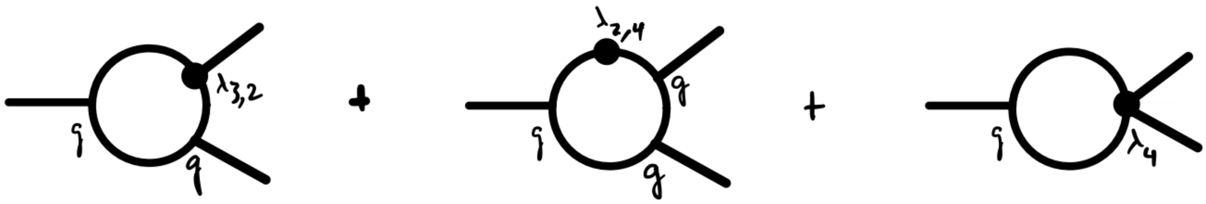
where we have written them together with dimensionless couplings, collectively referred to as $\vec{\lambda}$. $\phi^2\partial^2\phi$ is associated with contributions to 3-point 1PI diagrams proportional to p_{ext}^2. $(\partial^2\phi)^2$ is associated with contributions to the self-energy proportional to p_{ext}^4. Actually, to $\mathcal{O}(\vec{\lambda})$, $\phi^2\partial^2\phi$ and $(\partial^2\phi)^2$ can be both removed by field redefinitions, and are called "redundant." We will discuss this more below, but to fully renormalize the theory off-shell, we may need to keep them around.

At one-loop order, to first order in the $\vec{\lambda}$, the following diagrams and counterterms arise:



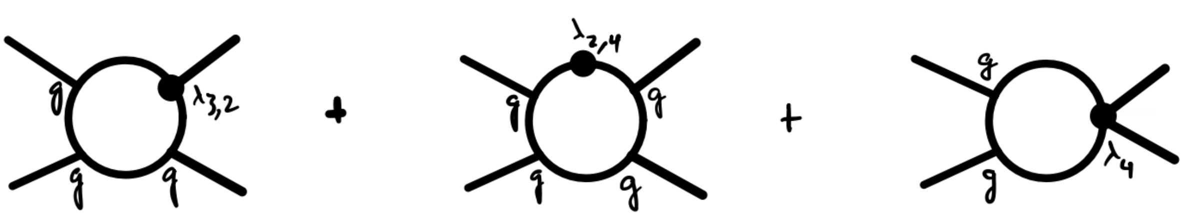
$$\Rightarrow \quad \delta_Z, \ \frac{\delta m^2}{m^2} \sim \mathcal{O}\left(\frac{\lambda_4 m^2}{M^2}, \ \frac{\lambda_{3,2}\, g m^2}{M^2}, \ \frac{\lambda_{2,4}\, g^2 m^2}{M^2}\right) \times \frac{1}{\epsilon}$$

$$\delta\lambda_{2,4} \sim \mathcal{O}\left(\lambda_{3,2}\, g, \ \lambda_{2,4}\, g^2\right) \times \frac{1}{\epsilon}$$



$$\Rightarrow \quad \delta g \sim \mathcal{O}\left(\frac{\lambda_{3,2}\, g^2 m^2}{M^2}, \ \frac{\lambda_{2,4}\, g^3 m^2}{M^2}\right) \times \frac{1}{\epsilon}$$

$$\delta\lambda_{3,2} \sim \mathcal{O}\left(\lambda_{3,2}\, g^2, \ \lambda_{2,4}\, g^3\right) \times \frac{1}{\epsilon}$$



$$\Rightarrow \quad \delta\lambda_4 \sim \mathcal{O}\left(\lambda_4\, g^2, \ \lambda_{3,2}\, g^3, \ \lambda_{2,4}\, g^4\right) \times \frac{1}{\epsilon}$$

Correspondingly, we expect to find new contributions to the beta functions of order

$$\begin{aligned}
&\beta_{m^2} \sim m^2 \times \mathcal{O}\left(\frac{\lambda_4 m^2}{M^2},\ \frac{g\lambda_{3,2} m^2}{M^2},\ \frac{g^2\lambda_{2,4} m^2}{M^2}\right) \\
&\beta_g \sim \mathcal{O}\left(\frac{g\lambda_4 m^2}{M^2},\ \frac{g^2\lambda_{3,2} m^2}{M^2},\ \frac{g^3\lambda_{2,4} m^2}{M^2}\right) \\
&\beta_{\lambda_4} \sim \mathcal{O}\left(g^2\lambda_4,\ g^3\lambda_{3,2},\ g^4\lambda_{2,4}\right) \\
&\beta_{\lambda_{3,2}} \sim \mathcal{O}\left(g\lambda_4,\ g^2\lambda_{3,2},\ g^3\lambda_{2,4}\right) \\
&\beta_{\lambda_{2,4}} \sim \mathcal{O}\left(g\lambda_{3,2},\ g^2\lambda_{2,4}\right).
\end{aligned} \tag{1.64}$$

We'll just work out the $\mathcal{O}(\lambda_4)$ contributions to β_{m^2}, β_g, and β_{λ_4}, without doing a complete analysis. It turns out that for this purpose it is sufficient to set external momenta to zero. Recall that

$$\int \frac{d^d k}{(k^2+\Delta)^\alpha (2\pi)^d} = \frac{1}{(4\pi)^{d/2}} \frac{\Gamma\left(\alpha - \frac{d}{2}\right)}{\Gamma(\alpha)} \left(\frac{1}{\Delta}\right)^{\alpha - \frac{d}{2}}. \tag{1.65}$$

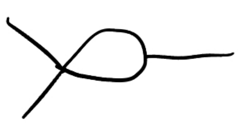
The first new diagrams is:

$$\begin{aligned}
&= -\frac{i\lambda_4}{2M^2} \int \frac{i d^6 k}{(2\pi)^6} \frac{i}{k^2+m^2} \qquad \text{(Euclidean)} \\
6 \to d = 6-2\epsilon \Rightarrow &\ \frac{i\lambda_4}{2(4\pi)^3} \left(\frac{4\pi\mu^2}{m^2}\right)^\epsilon \frac{m^4}{M^2} \underbrace{\Gamma(-2+\epsilon)}_{\frac{1}{2}\left(\frac{1}{\epsilon} - \gamma + 3/2 + \cdots\right)} \\
&\Rightarrow \frac{i\lambda_4}{256\pi^3} \frac{m^4}{M^2} \left(\frac{1}{\epsilon} + \log(\tilde{\mu}^2/m^2) + 3/2 + \ldots\right)
\end{aligned} \tag{1.66}$$

We can absorb the $1/\epsilon$ into δm^2:

$$\delta m^2 = -\frac{1}{128\pi^3}\left(g^2 - \frac{\lambda_4 m^2}{2M^2}\right) m^2 \frac{1}{\epsilon} \tag{1.67}$$

Next, we have:

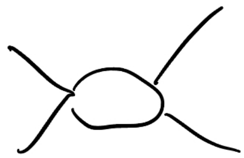


$$\begin{aligned}
&= \frac{3}{2}\left(-\frac{i\lambda_4}{M^2}\right)(-ig) \int \frac{i d^6 k}{(2\pi)^6} \left(\frac{i}{k^2+m^2}\right)^2 \\
&\Rightarrow \frac{3ig\lambda_4}{2(4\pi)^3} \left(\frac{4\pi\mu^2}{m^2}\right)^\epsilon \frac{m^2}{M^2} \frac{\Gamma(-1+\epsilon)}{\Gamma(2)} \\
&\Rightarrow \frac{3i\lambda_4 g}{128\pi^3} \frac{m^2}{M^2} \left(-\frac{1}{\epsilon} + \ldots\right)
\end{aligned} \tag{1.68}$$

Here $\alpha = 2$, so $\alpha - d/2 \to -1 + \epsilon$.[2] The modified counterterm:

$$\delta g = -\left(g^3 + 3g\lambda_4 \frac{m^2}{M^2}\right) \frac{1}{128\pi^3} \frac{1}{\epsilon} \tag{1.69}$$

And finally[3]



$$= 6 \times \left(-\frac{i\lambda_4}{M^2}\right) (-ig)^2 \int \frac{id^6k}{(2\pi)^6} (-1)^3 \left(\frac{i}{k^2 + m^2}\right)^3$$

$$= -\frac{6i\lambda_4 g^2}{M^2(4\pi)^3} \left(\frac{1}{\Gamma(3)\epsilon} + \dots\right) \tag{1.70}$$

$$\Rightarrow \delta\lambda_4 = -\frac{2\lambda_4 g^2}{(4\pi)^3} \frac{1}{\epsilon}. \tag{1.71}$$

(We only work to $O(\lambda_4)$, dropping diagrams of order λ_4^2.)

Now we can compute the modifications to the β functions. Repeating the previous procedure, we obtain

$$\begin{aligned}
\beta_{m^2} &= -\frac{5g^2m^2}{384\pi^3} + \frac{\lambda_4 m^4}{128M^2\pi^3} + \cdots \\
\beta_g &= -\frac{3g^3}{256\pi^3} - \frac{3g\lambda_4 m^2}{64\pi^3 M^2} + \cdots \\
\beta_{\lambda_4} &= \frac{g^2\lambda_4}{192\pi^3} + \cdots \\
&\dots
\end{aligned} \tag{1.72}$$

where ellipses denote contributions from $\lambda_{3,2}$ and $\lambda_{2,4}$ and their beta functions. These results are consistent with our squiggle-estimates of the necessary counterterms at this order.

Notable features of the result:

---

[2]The coefficient 3/2 is

$$\frac{1}{3! \cdot 4!} (3 \cdot 4 \cdot 3 \cdot 3 \cdot 2) = \frac{3}{2}$$

which comes from a symmetry factor a sum over three diagram topologies, which all give the same result because we compute the diagram at zero external momentum. I have about 60% confidence in this factor.

Here is everything I know about symmetry factors. If a given vertex appears n times in a diagram, we leave off the $1/n!$ it should have come with from the Taylor expansion of $e^{-iH_{int}t}$, under the assumption that it will cancel in the sum over all diagrams which differ only by a permutation of identical vertices. Then we draw all diagrams with fixed external legs which are distinct ("different topologies") under permutation of the vertices. Our Feynman rules leave off the 1/4! in the coupling because that will mostly cancel with different contractions which give the same diagram topology. On a napkin, we scribble the contraction combinatorics to make sure these rules didn't over/under-count, and then we multiply each diagram by a correction factor (one over the "symmetry factor"), as needed. Then we throw away the napkin and forget how we did it.

[3]The coefficient 6 is reported at 55% confidence. Again this includes a sum over diagram topologies, which all give the same result because we compute the diagram at zero external momentum.

- λ_4 only impacts β_g and β_{m^2} at order m^2/M^2, which $\to 0$ as $m^2 \to 0$. More generally, **higher dimension operators decouple from the beta functions of lower dimensional operators in mass independent subtraction schemes in the massless limit.** This is obvious from dimensional analysis, and it is very useful for studying RGEs of massless or light fields.

- $\beta_{m^2} \sim m^2$, not $\sim \mu^2$. If the dimensionless couplings are perturbative and $m^2 \ll M^2$, then all the beta functions are perturbative. No extra condition on μ^2 is required to control which operators get generated, and in particular taking $\mu \sim M$ is fine – this will be useful later.

These results are general. In a mass-independent scheme, the perturbative expansion is structured in a way that makes power counting much easier.

As discussed, to complete the computation of the beta functions of the dimension-8 operators we would need to obtain $\beta_{\lambda_{3,2}}$ and $\beta_{\lambda_{2,4}}$. However, at any renormalization scale, to $\mathcal{O}(\vec{\lambda})$, these operators can be removed by field redefinitions. This works as follows. Suppose we take

$$\phi \to \phi + \underbrace{a\partial^2\phi + b\phi^2 + c\phi}_{\delta\phi}, \qquad a, b, c \sim \mathcal{O}(\vec{\lambda}^2). \tag{1.73}$$

Then

$$\delta S = \int d^6x\,(\text{EOM})\,|_{\vec{\lambda}=0} \times \delta\phi + \mathcal{O}(\vec{\lambda}^2) \tag{1.74}$$

The equation of motion with $\vec{\lambda} = 0$ is $\partial^2\phi + m^2\phi + \frac{1}{2}g\phi^2 = 0$. So

$$\delta\mathcal{L} = a(\partial^2\phi)^2 + (ag/2 + b)\phi^2\partial^2\phi + (am^2 + c)\phi\partial^2\phi + (\text{no derivs}) \tag{1.75}$$

So if we choose

$$a = -\frac{\lambda_{2,4}}{M^2}, \qquad b = -\frac{1}{2}ag^2 - \frac{\lambda_{3,2}}{M^2}, \qquad c = -am^2 \tag{1.76}$$

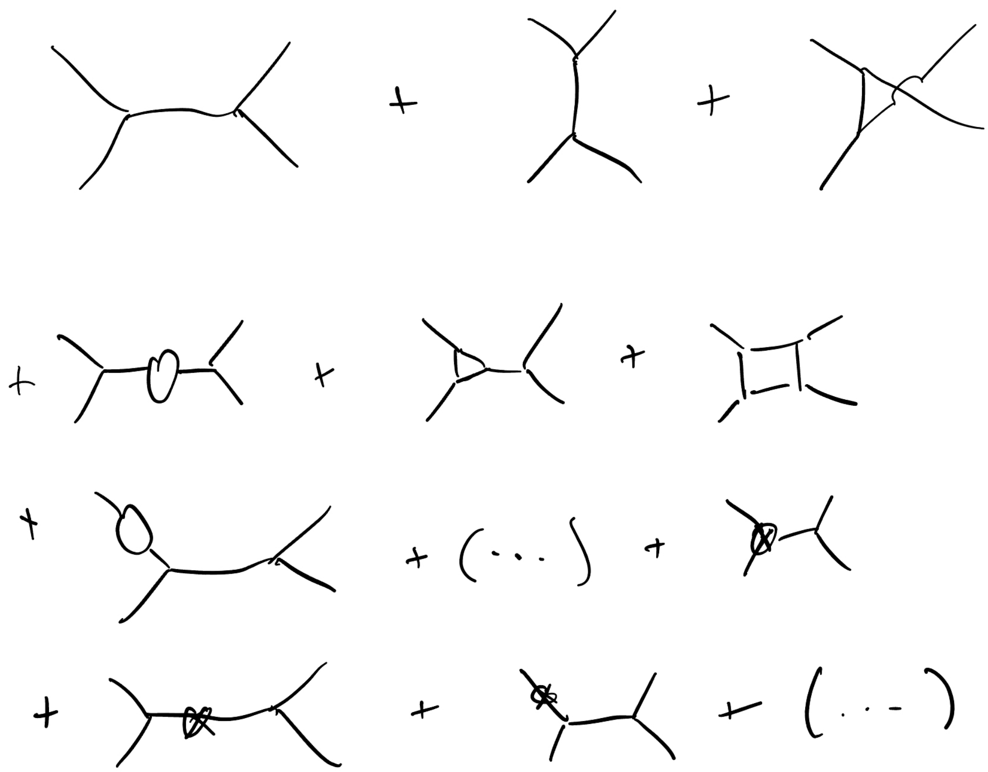
then we can remove $\lambda_{2,4}$ and $\lambda_{3,2}$, to $\mathcal{O}(\vec{\lambda}^2)$ and keep the kinetic term canonically normalized. This works in general for operators that are proportional to equations of motion (or equivalent to other operators plus equations of motion). Field redefinitions cannot affect physical observables like S-matrix elements, so these operators are called "redundant."

A final note – we dropped total derivative operators. This is because any such operator, comprised of n elementary fields, has a Feynman rule proportional to $\sum_{i=1}^n p_i^\mu$, and therefore vanishes by momentum conservation.

1.7 Large log resummation

Another way to think of the RG is that it is reorganizing the perturbative series as μ is changed.

To illustrate, consider $2 \to 2$ scattering, or a 4-point correlation function, at one loop order in φ^3 theory. We have:



The counterterm diagrams absorb the divergences in the one-loop diagrams into the renormalized parameters, and they also introduce explicit dependence on μ_0. There will be terms in the one-loop Feynman integrals proportional to $\log \Lambda^2/(p^2+m^2)$, which become $\log \mu_0^2/(p^2+m^2)$ upon addition of the counterterm diagrams. These "large logs" dominate the radiative corrections for $\mu_0^2 \gg p^2+m^2$. If we use the RG, we reorganize the perturbation series so that these corrections are folded in already at tree level: using the RGEs with boundary condition at μ_0, we then run down to some $\mu^2 \simeq p^2+m^2$. Then the tree level is $\mathcal{O}(g^2(\mu))$ and large logs have disappeared from the radiative corrections. This is a cheap way of increasing precision, once you know the RGEs.

Suppose we have a perturbative expansion of an amplitude

$$\begin{aligned} A(p) &= ag^2(\mu_0) + bg^4(\mu_0)\log p^2/\mu_0^2 + \dots \\ &= ag^2(\mu_1) + bg^4(\mu_1)\log p^2/\mu_1^2 + \dots \end{aligned} \tag{1.77}$$

The coupling at μ_1 is related to the coupling at μ_0 by the RGE, $\frac{\partial g}{\partial \log \mu} = \beta(g)$. Suppose $\mu_0 \gg \mu_1 \sim p$. Then the first line has large logarithms, but the second line does not: they have been swept up into the running coupling.

In fact it's even better than that. Take a solution to the RGEs. For example, if $\frac{dg}{d\log\mu} = -bg^3$, then

$$\int_{g_0}^{g} \frac{1}{g'^3} dg' = -b \log(\mu/\mu_0)$$
$$\Rightarrow g^2 = \frac{g_0^2}{1 + 2b g_0^2 \log(\mu/\mu_0)} \tag{1.78}$$

We will examine the physics of this expression more closely in a moment, but first, consider what we get if we compute a tree-level scattering amplitude with $g^2(\mu_0)$ and $g^2(\mu)$. The amplitudes just are proportional to $g^2(\mu_0)$ and $g^2(\mu)$, respectively. But

$$\begin{aligned} g^2(\mu) = g^2(\mu_0) &- g^4(\mu_0) 2b \log \mu/\mu_0 \\ &+ g^6(\mu_0) 4b^2 \log^2 \mu/\mu_0 \\ &- g^8(\mu_0) 8b^3 \log^3 \mu/\mu_0 \\ &+ \ldots \end{aligned} \tag{1.79}$$

For $\mu_0 \gg p$, there are large logs $\sim g^4 \log(\mu_0^2/p^2)$ in the radiative corrections to the first amplitude. As discussed above, for a different choice of scale μ, these corrections have been absorbed into the running coupling $g^2(\mu)$, up to $\log \mu^2/p^2$ corrections. The latter are small if μ^2 is chosen intelligently. But we also got capture $g^6 \log^2, g^8 \log^3, \ldots$. These are real $2, 3, 4, \ldots$ loop corrections in the μ_0 case! We get them *for free* from the RGE, in the tree level analysis with $\mu \sim p$. This is known as "large log resummation."

Caveats:

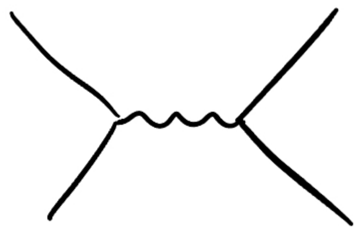
- Even with a well-chosen renormalization scale, we still have 1-loop corrections of the form $\log(\frac{\mu^2}{p^2+m^2})$ and non-logarithmic terms.
- We still have $(g^2)^a (g^2 \log(\mu/\mu_0))^n$ corrections with $a > 1$:

$$\begin{aligned} &a = 1 : \text{resummed by 1-loop } \beta \text{ function} \\ &a = 2 : \text{resummed by 2-loop } \beta \text{ function} \\ &a = 3 : \text{resummed by 3-loop } \beta \text{ function} \\ &\text{etc.} \end{aligned} \tag{1.80}$$

But in general RGEs are a good tool to soak up a lot of radiative corrections, if your couplings are accurately measured at one characteristic momentum scale and you want to make a prediction at a smaller momentum scale.

We will illustrate the utility of log resummation further in some EFT examples later. For now, let's just see that for the two-point function, it is closely related to bubble chain resummation.

Recall that nonrelativistic $2 \to 2$ scattering amplitude is proportional to $\tilde{V}(q)$, the Fourier transform of the potential one would put into the Schrödinger equation:

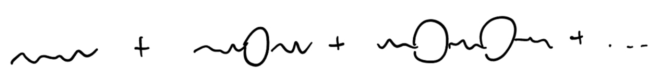


$$\sim \frac{e^2}{\vec{p}^2} \xrightarrow[FT]{} \frac{e^2}{4\pi r} = V(r) \tag{1.81}$$

We can also include the one loop corrections and find the (Fourier transform of the) potential:

$$\tilde{V}^{1-\text{loop}} = \frac{e^2}{p^2}\left(1 + \frac{e^2}{12\pi^2}\log p^2/p_0^2\right) \tag{1.82}$$

where the renormalization condition was $\tilde{V}(p_0)^{1-\text{loop}} = \frac{e^2}{p_0^2}$. Gauge invariance implies that the vertex and electron self-energy one loop corrections cancel, so only the vacuum polarization contributes to the corrected potential. The bubble chain



is a geometric series. We can sum it up:

$$\tilde{V}(p)^{\text{resummed}} = \frac{1}{p^2}\left[\frac{e^2}{1 - \frac{e^2}{12\pi^2}\log p^2/p_0^2}\right]. \tag{1.83}$$

However this could equally well be obtained by requiring $\tilde{V}(p)^{1-\text{loop}}$ to be independent of unphysical scale p_0:

$$\frac{d\tilde{V}^{1-\text{loop}}}{d\log p_0} = 0 \to \frac{de(p_0)}{d\log p_0} = \frac{e^3}{12\pi^2} \tag{1.84}$$

$$\to e^2(p) = \frac{e^2(p_0)}{1 - \frac{e^2(p_0)}{12\pi^2}\log p^2/p_0^2}. \tag{1.85}$$

then $\frac{e^2(p)}{p^2} = \tilde{V}(p)^{\text{resummed}}$ above.

Using RGE sums up diagrams and large logs without having to write the diagrams past one loop.

1.8 Composite operators

So far, we have only discussed correlation functions of the fundamental fields. Correlation functions with insertions of composite operators, e.g. φ^2, $(\varphi\partial_\mu\varphi)^2$, etc exhibit additional divergences. These can be absorbed by rescaling the operators. In general, the operators also mix under RG. We write a very general operator renormalization as:

$$\mathcal{O}_{0,i} \equiv Z_i^j \mathcal{O}_j \tag{1.86}$$

with sum over j implied. Here the $\mathcal{O}_j$ ($\mathcal{O}_{0,j}$) are some collection of renormalized (bare) operators. Then we define an "anomalous dimension matrix" as

$$\frac{d\mathcal{O}_i}{d\log\mu} = \frac{d(Z^{-1}\mathcal{O}_0)_i}{d\log\mu} = -\left(Z^{-1}\frac{dZ}{d\log\mu}Z^{-1}\mathcal{O}_0\right)_i = -\left(\frac{d\log Z}{d\log\mu}\right)_i^j \mathcal{O}_j \equiv -\gamma_i^j \mathcal{O}_j. \tag{1.87}$$

If some operators have protected dimensions (like conserved currents), or don't mix with each other (usually for symmetry reasons, or on dimensional grounds), there will be zeroes in this matrix. You can find simple examples demonstrating how to compute operator anomalous dimensions diagrammatically in most textbooks.

We'll do something else. We've already discussed how to renormalize theories with irrelevant operators in the action, to arbitrary fixed order in p/M. You might think that since the action is composed of local operators, the beta functions for the running couplings and the operator anomalous dimension matrix described above might be more or less the same thing. Let's derive the relationship.

Suppose that the Euclidean action contains a collection of operators

$$S \supset \int d^d x \sum_i z_i \lambda_i \mathcal{O}_i. \tag{1.88}$$

Now let us add

$$S \to S + \sum_i J_i \mathcal{O}_i. \tag{1.89}$$

The J_i are small constants. (More generally they can be slowly varying functions, but constants will be sufficient for our purposes.) Since the J's multiply operators that are already present in the action, no new counterterms are needed to renormalize the theory. Let us also assume that the field strength renormalization factors for the elementary fields are at least quadratic in the J_i.

We will use tildes to denote quantities in the theory with the J_i turned on, and remove the tildes to denote quantities in the theory with all J_i set to zero. For example,

$$\tilde{G}^n = G^n - \sum_k J_k G^{k;n} + O(J^2). \tag{1.90}$$

Here G^n is an n-point momentum space Euclidean correlation function of elementary fields, e.g. $G^n = \langle \varphi(p_1)\dots\varphi(p_n)\rangle$, and $G^{k;n}$ is the same correlation function with a zero-momentum insertion of $\mathcal{O}_k$, e.g. $G^{k;n} = \langle \mathcal{O}_k(0)\varphi(p_1)\dots\varphi(p_n)\rangle$. Here we have used that a derivative wrt J_i "brings down an $\int d^d x\, \mathcal{O}_i$" in the path integral – i.e. a zero-momentum $\mathcal{O}_i$ insertion in the correlation function. The beta functions are

$$\tilde{\beta}_{\lambda_i} = \beta_{\lambda_i} + \sum_k J_k \frac{\partial \beta_{\lambda_i}}{\partial \lambda_k} + O(J^2) \tag{1.91}$$

where we used $\tilde{\beta}_{\lambda_j}(\lambda_i) = \beta_{\lambda_j}(\lambda_i + J_i)$, and, by assumption, the anomalous dimensions of the elementary fields are $\tilde{\gamma} = \gamma + O(J^2)$.

Then the $O(J)$ term in the Callan-Symanzik equation for $\tilde{G}^n$ is

$$-\gamma_i^k G^{i;n} + \sum_i \frac{\partial \beta_{\lambda_i}}{\partial \lambda_k} \frac{\partial}{\partial \lambda_i} G^n = 0. \tag{1.92}$$

But $\partial G^n/\partial \lambda_i = \partial \tilde{G}^n/\partial J_i|_{J=0} = G^{i;n}$, so

$$\sum_i \left[\gamma_i^k - \frac{\partial \beta_{\lambda_i}}{\partial \lambda_k}\right] G^{i;n} = 0. \tag{1.93}$$

For this to hold for all $G^{i,n}$, we require

$$\gamma_i^k = \frac{\partial \beta_{\lambda_i}}{\partial \lambda_k}. \tag{1.94}$$

In other words, the operator anomalous dimension matrix just contains the same information about operator mixing as the running couplings in the effective action.

Consider, for example, $m^2\phi^2$ in a perturbative scalar field theory. Then

$$\frac{d(m^2\phi^2)}{d\log\mu} = \beta_{m^2}\phi^2 - m^2\gamma_{\phi^2}\phi^2. \tag{1.95}$$

But $\beta_{m^2} = m^2\partial_{m^2}\beta_{m^2}$ on dimensional grounds. So $m^2\phi^2$ is renormalization group invariant.

1.8.1 Example: Strange Physics

When we discuss EFT from the top down, we will describe how Fermi's electroweak theory can be obtained as a low-energy effective description of the Standard Model electroweak theory. The basic structure is that below $\mu \simeq m_W$, the Lagrangian of QED and QCD must be supplemented by a collection of 4-fermion operators mediating a variety of meson decays and mixings and lepton decays. Here we use some of the operators associated with kaons to illustrate higher-dimension operator mixing and renormalization.

Kaons are light mesons containing one strange quark and one light quark. Strangeness-number-changing operators mediate kaon production/decay ($\Delta S = 1$) and $K^0 - \bar{K}^0$ oscillations ($\Delta S = 2$). Only $\Delta S = 1$ operators are present in the effective Lagrangian at tree level at $\mu = m_W$.[4] These operators have the form

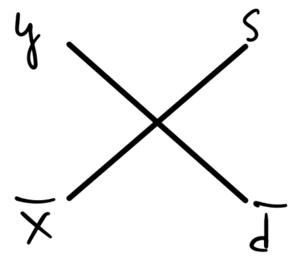
$$\begin{aligned}\mathcal{L} \supset& \frac{G_F}{\sqrt{2}} \sum_{x,y=u,c} V_{xs}V_{dy}^\dagger \left(\bar{x}_\alpha\gamma^\mu(1+\gamma_5)s_\alpha\right)\left(\bar{d}_\beta\gamma_\mu(1+\gamma_5)y_\beta\right)\\ =& \frac{G_F}{\sqrt{2}} \sum_{x,y=u,c} V_{xs}V_{dy}^\dagger \left(O_+^{\bar{x}y} + O_-^{\bar{x}y}\right)\end{aligned} \tag{1.96}$$

[4]As we will see, this is because single-W exchange can only change S by one unit.

where it is useful to define

$$2O_\pm^{\bar{x}y} = \bar{x}_\alpha\gamma^\mu(1+\gamma_5)s_\alpha\bar{d}_\beta\gamma_\mu(1+\gamma_5)y_\beta \pm \bar{x}_\alpha\gamma^\mu(1+\gamma_5)y_\alpha\bar{d}_\beta\gamma_\mu(1+\gamma_5)s_\beta$$
$$= \bar{x}_\alpha\gamma^\mu(1+\gamma_5)s_\alpha\bar{d}_\beta\gamma_\mu(1+\gamma_5)y_\beta \pm \bar{x}_\alpha\gamma^\mu(1+\gamma_5)s_\beta\bar{d}_\beta\gamma_\mu(1+\gamma_5)y_\alpha. \tag{1.97}$$

These are four-fermi operators,



(The particular linear combinations have nice transformation properties under the exchange of s and y color indices, and renormalize in a nice way.) In these interactions, the quark fields are written in the basis where the mass matrices are diagonal, and V is the unitary CKM matrix which arises due to the mismatch between the basis that diagonalizes the down-type and up-type quark mass matrices. α, β are color $SU(3)$ indices. The second line of Eq. (1.97) is a Fierz rearrangement of the first line.

Eq. (1.96) is an effective Lagrangian at renormalization scale $\mu = m_W$. Below m_W, radiative corrections cause the coefficients of the $O_\pm^{\bar{x}y}$ to run. If we ignore quark masses, we only need to introduce different effective couplings for O_+ and O_-, without dependence on $\bar{x}$ and y:

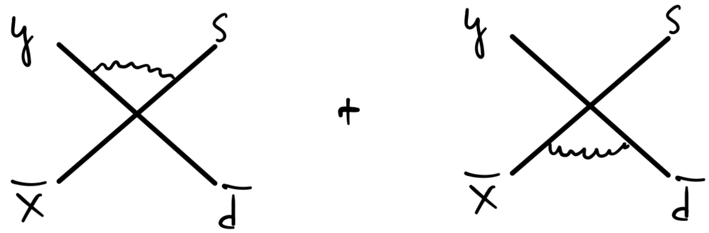
$$\mathcal{L} \supset = \frac{G_F}{\sqrt{2}} \sum_{x,y=u,c} V_{xs}V_{dy}^\dagger \left(h_+(\mu)O_+^{\bar{x}y} + h_-(\mu)O_-^{\bar{x}y}\right). \tag{1.98}$$

Since all of these operators are dimension 6, their couplings h_i are dimension -2. On dimensional grounds, continuing to ignore quark masses, the beta functions must have the form

$$\frac{dh_i}{d\log\mu} = \beta_i(h_j, g) = \gamma_{ij}(g)h_j. \tag{1.99}$$

Comparing with Eq. (1.94), we see that γ_{ij} is just the anomalous dimension matrix describing mixing of the dimension 6 operators. g denotes some gauge couplings, of which the dominant is the strong coupling g_s. At order $\alpha_s = g_s^2/4\pi$, γ is generated by attaching a gluon loop to the quark lines, e.g.[5]

[5]Here the labels on the external lines correspond to fields in an operator, rather than states in an amplitude. In all amplitude diagrams, the in state will be on the left and the out state on the right. So the "bars" on the s and the d would be swapped in these diagrams, if they were representing an amplitude.



So at this order the anomalous dimension matrix has the form $\gamma_{ij} = g_s^2 A_{ij}$, where A is a numerical matrix. Using $g_s^2 = -\beta_{g_s}/(bg_s)$ for the one-loop QCD coupling, the solution to Eq. (1.99) is

$$h_i(\mu) = \left(e^{-A/b \log(g_s/g_s(\mu_0))}\right)_{ij} h_j(\mu_0) \tag{1.100}$$

which can be easily verified. If we have a linear combination of operators that is a left-eigenvector of A, $\nu A = a\nu$, then that combination does not mix with other operators, and is only multiplicatively renormalized:

$$\nu \cdot h(\mu) = \left(\frac{g_s(\mu)}{g_s(\mu_0)}\right)^{-a/b} \nu \cdot h(\mu_0). \tag{1.101}$$

The linear combinations $O_\pm$ are such eigenvectors. For O_+, the exponent is

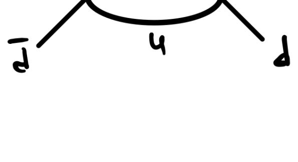
$$-a/b = -(1/(2\pi^2))/(11 - 2/3N_f)/(8\pi^2) = -12/23 \tag{1.102}$$

between m_W and m_b, where there are five light flavors active in the QCD beta function (discussed in more detail below). For O_-, the anomalous dimension is -2 times that of O_+, and so the exponent is $-a/b = 24/23$.

If we don't ignore quark masses, then effects of order h^2 and higher can appear in the RGE. These will be suppressed by $m_q^2 G_F$, which is not small for the top quark (in fact, $m_t > m_W$, so it is a bit delicate how to include the top quark at all), although its effects are instead suppressed by small CKM elements. Here we just note one effect, the generation of the $\Delta S = 2$ operator $O_+^{\bar{d}s}$ by the RGEs. The notation here is the same as for the $\Delta S = 1$ operators,

$$O_+^{\bar{d}s} = \frac{1}{2}\left(\bar{d}_\alpha \gamma^\mu (1+\gamma_5) s_\alpha \bar{d}_\beta \gamma_\mu (1+\gamma_5) s_\beta + \bar{d}_\alpha \gamma^\mu (1+\gamma_5) s_\beta \bar{d}_\beta \gamma_\mu (1+\gamma_5) s_\alpha\right). \tag{1.103}$$

$O_+^{\bar{d}s}$ is not present in the effective Lagrangian at $\mu = m_W$, but it can be generated by two $\Delta S = 1$ operators with mass insertions in the RGE flow down to the scales of meson physics. Some such diagrams are finite; we will see how to deal with them later. The divergent diagrams have the form

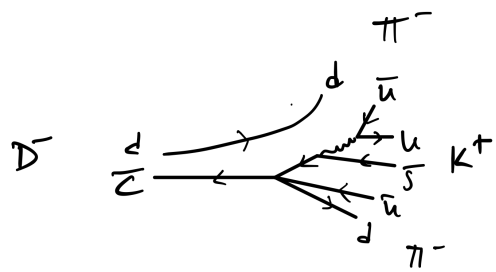


Divergences in this diagram contribute to a beta function for a coupling of the form $V_{cs}V_{dc}^{\dagger}V_{ts}V_{dt}^{\dagger}h_{ct}O_{+}^{\bar{d}s}$ (t-components of V appear by using unitarity.) The coefficient h_{ct} is also renormalized by QCD loops. In total, the dominant contribution to its beta function is of the form (Gilman & Wise 1983)

$$\frac{dh_{ct}}{d\log\mu} = \beta_{ct} = \frac{\alpha_s}{\pi}h_{ct} - \frac{G_F m_c^2}{8\pi^2}\left(\frac{3}{2}h_+^2 + \frac{1}{2}h_-^2 - h_+h_-\right) \tag{1.104}$$

where m_c is the charm mass and $h_{\pm}$ are the coefficients of the $\Delta S = 1$ operators described above. Both m_c and $h_{\pm}$ satisfy their own RGEs, so the whole system of coupled ODEs needs to be solved together.

Running these couplings down from $\mu = m_W$ to the scale of the charm quark or so, the $\Delta S = 1$ operators $O_{\pm}^{\bar{c}u}$ can be used to describe decays of D-mesons (mesons with one charm quark) into kaons, and, with rather more work, the decays of kaons into pions. (The analysis becomes complicated below the scale of the charm quark mass.)



The $\Delta S = 2$ operators can be used to describe $K^0 - \bar{K}^0$ mixing.[6] More precisely, the RG analyses provide good computations of the operator *coefficients*; additional input (e.g. lattice) is needed to determine operator *matrix elements* with the meson states, in order to connect to observables.

1.9 Wilsonian RG

At this stage it is useful to introduce a complementary perspective on the flow of couplings, "Wilsonian" or "exact" RG. We start from the generating functional of connected Euclidean correlation functions, which has the path integral expression:

$$Z_\Lambda[J] = \int \mathcal{D}_\Lambda \phi e^{-S_\Lambda[\phi] - \int J\phi} \tag{1.105}$$

[6] K^0 and $\bar{K}^0$ are flavor eigenstates, produced in QCD processes, and are mass-degenerate at leading order. The $\Delta S = 2$ operators produce an off-diagonal term in their mass matrix, causing the flavor eigenstates to oscillate. In the absence of CP violation, a good first approximation, the mass eigenstates are K_L^0 (CP odd) and K_S^0 (CP even). The splitting is on the order of 10^{-6} eV. The K_L^0 has a longer lifetime than K_S^0 because it can only decay into 3π (CP odd), while K_S^0 can decay to 2π (CP even). However, CP is violated, so sometimes K_L^0 decays to 2π.

from which

$$\frac{1}{Z}\frac{\delta^n Z[J]}{\delta J(x_1)\dots\delta J(x_n)}\bigg|_{J=0} = \langle\phi(x_1)\dots\phi(x_n)\rangle_{\text{connected}} \tag{1.106}$$

$$D_\Lambda\phi = \prod_{k^2<\Lambda^2} d\tilde{\phi}_k, \quad \tilde{\phi}_k = \int d^4x e^{ikx}\phi(x) \tag{1.107}$$

Here the path integral is written with an explicit momentum cutoff Λ.

Wilson: Try to "do the path integral" over $\Lambda'^2 < k^2 < \Lambda^2$ modes. Assuming were only interested in correlators for $k^2 \ll \Lambda^2$, we need $\tilde{J}_k \neq 0$ only for $k^2 \ll \Lambda$, and doing the path integral will lead to a new $Z_{\Lambda'}[J]$ that gives exactly the same predictions for the low-energy correlators:

$$Z_{\Lambda'}[J] = \int \mathcal{D}_{\Lambda'}\phi e^{-S_{\Lambda'}[\phi]-\int J\phi} \tag{1.108}$$

where

$$e^{-S_{\Lambda'}[\phi]} \equiv \int \prod_{\Lambda'^2<k^2<\Lambda^2} d\tilde{\phi}_k e^{-S[\phi]} \tag{1.109}$$

$S_{\Lambda'}$ is the "Wilsonian Effective Action" and the charge in couplings generated this way is the "Wilsonian RG."

In comparison to our previous discussion of RG, this is very similar, except now $\mu = \Lambda$, and we absorb all loop effects into running couplings, not just UV divergences. We expand on the relationship in Sec. 1.9.1 below.

In simple cases, $S_{\Lambda'}$ can be computed in perturbation theory. Here we will just sketch the procedure for ϕ^4 theory. First, we split the field as $\phi = \phi' + \phi''$. ϕ' is the real space field containing only momenta up to Λ', and ϕ'' contains momenta from Λ' to Λ:

$$\text{keep } \to \tilde{\phi}'_k = \begin{cases} \tilde{\phi}_k & 0 \leq k^2 < \Lambda'^2 \\ 0 & k^2 > \Lambda'^2 \end{cases} \tag{1.110}$$

$$\text{“integrate out” } \to \tilde{\phi}''_k = \begin{cases} 0 & 0 \leq k^2 < \Lambda'^2 \\ \tilde{\phi}_k & \Lambda'^2 < k^2 < \Lambda^2 \end{cases} \tag{1.111}$$

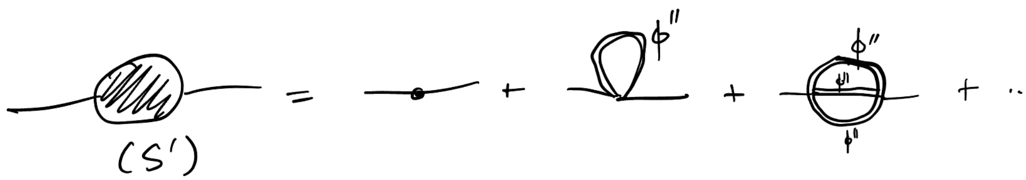
Then the action has the form

$$S_n = \int d^4x \left(\frac{1}{2}(\partial_\mu\phi)^2 + \frac{m^2}{2}\phi^2 + \frac{\lambda}{4!}\phi^4\right) \tag{1.112}$$

$$= \int d^4x \bigg(\frac{1}{2}(\partial_\mu\phi')^2 + \frac{1}{2}(\partial_\mu\phi'')^2 + \frac{m^2}{2}\phi'^2 + \frac{m^2}{2}\phi''^2 + \frac{\lambda}{4!}\phi'^4 + \frac{\lambda}{4!}\phi''^4 \tag{1.113}$$

$$+ \underbrace{\frac{\lambda}{3!}\phi'^3\phi'' + \frac{\lambda}{4}\phi'^2\phi''^2 + \frac{\lambda}{3!}\phi'\phi''^3}_{\text{high}-\text{low interactions}}\bigg) \tag{1.114}$$

How do we integrate out ϕ''? This will be our first example of "matching": $S_{\Lambda'}$ must be able to reproduce the ϕ' correlators computed using S_Λ. Take for example the two-point function:



Since loops only contain momenta $k > \Lambda'$ and external lines only contain $p < \Lambda'$, we can expand the integrals in $\frac{p,m}{\Lambda'}$. Thus this procedure generates a local $S_{\Lambda'}$.

We can make $\Lambda - \Lambda' = d\Lambda$ infinitesimal. Let's consider the potential couplings $\lambda_n \phi^n/n!$. We obtain an infinite set of ODES,

$$\frac{\Lambda d\lambda_4}{d\Lambda} = \beta_4\left(\lambda_4, \Lambda^2\lambda_6, \dots\right) \tag{1.115}$$

$$\frac{\Lambda d\lambda_6}{d\Lambda} = \Lambda^{-2}\beta_6\left(\lambda_4, \Lambda^2\lambda_6, \dots\right) \tag{1.116}$$

and so on. Define dimensionless couplings $\hat{\lambda}_4 \equiv \lambda_4$, $\hat{\lambda}_6 \equiv \Lambda^2\lambda_6$, in terms of which

$$\Lambda\frac{d\hat{\lambda}_4}{d\Lambda} = \beta_4\left(\hat{\lambda}_4\hat{\lambda}_6, \dots\right) \tag{1.117}$$

$$\Lambda\frac{d\hat{\lambda}_6}{d\Lambda} = 2\hat{\lambda}_6 + \beta_6\left(\hat{\lambda}_4, \hat{\lambda}_6, \dots\right) \tag{1.118}$$

By sketching diagrams it is easy to see that

$$\beta_4 = \frac{1}{16\pi^2}\left(a_1\hat{\lambda}_4^2 + a_2\hat{\lambda}_6\right) + \text{(higher loop)} \tag{1.119}$$

$$\beta_6 = \frac{1}{16\pi^2}\left(a_3\hat{\lambda}_4^3 + a_4\hat{\lambda}_4\hat{\lambda}_6 + a_5\hat{\lambda}_8\right) + \text{(higher loop)} \tag{1.120}$$

for some numerical coefficients a_i. For weak coupling and $\Lambda' \ll \Lambda$, the "classical scaling term" $+2\hat{\lambda}_6$ term dominates the $\hat{\lambda}_6$ RGE.

$$\Rightarrow \hat{\lambda}_6(\Lambda') \simeq \left(\frac{\Lambda'}{\Lambda}\right)^2 \hat{\lambda}_6(\Lambda) \tag{1.121}$$

equivalent to our scaling arguments with $[\lambda_6] = -2$.

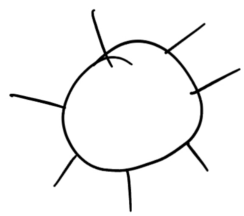
Before $\hat{\lambda}_6$ becomes negligible, however, it sources $\hat{\lambda}_4$. We also saw this in our earlier treatment of RG: UV parts of integrals involving HDOs could be absorbed by redefinitions of the local couplings.

This has only been a sketch, because the Wilsonian RG tends to be better as a conceptual picture of RG flows than a practical calculational method. To relate it to what we did previously, let us discuss:

1.9.1 A comparison of Wilsonian RG and the standard procedure

In Wilsonian RG, we hold IR physics fixed while we change the cutoff and bare parameters. In the standard procedure, we hold the bare parameters fixed while we change the renormalized parameters. These are essentially equivalent. Once we remove the regulator (take the cutoff to infinity or set ϵ to zero) in the standard computation with counterterms, the renormalization scale μ behaves similarly to the Wilsonian cutoff: the counterterms cancel/absorb the UV parts of loops down to scales of order μ. The main difference is how finite corrections are handled.

In our previous treatment, after choosing $\mu \approx E$, we still have UV - finite (μ-independent) loop corrections, for example



These are already absorbed into an effective φ^7 correction in Wilsonian RG if we take $\Lambda' \sim E$. Think of radiative corrections in an ordinary scheme as "finishing the Wilsonian RG." In this sense they are just two different schemes.

An additional difference will arise when we consider mass-independent renormalization schemes like minimal subtraction.

1.10 Decoupling

In the next two sections we discuss some important features of RG flows.

The Appelquist-Carazzone decoupling theorem states that **the contributions of heavy degrees of freedom to low-momentum amplitudes are suppressed ("decouple"), apart from contributions they make to the renormalized masses and couplings of the light degrees of freedom.**

The theorem is only manifest in the solutions to RGEs in mass-dependent renormalization schemes. In mass-independent schemes, it fails, and one consequence of this is that large logs can still appear when there are large hierarchies between "infrared" scales, e.g. $\log(m/p)$-type terms. On the other hand, we have seen that mass-independent schemes are very useful for dimensional analysis, since they restrict the dependence of correlation functions on the renormalization scale, and for organizing the perturbation series in a way that maintains a wider range of validity in the UV than mass-dependent schemes. So it would be advantageous to formulate a procedure to restore decoupling to computations carried out in mass-independent schemes. We will first illustrate how decoupling works in QED with a mass-dependent scheme. Later we will discuss Weinberg's "match and run" method to enforce decoupling by hand in a mass-independent scheme.

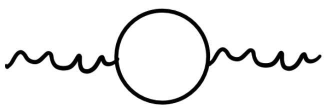
Consider QED in four dimensions:

$$\mathcal{L}_{4-2\epsilon} = -\frac{1}{4} z_\gamma F_{\mu\nu}^2 + i z_\psi \bar{\psi} \partial\!\!\!/ \psi - z_m m_\psi \bar{\psi}\psi - e z_e \mu^\epsilon \bar{\psi} A\!\!\!/ \psi$$
$$[e] = 0 \quad [\psi] = \frac{3}{2} - \epsilon \quad [A] = 1 - \epsilon \tag{1.122}$$

Here, as usual,

$$\begin{aligned} z_\gamma &= 1 + \delta_\gamma \\ z_\psi &= 1 + \delta_\psi \\ z_m &= 1 + \delta_m/m \\ z_e &= 1 + \delta_e/e. \end{aligned} \tag{1.123}$$

The one loop vacuum polarization has the structure



$$\equiv i\Pi_2^{\mu\nu}(q) = i(q^2 g^{\mu\nu} - q^\mu q^\nu)\Pi_2(q) \tag{1.124}$$

and in dimensional regularization one finds

$$\Pi_2(q) = -\frac{2e^2}{4\pi^2} \int_0^1 dx\, x(1-x) \left(\frac{1}{\epsilon} - \log \Delta/\tilde{\mu}^2 + \dots \right)$$
$$\Delta \equiv m^2 - x(1-x)q^2. \tag{1.125}$$

An example of a mass-dependent renormalization scheme is a momentum subtraction scheme, defined by

$$\delta_\gamma = \Pi_2(q_0) \tag{1.126}$$

for some spacelike q_0. (There is a -1 in the counterterm Feynman rule relative to the definition of the vacuum polarization, so this has the effect of replacing $\Pi_2(q) \to \Pi_2(q) - \Pi_2(q_0)$.) Similar subtractions can be used for the electron self-energy and electron-photon vertex, but for later convenience let us take an on-shell scheme for the electron mass renormalization. Then this parameter is the pole mass and is independent of q_0 in this scheme.

