

An Introduction to Effective Field Theories

Riccardo Penco*

Department of Physics, Carnegie Mellon University, Pittsburgh, PA

June 19, 2020

Abstract

The notes are an overview of effective field theory (EFT) methods. I discuss toy model EFTs, chiral perturbation theory, Fermi liquid theory, and non-relativistic QED, and use these examples to introduce a variety of EFT concepts, including: matching a tree and loop level; summation of large logarithms; naturalness problems; spurion fields; coset construction; Wess-Zumino-Witten terms; chiral and gauge anomalies; field rephasings; and method of regions. These lecture notes were prepared for the 2nd Joint ICTP-Trieste/ICTP-SAIFR school on Particle Physics, Sao Paulo, Brazil, June 22 - July 3, 2020.

*Please contact me at rpenco@cmu.edu for comments or corrections.

Contents

1	Introduction	3
1.1	Main ingredients of EFTs	3
1.2	Main advantages of EFTs	5
2	A toy effective field theory	7
2.1	Matching tree level amplitudes	8
2.2	Power counting and strong coupling scale	11
2.3	Matching at loop level	13
2.4	Summing large logarithms	18
2.5	A modern perspective on renormalizable theories	19
2.6	Naturalness	20
2.7	Additional resources	25
3	EFTs for Nambu-Goldstone modes	26
3.1	Chiral perturbation theory	26
3.2	Approximate symmetries and spurions	31
3.3	Naturalness and pseudo-Nambu-Goldstone modes	34
3.4	Coset construction	35
3.5	Inverse Higgs constraints	40
3.6	Wess-Zumino-Witten terms	41
3.7	Chiral anomalies	44
3.8	Gauge anomalies	45
3.9	Additional resources	47
4	EFTs for non-relativistic fermions	48
4.1	Non-relativistic fermions in vacuum	48
4.2	Contracted and accidental symmetries	50
4.3	Non-relativistic fermions at finite density	51
4.4	Non-relativistic QED	57
4.5	Additional resources	63

1 Introduction

A remarkable fact about nature is that interesting phenomena occur over a wide range of energy, length, and time scales. Over the last century alone, we have managed to stretch the scope of our investigations from microscopic scales as small as one billionth of the size of an atom (10^4 GeV) to macroscopic scales as large as the size of our observable universe (10^{-42} GeV). This huge energy range is key to the success of physical sciences. If all physical phenomena had taken place at the same scale, we would have likely had to understand them all at once: progress in physics would have been impossible! Instead, in a first approximation we can take advantage of this separation of scales to set to infinity (zero) all scales that are much larger (smaller) than the typical energy E of whatever physical process we happen to be interested in. In this limit, physics at scales much different from E becomes irrelevant and can be neglected; if needed, its effect can be reintroduced in perturbation theory. To this day, this basic strategy is adopted by physicists in all subfields in a more or less implicit way.

Effective field theories (EFTs) are a model building tool that explicitly implements the strategy outlined above and turns it into a precise, quantitative framework. Particle physicists like to think in terms of energy scales, but EFTs can be built by leveraging hierarchies between all sorts of dimensionful quantities: lengths, times, velocities, momenta, angular momenta, etc... In fact, one could go as far as claiming that, when properly formulated,

every theory in physics is an effective theory.

In what follows, I will introduce some of the main weapons in the EFT arsenal by showing how they work in a variety of different contexts. These notes are aimed at graduate students and junior postdocs with a high energy physics background. In fact, I tried to write the notes that I would have liked to read as a graduate student. Most of the content will therefore be familiar to the EFT practitioners, although I hope that they too might still find something valuable here and there.

1.1 Main ingredients of EFTs

I will not follow a historical order of presentation, which would likely start with a discussion of Wilson's approach to the renormalization group (RG). Instead, I will jump right in and briefly summarize our present understanding of EFTs, since the main principles are very easy to state. An EFT is defined by an effective action, which in turn is completely specified by the following three ingredients:

1. *Degrees of freedom.* The first step when building an EFT is to figure out what are the degrees of freedom that are relevant to describe the physical system one is interested

in. These are the variables that will appear in the effective action. The key word here is “relevant”: you can always complicate the description of any phenomenon by adding additional structures, but you’ll soon reach a point of diminishing return. Conversely, you can strive for the most economical description but, in the words of Albert Einstein, *“everything should be made as simple as possible, but no simpler”*. Sometimes the degrees of freedom to be used are suggested by symmetry considerations, as is the case for Nambu-Goldstone modes; more often, though, the choice of degrees of freedom is an independent input, and it’s a matter of art as much as science.

2. *Symmetries*. The second step in building an EFT consists in identifying the symmetries that constrain the form of the effective action, and therefore the dynamics of the system. Symmetries can come in many different flavors: they can be global, gauged, accidental, spontaneously broken, anomalous, approximate, contracted, etc... . We will encounter examples of all these possibilities in due time. For now, I will just mention that any term that is compatible with the symmetries of the system should in principle be included in the effective action. As a result, effective actions generically contain an infinite number of terms.
3. *Expansion parameters*. The key to handling an action with an infinite number of terms lies in the fact that all EFTs feature one or more expansion parameters. These are small quantities controlling the impact that the physics we choose to neglect could potentially have on the degrees of freedom we choose to keep. For example, in particle physics these expansion parameters are often ratios of energy scales E/Λ , where E is the characteristic energy scale of the process one is interested in, and Λ is the typical energy scale of the UV physics one is neglecting. Other examples of expansion parameters include ratios of velocities (*e.g.* v/c in a non-relativistic limit), angular momenta (as in a semi-classical expansion), or small dimensionless couplings. Observable quantities are calculated in perturbation theory as series in these small parameters. For this strategy to work, it is crucial to have an explicit power counting scheme, meaning that we should be able to assign a definite order in the expansion parameter to each term in the effective action. This ensures that only a finite number of terms contribute at any given order in perturbation theory, and that we can decide upfront which terms to keep in the action based on the desired level of accuracy.

These three elements are fairly easy to state, but ensuring that they are properly implemented in an EFT can be a subtle matter. The way in which this is achieved in practice often depends on the problem at hand. A great deal of know-how has been accumulated over the last half a century, and the goal of these notes is to introduce you to the main tools of the trade.

1.2 Main advantages of EFTs

Sometimes it is necessary to resort to an EFT description of a physical system because a more fundamental theory is lacking. Other times such a theory might be available, but extracting exact or even approximate predictions might be difficult, for example because the theory is strongly coupled. In these circumstances, EFTs can provide a more convenient, weakly coupled description of the same physics. Finally, even when EFTs are not forced upon us by necessity, they still remain a convenient framework. The main advantages of an EFT approach can be summarized as follows:

- EFTs drastically simplify calculations by focusing on the relevant degrees of freedom, and neglecting from the get-go those aspects that are not important for the problem at hand.
- When isolating the relevant degrees of freedom, new symmetries may become manifest that otherwise would have remained obscured. Often times, these symmetries can be used to draw general conclusions based on little or no calculations.
- By working with an EFT, we essentially factorize calculations into two parts: one that involves the degrees of freedom that we are keeping (the actual EFT calculation), and another one that depends on the physics we are neglecting (the matching calculation). By adopting this “modular” approach, one can avoid repeating similar calculations over and over again.
- When dealing with problems that feature several different scales, there may be observables that depend on logarithms of their ratios. When these scales are widely separated, the logarithms become large and can worsen the accuracy of perturbative calculations. By focusing on one scale at a time, EFTs can use RG running to sum up these large logarithms.
- EFTs contain a systematic parametrization of the neglected physics. When such physics is unknown or poorly understood, one can place model-independent constraints on it by comparing EFT predictions with experiments.
- Observable quantities in an EFT are series in one or more expansion parameters. How far one goes in perturbation theory determines the precision of EFT results.
- Related to this, EFTs are kind enough to let you know when you should stop trusting their predictions. When you enter a regime where the expansion parameters are of $\mathcal{O}(1)$, perturbation theory stops working. This means that the degrees of freedom you have been using are no longer the relevant ones, and you should come up with a different EFT.

Some of these remarks might sound a bit obscure at this point. If so, I encourage you to continue reading, and come back to this section after we've discussed a few examples of EFTs.

Conventions: I will be following the conventions of [1], unless otherwise specified. This means in particular that I will be using a metric with “mostly plus” signature and I will be working in units such that $\hbar = c = 1$. I will also work in Gaussian units for electromagnetism. A pair of equal indices should always be understood as contracted when one is raised and the other one is lowered. I will underline the key concepts I introduce throughout these lectures. I will only cite sources to provide additional background material or additional details, not to attribute credit. Apologies in advance to those who may feel that their contributions haven't been properly acknowledged. I will assume familiarity with quantum field theory, and refer the reader to [1] for standard material.

2 A toy effective field theory

In order to introduce some of the basic ideas and techniques surrounding effective field theories, we will start by considering the following toy model:

$$S = \int d^4x \left\{ -\frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} M^2 \phi^2 + \bar{\Psi} (i \gamma^\mu \partial_\mu - m) \Psi + i y \phi \bar{\Psi} \gamma^5 \Psi - \frac{1}{4!} \lambda \phi^4 \right\}. \quad (1)$$

This action is invariant under parity provided ϕ is a pseudoscalar, *i.e.* it transforms like $\phi \rightarrow -\phi$ under a parity transformation. A cubic self-interaction for ϕ would not be invariant and has therefore been omitted. We will consider a region of parameter space where $y, \lambda \ll 1$, so that we can work in perturbation theory. Moreover, we will assume that the scalar is much heavier than the fermion, *i.e.* $M \gg m$.

Imagine now that we are interested in describing a simple process such as the elastic scattering of two fermions. At lowest order in perturbation theory, *i.e.* at tree-level, this interaction is mediated by the exchange of a single scalar. The relevant Feynman diagrams are shown in Fig. 1, and the resulting scattering amplitude can be obtained using standard Feynman rules (see *e.g.* [1]):

$$i\mathcal{M} = (-y)^2 \left\{ \bar{u}_{\lambda_3, p_3} \gamma^5 u_{\lambda_1, p_1} \frac{-i}{(p_3 - p_1)^2 + M^2} \bar{u}_{\lambda_4, p_4} \gamma^5 u_{\lambda_2, p_2} - (3 \leftrightarrow 4) \right\}. \quad (2)$$

If the typical energy E of the fermions involved in this process is such that

$$m \ll E \ll M, \quad (3)$$

the scalar propagator can be expanded in powers of momenta, and at lowest order our amplitude reduces to

$$i\mathcal{M} \simeq -i \frac{y^2}{M^2} \left[\bar{u}_{\lambda_3, p_3} \gamma^5 u_{\lambda_1, p_1} \bar{u}_{\lambda_4, p_4} \gamma^5 u_{\lambda_2, p_2} - (3 \leftrightarrow 4) \right], \quad (4)$$

up to corrections of $\mathcal{O}(E^2/M^2)$. Interestingly, one could have also obtained this result starting from an *effective field theory* (EFT) described by the action

$$S_{\text{eff}} = \int d^4x \left\{ \bar{\Psi} (i \gamma^\mu \partial_\mu - m) \Psi - \frac{y^2}{2M^2} (\bar{\Psi} \gamma^5 \Psi)^2 \right\} \quad (5)$$

which makes no reference whatsoever to the heavy scalar field. This makes intuitive sense. Within the energy range in Eq. (3), conservation of energy prevents heavy scalars from being created, and if there are no scalars in the initial state we should be able to describe what's going on solely in terms of fermions. This intuition is based on our successful track record in modeling all sorts of physical phenomena without worrying about those aspects of Nature we haven't discovered yet. In the context of our toy model, this means that physicists who don't have access to energies above M should still be able to come up with a theory that describes experimental results.

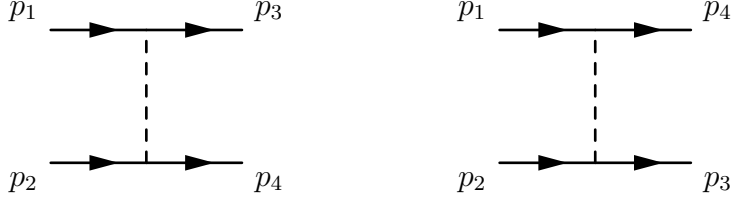


Figure 1: $2 \rightarrow 2$ fermion scattering at tree-level.

Importantly, as long as E^2/M^2 is smaller than the precision with which one can measure the scattering amplitude \mathcal{M} , the theories (1) and (5) will yield predictions that are virtually indistinguishable. Hence, by Occam’s razor, we might as well use the EFT (5), because it provides a simpler and equally accurate explanation for the scattering process under consideration. The ratio E/M controls how well our EFT approximates the full theory: it is our expansion parameter. At energies $E > M$, our EFT ceases to be useful and we need to resort to the full theory.

A remarkable feature of the effective action (5) is its *locality*: S_{eff} contains only products of Ψ and its derivative evaluated at the same point. Locality follows from the fact that ϕ mediates an interaction between the fermions which at lowest order in perturbation theory is described by the Yukawa potential

$$V(r) = -\frac{y^2}{4\pi} \frac{e^{-Mr}}{r} . \quad (6)$$

The processes we are considering are such that the momentum exchanged by the fermions is $q \sim E \ll M$. Therefore, we are probing the Yukawa potential with a resolution $\Delta r \sim 1/q \gg 1/M$ which is much larger than the range of our potential. In this limit, $V(r)$ reduces essentially to a contact interaction: fermions can only interact with each other if they are at the same point (within the limits imposed by the uncertainty principle).

2.1 Matching tree level amplitudes

The effective action (5) correctly reproduces the tree-level 4-point amplitude of fermions one would get from the theory (1) at lowest order in E/M . It is then natural to ask how far our effective theory approach can be pushed: can we also recover higher order corrections in E/M to the 4-point function? What about higher-point functions? And loop diagrams? In this section we will address the first two questions, postponing the third one until Sec. 2.3.

Based on the discussion in the previous section, our EFT in Eq. (5) should contain not just a handful of terms, but in fact all possible terms one can build out of Ψ and its derivatives while preserving all the symmetries. Each of these terms will enter the effective Lagrangian multiplied by an a priori arbitrary coupling constant, or *Wilson coefficient*. The infinite

number of these coefficients reflects the fact that there are in principle an infinite number of theories that at low enough energies contain a single Dirac field Ψ and have the same symmetries. The process of figuring out the precise values that the Wilson coefficients must take in order to reproduce the low-energy physics of the particular theory one is interested in—in our case, the toy model (1)—is known as *matching*. In practice, matching calculations are usually carried out at the level of individual amplitudes and up to the desired order in the loop expansion—similarly to what we did in the previous subsection for the 4-point function at tree level. It is however instructive to discuss how this procedure can be understood directly at the level of the generating functional.

Scattering amplitudes can be extracted from time ordered correlation functions of fields by using the LSZ reduction formula [1]. In turn, any time ordered correlator can be obtained from the generating functional

$$Z[J, \bar{\eta}, \eta] = \int D\phi D\bar{\Psi} D\Psi e^{iS[\phi, \Psi] + i \int (J\phi + \bar{\eta}\Psi + \bar{\Psi}\eta)}, \quad (7)$$

by taking a suitable number of derivatives with respect to the external currents $J, \bar{\eta}$ and η , and then setting them to zero. If we are only interested in scattering processes involving fermions, then all we need are correlation functions of Ψ and $\bar{\Psi}$. We can therefore set $J = 0$ in Eq. (7) from the get-go, in which case the path integral can be factored out as follows:

$$\begin{aligned} Z[J = 0, \bar{\eta}, \eta] &= \int D\bar{\Psi} D\Psi e^{i/\hbar \int (\bar{\eta}\Psi + \bar{\Psi}\eta)} \int D\phi e^{iS[\phi, \Psi]/\hbar} \\ &\equiv \int D\bar{\Psi} D\Psi e^{i/\hbar \int (\bar{\eta}\Psi + \bar{\Psi}\eta)} e^{iS_{\text{eff}}[\Psi]/\hbar}, \end{aligned} \quad (8)$$

where we have momentarily restored all factors of \hbar . From this perspective, the effective action for Ψ is obtained by *integrating out* the scalar field. Performing the integral over ϕ exactly captures the contribution of internal scalar lines to Feynman diagrams with external fermions. This leads to an effective action that admits an expansion in powers of \hbar of the form $S_{\text{eff}} = S_{\text{eff}}^{(0)} + \hbar S_{\text{eff}}^{(1)} + \mathcal{O}(\hbar^2)$, with the pieces proportional to \hbar coming from loops of scalars. The leading-order term in this expansion, $S_{\text{eff}}^{(0)}$, describes an EFT that reproduces all tree-level amplitudes of fermions in the full theory (1); the next-to-leading-order term, $S_{\text{eff}}^{(1)}$, reproduces all 1-loop corrections; and so on.

The tree-level effective action $S_{\text{eff}}^{(0)}$ is particularly easy to calculate, as it can be obtained by performing a saddle point approximation of the functional integral over ϕ , *i.e.*

$$e^{iS_{\text{eff}}^{(0)}[\Psi]/\hbar} = e^{iS[\hat{\phi}, \Psi]/\hbar}, \quad \text{with } \hat{\phi}[\Psi] \text{ a solution of } \frac{\delta S[\hat{\phi}, \Psi]}{\delta \hat{\phi}} = 0. \quad (9)$$

The solution $\hat{\phi}$ can be calculated perturbatively in the expansion parameter E/M of our EFT. The field Ψ plays the role of an external field for the purposes of this calculation.

Varying the action (1) with respect to ϕ we find the equation of motion

$$(\square - M^2)\phi + y\mathcal{J} - \frac{\lambda}{3!}\phi^3 = 0, \quad (10)$$

where we have streamlined our notation by defining $\mathcal{J} \equiv \bar{i}\Psi\gamma^5\Psi$. Postponing for now a detailed discussion about the systematics of the E/M expansion, we can solve perturbatively this equation by expanding the solution in inverse powers of M :

$$\hat{\phi} = \frac{y}{M^2} \left(1 + \frac{\square}{M^2} + \frac{\square^2}{M^4} + \frac{\square^3}{M^6} \right) \mathcal{J} - \frac{\lambda y^3}{6M^8} \mathcal{J}^3 + \dots \quad (11)$$

Plugging this solution back into the action for our toy model, and using Eq. (9), we find

$$S_{\text{eff}}^{(0)} = \int d^4x \left\{ \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi + \frac{y^2}{2M^2} \mathcal{J} \left(1 + \frac{\square}{M^2} + \frac{\square^2}{M^4} + \frac{\square^3}{M^6} \right) \mathcal{J} - \frac{\lambda y^4}{4!M^8} \mathcal{J}^4 + \dots \right\}. \quad (12)$$

This is an improved version of the effective action (5), which now captures all tree level amplitudes, not just the 4-point one. It is clear that this action contains an infinite number of terms—in fact, all possible terms that are compatible with the symmetries. The Wilson coefficients in front of the various terms are very specific functions of the parameters M, y, λ in the full theory. This particular choice of couplings ensures that the tree-level scattering amplitudes in our EFT are exactly the same as the ones we would have obtained using the full theory: we have successfully matched our EFT onto our toy model at tree-level.

We have been able to perform the matching procedure explicitly because (1) we knew the full theory at energies $E > M$, and (2) this theory was weakly coupled. The assumption of weak coupling, in particular, allowed us to read off the relevant degrees of freedom directly from the Lagrangian, and to figure out that the expansion parameter at low energies should be E/M . It is often the case that one of the two conditions above is not satisfied. One then needs to rely on experiments, brilliant theoretical intuition, or a combination of both, to figure out what are the relevant symmetries and degrees of freedom at low energies.

If the full theory is known and strongly coupled but the low-energy degrees of freedom are weakly coupled, as is the case for QCD, the matching procedure can be carried out numerically. If instead the full theory is unknown, as is the case for the Standard Model, one can still write down an effective action with arbitrary Wilson coefficients. Measuring these coefficients experimentally, or even just placing upper bounds on their values, can give us important clues about the full underlying theory. Thus, EFTs remain a helpful tool to describe the physics one is interested in, and to parametrize systematically the physics one is neglecting.

One last comment. Matching is ultimately about ensuring that the EFT and full theory make the same exact predictions for physical observables, *e.g.* on-shell scattering amplitudes.

Nevertheless, our generating functional approach showed that one doesn't necessarily have to compare observable quantities to read off the Wilson coefficients. You can match by comparing unphysical quantities that are easier to calculate (*e.g.* off-shell amplitudes with vanishing external momenta), and then use the matched EFT to derive physical predictions.

2.2 Power counting and strong coupling scale

At this point, some readers might feel a bit uneasy about actions that include an infinite number of terms. This is especially true if the Wilson coefficients are to be treated as arbitrary parameters. The plethora of models one encounters while studying physics are all described by finite Lagrangians with just a handful of parameters. How is one supposed to handle an infinite Lagrangian? An infinite number of measurements would seem necessary to determine all Wilson coefficients, in which case how can such a theory be useful, or even predictive? As we will now discuss, the answer to these questions lies in the third defining property of an EFT: its expansion parameter(s).

