An Introduction to Effective Field Theory

Thinking Effectively About Hierarchies of Scale

Preface

It is an everyday fact of life that Nature comes to us with a variety of scales: from quarks, nuclei and atoms through planets, stars and galaxies up to the overall Universal large-scale structure. Science progresses because we can understand each of these on its own terms, and need not understand all scales at once. This is possible because of a basic fact of Nature: most of the details of small distance physics are irrelevant for the description of longer-distance phenomena.

Our description of Nature's laws use quantum field theories, which share this property that short distances mostly decouple from larger ones. Effective Field Theories (EFTs) are the tools developed over the years to show why it does. These tools have immense practical value: knowing which scales are important and why the rest decouple allows hierarchies of scale to be used to simplify the description of many systems. This book provides an introduction to these tools, and to emphasize their great generality illustrates them using applications from many parts of physics: relativistic and nonrelativistic; few-body and many-body.

The book is broadly appropriate for an introductory graduate course, though some topics could be done in an upper-level course for advanced undergraduates. It should interest physicists interested in learning these techniques for practical purposes as well as those who enjoy the beauty of the unified picture of physics that emerges.

It is to emphasize this unity that a broad selection of applications is examined, although this also means no one topic is explored in as much depth as it deserves. The book's goal is to engage the reader's interest, but then to redirect to the appropriate literature for more details. To this end references in the main text are provided mostly just for the earliest papers (that I could find) on a given topic, with a broader – probably more useful – list of textbooks, reviews and other sources provided in the bibliography. There will be inevitable gems about which I am unaware or have forgotten to mention, and I apologize in advance to both their authors and to you the reader for their omission.

An introductory understanding of quantum and classical field theory is assumed, for which an appendix provides a basic summary of the main features. To reconcile the needs of readers with differing backgrounds — from complete newbies through to experts seeking applications outside their own areas — sections are included requiring differing amounts of sophistication. The background material in the appendices is also meant to help smooth out the transitions between these different levels of difficulty.

The various gradations of sophistication are flagged using the suits of playing cards: ♢, ♡, ♠ and ♣ in the titles of the chapter sections. The flag ♢ indicates good value and labels sections that carry key ideas that should not be missed by any student of effective theories. ♡ flags sections containing material common to most quantum field theory classes, whose familiarity may warm a reader's heart but can be skipped by aficianados in a hurry. The symbol ♠ indicates a section which may require a bit more digging for new students to digest, but which is reasonably self-contained and worth a bit of spadework. Finally,

readers wishing to beat their heads against sections containing more challenging topics should seek out those marked with ♣.

The lion's share of the book is aimed at applications, since this most effectively brings out both the utility and the unity of the approach. The examples also provide a pedagogical framework for introducing some specific techniques. Since many of these applications are independent of one another, a course can be built by starting with Part I's introductory material and picking and choosing amongst the later sections that are of most interest.

Acknowledgements

This book draws heavily on the insight and goodwill of many people: in particular my teachers of quantum and classical field theory – Bryce De Witt, Willy Fischler, Joe Polchinski and especially Steven Weinberg – who shaped the way I think about this subject.

Special thanks go to Fernando Quevedo for a life-long collaboration on these subjects and his comments over the years on many of the topics discussed herein.

I owe a debt to Patrick Labelle, Sung-Sik Lee, Alexander Penin and Ira Rothstein for clarifying issues to do with nonrelativistic EFTs; to John Donoghue for many insights on gravitational physics; to Thomas Becher for catching errors in early versions of the text; to Jim Cline for a better understanding of the practical implications of Goldstone boson infrared effects; to Claudia de Rham, Luis Lehner, Adam Solomon, Andrew Tolley and Mark Trodden for helping better understand applications to time-dependent systems; to Subodh Patil and Michael Horbatsch for helping unravel multiple scales in scalar cosmology; to Mike Trott for help understanding the subtleties of power-counting and SMEFT; to Peter Adshead, Richard Holman, Greg Kaplanek, Louis Leblond, Jerome Martin, Sarah Shandera, Gianmassimo Tasinato, Vincent Vennin and Richard Woodard for understanding EFTs in de Sitter space and their relation to open systems, and to Ross Diener, Peter Hayman, Doug Hoover, Leo van Nierop, Duncan Odell, Ryan Plestid, Markus Rummel, Matt Williams, and Laszlo Zalavari for helping clarify how EFTs work for massive first-quantized sources.

Collaborators and students too numerous to name have continued to help deepen my understanding in the course of many conversations about physics.

CERN, ICTP, KITP Santa Barbara and the Institute Henri Poincaré have at various times provided me with pleasant environs in which to focus undivided time on writing, and with stimulating discussions when taking a break from it. The book would not have been finished without them. The same is true of McMaster University and Perimeter Institute, whose flexible work environments allowed me to take on this project in the first place.

Heaven holds a special place for Simon Capelin and his fellow editors, both for encouraging the development of this book and for their enormous patience in awaiting it.