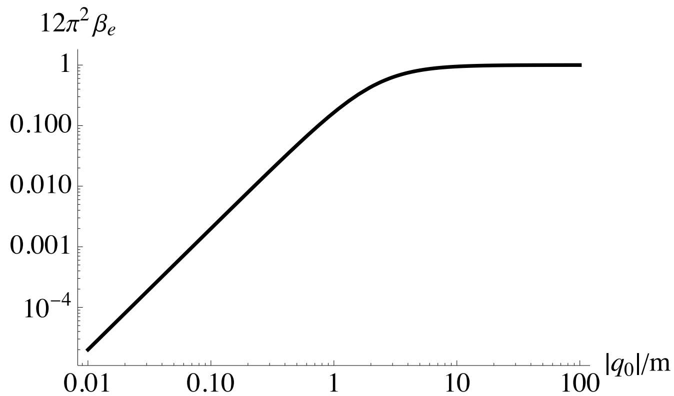
Now we can compute the one-loop β function for the fine structure constant by looking at, say, the one-loop contribution to 2-to-2 scattering. Only the vacuum polarization contributes. Requiring $e^2(q_0)(1 + \Pi_2(q) - \Pi_2(q_0))$ to be independent of q_0, we find

$$\beta_e = -\frac{e^3}{2\pi^2} \int_0^1 dx\, \frac{x^2(1-x)^2 q_0^2}{m^2 - (1-x)x q_0^2}. \tag{1.127}$$

Let us examine the beta function in two limits, $|q_0| \gg m$ and $|q_0| \ll m$. We find:

$$\beta_e \to \frac{e^3}{12\pi^2} \qquad (-q_0^2 \gg m^2)$$
$$\beta_e \to -\frac{e^3}{60\pi^2}\frac{q_0^2}{m^2} \qquad (-q_0^2 \ll m^2) \tag{1.128}$$

This behavior is an example of decoupling. At high RG scales (analogous to high transverse momentum exchange) the fluctuations of the electron field contribute to the running of the electromagnetic coupling. As we lower the RG scale, eventually we reach the electron mass, and it "decouples," leaving only the beta function of pure U(1) gauge theory (which vanishes, since the theory is free.) The full behavior is:



Unfortunately, in mass-independent schemes, decoupling no longer holds automatically. The QED counterterms in $\overline{MS}$ are

$$\delta_3 = \frac{e^2}{16\pi^2}\left[-\frac{4}{3\epsilon}\right]$$
$$\delta_2 = \frac{e^2}{16\pi^2}\left[-\frac{1}{\epsilon}\right]$$
$$\delta_1 = \frac{e^2}{16\pi^2}\left[-\frac{1}{\epsilon}\right]. \tag{1.129}$$

and so

$$\frac{d}{d\log\mu}\left[\underbrace{z_3^{-1/2} z_2^{-1} z_1}_{1+\frac{e^2}{16\pi^2}\frac{2}{3\epsilon}} e\mu^{\epsilon}\right] = 0$$
$$\Rightarrow \frac{de}{d\log\mu} = \beta_e = \frac{\left[-\frac{2e^3}{3\cdot 16\pi^2} - e\epsilon\right]}{1 + \frac{2e^2}{16\pi^2}\frac{1}{\epsilon}}$$
$$= \frac{e^3}{12\pi^2} - e\epsilon + \mathcal{O}(e^4). \tag{1.130}$$

The beta function is independent of the electron mass! This is ok if the electron is massless. But if it is massive, it should decouple. Fortunately, there is a simple fix: we will just remove it by hand. This procedure will be explained below.

The general solution to the RGE is

$$e(\mu) = \frac{e^2(\mu_i)}{1 - 2be_i^2 \log \mu/\mu_i} \qquad b > 0. \tag{1.131}$$

So the coupling is infrared free in the massless theory, and there is a Landau pole in the UV.

1.11 Universality, RG flows, and the continuum limit

Solutions to the RGEs may be thought of as flows between fixed points. A fixed point is a value of the couplings where the beta functions all vanish. Fixed points may be classified as Gaussian (free), trivial, and interacting. Here trivial means empty: the far infrared of a gapped theory is trivial, because all degrees of freedom have decoupled. The most important fixed point for high energy physics in 4D seems to be the Gaussian fixed point. If $\beta > 0$, the coupling is weaker a low energies, and the IR fixed point is free. If $\beta < 0$, on the other hand, the coupling weakens at high energies, flowing to the free UV fixed point. φ^3 in 6D has this property, as does QCD in 4D.

In general, if we run the couplings of a theory up to arbitrarily high scales, they may reach a singularity, or they may approach constants. The singularity case is called a Landau pole; more precisely the theory becomes strongly coupled, and we do not know what happens. The constant case means that there is a UV fixed point. Such theories are said to possess a **continuum limit.**[7]

In the standard perturbative RG, theories with only relevant, marginally relevant, and exactly marginal couplings flow to UV fixed points. In all the examples I am aware of in 4D the UV fixed point is Gaussian (asymptotic freedom). In a Wilsonian effective action, a theory can still possess a continuum limit if irrelevant operators are present, but they must appear in precise proportions: they must correspond precisely to what is generated by integrating out Fourier modes, up to infinite momentum, in a theory defined as a relevant perturbation of a UV fixed point. For example, φ^3 theory is asymptotically free, and thus possesses a continuum limit. The Wilsonian action with finite Λ does contain φ^4 and higher couplings, because these operators are sourced by φ^3 in the RGEs. Let $\hat{g}$ and $\hat{\lambda}$ be the φ^3 and φ^4 couplings in units of the cutoff. The RGEs have the qualitative form

$$\begin{aligned} \Lambda\partial_\Lambda \hat{g} &= -b_g \hat{g}^3 \\ \Lambda\partial_\Lambda \hat{\lambda} &= 2\hat{\lambda} - b_\lambda \hat{g}^4 \end{aligned} \tag{1.132}$$

[7]In the Landau pole case the theory is sometimes said to be "trivial," meaning that it only possesses a continuum limit if the couplings vanish (and then they vanish at all scales.) This terminology is suboptimal for two reasons. First, "trivial" is also used to describe empty theories containing no degrees of freedom. Second, a theory doesn't have to possess a continuum limit in order to be a useful, nontrivial effective theory. We will see many examples later, but QED and φ^4 theory are two examples where the couplings are marginally irrelevant, and so they run to strong coupling in the UV.

where we retain only those couplings necessary to illustrate the point, and $b_g > 0, b_\lambda > 0$. If $\hat{g}(\Lambda_0) = \hat{g}_0$ and $\hat{\lambda}(\Lambda_0) = \hat{\lambda}_0$, then in the far UV $\Lambda \gg \Lambda_0$,

$$\hat{\lambda} \sim \left(\frac{\Lambda}{\Lambda_0}\right)^2 \left(\hat{\lambda}_0 - \frac{1}{2} b_\lambda \hat{g}_0^4 \left(1 - 2b_g \hat{g}_0^2 \dots\right)\right). \tag{1.133}$$

So if we take an initial condition

$$\hat{\lambda}_0 = \frac{1}{2} b_\lambda \hat{g}_0^4 (1 - 2b_g \hat{g}_0^2 + \dots) \tag{1.134}$$

then $\hat{\lambda}$ will vanish as $\Lambda \to \infty$, and the theory possesses a continuum limit. The fact that φ^3 could source φ^4 in the RGE is another example of operator mixing. As we have noted already, in mass-independent schemes in massless theories, lower-dimensional operators cannot source higher dimensional operators, on dimensional grounds.

Relevant perturbations away from a UV fixed point induce RG flows into the infrared. If the relevant perturbations are weak, the flow is still governed by the UV fixed point for awhile, until they run strong. (This can take an exponentially long RG time for marginally relevant couplings.) Then the couplings are kicked away from the UV fixed point (say, because some fields decouple.) They may be governed by another nearby fixed point for a time, until another relevant perturbation grows strong, and they are kicked again. Eventually, the flow runs toward an IR fixed point which govern the extremely long-distance physics (including the ground state). Because massive degrees of freedom decouple and an infinite tower of irrelevant operators die away under RG flow to the IR, many different field theories flow to the same IR fixed points, and the physics of all these theories is the same at low enough energies. This is the phenomenon of **universality**.

We have mostly focused on weakly coupled theories, where perturbation theory around a Gaussian fixed point can be used to approximate the beta functions and anomalous dimensions. In these theories, to a good approximation, the classical scaling dimensions govern the importance of relevant and irrelevant operators across a large range of scales. The most important quantum effects in perturbative theories concern the classically marginal operators: since they sit on the line between relevant and irrelevant, a small perturbation has a big effect on them after many e-folds of RG evolution. For example, asymptotically free couplings often flow, after a long time, to a strong coupling regime. Alternatively, another relevant perturbation – like a mass – may change the trajectory, or in some special cases the coupling may flow to a weakly coupled IR fixed point.

In 4D physics, what often happens when a coupling runs strong in the infrared is that the degrees of freedom undergo a dramatic reordering, a mass gap develops, and the actual IR fixed point is either trivial (empty) or free. This is what happens in QCD and its relatives. In pure Yang-Mills, it is believed that a mass gap develops, the lightest degrees of freedom are "glueballs," (heuristically, bound states of many gluons) and the far IR is empty. In QCD with vanishing quark masses, it seems that the lightest degrees of freedom would be massless, IR-free Goldstone bosons associated with spontaneous chiral symmetry breaking. Remarkably, the mass scale of the glueballs in YM and the scale of the chiral condensate in QCD can still be estimated using the perturbative beta function. In both cases the 1-loop RGE has the form

$$\frac{dg}{d\log\mu} = -\frac{bg^3}{16\pi^2} \tag{1.135}$$

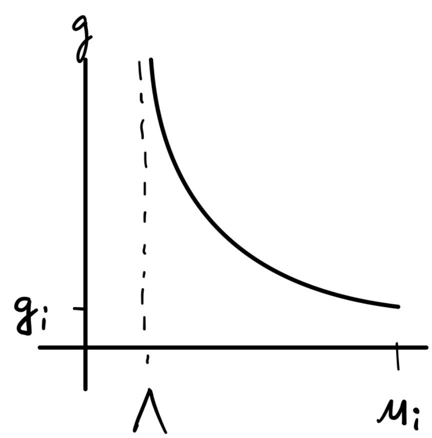
with $b > 0$. Then if the coupling is perturbative at some UV scale μ_i, it runs strong at some exponentially (but not infinitely) lower scale

$$\Lambda \equiv \mu_i e^{-\frac{8\pi^2}{2bg(\mu_i)^2}}. \tag{1.136}$$

(Note: Λ is not the hard momentum cutoff discussed previously. It is an unfortunate overloading of the symbol.) The important thing about Λ is that it is physical, i.e. RG invariant:

$$\frac{d\Lambda}{d\log\mu_i} = 0. \tag{1.137}$$

The existence of a new physical scale, the "dynamical scale," is usually a good prediction of the RG. Its precise value is not, since g naively diverges as $\mu \to \Lambda$, so perturbation theory is breaking down.



Nonetheless at the order-of-magnitude level it provides a reasonable estimate for the physical IR scales in many asymptotically free theories.

It is also an interesting possibility that the infrared of a theory is described by a strongly interacting fixed point. Empirically, this seems to be much more important for statistical mechanics in lower dimensions, and less important for particle physics in four dimensions. We will return to the discussion of "conformal windows" when we study supersymmetric gauge theory later.

Near a fixed point,

$$\gamma \to \gamma(g_i^*) \simeq \text{constant} \tag{1.138}$$

$$\Rightarrow e^{\frac{n}{2}\int_{\mu_0}^{\mu_1}\gamma\, d\log\mu} = \left(\frac{\mu_1}{\mu_0}\right)^{\gamma(g_i^*)n/2} = (x^{\gamma *})^{n/2} \tag{1.139}$$

(1.140)

This is why we considered the special case "γ is approximately constant" earlier, and the scaling is the origin of the name "anomalous dimension." In simple models like gauge theory or φ^4 or φ^3, we can rewrite the anomalous dimension factor in a useful way. Using the beta function, the integral in the exponent can be rewritten:

$$e^{\frac{n}{2}\int \gamma(g)\frac{dg}{\beta(g)}} \tag{1.141}$$

For example, in φ^3 theory, we have

$$e^{\frac{n}{2}\int(-\frac{1}{9g})dg} = \left(\frac{g(\mu_1)}{g(\mu_0)}\right)^{-\frac{n}{2\cdot 9}}. \tag{1.142}$$

We saw this sort of thing already in the discussion of four-Fermi operators.

1.11.1 Lattice gauge theory

In lattice gauge theory, a continuum gauge theory is approximated by discretizing spacetime. The lattice spacing a acts as a UV cutoff, and on shorter distance scales, the lattice theory is completely different from the continuum theory. Under what circumstances should the physics of the lattice theory on some length scale x well-approximate the physics of the continuum theory?

Generally speaking the lattice theory can be thought of as differing from the continuum theory at the renormalization scale $\mu \simeq 1/a$ by an infinite tower of higher dimension operators, with couplings set by powers of a. We know that as we flow to lower scales these operators will die away. So we certainly want $x/a \gg 1$, so that the operators die away enough. This is just classical scaling. We also want weak coupling, so that the irrelevant operators don't have much impact before they are suppressed. The lattice theory won't exactly lie on the renormalization trajectory to have a continuum limit, because that would require tuning an infinite number of operators, as in Eq. (1.134) but for all operators. But that doesn't matter if the theory starts near enough to the UV gaussian fixed point. It will still flow toward the surface of marginal and relevant perturbations before leaving the fixed point, and it will be in the same universality class as the desired continuum theory. Usually, to improve this convergence, "improved actions" are used, which cancel the first few irrelevant operators of lowest dimension.

1.12 Appendix: Doing Feynman Integrals in Dim Reg

First one introduces Feynman parameters and shifts variables get to a notationally symmetric integral, eg

$$\begin{aligned}\mathcal{I} &= \int_0^1 dx \int \frac{d^d k_E}{(2\pi)^d (k_E^2 + \Delta_x)^\alpha} \\ &= \int_0^1 dx \left(\int_0^\infty \frac{k_E^{d-1} dk_E}{(k_E^2+\Delta_x)^\alpha}\right) \frac{2\pi^{d/2}}{\Gamma(d/2)(2\pi)^d}\end{aligned} \tag{1.143}$$

using $\int d\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$. If $d < 2\alpha$, the integral in parenthesis is finite and gives

$$\left(\frac{1}{\Delta_x}\right)^{\alpha-\frac{d}{2}} \frac{\Gamma\left(\frac{d}{2}\right)\Gamma\left(\alpha-\frac{d}{2}\right)}{2\Gamma(\alpha)}. \tag{1.144}$$

Now let us specialize to four dimensions. We set $d = 4 - 2\epsilon$, and multiply by $\mu^{2\epsilon}$. We have

$$\mathcal{I}_{DR} = \int_0^1 dx \left[\mu^{2\epsilon}\left(\frac{1}{\Delta_x}\right)^{\alpha-2+\epsilon}\right] \frac{\Gamma(\alpha-2+\epsilon)}{(2\sqrt{\pi})^{4-2\epsilon}\Gamma(\alpha)} \tag{1.145}$$

This is valid for $\epsilon > \alpha - 2$, but we will continue back to $\epsilon = 0$ after subtractions.

Now let us look at some values of α.

- $\alpha = 2$. Our original momentum integral behaves in the UV like $\int d^4/k^4$, which is logarithmically divergent. Expanding the x integrand around $\epsilon = 0$, we have

$$\begin{aligned}
&\left(\frac{\mu^2}{\Delta}\right)^{\epsilon} \frac{\Gamma(\epsilon)}{(2\sqrt{\pi})^4\Gamma(2)} \\
&= \left[1 + \epsilon \log(4\pi\mu^2/\Delta) + \mathcal{O}(\epsilon^2)\right]\left[1/\epsilon - \gamma + \mathcal{O}(\epsilon)\right](2\sqrt{\pi})^4 \\
&= \left(\frac{1}{\epsilon} + \log\frac{4\pi\mu^2}{\Delta} - \gamma + o(\epsilon)\right)\frac{1}{16\pi^2} \\
&\to \frac{1}{16\pi^2}\left(\frac{1}{\epsilon} + \log\frac{\tilde{\mu}^2}{\Delta}\right), \quad \tilde{\mu}^2 \cong 4\pi e^{-\gamma}\mu^2
\end{aligned} \tag{1.146}$$

  (Each loop in 4D comes with a factor of $1/16\pi^2$.) We see that the $\log\Lambda$ UV divergence reappears in dim reg as a $1/\epsilon$ pole.

- $\alpha = 1$. Our original momentum integral behaves in the UV like $\int d^4/k^2$, which is quadratically divergent. Expanding the x integrand around $\epsilon = 0$, we have

$$\begin{aligned}
&\Delta\left(\frac{\mu^2 4\pi}{\Delta}\right)^{\epsilon} \frac{\Gamma(-1+\epsilon)}{16\pi^2\Gamma(1)} \\
&\Rightarrow \frac{\Delta}{16\pi^2}\left(1 + \epsilon\log\frac{4\pi\mu^2}{\Delta}\right)\left(\frac{\Gamma(\epsilon)}{\epsilon - 1} \simeq \left(\frac{1}{\epsilon} - \gamma\right)(-1-\epsilon)\right) \\
&\Rightarrow \frac{\Delta}{16\pi^2}\left[-\frac{1}{\epsilon} - \left(\log\frac{\tilde{\mu}^2}{\Delta} + 1\right)\right]
\end{aligned} \tag{1.147}$$

  The Λ^2 UV divergence has reappeared as a $-1/\epsilon$ pole. The only difference between the power law divergence and the logarithmic divergence in dim reg is the sign in front of $1/\epsilon$.

- Any α, and $\Delta = 0$. These are called scaleless integrals and can be set to zero self-consistently in dim reg. Eg $\int \frac{d^4k}{k^2} = \int \frac{d^4k}{k^4} = 0$. (For discussion, see Leibbrant review. It is easy to believe for $\int \frac{d^4k}{k^2}$ on dimensional grounds, but less obvious for $\int \frac{d^4k}{k^4}$, and easy but wrong "proofs" exist.)

1.13 Appendix: Subtleties with Divergent Integrals

In this appendix we discuss three examples of how divergent or ill-defined momentum integrals can have problems with changes of integration variables (commuting with limits). For example, when there is a shift in momentum in a divergent integral, we have to regulate the integral, and the shift may affect the regulator. (E.g. a momentum cutoff would shift.)

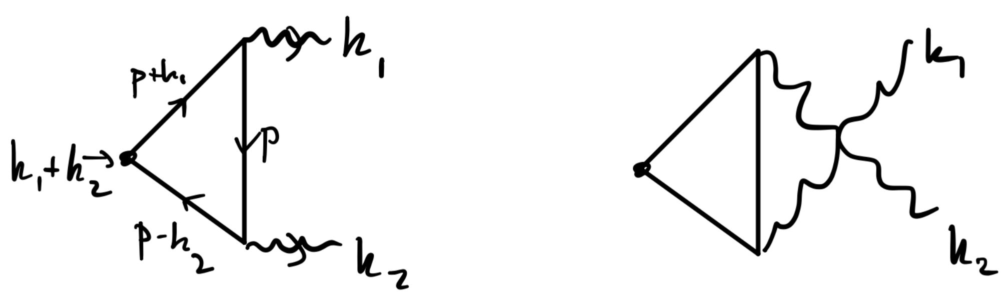
The simplest example is the vacuum polarization in QED. The Feynman integral is divergent and must be regulated before doing the momentum shift to put the integral in standard form. it can be cured by dimensional regularization, Pauli Villars (although many fermions are needed), or a cutoff with a counterterm to restore gauge invariance.

The next simplest example is in $h \to gg$, the Higgs decay to gauge bosons. The cross diagram $(p_1, \mu \leftrightarrow p_2, \nu)$ ends up contributing the same as the uncrossed diagram. There is a subtlety in that although the result for each diagram is finite after we do the first momentum shift, before we do the shift they are divergent. A regulator is needed to get an unambiguous answer. In the vacuum polarization case, the regulator introduces new scheme dependence. In the $h \to gg$ case, there is no counterterm (the operator is dimension 5) so it had better be regulator independent in the end, but we need some regulator to make the momentum shift in the first place. For example, in $d = 4 - 2\epsilon$ dimensions, the fermion trace in the loop has an $O(\epsilon)$ piece (from $\ell^\mu \ell^\nu \to \ell^2/dg^{\mu\nu}$) that sits in front of a UV divergent piece $\sim 1/\epsilon$. So ϵ/ϵ ~finite. By just working in $d = 4$ we would miss this finite bit (equivalent to taking $\epsilon \to 0$ before doing the integral.)

The last example is the chiral anomaly calculation. We compute

$$\Gamma^{AAJ} \equiv T \left\langle \partial_\mu J^{\mu 5} A_\rho A_\sigma \right\rangle \tag{1.148}$$

Feynman diagrams:



The corresponding Feynman integrals are

$$\Gamma^{AAJ} = \frac{g^2}{(2\pi)^4} \int d^4p \, \mathrm{Tr} \left(\not{q} \gamma_5 \frac{1}{\not{p} + \not{k}_1} \not{\epsilon}_1 \frac{1}{\not{p}} \not{\epsilon}_2 \frac{1}{\not{p} - \not{k}_2} + (1 \leftrightarrow 2) \right) \tag{1.149}$$

The integral is linearly divergent. The following procedure is tempting but incorrect. First, write

$$\not{q} \gamma_5 = -\gamma_5 \left(\not{k}_1 + \not{p} \right) - \left(\not{p} - \not{k}_2 \right) \gamma_5 \tag{1.150}$$

and use it to write the traces as

$$\mathrm{Tr}\big(-\gamma_5 \not{\epsilon}_1 \frac{1}{\not{p}} \not{\epsilon}_2 \frac{1}{\not{p} - \not{k}_2} - \gamma_5 \frac{1}{\not{p} + \not{k}_1} \not{\epsilon}_1 \frac{1}{\not{p}} \not{\epsilon}_2 - \gamma_5 \not{\epsilon}_2 \frac{1}{\not{p}} \not{\epsilon}_1 \frac{1}{\not{p} - \not{k}_1} - \gamma_5 \frac{1}{\not{p} + \not{k}_2} \not{\epsilon}_2 \frac{1}{\not{p}} \not{\epsilon}_1 \big). \tag{1.151}$$

Now shift $p \to p + k_2$ in term 1 and $p \to p - k_1$ in term 2, and use the anticommutativity of γ_5 and the cyclicity of the trace:

$$\mathrm{Tr}\Big(- \gamma_5 \not{\epsilon}_1 \frac{1}{\not{p} + \not{k}_2} \not{\epsilon}_2 \frac{1}{\not{p}} - \gamma_5 \frac{1}{\not{p}} \not{\epsilon}_1 \frac{1}{\not{p} - \not{k}_1} \not{\epsilon}_2 + \not{\epsilon}_2 \gamma_5 \frac{1}{\not{p}} \not{\epsilon}_1 \frac{1}{\not{p} - \not{k}_1} + \gamma_5 \not{\epsilon}_1 \frac{1}{\not{p} + \not{k}_2} \not{\epsilon}_2 \frac{1}{\not{p}} \Big). \tag{1.152}$$

Pairwise cancellation! By such formal manipulations, we might conclude that the graphs vanish.

But we committed a mathematical sin. The integrals are meaningless. They have to be regulated in order for the symbols to mean anything. It turns out that once we regulate the integrals, the cancellation is imperfect, and we get a finite contribution from the UV. For example, with Pauli-Villars, we would subtract the same expression but with heavy fermions of mass Λ in the loop:

$$-\frac{g^2}{(2\pi)^4} \int d^4 p \,\mathrm{Tr} \left(\not{q} \gamma_5 \frac{1}{\not{p} + \not{k}_1 - \Lambda} \not{\epsilon}_1 \frac{1}{\not{p} - \Lambda} \not{\epsilon}_2 \frac{1}{\not{p} - \not{k}_2 - \Lambda} + (1 \leftrightarrow 2) \right) \tag{1.153}$$

Adding these terms to the previous cuts off the linear divergence around Λ. The PV fields enable the momentum shifts above, but give new contributions as well. In the PV terms we can do a similar rearrangement:

$$\not{q}\gamma_5 = \underbrace{-\gamma_5 \left(\not{k}_1 + \not{p} - \Lambda \right) - \left(\not{p} - \not{k}_2 - \Lambda \right)}_{\text{these will drop out}} \gamma_5 - 2\Lambda \gamma_5 \tag{1.154}$$

so after doing the now-valid shifts for both the regular and PV terms, what survives is:

$$\frac{g^2}{(2\pi)^4} \int d^4 p \,\mathrm{Tr} \left(-2\Lambda \gamma_5 \frac{1}{\not{p} + \not{k}_1 - \Lambda} \not{\epsilon}_1 \frac{1}{\not{p} - \Lambda} \not{\epsilon}_2 \frac{1}{\not{p} - \not{k}_2 - \Lambda} + (1 \leftrightarrow 2) \right). \tag{1.155}$$

This we can tackle with Feynman parameters, etc. The leading term for large Λ is of the form

$$\frac{g^2 \Lambda^2}{(2\pi)^4} \epsilon_{\mu\nu\rho\sigma} k_1^\mu k_2^\nu \epsilon_1^\rho \epsilon_2^\sigma \int d^4 p \frac{1}{(p^2 - \Lambda^2)^3} \tag{1.156}$$

where the Levi-Civita symbol came from $\mathrm{Tr}(\gamma_5 \gamma_\mu \gamma_\nu \gamma_\rho \gamma_\sigma)$. In the large Λ limit the integral gives $\pi^2 / 2\Lambda^2$ that cancels the Λ^2 in front, leaving a finite, regulator-independent contribution. Inverting LSZ,

$$\epsilon_{\mu\nu\rho\sigma} k_1^\mu k_2^\nu \epsilon_1^\rho \epsilon_2^\sigma \sim F \tilde{F}. \tag{1.157}$$

As an operator, one finds

$$\partial_\mu J^{\mu 5} = \frac{g^2}{8\pi^2} F^{\mu\nu} \tilde{F}_{\mu\nu} \tag{1.158}$$

in the case of abelian gauge fields, or

$$\partial_\mu J^{\mu 5} = \frac{g^2}{16\pi^2} F^{a\mu\nu} \tilde{F}^a_{\mu\nu} \tag{1.159}$$

in the nonabelian case.

Anomalies arise due to an incompatibility between the symmetry and the UV regulator, as this example illustrates. We used Pauli-Villars, but one can also use dim reg. γ^5 is troublesome to define in d dimensions. ‘t Hooft and Veltman suggested $\gamma^0 \cdots \gamma^3$. Peskin & Schroeder calculate the anomaly this way.

Now an important subtlety:

$$F\tilde{F} = \partial^\mu K_\mu, \quad K_\mu = \epsilon_{\mu p\sigma} \left(A^a_\nu F^a_{\rho\sigma} - \frac{2}{3} f^{abc} A^a_\nu A^b_\rho A^c_\sigma \right). \tag{1.160}$$

So why not just redefine the current,

$$J^{\mu 5} \to J^{\mu 5} - K^\mu? \tag{1.161}$$

K_μ is not gauge invariant, so we cannot make this shift and obtain an observable current. Even it the abelian theory:

$$\begin{aligned} K^\mu &= \epsilon_{\mu\nu\rho\sigma} A_\nu \partial_\rho A_\sigma \\ \Delta K &= \epsilon_{\mu\nu\rho\sigma} (\underbrace{\partial_\nu \lambda \partial_\rho A_\sigma}_{\neq 0} + A_\nu \partial_\rho \partial_\sigma \lambda) \end{aligned} \tag{1.162}$$

But it is a total derivative: can it affect the path integral?

$$\delta \mathcal{L} = \alpha \partial_\mu J^{\mu 5} \propto \alpha F\tilde{F} \tag{1.163}$$

In the abelian case, finite Euclidean action $\int d^4x_E FF$ (sum of squares) implies $F_{\mu\nu} < 1/r^2$, and $F = dA$ implies $A < 1/r$, so $F\tilde{F} < 1/r^4$ and the fields go to zero fast enough on the boundary so that the integral of the Chern-Simons term vanishes. But in the nonabelian case, $F < 1/r^2$ does not imply $A < 1/r$. There can be cancellations between $\partial_\mu A^a_\nu$ and $f^{abc} A^b_\mu A^c_\nu$. We will see later that there are configurations, even semiclassically, for which $F \sim 1/r^4, A \to 1/r$. These give a finite boundary contribution $\int F\tilde{F} =$ finite. So the anomaly is physical.

1.14 Appendix: “Effective Actions”

Confusingly, the term “effective action” is used for closely related objects which are often described in such different language that their relationship is completely obscured.

In a QFT course, one usually first meets the “1PI effective action,” introduced by computing the generating functional (of connected correlators) from the path integral and then Legendre transforming it. This effective action is itself a generating functional of 1PI correlators. The Legendre transformation definition is equivalent to another definition, the “background field method,” which may be more intuitive. First, in the action, shift the fields by fixed classical backgrounds. Then, adjust sources so that the fields you are path integrating over have vanishing tadpoles. Finally, do the path integral. The resulting functional of the classical background fields is the effective action.

In slightly more detail, recall that the generating functional is

$$Z[J] = \int D\phi e^{i\int(\mathcal{L}[\phi]+J\phi)}. \tag{1.164}$$

Taking functional derivatives of $Z[J]$ wrt J generates correlators of the fields ϕ. Here the notation is general: ϕ could be scalars, fermions, gauge fields, and J denotes linear sources for all of them. We also define the object

$$W[J] = -i \log Z[J]. \tag{1.165}$$

Taking functional derivatives of $W[J]$ wrt J generates connected correlators of the fields ϕ. Now define the slightly modified objects

$$\begin{aligned} Z[J[\Phi],\Phi] &= \int D\phi e^{i\int(\mathcal{L}[\phi]+J[\Phi](\phi-\Phi))} = Z[J[\Phi]]e^{-i\int J[\Phi]\Phi} \\ \Gamma[\Phi] &= -i\log Z[J[\Phi],\Phi] \end{aligned} \tag{1.166}$$

where the source is adjusted so that ${}_J\langle 0|\phi|0\rangle_J = \Phi$. $\Gamma[\Phi]$ is the effective action. Intuitively, it is "the classical action plus quantum corrections" in the presence of sources that generate specified background fields Φ, without including an action contribution from the source term (because $J[\Phi](\phi - \Phi) \to J[\Phi](\langle\phi\rangle - \Phi) = 0$).

For low momenta this effective action is equivalent to a Wilsonian effective action with a low cutoff scale whenever the spectrum is gapped. If there are massless particles, the 1PI effective action typically has IR singularities (nonlocal terms) which are never present in a Wilsonian action, by construction.

What is $\Gamma[\Phi]$ good for, and how do we compute it? We'll look at these in reverse. There are two main methods for computing the effective action in perturbation theory. Both use the fact (true, but we have not proved it) that $\Gamma[\Phi]$ is the generating functional of 1PI graphs.

- The first method is valid to all orders in the background field. One defines the shifted field $\phi' = \phi - \Phi$, so that

  $$Z[J[\Phi],\Phi] = \int D\phi' e^{i\int(\mathcal{L}[\phi'+\Phi]+J[\Phi](\phi'))} = Z[J[\Phi]]e^{-i\int J[\Phi]\Phi} \tag{1.167}$$

  Derive the exact Feynman rules for ϕ', in perturbation theory around $\phi' = 0$, from the Lagrangian $\mathcal{L}[\phi' + \Phi]$, treating Φ as a contribution to the propagator and various couplings and dropping any tadpoles. Then, since there are no external ϕ' lines, sum up the 1PI *vacuum* graphs through a given loop order. This method gives the effective action to fixed order in $\hbar$ and all order in Φ. Generally it only works for very simple Φ.

- The second method treats the background field perturbatively. Again we derive Feynman rules for ϕ', in perturbation theory around $\phi' = 0$, from the Lagrangian $\mathcal{L}[\phi' + \Phi]$. But this time ϕ' has its ordinary propagator and interactions, plus interactions with Φ', and we treat Φ as separate field which can only appear in external lines. This works for any Φ, but it only generates the effective action as a series and derivative expansion in Φ, $\partial_\mu\Phi$, etc.

At one loop order, each method for computing the effective action is equivalent to expanding the fields in the path integral to quadratic order in small fluctuations and integrating over the fluctuations, which results in various functional determinants.

A common use of the effective action defined this way is to obtain values of the fields in simple states, including quantum corrections. This is because

$$\begin{aligned}\frac{\delta\Gamma}{\delta\Phi} &= {}_J\langle 0|\phi|0\rangle_J \frac{\delta J}{\delta\Phi} - \frac{\delta J}{\delta\Phi}\Phi - J[\Phi] \\ &= -J[\Phi].\end{aligned} \tag{1.168}$$

So, if we know Γ and we solve the "quantum corrected equation of motion" $\frac{\delta\Gamma}{\delta\Phi} = 0$ for fields Φ, the result gives us $\langle\phi\rangle$ in the absence of any sources. Often one is interested in vacuum states with constant (in spacetime) fields. Then one can define a related object called the effective potential,

$$\Gamma[\Phi] = -\int d^4x V_{eff}[\Phi]. \tag{1.169}$$

In other words, $V_{eff}[\Phi]$ is the quantum-corrected energy density in the presence of constant background fields. The minimum of the effective potential gives the ground state of ϕ.

Another use of the term "effective action" arises when some matter fields appear only quadratically in the action, and have couplings to *other* scalar, gauge, or gravitational fields. Since the action is quadratic in the matter fields, the path integral over them may be carried out exactly, without integrating over the other fields. For example, the Minkowski path integrals for a charged Dirac fermion and a charged scalar coupled to a gauge field are

$$Z[A] = \int \mathcal{D}\psi\mathcal{D}\bar{\psi} e^{i\int d^4x\bar{\psi}(i\not{D}-m)\psi} = \det(i\not{D} - m) \tag{1.170}$$

and

$$Z[A] = \int \mathcal{D}\phi\mathcal{D}\phi^* e^{i\int d^4x\phi^*(-D^2-m^2)\phi} = \left[\det(D^2 + m^2)\right]^{-1} . \tag{1.171}$$

Here $D = \partial + ieA$. All coordinates and derivatives are Minkowski, but these results are obtained by continuing $t \to -i\tau$, $\partial^2 \to -\partial_E^2$, $D^2 \to -D_E^2$, etc., doing the Gaussian integrals, and then continuing back. The effective action for the gauge fields is defined as

$$\Gamma[A] = -i \log Z[A] \propto \mathrm{Tr}(\log(G^{-1}[A])) \tag{1.172}$$

where $G^{-1}[A]$ is the inverse of the matter field propagator in the presence of the field A. So this is exactly like method (1) for computing the full 1PI effective action of some background gauge fields, at one loop order, except we haven't included contributions from self-interactions of the gauge fields. The full 1PI effective action, at one loop order, would include extra functional determinants from expanding the gauge field self-interactions to quadratic order in small fluctuations and integrating over the fluctuations. It is common to use the term "effective action" for either object, but this can be confusing, since we did not have to discuss shifted fields, Legendre transforms, sources, etc in (1.170)–(1.172). We just treat A as the background field ab initio. The equivalence to method (2) for computing the full 1PI effective action arises when the functional determinants are computed in a power series in A.

Some further comments...

Since the Dirac operator $i\not{D} - m$ is hermitian, its eigenvalues are real, and the operator $-i\not{D} - m$ has the same eigenvalues. Therefore we can rewrite the determinant as

$$\det(i\not{D} - m) = \sqrt{\det\left[(D^2 + m^2)1_{4\times 4}\right]} \tag{1.173}$$

In Euclidean signature we obtain

$$\int \mathcal{D}\psi\mathcal{D}\bar{\psi}e^{-\int d^4x\bar{\psi}(\not{D}+m)\psi} = \det(\not{D} + m). \tag{1.174}$$

The Euclidean effective action in each case is

$$\Gamma_E = -Z_1 = -\log Z = a\mathrm{Tr}\log(-D_E^2 + m^2) \tag{1.175}$$

with $a = -2$ for the Dirac fermion (1/2 from the log-square-root and 4 from the trace) and $a = +1$ for the complex boson.

An aside: one also sometimes notes that $\not{D}$ is "γ^5 hermitian," i.e. $\gamma^5\not{D}\gamma^5 = \not{D}^\dagger$, so e.g. $\gamma^5(\not{D} + m + \mu\gamma^0)\gamma^5 = (\not{D}^\dagger + m - \mu\gamma^0) = (\not{D} + m^* - \mu^*\gamma^0)^\dagger$. So

$$\det(\not{D} + m + \mu\gamma^0) = \det(\gamma^5(\not{D} + m + \mu\gamma^0)\gamma^5) = \det(\not{D} + m^* - \mu^*\gamma^0)^\dagger \tag{1.176}$$

which shows that both the theta term and the chemical potential lead to complex path integral measures for gauge fields.

These trace-logs can also be written in a worldline formalism. For the complex boson,

$$\begin{aligned} Z_1 &= -\mathrm{Tr}\log(-D^2 + m^2) \\ &= \int_0^\infty \frac{d\beta}{\beta}\, e^{-\beta m/2} \int \mathcal{D}x\, e^{-\mathcal{A}}\,, \end{aligned} \tag{1.177}$$

$$\mathcal{A} = \int_0^\beta d\lambda \left(\frac{1}{2}m(\partial_\lambda x^M \partial_\lambda x_M) - ieA_M\partial_\lambda x^M\right)\,. \tag{1.178}$$

This can be derived by starting from a quadratic formulation with a metric and gauge fixing, as done in the enormous textbook on path integrals by Kleinert. The integral over $d\beta/\beta$ can be thought of as lengths of the worldlines, divided by redundancy in the starting point. For the fermion a similar expression holds, although to really describe a fermion in QM we should also put in the appropriate Grassmann variables and couplings as Kleinert does.

Here is another approach to deriving the worldline representation, for a real scalar in curved Euclidean-signature space, following hep-th/9503016. (This is outside the main thread of these notes and can be skipped.)

$$\begin{aligned}\beta F &= \frac{1}{2}\log\det(-D^2+m^2)\\ &= -\frac{1}{2}\int d^n x\sqrt{g}\int_{\epsilon^2}^{\infty}\frac{ds}{s}e^{-sm^2}K(s,x,x)\end{aligned} \tag{1.179}$$

with the heat kernel trace defined as the

$$\int\int d^n x\sqrt{g}K(s,x,x) = \mathrm{Tr}(\langle x|e^{-s(-D^2)}|x'\rangle) = \int d^n x\sqrt{g}\langle x|e^{-s(-D^2)}|x\rangle. \tag{1.180}$$

The notation here is kind of annoying, but is supposed to mean "think of the operators D^2 or e^{sD^2} as matrices in the position basis." First imagine some basis-independent abstraction of the operator D^2. Call it H. H can act on a vector, $H|\psi\rangle$. We can insert complete sets,

$$\int d^n x\int d^n x'\sqrt{g(x)}\sqrt{g'(x)}|x'\rangle\langle x'|H|x\rangle\langle x|\psi\rangle = \int d^n x\int d^n x'\sqrt{g(x)}\sqrt{g'(x)}|x'\rangle H_{x'x}\psi(x) \tag{1.181}$$

and write the matrix elements $H_{x'x} = \langle x'|H|x\rangle = D_x^2\delta^n(x-x')/\sqrt{g(x)}$.

In the Schwinger proper length integral, ϵ^2 is a UV cutoff. To go from the first line to the second line, write

$$\begin{aligned}\log\det(-D^2+m^2) &= \mathrm{Tr}\log(-D^2+m^2)\\ &= \mathrm{Tr}\int dm^2(-D^2+m^2)^{-1}\\ &= \mathrm{Tr}\int dm^2\int_0^{\infty} dse^{-s(-D^2+m^2)}\\ &= \mathrm{Tr}\int_0^{\infty}\frac{ds}{s}e^{-s(-D^2+m^2)}\\ &= \int_{\epsilon^2}^{\infty}\frac{ds}{s}e^{-sm^2}K(s,x,x).\end{aligned} \tag{1.182}$$

Finally we use that $e^{-s(-D^2)}$ is the quantum mechanical Euclidean propagator for a particle moving in n dimensions with Euclidean time s. This has a path integral representation in the coordinate basis,

$$\langle x|e^{-s(-D^2)}|x'\rangle = \int_{x(0)=x,x(s)=x'} Dxe^{-\int_0^s ds'\frac{1}{4}g_{\mu\nu}\partial_s x^{\mu}\partial_s x^{\nu}} \tag{1.183}$$

For favorable metrics we can evaluate this path integral directly or semiclassically. Birrell and Davies instead mainly consider the adiabatic expansion of K for small s, where one gets a bunch of geometrical coefficients and powers of s that include some finite number of inverse powers.

1.15 Appendix: Yang-Mills theory

For semiclassical purposes, it is often useful to work with a gauge field normalization such that the (Lorentzian) action takes the form:

$$\begin{aligned}
S_{YM} &= -\frac{1}{2g^2}\int d^4x \,\mathrm{Tr}\left(F_{\mu\nu}F^{\mu\nu}\right) + \theta \text{ term} \\
F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu - i\left[A_\mu, A_\nu\right] \\
A_\mu &= A_\mu^a T^a \quad a = 1\ldots \dim(G) \\
T^\dagger &= T \quad \left[T^a, T^b\right] = i f^{abc} T^c.
\end{aligned} \tag{1.184}$$

Canonical normalization, useful for perturbation theory, can be obtained by rescaling $A \to gA$. For T^a in the defining rep, $\mathrm{Tr}\left(T^a T^b\right) = \frac{1}{2}\delta^{ab}$. The coefficient of δ^{ab} is called the Dynkin index of the rep. With this convection for the defining rep, the Dynkin index of other reps is fixed. For the adjoint, it is N.

It is sometimes convenient to work with a vector form of the gauge index on A and F, instead of the tensor form. In that case the formulas are

$$\begin{aligned}
S_{YM} &= -\frac{1}{4g^2}\int d^4x (F^a_{\mu\nu} F^{a\mu\nu}) + \theta \text{ term} \\
F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc} A^b_\mu A^c_\mu.
\end{aligned} \tag{1.185}$$

Gauge transformations act as:

$$A_\mu \to \Omega A_\mu \Omega^{-1} + i\Omega\partial_\mu\Omega^{-1} \quad \Omega : R^4 \to G. \tag{1.186}$$

These take

$$F_{\mu\nu} \to \Omega F_{\mu\nu}\Omega^{-1} \tag{1.187}$$

so S_{YM} is invariant. If we add matter, e.g. $\Delta S = \int \bar{\psi} i \not{D} \psi$, then the matter fields transform as

$$\psi \to D^r(\Omega)\psi \tag{1.188}$$

where D^r is a matrix representation.

The electric Wilson line operators furnish a useful class of operators:

$$\begin{aligned}
U^r\left(x_f, x_i, P\right) &\equiv P e^{i\int_{\tau_i}^{\tau_f} d\tau \frac{dx^\mu}{d\tau} A_\mu(x(\tau))} \\
&= P e^{i\int_{\tau_i}^{\tau_f} dx^\mu A_\mu}
\end{aligned} \tag{1.189}$$

Here $A_\mu = A^a_\mu (T^r)^a$ where T is a generator in any irrep r. P denotes the path ordered exponential, defined by series expansion, with later operators along the path written to the left:

$$
\begin{aligned}
&Pe^{i\int_{\tau_i}^{\tau_f} d\tau \frac{dx^\mu}{d\tau} A_\mu(x(\tau))} \\
&\equiv 1+\sum_{n=1}^{\infty}\frac{1}{n!}P\prod_{a=1}^{n}(\int_{\tau_i}^{\tau_f} d\lambda_a\, i\frac{dx^\mu}{d\tau}(\tau=\lambda_a)A_\mu(\tau=\lambda_a)) \\
&\equiv 1+\sum_{n=1}^{\infty}(i)^n\int_{\tau_i}^{\tau_f} d\lambda_1\int_{\tau_i}^{\lambda_1} d\lambda_2\cdots\int_{\tau_i}^{\lambda_{n-1}} d\lambda_n\,\frac{dx^\mu}{d\tau}(\tau=\lambda_1)A_\mu(\tau=\lambda_1)\ldots\frac{dx^\mu}{d\tau}(\tau=\lambda_n)A_\mu(\tau=\lambda_n)
\end{aligned}
\tag{1.190}
$$

Being the exponential map of group generators, the Wilson line is group valued, and can be in any rep. It can be shown to transform as

$$U \to D^r\left(\Omega\left(x_f\right)\right) U D^r\left(\Omega^\dagger\left(x_i\right)\right) \tag{1.191}$$

This makes it clear that we can make two gauge-invariant operators from it: the Wilson loop,

$$\mathrm{Tr}_r U^r(x,x,P) \tag{1.192}$$

and lines truncated by charges,

$$\bar{\psi}^r(x_f)U^r(x_f,x_i,P)\psi^r(x_i) \tag{1.193}$$

where r denotes the irrep.

1.15.1 Euclidean Yang-Mills

We take the Lorentzian signature metric to be $ds^2=(dx^0)^2-(dx_i)^2$. The Euclidean continuation is performed as follows (some of this is conventional.) Let $x^0\to -i\hat{x}^0$ with real $\hat{x}^0$. Then $ds^2\to (d\hat{x}^0)^2+(d\hat{x}_i)^2$, and the measure $d^4x\to -id^4\hat{x}$.

Gauge field. $A_\mu dx^\mu = A_0dx^0+A_idx^i \to -iA_0d\hat{x}^0+A_idx^i$, so we define the Euclidean continuation of the gauge fields $A_0\to i\hat{A}_0$ with real $\hat{A}_0$, and $A_i\to\hat{A}_i$. Then the components of the Euclidean one-form field $\hat{A}_\mu d\hat{x}^\mu$ are real.

Covariant derivative. $D_0=\partial_0+igA_0\to i\hat{\partial}_0-g\hat{A}_0=i\hat{D}_0$, and $D_i=\partial_i+igA_i\to\hat{\partial}_i+ig\hat{A}_i=\hat{D}_i$. Thus $D^2=D_0^2-D_i^2\to-\hat{D}^2=-(\hat{D}_0^2+\hat{D}_i^2)$.

Field strength tensor. $F_{0i}\to i\hat{\partial}_0\hat{A}_i-i\hat{\partial}_i\hat{A}_0-i[i\hat{A}_0,\hat{A}_i]=i\hat{F}_{0i}$, $F_{ij}\to\hat{F}_{ij}$. So $F_{\mu\nu}F^{\mu\nu}=-F_{0i}^2+F_{ij}^2\to\hat{F}_{0i}^2+\hat{F}_{ij}^2=\hat{F}_{\mu\nu}\hat{F}^{\mu\nu}$.