The 4-point amplitude in Eq. (4) was the leading term in an expansion in powers of E/M . In fact, all physical predictions of our EFT should be expressed as a series in E/M . How far one should push this expansion depends on how precise our measurements can be: there is no point in including relative corrections of order $(E/M)^n$ if these contributions cannot be measured. EFTs are a very pragmatic business.

A moment of reflection should be sufficient to realize that all the terms in the tree-level effective Lagrangian (12) that are suppressed by more than two powers of M either give rise to higher-point amplitudes, or contribute to the 4-point function at higher order in E/M . The interaction shown in Eq. (4) is the only one that can give a $1/M^2$ contribution to the 4-point function. Thus, if an experiment is only sensitive to the 4-point function at lowest order in E/M , then there is only one parameter to be measured.

More in general, our EFT is a useful framework as long as the outcome of any low-energy experiment with finite precision is described by a finite number of parameters. Then, only a finite number of measurements is needed to determine the values of these parameters and thus specify the EFT for all practical purposes. Our EFT is also a predictive framework because, after its parameters have been fixed, say, by measuring some scattering amplitudes at a certain energy, it can be used to make predictions for other observables (*e.g.*, to predict the values of those same amplitudes at different energies).

In practice, for any given experiment one must be able to determine at the outset which terms in the effective Lagrangian should be kept and which ones can be neglected. This can be achieved provided one can tell a priori at which order in E/M each term will contribute to observable quantities. The process of figuring out how each term in the effective action

scales with the expansion parameter(s) is known as power counting. Let's see how this works in practice for our toy EFT.

Within the energy range in Eq. (3), the 4-momentum of each fermion involved in a scattering process scales with energy like $p^\mu \sim E$. This is also how each derivative acting on Ψ should scale in this regime, *i.e.* $\partial_\mu \sim E$. Invoking the uncertainty principle, we can then deduce that $x_\mu \sim 1/E$ and $d^4x \sim 1/E^4$. Finally, for a weakly coupled effective theory it seems reasonable to assume that the kinetic term be the leading term in an E/M expansion. This requirement fixes the scaling of Ψ :

$$\int d^4x \bar{\Psi} i \gamma^\mu \partial_\mu \Psi \sim 1 \quad \implies \quad \Psi \sim E^{3/2}. \quad (13)$$

Using these scaling rules, we can associate a definite power of E/M to each operator in the Lagrangian. For example, the last term in Eq. (12) scales as follows:

$$\int d^4x \left\{ -\frac{\lambda y^4}{4! M^8} \mathcal{J}^4 \right\} \sim \left(\frac{E}{M} \right)^8. \quad (14)$$

In this simple example, the fact that the matched Wilson coefficient was proportional to $1/M^8$ already gave us a hint on how this term should scale with our expansion parameter. As we will see later on, figuring out the scaling of operators in the Lagrangian is often not so easy. The crucial point though is that each operator must scale like a definite power of the expansion parameter in order for the EFT to be well defined.

The scaling rules we have developed so far allow us to determine the relative importance of different terms in the Lagrangian. This sheds some light on why the Lagrangians you might have encountered so far contained only a handful of terms: for many applications, the expansion parameter is sufficiently small that considering only the most important terms is good enough.

When $E \sim M$, instead, all interactions become as important as the kinetic term, and perturbation theory stops working. For this reason, the scale M is usually referred to as the strong coupling scale of the EFT. More colloquially, many people will also call M the cutoff of the EFT—a hint at the fact that this is also the scale at which loops of slow modes are cut off in Wilson's approach to the renormalization group. In what follows I will refrain from using this term since, as we'll discuss in the next section, a consistent power counting scheme at loop level requires a mass independent regularization scheme such as dimensional regularization.

Let's now discuss how we can use our scaling rules to figure out the leading order contribution of each term in the Lagrangian to observable quantities. In the case of scattering amplitudes, we can use the LSZ reduction formula [1], which relates the probability amplitude for a scattering process $i \rightarrow f$ involving n fermions to an n -point function. For approximately

massless fermions, the LSZ formula can be written very schematically as

$$\langle f|i\rangle \equiv (2\pi)^4 \delta^4(p_f - p_i) i\mathcal{M}_n \sim \left\{ \prod_{j=1}^n \int d^4x_j e^{\pm i p_j \cdot x_j} \not{x}_j u_j \right\} \langle T\Psi(x_1) \dots \Psi(x_n) \rangle. \quad (15)$$

We've been very liberal in writing this equation, since some of the u_j 's could be \bar{u}_j 's, v_j 's or \bar{v}_j 's, and some of the Ψ 's could be $\bar{\Psi}$'s. Furthermore, we have suppressed all polarization indices and haven't worried too much about the order in which the various factors appear. All these details won't matter for the purposes of figuring out how the amplitude \mathcal{M}_n scales with E/M .

We need however to find out how the mode functions u_j scale with the energy. To this end, recall that the free operator Ψ can be decomposed into creation and annihilation operators as follows:

$$\Psi(x) = \sum_{\lambda} \int \frac{d^3p}{(2\pi)^3 2E_{\vec{p}}} \left[b_{\lambda, \vec{p}} u_{\lambda}(\vec{p}) e^{i\vec{p} \cdot x} + d_{\lambda, \vec{p}}^{\dagger} v_{\lambda}(\vec{p}) e^{-i\vec{p} \cdot x} \right], \quad (16)$$

with λ the polarization index and $[b_{\lambda, \vec{p}}, b_{\lambda', \vec{p}'}^{\dagger}] = [d_{\lambda, \vec{p}}, d_{\lambda', \vec{p}'}^{\dagger}] = 2E_{\vec{p}} (2\pi)^3 \delta(\vec{p} - \vec{p}') \delta_{\lambda\lambda'}$. These commutators determine the scaling of the creation and annihilation operators, and then we can use Eq. (13) to conclude that $u_j \sim E^{1/2}$.¹

Finally, the n -point function on the righthand side of Eq. (15) is given by

$$\langle T\Psi(x_1) \dots \Psi(x_n) \rangle = \frac{\int D\bar{\Psi} D\Psi \Psi(x_1) \dots \Psi(x_n) e^{iS_{\text{eff}}[\Psi]}}{\int D\bar{\Psi} D\Psi e^{iS_{\text{eff}}[\Psi]}}. \quad (17)$$

Combining this result with the LSZ formula and all the scaling rules we derived in this section, we can estimate how each term in the Lagrangian contributes to the amplitude \mathcal{M}_n in perturbation theory. For example, the leading order perturbative contribution to a n -point function coming from a term ΔS in the effective action with n fields will scale like $\Psi(x_1) \dots \Psi(x_n) \Delta S$.

Exercise 2.1: Show that contribution of a quartic interaction $\Delta S = \int d^4x \mathcal{O}(x)$ to the 4-point amplitude \mathcal{M}_4 of fermions scales like $\mathcal{O}(x)$.

Exercise 2.2: Determine how the 6-point amplitude \mathcal{M}_6 scales with energy at leading order in the effective theory in Eq. (12).

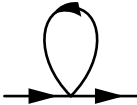
2.3 Matching at loop level

We have argued so far that an EFT described by the action (12) can reproduce all tree level scattering amplitudes calculated using the toy model (1) at low energies. We will now show

¹This is also consistent with the fact that $u_{\lambda}(\vec{p}) \bar{u}_{\lambda}(\vec{p}) \sim \not{p}$. [1]

how the matching procedure can be successfully carried out also at loop level. We will learn that loop diagrams built from the tree level effective action $S_{\text{eff}}^{(0)}$ aren't sufficient to reproduce all 1-loop corrections in the full theory. This is fine, of course, because $S_{\text{eff}}^{(0)}$ is just the tree-level part of the effective action. This apparent mismatch between loop corrections in the EFT and full theory is fixed by the 1-loop part of the effective action, $S_{\text{eff}}^{(1)}$. In fact, comparing loop diagrams is often a more efficient way of calculating $S_{\text{eff}}^{(1)}$ in practice, as opposed to calculating the path integral in Eq. (8). The advantage of this method becomes even more striking at higher loop order. To illustrate this procedure, we will focus for simplicity on the 1-loop corrections to the 2-point function of fermions, and we'll work at lowest order in the E/M expansion. This will suffice to introduce the main subtleties that arise when matching loop contributions.

Let's start by considering 1-loop corrections in the EFT. For our purposes, it will be sufficient to calculate 1PI diagrams rather than the full propagator. At lowest order in E/M , there is only one Feynman diagram that contributes to the 2-point function of fermions:



$$= -i \frac{y^2}{M^2} \int \frac{d^4 k}{(2\pi)^4} \gamma^5 \frac{-i(-\not{k} + m)}{k^2 + m^2 - i\epsilon} \gamma^5 = -\frac{y^2 m}{M^2} \int \frac{d^4 k}{(2\pi)^4} \frac{1}{k^2 + m^2 - i\epsilon}, \quad (18)$$

where in the last step we used $(\gamma^5)^2 = 1$, as well as the fact that the contribution linear in k must vanish by symmetry. This diagram clearly has a quadratic divergence in the UV. On physical grounds, one might be tempted to regularize this divergence by cutting off the integral over k at the scale M . After all, our EFT is only supposed to be valid at energies $E < M$, where our expansion parameter is smaller than one. Moreover, this approach would resonate with Wilson's perspective on the renormalization group, where low energy theories for coarse grained observables are obtained by integrating out "fast" Fourier modes [2]. And yet, this would be a bad idea. By doing so, the resulting factor of M^2 would cancel against the $1/M^2$ in the Wilson coefficient, giving a result that is independent of the UV scale M . This would directly contradict our power counting rules, according to which a quartic self-interaction should always be suppressed by a factor of E^2/M^2 .

The most common solution to this problem is to use instead dimensional regularization. This method presents multiple advantages: in a gauge theory, it respects gauge invariance; in conjunction with the modified minimal subtraction scheme (or \overline{MS} scheme, for short), it greatly simplifies the renormalization process; and, as we will now show, it doesn't spoil power counting. By switching to $d = 4 - \varepsilon$ dimensions², the righthand side of Eq. (18)

²We are using a "curly" epsilon (ε) to quantify deviations from $d = 4$. This should not be confused with the "script" epsilon (ϵ) that appears for instance in the denominator of Eq. (18).

Since our EFT calculation was valid only up to $\mathcal{O}(E^2/M^2)$, we should expand Eq. (23) in powers of m^2/M^2 and p^2/M^2 before making any comparison. Keeping all the terms up to $\mathcal{O}(M^{-2})$, we find

$$\begin{aligned}
 \text{Feynman diagram} &= \frac{iy^2}{16\pi^2} \left\{ -\frac{1}{\varepsilon}(\not{p} + 2m) + \frac{\not{p}}{2} \left[-\frac{1}{2} + \log\left(\frac{M^2}{\mu^2}\right) + \frac{m^2}{M^2} \right] \right. \\
 &\quad \left. + m \left[-1 + \log\left(\frac{M^2}{\mu^2}\right) + \frac{m^2}{M^2} \log\left(\frac{M^2}{m^2}\right) + \frac{p^2}{M^2} \right] + \mathcal{O}(\varepsilon, p^3, M^{-4}) \right\}.
 \end{aligned} \tag{24}$$

Exercise 2.4: Derive the result in Eq. (24).

Having calculated the 1-loop correction to the 2-point function of fermions both in the EFT, Eq. (19), and in the full theory, Eq. (24), we are now in a position to draw several conclusions:

1. The EFT loop integral in Eq. (18) can be obtained from the full theory loop integral in Eq. (21) by expanding the integrand in powers of $1/M$. At a pictorial level, this is encoded by the fact that the Feynman diagram in the EFT can be obtained from the one in the full theory by contracting the scalar propagator (dashed line) to a point.
2. Despite this connection, the final results in Eqs. (19) and (24) are different. This means that loop integration and series expansion are not interchangeable operations. Notice in particular that the EFT result misses all the terms that are non-analytic around $M = \infty$. This makes sense, since the very expansion in inverse powers of M relies on the assumption of analyticity.
3. The UV divergent pieces, proportional to $1/\varepsilon$, are also different in the EFT and the full theory. This implies that the mass parameters and fermion fields in the two theories have different anomalous dimensions, and thus should not be confused with each other.
4. On the contrary, the terms that are non-analytic around $m = 0$ are identical. This must always be the case, and in practice can be used as a consistency check. As we will discuss in the next section, this cancellation is crucial for EFTs to be able to separate IR and UV scales and tame large logarithms.
5. Last but not least, loop corrections in the EFT are automatically suppressed by the small ratio m^2/M^2 —see *e.g.* Eq. (19)—unlike in the full theory. This ensures that only a finite number of loop diagrams need to be calculated at any given order in E/M .

From now on, we'll carry out the renormalization procedure both in the EFT and the full theory using the \overline{MS} scheme. From a practical viewpoint this is tantamount to dropping

all factors of $1/\varepsilon$ from the final results. The discrepancy between the renormalized 1-loop diagrams in both theories can then be eliminated by introducing an additional tree-level diagram in the EFT equal to

$$\begin{aligned} \text{---} \bullet \text{---} \bigcirc \text{---} \bullet \text{---} &= \left[\begin{array}{c} \text{1-loop} \\ \text{full theory} \end{array} \right]_{\overline{MS}} - \left[\begin{array}{c} \text{1-loop} \\ \text{EFT} \end{array} \right]_{\overline{MS}} \\ &= \frac{iy^2}{16\pi^2} \left\{ -\not{p} \left[\frac{1}{4} - \frac{m^2}{M^2} - \log \left(\frac{M^2}{\mu^2} \right) \right] - m \left(1 + \frac{m^2}{M^2} \right) \left[1 - \log \left(\frac{M^2}{\mu^2} \right) \right] + \frac{m}{M^2} p^2 \right\}. \end{aligned} \quad (25)$$

This diagram arises from the 1-loop correction to the effective action, which therefore must have the form

$$\begin{aligned} S_{\text{eff}}^{(1)} = \int d^4x \frac{y^2}{16\pi^2} \left\{ \left[\frac{1}{4} - \frac{m^2}{M^2} - \log \left(\frac{M^2}{\mu^2} \right) \right] \bar{\Psi} i \gamma^\mu \partial_\mu \Psi \right. \\ \left. - m \left(1 + \frac{m^2}{M^2} \right) \left[1 - \log \left(\frac{M^2}{\mu^2} \right) \right] \bar{\Psi} \Psi - \frac{m}{M^2} \bar{\Psi} \square \Psi + \dots \right\}, \end{aligned} \quad (26)$$

where the dots stand for terms that are cubic and higher in Ψ . Notice that the first two terms can be combined with similar terms in the tree-level effective action (12). The last term instead is new: it did not show up at tree-level, but it is generated at loop level in accordance with the principle that all terms allowed by symmetries belong in the effective action.

One last comment: we have chosen to match off-shell—*i.e.*, for arbitrary values of p^2 —but we could have just as well worked on-shell—*i.e.* momenta such that $p^2 = -m^2$. This would have changed the last term in parentheses in Eq. (25) to $-m^3/M^2$. Consequently, the last term in the 1-loop effective action (26) would have been replaced by an additional contribution proportional to $\bar{\Psi}\Psi$. This new effective action is physically equivalent to the one we have derived. It can be obtained from ours via a (perturbative) field redefinition [3]—an operation that doesn't change the S -matrix [4]. This field redefinition is tantamount to using the lowest order equations of motion to remove higher derivative terms. Incidentally, this shows that some operators are redundant, in the sense that their effect can be captured by modifying the Wilson coefficients of other operators. Matching on-shell automatically eliminates redundant operators.

Exercise 2.5: Derive the field redefinition that connects the effective actions obtained by matching off- and on-shell.

To summarize, we've extracted the quadratic part of $S_{\text{eff}}^{(1)}$ by comparing 1-loop corrections to 2-point amplitudes in the EFT and the full theory. Similarly, one could compare higher-loop and higher-point amplitudes in the two theories to calculate respectively higher-loop and nonlinear corrections to the effective action.

2.4 Summing large logarithms

So far we have argued that EFTs are convenient to use because they focus on the degrees of freedom that are relevant for the observables one is interested in. Nevertheless, in the particular case of our toy example this meant going from a Lagrangian with just a handful of terms to an effective Lagrangian that in principle includes an infinite number of terms. Although we have argued that only a finite number of terms will contribute to any calculation with a finite precision, some readers might still feel that the price to pay is just too high. To cheer up those readers, we will now discuss another reason why EFTs are the smart way of organizing calculations in theories with multiple scales.

The introduction of a renormalization scale μ is an inevitable byproduct of the renormalization process. It is a priori an arbitrary scale, and therefore physical predictions cannot depend on it. To ensure this, the renormalized couplings must depend on μ in a way that is captured by the renormalization group (RG) equations.

In perturbative calculations, it is helpful to choose μ in such a way as to avoid the appearance of large logarithms in the final result. Consider for instance the full theory 1-loop correction in Eq. (24). The part proportional to \not{p} contains a term of the form $\frac{y^2}{16\pi^2} \log(M/\mu)$. Similarly, the n -th loop correction will introduce terms of order $\left[\frac{y^2}{16\pi^2} \log(M/\mu)\right]^n$, and therefore a naive application of perturbation theory will converge more slowly if $\log(M/\mu)$ is large. This can be avoided by choosing $\mu \sim M$.³ In theories with a large hierarchy of different scales, logarithms can also depend on ratios of UV and IR scales—see for instance the $\log(M^2/m^2)$ in the second line of Eq. (24)—and thus large logarithms could seem unavoidable. We will now argue that this problem can be easily circumvented using an EFT at low energies.

Consider again the effective action we calculated in the last section. Adding together the tree-level and 1-loop contributions, the quadratic part reads

$$S_{\text{eff}}^{(0)} + S_{\text{eff}}^{(1)} = \int d^4x \left[\mathcal{Z}(\mu) \bar{\Psi} i \gamma^\mu \partial_\mu \Psi - \mathcal{M}(\mu) \bar{\Psi} \Psi - \mathcal{N}(\mu) \bar{\Psi} \square \Psi + \dots \right], \quad (27)$$

where the dots stand once again for the non-linear part of the action, and we have introduced

$$\mathcal{Z}(\mu) = 1 + \frac{y^2}{16\pi^2} \left[\frac{1}{4} - \frac{m^2}{M^2} - \log \left(\frac{M^2}{\mu^2} \right) \right], \quad (28a)$$

$$\mathcal{M}(\mu) = m \left\{ 1 + \frac{y^2}{16\pi^2} \left(1 + \frac{m^2}{M^2} \right) \left[1 - \log \left(\frac{M^2}{\mu^2} \right) \right] \right\}, \quad (28b)$$

$$\mathcal{N}(\mu) = \frac{m}{M^2}. \quad (28c)$$

³This requires one to know the value of the couplings at the scale M . If these couplings are only known at a different scale, one can always extrapolate their values at $\mu \sim M$ by solving the RG equations.

We have argued above that this effective action exactly reproduces the 2-point function of the full theory at 1-loop up to $\mathcal{O}(E^2/M^2)$ included. And yet, something very interesting has happened.

On the one hand, the Wilson coefficients in Eqs. (28) only depend on $\log(M^2/\mu^2)$. Notice that this wouldn't have happened if the logarithms of the IR scale m hadn't dropped out in the matching—a remarkable fact we have already emphasized in the previous section. Therefore, large logarithms in the effective action can be avoided altogether by carrying out the matching procedure at $\mu = M$. In other words, we should calculate the values of the Wilson coefficients in the EFT at the UV scale M .

On the other hand, the EFT 1-loop correction in Eq. (19) only depends on $\log(m^2/\mu^2)$. Thus, large logarithms in the 2-point function can be avoided by calculating the 1-loop correction at the IR scale $\mu = m$. This of course requires knowing the values of the Wilson coefficients at the IR scale m . These can be obtained by solving the RG equations to run the Wilson coefficients from $\mu = M$ (where their value is obtained from matching) down to $\mu = m$.

To summarize, by switching from the full theory to an EFT at energies $E < M$ we are essentially decomposing large logarithms in the full theory in two parts:

$$\log\left(\frac{M^2}{m^2}\right) = \log\left(\frac{M^2}{\mu^2}\right) - \log\left(\frac{m^2}{\mu^2}\right) . \quad (29)$$

The first part gets absorbed in the Wilson coefficients of the EFT, while the second part is “resummed” by solving the RG equations. It is important to stress that it is the combination of *matching and running* that tames the large logarithms. RG running in the full theory wouldn't be sufficient to get the job done, because the RG group can only “resum” logarithms that depend on the renormalization scale μ .

2.5 A modern perspective on renormalizable theories

Historically, a great emphasis has been placed by the particle physics community on renormalizable quantum field theories. As a matter of fact, the toy model with we started with in Eq. (1) was renormalizable. This particular class of theories enjoys the special property that any UV divergence arising in a loop calculation can be absorbed into a redefinition of fields and a finite number of couplings. This was originally believed to be necessary in order for the theory to be predictive, so much so that renormalizability was often regarded as a fundamental requirement for particle physics models. By now, we understand that EFTs are equally predictive within their regime of applicability, despite the fact that they include an infinite number of “non-renormalizable” interactions. Following our discussion on matching at loop level, we are in a position to understand the special status of renormalizable theories.