Most importantly, I am grateful to my late parents for their gift of an early interest in science, and to my immediate family (Caroline, Andrew, Ian, Matthew and Michael) for their continuing support and tolerance of time taken from them for physics.

Contents

Illustrations

Tables

PART I

THEORETICAL FRAMEWORK

About Part I

This first part of the book sets up the basic framework of effective field theories (EFTs), developing along the way the main tools and formalism that is used throughout the remainder of the text. An effort is made to discuss topics that are sometimes left out in reviews of EFT methods, such as how to work with time-dependent backgrounds or in the presence of boundaries. This part of the book is meant to be relatively self-contained, and so can be studied on its own given limited time.

Discussions of formalism can easily descend into obscurity if not done with concrete questions in mind. To keep things focussed the first chapter here introduces a toy model in which most of the conceptual issues arise in a simple way. As each subsequent chapter in Part I introduces a new concept, its formal treatment is accompanied by a short illustrative discussion about how that particular issue arises within the toy model. Hopefully by the end of Part I the reader will be familiar with the main EFT tools, and will know the toy model inside and out.

The book's remaining major parts then work through practical examples of EFT reasoning throughout physics. Because the formalism is largely handled in Part I, the focus of the rest of the book is both on illustrating some of the techniques introduced in Part I, and on physical insights that emerge when these tools are used to study specific problems.

The later parts are grouped into four categories that share similar features:

- Parts II studies relativistic applications, studying first examples where both the low- and high-energy parts of the theory are well-understood and then switching to problems for which the high-energy sector is either unknown or is known but difficult to use precisely (such as if it involves strong interactions).
- Part III switches to non-relativistic applications, such as to slowly moving systems of a small number of particles, like atoms, or systems involving lots of particles for which only gross features like the centre-of-mass motion are of interest, like for planetary orbits in the solar system.
- Part IV then examines many-body and open systems, for which many particles are involved and more degrees of freedom appear in the coarse-grained theory. Part IV starts with calculations for which dissipative effects are chosen to be negligible, but closes with a discussion of open systems for which dissipation and decoherence can be important.

1 Decoupling and hierarchies of scale

The world around us contains a cornucopia of length scales, ranging (at the time of writing) down to quarks and leptons at the smallest and up to the universe as a whole at the largest, with qualitatively new kinds of structures — nuclei, atoms, molecules, cells, organisms, mountains, asteroids, planets, stars, galaxies, voids, and so on — seemingly arising at every few decades of scales in between. So it is remarkable that all of this diversity seems to be described in all of its complexity by a few simple laws.

How can this be possible? Even given that the simple laws exist, why should it be possible to winkle out an understanding of what goes on at one scale without having to understand everything all at once? The answer seems to be a very deep property of nature called *decoupling*, which states that most (but not all) of the details of very small-distance phenomena tend to be largely irrelevant for the description of much larger systems. For example, not much need be known about the detailed properties of nuclei (apart from their mass and electrical charge, and perhaps a few of their multipole moments) in order to understand in detail the properties of electronic energy levels in atoms.

Decoupling is a very good thing, since it means that the onion of knowledge can be peeled one layer at a time: our initial ignorance of nuclei need not impede the unravelling of atomic physics, just as ignorance about atoms does not stop working out the laws describing the motion of larger things, like the behaviour of fluids or motion of the moon.

It happens that this property of decoupling is also shared by the mathematics used to describe the laws of nature [1]. Since nowadays this description involves quantum field theories, it is gratifying that these theories as a group tend to predict that short distances generically decouple from long distances, in much the same way as happens in nature.

This book describes the way this happens in detail, with two main purposes in mind. One purpose is to display decoupling for its own sake since this is satisfying in its own right, and leads to deep insights into what precisely is being accomplished when writing down physical laws. But the second purpose is very practical; the simplicity offered by a timely exploitation of decoupling can often be the difference between being able to solve a problem or not. When exploring the consequences of a particular theory for short distance physics it is obviously useful to be able to identify efficiently those observables that are most sensitive to the theory's details and those from which they decouple. As a consequence the mathematical tools — *effective field theories* — for exploiting decoupling have become ubiquitous in some areas of theoretical physics, and are likely to become more common in many more.

The purpose of the rest of chapter 1 is twofold. One goal is to sketch the broad outlines of decoupling, effective lagrangians and the physical reason why they work, all in one place.

The second aim is to provide a toy model that can be used as a concrete example as the formalism built on decoupling is fleshed out in more detail in subsequent chapters.

1.1 An illustrative toy model ◇

The first step is to set up a simple concrete model to illustrate the main ideas. To be of interest this model must possess two kinds of particles, one of which is much heavier than the other, and these particles must interact in a simple yet nontrivial way. Our focus is on the interactions of the two particles, with a view towards showing precisely how the heavy particle decouples from the interactions of the light particle at low energies.