Dual field strength tensor. The dual is defined as $\tilde{F}_{\mu\nu}=\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$. If $\epsilon_{0123}=1$, then $\tilde{F}_{01}=F^{23}=F_{23}$ and $\tilde{F}_{23}=F^{01}=-F_{01}$, etc. So $\tilde{F}_{01}\to\hat{\tilde{F}}_{01}$ and $\tilde{F}_{23}\to i\hat{\tilde{F}}_{23}$. Therefore $F^{\mu\nu}\tilde{F}_{\mu\nu}\to i\hat{F}^{\mu\nu}\hat{\tilde{F}}_{\mu\nu}$.

Action.

$$
\begin{aligned}
iS &= i\left(-\frac{1}{2g^2}\int d^4x\,\mathrm{Tr}(F^2) + \frac{\theta}{16\pi^2}\int d^4x\,\mathrm{Tr}(F\tilde{F})\right) \\
&\to -\frac{1}{2g^2}\int d^4\hat{x}\,\mathrm{Tr}(\hat{F}^2) + i\frac{\theta}{16\pi^2}\int d^4\hat{x}\,\mathrm{Tr}(\hat{F}\tilde{\hat{F}}) \\
&\equiv -\hat{S}
\end{aligned}
\tag{1.194}
$$

(Whenever it is unspecified the trace is assumed to be in the fundamental representation.) For brevity, when it is clear that we are working in Euclidean signature, we will drop the hats on Euclidean coordinates and fields. To summarize, the Euclidean action of pure YM is

$$
S = \frac{1}{2g^2}\int d^4x\,\mathrm{Tr}(F^2) - i\frac{\theta}{16\pi^2}\int d^4x\,\mathrm{Tr}(F\tilde{F}). \tag{1.195}
$$

where the fields are all real and index contractions are performed with the four dimensional Euclidean metric. Note that the kinetic term is positive-definite, but the θ-term is imaginary.

Chapter 2

EFT

In the last chapter, we covered:

- Scaling and changing the renormalization scale with RG
- Renormalization with higher-dimension operators. Any local QFT is predictive, not just the renormalizable ones.
- Features of mass independent vs mass dependent schemes

Where do higher-dimension operators came from? Why are most QFTS in elementary particle physics renormalizable? (QED, QCD, Electroweak theory...) Examples of cases where higher-dimension operators involved:

- $G_F \psi_\mu \bar{\psi}_e \psi_{\nu e} \bar{\psi}_{\nu_\mu}$ dimension 6 Fermi theory
- $\frac{\pi^0}{f_\pi} F^{\mu\nu} F^{\rho\sigma} \epsilon_{\mu\nu\rho\sigma}$ dimension 5 pion decay
- $\frac{1}{M} H \psi_L H \psi_L$ dimension 5 neutrino mass

We will now introduce EFT and this will be clarified.
EFT: a Lagrangian theory containing

- All dynamical degrees of freedom light compared to a scale M
- All terms in $\mathcal{L}$ consistent with the prescribed symmetries, including higher-dimension operators with appropriate powers of M

Weinberg: such an $\mathcal{L}$ can describe the most general S-matrix for $E < M$. EFT is an organizational tool: tells you

- What sorts of interactions are possible
- How they depend on energy

This is a general "bottom up" view. Renormalization $\Rightarrow$ how to power count in full QFT, how to obtain predictions to fixed accuracy E. Operators of dimension $k+4$ contribute in the semiclassical limit as $(E/M)^k$. So if we demand

$$(E/M)^{k_\epsilon} = \epsilon \Rightarrow k_\epsilon \simeq \frac{\log(\epsilon)}{\log(E/M)}.$$

$k_E \longrightarrow 0$ as $E/M \to 0$: Renormalizable theories obtained

BUT: $k_E \longrightarrow \infty$ as $E/M \to 1$: sign that new physics needed. What is it?

What is missing from EFT is heavy particles. Turn the question around; how to go from a UV theory with both heavy and light fields, to an EFT, just in terms of the light fields? This is "top down EFT." The procedure is conceptually and practically useful.

2.1 Matching, Decoupling, and Integrating Out

Given a UV theory of some light degrees of freedom ψ_L and some heavy degrees of freedom ϕ_H, how do we produce an effective Lagrangian $\mathcal{L}_{eff}[\phi_L]$?

As with many techniques in physics, the terminology has grown somewhat diffuse over time. Informally the procedure is variously called "matching," or "integrating out" or "decoupling" the heavy fields. Somewhat more precisely, integrating out refers to doing a part of the path integral, similar to what we did in our Wilsonian treatment of renormalization. Given

$$\int D\phi_L D\phi_H e^{i\int d^4x \mathcal{L}[\phi_L,\phi_H]} \tag{2.1}$$

we do the path integral over *all* momentum shells for ϕ_H, and over ϕ_L modes with $k^2 \gtrsim m_H^2$.

For a simple example, let us consider doing the path integral in the leading-order saddle point approximation (Euclidean) or stationary phase approximation (Lorentzian). This will be a valid procedure at weak coupling. Generally we can find a saddle point of low Euclidean action where ϕ_H is constant in spacetime, i.e., we look for a constant solution to the equation of motion,

$$\partial V/\partial \phi_H = 0. \tag{2.2}$$

Suppose V admits a Taylor series expansion in the fields, which to quadratic order takes the form

$$V = \frac{1}{2}m_L^2\phi_L^2 + \frac{1}{2}m_H^2\phi_H^2 + \Delta^2\phi_L\phi_H + \dots \tag{2.3}$$

Then

$$\begin{aligned}
&\partial V/\partial \phi_H = m_H^2\phi_H + \Delta^2\phi_L = 0 \\
&\Rightarrow \phi_H = -\phi_L\Delta^2/m_H^2 \\
\text{Reinsert into } \mathcal{L}\text{: } V_{eff}[\phi_L] &= \frac{1}{2}\underbrace{(m_L^2 - \frac{\Delta^4}{m_H^2})}_{\tilde{m}_L^2}\phi_L^2 \\
&\Rightarrow \mathcal{L}_{eff}[\phi_L] = \frac{1}{2}(\partial\phi_L)^2 - \frac{1}{2}\tilde{m}_L^2\phi_L^2 + \mathcal{O}\left(\partial^2/m_H^2\right).
\end{aligned} \tag{2.4}$$

We can do a little bit better if we allow some backreaction: long-wavelength disturbances in ϕ_L will induce disturbances in ϕ_H, so we could allow for this in the saddle-point approximation by using the full equation of motion:

$$\begin{aligned}\Box\phi_H = -\partial V/\partial\phi \quad &\Rightarrow \quad \left(p^2 - m_H^2\right)\tilde{\phi}_H = \Delta^2\tilde{\phi}_L \\ &\Rightarrow \quad \tilde{\phi}_H = -\frac{\Delta^2}{m_H^2}\tilde{\phi}_L\left(1 + \frac{p^2}{p_H^2} + \cdots\right).\end{aligned} \tag{2.5}$$

which we can then insert to get higher derivative terms in the effective Lagrangian.

This is a useful technique in simple, weakly coupled cases where leading order (tree level accuracy) is good enough. It is exactly like Born-Oppenheimer in quantum mechanics: you solve for the "fast modes" (ϕ_H, or the electron) in the stationary background of the "slow modes" (ϕ_L, or the molecule) to get an effective theory of the latter.

We can increase the precision by "matching." Given $\mathcal{L}(\phi_L, \phi_H)$ and heavy scale m_H,

- write general $\mathcal{L}_{eff}(\phi_L)$ including higher dimension operators consistent with the symmetries
- fix the couplings in $\mathcal{L}_{eff}$ by requiring $\langle\phi_L\cdots\phi_L\rangle_{\mathcal{L}} = \langle\phi_L\cdots\phi_L\rangle_{\mathcal{L}_{eff}}$ for momenta $p \ll m_H$

Again this very similar to what we did in our Wilsonian treatment of renormalization, but we will use it in the standard continuum renormalization approach, with mass-independent schemes.

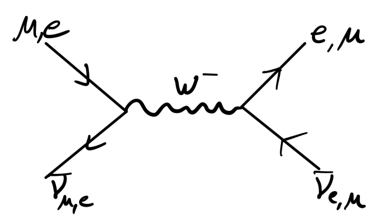
2.1.1 Example 1: the Fermi electroweak theory

One of the earliest examples of an EFT was introduced by Fermi to model muon decay. In the standard model, muon decay is governed by the charged current interactions,

$$\begin{aligned}\mathcal{L} &= \frac{e}{\sqrt{2}\sin\theta_W}\left(W^{+\mu}J_\mu^+ + W^{-\mu}J_\mu^-\right) \\ J_\mu^+ &= \bar{\nu}_{eL}\gamma^\mu e_L + \bar{\nu}_{\mu L}\gamma^\mu\mu_L + (\text{quarks, tau}) \\ J_\mu^- &= \bar{e}_L\gamma^\mu\nu_{eL} + \bar{\mu}_L\gamma^\mu\nu_{\mu L} + (\dots),\end{aligned} \tag{2.6}$$

Here $e/\sin\theta_W$ is a coupling, $W_\pm$ are heavy charged spin-1 gauge bosons, $\nu_{e,\mu}$, e, and μ are light leptons described by Dirac fermions, and $\nu_L = \frac{1}{2}(1-\gamma_5)\nu$ projects out LH Weyl components.

This interaction gives rise to tree-level scattering amplitudes of the form

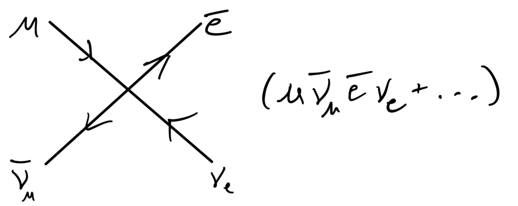


$$\sim \left(\bar{e}_L\gamma^\mu\nu_{eL} + \bar{\mu}_L\gamma^\mu\nu_{\mu L}\right)\left(\frac{ie}{\sqrt{2}\sin\theta_W}\right)^2 \frac{i\left(g^{\mu\nu} - \frac{p^\mu p^\nu}{m_W^2}\right)}{p^2 - m_W^2}\left(\bar{\nu}_{eL}\gamma_\nu e_L + \bar{\nu}_{\mu L}\gamma_\nu\mu_L\right) \tag{2.7}$$

and similar diagrams for the decay $\mu \to \nu_\mu e\bar{\nu}_e$. Now $m_\mu \simeq 105$ MeV is about a thousand times lighter than the 80 GeV W-boson, and the electron and neutrinos are even lighter than the muon.

So at center of mass energies $\sqrt{s} \ll m_W$, we should be able to write a theory of lepton scattering and decay that just includes $\mu, e, \nu_{\mu,e}$, with no W.

What interactions do we need? The diagram above tells us we need



We can trivially read off the effective Lagrangian, for $p^2 \ll m_W^2$,

$$\mathcal{L}_{eff} = \frac{e^2}{2\sin^2\theta_W m_W^2} \left(\bar{e}_L \gamma^\mu \nu_{eL} + \bar{\mu}_L \gamma^\mu \nu_{\mu L}\right)^2 + \mathcal{O}(\partial^2/m_W^2) \tag{2.8}$$

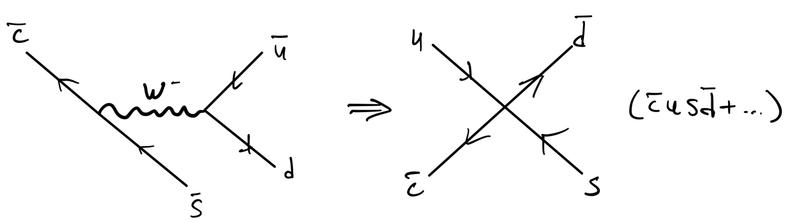
At low energies the leading term is a dimension 6 operator with no derivatives. Derivative corrections are highly suppressed.

Fermi's version had an undetermined coefficient that one could fit from measurements of the muon lifetime. It is now called the Fermi constant, G_F, and is related to the electroweak theory parameters by the matching procedure above. Including conventional numerical factors, it is

$$G_F = \frac{\sqrt{2}e^2}{8\sin^2\theta_W m_W^2}. \tag{2.9}$$

In the Standard Model, $G_F = 1/(2\sqrt{2}v^2)$, where $v = 174$ GeV is the Higgs vev. Muon decay is still the best way to measure this parameter, despite taking place three orders of magnitude lower in energy..

Similarly, integrating out the W produces a 4-quark charged current effective Lagrangian. This effective Lagrangian describes various flavor-changing processes, for example, the operators that we encountered in our discussion of $\Delta S = 1$ processes, cf. Eq (1.96). These were generated by matching the diagram



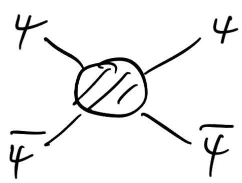
2.1.2 Example 2: Other 4-Fermi tensor structures

Now let's look at a 4-Fermi operator generated by scalar couplings. We start from the Lagrangian:

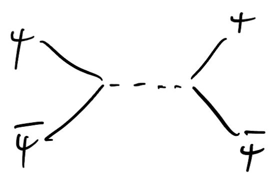
$$\mathcal{L} = \mathcal{L}_{\text{kinetic}} + y\phi\bar{\psi}\psi + \frac{1}{2}m_\phi^2\phi^2 \tag{2.10}$$

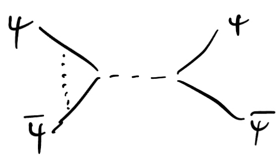
Here ψ is light and ϕ is heavy. What $\mathcal{L}$ describes low-energy ψ interactions?

$$\mathcal{L}_{eff} = \bar{\psi}\partial\!\!\!/\psi + m_\psi \bar{\psi}\psi + \frac{\lambda_1}{m_\phi^2}(\bar{\psi}\psi)^2 + \frac{\lambda_2}{m_\phi^2}(\bar{\psi}\gamma^\mu\psi)\left(\bar{\psi}\gamma_\mu\psi\right) + \dots \tag{2.11}$$

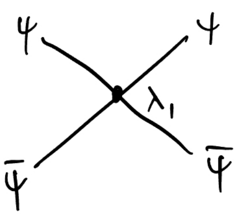
Consider 

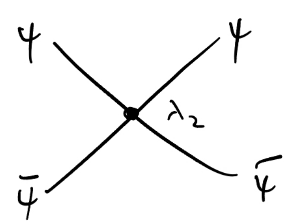
at $p^2 \ll m_\phi^2$ in $\mathcal{L}$ and $\mathcal{L}_{eff}$.

From $\mathcal{L}$: 

$+$ 

$+$ higher order.

From $\mathcal{L}_{eff}$: 

$+$ 

$+ \dots.$

At tree level, the matrix elements are of the form

$$\begin{aligned}
\mathcal{M}_{\mathcal{L}} &\sim (\bar{u}v)\frac{i(iy)^2}{s - m_{\phi^2}}(\bar{v}u) + (s \leftrightarrow u) + \dots \\
\mathcal{M}_{\mathcal{L}_{eff}} &\sim (\bar{u}v)\frac{i\lambda_1}{m_{\phi^2}}(\bar{v}u) + (\bar{u}\gamma^\nu v)\frac{i\lambda_2}{m_{\phi^2}}(\bar{v}\gamma_\mu u) + \dots.
\end{aligned} \tag{2.12}$$

For $s, u \ll m_\phi^2$ we see that matching requires $\lambda_1 = y^2$, $\lambda_2 = 0$. $\mathcal{L}_{eff}$ is another example of a 4-Fermi theory.

There are two types of corrections. First, as we noted in the electroweak theory, matching contributions of order (external momenta/m_ϕ)^n requires introducing higher-derivative operators into $\mathcal{L}_{eff}$. Second, when we include loops of the heavy field ϕ, the matching can receive corrections in powers of y. These corrections can be minimized, and often neglected in first approximation, by judicious use of the RG. Consider the loop diagram drawn above. It shifts the four-point amplitude to something of the form

$$\mathcal{M}_{\mathcal{L}} \sim (\bar{u}v)\frac{i(iy)^2}{s - m_{\phi^2}}\left[1 + c\frac{y^2}{16\pi^2}\log(\mu/m_\phi)\right](\bar{v}u) + (s \leftrightarrow u) + \dots . \tag{2.13}$$

Above we found that tree level matching sets $\lambda_1 = y^2$, but it doesn't specify the renormalization scale at which this relation holds. We see that the optimal choice is $\mu = m_\phi$: the loop corrections involving the heavy field are minimized (because large logs are resummed). **We "integrate out heavy fields at their thresholds."**

2.1.3 Match and Run EFT

These examples were simple toy models. They were of conceptual value, particular to understand where HDOs might "come from," if you don't know $\mathcal{L}$, as Fermi didn't.

But why bother with $\mathcal{L}_{eff}$ if you do know $\mathcal{L}$? It turns out to be very convenient for doing precise computations, using RG+matching. Easier than using full $\mathcal{L}$.

The typical structure of a match-and-run computation uses the lessons we've learned in our study of renormalization and RG. Here is the basic framework which governs a very wide range of EFT applications:

$$\begin{array}{ll} \mathcal{L}(\phi_L, \phi_H) & \\ \quad\Downarrow & \text{RG evolve } \mathcal{L} \\ \mu = m_H & \text{match: tweak coeffs in } \mathcal{L}_{eff}(\phi_L) \\ \quad\Downarrow & \text{RG evolve } \mathcal{L}_{eff} \\ \mu \simeq E & \end{array}$$

Mass independent schemes: can match anywhere. Higher-dimension operators only contribute to the β functions of lower-dimension operators by positive powers of m/m_H. $\mu = m_H$ especially convenient (no large logs).

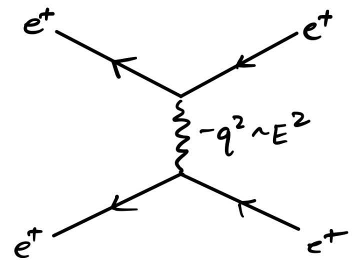
Mass dependent schemes: can only match at $\mu < M_H$, because higher dimension operators contribute to the β functions of lower dimension operators as positive powers of μ/m_H, which is uncontrolled as $\mu \to m_H$. This is inconvenient and misses the advantage of large log resummation.

$\Rightarrow DR+\overline{MS}$ is much more convenient for this type of computation. It also resums "infrared logs" that are difficult to capture in the full theory, as the next example illustrates.

2.1.4 Example 3: QED

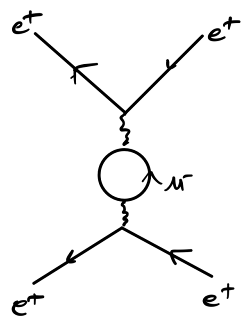
In our previous examples, we carried out the first two steps in the match-and-run flowchart above. The chart involves two periods of RG evolution, in two different theories: the UV theory, and the effective theory. Let's do one more example that illustrates what's happening in the second period of RG evolution.

Suppose we measure the electromagnetic coupling e by scattering positrons, $e^+e^+ \to e^+e^+$, at momentum transfers of about 500 MeV. The light particles are $e^\pm$, $\mu^\pm$. From



we obtain $e^2(\mu_0 \sim 500\text{MeV})$.

Now suppose we want to know how e^+e^+ scattering behaves at 10 MeV. Here the only light particle is $e^\pm$. If we use our $\overline{MS}$ β function and run e down to 10 MeV in the $e^\pm\mu^\pm$ theory, there will still be large 1-loop corrections from



$$\sim e^4 \log(m_\mu/E)$$

This is a general problem when attempting to resum logs by choosing an appropriate renormalization scale: what if there is a large hierarchy of scales? (This is only a problem for mass-independent schemes, which do not automatically implement the decoupling theorem, but we have seen these are very useful for simplifying the RG and expanding the validity of the perturbative expansion.) To address the problem, as in the heavy-scalar-light-fermion example, we use a careful choice of matching scale: we match onto a new EFT at each heavy-particle mass threshold.

In this QED example, matching onto an $e^\pm$ EFT is easy: we just delete the muon from the theory! No tree level modifications to the couplings are needed, because all electromagnetic couplings involve zero or two muons. (There are 1-loop effects, but by matching at $\mu = m_\mu$, these are small, as in the heavy-scalar-light-fermion example.)

So what changes in the computation? The electromagnetic β function! The second step of RG evolution, in the $e^\pm$ EFT, evolves the coupling more slowly:

$$\mathcal{L}: \;\; \beta_e = \frac{e^3}{12\pi^2} \times 2$$
$$\mathcal{L}_{eff}: \;\; \beta_e = \frac{e^3}{12\pi^2} \times 1 \tag{2.14}$$

The factor of 2 in the $\mathcal{L}$ theory comes from the electron and the muon loops. It is reduced to a factor of 1 in the $\mathcal{L}_{eff}$ theory, which contains only the electron. The coupling is continuous at the muon threshold, but the β function is not.

$$\begin{aligned} \text{RG1: } e(m_{\mu^\pm}) &= \frac{e^2(\mu_0)}{1 + \frac{e^2(\mu_0)}{6\pi^2}\log\left(\frac{\mu_0}{m_\mu}\right)} \\ \text{Match: } e^{\text{EFT}}(m_{\mu^\pm}) &= e(m_{\mu^\pm}) \\ \text{RG2: } e^{\text{EFT}}(10\text{ MeV}) &= \frac{e^2(m_{\mu^\pm})}{1 + \frac{e^2(m_{\mu^\pm})}{12\pi^2}\log\left(\frac{m_{\mu^\pm}}{10\text{ MeV}}\right)} \end{aligned} \tag{2.15}$$

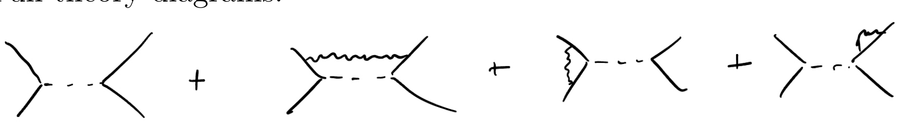
Using e^{EFT} (10 MeV), we can capture with a tree-level computation both $\log^n\left(\frac{500\text{ MeV}}{10\text{ MeV}}\right)$ ultraviolet logs from the full theory, plus $\log^n\left(\frac{m_\mu}{10\text{ MeV}}\right)$ "infrared logs" or "finite logs" from full theory. This is a general precision-advantage of using the EFT, instead of the full theory: **RG in the EFT can resum large logs that are not captured by RG in the full theory, in mass-independent schemes.**

We should be a little more precise. One can resum large infrared logs using RG in the full UV theory by using a mass-dependent scheme, where heavy particles decouple from β-functions automatically. So if you know the full theory, or want to test a hypothesis for a full theory, why not just do the low energy computation in that theory? The reason is it is easier in almost all technical applications. The EFT method breaks the computation into steps in energy scale, and focuses only on the important degrees of freedom and interactions at each scale.

Here is another example: QED corrections to our previous 4-fermion operator generated in Yukawa theory.

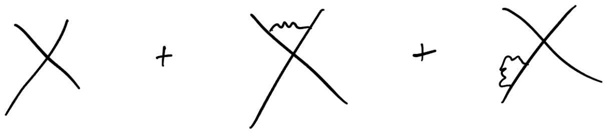
Say the fermion ψ is charged and we want the low energy cross section for $\psi\psi \to \psi\psi$, including quantum corrections from QED.

Full theory diagrams:[1]


$+$ permutations $+ \ldots$

[1]If we work in Landau gauge, the electron self-energy diagram does not generate an anomalous dimension, and the corresponding diagrams here and in the EFT expansion below can be dropped.

EFT diagrams:



$+ \text{permutations} + \dots$

The EFT loops renormalize the new 4-fermion coupling λ.

$$\beta_\lambda^{\overline{MS}} = -\frac{3e^2\lambda}{8\pi^2}, \quad \beta_e^{\overline{MS}} = \frac{e^3}{12\pi^2} \tag{2.16}$$

So

$$\begin{aligned} \frac{d\lambda}{d\log\mu} &= \lambda\frac{d\log e}{d\log\mu}\left(-\frac{3}{8\pi^2}\right)\left(12\pi^2\right) \\ \rightarrow \frac{d\log\lambda}{d\log\mu} &= \frac{d\log e}{d\log\mu} \times \left(-\frac{9}{2}\right) \\ \rightarrow \lambda(\mu) &= \lambda(\mu_0)\left(\frac{e(\mu)}{e(\mu_0)}\right)^{-\frac{9}{2}}. \end{aligned} \tag{2.17}$$

So if we know $\mathcal{L}$ at $\mu_0 = m_\phi$, we can match onto $\mathcal{L}_{eff}$ and run to $\mu \ll m_\phi$.
This is much easier than computing the full theory diagrams, which include finite large logs, divergent logs, box diagrams, etc. With the EFT method we miss nonlogarithmic 1-loop terms (from 1-loop matching) and finite p^2/m_ϕ^2 corrections. We gain $e^{2n}\log^n(p^2/m_\phi^2)$ resummation, and ease of computation.

2.1.5 Example 4: Strange physics

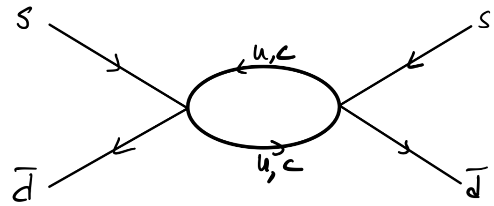
Actually, we already discussed – out of order – another example of matching and running in an EFT, the $\Delta S = 1$ and $\Delta S = 2$ 4-quark operators in the Fermi electroweak theory. We saw that the $\Delta S = 1$ operators are generated by tree level matching of diagrams involving W bosons, and prior to that, we discussed the running of both $\Delta S = 1$ and $\Delta S = 2$ operators in the EFT.

In this last example, we'll illustrate two more elaborations on matching: (1) one-loop matching, and (2) of an EFT onto another EFT. We'll use the $\Delta S = 2$ operator (1.103) of the Fermi theory, which we repeat here for convenience:

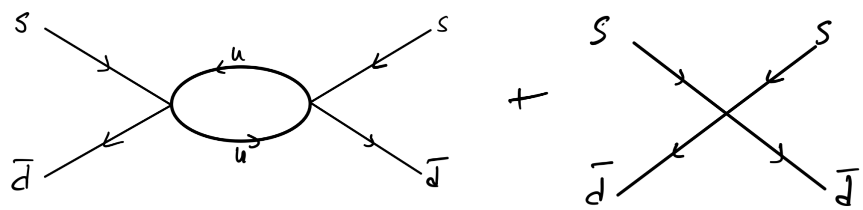
$$O_+^{\bar{d}s} = \frac{1}{2}\left(\bar{d}_\alpha\gamma^\mu(1+\gamma_5)s_\alpha\bar{d}_\beta\gamma_\mu(1+\gamma_5)s_\beta + \bar{d}_\alpha\gamma^\mu(1+\gamma_5)s_\beta\bar{d}_\beta\gamma_\mu(1+\gamma_5)s_\alpha\right). \tag{2.18}$$

If mixing with the third generation is neglected, then $O_+^{\bar{d}s}$ is not generated above the charm threshold m_c. (It can't be generated by integrating out the top, and the only contribution to the beta function is Eq. (1.104), which is attached to a flavor structure involving 3rd generation CKM elements V_{ts} and $V_{dt}^\dagger$.)

However, the operator is generated at the charm threshold by matching. In the theory with the charm quark, there are diagrams:



Below the charm threshold we have a new EFT with the charm quark removed, and the previous diagrams must be matched by



Thus the $\Delta S = 2$ effective operator is generated by matching at the charm threshold:

$$\delta\mathcal{L} = -\sin^2\theta_c \cos^2\theta_c m_c^2 \frac{G_F^2}{16\pi^2}\left[\frac{3}{2}h_+^2 - h_+h_- + \frac{1}{2}h_-^2\right] O_+^{\bar{d}s} \tag{2.19}$$

where θ_c is the Cabibbo angle, $\cos\theta_c = V_{ud}$ and $\sin\theta_c = V_{us}$, and the $\Delta S = 1$ couplings $h_\pm$ are renormalized at $\mu = m_c$.

This operator has a matrix element $\langle K^0|\delta\mathcal{L}|\bar{K}^0\rangle$ which contributes to the off-diagonal element of the $K^0-\bar{K}^0$ mass matrix. This off-diagonal element splits the mass eigenstates. If m_c is too large, the splitting would exceed the tiny observed value $\sim 3.5\times10^{-6}$ eV. This led to the prediction (Lee and Gaillard, 1974) that the charm quark should be found at the GeV scale. This was confirmed with the discovery of a 3.1 GeV $c\bar{c}$ bound state, the J/ψ meson, in November 1974.

2.1.6 Polemic interlude

$DR + \overline{MS}$ is easy and convenient. Nonetheless it is non-intuitive (modifies the UV and IR); occasionally misleading (questions of fine-tuning); and probably ill-defined nonperturbatively.

It is a tool, like any regulator/schemee, and must ultimately arrive at the same end, if well-defined.

Use it when it is the right tool for the job! Perturbation theory in gauge theories, match & run EFT.

EFT is also much more than a top-down calculational tool, a way to resum large logs, or a conceptual device. In some cases it is the only quantitative (analytic) theory we have:

- When heavy degrees of freedom are strongly coupled, as in QCD, so that matching is non-perturbative
- When we don't know the UV theory, for example when we want to parametrize deviations from the SM.

Shortly we will return to "bottom up" EFT. To write the most general Lagrangian, we need to identify the light degrees of freedom and the symmetries of the dynamics. Often these questions are related: the symmetries, and their realization, tell us something about what light DOF are expected.

2.2 Noether's Theorem

Consideration of the symmetries plays an essential role in effective field theory techniques. So, we will spend some time reviewing the basics.

There are many "types" of symmetries in QFT:

{exact,approximate}×{continuous,discrete} ×{global,gauge} × {ordinary,generalized}.

We will touch on most at different points in this course, but in this section we focus on ordinary continuous global symmetries. We begin by reviewing Noether's Theorem, then discuss spontaneous symmetry breaking and the Goldstone theorem, and finally introduce the CCWZ formalism for writing down an EFT of Nambu-Goldstone bosons.

Noether's theorem is an important result linking *continuous global symmetries* (either internal or spacetime) to conserved currents.

Classically, this link is derived as follows. Denote the set of fields as a vector Φ^i and the infinitesimal change in Φ^i under a transformation $\epsilon\delta_G\Phi^i$. Here ϵ is an infinitesimal parameter that we allow to vary over spacetime. If the action is a function of the fields and their gradients, $S = S[\Phi^i, \partial_\mu\Phi^i]$, then the variation is

$$\delta S = \int \left(\epsilon \left[\delta_G\Phi^i \frac{\delta S}{\delta\Phi^i} + \partial_\mu\delta_G\Phi^i \frac{\delta S}{\delta\partial_\mu\Phi^i} \right] + \partial_\mu\epsilon\delta_G\Phi^i \frac{\delta S}{\delta\partial_\mu\Phi^i} \right) . \tag{2.20}$$

At this stage, it is convenient to reorganize δS in slightly different ways depending on the goal.

First, note that the transformation $\delta_G\Phi^i$ is a *global symmetry* if δS is equal to a boundary term for constant ϵ. This implies the term in brackets is either zero or a total derivative, $\partial_\mu\mathcal{J}^\mu$. (A useful rule of thumb is that $\mathcal{J}^\mu$ is nonvanishing when the transformations relate fields at different points in spacetime. Another example is $\mathcal{J}^\mu$ is equal to the chern-simons current in axion electrodynamics if $\delta_G\Phi^i$ is the axion shift symmetry.) If we perform a global symmetry transformation with non-constant ϵ (but rapidly vanishing at infinity, so that we can integrate the last term in (2.20) by parts), then

$$\delta S = -\int \epsilon(x)\partial_\mu J^\mu , \qquad J^\mu \equiv \delta_G\Phi^i \frac{\delta S}{\delta\partial_\mu\Phi^i} - \mathcal{J}^\mu . \tag{2.21}$$

By definition, $\delta S = 0$ when the fields satisfy their equations of motion, even for non-constant ϵ. Therefore, classically, J^μ must be *conserved*:

$$\partial_\mu J^\mu = 0 . \tag{2.22}$$

Below, we will also see that this holds quantum mechanically (up to possible contact terms and anomalies); the reason is we may vary the generating functional wrt $\epsilon(x)$ to get correlators of $\partial_\mu J^\mu$, and these variations must vanish since ϵ just parametrizes changes of dummy (integration) variables.

The form of δS in Eq. (2.21) is useful for deriving what the conserved current is, and for obtaining a generating functional for J^μ correlators without assuming the EOM. However, it cannot tell us how $\mathcal{L}$ changes when the transformation is not an exact symmetry. (Although since we generally use the EOM, we know it will be some total derivative.)

Alternatively, let us define the current

$$K^\mu = \delta_G \Phi^i \frac{\delta S}{\delta \partial_\mu \Phi^i} \ . \tag{2.23}$$

Then, without assuming constant ϵ, or that the transformation is a symmetry, we have

$$\delta S = \int \left(\epsilon \delta_G \Phi^i \times [EOM] + \partial_\mu(\epsilon K^\mu)\right) . \tag{2.24}$$

If we use the equation of motion and take constant ϵ, this becomes

$$\delta S = \int \epsilon \partial_\mu K^\mu \ . \tag{2.25}$$

The form of δS in Eq. (2.25) is useful for deriving this $\delta \mathcal{L}$ (total derivative), and we see that it is equal to the conserved current if $\partial_\mu \mathcal{J}^\mu = 0$. But Eq. (2.25) cannot be used to derive the conserved current, because having assumed $\epsilon = const$, we do not need $\partial_\mu K^\mu = 0$. In typical cases with global symmetries, $K = J$ and neither are exact, so under the transformation the Lagrangian shifts by $+\epsilon \partial_\mu J^\mu$.

The above holds for classical symmetries. In QFT, Noether's theorem becomes a statement about correlation functions of the operator $\partial_\mu J^\mu$. Denote the generating functional of correlation functions as

$$Z[j] = \int \mathcal{D}\Phi \ e^{iS[\Phi^i] - i\int j^i \Phi^i} \ . \tag{2.26}$$

Let us treat the global symmetry transformation as a change of variables in the path integral. Since it is only a change of integration variables, the generating functional remains the same. We assume the measure is invariant (we will discuss anomalies later.) Then, using our previous results for the variation of the classical action, the path integral takes the form

$$Z[j] = \int \mathcal{D}\Phi \ e^{iS[\Phi^i] - i\int \epsilon \partial_\mu J^\mu - i\int j_i \Phi^i - i\int \epsilon j_i \delta_G \Phi^i} \ . \tag{2.27}$$

The simplest correlation function involving $\partial_\mu J^\mu$ is just

$$\langle \partial_\mu J^\mu \rangle = i \frac{1}{Z} \frac{\delta Z}{\delta \epsilon}\bigg|_{\epsilon = j = 0} = 0 \ . \tag{2.28}$$

More general correlation functions involving $\partial_\mu J^\mu$ and other field insertions have the form

$$(i)^n \frac{1}{Z}\frac{\delta \cdots \delta Z}{\delta j \cdots \delta j \delta\epsilon \cdots \delta\epsilon}\bigg|_{\epsilon=j=0} = \langle \partial_\mu J^\mu \cdots \partial_\mu J^\mu \Phi^i \cdots \Phi^i \rangle + \cdots = 0. \tag{2.29}$$

The first term does not completely vanish, being proportional to "contact terms" which arise from $\frac{\delta^2}{\delta j(x)\delta\epsilon(y)}\int dz \epsilon j \delta_G \Phi$. *****edited - used to not have the ellipses*****

2.2.1 Example: $O(N)$ model

Take a model with N scalar fields and action given by

$$S = \int \frac{1}{2}(\partial_\mu \phi^a)(\partial^\mu \phi^a) + V(\phi^a \phi^a) \ , \qquad a = 1 \ldots N \ . \tag{2.30}$$

The action is invariant under rotations of ϕ^a treated as an N-vector. ϕ^a transforms in the fundamental representation of $O(N)$,

$$\phi^a \to R^a_b(\epsilon)\phi^a \ , \quad R = e^{i\epsilon^A T^A} \tag{2.31}$$

where the (imaginary, $N \times N$, skew-symmetric) matrices T^A generate the group $O(N)$. Infinitesimally,

$$\phi^a \to \phi^a + i\epsilon^A (T^A)^a_b \phi^b \ . \tag{2.32}$$

Since the symmetry is internal, the Noether current is just

$$J^{A,\mu} = \delta\phi^a \frac{\delta S}{\delta \partial_\mu \phi^a} = \phi^a (T^A)^b_a \partial_\mu \phi_b \ . \tag{2.33}$$

There is one for each generator of rotations.

2.2.2 Example: Shift symmetries

$$S = \int \frac{1}{2}(\partial_\mu \theta)(\partial^\mu \theta) \ . \tag{2.34}$$

This action is invariant under a *nonlinear* symmetry, a shift $\theta \to \theta + \epsilon f$, where f is a parameter of the theory with the same dimensions of θ. The current is

$$J^\mu = f \partial^\mu \theta \ . \tag{2.35}$$

These types of symmetries and their generalizations play an important role in many EFTs. The reason is that a shift symmetry can forbid a mass term for a field (or keep it small if the amount of symmetry breaking is small), so it is a principled way to introduce light fields.

2.3 Spontaneously Broken Symmetries

Take the $O(N)$ model for $N = 2$. We can write this in terms of a complex scalar field ϕ by identifying $\phi^1 \to \mathrm{Re}(\phi)$, $\phi^2 \to \mathrm{Im}(\phi)$,

$$S = \int |\partial_\mu \phi|^2 - V(|\phi|) \ . \tag{2.36}$$

Then the action is invariant under $U(1)$ transformations, $\phi \to e^{i\alpha}\phi$, and the current is

$$J^\mu = -i(\phi^* \partial_\mu \phi - \phi \partial_\mu \phi^*) \ . \tag{2.37}$$

(Note that we have to treat ϕ and ϕ^* as independent variables to derive this current, but it follows naturally from working with the real and imaginary components.)

Now let us take a specific form for the potential:

$$V(\phi) = m^2|\phi|^2 + \lambda|\phi|^4 \ . \tag{2.38}$$

The potential has no impact on the form of the Noether current, regardless of whether it breaks the symmetry. Furthermore, this potential preserves the symmetry.

The minimum of the potential is $\langle\phi\rangle = 0$ when the parameter $m^2 > 0$. This vacuum transforms trivially under the $U(1)$ symmetry. When the parameter m^2 is negative, however, the $U(1)$ is spontaneously broken:

$$\langle|\phi|^2\rangle = -\frac{m^2}{2\lambda} \equiv \frac{v^2}{2} \ . \tag{2.39}$$

The minima of the potential form a continuous circle around the origin in field space, with radius $v/\sqrt{2}$. Any given point along this circle is not invariant under the $U(1)$; rather, points are taken into each other by the symmetry transformation.

Let us expand our field in small fluctuations around one point along the trough:

$$\phi(x) = \frac{1}{\sqrt{2}}\left(v + \delta v(x)\right) e^{i\alpha(x)} \tag{2.40}$$

where we have introduced real fields δv and α, and α is dimensionless and compact, $\alpha = \alpha + 2\pi$. The action now takes the form

$$S = \int \frac{1}{2}(\partial_\mu \delta v)^2 + \left(\frac{v^2}{2} + v\delta v + (\delta v^2)\right)(\partial_\mu \alpha)^2 - \lambda v^2 (\delta v)^2 + V_{int}(\delta v) + V_0 \tag{2.41}$$

where V_{int} is an interaction potential depending only on δv and V_0 is a constant. There are several things to note. Since $\lambda > 0$, the "radial fluctuation" δv is now massive, $m^2_{\delta v} = \lambda v^2$. The "angular fluctuation" is massless. Furthermore, it only interacts with itself and δv through derivative couplings. Let's look at the current in this new basis:

$$J^\mu = v^2 \partial^\mu \alpha \ , \tag{2.42}$$

exactly the form of a shift symmetry, with $f = v^2$ and $\theta = \alpha$. Indeed, if we started from the action (2.41), we would have concluded that the theory has a global shift symmetry for α, since it has only derivative couplings.

It is a more general fact that shift symmetries naturally appear from the spontaneous breaking of continuous global symmetries. The precise statement is the Goldstone Theorem, which plays a critical role in EFT. Goldstone Theorem: Spontaneous breaking of a global symmetry implies the presence of a massless particle with charges matching those of the Noether charge.

We will prove the statement with the classical action, but it can be easily generalized to the full quantum theory by replacing $S[\phi] \to \Gamma[\phi]$, the generating functional of 1-particle irreducible Green functions.

Take the field transformation to have the form

$$\Phi^a \to \Phi^a + i\epsilon T^a_b \Phi^b. \tag{2.43}$$

Then the transformation is a global internal symmetry if

$$\frac{\delta S}{\delta \Phi^a} T^a_b \Phi^b = 0 \ . \tag{2.44}$$

Let us evaluate the functional derivative of this wrt Φ^c:

$$\frac{\delta^2 S}{\delta \Phi^a \delta \Phi^c} T^a_b \Phi^b + \frac{\delta S}{\delta \Phi^a} T^a_c = 0 \ . \tag{2.45}$$

Evaluating this expression on constant fields at the minimum of the potential, Φ_0, the first variation $\delta S/\delta \Phi^a$ vanishes. Meanwhile at this point the second variation $\delta^2 S/\delta \Phi^2$ is the scalar mass-squared matrix. Therefore

$$(M^2)_{ac} T^a_b \Phi^b_0 = 0 \ . \tag{2.46}$$

So either $\Phi_0 = 0$ (unbroken symmetry) M^2 has a vanishing eigenvalue corresponding to eigenvector $T^a_b \Phi^b_0$ (broken symmetry). Furthermore we obtain one such eigenvector for every symmetry generator such that $T^a_b \Phi^b_0 \neq 0$.

The massless states are Nambu-Goldstone bosons (NGBs). (NG fermions are possible with spontaneous breaking of supersymmetry.)

2.4 Anomalies

Some familiarity with anomalies and their computation is assumed. This section is meant to serve as a telegraphic summary of about one percent of the dizzying panoply of anomalies discussed in the literature and a reference for some formulas. For an explicit computation of a chiral Adler-Bell-Jackiw (ABJ) anomaly in $d = 4$, see the appendix to chapter 1. For mixed anomalies in generalized global symmetries, anomalies in different dimensions, and gravitational anomalies, see a future version of these notes.

There are three basic types of anomalies: ‘t Hooft, ABJ, and gauge. They all have to do with the transformation of the effective action under global symmetries and gauge symmetries, respectively. Usually, you first meet these anomalies in a QFT class in the context of continuous symmetries in four dimensions, where they are associated with fermion triangle diagrams with zero, two, and three external gauge bosons. But they are more general than this, and arise in any dimensions, with continuous and discrete symmetries, both ordinary and generalized.

Let us start with a general approach for ordinary continuous symmetries. For all symmetries $S_A, S_B, S_C, \ldots$ of the action, we introduce background gauge fields $A, B, C, \ldots$. Here background means they are not yet integrated over in the path integral. They may be integrated over eventually, or not. Let $\Gamma(A_\mu, B_\mu, \ldots)$ be the effective action, obtained by path integrating over the matter fields. It depends on the gauge fields $A_\mu, B_\mu, \ldots$. The induced currents are the response of the effective action to changes in the gauge fields:

$$J^\mu_A = \frac{\delta\Gamma}{\delta A_\mu} \ . \tag{2.47}$$

An infinitesimal gauge transformation associated with symmetry S_A acts on the gauge field as $\delta_\Lambda A_\mu = D_\mu \Lambda$. Then the variation in the effective action is

$$\delta_\Lambda \Gamma \equiv G = \mathrm{Tr} \int dx \, D_\mu \Lambda(x) \frac{\delta\Gamma}{\delta A_\mu(x)} \ . \tag{2.48}$$

Here the trace and integral sum over all the degrees of freedom that vary. Integrating by parts,

$$G = \mathrm{Tr} \int dx \, \Lambda \, D_\mu J^\mu_A. \tag{2.49}$$

Then if the current is covariantly conserved, $D_\mu J^\mu_A = 0$, the effective action is invariant under symmetry S_A. Conversely, when the action is not invariant under S_A, the current is not covariantly conserved.

By explicit computation of the effective action we may obtain expressions for G. Here is a brief sketch – refer to the appendix of Chapter 1 for a full derivation of the triangle diagram with two external gauge fields.

- In 4D the anomaly is associated with various triangle diagrams. With three external gauge boson lines, we have a contribution to Γ involving three gauge fields. These may be different types of gauge fields, or the same type; let us refer to them all as “A” for brevity. We then make a gauge variation, in which $\delta A \to \partial\Lambda + \mathcal{O}(A)$ so that $G \sim A^2 + \mathcal{O}(A^3)$. In fact it is just part of $G \sim \int \mathrm{Tr}(\Lambda F \tilde{F})$.
- Another approach is to compute a triangle diagram with only two external gauge field lines and a current insertion in the fermion loop. This is not computing a term in the effective action, but rather the variation of the anomaly term with respect to A – it is showing that $J = \delta\Gamma/\delta A \sim AF$. We can contract J with $D\Lambda$ and integrate it to make G, and with integration by parts we again find $G \sim \int \mathrm{Tr}(\Lambda F \tilde{F})$.

- As alluded to above, many of these manipulations can be carried out at leading order in fields. E.g., we can linearize the YM field strength to $F = dA$. The triangle diagram is the lowest-order contribution to the anomaly in this sense. There are diagrams with more external gauge bosons that build up the nonlinear terms in the F's.

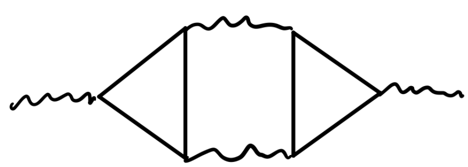
The upshot is that in four dimensions, with ordinary continuous symmetries, one finds that the divergence of a current J^μ associated with chiral symmetry S_A may be nonzero in the presence of background fields:

$$\partial_\mu J^\mu = c\frac{g^2}{8\pi^2}\mathrm{Tr}(F\tilde{F}) \qquad (\text{or } c\frac{g^2}{8\pi^2}F\tilde{F} \text{ if } F \text{ is abelian.}) \tag{2.50}$$

c is a constant that depends on the multiplicities and representations of the fermions in the loop, and g and F are the coupling and field strength associated with the background fields. In general, J^μ may couple to a background gauge field A, and F may be associated with a different gauge field $F = dB + \ldots$, or the same gauge field $F = dA + \ldots$.[2] The three types of anomalies correspond to whether A and/or B are dynamical – whether they are path-integrated over to obtain the partition function or correlation functions.

2.4.1 Gauge anomalies

Here both A and B are path-integrated over. This is a significant problem. The effective action is not gauge invariant. Gauge anomalies can imply a loss of unitarity. One way to see it is that they are couplings of the longitudinal gauge fields to transverse modes, and so there can be a longitudinal pole in the 4-point function of transverse gauge fields. Diagrammatically this is two triangle diagrams attached to each other:



Usually we just demand that all gauge anomalies cancel. This is sufficient, but slightly more than necessary – it is possible to concoct EFTs where there is a gauge anomaly that is fixed up in the UV. We will not discuss this case further, but see "Gauge anomalies in an effective field theory" (Preskill 1990) for a precise discussion.

Take the case where A and B are the same gauge field and the fermions are in representation r (which may not be irreducible). This is the basic anomaly of a gauge theory. The full structure of the anomaly equation is

$$\partial_\mu J^{\mu a} = \frac{g^2}{16\pi^2}\mathcal{A}^{abc}F^{\mu\nu b}\tilde{F}^c_{\mu\nu} \tag{2.51}$$

[2] Also, if F is associated with an Abelian gauge field, then it is gauge invariant by itself, and we can have objects like $F\tilde{F}'$, where F' is the field strength associated with a third gauge field $F = dC + \ldots$. This is not gauge invariant, and therefore does not arise, if S_B is nonabelian.

where

$$\mathcal{A}^{abc} \equiv \mathrm{Tr}(t_r^a\{t_r^b, t_r^c\}) \tag{2.52}$$

and t_r^a are the generators in representation r. Diagrammatically, the anticommutator appears due to addition of the crossed diagram.