The Wilson coefficients of our EFT depend on the UV scale M in two ways: (1) powers of the ratio m^2/M^2 , and (2) logarithms of the form of $\log(M^2/\mu^2)$. By evaluating the Wilson coefficients at the scale $\mu = M$, all the logarithms vanish, and we are left with the inverse powers of M . These suppress what would be regarded as “non-renormalizable” (or *irrelevant*) terms (see *e.g.* Eq. (28c)), and also provide corrections to the coefficients of “renormalizable” (or *relevant*) terms (see *e.g.* Eqs. (28a) and (28b)). Then, our EFT reduces to a (in our case, free) renormalizable theory when we take the formal limit $M \rightarrow \infty$ while keeping m and y fixed.

More in general, renormalizable theories arise in the limit where the UV scale is taken to be infinitely larger than all IR scales, and the UV physics decouples completely. The bare couplings of a renormalizable theory are then just a finite subset of Wilson coefficients evaluated at $M = \infty$. The divergences arising in loop calculations can now all be absorbed in a redefinition of the bare couplings and fields, because any other divergence would be suppressed by powers of $1/M \rightarrow 0$. (I am obviously being cavalier about the order of the limits $M \rightarrow \infty$ and $\varepsilon \rightarrow 0$ in order to get the main point across).

2.6 Naturalness

It is instructive to consider our toy model in Eq. (1) in a different region of parameter space, where the fermion is much heavier than the scalar, *i.e.* $m \gg M$. We can then follow the same logic we have developed so far in this section and integrate out the fermion to obtain an EFT for the scalar valid at energies $E \ll m$.

It is easy to see that the tree-level effective action $S_{\text{eff}}^{(0)}$ is equal to the action in Eq. (1) with $\Psi = 0$. To calculate the quadratic part of the 1-loop effective action $S_{\text{eff}}^{(1)}$, we compare the 1-loop correction to the scalar propagator obtained using the tree-level EFT and the full theory. In dimensional regularization, the EFT yields

$$\begin{array}{c} \text{---} \text{---} \text{---} \text{---} \end{array} = -\frac{i\lambda\tilde{\mu}^\varepsilon}{2} \int \frac{d^{4-\varepsilon}k}{(2\pi)^{4-\varepsilon}} \frac{-i}{k^2 + M^2 - i\epsilon} = \frac{i\lambda}{32\pi^2} M^2 \left[\frac{2}{\varepsilon} + 1 + \log\left(\frac{\mu^2}{M^2}\right) + \mathcal{O}(\varepsilon) \right]. \quad (30)$$

This same diagram appears also in the full theory, supplemented by a second one with a fermion running in the loop:

$$\begin{array}{c} \text{---} \text{---} \end{array} = -(-y)^2 \int \frac{d^4k}{(2\pi)^4} \text{Tr} \left[\frac{-i(-\not{k} - \not{p} + m)}{(k+p)^2 + m^2 - i\epsilon} \gamma^5 \frac{-i(-\not{k} + m)}{k^2 + m^2 - i\epsilon} \gamma^5 \right]. \quad (31)$$

for $\mu \sim \mathcal{O}(m)$. Unlike in the previous case, we now have a lower bound on $\mathcal{M}^2(m)$ for any fixed value of y . One could circumvent this bound by arranging a delicate *fine-tuning* of the tree-level and loop contributions at a particular scale μ . However, such fine-tuning would not survive if we changed the renormalization scale by a relative factor of $\mathcal{O}(1)$, and is thus regarded as an “unnatural” possibility. The “natural” expectation, instead, is that the scalar mass be in the neighborhood of the strong coupling scale m , within a window of size determined by the coupling y . When applied to realistic theories of Nature, this expectation is based on the sentiment that the properties of our universe should be fairly generic.

We should stress that the sensitivity of the scalar mass to the UV scale m is not an artifact of switching to an EFT at low energies. A similar result can be inferred using exclusively the full theory. In the \overline{MS} renormalization scheme, the physical mass of the scalar doesn’t coincide with the renormalized mass parameter M , but instead is defined as the location of the pole of the scalar propagator. Denoting with $i\Pi(k^2)$ the sum of all 1PI loop corrections, the physical mass is determined by the relation $M_{\text{ph}}^2 = M^2 - \Pi(-M_{\text{ph}}^2)$. At 1-loop, this yields

$$M_{\text{ph}}^2 = M^2 \left\{ 1 - \frac{\lambda}{32\pi^2} \left[1 + \log \left(\frac{\mu^2}{M^2} \right) \right] - \frac{y^2}{8\pi^2} \log \left(\frac{\mu^2}{M^2} \right) \right\} + \frac{y^2}{4\pi^2} m^2 \left[1 + \log \left(\frac{\mu^2}{m^2} \right) \right], \quad (37)$$

from which we conclude that a fine-tuning would be required for M_{ph}^2 to be much smaller than $\frac{y^2}{4\pi^2} m^2 \simeq \frac{y^2}{4\pi^2} m_{\text{ph}}^2$.

The issue of the *UV sensitivity* of a scalar mass is often discussed by regulating loops in the full theory with a hard cutoff Λ , and noticing that the scalar mass receives a correction that diverges quadratically with the cutoff. In the case of fermions, instead, the mass only depends logarithmically on Λ . Therefore, as the argument goes, scalar masses are more sensitive to UV physics. There is of course a kernel of truth behind this statement, but as a student I always found it confusing, since a cutoff regularization is just a trick we use to make sense of intermediate steps when calculating loops. Eventually, any trace of Λ gets absorbed in the renormalized fields and couplings, so why would it matter if the divergence is quadratic or logarithmic? The point is that imposing a hard cutoff in momentum space can be thought of as a (admittedly, highly unphysical) modification of the UV physics that kicks in at the scale Λ . From this perspective, the quadratic divergence can once again be interpreted as a sign that the natural size of a scalar mass is set by the “UV physics”. In dimensional regularization we only deal with physical mass scales, and this makes it clear that what ultimately matters is not how loops diverge, but rather how the physical mass of a scalar depends on physical UV scales.

The requirement that there be no unexplained fine-tuning can also be expressed in a different way. Clearly, for the bound in Eq. (36) to be satisfied it is sufficient to have

$$M^2 \gtrsim \frac{y^2}{4\pi^2} m^2. \quad (38)$$

Based on Eq. (35b), this requirement means that the 1-loop correction to the mass parameter in the EFT should not be larger than its tree-level value at the UV scale $\mu = m$. This observation allows us to introduce a slightly more precise notion of naturalness: the Wilson coefficients of an EFT are considered natural in a technical sense, or technically natural for short, if their magnitude is not altered significantly by loop corrections. When phrased this way, the requirement of *technical naturalness* is tantamount to demanding that the Wilson coefficients admit a well defined perturbative expansion.

We have seen that the toy model in Eq. (1) would be technically natural if $m^2 \ll M^2$, but not if $M^2 \ll \frac{y^2}{4\pi^2} m^2$. It all boils down to the fact that loop corrections to the fermion mass m are proportional to m itself, whereas loop corrections to the scalar mass squared M^2 are proportional to m^2 rather than M^2 . What is the origin of this difference? It turns out that our toy model acquires an additional discrete symmetry when $m = 0$, namely

$$\Psi \rightarrow \gamma^5 \Psi, \quad \phi \rightarrow -\phi. \quad (39)$$

Barring anomalies, which are not an issue in this model, loop corrections are supposed to preserve the symmetries of the theory. For this reason, corrections to the mass parameter can be non-zero only if m is already non-zero at tree-level. In other words, these corrections must vanish when $m \rightarrow 0$ to preserve the symmetry (39) in this limit. More broadly, it is natural for any dimensionless parameter (such as, in our case, m/M) to be much smaller than one if setting it to zero enhances the symmetry of the model. In this case, we say that the smallness of such parameter is “protected” by the additional symmetry. This particular form of the naturalness criterion is known as *’t Hooft naturalness* [5]. It is a slightly more stringent requirement than our notion of technical naturalness, since there exist theories in which certain Wilson coefficients do not receive *any* loop corrections without being protected by any symmetry.⁴ A particularly striking example of this is provided by Galileon theories [6].

Over the last few decades, naturalness has become one of the guiding principles in our search for physics beyond the Standard Model (SM). Given the important role that this concept has played in shaping the research agenda of the particle physics community, I would be remiss if I didn’t address briefly its broader implications. Before I do so, however, a disclaimer is in order: the following discussion reflects my own viewpoint on the subject. You are encouraged to develop your own perspective by consulting a variety of sources. To get you started, I have provided a few references in the following section.

I think it is important to stress that the naturalness criterion in Eq. (38) is specific to the particular toy model we have been considering: it is a constraint on the parameters of the

⁴I find the distinction between ’t Hooft and technical naturalness a useful one, albeit it is not common in the literature. The reader should be aware that what I am calling ’t Hooft naturalness is often referred to as technical naturalness.

action in Eq. (1). If we were to replace the heavy fermion with a different spectrum of heavy particles—*i.e.* choose a different UV completion for the scalar EFT—this constraint would take a different form. Thus, in order to establish whether the scalar mass is natural, one needs to (i) know the correct UV completion of the scalar EFT, and (ii) measure experimentally its parameters, *e.g.* by studying scattering processes involving both the light scalar and the heavy particles. This suggests that naturalness should be regarded as a constraint that a low-energy EFT imposes on its possible UV completions.⁵

The two most famous “naturalness problems” in particle physics are associated with the two UV sensitive parameters of the SM coupled to General Relativity (GR): the Higgs mass parameter, and the cosmological constant. The UV sensitivity of the Higgs mass is known as the hierarchy problem, and it is essentially the same phenomenon we encountered in our toy scalar EFT. The hierarchy problem certainly makes it more challenging to write down models of physics above the weak scale that are not finely tuned. Whether or not this turns out to be an actual problem—by which I mean, an aspect of Nature we don’t understand—will depend on what the correct UV completion of the SM happens to be.

For now, we should keep an open mind and search for new physics above the weak scale, remaining aware of the fact that none of the open problems in particle physics (*e.g.* baryogenesis, dark matter, neutrino masses, strong CP problem, ...) necessarily requires a solution in the UV (even though some of such solutions are particularly compelling). The strongest argument in favor of new UV physics is in my opinion the fact that GR is an effective theory that ceases to be valid at the Planck scale. Based on our experience with EFTs, we would expect new degrees of freedom to show up at or below this scale and give rise to a more complete description of gravitational interactions. It is also reasonable to expect this new physics to provide large quantum corrections to the Higgs mass parameter. Unfortunately, our present understanding of gravity at the Planck scale doesn’t allow us to perform a detailed matching calculation and explicitly verify this expectation.

The cosmological constant problem poses a much more serious challenge to our understanding of naturalness and its role in particle physics, for at least two reasons. First and foremost, unlike the hierarchy problem, it doesn’t hinge upon the existence of new UV physics. Integrating out any SM particle with mass m gives rise to a 1-loop correction to the cosmological constant of the form $\Delta\Lambda = \mathcal{O}(1)\frac{m^4}{16\pi^2 M_{\text{P}}^2}$. Even the electron would yield a contribution of $\mathcal{O}(10^{-52})\text{ GeV}^2$ which is much larger than the observed value of $\mathcal{O}(10^{-84})\text{ GeV}^2$. Second, the only known symmetry that can protect the cosmological constant from quantum corrections

⁵This should be contrasted with constraints that one can impose on low-energy EFTs by making some assumptions about the UV completion, *e.g.* connection between spin and statistics [4] and positivity constraints on the Wilson coefficients [7] in theories with a Lorentz-invariant UV completion, constraints on symmetries and particle content of EFTs coming from string theory [8], etc ...

is supersymmetry. Unfortunately there is so far no experimental evidence for supersymmetry, which means that, at best, it must be broken at a scale M_{susy} that is higher than the weak scale. In this scenario, the cosmological constant would receive corrections that are generically of order $\Delta\Lambda = \mathcal{O}(1) \frac{M_{\text{susy}}^4}{16\pi^2 M_{\text{p}}^2} \gtrsim \mathcal{O}(10^{-34}) \text{ GeV}^2$, which is again much larger than the observed value. This state of affair has led many to believe that a solution to the cosmological constant problem might point towards a new organizing principle in the fundamental laws of nature. Interesting possibilities include anthropic considerations [9] and cosmological relaxation mechanisms [10, 11].

Exercise 2.6: By convention, the cosmological constant enters the effective action as $S \supset \int \sqrt{-g} (-M_{\text{p}}^2 \Lambda)$. Calculate the 1-loop matching correction to the cosmological constant generated when integrating out the Higgs particle. Work at lowest order in the SM couplings. Compare your result with the vacuum expectation value of the stress energy tensor of the free Higgs field.

2.7 Additional resources

There are many excellent reviews on EFT methods, *e.g.* [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. In particular, a nice example of a matching calculation at 1-loop implemented at the level of the generating functional can be found in [17]. For a very pedagogical discussion of matching and running, see [22]. For interesting discussions on naturalness, see [23, 24]. For a provocative take on the hierarchy problem, see instead the appendix of [21]. The classic reference about the cosmological constant problem is [25].

3 EFTs for Nambu-Goldstone modes

We have argued that symmetries are one of the defining ingredients of any EFT: any term that is invariant should be included in the effective action. In practice, this is fairly easy to achieve when dealing with fields that transform linearly under the relevant symmetries, *i.e.* according to transformation rules of the form $\varphi^i \rightarrow M^i_j \varphi^j$ for some constant matrix M .⁶ Then it is just a matter of contracting indices in an invariant way. Spontaneously broken symmetries, however, are realized non-linearly on the ensuing Goldstone modes. In this case, writing down all possible invariant operators is not always straightforward. In the first part of this chapter, we will introduce non-linearly realized symmetries in the context of Chiral Perturbation Theory (CPT)—the EFT that describes the lightest QCD bound states. This EFT will also feature additional subtleties compared to the ones we discussed in the previous chapter. In the second part of this chapter, we will introduce the coset construction—a general technique to write down effective actions that are invariant under non-linearly realized symmetries. Along the way, we will also learn how to handle EFTs with approximate or anomalous symmetries.

3.1 Chiral perturbation theory

Imagine we are interested in studying QCD bound states that are lighter than proton and neutron. We might as well integrate out all SM particles with mass at or above the GeV scale and work the resulting effective theory. The remaining quarks— u , d and s —are described at low energies by the effective action

$$S = \int d^4x \left\{ \sum_{j=u,d,s} \bar{\Psi}_j (i\gamma^\mu D_\mu - m_j) \Psi_j - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} \text{Tr} G_{\mu\nu} G^{\mu\nu} + \dots \right\}, \quad (40)$$

with $D_\mu \Psi_j = (\partial_\mu - iq_j A_\mu - ig G_\mu) \Psi_j$, A_μ and G_μ the photon and gluon fields, and $F_{\mu\nu}$ and $G_{\mu\nu}$ their respective field strengths. Color indices won't play any role in what follows and will be suppressed throughout. The dots in the Lagrangian above stand for all sorts of “non-renormalizable” terms, which are suppressed at energies below the GeV. These include in particular interactions between quarks and light leptons.

The mass of all three remaining quarks are smaller than the only other scale in the problem, *i.e.* the QCD scale $\Lambda_{\text{QCD}} \simeq 300$ MeV associated with confinement. It seems therefore a reasonable approximation to study bound states by neglecting quark masses at a

⁶One can of course always complicate the transformation rules at will by performing an arbitrarily involved field redefinition. Steven Weinberg's “Third Law of Progress in Theoretical Physics” addresses this possibility: “*You may use any degrees of freedom you like to describe a physical system, but if you use the wrong ones, you'll be sorry*” [26].

first pass. Admittedly, this approximation works better for the u and d quarks, whose masses are of $\mathcal{O}(1)$ MeV, than it does for the s quark, since $m_s \simeq 100$ MeV. Still, with this caveat in mind, let's press on and double down by neglecting also the electromagnetic coupling. We will reintroduce the effect of finite masses and electromagnetism in the next section. In this limit, it is convenient to break up the quark fields into their left-handed and right-handed components, defined by

$$\Psi_L^j = \frac{1}{2}(1 - \gamma^5)\Psi^j, \quad \Psi_R^j = \frac{1}{2}(1 + \gamma^5)\Psi^j, \quad (41)$$

so that the action above reduces to

$$S \simeq \int d^4x \sum_{j=u,d,s} \left[\bar{\Psi}_L^j i\gamma^\mu (\partial_\mu - ig G_\mu) \Psi_L^j + \bar{\Psi}_R^j i\gamma^\mu (\partial_\mu - ig G_\mu) \Psi_R^j \right] - \frac{1}{2} \text{Tr} G_{\mu\nu} G^{\mu\nu} + \dots \quad (42)$$

This action is invariant under several global symmetry. First, it is invariant under conformal transformations. This follows from the fact that there is no mass scale in Eq. (42). Nevertheless, a preferred mass scale is generated dynamically, because quantum effects make the QCD coupling run with energy and interactions become strong at the scale Λ_{QCD} . In other words, conformal invariance is anomalous: it is broken by quantum effects and therefore is not a symmetry of the full theory.

The action (42) is also invariant under separate $U(3)$ transformations of the left-handed and right-handed quarks:

$$(\Psi_L^i)' = L^i_j \Psi_L^j, \quad (\Psi_R^i)' = R^i_j \Psi_R^j. \quad (43)$$

It is helpful to decompose this $U(3)_L \times U(3)_R$ symmetry group into the product $SU(3)_L \times SU(3)_R \times U(1)_V \times U(1)_A$, where $U(1)_V$ corresponds to global rephasings of all the quark fields, while $U(1)_A$ transforms left-handed and right-handed quarks with opposite phases, *i.e.*

$$\Psi_L' = e^{i\alpha} \Psi_L, \quad \Psi_R' = e^{-i\alpha} \Psi_R. \quad (44)$$

Each of these subgroups enjoys a different status. The axial symmetry $U(1)_A$ is anomalous, and therefore is not an actual symmetry of the quantum theory. More on this in Sec. 3.7. Moreover, strong interactions cause the operator $\bar{\Psi}_R^i \Psi_L^j$ to develop a non-trivial vacuum expectation value (vev) of the form

$$\langle \bar{\Psi}_R^i \Psi_L^j \rangle = v^3 \delta^{ij}, \quad (45)$$

with $v \sim \Lambda_{QCD}$. This vev breaks spontaneously the chiral symmetry $SU(3)_L \times SU(3)_R$ down to its diagonal subgroup $SU(3)_V$, such that left- and right-handed fermions transform in lockstep.⁷ (This vev would also break spontaneously the $U(1)_A$ and conformal symmetries,

⁷There is a very general argument based on anomalies that implies that $SU(3)_L \times SU(3)_R$ must be spontaneously broken down to some unspecified subgroup [5].

if it wasn't for the fact that they are not actual symmetries to begin with.) As a result, Goldstone's theorem dictates that the spectrum of this theory must contain (at least) eight massless bound states, *i.e.* one Nambu-Goldstone (NG) mode for each broken symmetry generator. Finally, the vev in Eq. (45) preserves the baryon symmetry $U(1)_V$.⁸ This means that bound states of quarks and gluons must have a definite charge under $U(1)_V$.

In what follows, we will derive an EFT for the NG modes associated with the spontaneous breaking of chiral symmetry. This EFT is known as Chiral Perturbation Theory, and its expansion parameter will once again be a ratio of the form E/Λ , with Λ some UV scale to be determined. Compared to the toy model we considered in the previous chapter, this EFT presents two novel features. First, the action (42) doesn't contain any mass scale, as we already pointed out. For this reason, it is not immediately obvious what the strong coupling scale Λ should be. Notice in particular that it doesn't have to coincide with the symmetry breaking scale v —although in practice it won't be too far from it. Second, the NG fields do not appear directly in the UV action (42). Therefore, we cannot derive our EFT simply by integrating out some heavy fields as we did in the previous chapter: strong interactions prevent us from carrying out an explicit matching calculation. We will therefore need to be guided by particle content and symmetries alone, and treat the EFT coefficients as free parameters to be fixed by experiments.