The statement that a gauge theory is "anomaly free" refers to the vanishing of the anomaly coefficient $\mathcal{A}^{abc}$. This is automatic for real and pseudoreal representations, where r is related to $\bar{r}$ by a unitary transformation, so $t_{\bar{r}}^a = -t_r^{a*} = -t_r^{aT} = U t_r^a U^\dagger$. Then since $\mathcal{A}^{abc}$ is invariant under unitary transformations we can write $\mathcal{A}^{abc} = \mathrm{Tr}(t_{\bar{r}}^a\{t_{\bar{r}}^b, t_{\bar{r}}^c\}) = -\mathrm{Tr}(t_r^{aT}\{t_r^{bT}, t_r^{cT}\}) = -\mathrm{Tr}(\{t_r^c, t_r^b\}t_r^a) = -\mathcal{A}^{abc}$, so $\mathcal{A}^{abc} = 0$.

A "vectorlike" gauge theory is one in which all irreps appear together with their conjugates – all fermions are Dirac – and so the total representation r is real. Therefore all the gauge anomalies cancel in vectorlike gauge theories. Likewise they cancel in theories where the fermionic matter is in a real irrep like the adjoint.

It is also interesting to consider chiral gauge theories, where anomaly cancellation demands a more intricate matter content. The standard example of an anomaly free chiral gauge theory is an $SU(N)$ gauge theory with one left-handed Weyl fermion in the symmetric (antisymmetric) two-index tensor representation and $N+4$ ($N-4$) left-handed Weyl fermions.[3]

2.4.2 ABJ anomalies

Here A is not path integrated over – J is a global current for symmetry S_A – but B is path integrated over, so symmetry S_B is gauged. This means J is not actually conserved, and correlation functions of $\partial_\mu J^\mu$ do not vanish, if there are any configurations of B such that $\mathrm{Tr}(F\tilde{F}) \neq 0$. We will discuss examples of such fields later. The basic result to remember is that the symmetry associated with A has been explicitly broken by the gauging of B.

Let the generators associated with the fermion representation R under S_A be T_R^i, and the generators associated the fermion representation with S_B be t_r^a. Then the anomaly equation is

$$\begin{aligned} \partial_\mu J^{\mu i} &= \frac{g^2}{16\pi^2}\mathcal{A}^{ibc} F^{\mu\nu b}\tilde{F}^c_{\mu\nu} \\ \mathcal{A}^{ibc} &= \mathrm{Tr}(T_R^i\{t_r^b, t_r^c\}) \end{aligned} \tag{2.53}$$

This goofy but standard notation for $\mathcal{A}$ has an implied tensor product, so it is really

$$\mathcal{A}^{ibc} = 2\mathrm{Tr}(T_R^i)\mathrm{Tr}(t_r^b t_r^c) = 2\mathrm{Tr}(T_R^i)\mathcal{I}(r)\delta^{bc} \tag{2.54}$$

[3] In $SU(N)$, the generators in the fundamental representation have anticommutator $\{t^a, t^b\} = \frac{1}{N}\delta^{ab} + d^{abc}t^c$. It turns out d^{abc} is the unique symmetric invariant tensor, so that $\mathcal{A}^{abc} = \frac{1}{2}A(r)d^{abc}$. Here $A(r)$ is called the anomaly coefficient of the representation r and it can be looked up in tables. The 1/2 is a conventional normalization. $A(r) = \pm 1$ for the fundamental/antifundamental and $A(r) = N \pm 4$ for the symmetric (+) and antisymmetric (−) two-index tensor irreps.

where $\mathcal{I}(r)$ is called the Dynkin index, and can be looked up in tables. (It is fairly but not completely standard to take the normalization $\mathcal{I}(r) = 1/2$ for the fundamental representation; then the adjoint representation for $SU(N)$ has $\mathcal{I}(A) = N$, etc.) The combination

$$2\mathrm{Tr}(T^i_R)\mathcal{I}(r) \tag{2.55}$$

is called the anomaly coefficient. If S_A is some global chiral $SU(n)$ symmetry, then the generators are traceless, and $\mathcal{A}^{ibc} = 0$.

More commonly S_A is a global chiral abelian symmetry, and so $\mathrm{Tr}(T^i_R)$ sums over the charges. For an abelian chiral symmetry (S_A) acting on N_f Dirac flavors in the fundamental representation of an $SU(N)$ gauge symmetry (S_B), $2\mathrm{Tr}(T^i_R)\mathcal{I}(r) = \mathrm{Tr}(T^i_R) = qN_f$, where q is the charge under S_A. So in this case

$$\partial_\mu J^\mu = qN_f \frac{g^2}{16\pi^2} F^{\mu\nu b}\tilde{F}^b_{\mu\nu} \tag{2.56}$$

For an abelian chiral symmetry (S_A) acting on N_f Dirac flavors with charge Q under an abelian gauge symmetry (S_B), the Dynkin index is replaced by Q^2. So in this case

$$\partial_\mu J^\mu = qQ^2 N_f \frac{g^2}{8\pi^2} F^{\mu\nu}\tilde{F}_{\mu\nu} \tag{2.57}$$

Finally, sometimes it is convenient to work with Majorana or Weyl fermions, treating them as separate flavors. In this case the anomaly coefficient is defined as $\mathrm{Tr}(T^i_R)\mathcal{I}(r)$, and the trace runs over all Majorana or Weyl species.

2.4.3 't Hooft anomalies

Here neither A nor B is path integrated over. If we set the background gauge fields to zero, we find the A-current is conserved, $\partial_\mu J^\mu$. So 't Hooft anomalies do not indicate a breaking of any symmetry. Instead, they are a contribution to the correlation function:

$$\langle (\partial_\mu J^\mu_A)(x) J^\nu_B(y) J^\rho_B(z)\rangle. \tag{2.58}$$

For separated x, y, z this vanishes, but in general it can have contact terms when y and/or z are coincident with x. This doesn't indicate a breaking of the symmetries, because at coincident points we no longer have the $\partial_\mu J^\mu$ operator in the correlator, we have some composite operator.

't Hooft anomalies are related to ABJ anomalies by the gauging of B. If we gauge B, or if we simply couple the system to background B fields, the 't Hooft anomaly becomes Eq. (2.53). We can define the 't Hooft anomaly coefficient for two different symmetries S_A and S_B by Eq. (2.54),(2.55). Similarly we can define the 't Hooft anomaly coefficient for a single symmetry S_A (with itself) by Eq. (2.52).

"'t Hooft anomaly matching" is a powerful constraint on EFTs. It says that 't Hooft anomalies have to be present at all scales. 't Hooft's argument is simple: if you have an anomaly in the form of a contact term contribution to (2.58) in the UV, you can weakly gauge the symmetries by coupling them with some couplings g to gauge fields, and add light spectator fermions to make the new gauge symmetries suitably anomaly free. Gauge anomaly cancellation must continue to hold at lower scales for unitarity of the EFT. But the contributions of the spectator fermions to the anomaly are independent from the contribution from the original theory – just the triangle diagram – at least as long as the gauge couplings of the symmetries remain weak. Then we may say that even if the original degrees of freedom have rearranged themselves (e.g. by spontaneous symmetry breaking or some strong dynamics), the EFT *without* the spectators and *vanishing* g has the same 't Hooft anomaly.

Note that this argument does not require arbitrarily weak couplings for the symmetries participating in the anomaly. It really just requires that the spectator fermions remain good, light degrees of freedom – they shouldn't get bound up into massive composites by some confining dynamics, for example. So we don't have to take $g \to 0$ at the end, and the argument applies even for ABJ-type anomalies, where some of the gauge fields are dynamical.

2.4.4 Example: Peccei-Quinn symmetry

Consider the model

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}(G^a_{\mu\nu}G^{\mu\nu a}) + |\partial_\mu\phi|^2 + iq^\dagger \not{D} q + i\bar{q}^\dagger \not{D}\bar{q} - (y(\phi)\bar{q}q + cc) - V(|\phi|). \tag{2.59}$$

The model contains an abelian gauge theory ("QED") with field strength F and a nonabelian gauge theory ("QCD") with field strength G. q, $\bar{q}$ are left-handed Weyl fermions carrying charge $\pm Q$ under QED and transforming in complex representations r, $\bar{r}$ of QCD. $\not{D}q = i\bar{\sigma}^\mu(\partial_\mu + ieQA_\mu + ig(T_r)^a_{ij}A^a_\mu$ and $\not{D}\bar{q} = i\bar{\sigma}^\mu(\partial_\mu - ieQA_\mu + ig(T_{\bar{r}})^a_{ij}A^a_\mu$. ϕ is a complex scalar field. There is a chiral "Peccei-Quinn" (PQ) symmetry,

$$q \to e^{i\alpha}q \quad \bar{q} \to e^{i\alpha}\bar{q} \quad \phi \to e^{-2i\alpha}\phi. \tag{2.60}$$

This symmetry has an ABJ anomaly with both QED and QCD:

$$\partial_\mu J^\mu_{PQ} = c\frac{e^2}{8\pi^2}F_{\mu\nu}\tilde{F}^{\mu\nu} + \frac{g^2}{16\pi^2}G^a_{\mu\nu}\tilde{G}^{\mu\nu a} \tag{2.61}$$

where the QED anomaly coefficient is

$$c = q\,\dim(r) \tag{2.62}$$

Now let us give ϕ has a Higgs-like potential, e.g. $V(\phi) = \frac{1}{4}\lambda\left(|\phi|^2 - f^2\right)^2$, so that the chiral symmetry is spontaneously broken with scale f. The fermions and the scalar radial mode get large masses from $\langle\phi\rangle$,

$$m_Q = yf, \quad m^2_\phi = \lambda f^2 \tag{2.63}$$

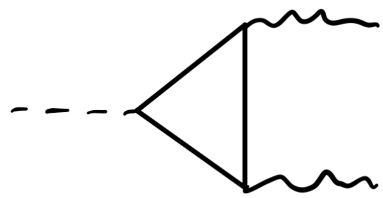
We can integrate them out and write an EFT for the low energy degree of freedom,

$$\phi \to fe^{ia(x)/f}, \tag{2.64}$$

where $a(x)$ is an NBG.

If $g(f)$ and $e(f)$ are both small, then we can treat the gauge fields as backgrounds when we integrate out the matter fields. Then both QED and QCD are weakly coupled at the matching scale, and we need to match the anomalies. How does it happen?

In general, anomaly matching requires light degrees of freedom in the EFT. Massless composite fermions could do the trick, via similar triangle diagrams, but we don't have any of those in this weakly-coupled theory. Instead, we have a Goldstone boson. In fact, it matches the anomaly through a dimension-5 effective coupling, generated by the diagram:



Including the effective couplings through dimension 5, the EFT is

$$\mathcal{L}_{eff} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu a} + (\partial_\mu a)^2 - c\frac{e^2}{8\pi^2 f}aF_{\mu\nu}\tilde{F}^{\mu\nu} - \frac{g^2}{16\pi^2 f}aG^a_{\mu\nu}\tilde{G}^{\mu\nu a} \tag{2.65}$$

In the UV theory, the PQ symmetry acts on ϕ as $\phi \to e^{-2i\alpha}\phi$, which is the same as saying it acts as a shift symmetry on the Goldstone, $a/f \to a/f - 2\alpha$. In the EFT, the classical current associated with the shift symmetry is

$$J^\mu = -2f\partial_\mu a. \tag{2.66}$$

The anomalous divergence is

$$\partial_\mu J^\mu = -2f\Box a = c\frac{e^2}{8\pi^2 f}aF_{\mu\nu}\tilde{F}^{\mu\nu} - \frac{g^2}{16\pi^2 f}aG^a_{\mu\nu}\tilde{G}^{\mu\nu a}. \tag{2.67}$$

So we see that the EFT has correctly matched the anomaly through the Goldstone couplings. We could also have read off $\partial_\mu J^\mu$ directly from the shift of the dimension-5 terms under $a \to a - 2\alpha f$, using the relationship between the divergence and the shift of the effective action under the symmetry transformation, Eq. (2.49).

Further comments:

- Fermions only contribute to an anomaly when their mass term is forbidden by the symmetry. Otherwise Pauli Villars can be used to regulate without breaking the symmetry. That is why anomalies are associated with chiral symmetries. Note, however, that the chiral symmetry can still be *explicitly* broken by a mass term, and the fermions can contribute to the anomaly in the approximate symmetry.
- Chiral anomalies are not the only type of anomalies and the spectator fermion technique is not the only way to cancel gauge anomalies in 't Hooft's argument. Another useful technique is allowing one's spacetime to be a boundary or codimension two surface in a higher dimensional spacetime. We will discuss this later.

- The term “trace anomaly” can refer to either an ABJ or a ‘t Hooft anomaly, of different types. The variation of the action with respect to the metric is the (Hilbert) stress tensor, and so the variation with respect to the conformal factor of the metric is the trace of the stress tensor. In this way the trace of the stress tensor is found to be the divergence of the dilatation current. The trace of the stress tensor in a classically scale-invariant QFT in flat space can contain (1) dimension-4 combinations of background gauge fields, which turn out to be of the form $a(Weyl)+c(Euler)$, and (2) dimension-4 combinations of dynamical gauge fields (like F^2, e.g. the operator associated with the Yang Mills beta function.) The former is a ‘t Hooft anomaly where the background gauge fields are gravitational. The latter is an ABJ anomaly.

2.5 The CCWZ prescription

Coleman, Callan, Wess, and Zumino gave a prescription for how to parametrize the fields in a theory with spontaneous symmetry breaking in order to separate out the NGB degrees of freedom from the rest. This will be a little abstract, but the technique is very useful, and we will see a concrete example of it later when we study chiral symmetry breaking and pion physics in QCD. It is also a vehicle to introduce gauge symmetry, which we will also study in future lectures.

Let the fields be ϕ^i. Let the path integral action and measure be invariant under a Lie group G that is spontaneously broken to a subgroup H by the vacuum. Let T^a be a set of generators of H (“preserved generators”) and X^b be a set generators mutually orthogonal to the T^a that together with T^a generate G (“broken generators.”) Then CCWZ showed that the π^i can be parametrized as

$$\phi^i = \Sigma^i_j(x)\varphi^j(x) \ , \qquad \Sigma(x) \equiv e^{i\pi\cdot X} \ , \tag{2.68}$$

where $\pi(x)$ are the NG fields (note that there is one for each broken generator X), and φ transforms linearly under H, $\varphi \to D(h)\varphi$ if ϕ is in representation D. There is some freedom in the decomposition (2.68), and it is convenient to choose φ so that

$$\varphi^\dagger_i (X^a)^i_j \langle \phi^j \rangle = 0 \ . \tag{2.69}$$

This condition just says φ is orthogonal to the broken directions of field space, or the vacuum manifold, in which our NG fluctuations will lie.

It is useful to know how the fields in the new basis transform under the original global symmetry group G. First consider a transformation by $h \in H$, $\phi \to D(h)\phi$. Then

$$\Sigma\varphi \to D(h)\Sigma D(h^{-1})D(h)\varphi = \Sigma'\varphi' \ , \tag{2.70}$$

where

$$\Sigma' = D(h)\Sigma D(h^{-1}) \ . \tag{2.71}$$

Therefore, if ϕ transforms in the fundamental representation of G, Σ transforms in the adjoint of H. Let us verify that Σ' may still be written as $e^{i\pi'\cdot X}$, for some transformed π fields. Using the properties of the adjoint and the exponential maps, we can write

$$\Sigma' = e^{i[D(h)X^a D(h)^{-1}]\pi^a} \ . \tag{2.72}$$

Furthermore the generators of G transform in the adjoint representation of G, so we can write

$$D(h)X^a D(h)^{-1} = M^a_b(h)X^b + N^a_b(h)T^b \tag{2.73}$$

for some M and N. But since the T generators transform in the adjoint representation of H, and the T and X can be taken to be mutually orthogonal,

$$N \sim \mathrm{Tr}[T^b D(h)X^a D(h)^{-1}] = (R_{adj})^b_c(h)\mathrm{Tr}[T^c X^a] = 0\ . \tag{2.74}$$

Thus $M = (R_{adj})$ and the π fields transform in the adjoint representation of H.

Now let us work out the transformations under a more general global transformation $g \in G$, which can be written as $g = e^{i\alpha^a X^a} e^{i\beta^b T^b}$ for elements of g continuously connected to the identity. Then

$$\Sigma\varphi \to g\Sigma\varphi = e^{i\alpha^a X^a} e^{i\beta^b T^b} e^{i\pi^c(x)X^c}\varphi\ , \tag{2.75}$$

and since $e^{i\alpha^a X^a} e^{i\beta^b T^b} e^{i\pi^c(x)X^c} \in G$, it can be rewritten

$$\begin{aligned} g\Sigma\varphi &= e^{i(\pi')^c(x)X^c} e^{i\gamma^b T^b}\varphi \\ &\equiv \Sigma'\varphi'\ . \end{aligned} \tag{2.76}$$

Here $\gamma = \gamma(g, \Sigma)$ specifies an element of H that should be inserted so that $\Sigma' = g\Sigma h(\gamma(g, \Sigma))^{-1}$ can still be written as $\Sigma' = e^{i\pi'\cdot X}$ for some transformed π fields. In general γ is nontrivial.

Thus far we have given a general discussion of the fields and their transformation laws according to the CCWZ prescription. Knowing the transformation laws, we can write down the most general G-invariant interactions, in the spirit of Weinberg's folk theorem. However, our analysis so far has been a bit abstract. Moreover in building an EFT the φ degrees of freedom are usually hierarchically separated from the NGBs, so we would like to keep just the Σ part fluctuating and set φ to its vev. Let's look at a concrete example.

2.5.1 Example: $O(N)$ model

We can rewrite the $O(N)$ model lagrangian in the broken phase as

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi^i\partial^\mu\phi^i - \lambda(\phi^i\phi^i - v^2)^2\ . \tag{2.77}$$

This is an example of what is called for historical reasons a Linear Sigma Model (LSM), because it exhibits invariance under a linearly realized group of transformations $G = O(N)$, and G is partly spontaneously broken,

$$\langle\phi^i\phi^i\rangle = -\frac{m^2}{2\lambda} \equiv \frac{v^2}{2}\ . \tag{2.78}$$

Up to an $O(N)$ rotation, we can take the vacuum to be

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ v \end{pmatrix}\ . \tag{2.79}$$

This vacuum is left invariant by transformations of the form

$$h = \begin{pmatrix} h' & 0 \\ 0 & 1 \end{pmatrix} \tag{2.80}$$

with $h' \in O(N-1)$. Thus the unbroken symmetry is $H = O(N-1)$ and the pattern of symmetry breaking is $G/H = O(N)/O(N-1) \simeq S^{N-1}$. There are $\frac{N(N-1)}{2} - \frac{(N-1)(N-2)}{2} = N-1$ broken generators and massless NGBs, leaving one additional non-NGB degree of freedom, corresponding to radial fluctuations. The CCWZ prescription is

$$\phi^i = \Sigma\varphi^i(x) \;, \qquad \Sigma = e^{i\pi^a(x)L^a} \;, \qquad \varphi^i(x) = (\rho(x)+v)) \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix} \;. \tag{2.81}$$

Inserting into the Lagrangian,

$$\mathcal{L} = \frac{1}{2}\partial_\mu\rho\partial^\mu\rho - \lambda(\rho^2+2\rho v)^2 + \frac{1}{2}(\rho+v)^2\left[\partial_\mu e^{-i\pi^a(x)L^a}\partial^\mu e^{i\pi^b(x)L^b}\right] \;. \tag{2.82}$$

(The minus sign in the exponent arises from taking the transpose of L, which is skew-symmetric.) As in the $O(2)$ example studied previously, the radial fluctuation is a massive field, $m_\rho^2 = 8\lambda v^2$. Let us therefore omit it and study only the massless NGB degrees of freedom. This is called the Nonlinear Sigma Model, or NLSM, because compared to the LSM in (2.77) the G-invariance will be nonlinearly realized through complicated transformations of the NGBs.

For $N=3$, we can visualize the vacuum manifold $\langle\phi^i\phi^i\rangle = v^2$ as the 2-sphere, so this will be the target space of the Σ fields. Choosing the vacuum to be $\langle\phi^i\rangle = (0,0,v)$, only the generator

$$L_3 = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.83}$$

of rotations about the 3 axis (in field space) is unbroken, $(L_3)_i^j\langle\phi^i\rangle = 0$ and $H = U(1)$. The X generators are thus L_1 and L_2, and

$$\Sigma = e^{i[L_1\pi_1(x)+L_2\pi_2(x)]} \;. \tag{2.84}$$

The adjoint representation of $U(1)$ is trivial, so the π fields do not transform under H. Under a general $O(3)$ transformation g, $\phi \to g\phi$. $g\Sigma$ does not in general have the form (2.84) for some $\pi'_{1,2}$, but can be written as

$$g\Sigma = \Sigma' h \tag{2.85}$$

where Σ' does have the form (2.84) and $h = h(g,\Sigma)$. This can be visualized by the parallel transport of a vector along the surface of a sphere. The object $g\Sigma$ transports the vector first from the north pole to a point A and then to a second point B. The result differs from the parallel transport of the vector directly from the north pole to B, because of curvature (this is curvature in *field space*), but they can be reconciled by first rotating the vector around the north pole axis (this is h in Eq. 2.85) and then parallel transporting it to B. Thus we take the transformation law to be

$$\begin{aligned}\Sigma\langle\phi\rangle &\to \left(g\Sigma h^{-1}(g,\Sigma)\right) h(g,\Sigma)\langle\phi\rangle \\ &= \left(g\Sigma h^{-1}(g,\Sigma)\right)\langle\phi\rangle \\ &= \Sigma'\langle\phi\rangle \ . \end{aligned} \tag{2.86}$$

In the $O(N)$ model, Σ is an orthogonal matrix, so $\Sigma^T\Sigma = \mathbf{1}$ and the only G-invariant operators must involve derivatives of Σ. The simplest is the 2-derivative operator,

$$\mathcal{L}_{eff}(\Sigma) = \frac{f^2}{2}\text{Tr}\left(\partial_\mu\Sigma^T\partial^\mu\Sigma\right) \ . \tag{2.87}$$

Comparing with the last term in Eq. (2.82), which is the only one that survives after we set $\rho \to \langle\rho\rangle = 0$, we see that the "UV completion" generates precisely this effective Lagrangian with $f = v$.

2.5.2 Spurions

So far we have considered only exact symmetries, but it is extremely useful to be able to accommodate small amounts of symmetry breaking into the preceding discussion. We might imagine that if a global symmetry were broken by a term in the action with a small coefficient, the NGBs would no longer be massless (thus "pseudo-NGBs", or PNGBs), but they might still be light compared to the rest of the DOF. Spurion analysis encodes small symmetry breaking in a systematic way.

Take an action $S_0[\phi]$ symmetric under a transformation $\phi \to D(g)\phi$ and perturb it by a small symmetry-violating term,

$$\Delta S = \int d^4x\lambda^a\mathcal{O}^a[\phi] \ . \tag{2.88}$$

Here $\mathcal{O}^a[\phi]$ is a functional of ϕ that transforms as $\mathcal{O}^a \to \tilde{D}^a_b\mathcal{O}^a$ under $\phi \to D(g)\phi$, and the λ^a are a set of small coupling constants. However, if we asserted that $\lambda \to \tilde{D}^{-1}\lambda$ under the transformation, then ΔS would be invariant. Now, this is only a trick: there is no unitary operator acting on the Hilbert space that "transforms" the coupling constant λ. However, this observation does show that whatever we compute in this theory has to be symmetric under the expanded transformation. λ is called a "spurion" for the broken symmetry, and if it is small, we can reliably compute results as λ-perturbations around the symmetric theory with $\lambda = 0$.

As an example, take the $O(2) \simeq U(1)$ model with a new type of "holomorphic" mass term:

$$S = \int |\partial_\mu\phi|^2 - \lambda(|\phi|^2 - v^2)^2 - \{\mu^2\phi^2 + c.c.\} \ . \tag{2.89}$$

If $\mu \to 0$, this action is invariant under $\phi \to e^{i\alpha}\phi$. Furthermore, if we promote μ^2 to a spurion transforming as $\mu^2 \to e^{-2i\alpha}\mu^2$, the whole action is invariant. Then we predict that

$$\begin{aligned}\langle\phi^*\phi^3\rangle &\propto \mu^{*2} \\ \langle\phi^*\phi^5\rangle &\propto \mu^{*4}\end{aligned} \tag{2.90}$$

etc.

2.6 QCD and the Chiral Lagrangian

QCD is described by the Lagrangian

$$\mathcal{L}_{QCD} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} + \bar{\psi}_i i \not{D} \psi_i - m_{ij} \bar{\psi}_i \psi_j \tag{2.91}$$

The 1-loop RGE for the gauge coupling is

$$\frac{\partial \alpha^{-1}}{\partial \log \mu} = \frac{b_0}{2\pi}, \qquad b_0 = \frac{11}{3} N - \frac{2}{3} N_f \tag{2.92}$$

where N_f is the number of Dirac fermions in the fundamental representation. The Standard Model has six quarks, so it is appropriate to take $N_f = 6$ at scales above the top quark mass. Then $b_0 = 11 - 4 = 7$, and α^{-1} decreases linearly as $\log \mu$ is lowered. At the top ($m_t \simeq 172\text{GeV}$), bottom ($m_b \simeq 4.5\text{GeV}$), and charm ($m_c \simeq 1.2\text{GeV}$) thresholds, we decouple each of these heavy quarks. To first approximation, this just means that below each threshold we drop the quark from the theory and decrement the beta function by $N_f \to N_f - 1$. Finally, around a few hundred MeV (well above the strange quark mass), the coupling runs strong, and perturbative QCD, as a theory of weakly coupled quarks and gluons, is no longer a good description of the physics. Empirically, we observe that strong dynamics at lower energies should be a theory of interacting mesons and baryons, but we do not know how to analytically integrate out degrees of freedom from QCD to produce a meson-baryon EFT.

But all is not lost! We can do bottom-up EFT thanks to some further theoretical and empirical observations:

1. For $m_{ij} = 0$, $\mathcal{L}_{QCD}$ has an $SU(N_f)_L \times SU(N_f)_R$ "flavor symmetry." There are $N_f = 3$ light quarks for which $m = 0$ might be a reasonable first approximation. Let ψ_i be a Dirac quark with flavor index i. It will be convenient to swap back and forth between two-component and four-component fermion notation in what follows. The flavor symmetry acts on the two-component quark fields as

$$\psi_i = \begin{pmatrix} \psi_{Li} \\ \psi_{Ri} \end{pmatrix} \longrightarrow \begin{pmatrix} \tilde{L}_{ij} & 0 \\ 0 & \tilde{R}_{ij} \end{pmatrix} \begin{pmatrix} \psi_{Lj} \\ \psi_{Rj} \end{pmatrix} \tag{2.93}$$

   and

$$\bar{\psi}_i = \begin{pmatrix} \psi^*_{Ri} & \psi^*_{Li} \end{pmatrix} \longrightarrow \begin{pmatrix} \psi^*_{Rj} & \psi^*_{Lj} \end{pmatrix} \begin{pmatrix} \tilde{R}^\dagger_{ji} & 0 \\ 0 & \tilde{L}^\dagger_{ji} \end{pmatrix} \tag{2.94}$$

   where $\tilde{L}$ and $\tilde{R}$ are $N_f \times N_f = 3 \times 3$ special unitary matrices acting on flavor indices. We can also write them as

$$\tilde{L} = e^{i\alpha^a_L T^a} \quad \tilde{R} = e^{i\alpha^a_R T^a} \tag{2.95}$$

where the T^a are traceless Hermitian $SU(N_f = 3)$ generators. In the following we will just set $N_f = 3$ and it will be clear from context how to generalize it to other choices of N_f. In terms of four-component quarks, we can also define

$$L \equiv \begin{pmatrix} \tilde{L} & 0 \\ 0 & 1_{3x3} \end{pmatrix}, \quad R \equiv \begin{pmatrix} 1_{3x3} & 0 \\ 0 & \tilde{R} \end{pmatrix} \tag{2.96}$$

which can be written in terms of traceless Hermitian generators as

$$L = e^{i\alpha_L^a T_L^a} \quad R = e^{i\alpha_R^a T_L^a} \tag{2.97}$$

where

$$T_L^a = \begin{pmatrix} T^a & 0 \\ 0 & 0 \end{pmatrix} \quad T_R^a = \begin{pmatrix} 0 & 0 \\ 0 & T^a \end{pmatrix}. \tag{2.98}$$

T_L^a and T_R^a span the $2(3^2-1) = 16$-dimensional Lie algebra; a general element is $\alpha_L^a T_L^a + \alpha_R^a T_R^a$. Then under $SU(3)_L \times SU(3)_R$, the four-component quark transforms as

$$\psi \to LR\psi. \tag{2.99}$$

2. There is a vacuum expectation value for a fermion bilinear which, in some basis, can be written $\langle \bar{\psi}\psi_j \rangle \sim \Lambda_{\text{QCD}}^3 \, \delta_{ij}$. Historically, this is putting the cart before the horse, but is easier to take this as a given.

   How does $\bar{\psi}_i \psi_j$ transform under flavor?

$$\bar{\psi}_i \psi_j = \psi_{Li}^\dagger \psi_{Rj} + cc \to \psi_{Lk}^\dagger \tilde{L}_{ki}^\dagger \tilde{R}_{jl} \psi_{Rl} + cc \tag{2.100}$$

   so in the special case where $\tilde{L} = \tilde{R}$,

$$\langle \bar{\psi}_i \psi_j \rangle \to \langle \bar{\psi}_i \psi_j \rangle \tag{2.101}$$

   because δ_{ij} is an invariant tensor. From this we conclude that transformations with $\alpha_L^a = \alpha_R^a$ are unbroken by the vacuum state. It is convenient to change the basis of generators so that some are aligned with the unbroken subgroup. In four-component notation, we write

$$\alpha_L^a T_L^a + \alpha_R^b T_R^b = \alpha_V^a \underbrace{(T_L^a + T_R^a)}_{\equiv T_V^a} + \beta^a X^a. \tag{2.102}$$

   There is freedom in the choice of the X generators: they cannot be degenerate with T_V, but they don't have to be orthogonal.

   The T_V^a generate the unbroken "vector" symmetries (vectors here means the transformations treat L and R fermion components the same.) The X^a generate the symmetries that are spontaneously broken by $\langle \bar{\psi}\psi \rangle$.

   Common choices for the X^a:

   - $X^a = T_L^a - T_R^a$, "axial" $SU(3)_A$. Here $\alpha_L = \alpha_V + \beta$ and $\alpha_R = \alpha_V - \beta$, or equivalently $\alpha_V = (\alpha_L + \alpha_R)/2$ and $\beta = (\alpha_L - \alpha_R)/2$.
   - $X^a = T_L^a$. Here $\alpha_L = \alpha_V + \beta$ and $\alpha_R = \alpha_V$, or equivalently $\alpha_V = \alpha_R$ and $\beta = \alpha_L - \alpha_R$.

2.6.1 CCWZ and the light pseudoscalar mesons

Now we proceed with the CCWZ construction of an EFT for the Goldstone bosons. We take the light fields to be local fluctuations in the directions of the vacuum manifold, parametrized by the broken generators acting on a particular (arbitrary) vacuum state. We take $\left\langle \psi_{Li}\psi^\dagger_{Rj}\right\rangle \propto \delta_{ij}$ as the particular state.

Then, as we showed above, the vev transforms under the flavor symmetries as

$$\left\langle \psi_i\psi^\dagger_{R_j}\right\rangle \to \tilde{L}_{il}\left\langle \psi_{Ll}\psi^\dagger_{Rk}\right\rangle \tilde{R}_{kj} \propto \left(\tilde{L}\tilde{R}^\dagger\right)_{ij}. \tag{2.103}$$

Now restrict to the broken transformations. We'll start with the second basis. In this basis, in two-component language, the broken transformations are parameterized by $\alpha_L = \alpha_V + \beta \to \beta,\ \ \alpha_R = \alpha_V \to 0$. So $\tilde{R} \Rightarrow 1_{3x3}$ and we write $\tilde{L}$ in a conventional form,

$$\tilde{L} = e^{2i\pi^a T^a/f_\pi}. \tag{2.104}$$

Here the 2 is conventional, f_π is a new scale to be determined, and we have replaced β^a by $\pi^a(x)$, a Goldstone field of mass dimension one. We define a matrix field

$$\Sigma(x) \equiv e^{2i\pi^a T^a/f_\pi} \tag{2.105}$$

which is valued in the $SU(3)_L \times SU(3)_R/SU(3)_V \simeq SU(3)$ coset. In four-component notation,

$$U \equiv e^{i\pi^a X^a/f_\pi} \equiv \begin{pmatrix} \Sigma(x) & 0 \\ 0 & 1_{3\times 3}\end{pmatrix} \tag{2.106}$$

and it is convenient to define also check

$$\bar{U} \equiv \begin{pmatrix} 1_{3\times 3} & 0 \\ 0 & \Sigma(x)^\dagger\end{pmatrix} \tag{2.107}$$

How does Σ transform under the flavor symmetries? In four-component language, $U\langle\psi\bar{\psi}\rangle\bar{U}$ is a state on the vacuum manifold. Under a flavor rotation,

$$U\langle\psi\bar{\psi}\rangle\bar{U} \to (LR)U\langle\psi\bar{\psi}\rangle\bar{U}(\overline{LR}) \tag{2.108}$$

where $\overline{LR} = \begin{pmatrix} \tilde{R}^\dagger & 0 \\ 0 & \tilde{L}^\dagger\end{pmatrix}$. Perhaps

$$U' \stackrel{?}{=} LRU = \begin{pmatrix} \tilde{L} & 0 \\ 0 & \tilde{R}\end{pmatrix}\begin{pmatrix} \Sigma & 0 \\ 0 & 1\end{pmatrix} = \begin{pmatrix} \tilde{L}\Sigma & 0 \\ 0 & \tilde{R}\end{pmatrix}. \tag{2.109}$$

But this is not in the form $\begin{pmatrix} \Sigma' & 0 \\ 0 & 1\end{pmatrix}$, so we can't read off a transformed Σ' from it. To fix this issue, note that we can insert an element D of the unbroken $SU(3)_V$, since $D\langle\psi\bar{\psi}\rangle\bar{D} = \langle\psi\bar{\psi}\rangle$. So let's try

$$U' \stackrel{?}{=} LRUD = \begin{pmatrix} \tilde{L}\Sigma D & 0 \\ 0 & \tilde{R}D\end{pmatrix} \tag{2.110}$$

So we see that if we set $D = \tilde{R}^\dagger$, we obtain

$$U' = \begin{pmatrix} \tilde{L}\Sigma\tilde{R}^\dagger & 0 \\ 0 & 1_{3\times 3} \end{pmatrix} \tag{2.111}$$

from which we read off the transformation law

$$\Sigma' = \tilde{L}\Sigma\tilde{R}^\dagger. \tag{2.112}$$

2.6.2 ξ-basis

What if we took the other basis?

$$\begin{aligned} &\tilde{R} \Rightarrow e^{-i\pi^a T^a/f_\pi} \equiv \xi^\dagger, \quad \tilde{L} \Rightarrow e^{i\pi^a T^a/f_\pi} \equiv \xi \\ &U = \begin{pmatrix} \xi & 0 \\ 0 & \xi^\dagger \end{pmatrix} \end{aligned} \tag{2.113}$$

As before, introducing an extra vector rotation allows us to assign a transformation rule to ξ. U transforms as

$$(LR)UD\langle\psi\bar{\psi}\rangle\bar{D}\bar{U}\overline{LR} \tag{2.114}$$

where

$$LRUD = \begin{pmatrix} \tilde{L} & 0 \\ 0 & \tilde{R} \end{pmatrix}\begin{pmatrix} \xi & 0 \\ 0 & \xi^\dagger \end{pmatrix}\begin{pmatrix} \tilde{D} & 0 \\ 0 & \tilde{D} \end{pmatrix} \tag{2.115}$$

So we require

$$\xi' = \tilde{L}\xi\tilde{D} = \tilde{D}^\dagger\xi\tilde{R}^\dagger \tag{2.116}$$

which we can regard as an implicit definition of $\tilde{D}$ in terms of $\tilde{L}, \tilde{R}, \xi$.

Then

$$\xi^2 \to \tilde{L}\xi\tilde{D}\tilde{D}^\dagger\xi\tilde{R}^\dagger = \tilde{L}\xi^2\tilde{R}^\dagger \tag{2.117}$$

so we may identify a map between the two bases:

$$\Sigma = \xi^2. \tag{2.118}$$

2.6.3 2-derivative Lagrangian

The π fields, $\pi^a T^a \equiv \pi$, describe the light mesons $\pi^\pm$ (139 MeV), π^0 (135 MeV), $K^0/\bar{K}^0$ (498 MeV), $K^\pm$ (494 MeV), and η (548 MeV) For fundamental Dynkin index $\mathrm{Tr}(T^a T^b) = 1/2\delta^{ab}$, the identification is

$$\pi = \frac{1}{\sqrt{2}}\begin{pmatrix} \pi/\sqrt{2} + \eta/\sqrt{6} & \pi^+ & K^+ \\ \pi- & -\pi^0/\sqrt{2} + \eta/\sqrt{6} & K^0 \\ K^- & \bar{K}^0 & -\frac{2}{\sqrt{6}}\eta \end{pmatrix} \tag{2.119}$$

The low energy $\mathcal{L}_{eff}$ is the most general Lagrangian describing the spontaneous symmetry breaking pattern $SU(3) \times SU(3) \to SU(3)$. Obviously the "Goldstones" aren't really massless – the symmetry is also explicitly broken by the light quark mass matrix.[4]

Most general term invariant under $\Sigma' = L\Sigma R^\dagger$ with no derivatives would be

$$\mathrm{Tr}\, \Sigma\Sigma^\dagger \ldots \Sigma\Sigma^\dagger \tag{2.120}$$

But $\Sigma\Sigma^\dagger = 1$, so these are all constants. We knew this - Goldstones are derivatively coupled.

The only 2 derivative term is:

$$\begin{aligned}\mathcal{L}_2 &= \frac{f_\pi^2}{4}\,\mathrm{Tr}\left(\partial_\mu \Sigma \partial^\mu \Sigma^\dagger\right) \\ &= \mathrm{Tr}\, \partial_\mu \pi \partial^\mu \pi + \frac{1}{3f_\pi^2}\,\mathrm{Tr}\,[\pi, \partial_\mu \pi]^2\end{aligned} \tag{2.121}$$

Since $\mathrm{Tr}\, T^a T^b = \frac{1}{2}\delta^{ab}$, the first term gives the canonically normalized kinetic terns

$$\frac{1}{2}\left(\partial_\mu \pi^a\right)^2 \tag{2.122}$$

All terms have an even number of π's - it is a pseudoscalar and parity is conserved.

$\mathcal{L}_2$ determines all scattering amplitudes to order p^2/f_π^2. $\pi - \pi$ scattering can be used to measure f_π, for example, although there are better ways.

2.6.4 Chiral currents

In response to a symmetry transformation of local infinitesimal parameter ϵ, Noether's theorem tells us the change in the action is of the form: $\delta\mathcal{L} = \partial_\mu j^\mu(x)$.

For $SU(3)_L$:

$$\begin{aligned}&\Sigma \to \Sigma + i\epsilon_L^a T^a \Sigma \\ &\Rightarrow \delta\mathcal{L} = \frac{i}{2} f_\pi^2 \partial_\mu \epsilon_L^a\,\mathrm{Tr}\left(T^a \Sigma \partial_\mu \Sigma^\dagger\right) \\ &\Rightarrow j_L^{\mu a} = \frac{i}{2} f_\pi^2\,\mathrm{Tr}\left(T^a \Sigma \partial^\mu \Sigma^\dagger\right)\end{aligned} \tag{2.123}$$

and similarly

$$j_R^{\mu a} = \frac{i}{2} f_\pi^2\,\mathrm{Tr}\left(T^a \Sigma^\dagger \partial^\mu \Sigma\right). \tag{2.124}$$

[4]We will see how to incorporate explicit symmetry breaking. Note that π^0, $\pi^\pm$ are much lighter than the other mesons. For this reason sometimes the theory is restricted to $SU(2)_L \times SU(2)_R / SU(2)_V$.

The spontaneously broken axial transformations have current

$$j_A^{\mu a} = j_R^{\mu a} - j_L^{\mu a} = -f_\pi \partial^\mu \pi^a + \dots \tag{2.125}$$

This is characteristic of spontaneously broken currents and implies the current can create a single Goldstone boson:

$$\langle 0 | j_A^{\mu a}(x) | \pi^b(p) \rangle = i f_\pi p^\mu \delta^{ab} e^{-ipx} \tag{2.126}$$

In nature, the $j^{\mu a}$ current is weakly gauged: $j_L^{\mu a} W_\mu^a$ which allows $\pi^+ \to \mu^+ \bar{\nu}$ weak decays $\propto \langle 0 | j^{\mu a} \left| \pi^b \right\rangle \propto f_\pi$. In practice this is how f_π is determined.

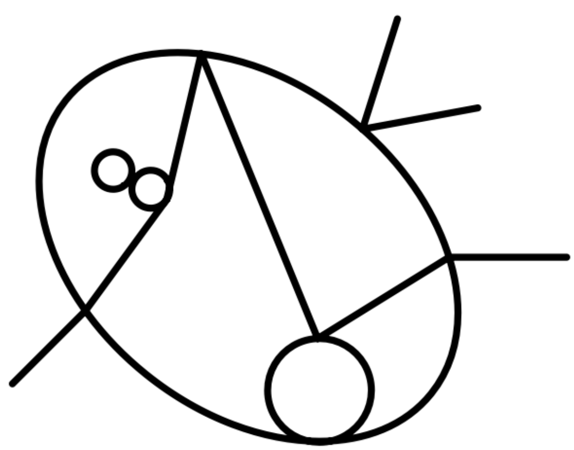
2.6.5 Power Counting

So far, we have discussed how to write the general $SU(3)_L \times SU(3)_R$-invariant $\mathcal{L}_{eff}$ for the NGBs of $SU(3)_L \times SU(3)_R / SU(3)_V$ SSB induced by $\langle \bar{\psi} \psi \rangle$ in QCD. This is called the chiral Lagrangian, because the SSB is of chiral symmetries, and describes the low-energy physics of the light mesons π, k, η. We saw the leading term in the derivative expansion was

$$\begin{aligned} \mathcal{L}_2 &= \frac{f_\pi^2}{4} \operatorname{Tr} \left(\partial_\mu \Sigma \partial^\mu \Sigma^\dagger \right) \\ &\simeq \underbrace{\operatorname{Tr} \partial_\mu \pi \partial^\mu \pi}_{\text{kinetic}} + \underbrace{\frac{1}{3 f_\pi^2} \operatorname{Tr} \left[\pi, \partial_\mu \pi \right]^2}_{\pi\pi \text{ scattering}} + \dots \end{aligned} \tag{2.127}$$

Next we will discuss power counting and naive dimensional analysis (NDA) for the chiral Lagrangian and try to work out the range of validity of chiral perturbation theory (ChPT).

The chiral Lagrangian is $\mathcal{L} = \sum_k L_k$ where k counts derivatives. Now consider a general complicated loop graph



The amplitude is some integral of the form

$$A = \int \left(d^4p\right)^L \left(\frac{1}{p^2}\right)^I \prod_k \left(p^k\right)^{M_k} \tag{2.128}$$

where

$L = \#$ loops
$I = \#$ internal lines
$m_k = \#$k-vertices (vertices containing k derivatives)

So the scaling dimension of the amplitude is

$$[A] = 4L - 2I + \sum_k k m_k \equiv D. \tag{2.129}$$

In a mass-independent renormalization scheme, on dimensional grounds, $A \sim p_{ext}^D \times f(p_{ext}/f_\pi$.

Now it can be shown that V(ertices)$-I$(nternal lines)$+L$(oops) $= 1$, where $V = \sum_k m_k$. So

$$D = 2 + 2L + \sum_k (k-2) m_k \tag{2.130}$$

eliminating I. L starts at order p^2, so all terms in the sum over k are ≥ 0. Therefore, to fixed order in p_{ext} (fixed D), we need only a finite number of the L_k.

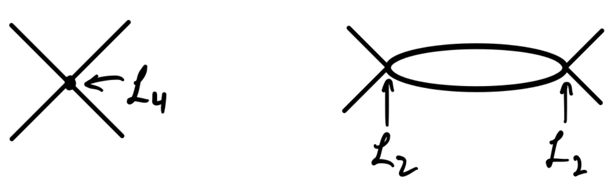
Example: to compute to $O(p_{ext}^4)$,

$$4 = 2 + 2L + \sum_k (k-2) m_k \tag{2.131}$$

At tree level ($L = 0$), we have $m_4 = 1$ and $m_{k>4} = 0$. At one loop ($L = 1$), $m_{k>2} = 0$. No contribution from higher loop. $\therefore$ To compute all scattering amplitudes to $\mathcal{O}(p_{ext}^4)$, we need only tree diagrams with vertices from $\mathcal{L} = \mathcal{L}_2 + \mathcal{L}_4$, and 1-loop diagrams with vertices from $L = L_2$. Acts like an ordinary renormalizable theory to fixed order in p_{ext}, as discussed previously.

2.6.6 Naive Dimensional Analysis

It turns out the p_{ext} expansion is somewhat better than p_{ext}/f_π. To get the idea, consider the $2 \to 2$ scattering to $O(p_{ext}^4)$.



The second diagram is

$$\int \frac{d^4k}{(2\pi)^4} \underbrace{\frac{k^2}{f_\pi^2}\frac{k^2}{f_\pi^2}}_{\text{vertices}} \underbrace{\frac{1}{k^2}\frac{1}{k^2}}_{\text{propagators}} \equiv \mathcal{I} \tag{2.132}$$

$$\left(\text{Recall}: \; L_2 = \mathrm{Tr}\,(\partial_\mu \pi)^2 + \underbrace{\frac{1}{3f_\pi^2}\mathrm{Tr}\,[\pi, \partial_\mu \pi]^2}_{\sim k^2/f_\pi^2 \text{ 4-point vertex}} + \ldots \right) \tag{2.133}$$

So

$$\mathcal{I} \simeq \frac{p^4}{16\pi^2}\frac{1}{f_\pi^4}\log\mu \qquad \text{in } DR + \overline{MS} \tag{2.134}$$

Meanwhile the tree level piece arises from $\mathcal{L}_4$.

$$\begin{aligned} L_4 \supset a\,\mathrm{Tr}\left[\partial_\mu \Sigma \partial^\mu \Sigma^\dagger \partial_\nu \Sigma \partial^\nu \Sigma^\dagger\right] \\ \to p_{ext}^4/f_\pi^4 \qquad \text{(4-pion vertices)} \end{aligned} \tag{2.135}$$

As usual, tree + loop is μ-independent because of implicit μ-dependence in the coupling a (RG).

$$= a\frac{p_{ext}^4}{f_\pi^4} + \frac{p_{ext}^4}{16\pi^2 f_\pi^4}\log\mu \tag{2.136}$$

So an $O(1)$ change in $\mu \Rightarrow \mathcal{O}\left(\frac{1}{16\pi^2}\right)$ change in a. Generically (meaning for a generic choice of μ, $|a| \geq |\delta a| \sim \frac{1}{16\pi\pi^2}$ barring accidental cancellations. Now write L as

$$\frac{f^2}{4}\left[\mathrm{Tr}\,\partial_\mu \Sigma \partial^\mu \Sigma^\dagger + \frac{1}{\Lambda_\chi^2}\mathcal{L}_4 + \frac{1}{\Lambda_\chi^4}\mathcal{L}_6 + \ldots\right] \tag{2.137}$$

Λ_χ is the typical scale of the effective Lagrangian, namely the derivative expansion is p/Λ_χ. Since $a \gtrsim \frac{1}{16\pi^2}, \Lambda_\chi \lesssim 4\pi f_\pi$, at least for the 2 -derivative term. In fact it is true for all terms, estimated in a similar way.

It turns out that in QCD, it is a very good rule to set $\Lambda_\chi \to 4\pi f_\pi$, saturating the inequality. Residual coefficients in $\mathcal{L}$ are mostly $O(1)$. Numerically, then, the cutoff is

$$4\pi f_\pi \sim 1\text{GeV}. \tag{2.138}$$

This ensures that ChPT works somewhat well for kaons ($M_k \sim 500$MeV) and very well for pions ($m_\pi \sim 135$MeV). If Λ_χ was $f_\pi \sim 100$MeV, ChPT would not even work for massive pions!

Naive Dimensional Analysis (NDA) is the statement that a generic term in $\mathcal{L}$ is of order

$$f_\pi^2 \Lambda_\chi^2 \left(\frac{\pi(x)}{f_\pi}\right)^n \left(\frac{\partial}{\Lambda_\chi}\right)^m \tag{2.139}$$

with n pions and m derivatives..

Examples:

$$f^2 \operatorname{Tr} \partial_\mu \Sigma \partial^\mu \Sigma^\dagger \xrightarrow[\substack{\text{kinetic}\\ \text{term}}]{} f^2 \partial^2 \frac{\pi^2}{f^2} \to f^2 \Lambda_\chi^2 \left(\frac{\partial}{\Lambda_\chi^2}\right)^2 \left(\frac{\pi}{f_\pi}\right)^2 \tag{2.140}$$

$$\operatorname{Tr} \partial_\mu \Sigma \partial^{mu} \Sigma^\dagger \partial_\nu \Sigma \partial^\nu \Sigma^\dagger \to f_\pi^2 \Lambda_\chi^2 \left(\frac{\partial}{\Lambda_\chi}\right)^4 \left(\frac{\pi}{f_\pi}\right)^4 \to \text{ c } \frac{f_\pi^2}{\Lambda_\chi^2} \sim 1/16\pi^2 \tag{2.141}$$

2.6.7 Explicit chiral symmetry breaking

As we have mentioned, the pions and kaons are massive. How can this be for Goldstone bosons?

The chiral symmetries are not exact - they are explicitly broken by quark masses

$$\begin{aligned} L &\supset -\psi_{Li}^\dagger M_{ij} \psi_{Rj} + c.c \\ M &= \begin{pmatrix} m_u & & \\ & m_d & \\ & & m_s \end{pmatrix} \end{aligned} \tag{2.142}$$

This is invariant if we also transform M as $M \to LMR^\dagger$. So $\mathcal{L}_\chi$ should also be invariant if we include M and impose this transformation. Then, when we treat M as a constant, the new terms encode the effects of symmetry breaking.

This is an example of spurion analysis, as discussed previously. Here M is a spurion for explicit χSB.

This is only useful if M is small, so that we can treat the $\mathcal{L}_x$ as an expansion in both $\partial/_\chi$ and m/Λ_χ.

To lowest order,

$$\mathcal{L}_m = \mu \frac{f_\pi^2}{2} \operatorname{Tr} \left(\Sigma^\dagger M + M^\dagger \Sigma\right). \tag{2.143}$$

This is invariant under the simultaneous chiral transformations of Σ and the spurion M. To understand μ better, note that this term can be thought of as taking $\mathcal{L}_{QCD}$ and setting $\bar{\psi}\psi$ to the condensate:

$$m_{ij}\psi^\dagger_{Li}\psi_{Rj} + cc \longrightarrow m_{ij} \underbrace{\tilde{L}^\dagger_{ki}}_{\Sigma^\dagger_{ki}} \underbrace{\langle\psi^\dagger_{Lk}\psi_{R\ell}\rangle}_{\delta_{k\ell}|\langle\bar\psi\psi\rangle|} \underbrace{\tilde{R}_{jl}}_{\delta_{j\ell}} + cc$$
$$\longrightarrow \mathrm{Tr}\left(M\Sigma^\dagger + cc\right)|\langle\bar\psi\psi\rangle|$$

so

$$\frac{\mu f_\pi^2}{2} = |\langle\bar\psi\psi\rangle|. \tag{2.144}$$

Non-derivative! So flat directions lifted. $\Sigma = $ const was the vacuum manifold before introducing M, now it is only approximately so. Taking M real,

$$\frac{\mu f_\pi^2}{2}\mathrm{Tr}\left(\Sigma^\dagger M + cc\right) \simeq \text{ const } - \frac{\mu f_\pi^2}{2}\frac{4\pi^a\pi^b}{f\pi^2}\mathrm{Tr}\left(MT^aT^b\right) + \ldots \tag{2.145}$$

Linear terms in π^a cancel $\Rightarrow \langle\pi\rangle = 0, \langle\Sigma\rangle = 1$.

Gell-Mann matrices:

$$\underbrace{T^1 = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad T^2 = \frac{1}{2}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad T^3 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{\text{SU(2) subgroup}}$$

$$T^4 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 \\ i & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad T^5 = \frac{1}{2}\begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \quad T^6 = \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$T^7 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad T^8 = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

$$TrT^AT^B = \frac{1}{2}\delta^{AB}$$

$$\begin{aligned} \pi^aT^a &= \frac{1}{2}\begin{pmatrix} \pi^3 + \pi^8/\sqrt{3} & \pi^1 - i\pi^2 & \pi^4 - i\pi^5 \\ \pi^1 + i\pi^2 & -\pi^3 + \pi^8/\sqrt{3} & \pi^6 - i\pi^7 \\ \pi^4 + i\pi^5 & \pi^6 + i\pi^7 & -2\pi^8/\sqrt{3} \end{pmatrix} \\ &\equiv \frac{1}{\sqrt{2}}\begin{pmatrix} \frac{\pi^0}{\sqrt{2}} + \eta/\sqrt{6} & \pi^+ & K^+ \\ \pi^- & -i\pi^0/\sqrt{2} + \eta/\sqrt{6} & K^0 \\ k^- & \bar{K}^0 & \frac{-2}{\sqrt{6}}\eta \end{pmatrix} \end{aligned} \tag{2.146}$$

The full 8×8 meson mass matrix is a little messy. But the $\pi^\pm, K^\pm, K^0, \bar{K}^0$ do not mix.

$$M^2_{\pi^\pm} = \mu\left(m_u + m_d\right) + \left(\Delta M^2\right) \leftarrow \text{1-loop EM correction} \tag{2.147}$$
$$M^2_{K^\pm} = \mu\left(m_u + m_s^2\right) + \left(\Delta M^2\right) \leftarrow \text{ identical EM corrections} \tag{2.148}$$
$$M^2_{K^0,\bar{K}^0} = \mu\left(m_0 + m_s\right) \tag{2.149}$$

The π^0 and η mix. In the basis π^0, η the mass matrix is

$$\mu \begin{pmatrix} m_u + m_d & \frac{m_u - m_d}{\sqrt{3}} \\ \frac{m_u - m_d}{\sqrt{3}} & \frac{1}{3}(m_u + m_d + 4m_s) \end{pmatrix} \tag{2.150}$$

$m_u = m_d$ is called "isospin breaking." Isospin $= SU(2)_V$. It is a small effect. The electromagnetic 1-loop corrections are also isospin breaking and of a similar magnitude. These are small effects. So

$$m_\pi^2 \simeq (m_u + m_d)\,\mu + O(m_u - m_d) \tag{2.151}$$

$$m_\eta^2 \simeq \frac{\mu}{3}(m_u + m_d + 4m_s) \tag{2.152}$$

We don't know μ. But it drops out in ratios:

$$\frac{m_u}{m_d} = \frac{m_{K^+}^2 - m_{K^0}^2 + 2m_{\pi^0}^2 - m_{\pi^+}^2}{m_{K^0}^2 - m_{K^+}^2 + m_{\pi^+}^2} \simeq 0.55 \tag{2.153}$$

$$\frac{m_s}{m_d} = \frac{m_{K^0}^2 + m_{K^+}^2 - m_{\pi^+}^2}{m_{K^0}^2 - m_{K^+}^2 + m_{\pi^+}^2} \simeq 20.1 \tag{2.154}$$

and the Gell-Mann Okubo (GMO) formula:

$$4m_{K^0}^2 = 3m_\eta^2 + m_\pi^2 \tag{2.155}$$

Empirically, the left-hand side is about 0.99GeV^2, and the right-hand side is about 0.92GeV^2.

This is leading order in M. Note that $m_\pi^2 \propto m_{\text{quark}}$. So in the chiral expression, since $p^2 \sim m_\pi^2$, we should treat m like p^2:

$$\mathcal{L}_2 \supset p^2, m \tag{2.156}$$

$$\mathcal{L}_4 \supset p^4, p^2 m, m^2 \tag{2.157}$$

Let's revisit $\pi - \pi$ scattering to $O(p^2)$: we have

$$L_2 = \frac{f_\pi^2}{4}\operatorname{Tr}\partial_\mu\Sigma\partial^\mu\Sigma^\dagger + \mu\frac{f_\pi^2}{2}\operatorname{Tr}\left(\Sigma^\dagger M + M^\dagger\Sigma\right).$$

We sur the $\frac{1}{3f_\pi^2}\operatorname{Tr}[\pi, \partial_\mu\pi]^2$ before. Now there is also $\frac{2}{3}\mu\operatorname{Tr}(M\pi^4)$.

One can also develop a formalism for the interactions of NGBs with heavy, quasi-static, on-shell particles like nucleons.

2.7 The Standard Model as an EFT

The SM is defined as a perturbatively renormalizable QFT. But it can be supplemented by HDOs. What operators are allowed? What scales should we expect?

2.7.1 Dimension 5

$$\mathcal{L}_5 \supset \frac{\gamma_{ff'}}{M_{BSM}}(\phi L_f)(\phi L_{f'}) + cc \tag{2.158}$$

where

$$L_f = \begin{pmatrix} \nu_f \\ l_f^- \end{pmatrix} \qquad f = e, \mu, \tau \text{ SM lepton doublets}$$
$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \qquad \text{SM Higgs doublet}$$

Here and in the rest of this section we use 2-component fermion notation. In this case, the SU(2)-invariant contraction is $\phi L \equiv \epsilon^{ab}\phi_a L_b$.

The SM is invariant under three "lepton number" symmetries,

$$L_f \to e^{-i\alpha_f} L_f, \qquad \bar{e}_f \to e^{i\alpha_f}\bar{e}_f \tag{2.159}$$

For example, the Higgs coupling is

$$\frac{1}{v}\begin{pmatrix} m_e & 0 & 0 \\ 0 & m_\mu & 0 \\ 0 & 0 & m_\tau \end{pmatrix}_{ff'} (\phi L)_f \, \bar{e}_{f'} \tag{2.160}$$

(Here the SU(2)-invariant contraction is $\phi^* L \equiv \delta^{ab}\phi_a^* L_b$.)

But $\mathcal{L}_5$ breaks these symmetries in general. So lepton number may simply be an "accidental symmetry" in the SM. **Accidental symmetries are symmetries of low-energy EFTs that are consequence of gauge invariance and renormalizability.** Dropping renormalizability, lepton number it is easily violated by 2 units, but the effect is suppressed by powers of the cutoff on the SM EFT.

The operator above provides "Majorana mass term" for neutrinos,

$$\left(\frac{\gamma'_{ff}}{M_{BSM}}v^2\right)\nu_f\nu_{f'}. \tag{2.161}$$

If $M_{BSM} \sim 10^{16}$ GeV, $m_\nu \sim 10^{-3}$ eV.

The operator can be generated by a UV renormalizable Lagrangian,

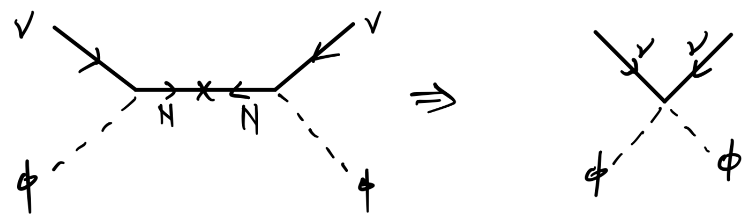
$$\mathcal{L} \supset -m_N NN - y(\phi L)N \tag{2.162}$$

where N are "sterile" (uncharged under electroweak symmetry) neutrinos with large m_N. After electroweak symmetry breaking the neutrino mass matrix is of the form

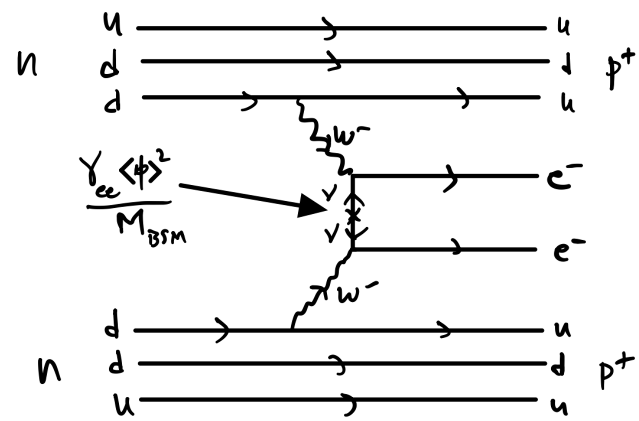
$$m_{\nu N} = \begin{pmatrix} m_N & yv \\ yv & 0 \end{pmatrix}. \tag{2.163}$$

If $m_N \gg v$, the eigenvalues are approximately m_N and $-(yv)^2/m_N$. This is the "seesaw mechanism" for generating light neutrino masses.

The EFT interpretation of the same physics is that we should integrate out N at m_N. This leaves behind a light, mostly-L neutrino with tiny mass from the dimension-5 operator



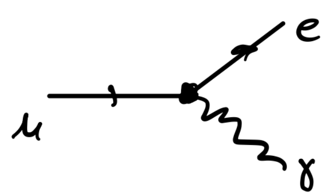
Low energy experiments probe this operator through neutrinoless double beta decay ("$0\nu\beta\beta$"):



2.7.2 Dimension 6

$$\mathcal{L}_{LFV} = \frac{\lambda_{ff'}}{M_{BSM}^2}(\phi^* L_f)\sigma^{\mu\nu}\bar{e}_{f'}F_{\mu\nu} \tag{2.164}$$

This operator induces lepton flavor-violating decays, for example, $\mu \to e\gamma$:

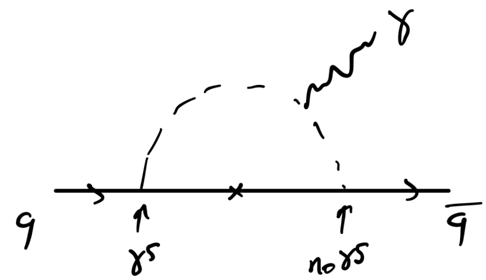


Experimental limits are of order $\mathrm{Br}(\mu \to e\gamma) \lesssim 10^{-11}$. Similar operators involving τs are less constrained.

Sometimes these are written as dimension-5 operators without the Higgs. This is sensible because the energy scale of the decay is a thousand times lower than the electroweak scale. However, it is important that the dimension-5 operator breaks chiral symmetries including $SU(2)_L$, so in order to embed it in the SM, we must eventually dress the L_f with a Higgs. This promotes the dimension 5 operator to dimension 6. Consequently, the natural size of the dimension 5 coupling is v/M_{BSM}^2 instead of $1/M_{BSM}$. The former may be much smaller.

$$\mathcal{L}_{qEDM} \subset \frac{\lambda_{ff'}}{M_{BSM}^2} \tilde{F}_{\mu\nu}(Q\phi)\sigma^{\mu\nu}\bar{u} + cc \tag{2.165}$$

Here $\tilde{F}_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\mu\mu\rho\sigma}F^{\rho\sigma}$. In the nonrelativistic limit this is an electric dipole moment operator for up-type quarks, $\sim \mu_u \vec{\sigma} \cdot \vec{E}$. It can be generated by integrating out heavy matter, for example



The quark EDM leads to a neutron EDM of order $-\frac{m_q}{M_{BSM}{}^2}$. Bounds on the neutron edm are of the order $10^{-26}e \cdot \text{cm}$, which corresponds to $M_{BSM} \gtrsim 100\text{TeV}$.

If we just remove the tilde from $\tilde{F}_{\mu\nu}$ we obtain magnetic dipole moment operators. For example, a typical contribution to the muon MDM is of the form

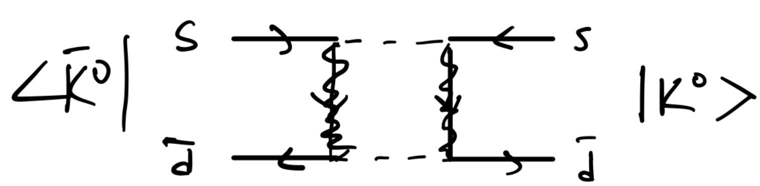
$$\mathcal{L}_{\mu g-2} = \frac{m_\mu}{M_{BSM}^2} F_{\mu\nu}\bar{\mu}\sigma^{\mu\nu}\mu. \tag{2.166}$$

The nonrelativistic limit is a contribution $-\mu_\mu \vec{\sigma} \cdot \vec{B}$ to the muon effective Hamiltonian.

Flavor-changing neutral current operators also arise at dimension 6. An example:

$$\mathcal{L}_{FCNC} = \frac{\varepsilon}{M_{BSM}^2}(s\sigma^\mu d^*)(s\sigma_\mu d^*) \tag{2.167}$$

This is a $\Delta S = 2$ operator and it contributes to $K - \bar{K}$ mixing. This process arises at one loop in the standard model, but is suppressed by small CKM elements. It can also receive contributions from new physics, e.g.:

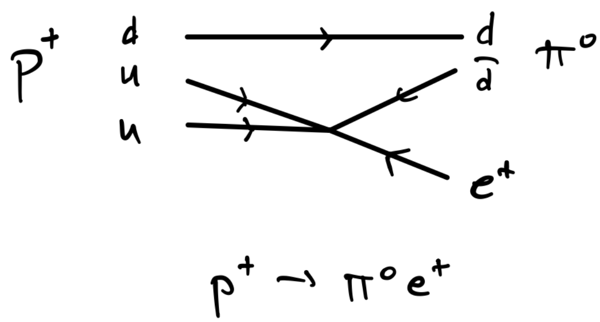


where the loop contains a new scalar and fermion.

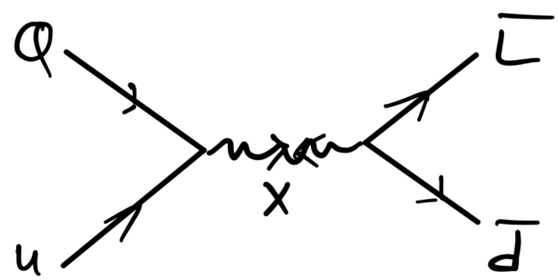
If the couplings are $\mathcal{O}(1)$, $K - \bar{K}$ mixing bounds $M_{BSM} \gtrsim 100$ TeV. However, if new physics approximately conserves a symmetry that the operator violates, there could be a small dimensionless coefficient. More properly, $M_{BSM} \gtrsim (100\sqrt{\varepsilon})$ TeV. New physics can be heavy, or weakly coupled, or both.

$$\mathcal{L}_p = \frac{1}{M_{BSM}^2} \left(Q\sigma^\mu \bar{u}^*\right) \left(L\sigma_\mu \bar{d}^*\right) \tag{2.168}$$

(Here we could also write $\bar{u}^* \equiv u_R$, $\bar{d}^* \equiv d_R$. The Q and L doublets are contracted in the $SU(2)$-invariant way.) This operator is $\Delta B = 1$, $\Delta L = 1$. It contributions to proton decay, e.g.



The bound on the proton lifetime comes from staring very hard at very large tanks of very pure water for a very long time. The current bound is of order $\tau_p \gtrsim 10^{33}$ years from which one infers $M_{BSM} \gtrsim 10^{15}$ GeV. A contribution to $\mathcal{L}_p$ could come from integrating out heavy GUT bosons, for example

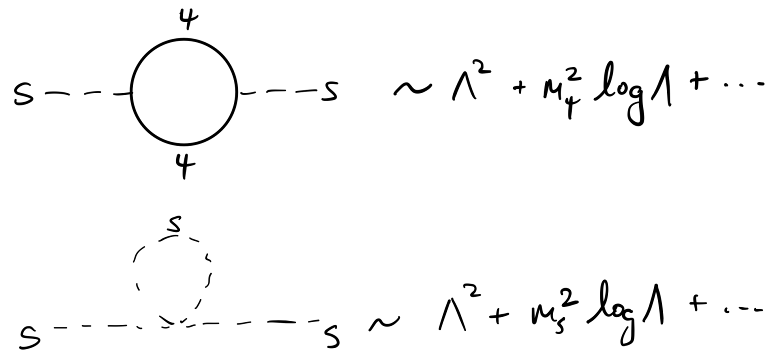


The unification scale is generally expected to be of order 10^{15-16} GeV, coincidentally comparable to the experimental reach. Unfortunately, since $\tau_p \sim M_{BSM}^{-4}$, sensitivity drops rapidly with unification scale.

In sum, the success of the SM *might* be explained if all new physics is very heavy.

2.7.3 The electroweak hierarchy problem

Ironically, it is not the nonrenormalizable operators that pose a conceptual problem, but the renormalizable ones. In the SM, there is only one: the (engineering) dimension-2 Higgs mass operator. Pretty generically, radiative corrections give "quadratically divergent" contributions to the counterterms for scalar mass parameters. For example,

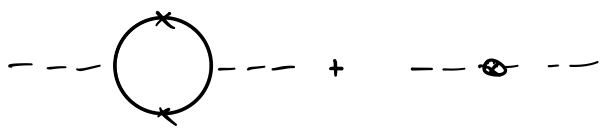


Quadratic divergences are regularization scheme dependent. However, they parametrize UV sensitivity, or the typical size of quantum corrections, if Λ is the cutoff scale of the SM EFT.

For example, in $DR+\overline{MS}$, there are no Λ^2 terms in the loop diagram. Mass-independent schemes never have such terms. But let us insert a physical UV completion. For simplicity, suppose we have a scalar S coupled to a heavy fermion ψ as

$$\mathcal{L}_{UV} = yS\bar{\psi}\psi + m_\psi \psi\bar{\psi} + \lambda S^4 \tag{2.169}$$

at some UV scale μ. Now let us integrate out ψ. At one loop, we have



$$\simeq \frac{y^2 m_\psi^2}{16\pi^2}\left(\log(\mu/m_\psi) + c\right) \tag{2.170}$$

This has to be matched onto the IR theory of just S:

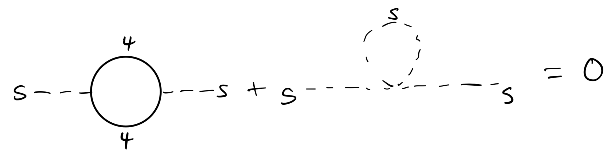
$$m_{IR}^2 = m_{UV}^2(\mu) + \frac{y^2 m_\psi^2}{16\pi^2}\left[\log\frac{\mu}{m_\psi} + c\right] \tag{2.171}$$

We started off with $m_{UV}^2(\mu) = 0$. The log can be interpreted as running m_{UV}^2 from μ, the sale where it vanishes, to the threshold m_ψ. It would be a coincidence if $\mu|_{m_{UV}^2=0} = m_\psi$! But even in that case, there is a finite threshold correction $\Delta m^2 = \frac{y^2 m_\psi^2}{16\pi^2}$ and $c \sim \mathcal{O}(1)$ in general.

For this reason, the fine-tuning problem of scalar mass parameters is often illustrated, qualitatively, by diagrams in the IR theory, just containing light fields in the loops. With $\Lambda^2 = m_\psi^2$, the quadratic divergence encodes sensitivity to new UV physical scales. Physically, it is just reflecting dimensional analysis and operator mixing under RG: if another relevant operator runs strong at some UV scale, interactions will mix it with scalar mass-squared operators and attempt to drive them strong as well.

So scalars "want" to be heavy. Symmetries can protect them:

1. Shift symmetries $\phi \to \phi + \delta$. This is the mechanism that keeps Goldstone bosons light. For example, the pions of ChPT don't have renormalizable couplings in the limit of exact chiral symmetry. Coupling to electromagnetism breaks the chiral symmetry explicitly and leads to quadratic divergences again.

2. Bose-Fermi cancellations in supersymmetry 

In the absence of symmetries or low cutoffs, relevant operators have to be fine-tuned in order to be weakly coupled. For scalars, this means the bare mass-squared parameter has to be fine-tuned to quadratic precision in order for the field to remain light. For example, if

$$m_{IR}^2 = m_{UV}^2 + \frac{y^2 m_\psi^2}{16\pi^2}(\log(\dots) + c) = 0 \tag{2.172}$$

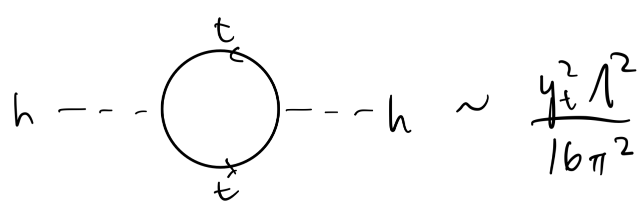
then m_{UV}^2 must be tuned against the *unrelated* parameters y and m_ψ in order to achieve $m_{IR} \ll y m_\psi$.

A fermion mass operator is also relevant (dimension 3), but in that case there is an (approximate) symmetry.

$$\Rightarrow \beta_{m_\psi} \sim m_\psi \tag{2.173}$$

which vanishes as $m_\psi \to 0$. This reflects the generic chiral symmetry $\psi \to e^{i\alpha\gamma_5}\psi$ which is restored as $m_\psi \to 0$. Thus these operators acquire only multiplicative logarithmic sensitivity to the cutoff.

Now let us look at the specific case of the SM Higgs boson, the only elementary scalar of the Standard Model. The quadratic divergence is usually sketched as a top loop,



As discussed, sensitivity to the cutoff from this diagram can be viewed as a qualitative estimate of various corrections at some high scale. For example, consider the seesaw mechanism of generating neutrino masses. The Higgs can couple to a heavy right-handed neutrino ν_R as

$$y_\nu (H \cdot L)\nu_R \qquad \left(H = \begin{pmatrix} \phi^+ \\ h \end{pmatrix}, \quad L = \begin{pmatrix} \nu_L \\ \ell^- \end{pmatrix}\right). \tag{2.174}$$

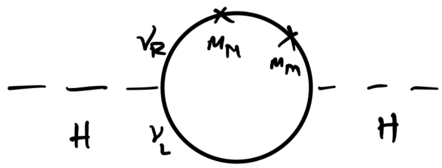
There is a heavy Majorana mass in the seesaw mechanism:

$$m_M \nu_R \nu_R + cc. \tag{2.175}$$

ν_R can be integrated out at m_M. We saw this generates the Weinberg operator,

$$\frac{y_\nu^2}{m_M}(H \cdot L)(H \cdot L) \tag{2.176}$$

which gives small masses to the ν_L. There is another important threshold correction:



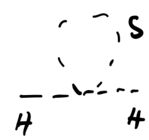
$$\sim \frac{y_\nu^2 m_M^2}{16\pi^2}\left(\log \mu/m_M + c\right).$$

The log contributes to $\beta_{m_H^2} \sim \frac{y_\nu^2 m_M^2}{16\pi^2}$. It causes m_H^2 to run above M_M^2. If we started from $m_H^2(\mu_0)$, the running shifts m_H^2 by $\Delta m_H^2 \simeq \frac{y_\nu^2 m_M^2}{16\pi^2}\log\frac{\mu_0}{m_M}$. The constant c is a residual threshold correction when we match onto an EFT with ν_R integrated out. So ultimately, the EFT contains a Higgs mass

$$m_{H,EFT}^2(m_M) = m_H^2(\mu_0) + \frac{y_\nu^2 m_M^2}{16\pi^2}\left[\log\left(\frac{\mu_0}{m_M}\right) + c\right]. \tag{2.177}$$

If $m_M^2 \gg (1\text{TeV})^2$, then $\Delta m_H^2 \gg (100\text{GeV})^2$ even if log and c are $\mathcal{O}(1)$. But at low energies, we need $m_{H,EFT}^2 \approx (100\text{GeV})^2$. So $m_H^2(\mu_0)$ has to delicately cancel Δm_H^2.

This problem is generic. The Higgs mass in the SM is not protected by any symmetries. All sorts of UV physics can strongly affect it, by dimensional analysis:

Higgs Portal: $\lambda H^\dagger H S^\dagger S$. Δm_H^2 from $\langle S^\dagger S\rangle$,  $\sim \frac{\lambda M_S^2}{16\pi^2}\log\ldots$

Z': $g^2 H^\dagger H Z'_\mu Z'^\mu$ Δm_H^2 from  $\sim \frac{g^2 m_{Z'}^2}{16\pi^2}\log\ldots$

indirect Z': $t \not{D}' \bar{t}$. Δm_H^2 from  $\sim \frac{g^2 y_t^2 m_{Z'}^2}{(16\pi^2)^2}\log\ldots$

If $\Delta m_H^2 \sim (10^{16}\text{ GeV})^2$, as in GUTs, then $m_H^2(10^{16}\text{ GeV})$ must cancel it to a part in 10^{26}. If the cutoff and correction are of order the Planck scale, then the cancellation must be good to a part in 10^{32}.

The radius of the solar system is like 10^{10} kilometers. A cancellation of a part in 10^{26} between two unrelated quantities is like subtracting the radii of two solar systems (according to some agreed-upon definition) and finding that they differ by a millimeter.

Small numbers like this cry out for dynamical explanation. An attractive solution is to introduce new physics that eliminates the corrections to m^2 from UV physics. Thus far no experimental evidence of such a mechanism has been forthcoming. Supersymmetry is one theory that can ameliorate the electroweak hierarchy problem, and we will study it later.

2.7.4 The strong CP problem

Back in Sec. 2.6, we claimed that QCD has a $SU(N_f)_L \times SU(N_f)_R$ flavor symmetry. This was incomplete. The classical symmetry is actually $U(N_f)_L \times U(N_f)_R$, which appears to have two additional $U(1)$'s worth of symmetry. The vector $U(1)$ is baryon number, which is not spontaneously broken by the chiral condensate. The axial $U(1)_A$ acts as

$$q_i \to e^{i\alpha\gamma^5} q_i \tag{2.178}$$

in four-component fermion notation (i is a flavor index, which is a spectator for this symmetry), or $q_{iL} \equiv q_i \to e^{i\alpha} q_i$, $q^*_{iR} \equiv \bar{q}_i \to e^{i\alpha}\bar{q}_i$ in two-component fermion notation. It is spontaneously broken along with all the other chiral symmetries by the chiral condensate. One can write a chiral current

$$j_5^\mu = \bar{q}\gamma_5\gamma_\mu q \quad \text{(4 comp)} \tag{2.179}$$

and classically the symmetry is only explicitly broken by the mass,

$$\partial_\mu j_5^\mu = m\bar{q}\gamma_5 q \tag{2.180}$$

using the Dirac equation. So the current seems conserved in the limit $m \to 0$ and we might expect a light goldstone boson, on the order of the kaon masses. But no light goldstone is observed. The closest candidate is the η' meson, which is almost a GeV.

This puzzle known as the "η' problem," which was ultimately resolved by the discovery of anomalies and instantons. The $U(1)_A$ current has an ABJ anomaly with QCD. Under a $U(1)_A$ transformation with parameter γ, the action shifts by

$$\delta\Gamma = \gamma\partial_\mu J^\mu = \gamma\left(n_q \frac{\alpha_s}{2\pi}\mathrm{Tr}(F\tilde{F}) + \bar{\psi}_i m_{ij}\gamma^5\psi_j\right) \tag{2.181}$$

where F is the gluon field strength and m is the quark mass matrix. As we will discuss later $\int \mathrm{Tr}(F\tilde{F})$ is nonzero on nonperturbative gluon fields, and so the $U(1)_A$ is strongly explicitly broken around the confinement scale. So there is no tension with Goldstone's theorem. There is no small parameter in QCD that we could dial down to restore the symmetry (no spurion); sending $m_{ij} \to 0$ is not enough.

However, this raises a different puzzle. $\mathrm{Tr}(F\tilde{F})$ is a marginal operator. It should be added to the QCD Lagrangian:

$$\Delta\mathcal{L} = \frac{\theta}{16\pi^2}\int \mathrm{Tr}(F\tilde{F})d^4x \tag{2.182}$$

We will see later that $\int F\tilde{F}$ is actually an integer-valued winding number, so the coupling θ is an angle. We can try to remove it by a field redefinition (an anomalous $U(1)_A$ transformation): in Eq. (2.181), let $\gamma = -\theta$. But what this actually does is move θ from the coefficient of $F\tilde{F}$ into the phase of the quark mass matrix. For a non-infinitesimal $U(1)_A$ transformation,

$$\theta \to \theta + \alpha, \ \arg\ \det m \to \arg\det m + \alpha \tag{2.183}$$

$$\tag{2.184}$$

and so for $\alpha = -\theta$ we have

$$m = \begin{pmatrix} m_u & & \\ & m_d & \\ & & m_s \end{pmatrix} \to e^{i\frac{\theta}{3}} \begin{pmatrix} m_u & & \\ & m_d & \\ & & m_s \end{pmatrix} \tag{2.185}$$

This is a useful technique to incorporate θ into ChPT . More generally, we can say that

$$\bar{\theta} \equiv \theta + \arg\det m \tag{2.186}$$

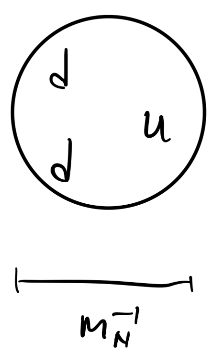
is a new field-redefinition invariant coupling/phase, which shows up in the renormalizable QCD Lagrangian. Subsequently we will refer to θ, but we really mean $\bar{\theta}$.

In ChPT, one prediction of a nonzero θ is

$$\Gamma(\eta \to \pi\pi) \sim m \sin\theta \tag{2.187}$$

For $\theta \to 0$, the Goldstones are P-odd and interactions conserve P. For $\theta \neq 0$, $P(F\tilde{F}) = -(F\tilde{F}) \Rightarrow$ P-violating processes like $\eta \to \pi\pi$ occur at $\mathcal{O}(\theta)$.

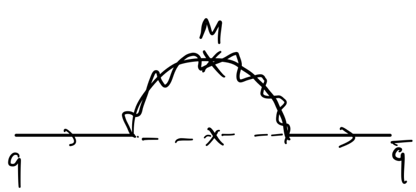
Observationally, nuclear, atomic, and molecular EDM experiments all show that θ is tiny. The strongest bounds come from the neutron EDM.



On dimensional grounds: $d_n \simeq e\theta m_n^{-1} \sim \theta \cdot 10^{-14}$ecm. But $d_n < 10^{-26}$ e cm. Sharper estimates include a $1/4\pi^2 \Rightarrow \theta \lesssim 10^{-10}$.

This is weird.

- It's easy to generate threshold corrections at tree level or one loops, e.g.



- CP is violated by the CKM phase, and by the obvious matter-antimatter asymmetry of the universe: CP cannot be a good symmetry of nature

Now that we understand θ is physical, it is a big puzzle why it is so small.

Solutions to the strong CP problem

(1) $m_u = 0$**.**

If $\det m = 0$, arg det m is unphysical. So we can "rotate away" θ by an anomalous $U(1)_A$ transformation. More precisely, $|m_u/m_d| \lesssim 10^{-10}$.

There are issues with this. Which m_u should be set to zero? $m_u = m_u(\mu)$. The chiral symmetry $m_u \to m_u e^{i\alpha}$ can protect a mass if it is a good symmetry. But it is composed of a non-anomalous $SU(3)_A$ transformation $e^{i(\beta_1 T^3 + \beta_2 T^8)\gamma^5}$, and an anomalous $U(1)_A$ transformation $e^{i\beta_3 \mathbb{I}\gamma^5}$: we need

$$\beta_1 T^3 + \beta_2 T^8 + \beta_3 \mathbb{I} = \text{diag}(\alpha, 0, 0) \tag{2.188}$$

which is three equations for three unknowns, and generically β_3 is nonzero. So it is not a real symmetry. And there is no reason not to expect UV contributions to $\text{Re}(m_u^{eff})$.

One can build models, but they are quite elaborate, involving large discrete flavor symmetries like $\mathbb{Z}_4 \times \mathbb{Z}_5$ (Banks Nir Seiberg).

Furthermore, we saw in (leading order) ChPT that the meson masses imply $m_u/m_d \simeq 1/2$. But at which scale? There are subtleties at second order in ChPT that amount to renormalization of m_u by small instantons, in a way that does *not* regenerate θ.

$$(\delta m_u)_{inst} \sim m_d m_s/\Lambda_{QCD} \tag{2.189}$$

with a cutoff on instanton sizes of order $\rho^{-1} \sim \Lambda_{QCD}$. But because of an IR divergence in the integration over instanton sizes, this is actually extremely sensitive to the IR cutoff: more precisely, it reads

$$(\delta m_u)_{inst} \sim (m_d m_s/\Lambda_{QCD}) \times (\rho_{IR}\Lambda_{QCD})^9 \tag{2.190}$$

So the only conclusive test is lattice QCD: Set a small lattice spacing $a \ll \Lambda_{QCD}^{-1}$, feed in the QCD lagrangian, and tweak m_u until m_π^2, m_K^2, f_π, m_{proton}, et cetera match observations. The result:

$$m_u(\mu \simeq 2\text{GeV}) \simeq 2 \pm 0.1\text{MeV} \tag{2.191}$$

So $m_u = 0$ is excluded to high confidence! Evidently the instanton contribution is small, and leading-order ChPT was a good guide:

$$\frac{m_u(2\text{GeV})}{m_d(2\text{GeV})} \simeq \frac{2}{4} \simeq 1/2. \tag{2.192}$$

(2) Spontaneous CPV

Maybe CP is secretly a good symmetry, but is spontaneously broken in the UV. IF SSB occurs in such a way that $\bar{\theta}$ is un affected, but CKM is generated, is strong CP solved?

We have to look at the radiative corrections. Remarkably, in the SM, $\beta_{\bar{\theta}}$ starts at 7 loop order. For any reasonable cutoff, $\Delta\bar{\theta} \ll 10^{-10}$ from running. This is the virtue of these models: small $\bar{\theta}$

is not fine-turned in the SM, in the sense that the radiative corrections are tiny. (There are finite corrections at lower loop order, but still high enough to be negligible.)

In the presence of new physics, however, all bets are off. It is generic that 1-loop effects recouple ϕ_{SCPV} to $F\tilde{F}$

$$\Rightarrow \quad \frac{\langle \phi_{\text{SCPV}} \rangle}{16\pi^2 M} = \Delta\theta$$

These models are complicated and radiatively unstable in most cases, but they are a popular model-building playground.

(3) Peccei-Quinn mechanism

We have already seen a toy model for the Peccei-Quinn mechanism in our discussion of anomalies in Sec. (2.4). The model is defined by Eq. (2.59), which we reproduce here:

$$\mathcal{L} = -\frac{1}{4e^2} F_{\mu\nu}\tilde{F}^{\mu\nu} - \frac{1}{4g^2}\text{Tr}(G_{\mu\nu}\tilde{G}^{\mu\nu}) + |\partial_\mu \phi|^2 + i\bar{Q}\not{D}Q - y\phi Q\bar{Q} - V(\phi). \tag{2.193}$$

After spontaneous symmetry breaking by the vev $\langle\phi\rangle = f$, we found a low-energy EFT description containing the gauge fields and a Goldstone boson, with couplings given by Eq. (2.65). In particular there was a coupling to QCD,

$$\sim \frac{c(yf)y}{(yf)^2 16\pi^2} a\text{Tr}(F\tilde{F}) = \frac{c}{16\pi^2 f} a\text{Tr}(F\tilde{F}) \in L_{eff} \tag{2.194}$$

This violates the NGB shift symmetry, but preserves an $a/f \to a/f + 2\pi k/c$ shift symmetry. Note that a essentially makes θ dynamical:

$$L \supset \theta\text{Tr}(F\tilde{F}) \to \left(\theta + c\frac{a}{f}\right)\text{Tr}(F\tilde{F}) \subset L_{\text{eff}} \tag{2.195}$$

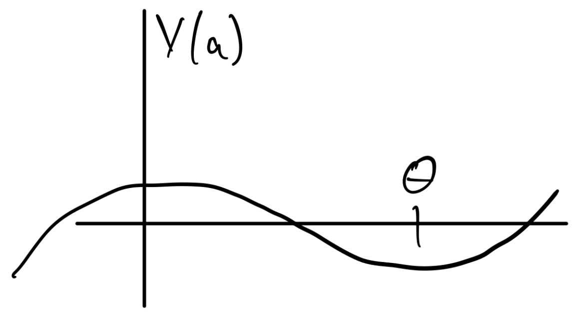
We can also infer this coupling from the anomalous current, $\partial_\mu J^\mu_{PQ} \sim \text{Tr}(F\tilde{F})$ and $J^\mu \sim f\partial^\mu a$.

We saw that in ChPT we could move θ into the pion potential. We can do the same with $\langle a \rangle$ to get a combined axion-pion potential.

$$\begin{aligned} &U(1)_A : \psi \to e^{i(\theta + c\langle a\rangle/f)\gamma^5}\psi \\ &\Rightarrow \mathcal{L} \supset -m_q e^{i\theta + \langle a\rangle/f}\langle\bar{\psi}\psi\rangle e^{2i\pi^a T^a/f_\pi} + cc \end{aligned} \tag{2.196}$$

For example, with only one light flavor, this would only be an axion potential (since the η' is heavy):

$$V(a) = -m_q\langle\bar{\psi}\psi\rangle\cos(\theta + ca/f) \tag{2.197}$$



The vev is $c\langle a\rangle/f = -\theta$. So, if we shift to the origin $a \to \langle a\rangle + \delta a$, then

$$L \supset (a/f + \theta)\text{Tr}(F\tilde{F}) \to \frac{\delta a}{f}\text{Tr}(F\tilde{F}) \tag{2.198}$$

θ is dynamically relaxed! The prediction of the theory is a new light boson:

$$\left.\frac{\partial^2 V}{\partial a^2}\right|_{a=\langle a\rangle} \sim \frac{(m_q \Lambda_{QCD}) f_\pi^2}{f_a^2} \to \frac{m_\pi^2 f_\pi^2}{f_a^2} \tag{2.199}$$

Let's plug in some numbers. $(m_\pi f_\pi)^2 \sim (100\text{MeV})^2$. To hide the new heavy fields (the color-charged fermions Q and the radial mode of the scalar ϕ), we need $f_a \gtrsim 1$ TeV. If $f_a \in [\text{TeV}, 10^{16}\text{GeV}]$, $m_a \in [10^{-9}\text{eV}, \text{keV}]$.

Note that this is a SM extension where the SM is not even the right EFT! There is a new light particle that must be included, with purely higher dimension couplings to SM fields.

Issues:

- $\frac{\delta_a}{f} F\tilde{F}$ allows axion production in nuclear environments. Axion radiation cools large stars too quickly. To suppress, we must increase $f_a : f \geq 10^9$GeV.
- In the hot early universe, can produce many axions. The simplest mechanism is called misalignment, which results in coherent production. EOM: $\ddot{a} + 3H\dot{a} + V(a) = 0$. For $H > m_a$, Hubble friction freezes a at some random value, typically $O(f)$. When $H < m_a$ as the universe cools, $H\dot{a}$ can be ignored: coherent oscillations begin.

$$\begin{aligned} \rho^a_{\text{osc}} &\sim m_a^2 a^2 \sim m_a^2 f^2 \\ \rho^{\text{rad}}_{\text{osc}} &\sim T^4_{\text{osc}} \sim H^2_{osc} M_p^2 \sim m_a^2 M_p^2 \end{aligned}$$

The coherent oscillations redshift like matter, $\rho^a/\rho^{\text{rad}} \sim 1/T$ Requiring no matter domination before 1 eV,

$$\frac{\rho^a_{\text{osc}}}{\rho^{osc}_{\text{rad}}} \sim \frac{1 \text{ eV}}{T_{\text{osc}}} \sim \frac{1 \text{ eV}}{\sqrt{m_a M_p}}$$
$$\text{or } \frac{m_a^2 f_a^2}{m_a^2 M_p^2} \sim \frac{1eV}{\sqrt{m_a M_p}} \sim \frac{1eV}{\sqrt{m_a f_\pi M_p / f_a}}$$

which implies $f_a \lesssim 10^{11}$ GeV. For $f_a \sim 10^{11}$ GeV the QCD axion can be all of dark matter. If f_a is larger, one has to do more work to avoid overclosure. Narrow window..

- We postulated an anomalous PQ symmetry. So it's not a symmetry. Why shouldn't there be other sources of PQ violation, other than the QCD anomaly? E.g., a piece of lore: quantum gravity breaks all global symmetries. If this breaking can be modeled in EFT via operators of the form $V \supset \left(\frac{\phi}{M_p}\right) \frac{|\phi|^{4+q-1}}{(M_p)^{q-1}} + cc$, (which breaks PQ by 1 unit) then $\Delta V_a \sim f_a^4 \left(\frac{f_a}{M_p}\right)^q \cos(a/f_a)$ The minimum is moved by this correction, away from $a = -\theta f$. We require, roughly,

$$f_a^2 \,(f_a/M_p)^q < 10^{-10} m_\pi f_\pi / f_a$$

Since f_a is so large and $m_\pi f_\pi$ so small, this requires elimination of such higher dimension, out to dimension 12 or so, for $f_a = 10^{11}$GeV. Why should PQ be of such high quality, only to be broken by QCD? It is possible to invent mechanisms in quantum gravity such that the violation of PQ by quantum gravity effects is nonperturbative, e.g.$\sim e^{-M_p^2/f_a^2}$. In such cases the axion does not have a quality problem.

Chapter 3

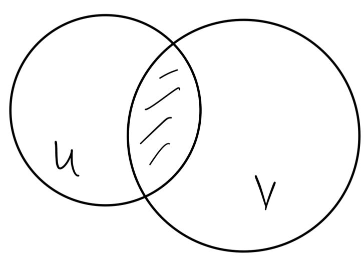
Topological objects

3.1 Mathematical preliminaries

Manifolds

M is a collection of points such that $\forall p \in M$, there is a neighborhood homeomorphic to an open set in R^n. We cover M with open sets U_i. Homeo $\Rightarrow \exists \phi_u : U \to D^n$, where D is an open disc and ϕ_u is invertible. ϕ is called a chart, and points in D^n are coordinates.

Where two sets overlap,



we can use the coordinates of either. "Transition functions" $\phi_U \circ \phi_V^{-1}$ transform V coordinates into U coordinates.

Differentiable manifold: all transition functions are C^∞. Transition functions are diffeomorphisms.

Example: $M = S^1$: we can use polar angle from y axis, θ, as a local coordinate. But we cannot cover S^1 in one chart, because θ and $\theta + 2\pi$ are the same point in S^1. θ is not one to one, so we need at least 2 charts.