In order to figure out how symmetries act on the NG fields, it is convenient to think of NG excitations as broken transformations of the vev that are “mildly” modulated in space and time. A constant broken transformation would, by definition, change the vev (45) to a different but physically equivalent vev; by contrast a NG excitation smoothly changes the vacuum at each point in spacetime. This is often expressed by saying that NG modes interpolate between different vacua. Based on this, we act on the vev (45) with a local $SU(3)_L \times SU(3)_R$ transformation such that $L(x) = R^\dagger(x) = e^{i\pi^a(x)Q_a}$, with $\pi^a(x)$ the eight NG fields, and the Q_a 's the generators of $SU(3)$. The result is

$$\langle \bar{\Psi}_R^i \Psi_L^j \rangle \rightarrow v^3 U^{ji}, \quad \text{with} \quad U(x) = L(x) \cdot R^\dagger(x) = e^{2i\pi^a(x)Q_a}. \quad (46)$$

Hence, NG excitations correspond to those particular local deformations of the 3×3 complex matrix $\langle \bar{\Psi}_R^i \Psi_L^j \rangle$ that can be described by an $SU(3)$ transformation $U(x)$. Notice that this transformation preserves the overall “size” of the matrix, *i.e.* its determinant. Now, a generic $SU(3)_L \times SU(3)_R$ transformation is such that $L(x) \rightarrow \tilde{L} \cdot L(x)$ and $R(x) \rightarrow \tilde{R} \cdot R(x)$, and

⁸To be precise, baryon symmetry is also anomalous when weak interactions are taken into account. It is only the difference between baryon and lepton number that is exactly conserved. However, the lepton symmetry acts trivially on quarks and gluons, and therefore the action (40) appears invariant under baryon symmetry alone. This means that baryon number is conserved by strong interactions.

therefore

$$U \rightarrow U' = \tilde{L} \cdot U \cdot \tilde{R}^\dagger. \quad (47)$$

Similarly, it is fairly easy to see that the π^a 's must carry zero baryonic charge, because the generators of $SU(3)_L \times SU(3)_R$ commute with the generator of $U(1)_V$.

It is not difficult to write down an effective action for U that is invariant under chiral transformations of the form (47), because these transformations are linear. We will need to use both U and U^\dagger , as well as some derivatives, since any product of the form $U^\dagger U$ is trivially invariant because U is unitary. A moment of reflection leads to the conclusion that there is only one invariant term we can write down with two derivatives, and therefore the effective action for the our NG bosons takes the form

$$S = \int d^4x \left\{ -\frac{f^2}{4} \text{Tr} (\partial_\mu U^\dagger \partial^\mu U) + \dots \right\}, \quad (48)$$

where the factor of $1/4$ has been introduced for later convenience, and the dots stand for all terms with more than two derivatives. This EFT displays many remarkable features:

- The effective action depends on a single parameter f at lowest order in the derivative expansion, despite the fact that it contains an infinite number of terms when expanded in powers of the NG fields. Up to quartic order, it reads

$$S = \int d^4x f^2 \left\{ -\frac{1}{2} \partial_\mu \pi_a \partial^\mu \pi^a + \frac{1}{6} f_{abc} f_{de}^c \pi^b \pi^e \partial_\mu \pi^a \partial^\mu \pi^d + \dots \right\}, \quad (49)$$

where f_{abc} are the structure constants of $SU(3)$, which can be found for instance in [27]. To derive this result, we used the fact that the normalization of the $SU(3)$ generators in the fundamental representation is such that $\text{Tr}(Q_a Q_b) = \frac{1}{2} \delta_{ab}$.

- Each term in Eq. (49) is manifestly invariant under linear $SU(3)$ transformations of the NG fields: this is the unbroken symmetry, $SU(3)_V$.
- The relative coefficients of the quadratic, quartic, and higher order terms are completely nailed by symmetry. This is a consequence of the spontaneously broken part of $SU(3)_L \times SU(3)_R$, which is realized non-linearly on the π 's. Consider for instance a broken transformation such that $\tilde{L} = e^{2i\epsilon^a Q_a}$ and $\tilde{R} = \mathbf{1}$. According to Eq. (47), the NG fields must change in such a way that

$$e^{2i\pi'_a Q^a} = e^{2i\epsilon_b Q^b} e^{2i\pi_a Q^a} = e^{2i(\pi_a + \epsilon_a - f_{abc} \epsilon^b \pi^c + \dots) Q^a}, \quad (50)$$

where in the last step we used the Baker-Campbell-Hausdorff formula to combine the exponentials, and the dots stand for an infinite number of terms that are higher order in π , ϵ , or both. The relative tuning of quadratic, quartic, and higher order terms is necessary for the effective action (49) to be invariant under these very complicated transformation rules.

- Incidentally, this shows why it's much easier to work with U , which transforms linearly as shown in Eq. (47), rather than directly with the π 's. It would have been much more difficult to come up with the action (49) starting from the transformation rules for the NG fields.
- Since $U_{ij}(x)$ was a $SU(3)$ matrix, the action (49) was obtained by calculating traces of products of generators in the fundamental representation. Had we worked with a different representation (*e.g.*, the adjoint one), the relative coefficients between quadratic, quartic, etc... terms would have remained the same. The only difference would have been the overall normalization, which can always be reabsorbed in the coupling f .⁹ Thus, the representation of the operator breaking the symmetry only affects the values of the Wilson coefficients. The allowed non-linear structures, instead, are completely determined by the symmetry breaking pattern, *i.e.* by which symmetry group is broken down to which subgroup.

By convention, the canonical normalization for a scalar field is such that the coefficient in front of the kinetic term is equal to $-1/2$ (with a mostly plus metric signature). Therefore, the canonically normalized NG fields are $\hat{\pi}^a = f\pi^a$. When expressed in terms of the $\hat{\pi}$'s, our effective action has the schematic form

$$S = \int \left\{ \sum_{n=0}^{\infty} c_n (\partial \hat{\pi})^2 \left(\frac{\hat{\pi}}{f} \right)^{2n} + \dots \right\}, \quad (51)$$

with c_n some coefficients determined by symmetries. Thus, additional powers of $\hat{\pi}$ in the Lagrangian are suppressed by the scale f . It is worth emphasizing that this is *not* equal to the strong coupling scale Λ of the EFT, which is instead the scale that suppresses additional derivatives in the Lagrangian, and at which perturbation theory breaks down. In other words, observable quantities such as n -point amplitudes will admit an expansion in powers of E/Λ , not E/f , and the effective action takes the schematic form

$$S = \int d^4x \Lambda^2 f^2 \mathcal{L}(U, U^\dagger, \partial_\mu/\Lambda). \quad (52)$$

We can derive an upper bound on the strong coupling scale Λ by invoking the naturalness criterion we discussed in the previous chapter. To this end, we will compare the size of tree-level and their 1-loop contributions to the 4-point function. At tree-level, the 4-point function scales like E^2/f^2 , as one can easily see from the action in Eq. (49). The 1-loop correction built out of two tree-level quartic vertices can be easily estimated in dimensional regularization: each vertex comes with a $1/f^2$ factor; each derivative contributes an external momentum of $\mathcal{O}(E)$, since there are no other scales around; finally, we have the ubiquitous “loop factor”

⁹This is essentially because $\text{Tr}(T_a T_b) = C \delta_{ab}$, with C some coefficient that depends on the representation.

$1/(4\pi)^2$. Putting it all together, we find schematically

$$\text{Diagram} \sim \frac{1}{(4\pi)^2} \times \left(\frac{1}{f^2}\right)^2 \times E^4 = \frac{E^2}{f^2} \times \frac{E^2}{(4\pi f)^2} . \quad (53)$$

This 1-loop diagram is suppressed compared the tree-level result by a factor of $E^2/(4\pi f)^2$. This suggests that perturbation theory will stop working at a scale $\Lambda \sim 4\pi f$, which is an order of magnitude larger than f . Notice that 4-derivative operators in the effective action will yield a tree-level contribution to the 4-point function that is also proportional to E^4 . Therefore, it is possible that perturbation theory breaks down at energies smaller than $4\pi f$ if these operators have Wilson coefficients larger than $1/(4\pi)^2$. On the contrary, having a strong coupling scale much larger than $4\pi f$ would require a delicate fine-tuning of the Wilson coefficients of higher-derivative terms, and is therefore considered an unnatural possibility. The statement that the strong coupling scale of this EFT should not be larger than $4\pi f$ is usually regarded as indication that new physics should appear at or below this scale.

Incidentally, the separation between f and the strong coupling scale is needed to ensure that chiral perturbation theory is actually useful. A comparison with experiments shows that $f \simeq 93 \text{ MeV}$,¹⁰ which is smaller than the mass of the lightest meson: if f was the strong coupling scale, we would be in trouble!

3.2 Approximate symmetries and spurions

Particle physics experiments have of course never discovered eight Goldstone bosons. This is because chiral symmetry is only an *approximate symmetry* of the action (40): it is explicitly broken by the quark mass terms and the electromagnetic interactions.¹¹ Approximate symmetries introduce additional expansion parameters in the EFT. Depending on the sources of explicit symmetry breaking, these can be either dimensionless couplings (*e.g.* quark charges), or ratios between the explicit symmetry breaking scales (*e.g.* quark masses) and the strong coupling scale. These new expansion parameters can be taken into account in a systematic way by resorting to a nifty trick: pretending that the symmetry breaking parameters are fictitious fields, usually referred to as *spurions*.

Let's see how this strategy works in practice in Chiral Perturbation Theory, and consider first the quark masses. We start by adding to the quark action in Eq. (42) the following term

$$\Delta S = \int d^4x \left\{ -\bar{\Psi}_R^j M_{jk} \Psi_L^k + \text{h.c.} \right\} , \quad (54)$$

¹⁰Notice that f is sometimes used in the literature to denote $\sqrt{2} \times 93 \text{ MeV} \simeq 130 \text{ MeV}$.

¹¹Weak interactions also break explicitly chiral symmetry but, in the effective action (40), do so via higher-dimensional operators whose effects are suppressed at low energies.

where M is the spurion field. This term is invariant under $SU(3)_L \times SU(3)_R$ (and, of course, under $U(1)_V$) provided the spurion transforms as follows:

$$M \rightarrow \tilde{R} \cdot M \cdot \tilde{L}^\dagger. \quad (55)$$

The term we just introduced reduces precisely to the mass terms we dropped in going from the action in Eq. (40) to the one in Eq. (42) when the spurion takes the following value

$$M_{jk} = \text{diag}(m_u, m_d, m_s). \quad (56)$$

Our strategy will consist of extending the effective action for the NG fields in Eq. (52) to include all possible couplings with the spurions that respect chiral symmetry. Chiral symmetry is a good approximation to the extent that the quark masses are small compared to the strong coupling scale. Therefore, we will treat M/Λ as an additional expansion parameter which controls the size of explicit symmetry breaking, so that our effective action now reads,

$$S = \int d^4x \Lambda^2 f^2 \mathcal{L}(U, U^\dagger, \partial_\mu/\Lambda, M/\Lambda), \quad (57)$$

where the Lagrangian contains in principle all possible operators that are invariant under chiral symmetry. After replacing the spurion M with its value in Eq. (56), the action (57) captures all possible ways in which non-zero quark masses modify the dynamics of the NG fields.

At lowest order in M/Λ , the spurion couples to the NG fields via the following operator in the effective Lagrangian:

$$\Delta S = \int d^4x f^2 \Lambda^2 \left\{ \frac{y}{2} \text{Tr} \left(\frac{M}{\Lambda} U \right) + \text{h.c.} \right\}, \quad (58)$$

with y some coupling assumed to be of $\mathcal{O}(1)$ on naturalness grounds. Expanding this term up to quadratic order in the π 's, we find that our NG fields are now endowed with mass terms. More precisely, it turns out that the eight NG modes arrange themselves into four pairs, three of which have masses

$$(\pi^+, \pi^-) : \quad m^2 = y \Lambda (m_u + m_d), \quad (59a)$$

$$(K^+, K^-) : \quad m^2 = y \Lambda (m_u + m_s), \quad (59b)$$

$$(K^0, \bar{K}^0) : \quad m^2 = y \Lambda (m_d + m_s), \quad (59c)$$

while the fourth one has a non-diagonal mass matrix of the form

$$(\pi^0, \eta) : \quad m^2 = y \Lambda \begin{pmatrix} m_u + m_d & m_u - m_d \\ m_u - m_d & \frac{1}{3}(m_u + m_d + 4m_s) \end{pmatrix}. \quad (60)$$

I have added in parentheses the mesons with which these mass eigenstates should be identified.

Exercise 3.1: Derive the results in Eqs. (59) and (60). **Hint:** the generators of $SU(3)$ in the fundamental representation are given by $Q_a = \frac{1}{2}\lambda_a$, where the λ 's are the eight Gell Mann matrices. These matrices satisfy the following properties:

$$\text{Tr}(\lambda_a) = 0, \quad \text{Tr}(\lambda_a \lambda_b) = 2\delta_{ab}, \quad \{\lambda_a, \lambda_b\} = \frac{4}{3}\delta_{ab}\mathbf{1} + 2d_{abc}\lambda^c. \quad (61)$$

Explicit expressions for the Gell Mann matrices and the totally symmetric coefficients d_{abc} can be found for instance in [27].

The results in Eqs. (59) and (60) are non-trivial predictions of Chiral Perturbation Theory: a priori, one could have imagined that each NG field would develop a different mass once chiral symmetry was broken!

We can further refine our discussion of explicit breaking due to quark masses by introducing an additional expansion parameter, namely $(m_d - m_u)/\Lambda$. At lowest order in this expansion, $m_d \simeq m_u$ and the spurion value in Eq. (56) preserves a $SU(2)$ subgroup of $SU(3)_V$: this is known as *isospin symmetry*. In this limit, η , (K^+, K^0) , (K^-, \bar{K}^0) , (π^0, π^+, π^-) , form one spin 0, two spin 1/2 (conjugate to each other), and one spin 1 representations of isospin, and their masses satisfy the *Gell-Mann-Okubo relation*:

$$m_\pi^2 + 3m_\eta^2 = 4m_K^2. \quad (62)$$

This is of course not end of the story, since the chiral symmetry is also broken explicitly by electromagnetic interactions. This breaking can be described once again by introducing some fictitious fields, although there are some differences compared to the case of quark masses we just considered, so it is worth discussing this in detail. Reintroducing electromagnetic interactions means adding the following term to the quark action in Eq. (42):

$$\Delta S = \int d^4x \left\{ eA_\mu \bar{\Psi}_L^j \gamma^\mu Q_{jk}^L \Psi_L^k + eA_\mu \bar{\Psi}_R^j \gamma^\mu Q_{jk}^R \Psi_R^k \right\}, \quad (63)$$

with

$$Q_{ij}^L = Q_{ij}^R = \text{diag} \left(\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3} \right). \quad (64)$$

This provides us with a new expansion parameter: the dimensionless electromagnetic coupling, e . We will now treat the symmetry generators Q^L and Q^R as spurion fields, and exploit the fact that ΔS remains invariant under chiral transformations if these spurions transform as

$$Q^L \rightarrow \tilde{L} \cdot Q^L \cdot \tilde{L}^\dagger, \quad Q^R \rightarrow \tilde{R} \cdot Q^R \cdot \tilde{R}^\dagger. \quad (65)$$

Following the same strategy we used for the quark masses, we then modify the effective action for chiral perturbation theory by including all possible couplings between the NG fields, the gauge field A_μ , and the spurions $Q^{L,R}$ that preserve gauge invariance and chiral symmetry.

Gauge invariance can be easily enforced by replacing ordinary derivative $\partial_\mu U$ with co-variant ones, defined by $D_\mu U = \partial_\mu U - ieA_\mu Q^L U + ieA_\mu Q^R U$. We can also write down interactions between NG fields and spurions that do not involve the gauge field. Notice that each spurion must appear with a factor of “ e ”, since the quark action is invariant under $e \rightarrow \lambda e$, $Q^{L,R} \rightarrow Q^{L,R}/\lambda$. At lowest order in e , there is only one operator that is allowed by the symmetries, namely

$$\Delta S = \int d^4x f^2 \Lambda^2 \left\{ \frac{e^2 y'}{2} \text{Tr} \left(Q^L U Q^R U^\dagger \right) + \text{h.c.} \right\}, \quad (66)$$

with y' another arbitrary coupling. When expanded in powers of π 's, this operator generates an infinite number of self-interactions that, unlike the ones in Eq. (49), do not involve any derivative. In particular, at quadratic order one finds that Eq. (66) yields the following corrections to the meson masses:

$$(\pi^+, \pi^-) : \quad m^2 \rightarrow y \Lambda (m_u + m_d) + y' e^2 f^2, \quad (67)$$

$$(K^+, K^-) : \quad m^2 \rightarrow y \Lambda (m_u + m_s) + y' e^2 f^2, \quad (68)$$

while the other masses are not affected. The fact that the fields π^\pm and K^\pm receive exactly the same correction at lowest order is another non-trivial result that goes under the name of *Dashen's theorem*. Neglecting isospin breaking, this correction is entirely responsible for the mass splitting between, say, π^0 and π^\pm .

Exercise 3.2: Derive the result in Eq. (67). **Hint:** think about which generators of $SU(3)_{L,R}$ commute with $Q^{L,R}$. You might want to use again the explicit form of the Gell-Mann matrices given in [27].

3.3 Naturalness and pseudo-Nambu-Goldstone modes

The discussion in the previous section provides a different viewpoint on the issues of naturalness of scalar masses and symmetry protection. Take for instance the mass of the π^\pm mesons, which using $\Lambda \simeq 4\pi f$ can be written as

$$m^2 = \left[y \frac{m_u + m_d}{\Lambda} + y' \frac{e^2}{(4\pi)^2} \right] \Lambda^2. \quad (69)$$

At face value, this doesn't look too different from the natural value of the scalar mass squared in the toy model of the previous chapter: it has the same form, namely (small parameters) \times (strong coupling scale)². There is however an important qualitative difference. In the toy model of the previous chapter, the small parameter was the coupling y^2 which determined the strength of interactions between the IR degrees of freedom (the scalar) and the UV ones

(the fermion). This quantity has a well defined meaning only in the context of the full theory: it is an unknown quantity from the perspective of the low energy EFT.

The small parameters that appear in Eq. (69), instead, have a well defined meaning from a purely low-energy perspective: they are expansion parameters of the EFT. These can be measured experimentally without knowing anything about the physics above the strong coupling scale, and must be small *by definition* in order for the EFT to be a valid perturbative framework. In other words, the masses of the NG fields are guaranteed to be parametrically smaller than the strong coupling scale within the regime of validity of the EFT.

Something similar happened in the EFT for light fermions discussed in the previous chapter. In fact, we could have expressed the natural value of the fermion mass as

$$m = \left(\frac{m}{M}\right) \times M, \quad (70)$$

with M the strong coupling scale, and m/M the expansion parameter that controls the explicit breaking of the discrete symmetry in Eq. (39). This expression is now very similar to the one for the scalar mass in Eq. (69). The fermion mass was protected by the discrete symmetry (39), the meson masses are protected by the non-linearly realized symmetries (50). Massive scalars whose mass is protected by symmetries realized non-linearly are called *pseudo-Nambu-Goldstone bosons* (PNGBs). Presently, we only know of one other way of protecting scalar masses: imposing invariance under supersymmetry, so that scalar masses can inherit the symmetry protection of fermion masses.

3.4 Coset construction

In Sec. 3.1, we pointed out that the form of the effective action (48) is independent of the representation of the operator that spontaneously breaks chiral symmetry. Knowing that this operator was $\bar{\Psi}_R^i \Psi_L^j$ allowed us to guess that $U = e^{2i\pi^a Q_a}$ was a particularly convenient quantity to work with. However, the only thing that ultimately mattered was the symmetry breaking pattern, *i.e.* what was the full symmetry group G of the EFT, and what subgroup H was realized nonlinearly. This of course resonates with the overarching philosophy behind EFTs: the symmetry breaking mechanism is a UV detail, and we shouldn't need to know what physics looks like in the UV to describe what's going on in the IR. This suggests that there should be a way to formulate effective theories of NG bosons based on the symmetry breaking pattern alone: such an approach goes under the name of *coset construction*.¹² It was first developed by Callan, Coleman, Wess and Zumino [28] for spontaneously broken internal symmetries, and later extended to space-time symmetries by Volkov [29] and Ogievetsky [30].

¹²This name comes from the fact that, when a symmetry group G is spontaneously broken down to a subgroup H , the broken transformations belong to the *coset space* G/H .

In this section, we are going to review the basic ideas behind the coset construction in a way that applies both to internal and space-time symmetries.

Consider an arbitrary symmetry breaking pattern of the form $G \rightarrow H$. We will start by grouping the generators of G into three categories:

1. the generators of *unbroken translations*: P_α
2. the generators of all other *unbroken symmetries*: T_A
3. the generators of *broken symmetries*: X_a

This classification is always possible, albeit it is not unique since one can always redefine the broken generators as $X_a \rightarrow X_a + c_{aA}T^A + c_{a\alpha}P^\alpha$, with c_{aA} and $c_{a\alpha}$ some arbitrary coefficients.

In chiral perturbation theory, for example, all translations are unbroken; the T_A 's are the generators of $SU(3)_V \times U(1)_V$; the broken generators are $X_a = 2Q_a$.¹³ With this notation, the building block U_{ij} we used in chiral perturbation theory was just a particular representation of the abstract group element $g = e^{i\pi^a(x)X_a}$. A natural guess would be to try using this quantity for any symmetry breaking pattern. For reasons that will become more clear in a moment, however, it is actually better to work instead with the following quantity:

$$g = e^{ix^\alpha P_\alpha} e^{i\pi^a(x)X_a} . \quad (71)$$

The element g is usually referred to as a coset parametrization. The extra factor of $e^{ix^\alpha P_\alpha}$, describing a translation from the origin of our coordinate system to the point x^α at which the NG fields $\pi^a(x)$ are evaluated, will ensure that the π 's transform as expected under spatial translations. This is especially non-trivial in the presence of spontaneously broken space-time symmetries. But we are getting ahead of ourselves: we still need to define how *any* symmetry acts on the NG fields, let alone translations.