Tangent Space

$\forall p \in M$, there is an associated vector space T_pM. The elements of T_pM are the directional derivative operators. A basis is given by $\frac{\partial}{\partial x^i}$. The tangent bundle is the union of these vector spaces, $TM \equiv \bigcup_p T_pM$. A vector field, or a section of the tangent bundle, is a choice of element of $T_pM \forall p$: $x = \xi^i(x) \frac{\partial}{\partial x^i}$.

Cotangent Space

The cotangent space T_p^*M is the set of maps $T_pM \to \mathbb{R}$. It is a vector space and the elements are called "one forms." We define a coordinate basis dx^i by the rule:

$$\left(dx^j, \frac{\partial}{\partial x^i}\right) = \delta_i^j. \tag{3.1}$$

We define an interior product i_x as

$$\begin{aligned} i_x\omega &\equiv \left(\omega_j dx^j, \xi^i \frac{\partial}{\partial x^i}\right) \\ &= \omega_i \xi^i. \end{aligned} \tag{3.2}$$

Differential k-forms

We can define the k-fold antisymmetric tensor product of T_p^*M. An element of this vector space is a differental k-form:

$$\omega = \frac{1}{k!}\omega_{i_1,\ldots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k} \tag{3.3}$$

Here $\wedge$ denotes an antisymmetric "wedge" product. The wedge product of a k-form α and a p-form β is

$$\alpha \wedge \beta = \frac{1}{k!p!}\alpha_{[i_1\ldots i_k}\beta_{i_{k+1}\ldots i_{p+k}]} dx^{i_1} \wedge \ldots \wedge dx^{i_{k+p}} \tag{3.4}$$

The $[\ldots]$ notation on the indices denotes antisymmetrization, which includes a $\frac{1}{(p+k)!}$. It can be added or removed for free because it is redundant with the antisymmetrization in the wedge products. It can be convenient to carry around when working just with the components.

All $k > n$ forms vanish on manifolds of dimension n, because $dx \wedge dx = 0$ by antisymmetry.

Exterior derivative

The exterior derivative d is an operator that takes k-forms to $k+1$-forms.

$$d\omega \equiv \frac{1}{k!} \quad \frac{\partial \omega_{i_1} \ldots i_k}{\partial x^j} dx^j \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} \tag{3.5}$$

If f is a 0-form (a function) and ω is a one-form, then

$$df = \frac{\partial f}{\partial x^i} dx^i \tag{3.6}$$

$$d\omega = \frac{\partial \omega_i}{\partial x^j} dx^j \wedge dx^i = \frac{1}{2}\left(\partial_i \omega_j - \partial_j \omega_i\right) dx^i \wedge dx^j \tag{3.7}$$

On any form,

$$d^2 \propto \partial_i \partial_j - \partial_j \partial_i = 0. \tag{3.8}$$

If α is a k-form, then

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta. \tag{3.9}$$

Integration of forms

One sometimes reads the phase "it is natural to integrate k-forms on k-dim submanifolds of M." What this means is the following. First, construct a k-tensor of vectors called the coordinate k-paralleliped. In three dimensions, for example, this is the 3-form:

$$v = \Delta x \Delta y \Delta z \partial_x \otimes \partial_y \otimes \partial_z \tag{3.10}$$

Now contract it with a k-form $\omega = \frac{1}{k!}\omega_{i_1 \ldots i_k} dx^{i_1} \ldots dx^{i_k}$:

$$i_v \omega = \frac{1}{k!}\omega_{i_1 \ldots i_k} \epsilon^{i_1 \ldots i_k} \Delta x^1 \ldots \Delta x^k. \tag{3.11}$$

The contraction is coordinate invariant. Then we replace $\Delta x^1 \ldots \Delta x^k \equiv d^k x$, the ordinary integration measure. Then we integrate.

One may think of this construction as the k-form integration rule:

$$dx^{i_1} \wedge \ldots \wedge dx^{i_k} \Rightarrow \epsilon^{i_1 \ldots i_k} d^k x. \tag{3.12}$$

We can also integrate functions (0-forms) if we have a "volume form". On n-dimensional Riemannian manifolds there is a metric and "canonical volume form." To define it, first we need the Hodge $\star$ operation, which is a map from k-forms to $(n-k)$-forms:

$$\star\omega \equiv \frac{1}{(n-k)!}\frac{1}{k!}\sqrt{g}\epsilon_{i_1 \ldots i_n} g^{i_1 j_1} \ldots g^{i_k j_k} \omega_{j_1 \ldots j_k} dx^{i_{k+1}} \wedge \ldots \wedge dx^{i_n}. \tag{3.13}$$

Then the volume form e is defined as

$$e \equiv \star 1 \tag{3.14}$$

$$= \frac{1}{n!}\sqrt{g}\epsilon_{i_1 \ldots i_n} dx^{i} \wedge \ldots \wedge dx^{i_n} \tag{3.15}$$

$$= \sqrt{g} d^n x \tag{3.16}$$

where, in going to the last line, we used the definition of how to integrate forms. We recognize here the coordinate-invariant measure.

Essentially, k-forms transform inversely to $d^n x$, which is why our integration of forms is coordinate invariant. In the canonical volume form, the $\sqrt{g}$ takes care of this. It is there in the definition of $\star$ so that the $(n-k)$ form components have the right transformation property given the transformation property of the k-form components.

In practice it is convenient to integrate a p-form over a p-manifold like this:

$$\int F = \int (\star_p F) dV_p. \tag{3.17}$$

Here $\star_p F$ is the dual on the p-manifold. With induced metric h,

$$\begin{aligned}\star_p F &= \frac{1}{p!}\frac{1}{(p-p)!}\sqrt{h}h^{\mu_1\nu_1}\dots h^{\mu_p\nu_p}\epsilon_{\nu_1\dots\nu_p}F_{\mu_1\dots\mu_p} \\ &= \frac{1}{p!\sqrt{h}}\epsilon^{\mu_1\dots\mu_p}F_{\mu_1\dots\mu_p}. \end{aligned} \tag{3.18}$$

Alternatively, if we know that $F = qe$ where e is the volume form, then the integration is simple and gives qV_p.

Other useful facts:

$$d(M\wedge N) = dM\wedge N + M\wedge dN(-1)^m \tag{3.19}$$

$$M\wedge N = N\wedge M(-1)^{mn} \tag{3.20}$$

$$M\wedge\star M = \langle M, M\rangle dV \tag{3.21}$$

where $\langle M, M\rangle = \frac{1}{p!}M_{i_1\dots i_p}M^{i_1\dots i_p}$. Only the wedge product of two identical one-forms vanishes; in general two identical p-forms don't have vanishing wedge product.

Closed & Exact forms

$$d\omega = 0 \Rightarrow \omega \text{ is closed} \tag{3.22}$$

$$\omega = d\alpha \Rightarrow \omega \text{ is exact} \tag{3.23}$$

More precisely, ω is exact if it can be written as $d\alpha$ for a well-defined $k-1$ for α. "well-defined" will be important later.

- $d^2 = 0 \Rightarrow$ all exact forms are closed.
- Poincaré lemma: all closed forms are locally exact.

Of particular interest are closed forms that fail to be globally exact. The simplest examples are volume forms on closed, boundary-less manifolds. As n-forms on n-dimensional manifolds, $de = 0$ holds automatically. Are they exact? (is $e = db$ for some $(n-1)$-form b?) Stokes' Theorem is a useful/ diagnostic:

$$\int_C d\omega = \int_{\partial C}\omega \tag{3.24}$$

where C is a $k+1$-dimensional subspace of M, ∂C is the boundary of C, and ω is a k-form.

So, if the volume form was exact, we would have

$$\int_M e \stackrel{?}{=} \int_M db = \int_{\partial M} b = 0 \tag{3.25}$$

because $\partial M = 0$. But the first integral is nonzero because it compute the volume. So the volume form on such M is always closed and never exact.

Example 1: $M = S^1$. The volume form is often written "$e = d\theta$." This is crummy notation for a 1-form that is closed but not exact, because the "0-form" θ is not globally defined (it is not a function on the circle.) We have the volume

$$\int d\theta = 2\pi \tag{3.26}$$

What this means it we remove a point from the circle and do the integral on this patch where θ is defined. The point contributes zero to the integral.

Example 2: $M = S^2$. We write $e = \sin\theta d\theta \wedge d\varphi$. This is globally defined on S^2. Locally, $e = d(-\cos\theta d\varphi)$, but $\cos\theta d\varphi$ is not a globally defined 1-form because $\varphi, d\varphi$ are not defined at $\theta = 0, \pi$.

$$\int_{S^2} e = 4\pi. \tag{3.27}$$

de Rham cohomology

The k^{th} cohomology group is

$$H^k(M) \equiv C^k(M)/Z^k(M) \tag{3.28}$$

where $C^k(M)$ is the space of closed k-forms on M and $Z^k(M)$ is the space of exact k-forms. $H^k(M)$ forms an abelian group with group multiplication defined as addition of the forms.

Homotopy

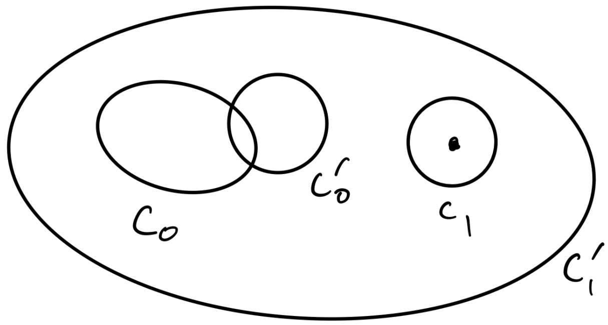
Let f_1, f_2 be two maps:

$$f_1(x) : M \to N \tag{3.29}$$
$$f_2(x) : M \to N \tag{3.30}$$

where M, N are manifolds. Then f_1 is homotopic to f_2 if there is a continuous map $f(x,\tau), \tau \in [0,1]$ such that $f(x,0) = f_1(x)$ and $f(x,1) = f_2(x)$. We say that f_1 and f_2 are "equivalent under continuous deformations."

Spheres provide an important class of example manifolds. $\pi_k(N)$ denotes the homotopy-equivalent classes of maps from $S^k \to N$. $\pi_1(N)$: is called the "fundamental group" of N, and contains equivalence classes of closed loops on N, where two loops are equivalent if they are deformable into each other.

Example: $\pi_1\left(R^2 - \{0\}\right)$.



We see that $C_0 \sim C_0'$, $\quad C_1 \sim C_1'$, $\quad C_0 \nsim C_1$.

Define addition of two loops by deforming them to have the same starting point and then traversing both sequentially. Thus $C_0 + C_0 \sim C_0, C_0 + C_1 \sim C_1, C_1 + C_1 \sim C_2$.

So loops $\cong$ the group of integers, with $C_0 = 0$. We conclude

$$\pi_1(\mathbb{R}^2 - \{0\}) = \mathbb{Z} \tag{3.31}$$

and group multiplication = map composition. This sort of definition can be extended to other π_k - hence "homotopy groups."

When the group is $\mathbb{Z}$ it is often possible to use elements of a cohomology group to define winding numbers, which compute the homotopy class.[1] This is a powerful technique.

Example: $\pi_n(S^n)$

The maps are $n^a(x) : S^n_M \to S^n_N$, with $a = 1 \ldots n+1$ and $n^a n^a = 1$. n^a are cartesian coordinates on the target sphere.

The relevant element of $H^n(S^n)$ is the volume form. The volume form on the sphere in these coordinates is:

$$dV_{S^n} = i_{\partial_r} dV_{\mathbb{R}^{n+1}} = \left(\partial_r, \frac{1}{(n+1)!}\epsilon_{a_1,\ldots a_{n+1}} dn^{a_1} \wedge \ldots \wedge d_n^{a_{n+1}}\right). \tag{3.32}$$

Let $n^a = \bar{n}^a r$. Then $dn^a = r d\bar{n}^a + \bar{n}^a dr$, so

$$\begin{aligned} dV_{S^n} &= \frac{1}{(n+1)!}\epsilon_{a_1\ldots a_{n+1}} \left(\bar{n}^{a_1} dn^{a_2} \wedge \ldots \wedge dn^{a_{n+1}} \underbrace{\pm \ldots\ldots\ldots\ldots}_{n+1 \text{ equiv terms}}\right) \\ (\text{set } r=1) \quad &= \frac{1}{n!}\epsilon_{a_1\ldots a_{n+1}} n^{a_1} d_n^{a_2} \wedge \ldots \wedge d_n^{a_{n+1}}. \end{aligned} \tag{3.33}$$

[1]Technically, for $\pi_k(N) = \mathbb{Z}$, we look for an element of $H^k(N)$ - that is, a form ω on N, which we pull back to S^k and integrate. A pullback of a form is defined as follows. Given a map $f : M \to N$ and $\omega \in T^*_{f(p)}N$, we define $f^*\omega \in T^*_p M$ by $(f^*\omega, v \in T_p M) = \left(w, f_* v \in T_{f(p)}N\right)$ where $(f_* v)(h) = v(h(f))$ for any function h on N.

Now plug in $n^a \to n^a(x)$ where x are coordinates on S^n_M. Then

$$\frac{1}{Vol(S^n)} \int_{S^n_M} \frac{1}{n!} \epsilon_{a_1 \dots a_{n+1}} n^{a_1}(x) dn^{a_2}(x) \wedge \dots \wedge dn^{a_{m+1}}(x) \equiv Q[n]. \tag{3.34}$$

$Q[n]$ is the winding number of the map. It is a homotopy invariant: $Q[n+\delta n(x)] - Q[n] = 0$. To see it, we rearrange a little bit,

$$\begin{aligned} Q[n+\delta n(x)] - Q[n] = \frac{1}{n! Vol(S^n)} \int \Big[&(n+1) \delta n^{a_1} \epsilon_{a_1 \dots a_{n+1}} dn^{a_2} \wedge \dots \wedge dn^{a_{n+1}} \\ &- nd \left(\epsilon_{a_1 \dots a_{n+1}} \delta n^{a_1} n^{a_2} dn^{a_3} \wedge \dots \wedge dn^{a_{n+1}} \right) \Big]. \end{aligned} \tag{3.35}$$

The exact form integrates to zero on M. The first term vanishes because $n^a n_a = 1 \to \delta n^a n_a = 0$ and $\epsilon_{a_1} \dots a_{n+1} dn^{a_2} \wedge \dots \wedge dn^{a_{n+1}} \propto n_{a_1}$.

The winding number computes how many times the sphere S^n_M is wrapped over S^n_N.
Example: the map $f : \theta, \varphi \to \theta' = \theta, \varphi' = 2\varphi$ wraps $S^2_{(N)}$ twice over $S^2_{(M)}$. Change by on exact form: same answer.

A gauge theory can be defined by a specification of bundles to include in the path integral. A gauge field is a connection on a principal bundle. We will not need too much of this machinery. Main things we need:

- Gauge fields $\longleftrightarrow$ forms on manifolds
- Gauge group $G \leftrightarrow$ gauge fields are only defined up to transition functions in G.
- Cohomology classes $\leftrightarrow$ interesting/topologically nontrivial gauge field configurations, often associated with a winding number.

Abelian gauge theory in form language

Define a 1-form $A \equiv A_\mu dx^\mu$. Then

$$\begin{aligned} dA &= \partial_\mu A_\nu dx^\mu \wedge dx^\nu \\ &= \frac{1}{2} \left(\partial_\mu A_\nu - \partial_\nu A_\mu \right) dx^\mu \wedge dx^\nu \\ &= \frac{1}{2} F_{\mu\nu} dx^\nu \wedge dx^\nu \\ &\equiv F, \text{ a 2-form.} \end{aligned} \tag{3.36}$$

Also

$$F = E_i dx^0 \wedge dx^i + \frac{1}{2} \epsilon_{ijk} B^k dx^i \wedge dx^j. \tag{3.37}$$

Since $F = dA$ locally, $dF = 0$. This encodes Bianchi and Maxwell, $\nabla \cdot B = 0$ and $\nabla \times E = -\partial_t B$.

In 4D the dual of a 2-form is a $4-2=2$ form:

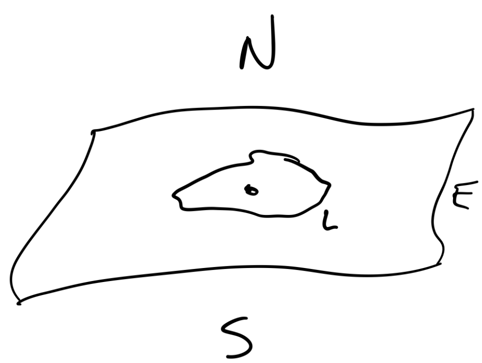
$$\star F = B_i dx^0 \wedge dx^i - \frac{1}{2}\epsilon_{ijk}E^k dx^i \wedge dx^j \tag{3.38}$$

So Hodge duality swaps $E \to B, B \to -E$. $d\star F = \star J$ are the other two Maxwell equations. The covariant action is

$$\begin{aligned} S &= -\frac{1}{2}\int_M F\wedge \star F + \int_M A \wedge \star J \\ &= \int d^4x\sqrt{g}\left\{-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + A_\mu J^\mu\right\} \end{aligned} \tag{3.39}$$

3.2 Wu-Yang Magnetic Monopole

The fields of a point charge are defined on $\mathbb{R}^3 - \{0\}$. Now consider two patches of space, separated by a surface E that is topologically $\mathbb{R}^2 - \{0\}$.



The gauge fields on $N \cap S = E$ orly have to agree up to a gauge transformation:

$$A_N = A_S + \frac{1}{e}d\lambda \tag{3.40}$$

for some function $\exp(i\lambda(E)) \in U(1)$.

Now look at non-contractible loops L in E. Since $U(1) = S^1$ and $\pi_1(S^1) = \mathbb{Z}$ there should be classes of homotopically-inequivalent maps from $L \to U(1)$ that are classified by a winding number, e.g.,

$$\lambda = n\theta \quad n \in \mathbb{Z} \tag{3.41}$$

The class is computed by the cohomology element e, the volume form on S^1, pulled back to L $\lambda =$ angle on $U(1) \Rightarrow d\lambda =$ volume element. So the object of interest is $\int_L d\lambda$ (where $d\lambda$ is not an exact form on L.) For the example maps above, it is $\int_L d\lambda = n\int_L d\theta = 2\pi n$.

Suppose we can find suitable gauge fields A_N and A_S that differ on E by a gauge function λ with winding number n. Then

$$\begin{aligned}2\pi n &= \int_L d\lambda \\ &= e\left(\int_L A_N - \int_L A_S\right) \\ &= e\left(\int_N F_N + \int_S F_S\right) \\ &= e\int_S F \\ &= e \times \text{magnetic charge, using magnetic Gauss law}\end{aligned} \tag{3.42}$$

Let the magnetic charge be m. Then we have found

$$\frac{em}{2\pi} \in \mathbb{Z}. \tag{3.43}$$

This is the Dirac quantization condition[2], and here we have obtained it as a consequence of the topology of the gauge group. The unit magnetic charge is thus $2\pi/e$.

We can also write down some explicit gauge fields that do the job. For example,

$$\begin{aligned}A_N &\equiv \frac{m}{4\pi}(1-\cos\theta)d\phi \\ A_S &\equiv \frac{m}{4\pi}(-1-\cos\theta)d\phi\end{aligned} \tag{3.44}$$

The point of the "1" in A_N and the "-1" in A_S is so that $A_N = 0$ at $\theta = 0$ and $A_S = 0$ at $\theta = \pi$. Otherwise, they are not defined at $\theta = 0, \pi$ because ϕ and $d\phi$ are not defined there. With the $\pm 1, A_N$ is well-defined on all of N, and A_S is well-defined on all of S.

Then $dA_N = dA_S = \frac{m}{4\pi}\sin\theta d\theta \wedge d\phi = F$. On $\mathbb{R}^3 - \{0\}$ the field strength is a closed 2-form, $dF = 0$, and it is proportional to the volume form on any S^2 surrounding the origin. So

$$\int_{S^2} F = m \tag{3.45}$$

and more generally

$$\int_{M_2} F = m \tag{3.46}$$

on any M_2 homotopic to S^2. This is the magnetic charge enclosed, by the magnetic version of the Gauss law.

The magnetic field is

$$B = {}^{(3)}\star F = \frac{m}{4\pi r^2}dr \tag{3.47}$$

appropriate for a monopole of magnetic charge m.

[2]In natural units. Sometimes the quantization condition is reported in Gaussian units, where both types of charge are scaled up by $\sqrt{4\pi}$, so that the quantization condition (still suppressing $\hbar$s and cs) reads $em \in \mathbb{Z}/2$.

3.3 Monopoles in nonabelian gauge theories

3.3.1 ‘t Hooft-Polyakov monopole

The analysis of monopoles in nonabelian gauge theories is closest to that of the abelian theory in cases where the gauge symmetry is "adjoint Higgsed" to a product of $U(1)$s. Add an adjoint scalar ϕ^a. Then generically $\langle \phi^a \rangle$ breaks $SU(N) \to U(1)^{N-1}$ at low energies. (To see this, write $\phi_{ij} = \phi^a T^a_{ij}$, with $T \in$ Cartan. This can always be done by an $SU(N)$ transformation. Then ϕ_{ij} commutes with all $N-1$ commuting elements of Cartan, but generically no other generators.)

These models have nonsingular monopole solutions. Simplest example: Georgi-Glashow model $SU(2) \to U(1)$. ϕ^a is a triplet of $SU(2)$.

$$\begin{aligned} S &= \int d^4x \left[-\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu a} + \frac{1}{2} (D_\mu \phi)^a (D^\mu \phi)^a - V(\phi) \right] \\ (D_\mu \phi)^a &= \partial_\mu \phi^a + e \epsilon^{abc} A^b_\mu \phi^c \\ V(\phi) &= \frac{\lambda}{4} \left(\phi^a \phi^a - v^2 \right)^2 \end{aligned} \tag{3.48}$$

We look for static, spherically symmetric solutions. IR-finiteness of the energy implies the solutions must fall off like $F^a_{ij} \sim 1/r^2$, $(D^\mu \phi)^a \sim 1/r^2$. Because of the scalar potential, $\phi \to v\hat{\phi}$ exponentially fast, where $\hat{\phi}^a \hat{\phi}^a = 1$. So on a sphere far from the center of the solution, we have

$$\int_{S^2} \epsilon^{abc} \hat{\phi}^a (D\hat{\phi})^b \wedge (D\hat{\phi})^c \Rightarrow 0. \tag{3.49}$$

Expanding it out, one can show that this implies

$$2e \int_{S^2} \hat{\phi}^a F^a = 2e \int_{S^2} \epsilon^{abc} \hat{\phi}^a d\hat{\phi}^b \wedge d\hat{\phi}^c = 8\pi Q[\hat{\phi}] \tag{3.50}$$

The quantity $\hat{\phi}^a F^a$ on the left-hand side is the unbroken $U(1)$ magnetic field. The integral on the right-hand side counts the winding number $Q[\hat{\phi}] \in \mathbb{Z}$ of the S^2 vacuum manifold, parametrized by $\hat{\phi}$ (recall $SU(2)/U(1) \simeq S^2$) over the S^2 at spatial infinity. The winding number is valued in $\mathbb{Z}$ because $\pi_2(S^2) \simeq \mathbb{Z}$. So:

$$\int_{S^2} F_{U(1)} = \frac{4\pi Q}{e} \tag{3.51}$$

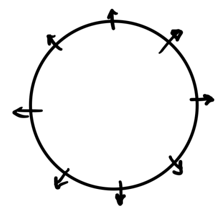
or

$$me = 4\pi\mathbb{Z}. \tag{3.52}$$

where $\int F = m$ is the magnetic charge. Note that in our discussion of the Wu-Yang monopole, 2π appeared on the right-hand side instead of 4π. Evidently, not all monopole charges consistent with Dirac quantization are produced by distinct topological classes in this model. More on this shortly.

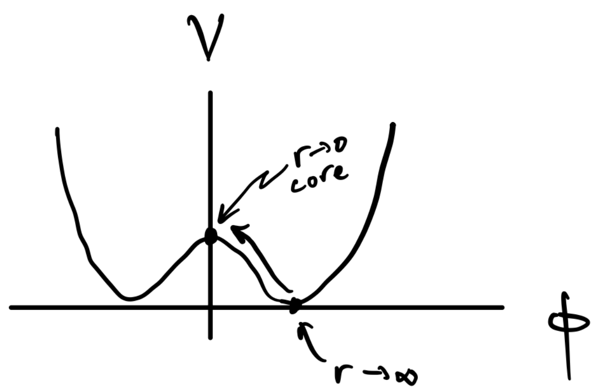
So far we have just discussed the monopole fields far from the core. In the Georgi-Glashow model, something interesting happens: the fields are smooth in the core, and the total energy is finite.

For a monopole of unit magnetic charge, $\hat{\phi}$ is in a "hedgehog" configuration at infinity – everywhere pointing away from the core:



For smoothness, we require $\phi \to 0$ somewhere. If it is only zero at a point, then by spherical symmetry it must be at the origin. So right in the center of the core, the $SU(2)$ is unbroken.

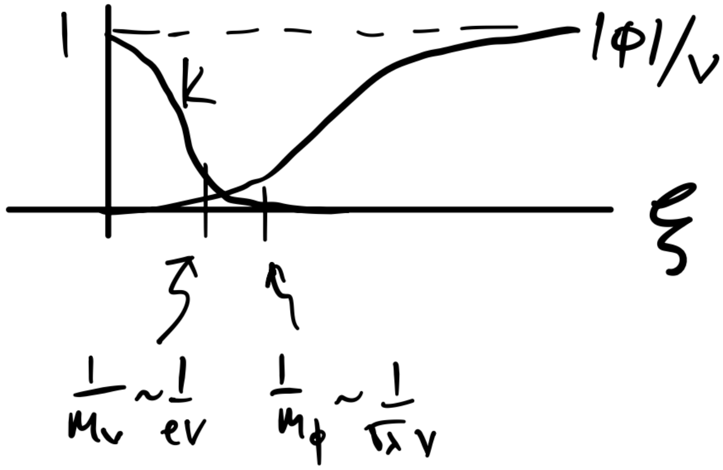
The size of the core can be understood physically as a competition between different terms in the energy. Unbreaking the $SU(2)$ costs energy density $\sim \lambda v^4$. But if ϕ were $v\hat{\phi}$ everywhere, then the gradient energy $\sim v^2/r^2$ would be singular at the origin. As we move inward in radius, the gradient energy grows, and it becomes favorable to "shrink v" – i.e. let $\phi \to 0$.



The field profiles take the form

$$\phi^a = -v\frac{x^a}{r}\left(\frac{H(\xi)}{\xi}\right)$$
$$A_i^a = \epsilon_{aik}\frac{x^k}{er^2}[1-K(\xi)] \tag{3.53}$$

where the functions H and K are functions of the dimensionless variable $\xi = evr$. In general they must be obtained numerically by solving the second order field equations.



The energy is bounded by

$$\begin{aligned}
E &\geq \int d^3x \left[\frac{1}{2}(B_i^a)^2 + \frac{1}{2}((D_i\phi)^a)^2\right] \\
&= \int d^3x \left[\frac{1}{2}(B_i^a - D_i\phi^a)^2 + B_i^a D_i\phi^a\right] \\
&\geq \int d^3x \left[B_i^a D_i\phi^a\right] \\
&= \oint d^2x\, \hat{n}^i B_i^a \phi^a \quad \text{(IBP, use Bianchi } (D_iB_i)^a = 0) \\
&= v \oint F^a \hat{\phi}^a \\
&\geq \frac{4\pi v}{e}.
\end{aligned} \tag{3.54}$$

This is called the BPS bound. It is saturated by the monopole solution in the limiting case $\lambda \to 0$, and then the fields satisfy the first order BPS equations,

$$B_i^a = (D_i\phi)^a. \tag{3.55}$$

These (nine) nonlinear equations actually do admit analytic solutions.

3.3.2 Other monopoles

GNO classification. In unbroken $SU(N)$ we can write down solutions that are monopole-like. It turns out that up to a gauge transformation, all we have to do is take an abelian monopole and multiply it by a generator matrix. However, they are not stable solutions.

First, we should say a word about the use of classical equations of motion. The weak coupling limit is a semiclassical limit. This can be seen by rescaling the canonically normalized gauge field by $A \to A/g$, after which the Yang-Mills action is rescaled from $-\frac{1}{2}\mathrm{Tr}\int d^4xF^2 \to -\frac{1}{2g^2}\mathrm{Tr}\int d^4xF^2$. All of the g dependence sits explicitly in front of the action, where it joins with $\hbar$ to control the semiclassical limit.

Now the YM gauge coupling runs, and in asymptotically free theories it will reach strong coupling eventually. There are two circumstances where we can use classical equations of motion. The first case is when the theory is Higgsed at a scale where the coupling is weak. For example, in the Georgi-Glashow discussion above, the adjoint Higgsing is presumed to take place at weak coupling. The $U(1)$ coupling is matched onto the $SU(2)$ coupling at the scale v, and then since there are no light charged particles, it does not run further in the IR. The second case where can still use classical equations of motion is in regimes of high energy density. A semiclassical description of monopole solutions in pure, unbroken Yang-Mills theory will be valid in an intermediate regime, sufficiently far outside the monopole core where we can take the fields to fall off like $1/r$, but sufficiently close to the core that the energy densities still exceed the confinement scale. Since the confinement scale is exponentially small, there can be a large range of distance scales over which weak coupling and far field approximations are simultaneously valid.

We work in A_0=0 gauge and we seek static solutions to the equations of motion. Actually, it is not essential to fix $A_0 = 0$ gauge; we can assume A_0 vanishes if we restrict attention to time-reversal-invariant static configurations. Then residual time-independent gauge transformations can be used to set $A_r = 0$. We seek $\Omega(r)$ such that

$$\Omega A_r \Omega^{-1} + i\Omega\partial_r\Omega^{-1} = 0. \tag{3.56}$$

A solution is[3]

$$\Omega = Pe^{i\int^r drA_r} \tag{3.57}$$

P denotes path ordering. Now in the far field regime, the gauge field is of the form

$$A_\mu dx^\mu = \frac{1}{r}\left(a_\theta(\theta,\phi)d\theta + a_\phi(\theta,\phi)d\phi\right) + \mathcal{O}(1/r^2) \tag{3.58}$$

where $a_\phi(0,\phi) = a_\phi(\pi,\phi) = 0$ in non-singular configurations. Finally, we can set $a_\theta = 0$ by a θ,ϕ-dependent gauge transformation:

$$\Omega a_\theta \Omega^{-1} + i\Omega\partial_\theta\Omega^{-1} = 0. \tag{3.59}$$

We define the gauge transformation on two patches. In the northern (southern) hemisphere, we integrate from the north (south) pole:

$$\begin{aligned} \Omega_N &= Pe^{i\int_{\theta=0}^{\theta} d\theta a_\theta} \\ \Omega_S &= Pe^{i\int_{\theta=\pi}^{\theta} d\theta a_\theta}. \end{aligned} \tag{3.60}$$

[3]The path ordered exponential is defined as

$$Pe^{i\int^r drA_r} \equiv 1 + \sum_{n=1}^{\infty}\frac{1}{n!}P\prod_{i=1}^{n}\left(\int^r dr_i\, iA_r(r_i)\right) \equiv 1 + \sum_{n=1}^{\infty}(i)^n\int^r dr_1\int^{r_1}dr_2\cdots\int^{r_{n-1}}dr_n\, A_r(r_1)\ldots A_r(r_n)$$

from which it is clear that

$$\partial_r\Omega^{-1} = -iA_r(r)\Omega^{-1}.$$

We are left with $a_\phi(\theta, \phi)$. The general solution of the equations of motion $D_\mu F^{\mu\nu} = 0$ is

$$\begin{aligned} a_\phi^N &= q(1 - \cos\theta) \\ a_\phi^S &= -q(1 + \cos\theta) \end{aligned} \tag{3.61}$$

with constant matrix q. By a global $SU(N)$ transformation, we can diagonalize q, in which case it is a linear combination of Cartan generators.

Thus the general nonabelian monopole solution can be got by taking an Abelian solution and multiplying by a diagonal generator matrix. The matching condition at $\theta = \pi$ is

$$e^{4\pi i q} = \mathbb{I} \tag{3.62}$$

which quantizes the magnetic charge.

But $\pi_1(SU(N)) = 0$, so these are not topologically stable. They are only stabilized if we adjoint Higgs the theory. Nonetheless they play a role in the classification of line and surface operators in the theory.

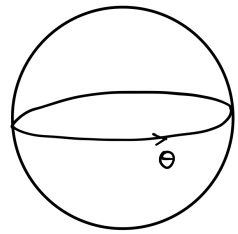
Gauging elements of the center. $\mathbb{Z}_N \in SU(N)$ are elements of the group that commute with all others. We can construct these, for example, by looking at the generator:

$$\begin{aligned} T^{N-1} &\equiv \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1-N \end{pmatrix} \in \mathrm{su}(N) \text{ Cartan} \\ \exp\left(i\alpha T^{N-1}\right) &= \begin{pmatrix} e^{i\alpha} & & & \\ & \ddots & & \\ & & e^{i\alpha} & \\ & & & e^{i\alpha(1-N)} \end{pmatrix} \end{aligned} \tag{3.63}$$

Then for $\alpha = 2\pi k/N$, we find $\exp\left(i\alpha T^{N-1}\right) = e^{2\pi i k/N}\mathbb{I}_{N\times N}$.

"Gauging center" means two elements of the gauge group g_1, g_2 are identified with each other if they satisfy $g_1 = g_2 g_3$, for $g_3 \in \mathbb{Z}_N$. We can gauge all of center – this theory is denoted $SU(N)/Z_N$ – or just a subgroup. An important global difference between $SU(N)$ theory, and $SU(N)/Z_N$ theory, is $\pi_1(SU(N)/Z_N) = Z_N$. So we can form closed, noncontractible loops in the group manifold again, and use them to construct topologically stable monopoles:

$$\begin{aligned} A_\mu^N &= U A_\mu^S U^{-1} + iU\partial_\mu U^{-1} \\ U &= e^{i\theta \frac{k}{N} T^{N-1}} \\ U(\theta = 0) &= U(\theta = 2\pi)\underbrace{e^{-i\pi i k T^{N-1}/N}}_{\in \mathbb{Z}_N.} \end{aligned} \tag{3.64}$$



This is how to recover the original Dirac quantization condition in the Georgi-Glashow model: we have to gauge the center, changing the gauge group to $SU(2)/Z_2 \simeq SO(3)$. This introduces new, topologically stable monopoles with half-integral magnetic charge relative to the $SU(2)$ theory.

The Z_N charge of these monopoles is sometimes called 't Hooft charge, and the charge mod Z_N is sometimes called the GNO charge.

3.4 Polyakov Model

The model we will study is defined in the ultraviolet as $SU(2)$ gauge theory weakly coupled to a Higgs field in the adjoint (vector) representation. This is sometimes called the Georgi-Glashow model, or in 2+1 dimensions, the Polyakov model. The Euclidean action is

$$S = \int d^3x \left(\frac{1}{4e^2}(F^a_{\mu\nu})^2 + (D_\mu \Phi^a)^2 + \frac{1}{4}\lambda((\Phi^a)^2 - v^2)^2 \right). \tag{3.65}$$

Here we use the vector notation for the adjoint rep gauge indices rather than the matrix notation. The covariant derivative is $D_\mu \Phi^a = \partial_\mu \Phi^a + \epsilon^a_{bc} A^b_\mu \Phi^c$. The gauge field and e^2 are of mass dimension one. The scalar field and parameter v are of mass dimension 1/2, and λ is of mass dimension one.

The vacuum manifold is $\Phi^a \Phi^a - v^2 = 0$, or S^2. Without loss of generality let us take $\langle \Phi^a \rangle = v\delta^{a3}$. Then $(D_\mu \Phi^a)^2 \to ((A^1_\mu)^2 + (A^2_\mu)^2)v^2$ in the vacuum. In this case there is a massless photon field A^3_μ, a massive W boson, $W^\pm_\mu = A^1_\mu \pm i A^2_\mu$, $m_W = gv$, and a massive scalar "Higgs boson" generated by radial fluctuations of Φ, of mass $\sqrt{2\lambda}v$. Angular fluctuations of Φ are Goldstone bosons eaten to provide the longitudinal components of the massive Ws.

Perturbatively, the theory appears simple in the IR. The massive W and Higgs decouple and we are left with a free 2+1 dimensional electromagnetism. Nonperturbatively, the theory is much more interesting.

We have discussed the magnetic monopoles of the 3+1 dimensional Georgi-Glashow model. They are static solutions, independent of time. Therefore we can delete the time axis and automatically instanton solutions to the 2+1 dimensional Euclidean equations of motion. These are sometimes called "monopole-instantons" – we are just reinterpreting one of the spatial Cartesian axes of the original monopole solution, say x^3, as Euclidean time. Polyakov calculated the contribution of instanton saddles to correlation functions of Wilson loops in the long-distance $U(1)$ theory. We will follow this derivation.

We will use the singular gauge form of the solutions, which is convenient for superposing multiple instantons. Far from the core at $x_\mu = 0$, we have

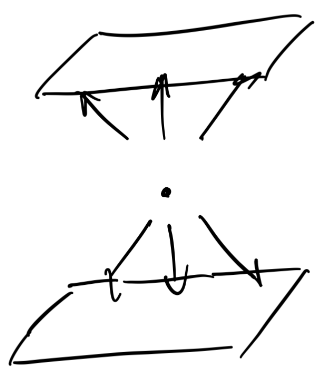
$$\Phi^a \to v\delta^{a3}, \quad W^\pm_\mu \to 0,$$
$$B_\mu \equiv \tilde{F}^3_\mu = \epsilon_{\mu\nu\rho} F^3_{\nu\rho} \to \frac{x_\mu}{2r^3} - 2\pi\delta_{\mu 3}\theta(x_3)\delta(x_2)\delta(x_1) \tag{3.66}$$

The dual field strength $B_\mu \equiv \tilde{F}^3_\mu$ is a vector in 2+1 dimensions. It falls off like $1/r^2$ apart from a Dirac string singularity running along the positive x_3 axis. The massive fields relax to the vacuum.

An aside: if instantons describe tunneling, what states do the monopole instantons tunnel between? In Lorentzian signature, continuing $x^3 \to ix^3$, the $\mu = 3$ component becomes a scalar magnetic field. We can integrate it over a spatial slice. Let $n^\mu = \delta^{\mu 3}$ be the unit normal to the time slices. Then for large $|x_3|$,

$$\frac{1}{2\pi}\int d^2x n^\mu \tilde{F}^3_\mu \longrightarrow -\theta(x_3). \tag{3.67}$$

The instanton is tunneling between states of different (integer valued) magnetic flux.



Each instanton represents a separate saddle point contribution to the path integral. We also expect to have approximate saddles when we superpose widely-separated instantons– this is the dilute instanton gas. Here 'widely-separated' means compared to the core size. We will have to sum over all of these saddles. Suppose we have a dilute gas of monopoles and antimonopoles at positions x^μ_a, $a = 1 \ldots N$, with $|x_a - x_b| \gg 1/m_W$. Then the "magnetic field" is approximately

$$B_\mu = \sum_{a=1}^N \frac{1}{2} q_a \frac{(x^\mu - x^\mu_a)}{|x - x_a|^3} - 2\pi\delta_{\mu 3}\sum_{a=1}^N q_a \theta(x^3 - x^3_a)\delta(x^2 - x^2_a)\delta(x^1 - x^1_a). \tag{3.68}$$

Here $q_a = +1$ for monopoles and $q_a = -1$ for antimonopoles. The euclidean action is the static energy of the monopole gas in 3+1, or the volume integral of $(F^3_{\mu\nu})^2$. There are two contributions: the self-energies of each separate monopole, and the magnetic coulomb potential energy of each pair. The self energies are of the form $\int d^3x 1/r^4$ and are thus UV dominated, cut off by the smooth nonabelian fields in the core. We already calculated this: the monopole mass is $\sim v^2$. The interaction energies are

$$\sum_{a\neq b} \frac{q_a q_b}{8e^2}\int d^3x \frac{(x^\mu - x^\mu_a)}{|x - x_a|^3}\frac{(x^\mu - x^\mu_b)}{|x - x_b|^3} = \frac{\pi}{2g^2}\sum_{a\neq b}\frac{q_a q_b}{|x_a - x_b|}. \tag{3.69}$$

Putting the pieces together, the action of the collection is

$$S \approx N\frac{m_W}{g^2}f(\lambda/e^2) + \frac{\pi}{2e^2}\sum_{a,b,a\neq b}\frac{q_a q_b}{|x_a - x_b|}. \tag{3.70}$$

The path integral sums over all N, charges q_a, and locations x_a^μ. There is an interesting technical challenge in figuring out the measure for the x_a^μ integration. We will postpone sorting this out and just write it as cd^3x_a. Then the dilute instanton gas sum contributes to the total partition function:

$$\begin{aligned} Z &= \sum_{N,\{q_a\}}\frac{1}{N!}\left(\int cd^3x_1\cdots\int cd^3x_N e^{-S}\right) \\ &= \sum_{N,\{q_a\}}\frac{\left(ce^{-\frac{m_W}{e^2}f(\lambda/e^2)}\right)^N}{N!}\left(\int d^3x_1\cdots\int d^3x_N e^{-\frac{\pi}{2e^2}\sum_{a,b,a\neq b}\frac{q_a q_b}{|x_a - x_b|}}\right) \end{aligned} \tag{3.71}$$

We can think of the Coulomb term in parenthesis as JGJ, where J is the charge density and G is a Green function. In this case, since the charges are magnetic, G is not the two-point function of the ordinary vector potential, which couples to electric charges, but rather the two-point function of some dual magnetic vector potential. We will see this intuition borne out more precisely momentarily. Mathematically we proceed as follows. Recall that the generating functional for connected Euclidean correlation functions of a free scalar field is

$$\begin{aligned} -\log Z[J] &= -\log\int\mathcal{D}\varphi e^{-\int_M\left(\frac{1}{2}(\nabla\varphi)^2 + J\varphi\right)} \\ &= -\frac{1}{2}\mathrm{Tr}\log(-\nabla^2) + \int\int JGJ. \end{aligned} \tag{3.72}$$

Let us divide $Z[J]/Z[0]$, or simply redefine the measure $\mathcal{D}\varphi$ in $Z[J]$ with a normalization factor such that $Z[0] = 1$. Then, with this normalization,

$$-\log Z[J] = \int\int JGJ. \tag{3.73}$$

Now the Green function for the Laplace operator on R^3 is

$$G(x;y) = -\frac{1}{4\pi|x-y|} \tag{3.74}$$

so, on R^3,

$$Z[J] = e^{\frac{1}{4\pi}\int d^3x d^3y\frac{J(x)J(y)}{|x-y|}}. \tag{3.75}$$

So if we let

$$\begin{aligned} J &\to i\frac{\sqrt{2}\pi}{e}\rho(x) \\ \rho(x) &\equiv \sum_a q_a\delta^3(x - x_a) \end{aligned} \tag{3.76}$$

where $\rho(x)$ is the monopole charge density, we can rewrite the Coulomb term in the partition function as

$$\int d^3x_1 \cdots \int d^3x_N e^{-\frac{\pi}{2e^2}\sum_{a,b,a\neq b}\frac{q_aq_b}{|x_a-x_b|}} \to \int d^3x_1 \cdots \int d^3x_N \int \mathcal{D}\varphi e^{-\int d^3x \frac{1}{2}(\nabla\varphi)^2} e^{i\frac{\pi}{e}\sum_a q_a\varphi(x_a)}$$
$$= \int \mathcal{D}\varphi e^{-\int d^3x\frac{1}{2}(\nabla\varphi)^2}\prod_{a=1}^{N}\int d^3x_a e^{i\frac{\sqrt{2}\pi}{e}q_a\varphi(x_a)}. \quad (3.77)$$

We have used an arrow instead of an equals sign because the two sides are not quite the same: the sum in the left-hand side omits $a = b$, while the right hand side gives rise to an $\int JGJ$ of the form $-\int d^3xd^3y\sum_{a,b} q_aq_b\delta^3(x-x_a)\delta^3(y-x_b)\frac{1}{|x-y|}$, where $a = b$ terms are included, and quite divergent. Fortunately, such terms are just monopole self-energy contributions. We can regulate the delta functions (in an ad hoc way, or by considering the full $SU(2)$ theory) and then we see that the $a = b$ terms are translation invariant and just provide a magnetostatic contribution to the mass of each monopole (or the action of a single monopole-instanton.) Physically, the N divergences are cut off at the finite core radius, and they are absorbed by renormalization of the Boltzmann term $e^{-N\frac{m_W}{e^2}f(\lambda/e^2)}$ in the partition function. Henceforth we take it as a definition of the functional integral that these divergences will be subtracted.

Suppose we also add an external (background) gauge field $\tilde{A}$ coupling to the monopole charge density ρ. Then Eq. (3.78) becomes

$$\int d^3x_1 \cdots \int d^3x_N e^{-\frac{\pi}{2e^2}\sum_{a,b,a\neq b}\frac{q_aq_b}{|x_a-x_b|}} e^{i\int d^3x\tilde{A}(x)\rho(x)} = \int \mathcal{D}\varphi e^{-\int d^3x\frac{1}{2}(\nabla\varphi)^2}\prod_{a=1}^{N}\int d^3x_a e^{i\frac{\sqrt{2}\pi}{e}q_a(\varphi(x_a)+\tilde{A}(x_a))}. \quad (3.78)$$

This makes it clear that φ is indeed the dynamical part of the background "magnetic" gauge field $\tilde{A}$ (which is of mass dimension zero, if you're checking dimensions.)

Plugging into Eq. (3.71) (to which we also add the background $\tilde{A}$), we have

$$\begin{aligned}
Z[\tilde{A}] &= \sum_{N,\{q_a\}}\frac{\left(ce^{-\frac{m_W}{e^2}f(\lambda/g^2)}\right)^N}{N!}\int \mathcal{D}\varphi e^{-\int d^3x\frac{1}{2}(\nabla\varphi)^2}\prod_{a=1}^{N}\int d^3x_a e^{i\frac{\sqrt{2}\pi}{e}q_a(\varphi(x_a)+\tilde{A}(x_a))} \\
&= \sum_{N}\frac{\left(ce^{-\frac{m_W}{e^2}f(\lambda/g^2)}\right)^N}{N!}\int \mathcal{D}\varphi e^{-\int d^3x\frac{1}{2}(\nabla\varphi)^2}\prod_{a=1}^{N}\int d^3x_a 2\cos\left(\frac{\sqrt{2}\pi}{e}\varphi(x_a)+\tilde{A}(x_a)\right) \\
&= \int \mathcal{D}\varphi e^{-\int d^3x\frac{1}{2}(\nabla\varphi)^2}\sum_{N}\frac{\left(2ce^{-\frac{m_W}{e^2}f(\lambda/g^2)\int d^3x\,\cos\left(\frac{\sqrt{2}\pi}{e}\varphi(x)+\tilde{A}(x)\right)}\right)^N}{N!} \\
&= \int \mathcal{D}\varphi e^{-\int d^3x\left[\frac{1}{2}(\nabla\varphi)^2-2ce^{-\frac{m_W}{e^2}f(\lambda/g^2)}\cos\left(\frac{\sqrt{2}\pi}{e}\varphi(x)+\tilde{A}(x)\right)\right]} \\
&= \int \mathcal{D}\varphi e^{-\frac{e^2}{2\pi^2}\int d^3x\left[\frac{1}{2}(\nabla(\varphi-\tilde{A}))^2-M^2\cos(\varphi)\right]}
\end{aligned} \quad (3.79)$$

In the last line we have shifted and rescaled the integration variable φ so that it is dimensionless, dropping a constant rescaling of the measure which will not matter for our purposes. Remarkably, the partition function is equivalent to that of a scalar field with a cosine potential, and the scalar has an interpretation as a sort of dual gauge field, coupling to magnetic charge instead of electric. The magnitude of the potential, M^2, is exponentially small and nonperturbative, $M^2 \sim ce^{-\frac{m_W}{e^2}f(\lambda/e^2)}$.

Had we not known about monopole-instantons, we would have expected the infrared $U(1)$ functional integral to look like that of a free massless scalar field. The reason is we could dualize the Maxwell field to a vector, $B_\mu = \epsilon_{\mu\nu\rho}F^{\nu\rho}$, and then locally write $B_\mu = \partial_\mu\varphi$. φ is the dual gauge field coupling to monopole charge. The effect of the instantons is to make the dual gauge field massive.

We might expect that if the magnetic gauge field is massive, the electric gauge field is confining, realizing 't Hooft's idea of "confinement as a dual Meissner effect." To show this explicitly we can calculate the expectation value of the electric $U(1)$ Wilson operator on a loop C bounding a region S in the $x^1 - x^2$ plane,

$$\begin{aligned} W[C] &= \langle e^{i\int_C A^3_\mu dx^\mu}\rangle \\ &= \langle e^{i\int_S dx^1dx^2 B^3}\rangle \\ &= \langle e^{i\int_S dx^1dx^2 \int d^3y G(\vec{x},\vec{y})\rho(\vec{y})}\rangle \\ &= \langle e^{i\int d^3y \tilde{A}(\vec{y})\rho(\vec{y})}\rangle \end{aligned} \tag{3.80}$$

where

$$\tilde{A}(\vec{y}) = \int_S dx^1dx^2 G(\vec{x},\vec{y}) = \int_S dx^1dx^2 \frac{(x^3-y^3)}{|\vec{x}-\vec{y}|^3} \tag{3.81}$$

Apparently, in order to evaluate $W[C]$, all we need to do is evaluate the generating functional (3.79) for the particular field (3.81). In the semiclassical limit, we must solve the Sine-Gordon-type equation:

$$\nabla^2(\varphi - \tilde{A}) = M^2\sin(\varphi). \tag{3.82}$$

The source term is

$$\nabla^2_y \int_S dx^1dx^2 \frac{(x^3-y^3)}{|\vec{x}-\vec{y}|^3} = 2\pi\delta(y^3)\theta_S(y^1,y^2) \tag{3.83}$$

where $\theta_S(y^1,y^2) = 1$ for points in S, and zero otherwise. Deep in the interior of S, far from its boundary C, we may approximate the equation of motion by a one-dimensional equation,

$$\partial_3^2\varphi(y^3) - M^2\sin(\varphi(y^3)) = -2\pi\delta(y^3) \tag{3.84}$$

The solution is

$$\varphi(y^3) = 4\mathrm{sgn}(y^3)\tan^{-1}\left(e^{-M|y^3|}\right). \tag{3.85}$$

It is straightforward to verify that $W[C] \sim e^{-aMA_S}$, where A_S is the area of S, a is a constant, and M is the constant above, of order e^{-m_W/e^2}.

So Maxwell theory, supplemented by monopoles, is a linearly confining theory in 2+1 dimensions.

3.5 Dyons

We are studying heavy, stable objects. EFT methods can also be used to study the kinematics and dynamics of low-energy excitations around some fixed number of these objects. Here we will discuss the low-lying excitations of the 't Hooft-Polyakov monopole.

The starting point is symmetry. We can find low-energy excitations by reasoning analogous to Goldstone's theorem. Here is the procedure:

- Identify continuous symmetries that act trivially on the vacuum, but nontrivially on the heavy state. In the present case if the static monopole solution is not invariant under some of the symmetries of the vacuum, then acting on the monopole state with those transformations will produce a new, degenerate solution. The parameters of such transformations are called "moduli."
- Allow the symmetry transformation parameters to vary slowly with the worldvolume coordinates of the solution. A monopole is a particle, so this means we let the symmetry transformation parameters acquire some slow time dependence. Its moduli are now some quantum mechanical degrees of freedom.
- Plug the transformation, with slowly varying parameters, into the action, and integrate over coordinates orthogonal to the worldvolume of the object. The result is an effective action for the moduli, from which we can derive an effective Hamiltonian. Classically, the moduli excitations are ungapped, due to the symmetry. Quantum mechanically, we have to determine the spectrum of the effective Hamiltonian.

In the Georgi-Glashow model, the vacuum preserves a global $U(1)$ symmetry (electromagnetism) and the Poincare group. Of these symmetries, the monopole preserves only rotations and time translations. Its moduli are associated with spatial translations – moving the monopole produces an equally good monopole solutions – and, less trivially, $U(1)_{EM}$. The reason it is not invariant under $U(1)_{EM}$ is that the heavy charge W boson fields are excited in the core.

Another way to go about studying excitations is to consider an arbitrary variation,

$$A^a_\mu = A^{a(0)}_\mu + \epsilon a^a_\mu, \quad \phi^a = \phi^{a(0)} + \epsilon \varphi^a \tag{3.86}$$

where $A^{a(0)}_\mu$ and $\phi^{a(0)}$ are the 't Hooft-Polyakov solution, and a^a_μ and φ^a are small fluctuations with bookkeeping parameter ϵ. We then plug (3.86) into the Georgi-Glashow action and expand to quadratic order in ϵ. This procedure finds all excitations, gapped and ungapped. (Care must be taken to gauge fix, so that unphysical ungapped excitations are removed.) By focusing on the moduli, we get just the lowest energy excitations, which are protected by the symmetries.

The translation moduli of the monopole are incorporated by the shift $\vec{x} \to \vec{x} + \vec{x}_0(t)$. $\vec{x}_0(t)$ is the slowly-varying "collective coordinate" describing the center of mass of the monopole. We can guess the answer for its effective action, although it takes some work to actually derive it:

$$L_{eff}[\vec{x}_0(t)] = \frac{1}{2}\Lambda \dot{x}_0^2 \tag{3.87}$$

where Λ is the monopole mass. The low-lying excitations associated with translations just describe free motion of the monopole, and the eigenstates of the effective Hamiltonian are labeled by momentum, with nonrelativistic dispersion relation.

An infinitesimal global electromagnetic transformation is of the form (3.86) with

$$a_\mu^a = \frac{\alpha}{v}(D_\mu^{(0)}\phi^{(0)})^a, \qquad \varphi = 0. \tag{3.88}$$

with constant, $U(1)$-valued α. We promote $\alpha \to \alpha(t)$ and plug into the action. Terms of order α vanish by the equations of motion, satisfied by the monopole fields $A_\mu^{a(0)}$, $\phi^{a(0)}$. Terms of order α^2 almost all vanish because the transformation is a symmetry for constant α.[4] The only terms that survive contain time derivatives of α. These come from terms in F^2 of the form

$$\left[\partial_\mu(\alpha/v(D_\nu^{(0)}\phi^{(0)})^a) - \partial_\nu(\alpha/v(D_\mu^{(0)}\phi^{(0)})^a)\right]^2. \tag{3.89}$$

The cross terms in this expression vanish, because α is independent of space, $\phi^{(0)}$ is independent of time, and $A_0^{(0)} = 0$. All that survives is

$$\delta\mathcal{L} = \frac{1}{2v^2}(\dot{\alpha})^2((D_i^{(0)}\phi^{(0)})^a)^2. \tag{3.90}$$

The effective action for α is

$$\begin{aligned} S[\alpha] &= \frac{c}{2}\int dt\,(\dot{\alpha})^2 \\ c &\equiv \frac{1}{v^2}\int d^3x\,((D_i^{(0)}\phi^{(0)})^a)^2 \end{aligned} \tag{3.91}$$

In the BPS case, the spatial integral is just Λ/e^2, so $c = \Lambda/m_W^2$. We have learned that there is a low-energy excitation of the monopole that behaves like a quantum mechanical rotor (since $\alpha \sim \alpha + 2\pi$)!

The meaning of the rotor excitations is quite interesting. First, they are quantized. The conjugate momentum is

$$\pi_\alpha = \frac{\delta L_\alpha}{\delta\dot{\alpha}} = c\dot{\alpha} \tag{3.92}$$

so the effective Hamiltonian is

$$H_\alpha = \frac{1}{2c}\pi_\alpha^2. \tag{3.93}$$

[4] The transformation (3.88) has corrections of order α^2, but to obtain the change in the action to order α^2, only the correction to the fields of order α is needed, since we are expanding around a solution.

On the Hilbert space $L^2(S^1)$ the momentum acts as $-i\partial_\alpha$. So the eigenstates are $\psi_n \propto e^{in\alpha}$ with integer n, and the energy spectrum is $E_n = \frac{1}{2c}n^2$. Quantum mechanically, there is a gap, but it is very small at weak coupling, since $m_W/\Lambda \sim e^2/4\pi \ll 1$..

Secondly, the rotor excitations generate an *electric* field. Since $A_0^a = 0$, the unbroken electric field is

$$
\begin{aligned}
\vec{E}_{EM} &= \frac{1}{v}\vec{E}^a\phi^a \\
&= \frac{1}{v}\dot{\vec{A}}^a\phi^a \\
&= \frac{1}{v^2}\dot{\alpha}\phi^{a(0)}(D_i^{(0)}\phi^{(0)})^a \\
&= \frac{1}{cv^2}\pi_\alpha\phi^{a(0)}(D_i^{(0)}\phi^{(0)})^a \\
&\to \frac{n}{cv^2}\phi^{a(0)}(D_i^{(0)}\phi^{(0)})^a
\end{aligned}
\tag{3.94}
$$

where, in the last line, we replace π_α by its eigenvalue in the rotor energy eigenstates. Evidently, the rotor excitations endow the monopole with electric charge! These objects are called dyons.

In the BPS case, $\phi^{a(0)}(D_i^{(0)}\phi^{(0)})^a = \phi^{a(0)}B_i^{(0)}$, the unbroken magnetic field. So we may obtain the charge from the electric and magnetic Gauss laws,

$$
q_e = \frac{nm}{cv} = ne. \tag{3.95}
$$

where m is the magnetic charge and we have used the minimal BPS monopole charge, $m = 4\pi/e$. n is just telling us the (integer) charge of the dyon! So the $SU(2)$ GG model contains magnetic charges, electric charges, and objects charged under both.

There is a dyon quantization condition, generalizing Dirac quantization, due to Schwinger and Zwanziger. The global structure of the gauge group plays a somewhat subtle role, which we will not go into in detail; for a precise discussion, see `https://arxiv.org/pdf/1305.0318`. One way to obtain the condition is to look at a dyon of charge q_e, q_m moving the background of another dyon of charge q_e', q_m', compute the angular momentum, and demand that it is quantized in units of $2\pi\hbar$. The result is

$$
q_e q_m' - q_e' q_m = 2\pi\mathbb{Z}. \tag{3.96}
$$

For example, take the 't Hooft-Polyakov monopole in $SU(2)$, of magnetic charge $q_m = 4\pi/e$ and electric charge $q_e = 0$. Here the quantization condition is compatible with elementary electric charges of charge $q_e = \mathbb{Z}e/2$: these are exactly what we would have found, at low energies, if we had added matter fields carrying the fundamental rep of $SU(2)$. So the allowed charge lattice of the theory is $(q_e/e, q_m e/2\pi) = (\mathbb{Z}/2, 2\mathbb{Z})$.

If the gauge group is $SU(2)/Z_2 \simeq SO(3)$, only elementary electric charges of charge $q_e = \mathbb{Z}e$ are allowed (e.g. the adjoint of $SU(2)$, which upon higgsing produces charge-1 W bosons.) The theory contains a monopole of $(q_e/e = 0, q_m e/2\pi = 1)$ ('t Hooft charge 1, in the language of our previous discussion). Together these states saturate the quantization condition and give rise to a charge lattice $(q_e/e, q_m e/2\pi) = (\mathbb{Z}, \mathbb{Z})$.

The charge quantization condition suggests another possibility: to have only dyons of charge $(q_e/e = 1/2, q_m e/2\pi = 1)$ together with monopoles of charge $(q_e/e = 0, q_m e/2\pi = 2)$, as generators of the charge lattice. Is this realized in a model?

3.5.1 Witten effect

Suppose we add a θ term to the Georgi-Glashow model. What impact does it have on the monopole?

The θ term is topological, so it does not affect the equations of motion, and any solution – including the monopole – obtained for $\theta = 0$ remains a solution for nonzero θ. Likewise, it appears as a total derivative in the effective α theory:

$$\begin{aligned} \mathcal{L} &\supset -\frac{\theta}{8\pi^2} \\ \Rightarrow L_\alpha &\supset -\frac{\theta\dot{\alpha}}{8\pi^2 v}\int d^3x\,(D_i\phi)^a B^{ia} \\ &= \frac{\theta\dot{\alpha}}{8\pi^2}\oint d^2x\,\hat{\phi}^a B^{ia} \\ &= \frac{\theta\dot{\alpha}}{2\pi}. \end{aligned} \tag{3.97}$$

So

$$S[\alpha] = \int dt\,[\frac{1}{2m}(\dot{\alpha})^2 + \frac{\theta}{2\pi}\dot{\alpha}]. \tag{3.98}$$

What the θ term does affect is the conjugate momentum, and thus the spectrum. The velocity-momentum relation is now

$$\pi_\alpha = \frac{\dot{\alpha}}{m} + \frac{\theta}{2\pi} \tag{3.99}$$

so our electric field computation is modified:

$$\begin{aligned} \vec{E}_{EM} &= \frac{1}{v}\vec{E}^a\phi^a \\ &= \frac{1}{v}\dot{\vec{A}}^a\phi^a \\ &= \frac{1}{v^2}\dot{\alpha}\phi^{a(0)}(D_i^{(0)}\phi^{(0)})^a \\ &= \frac{m}{v^2}\left(\pi_\alpha - \frac{\theta}{2\pi}\right)\phi^{a(0)}(D_i^{(0)}\phi^{(0)})^a \\ &\to \frac{1}{cv^2}\left(n - \frac{\theta}{2\pi}\right)\phi^{a(0)}(D_i^{(0)}\phi^{(0)})^a. \end{aligned} \tag{3.100}$$

Therefore

$$q_e = \frac{n}{cv}(n - \theta/2\pi) = .e(n - \theta/2\pi) \tag{3.101}$$

again using the minimal BPS monopole charge $m = 4\pi/e$. Evidently, θ automatically endows the monopole with electric charge, even for $n = 0$!

The electric-magnetic charge lattice exhibits an interesting periodicity. θ is an angle in weakly-coupled theories like the GG model. The periodicity $\theta \sim \theta + 2\pi$ is realized by a monodromy in the charge lattice. For $\theta = 0$, for example, under $\theta \to \theta + 2\pi$, an ordinary monopole becomes a dyon with electric charge $+e$, a dyon of electric charge $-e$ becomes an ordinary monopole, and so on.

Something slightly different happens in the $SO(3)$ theory. Here,

$$q_e = \frac{n}{cv}(n - \theta/2\pi) = e/2(n - \theta/2\pi) \tag{3.102}$$

using the minimal monopole charge $m = 2\pi/e$. In this theory, under $\theta \to \theta + 2\pi$, the monopole of magnetic charge $2\pi/e$ becomes a dyon with electric charge $+e/2$. There were no half-integral electric charges, fundamental or dyonic, in the original theory! This is a different quantum theory with a different charge lattice of states.

This provides an answer to the question raised above: if we start in the $SO(3)$ theory (sometimes called $SO(3)_+$) with electric charges $= e$ and monopoles of charge $q_m e/2\pi = 1$, then we take $\theta \to \theta + 2\pi$, we will arrive at the other consistent theory ($SO(3)_-$) that we anticipated from the quantization condition.

If we do another $\theta \to \theta + 2\pi$ shift, we are back to states of integral electric charge and the charge lattice of $SO(3)_+$ again. In $SO(3)$ gauge theory, θ is 4π-periodic.

3.6 Instantons in Yang-Mills Theory and QCD

In our discussion of the Polyakov model, we saw that the static monopole solutions of 3+1 dimensional theories can become instantons – localized, finite-action solutions of the Euclidean field equations – if we simply remove the time direction. In 2+1 dimensions monopoles can cause confinement. Are there similar solutions to the 3+1D Euclidean field equations? What are their roles?

First, topology. Classically, the vacuum is $A_\mu = 0$, up to a gauge transformation. A finite-action solution must approach $A_\mu = i\Omega\partial_\mu\Omega^{-1}$ at Euclidean infinity. Let's look at pure $SU(2)$ gauge theory. The group manifold is $SU(2) \simeq S^3$. On $\mathbb{R}^4$, Euclidean infinity is also an S^3. So we should consider homotopy classes of maps from the S^3 at infinity to the S^3 group manifold. As we have seen $\pi_3(S^3) = \mathbb{Z}$, so we should look for classes of solutions that differ by an integer winding number.

This number is computed by the integral of a "topological charge density" $F\tilde{F}$:

$$\int d^4x\, F^{a\mu\nu}\tilde{F}^a_{\mu\nu} = \int d^4x\, \partial_\mu K^\mu = \int_{S^3} n_\mu K^\mu \tag{3.103}$$

where the Chern-Simons form is

$$K^\mu = \epsilon^{\mu\nu\rho\sigma}\left(A^a_\nu F^a_{\rho\sigma} - \frac{2}{3}f^{abc}A^a_\nu A^b_\rho A^c_\sigma\right). \tag{3.104}$$

The flux of K integrated over the sphere at infinity computes the winding number, which is called the topological charge,

$$\frac{1}{32\pi^2}\int d^4x (F\tilde{F}) = n. \tag{3.105}$$

Now $\mathrm{Re}(S_E) = \int d^4x \frac{1}{4g^2}FF$, and since

$$\begin{aligned}\int (F \pm \tilde{F})^2 &= \int \left(F^2 + \tilde{F}^2 \pm 2F\tilde{F}\right) \\ &= \int \left(2F^2 \pm 2F\tilde{F}\right) \geq 0\end{aligned} \tag{3.106}$$

we learn that $\int F^2 \geq \pm \int F\tilde{F}$, which must be true for either sign, so $\int F^2 \geq \left|\int F\tilde{F}\right|$ or

$$\mathrm{Re}(S_E) \geq \frac{8\pi^2|n|}{g^2}. \tag{3.107}$$

The bound is saturated for self-dual (anti-self-dual) configurations, for which $F = \pm\tilde{F}$. These are automatically solutions to EOM since they saturate the bound on $\mathrm{Re}S_E$ in a given topological sector – the action cannot be reduced by small variations, so it is automatically a stationary point.

A solution with $n = 1$, centered at the origin, is

$$A_\mu = g_1^{-1}(\hat{x})\partial_\mu g_1(\hat{x})\frac{x^2}{x^2+\rho^2} \tag{3.108}$$

$$g_1 = \hat{x}_\mu \sigma^\mu. \tag{3.109}$$

As discussed previously, $A \sim 1/x$ at large $|x|$, but the field strength behaves as

$$F \sim \frac{\rho^2}{(x^2+\rho^2)^2} \sim \frac{1}{x^4} \tag{3.110}$$

so the action is finite. This is called the “BPST” instanton. It is a solution of the nonlinear Euclidean field equations of $SU(2)$ Yang-Mills theory. It automatically supplies solutions for other gauge groups, by identifying $SU(2)$ subgroups embedded in them. ρ is an arbitrary parameter called the size modulus. It appears because the classical theory is scale invariant.

In the purely bosonic theory, what does the instanton describe? Consider $\mathbb{R}^3 \times \mathbb{R}^1$. States on $\mathbb{R}^3$ are finite energy $A^i \to$ const at spatial infinity in $A^0 = 0$ gauge. If $A^i \to$ const at spatial ∞, we can treat ∞ as a single point! This compactifies $R^3 \to S^3$. (This is a different S^3 from Euclidean infinity.) The pure-gauge states that can't be deformed into each other are vacua of different $k \in \pi_3(S^3)$, where this is the "spatial S^3": The BPST instanton describes tunneling from $k \to k+1$ (more generally, $(k \to k+n)$.)

Instanton amplitudes are proportional to e^{-S_E}. Above we computed $\text{Re}(S_E) = 8\pi^2|n|/g^2$. There is also an imaginary contribution from the theta term:

$$S_E \supset -i\frac{\theta}{32\pi^2}\int(F^a\tilde{F}^a) = -in\theta \tag{3.111}$$

so each topological sector is weighted by

$$e^{-\frac{8\pi^2|n|}{g^2}+in\theta}. \tag{3.112}$$

Fermions

In a theory with vectorlike fermions, like QCD, we have the chiral anomaly $\partial_\mu J^{\mu 5} = \frac{g^2 F^a\tilde{F}_a}{32\pi^2}$. Then the instanton amplitudes explicitly violate the anomalous symmetry:

$$\begin{aligned}&\int \partial_\mu J^{\mu 5} d^4x = \int d^3x J^{05}\Big|_{t_i}^{t_f} + (\text{ zero if no current through spatial } \infty)\\ &= \Delta Q^5 = n\end{aligned} \tag{3.113}$$

Note that while we could see $\partial_\mu J^{\mu 5} \neq 0$ via a one-loop computation, because the result was a total derivative, a nonperturbative computation was required to see that the ABJ anomaly actually results in nonconservation of the charge.

The Atiyah-Singer index theorem states that:

$(n_L - n_R)$ zero modes of Dirac operator = (Dynkin index of irrep) $\times$ (Pontrjagin number of the background gauge field.)

$$(n_L - n_R)\Delta Q = \int \partial_\mu J^\mu = \frac{c}{16\pi^2}\int F^a\tilde{F}^a = 2cn \tag{3.114}$$

So the anomaly coefficient $c = (\#zm)/2n$.

Example: one SU(2)-fundamental Dirac fermion of charge 1 under an anomalous global axial $U(1)$:

$$\begin{aligned}&c = \text{Tr}(QT^aT^b) = \text{Tr}(Q)\text{Tr}(T^aT^b) = (2)(1/2) = 1\\ &\#zm = 2 \text{ for } n = 1.\end{aligned} \tag{3.115}$$

The presence of Dirac zero modes means that instanton amplitudes vanish unless there are enough insertions of fermion operators to soak up the zero modes. These can be put in by hand, to make a correlation function, or they can come from mass insertions, etc. We can also view small instantons as generating effective vertex:

$$\mathcal{L}^{eff} \supset e^{-\frac{8\pi^2}{g^2}+i\theta} \det(\psi_a \psi_b) \tag{3.116}$$

for left-handed Weyl fermions ψ. Tying off zero modes with Dirac masses, the vacuum-vacuum amplitude is proportional to $e^{-\frac{8\pi^2}{g^2}+i\theta} \det(m^*_{ab})$.

So the vacuum energy depends on the physical combination $\bar{\theta} = \theta - \arg\det m$, as we anticipated previously. flipped sign?

Can we compute $m^2_{\eta'}$ from instantons? Recall that we have to sum over all semiclassical solutions. This sum includes $\int d_\rho$, where ρ is the size modulus:

$$\begin{aligned}
\Delta m^2_{\eta'} &\sim \int_0^\infty d\rho e^{-8\pi^2/g^2(\rho^{-1})} \rho^{-3} \\
&= \int_0^\infty d\rho \, (\Lambda\rho)^{b_0} \, \rho^{-3} \\
&= \Delta m^2_{\eta'} \sim \int_0^\infty d\rho \Lambda^9 \rho^6
\end{aligned} \tag{3.117}$$

This is strongly IR divergent.

However, QCD is not semiclassical on long distance scales. The structure of the vacuum state is not close to the classical vacuum $A_\mu = 0$ on distance scales longer than $1/\Lambda$, and instanton configurations are not distinguished compared to other configurations. We can cut off the size modulus integral at $\rho \sim 1/\Lambda$, but we must conclude that instantons only suggest qualitative behaviors of the QCD vacuum. However, nonperturbative effects at strong coupling are clearly violating $U(1)_A$, since η' is so heavy.

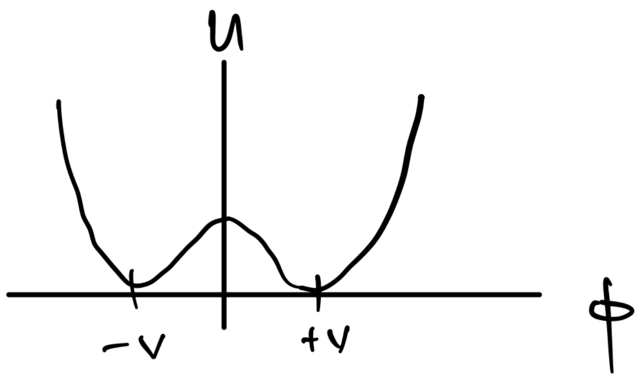
Higgs theories

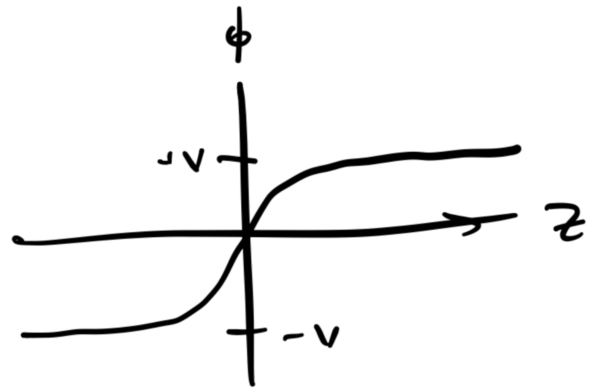
There are circumstances where instanton computations are reliable. There are some axial symmetry violating correlation functions in QCD which are dominated by small instantons, $\rho \ll 1/\Lambda$, where the coupling is small and semiclassical approximations are valid. Instantons are also meaningful in Higgs theories, where the gauge bosons gain mass at weak coupling.

3.7 Domain walls

Domain walls are finite-tension (energy/unit area) objects interpolating between degenerate vacua. Stable walls are associated with spontaneously borken discrete symmetries. The canonical example is a real scalar field with a global Z_2 symmetry:

$$S = \int \frac{1}{2}(\partial\phi)^2 - \frac{g^2}{4}(\phi^2 - v^2)^2. \tag{3.118}$$

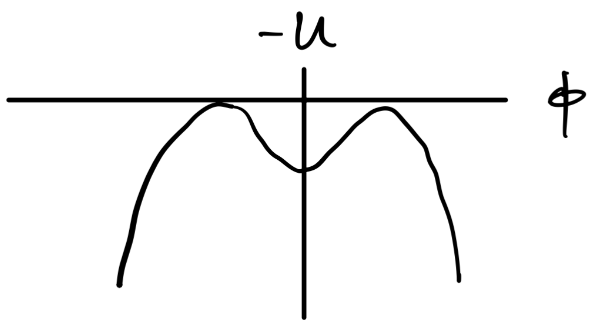




The $\phi \to -\phi$ symmetry is spontaneously broken in the vacuum state. The equation of motion is $\partial_t^2\phi - \nabla^2\phi = -U'(\phi)$. We look for static, 1D solutions interpolating from $\pm v$ to $\mp v$. The equation of motion reduces to:

$$\partial_z^2\phi = U'(\phi) \tag{3.119}$$

which is Newton's law with $z =$"time" and an inverted potential $U \to -U$.



There is a conserved "energy,"

$$\frac{1}{2}(\partial_z\phi)^2 + (-U) = 0 \tag{3.120}$$

which means we can work instead with the first order equation

$$\frac{d\phi}{dz} = \pm\frac{g}{\sqrt{2}}(\phi^2 - v^2). \tag{3.121}$$

We'll choose "-" to interpolate from $-v$ to v. The equation is easily integrated:

$$-\frac{\sqrt{2}}{g}\int\frac{d\phi}{\phi^2 - v^2} = z - z_0 = \frac{\sqrt{2}}{gv}\tanh^{-1}(\phi/v) \tag{3.122}$$

so

$$\phi(z) = v\tanh\left(\frac{gv}{\sqrt{2}}(z - z0)\right) \tag{3.123}$$

like we drew. The tension is

$$\sigma = \int_{-\infty}^{\infty} \left[\frac{1}{2}(\partial_z\phi)^2 + \frac{g^2}{4}(\phi^2 - v^2)^2\right] \tag{3.124}$$

Clearly z_0 is a collective coordinate related to the translation symmetry broken by the domain wall. If we promote it to $z_0(x, y, t)$, we can describe small excitations of the domain wall. For the moment, we hold it fixed and continue with the tension computation. Note that

$$U = \frac{1}{2}\left(\frac{\partial W}{\partial \phi}\right)^2, \quad W = \frac{g}{\sqrt{2}}\left(\frac{1}{3}\phi^3 - v^2\phi\right) \tag{3.125}$$

so the energy density is a sum of squares. As we did with the energy of the BPS monopole and the action of the self-dual instanton, we rewrite the domain wlal tension as

$$\sigma = \int_{-\infty}^{\infty} dz \left[\frac{1}{2}(\partial_z\phi + W')^2 - \frac{\partial\phi}{\partial z}\frac{\partial W}{\partial \phi}\right]. \tag{3.126}$$

The second term is a total derivative, $\frac{\partial\phi}{\partial z}\frac{\partial W}{\partial\phi} = \frac{\partial W}{\partial z}$, so we can integrate it to give

$$\sigma = \underbrace{W(v)}_{z\to\infty} - \underbrace{W(-v)}_{z\to-\infty} + \underbrace{\int_{-\infty}^{\infty} dz\, \frac{1}{2}(\partial_z\phi + W')^2}_{\geq 0} \tag{3.127}$$

so

$$\sigma \geq W|_{-v}^{v} \tag{3.128}$$

which is called the Bogolmol'nyi bound. The bound is saturated for $\partial_z\phi + W' = 0$, or $\partial_z\phi = \sqrt{-2U}$, which is the same as the first order equation we found previously. In general, the domain wall extremizes the tension functional, and in this model, it saturates the lower bound.

Now let's look at fluctuations of the moduli. Write $\phi = \phi_0(z - z_0(x, y, t))$, where ϕ_0 is given by Eq. (3.123). Plug it into the action (3.118). The result is

$$S = \int d^4x\, \frac{1}{4}g^2v^4\text{sech}^4\left(\frac{1}{\sqrt{2}}gv(z - z_0)\right)\left((\partial_t z_0)^2 - (\partial_x z_0)^2 - (\partial_y z_0)^2\right). \tag{3.129}$$

We can perform the z integral immediately, obtaining

$$S = \int d^3x\, \frac{\sqrt{2}}{3}gv^3\left((\partial_t z_0)^2 - (\partial_x z_0)^2 - (\partial_y z_0)^2\right). \tag{3.130}$$

Thus the infrared theory contains excitations localized on the domain wall, corresponding to local fluctuations in the domain wall position. They are described by a three dimensional massless scalar field theory.

This is equivalent to a (generalization of) the Polyakov string worldsheet action in a particular gauge. It can be generalized further by coupling to fermions, electromagnetism, and gravity, and considering junctions and networks.

3.8 Strings

We will discuss three types of strings, following the classifications of Polchinski and Banks-Seiberg:

- global
- local/gauge/Abrikosov-Nielsen-Olesen (ANO)
- Aharonov-Bohm (AB)

The "local/global" terminology historical rather than optimal, which will become apparent shortly.

Global strings

QED is a "zero-form" U(1) gauge theory, meaning the transformation parameter is a zero-form, a function. In zero-form gauge theories the gauge field is a $0+1 = 1$-form. A 1-form may be integrated along lines, which can be interpreted as worldlines of particles. Thus, the electrically charged objects in a zero-form gauge theory are particles. They couple as

$$S \supset q \int_{WL} A \tag{3.131}$$

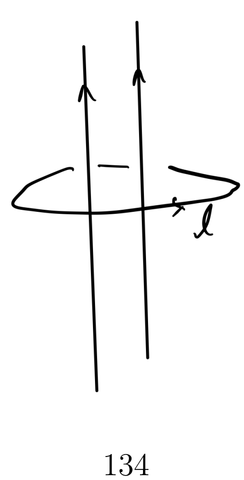
where WL is a 1-dimensional worldline. Charged particles also source $F = dA$, so we can detect the charges far away using the Gauss law: $\int_{M_2} \star F \propto Q_{encl}$.

These ideas generalize to higher-form gauge theories. A 1-form U(1) gauge theory has 1-form transformation parameters and 2-form gauge fields B. A 2-form may be integrated over surfaces, which can be interpreted as the worldsheets of strings. Thus, the electrically charged objects in a zero-form gauge theory are particles. They couple as

$$S \supset q \int_{WS} B \tag{3.132}$$

where WS is a 2-dimensional worldsheet. The gauge transformation is $B \to B + dA$, where the gauge transformation function A is itself a 1-form U(1) gauge field. The gauge invariant field strength is a 3-form $C = dB$. The analog of the Gauss law is

$$\int_{\ell} \star C \propto \#\text{strings in } \ell \tag{3.133}$$



Here $\star C$ is a 1-form. In a static gauge, with cylindrical coordinates, the field configuration of the string is $B_{tz} \sim r$, $C_{trz} \sim 1$, and $\star C$ is a 1-form with constant θ component. Because $d \star C = 0$, the line integral around ℓ is topological.

Unfortunately, these are the objects called global strings. The reason is there is a more common dual formulation. All such dualities have the form

$$\star d(\text{gauge field}) = d(\text{dual gauge field}). \tag{3.134}$$

In the present case, we would write

$$\star C = d\alpha \tag{3.135}$$

where α is the dual zero-form (scalar). However, α cannot be well-defined everywhere– it must be a zero form gauge field. This is necessary to satisfy the gauss law:

$$\int_\ell \star C = \int_\ell d\alpha \propto \#\text{strings in } \ell \tag{3.136}$$

which can only happen if α is compact – it is a scalar valued on the circle, which is what is means to be a zero form U(1) gauge field. Then $d\alpha$ is a closed 1-form, but not exact, because α winds around ℓ.

For example, for one infinite straight string, set up a cylindrical coordinate system with azimuthal angle θ. Then the string configuration, in the dual formulation, is simply

$$\alpha = \theta. \tag{3.137}$$

The string number is computed by the topological charge,

$$\frac{1}{2\pi} \int_\ell d\alpha = 1. \tag{3.138}$$

This is our friend $\pi_1(S^1)$ again, here associated with maps from real-space loops surrounding the string to the $U(1)$-valued field α.

We can write field theories that have these strings in their dynamical spectrum. For this the dual formulation is more familiar and useful. The textbook example is the global Abelian Higgs model,

$$\mathcal{L} = |\partial_\mu \phi)|^2 - \lambda(|\phi|^2 - v^2)^2 \tag{3.139}$$

where ϕ is a complex scalar. This has a U(1) global symmetry that is completely spontaneously broken. At low energies, $\phi \to ve^{i\alpha}$, where α is a real compact scalar Goldstone boson. The model contains "global strings" which are defects around which α winds by $2\pi k$. This is the dynamical realization of the previous.

From the IR perspective, the defect is singular. The tension is

$$T = \int_\perp d^2x[|\partial_i \phi|^2 - U(\phi)] \tag{3.140}$$

where the integral runs over the transverse directions. If we plug in $\phi = ve^{ik\theta}$,

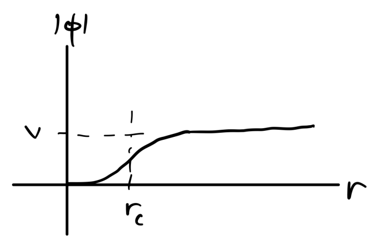
$$T \to 2\pi \int r dr \left| \frac{1}{r} \partial_\theta (ve^{ik\theta}) \right|^2$$
$$= 2\pi v^2 k^2 \int \frac{dr}{r} \tag{3.141}$$

which is log divergent in both the core and far from the string.

The field theory itself regulates the UV divergence in the core, by letting $|\phi| \to 0$ as $r \to 0$. This is symmetry restoration in the core, which we saw previously in our discussion of monopoles. Then

$$r \left| \frac{1}{r} \partial_\theta (\phi(r) e^{ik\theta}) \right|^2 \sim \frac{|\phi|^2}{r} \tag{3.142}$$

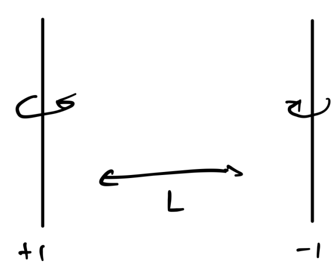
which is integrable as $r \to 0$ as long as ϕ goes to zero as $\phi \sim 1/\log(r)$ or faster.



We can estimate the core size by an energy balance between the gradient terms and the potential:

$$r_c^2 \lambda v^4 \sim \frac{v^2}{k^2} r_c^2 r_c^2$$
$$\Rightarrow r_c \sim \frac{k}{v\sqrt{\lambda}}. \tag{3.143}$$

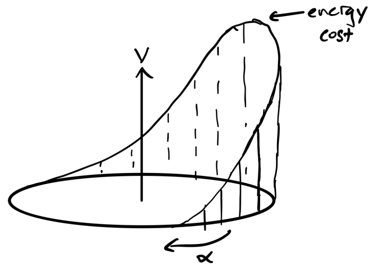
The IR divergence, however, is still there and signals something interesting: confinement.



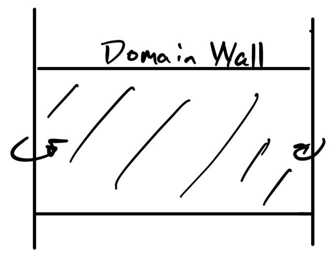
So the energy density grows $\sim \log L$. Loops want to collapse.

In any case, we see that strings are a dynamical set of nonperturbative excitations in these models. They source/couple to a 2-form gauge field in the dual description, and can be detected far away by a generalized Gauss law.

We must emphasize that the "2-form electric" U(1) gauge symmetry $B \to B + dA$ is different from the "zero-form magnetic" U(1) symmetry $\alpha \to \alpha + const$. The latter is really a global symmetry and might be explicitly broken. For example, we could add a term $\epsilon(\phi + \phi^*)$ to the scalar potential in the AHM. The leading effect is to general a potential for α:



In the presence of this potential, $\alpha = \theta$ is no longer a solution of the equations of motion. The field can still wind around an axis, but it needs to "move quickly" through the high potential region. The result is a domain wall that ends on the string.



Gauge and AB strings

Now let us gauge the U(1) global symmetry of the AHM, giving ϕ a charge q.

$$\mathcal{L} = -\frac{1}{4g^2} F_{\mu\nu} F^{\mu\nu} + |D\phi|^2 - U(\phi) \tag{3.144}$$

We look again for a minimal energy string configuration, with $\phi = ve^{i\alpha(\theta)}$ far from the core. The energy density far from the core will be

$$\mathcal{E} \sim E^2 + B^2 + |D\phi|^2 \tag{3.145}$$

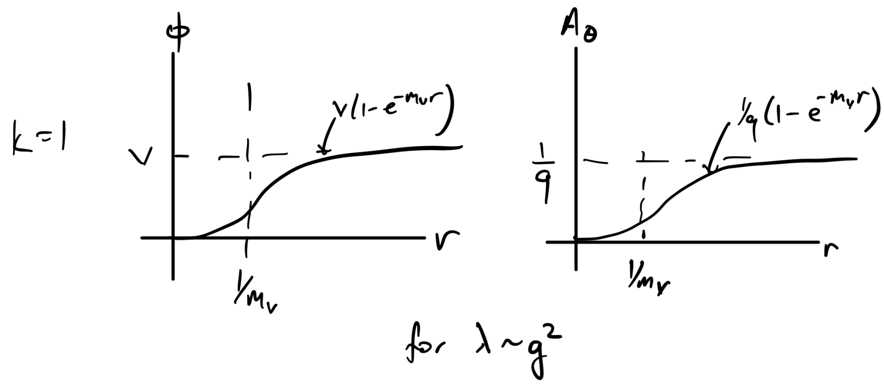
since $U(|\phi| = v) = 0$. Now

$$|D\phi|^2 \propto (\partial_\theta \alpha - qA_\theta)^2 \tag{3.146}$$

so if $\alpha = \theta k$, then we can minimize this term by taking $A_\theta = k/q$. Since locally we can write $A = d\lambda$, this is pure gauge and $E = B = 0$. So the energy density vanishes far from the core, quite different from the global string. However,

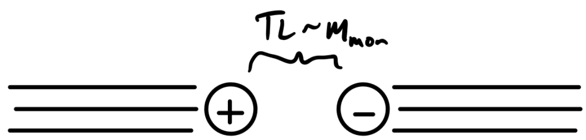
$$\int_\ell A = \int A_\theta d\theta = 2\pi k/q \tag{3.147}$$

which equals the magnetic flux enclosed by the loop. Thus the energy density is not zero everywhere: the string core contains a magnetic flux tube. It is smooth because of symmetry restoration. The fields have the form:



The tension is UV and IR finite in this case. Roughly, it is $r_c^2 U(0) \sim v^2$. A more precise estimate finds $v^2 \log(\lambda/g^2)$, which shows that the log divergence in the global string tension re-emerges in the "global limit" $g \to 0$.

For $q = 1$, the string number is completely ungauged. There is no analog of a long-range 2-form field B, and we cannot detect the string far away. There is no Aharonov-Bohm effect here because the magnetic flux is quantized. Furthermore the string number need not be conserved:



Or we can make flux tubes by separating monopoles:

The energy grows with the separation L. Linear confinement of magnetic charges $\leftrightarrow$ electric higgs phase. This is the Meissner effect, dual to what we discussed in the Polyakov model.

For $q > 1$, the theory has an unbroken Z_q gauge symmetry. $\phi \to e^{iq\lambda}\phi$, so $\lambda = 2\pi n/q$ with $n = 0 \ldots q-1$ leaves the vacuum invariant. Then the Wilson line around the k-string is

$$e^{i\int_\ell A} = e^{2\pi i k/q} \tag{3.148}$$

This phase can be detected in Aharonov-Bohm experiments with electric charge-1 probes. We see that there are q-types of string corresponding to $k = 0 \ldots q-1$.

3.9 Fermions in global string backgrounds

We'll work in four dimensions, although a number of the results apply in even dimensions. Let $\phi = \phi_1 + i\phi_2$ be a complex scalar field with a vev f and a flat direction, as in the global abelian Higgs model. We couple it to a Dirac fermion via

$$\mathcal{L} \supset -\bar{\psi}(\phi_1 + i\gamma^5\phi_2)\psi. \tag{3.149}$$

By field redefinitions we can choose y to be real and positive. We want to solve the Dirac equation on a global string background $\phi = f(r)e^{\pm i\theta}$, where $f(0) = 0$ and $f(\infty) = f$. Orient the string so that its worldsheet is at fixed spacetime coordinates $x_1 = r\cos\theta, x_2 = r\cos\theta$. Then spacetime coordinates x_0, x_3 can be used as worldsheet coordinates. Decompose the fermion into subspaces of definite chirality by writing $\psi = \psi_+ + \psi_-$, where $\gamma^5\psi_\pm = \pm\psi_\pm$ and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The Dirac equation is

$$i\partial\!\!\!/\psi - y(\phi_1 + i\gamma^5\phi_2)\psi = 0 \tag{3.150}$$

which, projected onto the $\pm$ eigenspaces of γ^5, splits into

$$\begin{aligned} i\partial\!\!\!/\psi_- - y\phi\psi_+ &= 0 \\ i\partial\!\!\!/\psi_+ - y\phi^*\psi_- &= 0. \end{aligned} \tag{3.151}$$

Since ϕ introduces explicit dependence on θ, it is convenient to decompose $\partial\!\!\!/$ into the transverse and longitudinal contributions. Let $\partial\!\!\!/^{\text{int}} = \gamma^a\partial_a$ for $a = 0, 3$. $i\partial\!\!\!/^{\text{int}}$ is the Dirac operator on the string. Similarly define a hermitian matrix $\gamma^{5,\text{int}} = \gamma^0\gamma^3$. Let us also look for solutions that are independent of θ (they respect the rotation invariance of the string.) All together, we have the system

$$\begin{aligned} i\partial\!\!\!/^{\text{int}}\psi_- + i(\gamma^1\cos\theta + \gamma^2\sin\theta)\partial_r\psi_- - y\phi\psi_+ &= 0 \\ i\partial\!\!\!/^{\text{int}}\psi_+ + i(\gamma^1\cos\theta + \gamma^2\sin\theta)\partial_r\psi_+ - y\phi^*\psi_- &= 0. \end{aligned} \tag{3.152}$$

Remarkably, it is straightforward to verify that there are chiral zero-mode solutions to Eq. (3.152) bound to the string. What this means is, suppose we have a solution to the massless 2D Dirac equation $\partial\!\!\!/^{\text{int}}\psi_- = 0$ (a "zero mode"). Such solutions can be chosen to satisfy $\gamma^{5,\text{int}}\psi_- = \pm\psi_-$ ("chiral"). Then, for one choice of this chirality, the solution can be embedded in a complete solution to Eq. (3.152), where the support of the complete solution in the radial direction falls off rapidly ("bound").

To find these solutions explicitly, separate the fields as $\psi_- = \eta(x_0, x_1)h(r)$, $\psi_+ = \chi(x_0, x_1)h(r)$, where η and χ are spinors and h is a function. Then the first equation in (3.152) becomes

$$i(\gamma^1\cos\theta + \gamma^2\sin\theta)\eta\partial_r h - y\phi\chi h = 0 \tag{3.153}$$

using the ansatz $\partial\!\!\!/^{\text{int}}\eta = 0$. Clearly we want $h' = -yf(r)h$, or

$$h(r) = e^{-y\int_0^r dr' f(r')}, \tag{3.154}$$

after which we have left

$$i(\gamma^1 \cos\theta + \gamma^2 \sin\theta)\eta = -e^{\pm i\theta}\chi. \tag{3.155}$$

Now $\gamma^5\eta = -\eta$, and $\gamma^5 = i\gamma^{5,\text{int}}\gamma^1\gamma^2$. So the chiral ansatz $\gamma^{5,\text{int}}\eta = b\eta$, $b = \pm 1$, is equivalent to $\gamma^1\gamma^2\eta = ib\eta$. Therefore

$$i\gamma^1(\cos\theta - \gamma^1\gamma^2 \sin\theta)\eta = i\gamma^1(\cos\theta - ib\sin\theta)\eta = -e^{\pm i\theta}\chi. \tag{3.156}$$

We conclude that we must choose the two-dimensional chirality b, to be the opposite of the the string charge ± 1, and also set

$$\chi = -i\gamma^1\eta. \tag{3.157}$$

This completes the solution. One can readily check that the second equation in Eq. (3.152) is satisfied. The solution describes a chiral fermionic excitation propagating along the string according to the 2D massless Dirac equation

$$(\gamma^0\partial_0 + \gamma^3\partial_3)\eta = 0, \quad \gamma^0\gamma^3\eta = b\eta. \tag{3.158}$$

Suppose the string charge is $+1$, so $b = -1$. Then in the Weyl basis for the gamma matrices,

$$\eta = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \tilde{\eta} \end{pmatrix}. \tag{3.159}$$

Similarly if the string charge is -1, $b = +1$ and

$$\eta = \begin{pmatrix} 0 \\ 0 \\ \tilde{\eta} \\ 0 \end{pmatrix}. \tag{3.160}$$

So the equation of motion reduces to

$$(\partial_0 + b\partial_3)\tilde{\eta} = 0, \tag{3.161}$$

describing an excitation propagating at the speed of light in a direction correlated with the string charge.

In the bulk, the fermions and the radial mode of the scalar are gapped by the scalar vev. Therefore the EFT describing the low-energy limit of the string contains a massless bulk Goldstone boson, a massless 2D scalar describing small fluctuations of the string, and a massless 2D chiral fermion bound to the string!