Using the group structure, we define the action of a symmetry transformation \tilde{g} on the quantity g as the product of the two group elements, *i.e.* $g \rightarrow \tilde{g} \cdot g$. What does this imply for the coordinates and the NG fields? In general the group element $\tilde{g} \cdot g$ will not have the particular form in Eq. (71), since it will also depend on unbroken generators. However, this element can always be decomposed uniquely into the product of a term that looks like g (albeit with different values of x^α and π^a) and an element h of the unbroken group H (which in general will depend on the π 's and on \tilde{g}):

$$\tilde{g} \cdot g(x, \pi) = g(\tilde{x}, \tilde{\pi}) \cdot h(\pi, \tilde{g}). \quad (72)$$

¹³To be more precise, if $Q_a^{L,R}$ are the generators of $SU(3)_{L,R}$, then broken generators we worked with are $Q_L^a - Q_R^a$. The quantity U_{ij} transforms in the **3** (**$\bar{3}$**) representation of $SU(3)_L$ ($SU(3)_R$), in which case $Q_L^a = -Q_R^a = Q^a$ with $Q^a = \frac{1}{2}\lambda^a$ and λ^a the Gell-Mann matrices.

This equation defines the transformation rules of coordinates and NG fields, in the sense that $(x^\alpha, \pi^a(x)) \xrightarrow{\tilde{g}} (\tilde{x}^\alpha, \tilde{\pi}^a(\tilde{x}))$. These transformation rules are in general highly non-linear. The explicit form of \tilde{x} , $\tilde{\pi}$, and $h(\pi, \tilde{g})$ can in principle be calculated using the algebra of the symmetry generators. In practice, though, one seldom needs this information.

Exercise 3.3: Use Eq. (72) to show that the NG fields of spontaneously broken internal symmetries transform like scalars under Poincaré transformations.

The need for the factor $h(\pi, \tilde{g})$ in Eq. (72) might not seem obvious based on our discussion of chiral perturbation theory. This is because, in that example, we worked in a particular representation where the generators of $SU(3)_L$ and $SU(3)_R$ were the same up to a sign. As a result, $U = e^{2i\pi^a Q_a}$ was the most general parametrization of an $SU(3)$ element, and therefore it kept the same form no matter which $SU(3)$ transformation acted on it. In other words, in this particular representation one could choose to absorb $h(\pi, \tilde{g})$ in the transformation rules of the NG fields—and that’s exactly what we did in Eq. (50). This is however not possible for a generic representation, and the simple strategy we followed to write down the effective action for chiral perturbation theory will not work. We will have to be more clever.

Our plan will be to derive some fundamental building blocks that depend on the NG fields and have relatively simple transformation rules. NG fields are derivatively coupled, meaning that each term in their Lagrangian contains some derivatives. We therefore start by taking a derivative of g , or even better, by introducing the Maurer-Cartan form $dx^\mu g^{-1} \partial_\mu g$. This quantity is particularly convenient because it is an element of the algebra of G , and therefore can be expressed as a linear combination of all the generators, which we can always cast in the following form:

$$dx^\mu g^{-1} \partial_\mu g = i dx^\mu (e_\mu^\alpha P_\alpha + e_\mu^\alpha \nabla_\alpha \pi^a X_a + C_\mu^A T_A). \quad (73)$$

The coefficients e_μ^α , $\nabla_\alpha \pi^a$, and C_μ^A are non-linear functions of the Goldstone π ’s, and are the main building blocks that we’ll use to build the effective action for NG fields. Their explicit form can be calculated using the algebra of the group G . We will not need to know the exact commutation relations in what follows, but we will assume that

$$[T_A, X_a] = i f_{Aab} X^b, \quad (74a)$$

$$[T_A, P_\alpha] = i f_{A\alpha\beta} P^\beta, \quad (74b)$$

meaning that the generators of translations and broken transformations form a (possibly reducible) representation of the unbroken group H . This requirement is not very restrictive; for example, Eq. (74a) is satisfied by any compact, internal symmetry group, as well as in many other physically interesting situations.

Exercise 3.4: For compact groups, one can always choose the generators in such a way that the structure constants are totally antisymmetric [31]. Use this result together with the fact that the T_A 's span a subgroup to prove Eq. (74a). **Hint:** this amounts to showing that the righthand side of (74a) cannot depend on the T_A 's.

Using now the transformation rule (72) and the commutation relations (74), it is easy to show that the coefficients of the Maurer-Cartan form must transform as follows under the action of \tilde{g} :

$$e_\mu^\alpha \rightarrow h(\pi, \tilde{g})^\alpha_\beta e_\mu^\beta, \quad (75a)$$

$$\nabla_\alpha \pi^a \rightarrow h(\pi, \tilde{g})^a_b \nabla_\beta \pi^b h^{-1}(\pi, \tilde{g})^\beta_\alpha, \quad (75b)$$

$$C_\mu^A T_A \rightarrow h(\pi, \tilde{g}) C_\mu^A T_A h^{-1}(\pi, \tilde{g}) - i h(\pi, \tilde{g}) \partial_\mu h^{-1}(\pi, \tilde{g}), \quad (75c)$$

where $h(\pi, \tilde{g})^\alpha_\beta$ and $h(\pi, \tilde{g})^a_b$ are different representations of the abstract group element $h(\pi, \tilde{g})$. These transformation rules are not as complicated as they might appear at first sight. They are highly non-linear in the NG fields, but the non-linearity is “packaged” inside $h(\pi, \tilde{g})$, e_μ^α , $\nabla_\alpha \pi^a$ and C_μ^A . The generic group element \tilde{g} only enters through the unbroken element $h(\pi, \tilde{g})$, which in turn acts on the coefficients of the Maurer-Cartan form in fairly simple ways, suggesting the role that each of these building blocks will play.

The coefficients e_μ^α are known as the *coset vierbein*, because their transformation properties resemble those of the vierbein in General Relativity under local Lorentz transformations [32]. If the X 's are generators of internal symmetries, the coset vierbein is just a Kronecker delta and $h(\pi, \tilde{g})$ can only be a global Lorentz transformation (or Galilei transformation, in non-relativistic systems). Things are more interesting when some space-time symmetries are spontaneously broken, in which case e_μ^α and $h(\pi, \tilde{g})^\alpha_\beta$ acquire a non-trivial dependence on the NG fields. The latter however remains a tensor with unit determinant, and therefore the measure of integration $d^4x \det e$ is invariant under the transformations of the form (75a). As a bonus, this measure is also invariant under arbitrary redefinitions of the coordinates x^μ . This allows us in principle to work not just in the Cartesian coordinates implicitly used in Eq. (71), but in fact in arbitrary coordinate systems.

The quantities $\nabla_\alpha \pi^a$ are known as *coset covariant derivatives*. Despite the notation, we shouldn't think of these covariant derivatives as the action of some derivative operator ∇_α on the fields π^a . These are instead particularly clever non-linear combinations of the NG fields of the form

$$\nabla_\alpha \pi^a = \partial_\alpha \pi^a + \mathcal{O}(\pi^2). \quad (76)$$

The non-linear terms ensure that $\nabla_\alpha \pi^a$ transforms in a linear representation of $h(\pi, \tilde{g})$ —unlike the $\partial_\alpha \pi^a$, which by itself would transform in a much more complicated way. It is very

easy to build invariant operators out of quantities that transform linearly: we just need to contract all the indices in the appropriate way. The amazing property of the coset covariant derivatives is that contractions that are manifestly invariant under unbroken transformations h , are also guaranteed to be secretly invariant under broken transformations!

Finally, we will call C_μ^A the coset connection, because it transforms exactly like a gauge field under $h(\pi, \tilde{g})$ transformations. We can use it to define a covariant derivative

$$D_\alpha = (e^{-1})_\alpha{}^\mu (\partial_\mu + i C_\mu^A T_A), \quad (77)$$

which preserves the linearity of the transformation rules when acting on $\nabla_\nu \pi^a$ —or, for that matter, on any other field that transforms at $\Psi^i \rightarrow h(\pi, \tilde{g})^i{}_j \Psi^j$.

Putting it all together, we have found that effective action for the NG fields associated with any symmetry breaking pattern $G \rightarrow H$ takes the form

$$S = \int d^4x \det e \mathcal{L}(\nabla_\alpha \pi^a, D_\alpha), \quad (78)$$

where the Lagrangian \mathcal{L} contains all possible operators that are manifestly invariant under the unbroken transformations. The main building blocks of this action are defined in Eq. (73), and can be calculated using only the algebra of the symmetry generators. This procedure is in principle straightforward, but as usual the devil is in the details, so let's spell out a few of them:

- We have already mentioned that spontaneously broken generators are only defined up to the addition unbroken generators. Different choices of broken generators correspond to field redefinitions of the NG fields. This freedom can always be leveraged to simplify the calculations of the Maurer-Cartan form. For example, in chiral perturbation theory it is especially convenient to work with the broken generators $X_a = Q_a^L = \frac{1}{2}(Q_a^L - Q_a^R) + \frac{1}{2}(Q_a^L + Q_a^R)$, because this choice makes the coset connection vanish.
- It is usually convenient to break up the broken generators X_a into irreducible representations $X_a^{(j)}$ of the unbroken symmetries, and to use a coset parametrization given by a product of several factors, *i.e.* $e^{i\pi^a(x)X_a^{(1)}} e^{i\pi^b(x)X_b^{(2)}} \dots$. These exponentials should be ordered in such a way that the $X^{(1)}$'s form a representation of the symmetries generated by the $X^{(2)}$'s (*i.e.* $[X_a^{(1)}, X_b^{(2)}] = if_{abc}X_c^{(1)}$), and so on. Factorizing the coset parametrization in this way amounts once again to a field redefinition of the π 's. From this perspective, chiral perturbation theory is a particularly simple example because the broken generators form an irreducible representation of $SU(3)_V$.

- A couple of identities you might find helpful when performing explicit calculations are:

$$e^{-Y} X e^Y = X + [X, Y] + \frac{1}{2!} [[X, Y], Y] + \frac{1}{3!} [[[X, Y], Y], Y] + \dots, \quad (79a)$$

$$e^{-i\pi^a X_a} \partial_\mu e^{i\pi^a X_a} = i \partial_\mu \pi^b \int_0^1 ds e^{-i(1-s)\pi^a X_a} X_b e^{i(1-s)\pi^c X_c} \quad (79b)$$

$$= i \partial_\mu \pi^a \left(X_a + \frac{i}{2!} \pi^b [X_a, X_b] + \frac{i^2}{3!} \pi^b \pi^c [[X_a, X_b], X_c] + \dots \right). \quad (79c)$$

Calculating the first few terms in these expansions is usually sufficient to notice a pattern and re-sum the entire series.

- The effective action (78) can be easily generalized by including additional matter fields that transform under irreducible representations of the unbroken symmetries. Interaction terms that are manifestly invariant under all unbroken symmetries will also preserve the broken ones as long as the matter fields transform as $\Psi \rightarrow h(\pi, \tilde{g}) \Psi$. In chiral perturbation theory, one can describe in this way processes that involve both mesons and light leptons.
- Finally, one can also take into introduce gauge fields by promoting the ordinary derivative appearing in the Maurer-Cartan form to a gauge covariant derivative:

$$dx^\mu g^{-1} \partial_\mu g \rightarrow dx^\mu g^{-1} (\partial_\mu - ie A_\mu^i Q_i) g, \quad (80)$$

where the Q_i 's can be a combination of broken and unbroken generators. Our coset building blocks will now depend not only on the NG fields, but also on the gauge fields A_μ^i . For any gauge generator that is broken, one can always set to zero the corresponding NG field in the coset parametrization by working in unitary gauge.

Exercise 3.5: Rederive the effective action for chiral perturbation theory in Eq. (48) using the coset construction.

3.5 Inverse Higgs constraints

The textbook version of Goldstone's theorem, asserting the existence of one NG mode for each broken generator, only applies to internal symmetries. When space-time symmetries are spontaneously broken, there is often a mismatch between the number of broken generators and that of NG modes. Think for instance of a relativistic 4D brane in 5 dimensions, whose embedding is defined by the condition $X^5 = 0$. This constraint breaks spontaneously translations in the 5-direction (generated by P_5) as well as Lorentz transformations that mix the 5-direction with directions along the brane (generated by $J_{\alpha 5}$). And yet, this system has a single massless mode corresponding to local deformations of the brane in the normal direction.

How does the coset construction capture this mismatch between NG modes and broken generators? The answer is that, when space-time symmetries are spontaneously broken, it is possible to impose additional local constraints that preserve all the symmetries and can be solved to express some of the NG modes in terms of others. These are known as inverse Higgs constraints, and amount to setting to zero one or more multiplets (under H) of coset covariant derivatives. The rule of thumb to determine which constraints can be imposed is that, if $X_a^{(1)}$ and $X_a^{(2)}$ are two multiplets of broken generators such that

$$[P_\alpha, X_a^{(1)}] \supset i f_{\alpha ab} X_{(2)}^b, \quad (81)$$

then we can set $f^{\alpha ab} \nabla_\alpha \pi_b^{(2)} = 0$ to eliminate the $\pi_a^{(1)}$'s from the effective theory. Notice that the righthand side of Eq. (81) can be non-zero only if the $X_a^{(1)}$'s generate space-time symmetries. It is easy to see that our rule of thumb is indeed satisfied by the 4D brane we mentioned above, since $[P_\alpha, J_{5\beta}] = i \eta_{\alpha\beta} P_5$.

What is the rationale behind the criterion in Eq. (81)? Whenever this condition is satisfied, one can show that

$$f^{\alpha ab} \nabla_\alpha \pi_b^{(2)} = f^{\alpha ab} \partial_\alpha \pi_b^{(2)} + \kappa \pi_{(1)}^a + \dots, \quad (82)$$

where κ is some constant and the dots stand for terms non-linear in the NG fields. The inverse Higgs constraint can then be solved order-by-order in the π 's to obtain a local expression for $\pi_{(1)}^a$ in terms of $\pi_b^{(2)}$, their derivatives, and possibly other NG fields.

By imposing all possible inverse Higgs constraints, one obtains the most minimal field content needed to realize a given symmetry breaking pattern. We can choose not to impose some inverse Higgs constraints, but then Eq. (82) shows that operators of the form $(f^{\alpha ab} \nabla_\alpha \pi_b^{(2)})^2$ will give rise to a mass term for the $\pi_{(1)}^a$'s. Thus, those NG modes that could be removed by inverse Higgs constraints are always massive. Keeping them in the effective theory is possible, but it's akin to including additional matter fields. This can be seen explicitly by performing a local field redefinition to trade the $\pi_{(1)}^a$'s for $\Psi^a = f^{\alpha ab} \nabla_\alpha \pi_b^{(2)}$, which transform exactly like matter fields, *i.e.* $\Psi^a \rightarrow h(\pi, \tilde{g})^a_b \Psi^b$.

Exercise 3.6: Consider a relativistic point particle at rest. What space-time symmetries does it break? Use the coset construction and impose all possible inverse Higgs constraints to derive the usual world-line action for a relativistic point particle.

3.6 Wess-Zumino-Witten terms

Let's turn our attention back to chiral perturbation theory for a moment. Remarkably, all the terms in the effective action (48) are invariant under the symmetry $U \leftrightarrow U^\dagger$, which is

also equivalent to $\pi^a \rightarrow -\pi^a$. Even those terms that explicitly break chiral symmetry, such as those in Eqs. (58) and (66), preserve this discrete symmetry. This symmetry forbids scattering processes involving an odd number of mesons. Unfortunately, this is in conflict by experiments, where processes such as $K^+K^- \rightarrow \pi^+\pi^0\pi^-$ are observed.

This apparent contradiction is due to the fact that all the operators we have considered so far are exactly invariant under chiral symmetry¹⁴—as opposed to invariant up to a total derivative. The coset construction makes this very clear: by contracting indices in a way that preserves the unbroken symmetries, we can only build operators that are exactly invariant. However, terms that are invariant only up to a total derivative also belong in the effective action. These terms are usually referred to as *Wess-Zumino-Witten* (WZW) terms in the context of effective theories for NG modes.

In the case of chiral perturbation theory, there happens to be only one WZW term. In order to write it down, we will resort to a clever trick. Let's extend spacetime to 5 dimensions, in such a way that ordinary 4D space-time is the boundary of this 5D region. We will denote with X^M the 5D coordinates. Consider now the following 5D action:

$$S_{WZW} = \int d^5X \left\{ -\frac{2i\lambda}{15\pi^2} \epsilon^{MNPQR} \text{Tr} \left[U^\dagger (\partial_M U) U^\dagger (\partial_N U) U^\dagger (\partial_P U) U^\dagger (\partial_Q U) U^\dagger (\partial_R U) \right] \right\}, \quad (83)$$

where the definition of the coupling λ has been chosen for later convenience. It is exactly invariant under chiral transformations, and its integrand has a special property: when expanded in powers of NG fields, each order is a total derivative. This means that, despite the appearances, this is actually a boundary term, *i.e.* an integral over 4D space-time. At lowest order in the canonically normalized NG fields, it reads

$$S_{WZW} = \int d^4x \frac{64\lambda}{15\pi^2 f^5} \epsilon^{\mu\nu\lambda\rho} \hat{\pi}^a \partial_\mu \hat{\pi}^b \partial_\nu \hat{\pi}^c \partial_\lambda \hat{\pi}^d \partial_\rho \hat{\pi}^e \text{Tr}(Q_a Q_b Q_c Q_d Q_e) + \dots, \quad (84)$$

and therefore it violates the $\pi^a \rightarrow -\pi^a$ symmetry. At lowest order, the broken chiral transformations act on the NG fields as constant shifts: $\pi^a \rightarrow \pi^a + \epsilon^a + \dots$. It is easy to check that the leading order term in Eq. (84) is in fact invariant up to a total derivative under these transformations. The quintic coupling in Eq. (84) accounts precisely for processes such as $K^+K^- \rightarrow \pi^+\pi^0\pi^-$.

WZW terms enjoy very special non-renormalization properties, in that they do not get renormalized by operators that are exactly invariant. To prove this, imagine promoting the Wilson coefficients of all the exactly invariant terms to spurion fields. These terms will remain invariant under chiral symmetry provided the spurions do not transform. However, if the couplings in front of WZW terms were to receive quantum corrections proportional to

¹⁴Assuming that the spurions transform appropriately.

the spurions, they would effectively become functions of coordinates, spoiling invariance up to a total derivative. We conclude therefore that WZW terms cannot get renormalized by exactly invariant operators.

In the particular case of chiral perturbation theory, one could have of course reached the same conclusion by simply invoking the fact that loops built out of exactly invariant operators should preserve the $\pi^a \rightarrow -\pi^a$ symmetry, and thus cannot give rise to the WZW term in Eq. (84). In fact, this term is even more robust, in that doesn't get renormalized at all: it is protected for topological reasons, which require λ to be an integer number [33].¹⁵

The procedure we followed to introduce the WZW term for chiral perturbation theory might have seemed very *ad hoc*. However, it is just a particular example of a very general method, based once again on the coset construction. To conclude this section, we will briefly review the main idea, pointing out along the way a connection with a beautiful mathematical structure. Unfortunately, we will only be able to scratch the surface.

Let's forget for a moment that the NG fields are functions of the space-time coordinates, and treat them as independent quantities. This allows us to build exactly invariant 5-forms by wedging together the coefficients of the Maurer-Cartan from $g^{-1}dg$, each of which is a 1-form, and contracting all the indices appropriately. In order to build WZW terms, we are interested in those 5-forms α that are exact, *i.e.* such that $\alpha = d\beta$ with β some 4-form. These 5-forms should be built *before* imposing any inverse Higgs constraint. The form β does not need to be exactly invariant, but it can at most shift by a total derivative under a symmetry transformation since α must remain invariant:

$$\beta \rightarrow \beta + d\gamma. \quad (85)$$

If β does shift by a total derivative, then its integral over 4D space-time is a WZW term.

According to this definition, β is only defined up to the addition of exactly invariant terms. How many truly independent WZW terms are there for a given symmetry breaking pattern $G \rightarrow H$? It turns out that there is a one-to-one correspondence between the number of WZW terms and the elements of the relative Lie algebra cohomology of G/H . See [34] and references therein for many more details.

Exercise 3.7: The standard kinetic term for a non-relativistic point particle is invariant under boosts only up to a total derivative. Using the algebra of the Galilei group (which you can find at the end of Sec. 2.4 of [4]), carry out the coset construction and follow the procedure outlined above to write the kinetic term as a WZW term. **Hint:** The EFT for

¹⁵I should stress that this is a special property of the WZW term in Eq. (83). It follows from the fact that, in this case, the homotopy group $\pi_4(G/H)$ is non-trivial. WZW terms do not necessarily have Wilson coefficients that are quantized.

a point particle is a (0+1)-dimensional field theory. Therefore, WZW terms are integrals of exact 2-forms.

3.7 Chiral anomalies

The WZW term in Eq. (84) has deep connections with anomalies, which unfortunately we will not be able to explore in full detail. Still, I cannot resist the temptation of saying at least a few words about chiral anomalies. One can introduce the coupling with electromagnetism by supplementing the terms in Eq. (84) with

$$\begin{aligned} \Delta S_{WZW} = \int d^4x \Big\{ & -\frac{e\lambda}{48\pi^2} \epsilon^{\mu\nu\lambda\rho} A_\mu \text{Tr} [Q_L(\partial_\nu U)U^\dagger(\partial_\lambda U)U^\dagger(\partial_\rho U)U^\dagger \\ & + Q_R U^\dagger(\partial_\nu U)U^\dagger(\partial_\lambda U)U^\dagger(\partial_\rho U)] \\ & + \frac{ie^2\lambda}{24\pi^2} \epsilon^{\mu\nu\lambda\rho} (\partial_\mu A_\nu) A_\lambda \text{Tr} [Q_L^2(\partial_\rho U)U^\dagger + Q_R^2 U^\dagger(\partial_\rho U) + Q_L U Q_R U^\dagger(\partial_\rho U)U^\dagger] \Big\}, \end{aligned} \quad (86)$$

with Q^L and Q^R defined in Eq. (65). Notice that in this case gauging is not as straightforward as usual because the WZW term is not exactly invariant. When expanded in powers of the NG fields, the righthand side of Eq. (86) contains a term of the form

$$\Delta S_{WZW} \supset \int d^4x \frac{\lambda e^2}{48\pi^2 f} \hat{\pi}^0 \epsilon^{\mu\nu\lambda\rho} F_{\mu\nu} F_{\lambda\rho} \quad (87)$$

with $\hat{\pi}^0$ the canonically normalized field that describes neutral pions. This interaction is responsible for the decay process $\pi^0 \rightarrow \gamma\gamma$, which in turn is a consequence of the fact that the non-linearly realized symmetry associated with π^0 is anomalous. To see this, we can calculate the Noether current $J_{3,A}^\mu$ for this symmetry¹⁶ following the standard procedure. We find that its divergence doesn't vanish even in the limit of zero quark masses:

$$\partial_\mu J_{3,A}^\mu = -\frac{\lambda e^2}{48\pi^2} \epsilon^{\mu\nu\lambda\rho} F_{\mu\nu} F_{\lambda\rho}. \quad (88)$$

The righthand side arises precisely from the term in Eq. (86). This result is an exact relation between operators that are well defined at all energies. As such, it can be derived not only using chiral perturbation theory, but also from a one-loop calculation in QCD, where one finds that λ is equal to the number of colors, *i.e.* $\lambda = 3$.

It is worth stressing that ΔS_{WZW} is the *only* term that prevents the divergence of $J_{3,A}^\mu$ from vanishing: in other words, it is the only term that breaks the associated symmetry. From the perspective of QCD, this is a manifestation of the fact that anomalies are 1-loop exact, meaning that the righthand side of (88) doesn't receive contributions from higher-loops [35]. By contrast, when a symmetry is explicitly broken, the effective action contains

¹⁶The subscript A on the current stands for “axial”, meaning that $J_{3,A}^\mu = J_{3,L}^\mu - J_{3,R}^\mu$; the subscript “3” stands for the fact that the symmetry generator associated with π^0 is $Q_3 = \frac{1}{2}\lambda_3$.

an infinite number of symmetry-breaking operators. Anomalies are therefore a very special way of breaking symmetries.

The reader might remember that we invoked another anomaly at the beginning of this chapter to argue away the axial symmetry $U(1)_A$, and therefore the existence of a corresponding NG mode. This could seem at odds with the fact that $J_{3,A}^\mu$ is also anomalous, and yet its spontaneous breaking gives rise to the NG field we identify with the π^0 meson. The difference between these two anomalies lies in the fact that, had we done a 1-loop calculation in QCD, we would have found that

$$\partial_\mu J_A^\mu = -\frac{g^2}{16\pi^2} \epsilon^{\mu\nu\lambda\rho} \text{Tr}(G_{\mu\nu} G_{\lambda\rho}). \quad (89)$$

Unlike in Eq. (88), the anomaly in the conservation of J_A^μ is now proportional to the field strength of strong interactions. This is once again a relation among operators, and therefore should hold at all scales. However, at low energies all quarks are confined, and there is no trace of QCD; therefore, there shouldn't be any trace of the $U(1)_A$ symmetry either.

3.8 Gauge anomalies

Since we have opened up the pandora's box of anomalies, I would like to also say a few words about anomalous gauge symmetries. The standard lore is that gauge anomalies should be avoided, otherwise the resulting gauge theory doesn't make sense. In fact, absence of gauge anomalies is often used as a constraint on the particle content of EFTs. And yet, I am now going to argue that anomalous gauge theories are just fine from an EFT perspective as long as the gauge boson is massive [36].

Let's consider a concrete example, namely a $U(1)$ gauge theory with N left-handed Weyl fermions described by the action

$$S = \int d^4x \left\{ -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \sum_{j=1}^N i \psi_j^\dagger \bar{\sigma}^\mu (\partial_\mu - ie q_j A_\mu) \psi_j \right\}. \quad (90)$$

Without going too much into details—which you can find for instance in [1]—it turns out that the $U(1)$ symmetry is anomalous unless the following condition is satisfied:

$$\sum_j q_j^3 = 0. \quad (91)$$

This combination of charges is usually referred to as the *anomaly coefficient*. Let's assume therefore that such requirement is met, and now add to the theory a complex scalar field with the following interactions:

$$\Delta S = \int d^4x \left\{ -|(\partial_\mu - ie A_\mu) \Phi|^2 - \frac{\lambda}{2} (|\Phi|^2 - v^2)^2 - (y \Phi^\dagger \psi_N \cdot \psi_N + \text{h.c.}) \right\}. \quad (92)$$

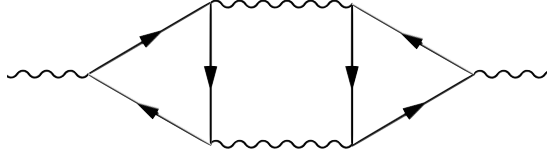


Figure 2: Feynman diagram contributing to the gauge boson mass.

The Yukawa interaction between Φ and ψ_N is gauge invariant provided $q_N = \frac{1}{2}$; moreover, it is technically natural for all other Yukawa interactions to be absent as long as the charges of the remaining Weyl fermions do not have this value. Expanding around the vev $\langle \Phi \rangle = v$, we find that the gauge field, the scalar, and the N -th fermion acquire masses $m_A = 2ev$, $m_\Phi = \sqrt{2\lambda}v$ and $m_N = 2yv$ respectively. If $\lambda, y \gg e$, we can choose to integrate out the scalar fluctuation and the N -th fermion and work with a low-energy EFT that only includes the (massive) gauge boson and $N - 1$ fermions. By construction, this EFT doesn't satisfy Eq. (91), and yet we know that this EFT must be fine because it originated from an anomaly-free, renormalizable theory. In this construction, we violated (91) at low energies at the expense of the gauge bosons becoming massive. This is actually a general result: since the masslessness of gauge bosons is enforced by the gauge symmetry, it make sense for A_μ to have a mass if the $U(1)$ is anomalous and hence not an actual symmetry.

We can even derive a lower bound on the gauge boson mass by resorting to a naturalness argument. Imagine starting with a massless gauge boson at tree level, and consider the Feynman diagram shown in Fig. 2, with the N -th quark running in the triangular loops. A reader well-versed in anomalies will recognize that this graph is equal to two copies of a triangular fermion diagram—the hallmark of chiral anomalies—glued together by two photon lines. This 3-loop diagram is UV divergent, but there is no way to regulate it in such a way that gauge invariance is preserved. In order to match this diagram in the low-energy EFT one needs to add, among others, a mass term for the gauge boson with

$$m_A^2 \sim \left[\frac{1}{(4\pi)^2} \right]^3 \times (eq_N)^6 \times (2yv)^2. \quad (93)$$

The origin of the three factors on the right-hand side should be clear: one loop factor $1/(4\pi)^2$ for each loop, one factor of eq_N for each vertex, and the fermion mass $2yv$ is the only mass scale appearing in this diagram since we assumed the gauge boson mass to be zero at tree level. We can also rewrite Eq. (93) in terms of quantities that are defined purely in the context of the low-energy EFT:

$$m_A^2 \sim \left(\frac{e}{4\pi} \right)^6 \times \left(\sum_{j=1}^{N-1} q_j^3 \right)^2 \times \Lambda^2, \quad (94)$$

with Λ the strong coupling scale of the EFT. *A posteriori*, we could have actually guessed this result based on the following observations:

1. the strong coupling scale of the EFT is the only scale in the problem;
2. physical quantities can only depend on the square of e , because the Lagrangian is invariant under $e \rightarrow -e, A_\mu \rightarrow -A_\mu$;
3. it makes sense for the mass to be proportional to the anomaly coefficient, since this is the quantity that controls whether the $U(1)$ is actually a symmetry;
4. each factor of e should come with a factor of q_j , since the Lagrangian is invariant under $e \rightarrow -e, q_j \rightarrow -q_j$.

The expression in Eq. (94) is actually quite general and, on naturalness grounds, provides a lower bound on the mass of an abelian gauge field in an EFT with anomalous fermion content.

3.9 Additional resources

More in-depth reviews on chiral perturbation theory are [37, 38, 27]. General EFTs for Goldstone modes are the focus of the reviews [39, 40]. You can read much more about anomalies in [35, 41], and about anomalous Yang-Mills theories in [36].

4 EFTs for non-relativistic fermions

4.1 Non-relativistic fermions in vacuum

Power counting was particularly easy in the toy model of Sec. 2 because the fermions were relativistic, *i.e.* they had energies $E \gg mc^2$.¹⁷ As a result, each component of a partial derivative acting on Ψ scaled the same way, namely $\partial_\mu \sim E/c$. This is however no longer true for scattering processes involving non-relativistic fermions, which have a momentum $p \ll mc$. In this case,

$$\partial_\mu \Psi \sim i(-mc, \vec{p}) \Psi, \quad (95)$$

and the zero-th component is now much larger than the spatial components. To be clear, the fact that different derivatives scale differently is not necessarily an issue—as we will see, this happens in many EFTs. The problem here is that the zero-th component doesn't scale at all with the momentum (or equivalently, the energy) of the fermion. For this reason, higher time derivatives of Ψ are not necessarily more suppressed at low energies, and therefore operators in the Lagrangian are not clearly organized in powers of momentum. Fortunately, this can be easily remedied by introducing a new field $\tilde{\Psi}$ via the following *rephasing*:

$$\Psi(t, \vec{x}) \equiv e^{-imc^2 t} \tilde{\Psi}(t, \vec{x}) \quad (96)$$

Derivatives of $\tilde{\Psi}$ now scale like

$$\partial_\mu \tilde{\Psi} = \partial_\mu \left(e^{imc^2 t} \Psi \right) \sim i \left(-\frac{p^2}{2mc}, \vec{p} \right) \tilde{\Psi} \quad (97)$$

and therefore operators with additional derivatives are more suppressed for small momenta (compared to the strong coupling scale of the effective theory).

It is instructive to implement the field redefinition directly at the level of the Lagrangian. Take for instance the action in Eq. (5) with $(\bar{\Psi} \gamma^5 \Psi)^2 \rightarrow (\bar{\Psi} \Psi)^2$. We can write the action directly in terms of $\tilde{\Psi}$ with all the powers of c and gamma matrices spelled out as follows:

$$S_{\text{eff}} = \int dt d^3x c \left\{ \tilde{\Psi}^\dagger \left[\frac{i\partial_t}{c} + i\gamma^0 \gamma^i \partial_i + (1 - \gamma^0) mc \right] \tilde{\Psi} - \frac{y^2}{2M^2 c} (\tilde{\Psi}^\dagger \gamma^0 \tilde{\Psi})^2 \right\}. \quad (98)$$

The factor of $1/c$ in front of the quartic interaction can always be reabsorbed by redefining the coupling constant, but has been introduced for later convenience. According to the power counting rule (97) for the derivatives of $\tilde{\Psi}$, the dominant term in square brackets is the one proportional to mc . This suggests that we introduce the projection matrices¹⁸ $P_\pm \equiv \frac{1}{2}(1 \pm \gamma^0)$

¹⁷In this section we will reintroduce all factors of c ; this means in particular that $\partial_0 = \partial_t/c$.

¹⁸It is easy to check that P_\pm indeed satisfy the defining properties of projectors ($P_\pm^2 = P_\pm$). Moreover, they are orthogonal ($P_+ P_- = P_- P_+ = 0$) and form a complete set ($P_+ + P_- = 1$).

to decompose $\tilde{\Psi}$ as a sum of $\tilde{\Psi}_{\pm} \equiv P_{\pm} \tilde{\Psi}$. Because the leading quadratic term in the Lagrangian is $mc^2 \tilde{\Psi}_+^\dagger \tilde{\Psi}_+$, we can integrate out $\tilde{\Psi}_+$ by working perturbatively in p/mc —the expansion parameter of our non-relativistic EFT—and obtain an effective Lagrangian for $\tilde{\Psi}_-$ alone. Notice that, since for non-relativistic particles we have $p = mv$, our expansion parameter can also be expressed as v/c .

Exercise 4.1: integrate out $\tilde{\Psi}_+$ at tree-level and derive an effective action for $\tilde{\Psi}_-$ at lowest order in p/mc .

Physically, we can integrate out half of the modes in the non-relativistic regime because low-energy scattering events involving only, say, particles, are unable to produce anti-particles by conservation of energy. We can make this more explicit by working with the Dirac representation of the Dirac matrices [42], in which γ^0 is diagonal:¹⁹

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \Longrightarrow \quad \tilde{\Psi}_- = \begin{pmatrix} 0 \\ \psi \end{pmatrix}. \quad (99)$$

Then, the leading order Lagrangian for the 2-component spinor ψ is

$$S_{\text{eff}} = \int dt d^3x \left\{ \psi^\dagger \left(i\partial_t + \frac{\nabla^2}{2m} \right) \psi - \frac{y^2}{2M^2} (\psi^\dagger \psi)^2 \right\}. \quad (100)$$

It is now time to do some more power counting. From the scaling of derivatives in Eq. (97) we infer that $dt d^3x \sim 1/p^5$. Notice that, unlike the examples discussed in the previous two sections, the scaling with momentum can no longer be inferred from dimensional analysis in natural units. The scaling of ψ is determined as usual by the requirement that the free action doesn't scale, which implies $\psi \sim p^{3/2}$. Therefore, we see that the quartic interaction we have kept scales like $\sim p$ and is therefore irrelevant at small momenta.²⁰ In fact, *any* interaction involving just fermions in vacuum turns out to be irrelevant. We will see in a moment that this is not the case for fermions at finite density. Before, though, let's discuss how the symmetries of our effective action (100) relate to those of the relativistic action we started from.

¹⁹This basis is more convenient in the non-relativistic limit compared to the Weyl basis used in [1], in which γ^5 is diagonal. Spinors in the two basis are connected by a linear transformation $\Psi_{\text{Dirac}} = S \Psi_{\text{Weyl}}$ with $S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$.

²⁰You can use Eq. (17) and the non-relativistic version of the LSZ equation, in which $\not{\partial} u_j \rightarrow i\partial_t + \frac{\nabla^2}{2m}$, to convince yourself that this scaling is consistent with the fact that the 4-point amplitude is a constant at lowest order in perturbation theory.

4.2 Contracted and accidental symmetries

Our relativistic Lagrangian is invariant under Lorentz transformations $x^{\mu'} = \Lambda^{\mu}_{\nu} x^{\nu}$ provided Ψ transforms as follows (in the Dirac representation):

$$\Psi'(x') = \frac{1}{2} \begin{pmatrix} e^{\frac{i}{2}\vec{\sigma}\cdot(\vec{\alpha}+i\vec{\beta})} + e^{\frac{i}{2}\vec{\sigma}\cdot(\vec{\alpha}-i\vec{\beta})} & e^{\frac{i}{2}\vec{\sigma}\cdot(\vec{\alpha}+i\vec{\beta})} - e^{\frac{i}{2}\vec{\sigma}\cdot(\vec{\alpha}-i\vec{\beta})} \\ e^{\frac{i}{2}\vec{\sigma}\cdot(\vec{\alpha}+i\vec{\beta})} - e^{\frac{i}{2}\vec{\sigma}\cdot(\vec{\alpha}-i\vec{\beta})} & e^{\frac{i}{2}\vec{\sigma}\cdot(\vec{\alpha}+i\vec{\beta})} + e^{\frac{i}{2}\vec{\sigma}\cdot(\vec{\alpha}-i\vec{\beta})} \end{pmatrix} \Psi(x), \quad (101)$$

where the α 's parametrize rotations and the vector $\vec{\beta} \equiv \vec{v}/c$ parametrizes Lorentz boosts. At lowest order in a p/mc expansion, Eq. (101) implies the following transformation rule for the 2-component spinor ψ :

$$\psi(t, \vec{x}') = e^{im(v^2 t/2 - \vec{v}\cdot\vec{x})} e^{i\vec{\sigma}\cdot\vec{\alpha}/2} \psi(t, \vec{x}), \quad (102)$$

while coordinates transform as usual under rotations and Galilean boosts.

Exercise 4.2: show that the differential operator $D \equiv i\partial_t + \nabla^2/(2m)$ is a covariant derivative with respect to boosts, in the sense that $D'\psi'(t, \vec{x}') = e^{im(v^2 t/2 - \vec{v}\cdot\vec{x})} D\psi(t, \vec{x})$ with $\vec{x}' = \vec{x} - \vec{v}t$.

By using this result, we can see immediately that the non-relativistic Lagrangian (100) is invariant under rotations as well as Galilean boosts. This is of course just the $c \rightarrow \infty$ limit of Lorentz invariance. For lack of a better term, I will call this a contracted symmetry, since the algebra of Galilean transformations follows from a contraction of the Lorentz algebra [4].

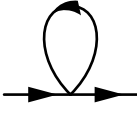
There is also an internal $U(2) = SU(2) \times U(1)$ symmetry transformation that acts on ψ without changing the coordinates. The $U(1)$ subgroup is just a remnant of the $U(1)$ symmetry of the relativistic action—except that at low energies it is the number of particles, rather than particles minus antiparticles, that is conserved by kinematics. The $SU(2)$ subgroup instead doesn't have a relativistic counterpart, and is a manifestation of the well known fact that spin and coordinates can rotate independently in a non-relativistic theory.

This additional $SU(2)$ is good example of accidental symmetry. As the name suggests, these are symmetries that arise accidentally when the particle content and fundamental symmetries of an EFT are so constraining that one can only write down invariant operators up to some order in the EFT expansion. Accidental symmetries are eventually broken by higher order corrections in the EFT. Thus, one can in principle distinguish between accidental and fundamental symmetries by testing the EFT with high enough accuracy. Some examples of accidental symmetries in the Standard Model are the ones associated with baryon and lepton number conservation, or the custodial symmetry of the Higgs sector.

Looking back at the lowest-order effective action (100), we now realize that we could have *almost* guessed its form based on symmetry arguments alone. Invariance under spatial

rotations requires the spin index of each ψ to be contracted with an index carried by a ψ^\dagger , while invariance under Galilean boosts requires us to use the covariant derivative D . Granted, at first order in derivatives we could have also added a term of the form $(D\psi)^\dagger\psi$, but it is easy to show that this is equivalent to $\psi^\dagger D\psi$ up to integrations by parts. More puzzling instead is the absence of a quadratic term $\psi^\dagger\psi$. This term would be compatible with all the symmetries we've discussed so far, and its absence doesn't lead to a symmetry enhancement. In light of our previous discussions on naturalness, you might now be worried that our effective action (100) is fine-tuned. In fact, it is easy to see that such a term doesn't get renormalized, and therefore it is self-consistent to set it to zero.

Consider for instance the following loop correction, which would renormalize the mass term of a relativistic scalar:



$$= -i \frac{y^2}{M^2} \int \frac{dE d^3p}{(2\pi)^3} \frac{i}{E - p^2/2m + i\epsilon} = -i \frac{y^2}{M^2} \int \frac{dQ^2 d^3p}{(2\pi)^3} \frac{i}{Q^2 - p^2 + i\epsilon}. \quad (103)$$

This integral is UV divergent, but after redefining the integration variable E it doesn't involve any length scale, and therefore vanishes in dimensional regularization. A similar argument applies also to higher loop corrections to the 2-point function. After absorbing the factor of $2m$ in the energies, the only scales in the loop integrals are the external energy and momentum. This means that, in dimensional regularizations, loops will only renormalize quadratic operators with derivatives.