The theory has a conserved current associated with the global $U(1)_F$ fermion number symmetry. The matter content is vectorlike, so there is no $U(1)_F^3$ 't Hooft anomaly. We can couple this current to electromagnetic fields. This is fine in the 4D description: it is one-flavor QED with a

Yukawa coupling to a neutral scalar, from which the fermions gain a mass. But there is a puzzle: we have found that the low-energy EFT contains a 1+1 dimensional chiral fermion living on the string. The fermion number current in this theory has a ‘t Hooft anomaly, so 1+1 dimensional QED with one Weyl fermion has a gauge anomaly,

$$\partial_a J^a = \frac{e}{4\pi}\epsilon^{ab}F_{ab} \tag{3.162}$$

where a, b are indices corresponding to the string worldsheet coordinates x^0, x^3. (This is half the usual 2D chiral anomaly, since there is only one Weyl fermion.) Does anomaly matching fail? Is the current conserved in the presence of background EM fields, nor not? Is the theory with gauged fermion number consistent?

This puzzle is resolved by a mechanism called “anomaly inflow.” Charge can flow from the bulk onto the string. Here is how it works.

By coupling to slowly varying background electromagnetic fields, the bulk EFT has an additional coupling. It is just the $aF\tilde{F}$ coupling we saw in our discussion of axion electrodynamics,

$$\begin{aligned}\mathcal{S} &\supset \frac{\alpha}{2\pi}\int \theta F\wedge F \\ &= \frac{\alpha}{2\pi}\int \varphi dA\wedge dA.\end{aligned} \tag{3.163}$$

where $\varphi = a/f$ is the angle-valued Goldstone field. The current induced by the background fields is

$$e\langle J^\lambda\rangle = \frac{\delta S_{eff}}{\delta A_\lambda} = -\frac{\alpha}{\pi}\epsilon^{\lambda\gamma\rho\sigma}(\partial_\gamma\varphi)(\partial_\rho A_\sigma). \tag{3.164}$$

In the string background $\phi = f(r)e^{i\theta}$, $\partial_\lambda J^\lambda$ vanishes almost everywhere due to antisymmetry, but the string is a singular line where neither the axion $\varphi = \arg(\phi)$ nor the form $d\varphi$ are defined, so we need to be more careful there. The Stokes theorem tells us

$$\frac{1}{2\pi}\int_{\partial\Sigma} d\varphi = 1 = \frac{1}{2\pi}\int_{\partial\Sigma} dd\varphi \tag{3.165}$$

so

$$dd\varphi = 2\pi\delta(x^1)\delta(x^2)dx^0\wedge dx^3 \tag{3.166}$$

and

$$\partial_\lambda J^\lambda = -\frac{e}{4\pi}\delta(x^1)\delta(x^2)\epsilon^{ab}F_{ab} \tag{3.167}$$

where a, b are indices on the string worldsheet. We see that an applied electric field *longitudinal* to the string induces a *radial* current (3.167), and current conservation is rescued. This is an example of a general mechanism: an anomaly can be canceled by coupling the theory to a higher dimensional bulk.

This was a case where anomalous theory lives on a codimension-2 surface in the bulk. Another interesting example is:

3.10 Fermions in domain wall backgrounds

Let

$$\mathcal{L} = \bar{\psi} i \partial\!\!\!/ \psi + \frac{1}{2}(\partial_\mu \phi)^2 - y\phi\bar{\psi}\psi - \frac{1}{4}(\phi^2 - f^2)^2. \tag{3.168}$$

ϕ is a real scalar. In this case we consider the theory in five spacetime dimensions. There is a domain wall solution interpolating between $\phi = \pm f$. The worldvolume of the domain wall is four dimensional, so it is useful to think of it as our spacetime living inside a higher dimensional bulk.

At the wall, $\phi \to 0$ and the fermion mass vanishes. Crossing the wall, the fermion mass changes sign. Choose coordinates so that the wall is at fixed x^4 and $x^{0,1,2,3}$ are internal coordinates on the wall. The Dirac equation is

$$i\partial\!\!\!/^{\text{int}} \psi - i\gamma^5 \partial_4 \psi = y\phi\psi. \tag{3.169}$$

where γ^5 is the extra Dirac matrix included in the Clifford algebra in five dimensions. Again there is a zero mode of definite chirality,

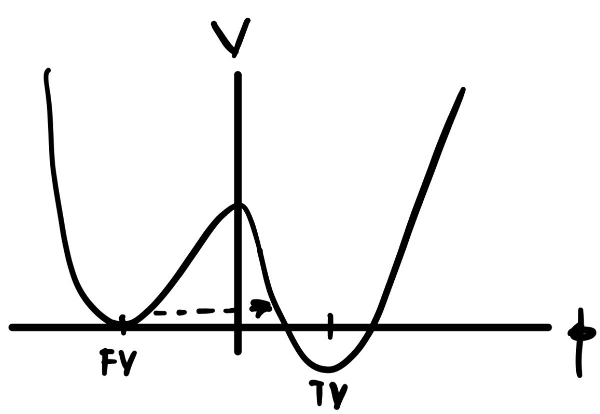
$$\psi = \eta(x^{\text{int}}) e^{-y\int_0^{x^4} dx' \phi(x')}\ , \qquad \partial\!\!\!/^{\text{int}} \eta = 0, \gamma^5 \eta = \eta. \tag{3.170}$$

Now a similar puzzle arises: the low energy theory contains a 4D Weyl fermion, some 4D scalars, and a 5D bulk Goldstone boson. The 4D $U(1)_F$ fermion number symmetry has a $U(1)_F^3$ anomaly which can be expressed as $\partial_\mu J^\mu \propto F\tilde{F}$. The theory (3.168) does not have this anomaly. If we couple the $U(1)_F$ current to background gauge fields, what happens? In this case, the 5D EFT has a Chern-Simons term coupling to the sign of ϕ. A similar calculation reveals an induced current flowing onto the wall from the bulk.

3.11 Vacuum Decay

This section is somewhat outside the theme of the chapter, since topology plays no role. However, it is morally closely related to the discussion of domain walls and instantons, and uses similar semiclassical techniques.

We will discuss the bubble nucleation processes by which metastable field theory vacua can decay. For example, the scalar potential below exhibits a metastable vacuum state:



In the standard semiclassical treatment of vacuum decay, one searches for saddle points of the Euclidean path integral that describe bubble nucleation. These are solutions to the equations of motion in which the fields asymptote to the false vacuum in all directions at Euclidean infinity, but interpolate to somewhere around the true vacuum inside a finite hyperspherical "bubble." Since the fields are only excited in a finite region of Euclidean spacetime, like the BPST instanton, these are also called instanton, or sometimes "bounce," solutions.

To illustrate another application of effective field theory, we won't consider the most general field theoretic problem. What we lose in generality we gain in simplicity. In fact, we are only going to study the quantum mechanics of a single degree of freedom R, which will be identified with a collective coordinate – the radius of the bubble of true vacuum, nucleating inside the sea of false vacuum. $R = 0$, the "bubble of zero radius," is identified with the metastable ground state in the semiclassical limit, the empty false vacuum.

We start from a simple, concrete field theory setting: consider a real scalar field in 3+1 dimensions with a double-well potential and a small symmetry breaking parameter ϵ,

$$V(\phi) = \frac{1}{2}\lambda(\phi^2 - a^2)^2 - (\epsilon/2a)\phi. \tag{3.171}$$

This potential, like the figure above, exhibits a metastable vacuum state. For $\epsilon \to 0$, the theory would possess two degenerate vacua and heavy, stable domain wall solutions interpolating between them. For small finite ϵ, we can imagine bending a domain wall into a large spherical bubble, with the true (lower energy density) vacuum on the inside. Playing the vacuum energy off the domain wall tension, we can arrange for a configuration with the same energy as the false vacuum. This is a candidate "decay product" for the false vacuum, since it conserves energy. We have to see, however, how to compute the decay rate.

Since the spherical bubble will need to be very large for small ϵ, we may neglect the slow accelerations driven by curvature and the energy gap ϵ, the equation of motion describing the heavy domain wall degree of freedom is $\partial_t^2\phi - \partial_r^2\phi + 2\lambda(\phi^2 - a^2)\phi = 0$. The relevant solutions are

$$\phi \approx a \tanh\left[\frac{\sqrt{\lambda}a(R + vt - r)}{\sqrt{1 - v^2}}\right]. \tag{3.172}$$

Here R is a large radius, v is the wall velocity, the wall tension is given by $\sigma \propto \sqrt{\lambda}a^3$, and the factor of $\frac{1}{\sqrt{1-v^2}}$ accounts for length contraction of the wall thickness.

To obtain an effective action for slow variations of v driven by curvature and pressure, we replace $R + vt \to R(t)$, $v \to \dot{R}(t)$, and plug these field configurations into the Klein-Gordon action. Upon so doing each term in the Lagrangian is found to be proportional to $f(r)\,\mathrm{sech}^4\left[\frac{\sqrt{\lambda}a(R-r)}{\sqrt{1-\dot{R}^2}}\right]$. In the thin-wall limit $\frac{\sqrt{\lambda}a}{\sqrt{1-\dot{R}^2}}R \gg 1$ the sech^4 is sharply peaked around $r = R$, so we may replace $f(r) \to f(R)$ and do the integral over space. One obtains an effective Lagrangian for the collective coordinate $R(t)$, given by Eq. (3.173) below with $p = 3$ and $R_0 = 3\sigma/\epsilon$:

$$L/\epsilon = c\epsilon\left(-R_0 R^{p-1}\sqrt{1 - (\partial_t R)^2} + R^p\right). \tag{3.173}$$

c is a numerical constant. In fact a large class of field theoretic vacuum decay processes are described by this effective Lagrangian at the leading semiclassical order, including thin-wall bubble nucleation in $d = p$ spatial dimensions, Schwinger pair production in a constant background electric field for $p = 1$, and bubble of nothing decays for $p = d - 1$. (Therefore R is assumed to be ≥ 0, since it plays the role of a radial collective coordinate.)

Having obtained an effective quantum mechanical theory (3.173), we would like to use it to compute a tunneling rate. We formulate it as follows: we study quantum mechanical transition amplitudes of the form

$$\langle\psi_f(R,T_f)|\psi_i(R,T_i)\rangle = N\int dR_i dR_f \int_{R(T_i)=R_i}^{R(T_f)=R_f} DR\, e^{iS/\hbar}\,,$$
$$S = S_i(R_i) - S_f(R_f) + \int_{T_i}^{T_f} dt\, L(\partial_t R, R, t). \tag{3.174}$$

The initial and final states are given by wavefunctions parametrized as $\psi(R) = e^{iS_{i,f}(R)/\hbar}$, where $S_{i,f}$ are complex functions. In tunneling problems these wavefunctions should be localized inside and outside of a false vacuum well. We have written explicit factors of $\hbar$ to indicate that we are interested in states where both the real and imaginary parts of $S_{i,f}$ contain contributions of order one in the $\hbar$ expansion, like coherent states, but subsequently we will drop the $\hbar$s.

Next, we perform a Wick rotation of the time contour, $t \to -i\tau$. Now the path integral describing the amplitude is

$$\int dR_i dR_f \int_{R(T_i)=R_i}^{R(T_f)=R_f} DR\, e^{-S_c}\,,$$
$$S_c = -iS_i(R_i) + iS_f(R_f) - i\int_{T_i}^{T_f} d\tau\, L_c(\partial_\tau R, R, \tau)$$
$$L_c(\partial_\tau R, R, \tau) = -iL(i\partial_\tau R, R, -i\tau). \tag{3.175}$$

In the leading semiclassical approximation we are interested in solutions to the bulk Euler-Lagrange equation, $\delta L_c/\delta R - \partial_\tau(\delta L_c/\delta\partial_\tau R) = 0$. The equation of motion can be replaced by the conservation of energy condition if L does not depend explicitly on t.

Finally, we relax the Dirichlet boundary conditions used in ordinary semiclassical treatments of the path integral. Since we have introduced general initial and final states, we can consider unrestricted boundary variations. Stationarity of the action then gives rise to the following boundary relations:

$$p_i \equiv \left.\frac{\delta L_c}{\delta\partial_\tau R}\right|_{\tau=T_i} = S_i'(R_i)$$
$$p_f \equiv \left.\frac{\delta L_c}{\delta\partial_\tau R}\right|_{\tau=T_f} = S_f'(R_f). \tag{3.176}$$

Eq. (3.176) relates the initial and final momenta of the semiclassical trajectories to the gradient of the wavefunction at the initial and final positions. A totally generic initial and final wavefunction, and a given solution $R(\tau)$, will not in general satisfy Eq. (3.176), and so will not provide a single saddle point approximation to the full path integral, including the initial and final states as described above. One can instead find solutions to the bulk equations of motion first, and then use Eq. (3.176) to infer wavefunctions with consistent properties, such that those solutions provide genuine saddle points of the full path integral. These are generally weak constraints on the initial and final wavefunctions, only constraining their gradients at points, but we must determine them after the fact.

These self-consistency conditions have very natural physical interpretations. For example, the real part of the momentum at the semiclassical endpoints has to match the momentum of the initial and final state wavefunctions in the eikonal approximation.

Although this is an acceptable formulation of a tunneling problem, it is not generally how the computation is carried out in standard problems. Since (3.173) does not depend explicitly on time, there is a conserved energy. There is a false vacuum at $R_{FV} = 0$ with $E_{FV} = 0$, and the classical turning point is $R_{TP} = R_0$. So one can instead solve a variational problem with fixed energy:

$$E_E = E_{FV} \tag{3.177}$$

where E_E is the Euclidean energy and $E_{FV} = V(R_{FV})$ is the value of the potential in the classical false vacuum R_{FV}. Typically there is a solution to Eq. (3.177) that starts at R_{FV} and reaches the classical turning point $R = R_{TP}$ with zero momentum at some later time, which is determined by the solution. We are taking a different approach which offers some useful generalizations and a different perspective, but for the application below the approaches are equivalent.

We note that it is important *not* to consider a path integral with fixed initial and final times and fixed initial and final values of R. Position eigenstates have completely uncertain momentum, and there will always be solutions, unrelated to tunneling, that have enough energy to summit the barrier classically. The fixed-energy formulation avoids these complications, as does the wavepacket formulation, as long as the gradients in the wavepacket are not too large.

For $p > 1$, the solutions relevant for tunneling are quite simple. In the Euclidean case $t \to -i\tau$, they are

$$R = \Theta(\tau + R_0)\sqrt{R_0^2 - \tau^2}. \tag{3.178}$$

The solution is real in the relevant range of τ, and the nucleation point is $R(0) = R_0 = 3\sigma/\epsilon$, where the momentum vanishes. It is consistent with initial and final wavepacket states peaked at $R_i = 0$ and $R_f = R_0$.

The on-shell action gives the tunneling exponent. For example, for $p = 3$, one finds

$$\text{Re}\left[-i\int_{T_i}^{T_f} d\tau\, L_c(\partial_\tau R, R, \tau)\right] = \frac{c\pi R_0^4 \epsilon}{16} = \frac{81 c\pi\sigma^4}{16\epsilon^3} \tag{3.179}$$

So the tunneling amplitude is proportional to

$$e^{-\frac{81c\pi\sigma^4}{16\epsilon^3}}. \tag{3.180}$$

More precisely, to really get an answer to our Lorentzian transition amplitude question, we need to analytically continue the amplitude in the complex T_i-plane back to the real time axis, for *fixed* final states. Here we will be content with a plausibility argument. We have already seen that the tunneling exponent is independent of T_i (if $|T_i| > R_0$, so that the trajectory described by the instanton has enough time to complete.) Furthermore, on physical grounds, the decay rate is expected to be constant over timescales long compared to the perturbative timescales in the false vacuum and short compared to the lifetime. Therefore we make the plausible assumption that T_i can be continued back to the real axis trivially, and (3.180) is the leading semiclassical estimate for the transition rate. This can also be justified by the fully Euclidean computation of Coleman and Callan, which uses the optical theorem.

After nucleation, the semiclassical state is a bubble of radius R_0 at rest. Its subsequent evolution is easily obtained by analytic continuation of (3.178):

$$R = \sqrt{R_0^2 + t^2}, \quad t \geq 0. \tag{3.181}$$

The bubble expands outward, and on a timescale of order a few R_0 it is expanding at the speed of light.

Chapter 4

Supersymmetry

4.1 Supersymmetry

Poincare Algebra:

$$
\begin{aligned}
&[P^\mu, P^\nu] = 0 \quad \text{Translations} \\
&[M_{\mu\nu}, M_{\rho\sigma}] = ig_{\nu\rho}M_{\mu\sigma} - ig_{\mu\rho}M_{\nu\sigma} - ig_{\nu\sigma}M_{\mu\rho} + ig_{\mu\sigma}M_{\nu\rho} \\
&[M_{\mu\nu}, P_\rho] = -ig_{\rho\mu}P_\nu + ig_{\rho\nu}P_\mu.
\end{aligned}
\tag{4.1}
$$

Here $M_{\mu\nu} = -M_{\nu\mu}$, $M_{0i} = K_i$ are the boosts, and $M_{ij} = \epsilon_{ijk}J_k$ are the rotations.

There is an extension to include generators with spinor indices:

$$
\begin{aligned}
&[M^{\mu\nu}, Q_\alpha] = \frac{1}{2}\left(\sigma^{\mu\nu}\right)_\alpha^\beta Q_\beta \qquad \left(\sigma^{\mu\nu} = \frac{1}{2}\left(\sigma^\mu_{\alpha\dot\alpha}(\bar\sigma^\nu)^{\dot\alpha\beta} - \sigma^\nu_{\alpha\dot\alpha}(\bar\sigma^\mu)^{\dot\alpha\beta}\right)\right) \\
&\left[Q^A_\alpha, P^\mu\right] = 0 \\
&\left\{Q^A_\alpha, Q^{*B}_{\dot\beta}\right\} = 2\sigma^\mu_{\alpha\dot\beta}P_\mu\delta^{AB} \quad A, B = 1\ldots N.
\end{aligned}
\tag{4.2}
$$

All anti/commutators vanish, apart from possible central charges.

The spinor nature of the Q's and the anticommutator the result of a pair of deep theorems which we will only quote:

Coleman-Mandula:
Given (1) Local, relativistic, 4D QFT with S-matrix
(2) Finite particle species
(3) Gap Then the most general Lie algebra of symmetries of the S-matrix is the Poincare algebra $\times$ internal global symmetries.

Haag Lopuszanski Sohnius: If anticommutators are allowed (graded Lie algebra), then the previous structure is the most general allowed.

$N = 1$ is the most plausible for low energy phenomenology, because it is the only case consistent with chiral fermions.

In ordinary QFT, fields come in finite-dimensional representations of Poincare: $A^\mu, \psi^\alpha, \phi \ldots$

In SUSY, we will see that fields are in finite-dim reps of super-Poincare that include fields of different spin as components

Chiral superfields: (ϕ, ψ_α)
Vector superfields: $(\lambda_\alpha A_\mu)$
Gravity superfield: $(\ \psi_{\mu\alpha}, g_{\mu\nu})$

4.1.1 Superspace

Superspace is a convenient way to write field content in SUSY and build invariant actions. We need some definitions.

First, we extend the spacetime coordinates:

$$x^\mu \to x^\mu,\ \theta_\alpha,\ \theta^*_{\dot\alpha}(\equiv \bar\theta_{\dot\alpha}) \tag{4.3}$$

The θ's are Grassmann-valued two-component spinors:

$$\{\theta_\alpha, \theta_\beta\} = \{\bar\theta_{\dot\alpha}, \bar\theta_{\dot\beta}\} = \{\theta_\alpha, \bar\theta_{\dot\beta}\} = 0 \tag{4.4}$$

$$(\text{unlike } [x^\mu, x^\nu] = 0) \tag{4.5}$$

Since $[\theta_\mu, \theta_\alpha] = 0$ and $\{\theta_\alpha, \theta_\alpha\} = 0$,

$$\theta_\alpha\theta_\alpha = 0. \tag{4.6}$$

This makes Taylor series simple.

Derivatives also satisfy anticommutation:

$$\left\{\frac{\partial}{\partial\theta_\alpha}, \frac{\partial}{\partial\bar\theta_\beta}\right\} = 0,\ \text{etc.} \tag{4.7}$$

Integration is defined by requiring

$$\int d\theta f(\theta+\epsilon) = \int d\theta f(\theta) \tag{4.8}$$

which is a generalization to Grassmann integration of the more familiar

$$\int_{-\infty}^{\infty} dx f(x) = \int_{-\infty}^{\infty} dx f(x+a). \tag{4.9}$$

Since, for a single Grassmann variable, the Taylor series terminates at linear order,

$$f(\theta) = f(0) + \theta f'(0)$$
$$f(\theta + \epsilon) = f(0) + (\theta + \epsilon) f'(0) \tag{4.10}$$

we will satisfy Eq. (4.8) if we assign the rules

$$\int d\theta \, 1 = 0$$
$$\int d\theta \, \theta = 1 \tag{4.11}$$

Simple!

For Grassmann-valued spinors $\theta_\alpha, \bar{\theta}_{\dot{\gamma}}$, the rules become

$$\int d^2\theta \theta_\alpha \theta_\beta \epsilon^{\alpha\beta} = 1 \equiv \int d^2\theta \theta^2$$
$$\int d^2\bar{\theta} \, \bar{\theta}^2 = 1 \tag{4.12}$$

and all others vanish.[1]

In superspace, the $Q, \bar{Q}$ generators of the super-Poincare algebra are represented by differential operators, much like $P_\mu \to i\partial_\mu$ when acting in functions in ordinary space.

$$Q_\alpha = \frac{\partial}{\partial \theta_\alpha} - i\sigma^\mu_{\alpha\dot{\alpha}} \bar{\theta}^{\dot{\alpha}} \partial_\mu$$
$$\bar{Q}_{\dot{\alpha}} = -\frac{\partial}{\partial \bar{\theta}^\alpha} + i\theta^\alpha \sigma^\mu_{\alpha\dot{\alpha}} \partial_\mu. \tag{4.13}$$

Note: since $[P] = 1$ and $Q^2 \sim P$, the dimensions of Q and θ are $[Q] = 1/2$, $[\theta] = -1/2$.

These differential operators obey the super-Poincare algebra. This is why superspace is so convenient!
Much like e^{iHt} generates finite time translations,

$$e^{\epsilon Q + \bar{\epsilon}\bar{Q}} \Phi(x^\mu, \bar{\theta}, \bar{\theta}) = \Phi(x^\mu - i\epsilon\sigma^\mu\bar{\theta} + i\theta\sigma^\mu\bar{\epsilon}, \theta + \epsilon, \bar{\theta} + \bar{\epsilon}) \tag{4.14}$$

generates *superspace* translations.

To get representations (function spaces on which the Q's act irreducibly), it is useful to introduce $D_\alpha, \bar{D}_{\dot{\alpha}}$, defined as

$$D_\alpha = \partial_\alpha + i\sigma^\mu_{\alpha\dot{\alpha}} \dot{\theta}^{\dot{\alpha}} \partial_\mu, \qquad \bar{D}_{\dot{\alpha}} = -\partial_{\dot{\alpha}} - i\theta^\alpha \sigma^\mu_{\alpha\dot{\alpha}} \partial_\mu \tag{4.15}$$

[1] ϵ appears here because it is an $SL(2, C)$ invariant tensor, meaning $M^\gamma_\alpha M^\delta_\beta \epsilon_{\gamma\delta} = \epsilon_{\alpha\beta}$ for $M \in SL(2, C)$, which is the Weyl representation of the Lorentz group.

which can be shown to satisfy

$$\left\{D_\alpha, \bar{D}_{\dot{\alpha}}\right\} = 0, \qquad \{D, D\} = 0, \qquad [D, Q] = 0. \tag{4.16}$$

Since $[D, Q] = 0$, we find the useful result that $\bar{D}\Phi = 0$ and $D\bar{\Phi} = 0$ are super-Poincare-invariant equations. Fields that satisfy the first (second) of these equations are called chiral (antichiral) superfields.

Chiral superfields

Now let us construct chiral superfields. Note that

$$\bar{D}_{\dot{\alpha}} \left(y^\mu \equiv x^\mu + i\theta\sigma^\mu\bar{\theta}\right) = 0, \quad \bar{D}_{\dot{\alpha}}\theta = 0. \tag{4.17}$$

Suppose $\Phi = \Phi(y, \theta^\alpha)$. Then $\bar{D}\Phi = 0$ automatically. Since any function of θ can be expanded in a series that truncates at $\theta_\alpha\theta^\alpha$, we can write $\Phi = \phi(y) + \sqrt{2}\theta_\alpha\psi^\alpha(y) + \underbrace{\theta^\alpha\theta_\alpha}_{\theta^2} F(y)$. If we further expand y, this is

$$\begin{aligned}\Phi(y, \theta^\alpha) =& \phi(x) + i\theta\sigma^\mu\bar{\theta}\partial_\mu\phi + \frac{1}{4}\theta^2\bar{\theta}^2\partial^2\phi \\ &+ \sqrt{2}\theta\psi - \frac{i}{\sqrt{2}}(\theta\theta)\partial_\mu\psi\sigma^\mu\bar{\theta} + \theta^2\bar{F}.\end{aligned} \tag{4.18}$$

.

The transformation laws of the component fields arise from $\delta\Phi = \left(\epsilon^\alpha Q_\alpha + \bar{\epsilon}_{\dot{\alpha}}\bar{Q}^{\dot{\alpha}}\right)\Phi$. Expanding out the derivatives, one can show that

$$\begin{aligned}&\delta\phi = \sqrt{2}\epsilon\psi \\ &\delta\psi = \sqrt{2}\epsilon F + \sqrt{2}i\sigma^\mu\epsilon^*\partial_\mu\phi \\ &\partial F = i\sqrt{2}\epsilon^*\bar{\sigma}^\mu\partial_\mu\psi.\end{aligned} \tag{4.19}$$

Spinors and scalars mix under the supersymmetry transformations! The chiral superfield Φ contains both a complex scalar ϕ and a Weyl fermion ψ. We will see that the complex scalar F is non-dynamical.

Holomorphic functions of chiral superfields are chiral, $\bar{D}\Phi^n = 0$, by the chain rule, a fact which we will use shortly.

Vector superfields

If we impose a reality condition on superspace functions, $V = V^\dagger$, it is preserved by $Q, \bar{Q}$. Carrying out the expansion,

$$V = i\chi - i\chi^\dagger - \theta\sigma^\mu\bar{\theta}A_\mu + i\theta^2\bar{\theta}\bar{\lambda} - i\bar{\theta}^2\theta\lambda + \frac{1}{2}\theta^2\bar{\theta}^2 D. \tag{4.20}$$

D will turn out to be non-dynamical. χ can be removed by a gauge transformation,

$$V \to V + i\Lambda - i\Lambda^\dagger \tag{4.21}$$

where Λ is itself a chiral superfield. This transformation generalizes $A \to A\partial_\mu\lambda$. (Compare $\partial_\mu\lambda$ with the $\theta\sigma^\mu\bar{\theta}$ term of Eq. (4.18).)

λ is a spin-1/2 "gaugino" and A is an abelian gauge field.

To construct the U(1) field strength, one builds

$$W_\alpha = - - \frac{1}{4}\bar{D}^2 D_\alpha V \qquad = -i\lambda_\alpha + \theta_\alpha D - (\sigma^{\mu\nu})_\alpha^\beta F_{\mu\nu}\theta_\beta + \theta^2\sigma^\mu_{\alpha\dot{\beta}}\partial_\mu\lambda^{*\dot{\beta}} \tag{4.22}$$

in $\chi = 0$ (Wess-Zumino) gauge.

A gauge transformation acts on charged fields as $\Phi \to e^{-iq\Lambda}\Phi$ where, again, Λ is a chiral superfield. Then $\Phi^\dagger e^{qV}\Phi$ is gauge invariant, for example. The gauge-covariant superderivative is

$$\mathcal{D}_\alpha\Phi = D_\alpha\Phi + D_\alpha V\Phi \tag{4.23}$$

Building invariant actions

Under an ordinary infinitesimal translation, $\mathcal{L} \to \mathcal{L} + \alpha_\mu\partial^\mu\mathcal{L}$. The action $\int d^4x\,\mathcal{L}$ is translation invariant. Similarly, we will obtain supersymmetry-invariant actions by integrating operators built out of superfields over superspace:

$$\int d^4x d^4\theta\, h(\Phi, \Phi^\dagger, V) \tag{4.24}$$

In slightly more detail: Acting on h with $Q \sim \partial_\theta - i\bar{\theta}\partial_\mu$ and $\bar{Q} \sim -\partial_{\bar{\theta}} + i\theta\partial_\mu$, all the terms are of the form

$$\int d^4\theta\partial_\theta(\text{anything}) = 0 \quad \text{or}$$
$$\int d^4x\partial_\mu(\text{anything}) = 0 \tag{4.25}$$

for fields vanishing fast enough at infinity. So $\int d^4x h_{\theta\theta\bar{\theta}\bar{\theta}} \equiv \int d^4\theta d^4x h$ is susy-invariant. The $\theta\theta\bar{\theta}\bar{\theta}$ component of a superfield is generally called a "D-term".

There is also a way to make an invariant with $\int d^2\theta$. Note that the $\theta\theta$ component of a chiral superfield transfers into a total spacetime derivative:

$$\delta F = i\sqrt{2}\epsilon^*\bar{\sigma}^\mu\partial_\mu\psi. \tag{4.26}$$

So $\int d^2\theta \int d^4x(\text{anychiralsuperfield})$ is invariant. Recalling that holomorphic functions of chiral superfields are chiral superfields, we can write another invariant,

$$\int d^2\theta \int d^4x\, f(\Phi). \tag{4.27}$$

The $\theta\theta$ component of a chiral superfield is generally called an "F-term".

Most interesting Lagrangians are built out of three terms:

$$\mathcal{L} = \mathcal{L}_{chiral\ kinetic} + \mathcal{L}_{gauge\ kinetic} + \mathcal{L}_{superpotential} + cc. \tag{4.28}$$

The first is a D-term and the second and third are F-terms.

$\mathcal{L}_{chiral\ kinetic}$:

$$\int d^4\theta \sum_i \Phi_i^\dagger e^V \Phi_i \sim \int d^4\theta\, (\phi + \theta\bar{\theta}\partial\phi + \dots)^\dagger \left(1 + \theta\bar{\theta}A + \theta\bar{\theta}\theta\bar{\theta}AA + \dots\right)(\phi + \theta\bar{\theta}\partial\phi + \dots) \tag{4.29}$$

so we can see there are D-terms like $(\partial\phi)^2$, $\partial\phi A\phi$, $\phi A\phi A$. These are building up $|(\partial + A)\phi|^2$ – the ordinary scalar kinetic term. Similarly for ψ.

$\mathcal{L}_{gauge\ kinetic}$:

$$\frac{1}{g^2}\int d^2\theta\, W_\alpha W^\alpha \sim \frac{1}{g^2}\int d^2\theta\, F\theta F\theta + \dots \sim \frac{1}{g^2}F^2 \tag{4.30}$$

where here F stands for the gauge field strength. So we can see the F-term is building up the ordinary gauge kinetic term.

$\mathcal{L}_{superpotential}$:

$$\int d^2\theta\, W(\Phi). \tag{4.31}$$

$d^2\theta$ is dimension 1, so W is dimension 3. Therefore the renormalizable superpotentials are of the form

$$W = \frac{1}{2}m_{ij}\Phi_i\Phi_j + \frac{1}{3}\Gamma_{ijk}\Phi_i\Phi_j\Phi_k \tag{4.32}$$

(We will understand linear terms later.) As mentioned, the F-component of Φ is a non-dynamical complex scalar. This is because

$$\mathcal{L}_{\text{kinetic}} \sim \int d^4\theta\, (\Phi \to \theta\theta F)^\dagger(\Phi \to \theta\theta F) \sim F^\dagger F \tag{4.33}$$

so there are no derivatives (no kinetic term) for F.

We also have, from $\Phi \sim \phi + \theta\psi + \theta^2 F$,

$$\int d^2\theta\, W = \left.\frac{\partial W}{\partial \Phi_i}\right|_{\Phi=\phi} F_i + \left.\frac{\partial^2 W}{\partial \Phi_i \partial \Phi_j}\right|_{\Phi=\phi} \psi_i \psi_j. \tag{4.34}$$

So the F_i-dependent terms in $\mathcal{L}$ are:

$$F_i^* F_i + \frac{\partial W}{\partial \Phi_i} F_i + \frac{\partial W^*}{\partial \Phi_i^!} F_i^*. \tag{4.35}$$

We can integrate out F_i by solving the equation of motion:

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta F} &= F_i^* + \frac{\partial W}{\partial F_i} = 0 \\ \Rightarrow F_i &= -\left(\frac{\partial W}{\partial \Phi_i}\right)^*. \end{aligned} \tag{4.36}$$

In a non-gauge theory, the F-dependent terms above are the only scalar potential terms. Substituting back, we have

$$V_F(\phi) = \sum_i |F_i|^2 = \sum_i \left|\frac{\partial W}{\partial \phi_i}\right|^2 \tag{4.37}$$

which is called the F-term potential.

In gauge theories there is also a scalar potential from gauge-auxiliary fields

$$\begin{aligned} &\frac{1}{g^2}\int d^2\theta\, W_\alpha^2 \to \frac{1}{g^2}\int d^2\theta\, (D^a)^2\theta^2 \to \frac{1}{g^2} D^a D^a, \\ &\int d^4\theta\, \Phi_i^+ V \Phi_i \to \phi_i^! T^a \phi_i \int d^4\theta\, \theta\theta\bar{\theta}\bar{\theta} D^a \to \phi_i^! T^a \phi_i D^a. \end{aligned} \tag{4.38}$$

(Note: in a nonabelian gauge theory, $V = V^a T^a$.)

D^a is nondynamical, so we can solve

$$\frac{\delta \mathcal{L}}{\delta D^a} = 0 \to D^a = g^2 \phi_i^\dagger T^a \phi_i \tag{4.39}$$

which produces, upon substituting back in,

$$V_D(\phi) = g^2 \left(\sum_i \phi_i^* T^a \phi_i\right)^2 \sim D^2 \tag{4.40}$$

and the total scalar potential has the form

$$V = |F|^2 + D^2 > 0. \tag{4.41}$$

Thus the ground state energy is positive-definite. Another way to see it: $\left\{Q_\alpha, \bar{Q}_\beta\right\} = 2P_\mu \sigma^\mu_{\alpha\dot{\beta}}$, and so

$$\mathrm{Tr}\left(\bar{\sigma}^0\{Q, \bar{Q}\}\right) = 2\,\mathrm{Tr}\, P_\mu \sigma^\mu \bar{\sigma}^0 \tag{4.42}$$

$$\Rightarrow \frac{1}{4}\left(Q_\alpha \bar{Q}_{\dot{\alpha}} + \bar{Q}_{\dot{\alpha}} Q_\alpha\right) = P_0 = E. \tag{4.43}$$

We can take the expectation value of both sides to conclude

$$\frac{1}{4}\left(\|Q|0\rangle\,\|^2 + \|\,\bar{Q}|0\rangle\|^2\right) = \langle E\rangle > 0 \tag{4.44}$$

Moreover, E_{ground} is an order parameter for SUSY breaking!

$$\begin{aligned} &Q|0\rangle = 0 \rightarrow \text{ SUSY unbroken, } E = 0 \\ &Q|0\rangle \neq 0 \rightarrow \text{ SUSY spontaneously broken, } E > 0 \end{aligned} \tag{4.45}$$

Now $\delta\psi \sim \epsilon F, \quad \delta\lambda \sim \epsilon D$ and $\delta\psi = i\{Q, \psi\}, \delta\lambda = i\{Q, \lambda\}$. So $\langle\delta\psi\rangle \neq 0$ or $\langle\delta\lambda\rangle \neq 0 \Rightarrow SSB$ $\Leftrightarrow \langle F\rangle, \langle D\rangle$ nonzero.

SSB leads to a goldstone fermion, the goldstino.

We an already see susy's impact on one hierarchy problem! If global SUSY $\leftrightarrow E_{\text{ground}} = 0$, then SUSY solves the c.c. problem. Unfortunately, it can't be an exact symmetry.

4.1.2 Simple models

Wess-Zumino Model

Now let's study the simplest 4D supersymmetric QFT, a model with one chiral ϕ and no gauge multiplets. The superpotential is

$$W = \frac{1}{2} m\phi^2 + \frac{\lambda}{3}\phi^3 \tag{4.46}$$

so the scalar potential is (using ϕ to denote both the superfield and its lowest component):

$$V = |F|^2 = \left|m\phi + \lambda\phi^2\right|^2 \tag{4.47}$$

from which we read off the scalar mass,

$$m_\phi^2 = |m|^2. \tag{4.48}$$

The fermion terms in the Lagrangian are:

$$\frac{1}{2}\frac{\partial^2 W}{\partial\phi^2}\psi\psi + cc \rightarrow \frac{1}{2}m\psi\psi + cc + \text{ Yukawas} \tag{4.49}$$

So $m_\psi = m$ as well. There is a Bose-Fermi degeneracy.

What are the symmetries? First, set $m = 0$. Then there is what is known as a $U(1)$ "R symmetry." It is a symmetry under which the θ's transform. Let $R_\theta = 1$. Then $\int d\theta\,\theta = 1 \Rightarrow R_{d\theta} = -1$, which in turn implies

$$\begin{aligned} L &\sim \int d^2\theta\, W \Rightarrow R_W = +2 \\ W &= \frac{1}{3}\phi^3 \Rightarrow R_\phi = 2/3. \end{aligned} \tag{4.50}$$

The superfield components transform as $\phi \rightarrow e^{i2/3\alpha}\phi, \psi \rightarrow e^{i(2/3-1)\alpha}\psi, F \rightarrow e^{i(2/3-2)\alpha}F$. Now a quadratic term in the superpotential would cary $R(\phi^2) = 4/3$, and therefore breaks the R symmetry. It cannot be generated radiatively. If $m \neq 0$, it is a spurion for R-breaking.

So this is a strange thing – the vanishing scalar mass is radiatively stable!

How does it work? For fermion, no diagram. For the boson,

diagram

We will see this extends beyond 1− loop and to finite m.

$U(1)$ gauge theory

Next we study supersymmetric QED with two charged chirals, ϕ^+, ϕ^-. First, take $W = 0$.

Then, using

$$D \sim (\phi^+)^*(1)\phi^+ + (\phi^-)^*(-1)\phi^- \tag{4.51}$$

we find the scalar potential

$$V(\phi^\pm) = \frac{1}{2}D^2 = \frac{g^2}{2}\left(\left|\phi^+\right|^2 - \left|\phi^-\right|^2\right)^2. \tag{4.52}$$

Ground states have $\langle D\rangle = 0$. But this does not imply $\phi^+ = \phi^- = 0$: we can have any $\phi^+ = v, \phi^- = ve^{i\alpha}$, up to a gauge transformation.

Thus we have a continuous infinity of distinct, degenerate vacua.

Pick some v. Then we have SSB a la Higgs.

$$\sum_{+,-} |(\partial_\mu + igA_\mu)\phi|^2 \to 2g^2v^2A^2 \tag{4.53}$$

so

$$m_A^2 = 4g^2v^2. \tag{4.54}$$

The gauge interactions also include

$$\int d^4\theta\, \phi^+ e^V \phi \supset \int d^4\theta\, (\bar{\theta}\psi^\dagger)(\theta^2\bar{\theta}\lambda^\dagger)(\phi) \tag{4.55}$$

which are mixed gaugino-fermion yukawa couplings. Upon Higgsing, these terms give rise to fermion Dirac masses,

$$\sqrt{2}gv\lambda\left(\psi_{\phi^+} - \psi_{\phi^-}\right). \tag{4.56}$$

The matter fermion and the gaugino pair up to make a massive vector multiplet.

Due to the D-flat directions in the scalar potential, parametrized by the arbitrary *magnitude* of v, there is also a massless multiplet unconnected to the Goldstone phenomenon!

Fluctuaations in v don't see any potential. The linear combination $\delta\phi^+ + \delta\phi^-$ is massless. Likewise, the phase α is undetermined. Together, these make up a massless complex scalar. Furthermore $\psi_{\phi^+} + \psi_{\phi^-}$ is also massless. So together they make up a whole massless chiral multiplet.

It had to be so: SUSY is unbroken and all components of a multiplet must share the same mass, because of the algebra:

$$\left[P^2, Q\right] = 0 \tag{4.57}$$

so *one* scalar flat direction $\Rightarrow$ a whole massless multiplet.

Such multiplets make up what is called the *moduli space*. It can always be parametrized by gauge invariant operators, in this case

$$\phi^+\phi^- \cong v^2 + v\left(\delta\phi^+ + \delta\phi^-\right). \tag{4.58}$$

4.1.3 Non-renormalization theorems

A more general supersymmetric EFT takes the form

$$\mathcal{L} = \int d^4\theta \underbrace{K(\phi_i, \phi_i^\dagger)}_{\text{Kahlerpot'l}} + \int d^2\theta \underbrace{W(\phi_i)}_{superpot'l} + \int d^2\theta \underbrace{f_a(\phi_i) W_\alpha^2}_{gauge\, coupling\, fn} + cc \tag{4.59}$$

Some important results:

- W is not corrected in perturbation theory beyond tree level.
- f is renormalized only at 1 loop.

The original proofs were diagrammatic, but in most cases simpler proofs we available, based on holomorphy and spurious symmetries. We'll illustrate the ideas.

First, consider the massive WZ model. Take $R_\phi = 1$. Then $R_\lambda = -1$, and we can treat λ as a spurion for R-breaking. Now consider possible renormalizations of m. $\lambda^*\lambda$ could appear consistent with the R-symmetry. But if we promote λ to the vev of a background chiral superfield, $\lambda \to \lambda(y, \theta)$, it is clear that λ^* cannot appear: W must be holomorphic in superfields.

What about other combinations? $\lambda\lambda$ has the wrong R-charge, $(\lambda^*\lambda)\,\lambda$ is non-holomorphic, etc. Dimension 5 operators? $k\phi\phi\phi\phi$ looks ok, if $k \sim \lambda^2$. But actually there is another symmetry (non-R): $\phi \to e^{i\beta}\phi$. Under this symmetry m and λ are spurions of charge $q_m = -2$ and $q_\lambda = -3$. The most general term in the effective superpotential would have the form

$$m\phi^2 F\left(\frac{\lambda\phi}{m}\right) \tag{4.60}$$

for general F, because the combination $\frac{\lambda\phi}{m}$ is neutral under both symmetries.

Now the coefficient of of $\phi^n \sim \frac{\lambda^{n-2}}{m^{n-4}}$. $n = 8 : \frac{d^6}{m^4}$. But these just correspond to tree graphs! E.g. diagram These are not part of the effective action. All higher orders in λ (loop graphs) are inconsistent with the form $m\phi^2 F\left(\frac{\lambda\phi}{m}\right)$. We conclude that

$$W_{\text{eff}} = W_{\text{tree}} \tag{4.61}$$

is exact.

K is not holomorphic, and can have general $\lambda^*\lambda$ renormalizations.

Now consider the gauge coupling function. Again we will treat g^{-2} as part of a chiral field. Define

$$S = \frac{8\pi^2}{g^2} + ia + \mathcal{O}(\theta^\alpha). \tag{4.62}$$

The imaginary part of the lowest component, a, couples to $F\tilde{F}$. The real part, $\sim 1/g^2$, couples to FF.

In perturbation theory, $a \to a + c$ does not affect physics, because $F\tilde{F}$ is a total derivative. The effective action should reflect this as well.

Now the gauge coupling function is holomorphic, so

$$f(g^2) \to f(S) = S + const. \tag{4.63}$$

This is the only form consistent with the shift symmetry of $a = \text{Im}(S)$ in perturbation theory.

Now, at one loop,

$$\frac{8\pi^2}{g^2(\mu)} = \frac{8\pi^2}{g^2(M)} + \beta \log \mu/M. \tag{4.64}$$

So the constant in f can get renormalized at one loop. But higher terms are $\mathcal{O}(g^2)$ and are forbidden.

We will use the nonrenormalization of W to solve the electroweak hierarchy problem in part. But it will not be exact. To understand this, we have to discuss susy breaking. If realized in nature at all, susy must be broken, because we do not observe scalar electrons, etc,

So we turn now to the subject of susy breaking.

4.1.4 SUSY Breaking

O'Raifeartaigh Models

Simple example of a spontaneous SUSY-breaking model:

$$\begin{aligned} W &= \lambda A \left(X^2 - \mu^2\right) + mBX \\ F_A &= \frac{\partial W}{\partial A} = \lambda \left(X^2 - \mu^2\right) \\ F_B &= \frac{\partial W}{\partial B} = mX \\ F_X &= \frac{\partial W}{\partial X} = 2\lambda AX + mB. \end{aligned} \tag{4.65}$$

From these F-terms we see that $F_A = 0$ and $F_B = 0$ are incompatible with each other

Rather than workwith SSB models, it is convenient to introduce explicit susy breaking in a controlled, or "soft" way. Take the WZ model with $W = \frac{1}{3}\lambda\Phi^3,\ K = \Phi^\dagger\Phi$.

SUSY implies $m_\psi = m_\phi$ (and both vanish in this model), and W is not renormalized, so no mass term $m\Phi\Phi$ is generated. In particular the scalar is massless to all orders in the SUSY limit. This is surprising from the point of view of generic EFT! Now add an explicit SUSY breaking "soft" scalar mass to the potential:

$$V_\phi = |F_\phi|^2 + m^2_{\text{soft}}\,|\phi|^2. \tag{4.66}$$

Since susy is broken, the nonrenormalization theorems fail. What is Δm^2 ?

$$F_\Phi = \partial W/\partial\Phi = \lambda\phi^2 \Rightarrow V = \lambda^2|\phi|^4 + m^2_{\text{soft}}\,|\phi|^2. \tag{4.67}$$

We also have the interaction $\int d^2\theta W \supset \lambda\psi\psi\phi$. So there are the following self-energy contributions:

diagram

$$\sim -\frac{\lambda^2\Lambda^2}{16\pi^2} - \frac{\lambda^2 m^2_{\text{soft}}}{16\pi^2}\log\left(\Lambda/m_s\right) + \dots \tag{4.68}$$

diagram

$$\sim \frac{\lambda^2\Lambda^2}{16\pi^2} \tag{4.69}$$

The sum is

$$-\frac{\lambda^2 m^2_{\text{soft}}}{16\pi^2}\log(\Lambda/m_S) \tag{4.70}$$

So the far-UV sensitivity, the quadratic divergence $\sim \Lambda^2$ still cancels! But there is a non-canceling part that "knows" about SUSY breaking. The renormalization proportional to m^2_{soft} is still small if m^2_{soft} is small. Dimensional analysis: this mass squared was a spurion for susy breaking, so it must appear in all susy breaking effects: no room for Λ^2, except in logarithms.

More generally, scalar masses are not sensitive to arbitrary UV scales if the sources of susy breaking are parametrized by dimensionful soft terms:

Soft scalar masses: $m^2|\phi|^2 + B(\phi\phi + cc)$
Gaugino mass: $m_\lambda\lambda\lambda + cc$ (splits gaugino from A^μ)
trilinear scalar couplings: $\Gamma\phi\phi\phi + cc$

Hard susy breaking, on the other hand, generates UV sensitivity. E.g., suppose

$$\mathcal{L} \sim -\lambda^2|\phi|^4 - \tilde{\lambda}\phi\psi\psi + cc. \tag{4.71}$$

Here hard susy breaking is present whenever the dimensionless couplings λ and $\tilde{\lambda}$ are unequal. The sum of the one-loop scalar self energy diagrams is of order

$$\Delta m^2 \sim \frac{(\tilde{\lambda}^2 - \lambda^2)}{16\pi^2}\Lambda^2 \tag{4.72}$$

So the nonrenormalization theory is badly broken in this case.

Let's finish by briefly discussing extending the SM to a minimal supersymmetric standard model (MSSM).