The fact that the operator $\psi^\dagger\psi$ doesn't get renormalized admits an intuitive explanation. Using standard arguments to calculate the $U(1)$ Noether current from the action (100), one finds that operator $\psi^\dagger\psi$ is the density of $U(1)$ charge. Thus, adding a term of the form $\Delta L = \mu_{\text{nr}}\psi^\dagger\psi$ corresponds to turning on a source for the charge density—*i.e.*, a chemical potential. This is something that an experimenter should be able to do at will, without having to worry about loop corrections!

4.3 Non-relativistic fermions at finite density

Let's now consider a state with a finite density of fermions. This can be achieved by turning on a finite chemical potential $\mu_{\text{nr}} \ll mc^2$, and thus introducing another scale in the problem. At zero temperature, all the states with energies less than or equal to the Fermi energy $E_F \equiv \mu_{\text{nr}}$ will be occupied, while all the other ones will be empty. This condition defines a closed surface in momentum space—the *Fermi surface*—separating occupied from unoccupied momentum states. In the case of free fermions, this is just a sphere with radius $p_F = \sqrt{2mE_F}$. In the

presence of interactions this sphere gets deformed and the Fermi surface can display more complicated shapes.

The theory of *Fermi liquids* is based on the assumption that the relevant collective excitations at zero temperature and finite density remain fermions. This makes intuitive sense in the weak coupling regime, where one can visualize these excitations as fermions that are “dressed” by interactions²¹ and excited slightly above the Fermi surface. Remarkably, though, this assumption proves to be empirically correct also in many strongly coupled systems. These fermionic excitations are often referred to as quasi-particles, to distinguish them from the actual fermions that make up the system.

I should point out that fermionic quasi-particles are not the only option. For example, fermions could pair up into bosons which then condense at low energies, as is the case in Helium 3 or in ordinary superconductors at low temperatures. We will briefly discuss this possibility at the end of this section. Some strongly coupled systems of fermions can even lack quasi-particle excitations altogether, in which case they are simply referred to as *non-Fermi liquids*. In what follows, we will restrict our attention to Fermi liquids, and derive an EFT description of their fermionic quasi-particles.

Quasi-particles that are slightly excited above the Fermi surface are fermions whose energy differs slightly from E_F . This should be reminiscent of our previous discussion of non-relativistic fermions, whose energy deviated slightly from mc^2 . The difference here is that particles on the Fermi surface have also finite momenta \vec{p}_F . To handle that, we will generalize the procedure followed in Sec. 4.1 and parametrize our non-relativistic fermionic field ψ as follows:

$$\psi(t, \vec{x}) \equiv e^{-iE_F t} \sum_{\vec{p}_F} e^{i\vec{p}_F \cdot \vec{x}} \psi_{\vec{p}_F}(t, \vec{x}). \quad (104)$$

Thus, by extracting a phase that depends not only on E_F but also on \vec{p}_F , we have effectively traded a single field ψ for a collection of fields $\psi_{\vec{p}_F}$ labeled by all possible momenta on the Fermi surface. By construction, these fields describe excitations such that $\partial_t \psi_{\vec{p}_F} \sim \epsilon \psi_{\vec{p}_F}$ with $\epsilon \ll E_F$. Therefore, the natural expansion parameter of our EFT is the ratio ϵ/E_F .

Let’s now turn to the business of writing down an effective action for the $\psi_{\vec{p}_F}$ ’s. For strongly interacting systems we cannot derive our EFT directly from the underlying theory, as we did for instance for the non-relativistic model discussed in Sec. 4.1. For this reason, we will rely instead on symmetries as our guideline. Starting from the transformation properties of ψ , it is easy to show that the fields $\psi_{\vec{p}_F}$ must transform under space-time symmetries

²¹By this I mean that interactions will change the dispersion relation of one-particle states near the Fermi surface.

as follows:

$$\text{time translations:} \quad \psi'_{\vec{p}_F}(t + c, \vec{x}) = e^{iE_F c} \psi_{\vec{p}_F}(t, \vec{x}) \quad (105a)$$

$$\text{spatial translations:} \quad \psi'_{\vec{p}_F}(t, \vec{x} + \vec{c}) = e^{-i\vec{p}_F \cdot \vec{c}} \psi_{\vec{p}_F}(t, \vec{x}) \quad (105b)$$

$$\text{spatial rotations:} \quad \psi'_{R \cdot \vec{p}_F}(t, R \cdot \vec{x}) = M(R) \psi_{\vec{p}_F}(t, \vec{x}) \quad (105c)$$

$$\text{Galilean boosts:} \quad \psi'_{\vec{p}_F - m\vec{v}}(t, \vec{x} - \vec{v}t) = \psi_{\vec{p}_F}(t, \vec{x}), \quad (105d)$$

where $M(R)$ is the spin 1/2 representation of the rotation R .

Exercise 4.4: use the algebra of the Galilei group, and in particular the commutators

$$[H, K_j] = iP_j, \quad [P_i, K_j] = iM\delta_{ij}, \quad [M, K_j] = 0, \quad (106)$$

to argue that, under a boost transformation such that $\vec{x}' = \vec{x} - \vec{v}t$, the Fermi momentum and energy change as follows:

$$\vec{p}'_F = \vec{p}_F - m\vec{v}, \quad E'_F = E_F - \vec{p}_F \cdot \vec{v} + \frac{mv^2}{2}. \quad (107)$$

Use then these results to derive the transformation rule (105d). Finally, convince yourself that Galilean boosts change the position of the Fermi surface in momentum space but do not alter its shape.

To the extent that relativistic corrections are negligible, one should also impose the accidental symmetry under spin rotations:

$$\text{spin rotations:} \quad \psi'_{\vec{p}_F}(t, \vec{x}) = U \psi_{\vec{p}_F}(t, \vec{x}). \quad (108)$$

Which of the above symmetries we eventually impose depends on the physical system under consideration. For instance, in the case of electrons in a metal, the underlying lattice spontaneously breaks boosts and rotations. These symmetries are non-linearly realized by the phonons, but appear to be explicitly broken if couplings with phonons are not taken into consideration. In what follows, we will consider the simplest possible scenario, namely one in which there are only fermions and there is no underlying lattice. In this case, our effective action should be invariant under all the transformations discussed above, and the Fermi surface remains a sphere.

Taken all together, the transformation rules (105) and (108) place strong constraints on the form of our effective action. In particular, invariance under time translations imply that each $\psi_{\vec{p}_F}$ should appear together with a $\psi_{\vec{p}_F}^\dagger$. Spatial translations require instead that the sum of all the labels of the $\psi_{\vec{p}_F}$'s minus those of the $\psi_{\vec{p}_F}^\dagger$'s add up to zero. Finally, spin rotations is ensured provided all spinor indices are contracted.

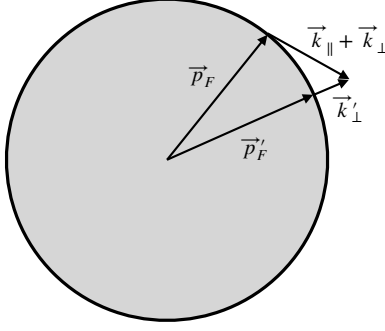


Figure 3: Momenta parallel and perpendicular to the Fermi surface.

The transformation properties of the $\psi_{\vec{p}_F}$'s under boosts require us to introduce covariant derivatives that are invariant under boosts. To this end, it is helpful to define at each point on the Fermi surface the normal vector

$$\vec{v}_F \equiv \frac{\partial E_F}{\partial \vec{p}_F}, \quad (109)$$

which is usually referred to as *Fermi velocity*. It is easy to see that this vector shifts as expected under boosts, *i.e.* $\vec{v}'_F = \vec{v}_F - \vec{v}$, and therefore that the differential operator $\mathcal{D} \equiv \partial_t + \vec{v}_F \cdot \vec{\nabla}$ is invariant. At lowest order in derivatives there is only one kinetic term that we can write down while preserving all the symmetries, which is

$$S_{\text{kin}} = \sum_{\vec{p}_F} \int dt d^3x \psi_{\vec{p}_F}^\dagger i\mathcal{D} \psi_{\vec{p}_F}, \quad (110)$$

where the spinor indices are understood to be contracted.

Now that we have determined the leading kinetic term, we are in a position to derive the power counting rules of our EFT. First, in order for the invariant differential operator to scale homogeneously we must have

$$\vec{v}_F \cdot \vec{\nabla} \sim \partial_t \sim \epsilon. \quad (111)$$

The kinetic term determines the scaling of only one component of spatial gradients—the one perpendicular to the Fermi surface. This is unlike all other models we have considered so far. To understand why that's the case, we can decompose the momentum \vec{k} of each Fourier mode $\psi_{\vec{p}_F}(t, \vec{k})$ along directions parallel and perpendicular to the Fermi surface at \vec{p}_F :

$$\vec{k} = \hat{v}_F(\hat{v}_F \cdot \vec{k}) + \hat{v}_F \times (\vec{k} \times \hat{v}_F) \equiv \vec{k}_\perp + \vec{k}_\parallel. \quad (112)$$

Then, as shown in Fig 3, one can always redefine the field label to absorb any component of the momentum parallel to the Fermi surface. To be more precise, we have the following relation between the Fourier modes of fields at different points on the Fermi sphere:

$$\psi_{\vec{p}_F}(t, \vec{k}_\parallel + \vec{k}_\perp) = \psi_{\vec{p}'_F}(t, \vec{k}'_\perp), \quad (113)$$

with

$$k'_\perp = |\vec{p}_F + \vec{k}_\parallel + \vec{k}_\perp|, \quad \hat{k}'_\perp = \hat{p}'_F = \frac{\vec{p}_F + \vec{k}_\parallel + \vec{k}_\perp}{|\vec{p}_F + \vec{k}_\parallel + \vec{k}_\perp|}. \quad (114)$$

We can therefore always assume that our fields $\psi_{\vec{p}_F}$ only carry momentum in the direction perpendicular to the Fermi surface. The integration measure and our fields must scale therefore like

$$dt d^3x \sim dt dx_\perp d^2x_\parallel \sim \epsilon^{-2}, \quad \psi_{\vec{p}_F} \sim \epsilon^{1/2}, \quad (115)$$

to ensure that the kinetic term (110) is of $\mathcal{O}(1)$ in the derivative expansion.

Armed with the power counting rules (111) and (115), we are now in a position to discuss interactions among quasi-particles. First, notice that there is in principle one relevant quadratic term that we could write down, namely

$$-\sum_{\vec{p}_F} \int dt d^3x \Delta \psi_{\vec{p}_F}^\dagger \psi_{\vec{p}_F} \sim \epsilon^{-1}, \quad (116)$$

where Δ is a parameter that must be independent of \vec{p}_F to preserve boost invariance. This term would modify the dispersion relation of quasi-particles by adding an energy gap, *i.e.* $\epsilon \rightarrow \vec{v}_F \cdot \vec{k} + \Delta$. In other words, this operator shifts the energy of all quasi-particles by a constant amount. Such shift can always be absorbed into a redefinition of the Fermi surface, and that is why the quadratic operator (116) can always be removed by re-phasing our fields: $\psi'_{\vec{p}_F} \equiv e^{i\Delta t} \psi_{\vec{p}_F}$.

More interesting is the fact that there exist some quartic marginal interactions that have the schematic form

$$S_{\text{int}} = \sum'_{\vec{p}_{F,i}} \int dt d^3x g \psi_{-\vec{p}_{F,3}}^\dagger \psi_{\vec{p}_{F,1}} \psi_{-\vec{p}_{F,4}}^\dagger \psi_{\vec{p}_{F,2}} \sim 1, \quad (117)$$

where the prime on the sum is a reminder of the constraint $\vec{p}_{F,1} + \vec{p}_{F,2} + \vec{p}_{F,3} + \vec{p}_{F,4} = 0$ imposed by translational invariance. Moreover, the coefficient g carries an index structure to ensure that all spinor indices are contracted in a $SU(2)$ -invariant way, and can depend on the labels $\vec{p}_{F,i}$ in the most general way consistent with rotational and boost invariance.

Let's first discuss the momentum dependence. The requirement of translational invariance is particularly stringent given that all the $\vec{p}_{F,i}$'s must simultaneously be on the Fermi surface. To see this, let's choose $\vec{p}_{F,1}$ and $\vec{p}_{F,2}$ along two arbitrary directions in momentum space. These two vectors will lie on a cone with axis in the direction of $\vec{p}_{F,1} + \vec{p}_{F,2}$, as shown in the right panel of Fig. 4. Conservation of momentum requires $\vec{p}_{F,3}$ and $\vec{p}_{F,4}$ to lie on a cone with the same axis but opposite orientation with respect to the origin. Up to a global rotation, this configuration is fully described by just two angles, *i.e.* the angle θ between $\vec{p}_{F,1}$ and $\vec{p}_{F,2}$,

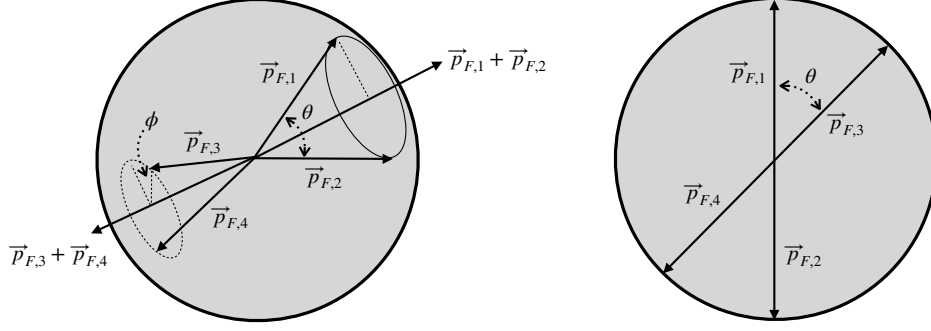


Figure 4: Interactions that conserve momentum on the Fermi surface.

and the angle ϕ between the planes spanned by $\vec{p}_{F,1}, \vec{p}_{F,2}$ and $\vec{p}_{F,3}, \vec{p}_{F,4}$. Following [43], we will refer to this configuration of momenta as the *zero-sound channel*.

This channel becomes ill-defined in the special case where $p_{F,1} = -\vec{p}_{F,2}$. Momentum conservation then implies that $p_{F,3} = -\vec{p}_{F,4}$. These “back-to-back” interactions are characterized by a single angle θ , as shown in the right panel of Fig 4. We will refer to this type of interactions as the *BCS channel*.

Exercise 4.5: Express the angles (θ, ϕ) of the zero-sound channel and the angle θ of the BCS channel in terms of the labels $\vec{p}_{F,i}$.

Let’s now restrict our attention to the BCS channel, for which the quartic interaction takes the form

$$S_{\text{BCS}} = \sum_{\vec{p}, \vec{p}'} \int dt d^3x g_{abcd}(\theta) \psi_{\vec{p}'}^{a\dagger} \psi_{\vec{p}}^b \psi_{-\vec{p}'}^{c\dagger} \psi_{-\vec{p}}^d, \quad (118)$$

with θ the angle between \vec{p} and \vec{p}' . It is easy to check that the only two independent index structures that are allowed by the $SU(2)$ spin symmetry are

$$g^{abcd}(\theta) = g_1(\theta) \delta^{ab} \delta^{cd} + g_2(\theta) \vec{\sigma}^{ab} \cdot \vec{\sigma}^{cd}. \quad (119)$$

The functions $g_i(\theta)$ can be further decomposed into Legendre polynomials, effectively introducing an infinite number of coupling constants $g_{i,\ell}$:

$$g_i(\theta) = \sum_{\ell} g_{i,\ell} P_{\ell}(\cos \theta). \quad (120)$$

Taken at face value, this seems to defy one of the standard lores about EFTs, namely that only a finite number of coupling constants should appear in the action at any given order in the expansion parameter. This is because the standard lore is based on the implicit assumption that the EFT contains a finite number of fields. In our case, we have effectively traded the ψ for an infinite number of fields $\psi_{\vec{p}}$. Notice however that predictivity is not lost: for any fixed

value of θ , the coupling $g_i(\theta)$ describes processes with arbitrary kinematical configurations of the “soft” momenta \vec{k} .

Exercise 4.6: Use the results derived in Exercise 4.5 to show that the angles in the zero-sound channel transform like $(\theta, \phi) \rightarrow (\theta, \pi + \phi)$ under a parity transformation. Then, show that the only possible tensor structures in the zero-sound channel are also of the form (119) if the theory is invariant under parity.

In summary, we have seen that having a finite density of fermions drastically changes the power counting rules compared to the vacuum case and allows for the existence of marginal interactions of the form (117). These interactions can be divided into two channels—zero-sound and BCS—and are in principle described by an infinite number of couplings. As usual, quantum corrections can make these interactions either relevant or irrelevant, depending on the anomalous dimension of the couplings. It turns out that attractive interactions become relevant whereas repulsive interactions become irrelevant [43].

Whether our couplings are positive or negative for any given system can only be determined by experiments or by matching onto an underlying UV theory (whose couplings have been determined by other experiments). On general grounds, though, we can expect the couplings that turn relevant to give rise to remarkable phenomena at low energies. For example, in Helium 3 couplings in the BCS channel with $\ell = 1$ are marginally relevant, prompting Helium atoms to pair up and condense at low temperatures, giving rise to superfluidity. In metals, instead, it is the BCS channel with $\ell = 0$ which grows relevant and is responsible for ordinary superconductivity. A detailed discussion of these phenomena goes well beyond the scope of these lectures, but the interested reader can find some useful resources listed at the end of this section. Overall, I hope to have conveyed the idea that, like any EFT, Fermi liquid theory relies on a particular choice of degrees of freedom, expansion parameter and symmetries. Lots of interesting physics follows from just these few assumptions.

4.4 Non-relativistic QED

We will now consider a slightly more involved example, namely non-relativistic QED (NRQED). This is an EFT where non-relativistic fermions experience long range electromagnetic interactions rather than the contact interaction in Eq. (100). Unlike the EFTs we have discussed until now, NRQED features more than one relevant kinematic region, and as a result power counting becomes considerably more complicated. NRQED can be used for instance to study bound states made of equal-mass constituents, with the goal of calculating observable quantities in a systematic expansion in powers of the velocity v .²² It is in fact easy to argue that v

²²In this section we will work again in units where $c = 1$.

must be small in a bound state by combining the virial theorem and the uncertainty principle:

$$\frac{mv^2}{2} \sim \frac{\alpha}{r} \lesssim \alpha mv \quad \implies \quad v \lesssim \alpha \ll 1. \quad (121)$$

It is worth stressing that the motivation behind NRQED is different compared to other EFTs we have encountered so far in these lectures. Here we are not integrating out particles that are too heavy to be produced at low-energies, nor are we preoccupied with modeling the low-energy consequences of strongly coupled or unknown physics in the UV. Instead, we are interested in using a well understood, perturbative theory like QED to calculate some observable quantities in an expansion in powers of the small velocity v . The most efficient way to achieve this is to switch to an EFT such as NRQED, where power counting in v is implemented directly at the level of the action.

Our starting point is the usual action for QED,

$$S = \int d^4x \left\{ \bar{\Psi} [i\gamma^\mu (\partial_\mu - ieA_\mu) - m] \Psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2\xi} (\partial_\mu A^\mu)^2 \right\}, \quad (122)$$

where we have already added explicitly a gauge fixing term. In what follows we will work in the Feynman gauge, where $\xi = 1$. At energies much smaller than m , we can rephase the fermion field and integrate out the anti-particles exactly as we did in Sec. 4.1. Denoting $A^\mu = (\Phi, \vec{A})$, we obtain the following tree-level effective action:

$$S_{\text{eff}}^{(0)} = \int dt d^3x \left\{ \psi^\dagger \left(i\partial_t - e\Phi + \frac{\vec{D}^2}{2m} + \frac{e}{2m} \vec{B} \cdot \vec{\sigma} \right) \psi - \frac{1}{2} \partial_\mu A_\nu \partial^\mu A^\nu + \mathcal{O}(1/m^2) \right\}, \quad (123)$$

with the 2-component spinor ψ defined in Eq. (99), the electric and magnetic fields are defined as usual in terms of the scalar and vector potentials [44], and $\vec{D} = \vec{\nabla} - ie\vec{A}$. This action is the analogue of the one in Eq. (100), with the contact interaction replaced by one mediated by the photon.

Exercise 4.7: Derive the tree-level effective action in Eq. (123).

Notice that the term linear in \vec{B} breaks the accidental $SU(2)$ spin symmetry and spatial rotations down to their diagonal subgroup. This term has the form $\vec{\mu} \cdot \vec{B}$, with $\vec{\mu} = 2\mu_B \vec{S}$ the density of magnetic moment, $\mu_B = q/(2mc)$ the Bohr magneton, and $\vec{S} = \frac{1}{2} \tilde{\Psi}_-^\dagger \vec{\sigma} \tilde{\Psi}_-$ the spin density: it is (minus) the familiar potential energy of a magnetic dipole density in a magnetic field.

Gauge invariance demands that each derivative in the action (123) appears together with a gauge potential. For this reason, it is no longer clear that expanding the effective action in inverse powers of m is equivalent to expanding in powers of $p/m = v$. For this to be the case, the potentials would also need to scale like v . Checking this requires in turn knowing

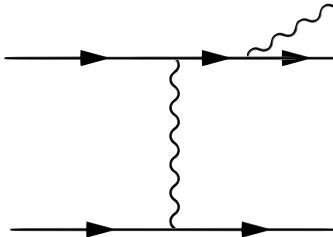


Figure 5: Sample tree-level diagram in QED.

how time and space derivatives act on the potentials, because then the scaling rule for the fields will follow as usual by demanding that the kinetic be a quantity of $\mathcal{O}(1)$. To this end, let's consider the diagram in Fig. 5. All external fermion lines have energy and momentum of order (mv^2, mv) . Thus, the energy and momentum exchanged in the scattering process, which is carried by the internal photon line, must generically be of order (mv^2, mv) as well. We will call this kinematic region the *potential* region. Consider now the photon emitted in this process: by conservation of energy it can at most have energy of $\mathcal{O}(mv^2)$. Since the corresponding external line must be on-shell, it must therefore carry energy and momentum of order (mv^2, mv^2) . We will refer to this kinematic region as the *ultrasoft* region.²³ This shows that the momentum carried by A^μ depends on the role that the photon line is playing in a Feynman diagram. Taken at face value, this result seems to prevent us from assigning a definite in v to the gauge potential. We will show in a moment how to address this problem and implement a well defined power counting in v .

It turns out that the potential and ultrasoft kinematic regions are not the end of the story. To see this, we'll need to first make a little detour and discuss one more EFT technique, known as the *method of regions*. We will first introduce it using a simple example, following [45]. Consider the following one-dimensional integral,

$$I(a, b) = \int_{-\infty}^{+\infty} dk \frac{1}{k^2 - a^2 + i\epsilon} \frac{1}{k^2 - b^2 + i\epsilon}, \quad (124)$$

whose structure is reminiscent of certain 1-loop integrals in relativistic field theories. We will be interested in a situation where the location of the two poles are widely separated, *i.e.* $a/b \ll 1$. Then, $I(a, b)$ can be calculated as an expansion in powers of the small parameter a/b . It seems reasonable to assume that the dominant contributions to this integral will come from the regions around the poles. Therefore we are going to approximate the full integral as a sum of two integrals centered around $k = \pm a$ and $k = \pm b$.

$$I(a, b) \simeq \left[\int_{k^2 \sim a^2} dk + \int_{k^2 \sim b^2} dk \right] \frac{1}{k^2 - a^2 + i\epsilon} \frac{1}{k^2 - b^2 + i\epsilon}. \quad (125)$$

²³This is also known as the *radiation* region.

In the first region we can expand the second term in the integrand in powers of k^2/b^2 , whereas in the second one we can expand the first term in powers of a^2/k^2 , to obtain

$$I(a, b) \simeq \sum_{n=0}^{\infty} \left[-\frac{1}{b^2} \int_{k^2 \sim a^2} dk \frac{1}{k^2 - a^2 + i\epsilon} \frac{k^{2n}}{b^{2n}} + \int_{k^2 \sim b^2} \frac{dk}{k^2} \frac{1}{k^2 - b^2 + i\epsilon} \frac{a^{2n}}{k^{2n}} \right]. \quad (126)$$

The exact values of these two integrals will depend in general on how we choose the regions around $k = \pm a$ and $k = \pm b$ —something we’ve been purposefully vague about until now. What we are going to do is extend these two regions to the entire real axis. At first this might seem like a very bad idea, since far away from the two poles the series expansions we have used do not seem justified. Moreover, the first integral is UV divergent for $n \neq 0$, whereas the original integral in Eq. (124) was not. We can easily address this, however, by using dimensional regularization. Then, the two integrals over the entire real axis can be calculated exactly to obtain

$$I(a, b) \simeq \frac{i\pi}{a} \frac{1}{b^2 - a^2} - \frac{i\pi}{b} \frac{1}{b^2 - a^2} = \frac{i\pi}{ab(a + b)}. \quad (127)$$

Something remarkable has happened: despite our seemingly questionable intermediate step, our final result turns out to be not just a good approximation, but in fact the *exact* value of $I(a, b)$! (This can be checked by calculating directly the integral in Eq. (124) using standard contour integration methods.) The reason why we were able to get away with extending the integrals to the entire real axis is that, in regions far away from each pole, all scales in the integrand can be neglected, and scale-less integrals vanish in dimensional regularization.

To summarize, we have shown that the integral in Eq. (124) can be calculated by identifying all the poles in the integrand, and then adding up the contributions that arise when expanding around each pole. This strategy is known as the method of regions, and it is a very helpful tool when applied to loop integrals in EFTs. In our simple example we have summed the series and obtained an exact result to give you some confidence that this method actually works. In an EFT, however, we are more interested in the fact that the integrals in (126) are explicitly organized in an expansion in the small parameter a/b .

In fact, if we think of the two factors in the integrand in Eq. (124) as the propagators of two different particles, the expression (126) lends itself to a very interesting interpretation. The first integral is equal to a loop made of one propagator $1/(k^2 - a^2 + i\epsilon)$, n static propagators $1/b^2$, and n field insertions k^2 . These are propagators and quadratic vertices that involve fields with momentum in the region $k^2 \sim a^2$. The second integral admits an analogous interpretation, with the relevant region being $k^2 \sim b^2$. The lesson is that, in order to organize $I(a, b)$ as an expansion in powers of a/b , we need to formally introduce a copy of each field for each kinematic region they are involved with, as well as additional interactions among all these copies.

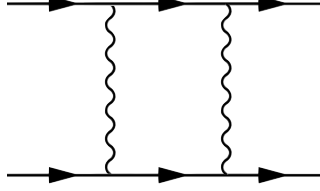


Figure 6: Sample one-loop diagram in QED.

Let's now return to our discussion of NRQED, and see how this idea can be implemented in practice. We have already identified two kinematic regions that play a role in NRQED (potential and ultrasoft), but it turns out that there is one more region we should consider. Take for instance the QED loop diagram shown in Fig. (6). We will work in the center-of-mass frame, and are interested in the situation where the external fermions are all non-relativistic. Up to an overall constant factor, this loop diagram is equal to the following integral:

$$\int \frac{d^4k}{(2\pi)^4} N \frac{1}{k_0^2 - \vec{k}^2 + i\epsilon} \frac{1}{(p_0 + k_0 - p'_0)^2 - (\vec{p} + \vec{k} - \vec{p}')^2 + i\epsilon} \frac{1}{(p_0 - k_0)^2 - (\vec{p} + \vec{k})^2 - m^2 + i\epsilon} \frac{1}{(p_0 + k_0)^2 - (\vec{p} + \vec{k})^2 - m^2 + i\epsilon}, \quad (128)$$

with $\pm\vec{p}$ ($\pm\vec{p}'$) the 3-momentum of the incoming (outgoing) particles, and N an analytic function of the momenta that won't play a role in our discussion. It is convenient to rewrite the integrand in terms of the kinetic energy of the fermions, $E = p^0 - m$, and to drop the factors of $i\epsilon$ to simplify the notation. The result is:

$$\int \frac{d^4k}{(2\pi)^4} N \frac{1}{k_0^2 - \vec{k}^2 + i\epsilon} \frac{1}{(E - E' + k^0)^2 - (\vec{p} + \vec{k} - \vec{p}')^2 + i\epsilon} \frac{1}{E^2 - 2Ek^0 - (\vec{p} + \vec{k})^2 + i\epsilon} \frac{1}{E^2 + 2Ek^0 - (\vec{p} + \vec{k})^2 + i\epsilon}. \quad (129)$$

According to our previous discussion, the location of any pole in the integrand corresponds to a relevant kinematic region. Given that $E, E' \sim mv^2$ and $p, p' \sim mv$, the second propagator has a pole for $(k^0, \vec{k}) \sim (mv, mv)$. We will call this the *soft* region. Notice that when the internal photon lines carry a soft 4-momentum, the internal fermion lines must do so as well by conservation of energy. The need for a soft region arises only when quantum loops are considered. In fact, it doesn't play any role in non-relativistic General Relativity (NRGR) [46, 47], where one is interested in *classical* gravitational bound states.

It turns out that the soft, potential, and ultrasoft regions are sufficient to systematically implement a velocity expansion in NRQED. To this end, we are going to decompose the fields

ψ and A^μ as follows:

$$\psi = \psi_S + \psi_P \quad (130a)$$

$$A^\mu = A_S^\mu + A_P^\mu + A_U^\mu, \quad (130b)$$

were the subscripts denote the corresponding regions. Notice that there is no need to introduce an ultrasoft fermion. Interactions between the various fields introduced above must preserve energy and momentum with an accuracy of $\mathcal{O}(mv^2)$. At this stage it is not obvious how to implement this, given the fields in the soft and potential region have energy and/or momentum of $\mathcal{O}(mv)$. Moreover, it is not clear how one should power count interactions between fields in different regions: how is the integration measure d^4x supposed to scale? To address these problems, we can rephase our soft and potential fields to extract a phase of $\mathcal{O}(mv)$:

$$\psi_S(x) = \sum_p e^{ip \cdot x} \psi_{S,p}(x) \quad (131a)$$

$$A_S^\mu(x) = \sum_p e^{ip \cdot x} A_{S,p}^\mu(x) \quad (131b)$$

$$\psi_P(x) = \sum_{\vec{p}} e^{i\vec{p} \cdot \vec{x}} \psi_{P,\vec{p}}(x) \quad (131c)$$

$$A_P^\mu(x) = \sum_{\vec{p}} e^{i\vec{p} \cdot \vec{x}} A_{P,\vec{p}}^\mu(x). \quad (131d)$$

Note that we have extracted a four-momentum from the soft fields, and a three-momentum from the potential fields. The procedure outlined above is not that dissimilar from what we did in Eq. (104) when we expressed the fermion field ψ as a sum of many fields $\psi_{\vec{p}_F}$, except that now we had to take into account the existence of multiple regions. Incidentally, the extraction of a soft phase from the soft and potential fields can also be thought of as some sort of multipole expansion [48, 49]. This makes intuitive sense, since ultrasoft fields can only probe the long wavelength behavior of soft and potential fields.

By plugging the expressions (130) and (131) in the the effective action (123) we can derive the dominant kinetic terms for the fields $\psi_{S,p}, \psi_{P,\vec{p}}, A_{S,p}^\mu, A_{P,\vec{p}}^\mu$, and A_U^μ , as well as their interactions. Because these fields only contain Fourier modes with residual energies and momenta of $\mathcal{O}(mv^2)$, conservation of four-momentum up to $\mathcal{O}(mv^2)$ is automatically guaranteed. Moreover, the appropriate scaling for the integration measure is now clearly $d^4x \sim 1/v^8$. It is then easy to figure out how the various fields must scale with v , and to calculate their propagators. The results are summarized in Table 1, with soft (*i.e.* label) momenta denoted with p and ultrasoft (*i.e.* residual) momenta with k . Notice in particular that the ultrasoft gauge fields scales like partial derivatives—they are both of $\mathcal{O}(mv^2)$ —and therefore can be used to build covariant derivatives that scale homogeneously: $D_\mu = \partial_\mu - ieA_\mu^U$

Table 1: Fields in NRQED and their properties.

Region	Field	Scaling	Propagator
Soft	$\psi_{S,p}$	$v^{3/2}$	$i/(p^0 - i\epsilon)$
	$A_{S,p}^\mu$	v	$-i\eta^{\mu\nu}/(p^2 - i\epsilon)$
Potential	$\psi_{P,\vec{p}}$	$v^{3/2}$	$i/(k^0 - \vec{p}^2/2m + i\epsilon)$
	$A_{P,\vec{p}}^\mu$	$v^{3/2}$	$-i\eta^{\mu\nu}/(\vec{p}^2 - i\epsilon)$
Ultrasoft	A_U^μ	v^2	$-i\eta^{\mu\nu}/(k^2 - i\epsilon)$

Exercise 4.8: Derive the results summarized in Table 1.

A detailed description of all Feynman rules and practical applications of NRQED is beyond the scope of this section, whose main goal was to introduce the method of regions. You can find more details in the references provided in the following section.

4.5 Additional resources

The standard reference for the EFT of Fermi liquids are Polchinski's TASI lectures [50]. For a more exhaustive treatment of fermions at finite density based on the renormalization group, see [43]. A pedagogical early review of Helium 3 superfluidity is [51]. The use of NRQED to study bound states was first advocated in [52]. Most of the intricacies of power counting in non-relativistic gauge theories have been sorted out in the context of non-relativistic QCD (NRQCD). For more details on NRQCD, see for instance [15, 45]. You can also find a calculation of the Lamb shift based on NRQED in [15].

Acknowledgments

I would like to thank the organizers of the 2nd Joint ICTP-Trieste/ICTP-SAIFR school on Particle Physics (Enrico Bertuzzo, Joan Elias Miró, Rogerio Rosenfeld, and Giovanni Villadoro) for inviting me to give these lectures. In preparing these notes, I have benefited greatly from reading the many excellent reviews on the subject, most of which I have cited where appropriate. I would also like to thank Angelo Esposito, Garrett Goon, Alberto Nicolis, Rachel Rosen, Ira Rothstein, and Adam Solomon for helpful discussions and comments on earlier drafts of these notes. This work was supported in part by the National Science Foundation under Grant No. PHY-1915611.

References

- [1] Mark Srednicki. *Quantum Field Theory*. Cambridge University Press, 2007.
- [2] K.G. Wilson. The renormalization group and critical phenomena. *Rev.Mod.Phys.*, 55:583–600, 1983.
- [3] Howard Georgi. On-shell effective field theory. *Nucl. Phys. B*, 361:339–350, 1991.
- [4] Steven Weinberg. *The quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 1995.
- [5] Gerard 't Hooft. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. *NATO Adv.Study Inst.Ser.B Phys.*, 59:135, 1980.
- [6] Garrett Goon, Kurt Hinterbichler, Austin Joyce, and Mark Trodden. Aspects of Galileon Non-Renormalization. *JHEP*, 11:100, 2016, 1606.02295.
- [7] Allan Adams, Nima Arkani-Hamed, Sergei Dubovsky, Alberto Nicolis, and Riccardo Rattazzi. Causality, analyticity and an IR obstruction to UV completion. *JHEP*, 0610:014, 2006, hep-th/0602178.
- [8] Cumrun Vafa. The String landscape and the swampland. 2005, hep-th/0509212.
- [9] Steven Weinberg. Anthropic Bound on the Cosmological Constant. *Phys.Rev.Lett.*, 59:2607, 1987.
- [10] Lasma Alberte, Paolo Creminelli, Andrei Khmelnitsky, David Pirtskhalava, and Enrico Trincherini. Relaxing the Cosmological Constant: a Proof of Concept. *JHEP*, 12:022, 2016, 1608.05715.
- [11] Peter W. Graham, David E. Kaplan, and Surjeet Rajendran. Relaxation of the Cosmological Constant. *Phys. Rev. D*, 100(1):015048, 2019, 1902.06793.

- [12] H. Georgi. Effective field theory. *Ann.Rev.Nucl.Part.Sci.*, 43:209–252, 1993.
- [13] Aneesh V. Manohar. Effective field theories. 1996, hep-ph/9606222.
- [14] Antonio Pich. Effective field theory. 1998, hep-ph/9806303.
- [15] Ira Z. Rothstein. TASI lectures on effective field theories. 2003, hep-ph/0308266.
- [16] David B. Kaplan. Five lectures on effective field theory. 2005, nucl-th/0510023.
- [17] C.P. Burgess. Introduction to Effective Field Theory. *Ann.Rev.Nucl.Part.Sci.*, 57:329–362, 2007, hep-th/0701053.
- [18] Witold Skiba. TASI Lectures on Effective Field Theory and Precision Electroweak Measurements. 2010, 1006.2142.
- [19] Ben Gripaios. Lectures on Effective Field Theory. 2015, 1506.05039.
- [20] Alexey A. Petrov and Andrew E. Blechman. *Effective Field Theories*. WSP, 2016.
- [21] Aneesh V. Manohar. Introduction to Effective Field Theories. In *Les Houches summer school: EFT in Particle Physics and Cosmology Les Houches, Chamonix Valley, France, July 3-28, 2017*, 2018, 1804.05863.
- [22] Timothy Cohen. As Scales Become Separated: Lectures on Effective Field Theory. *PoS, TASI2018:011*, 2019, 1903.03622.
- [23] Michael Dine. Naturalness Under Stress. *Ann. Rev. Nucl. Part. Sci.*, 65:43–62, 2015, 1501.01035.
- [24] Gian Francesco Giudice. The Dawn of the Post-Naturalness Era. 2017, 1710.07663.
- [25] Steven Weinberg. The Cosmological Constant Problem. *Rev.Mod.Phys.*, 61:1–23, 1989.
- [26] Steven Weinberg. Why the renormalization group is a good thing. In Alan H. Guth, Kerson Huang, and Robert L. Jaffe, editors, *Asymptotic Realms of Physics, Essays in Honor of Francis E. Low*, page 1. MIT Press, 1983.
- [27] Stefan Scherer. Introduction to chiral perturbation theory. *Adv. Nucl. Phys.*, 27:277, 2003, hep-ph/0210398.
- [28] Jr. Callan, C. G., S. R. Coleman, J. Wess, and B. Zumino. Structure of phenomenological Lagrangians. 2. *Phys.Rev.*, 177:2247–2250, 1969.
- [29] Dmitri V. Volkov. Phenomenological Lagrangians. *Fiz. Elem. Chast. Atom. Yadra*, 4:3–41, 1973.
- [30] V. I. Ogievetsky. Nonlinear realizations of internal and space-time symmetries. In *X-th winter school of theoretical physics in Karpacz, Poland*, 1974.
- [31] Steven Weinberg. *The quantum theory of fields. Vol. 2: Modern applications*. Cambridge University Press, 1996.

- [32] Sean M. Carroll. *Spacetime and geometry: An introduction to general relativity*. Benjamin Cummings, 2004.
- [33] Edward Witten. Global Aspects of Current Algebra. *Nucl.Phys.*, B223:422–432, 1983.
- [34] Garrett Goon, Kurt Hinterbichler, Austin Joyce, and Mark Trodden. Galileons as Wess-Zumino Terms. *JHEP*, 1206:004, 2012, 1203.3191.
- [35] Jeffrey A. Harvey. TASI 2003 lectures on anomalies. 2005, hep-th/0509097.
- [36] John Preskill. Gauge anomalies in an effective field theory. *Annals Phys.*, 210:323–379, 1991.
- [37] G. Ecker. Chiral perturbation theory. *Prog. Part. Nucl. Phys.*, 35:1–80, 1995, hep-ph/9501357.
- [38] A. Pich. Chiral perturbation theory. *Rept. Prog. Phys.*, 58:563–610, 1995, hep-ph/9502366.
- [39] C. P. Burgess. Goldstone and pseudoGoldstone bosons in nuclear, particle and condensed matter physics. *Phys. Rept.*, 330:193–261, 2000, hep-th/9808176.
- [40] Antonio Pich. Effective Field Theory with Nambu-Goldstone Modes. In *Les Houches summer school: EFT in Particle Physics and Cosmology*, 4 2018, 1804.05664.
- [41] Adel Bilal. Lectures on Anomalies. 2 2008, 0802.0634.
- [42] A. Zee. *Quantum field theory in a nutshell*. Princeton University Press, 2010.
- [43] R. Shankar. Renormalization group approach to interacting fermions. *Rev. Mod. Phys.*, 66:129–192, 1994.
- [44] John David Jackson. *Classical Electrodynamics*. Wiley, 1998.
- [45] Harald W. Griesshammer. Threshold expansion and dimensionally regularized NRQCD. *Phys.Rev.*, D58:094027, 1998, hep-ph/9712467.
- [46] Walter D. Goldberger. Les Houches lectures on effective field theories and gravitational radiation. 2007, hep-ph/0701129.
- [47] Rafael A. Porto. The effective field theorist’s approach to gravitational dynamics. *Phys. Rept.*, 633:1–104, 2016, 1601.04914.
- [48] Patrick Labelle. Effective field theories for QED bound states: Extending nonrelativistic QED to study retardation effects. *Phys. Rev. D*, 58:093013, 1998, hep-ph/9608491.
- [49] Benjamin Grinstein and Ira Z. Rothstein. Effective field theory and matching in nonrelativistic gauge theories. *Phys. Rev. D*, 57:78–82, 1998, hep-ph/9703298.
- [50] Joseph Polchinski. Effective field theory and the Fermi surface. 1992, hep-th/9210046.

- [51] Anthony J. Leggett. A theoretical description of the new phases of liquid He-3. *Rev. Mod. Phys.*, 47:331–414, 1975. [Erratum: *Rev. Mod. Phys.*48,357(1976)].
- [52] W.E. Caswell and G.P. Lepage. Effective Lagrangians for Bound State Problems in QED, QCD, and Other Field Theories. *Phys. Lett. B*, 167:437–442, 1986.