To this end consider a complex scalar field, ϕ, with action[1]

$$S := -\int \mathrm{d}^4x \left[\partial_\mu \phi^* \partial^\mu \phi + V(\phi^*\phi)\right], \tag{1.1}$$

whose self-interactions are described by a simple quartic potential,

$$V(\phi^*\phi) = \frac{\lambda}{4}\left(\phi^*\phi - v^2\right)^2 , \tag{1.2}$$

where λ and v^2 are positive real constants. The shape of this potential is shown in Fig. 1.1

1.1.1 Semiclassical spectrum

The simplest regime in which to explore the model's predictions is when $\lambda \ll 1$ and both v and $|\phi|$ are $O\left(\lambda^{-1/2}\right)$. This regime is simple because it is one for which the semiclassical approximation provides an accurate description. (The relevance of the semiclassical limit in this regime can be seen by writing $\phi := \varphi/\lambda^{1/2}$ and $v := \mu/\lambda^{1/2}$ with φ and μ held fixed as $\lambda \to 0$. In this case the action depends on λ only through an overall factor: $S[\phi, v, \lambda] = (1/\lambda)S[\varphi, \mu]$. This is significant because the action appears in observables only in the combination $S/\hbar$, and so the small-λ limit is equivalent to the small-$\hbar$ (classical) limit.[2])

In the classical limit the ground state of this system is the field configuration that minimizes the classical energy,

$$E = \int \mathrm{d}^3x \left[\partial_t \phi^* \partial_t \phi + \nabla\phi^* \cdot \nabla\phi + V(\phi^*\phi)\right]. \tag{1.3}$$

Since this is the sum of positive terms it is minimized by setting each to zero; the classical ground state is any constant configuration (so $\partial_t\phi = \nabla\phi = 0$), with $|\phi| = v$ (so $V = 0$).

[1] Although this book presupposes some familiarity with quantum field theory, see Appendix C for a compressed summary of some of the relevant ideas and notation used throughout. Unless specifically stated otherwise, units are adopted for which $\hbar = c = 1$, so that time ~ length and energy ~ mass ~ 1/length, as described in more detail in Appendix A.

[2] The connection between small coupling and the semi-classical limit is explored more fully once power-counting techniques are discussed in §3.

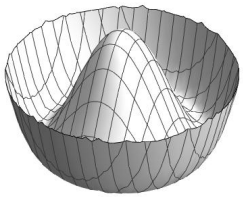


Fig. 1.1 The toy model's potential $V(\phi_R, \phi_I)$, showing its sombrero shape and the circular line of minima at $|\phi| = v$.

In the semi-classical regime particle states are obtained by expanding the action about the classical vacuum, $\phi = v + \tilde{\phi}$,

$$S = -\int \mathrm{d}^4x \, \left\{ \partial_\mu \tilde{\phi}^* \partial^\mu \tilde{\phi} + \frac{\lambda}{4} \left[v(\tilde{\phi} + \tilde{\phi}^*) + \tilde{\phi}^* \tilde{\phi} \right]^2 \right\} , \tag{1.4}$$

and keeping the leading (quadratic) order in the quantum fluctuation $\tilde{\phi}$. In terms of the field's real and imaginary parts, $\tilde{\phi} = \frac{1}{\sqrt{2}}(\tilde{\phi}_R + \mathrm{i}\tilde{\phi}_I)$, the leading term in the expansion of S is

$$S_0 = -\frac{1}{2} \int \mathrm{d}^4x \, \left[\partial_\mu \tilde{\phi}_R \, \partial^\mu \tilde{\phi}_R + \partial_\mu \tilde{\phi}_I \, \partial^\mu \tilde{\phi}_I + \lambda v^2 \, \tilde{\phi}_R^2 \right] . \tag{1.5}$$

The standard form (see §C.3.1) for the action of a free, real scalar field of mass m is proportional to $\partial_\mu \psi \, \partial^\mu \psi + m^2 \psi^2$, and so comparing with eq. (1.5) shows $\tilde{\phi}_R$ represents a particle with mass $m_R^2 = \lambda v^2$ while $\tilde{\phi}_I$ represents a particle with mass $m_I^2 = 0$. These are the heavy and light particles whose masses provide a hierarchy of scales.

1.1.2 Scattering

For small λ the interactions amongst these particles are well-described in perturbation theory, by writing $S = S_0 + S_{\mathrm{int}}$ and perturbing in the interactions

$$S_{\mathrm{int}} = -\int \mathrm{d}^4x \, \left[\frac{\lambda v}{2\sqrt{2}} \, \tilde{\phi}_R \left(\tilde{\phi}_R^2 + \tilde{\phi}_I^2 \right) + \frac{\lambda}{16} \left(\tilde{\phi}_R^2 + \tilde{\phi}_I^2 \right)^2 \right] . \tag{1.6}$$

Using this interaction a straightforward calculation – for a summary of the steps involved see Appendix B – gives any desired scattering amplitude order-by-order in λ. Since small λ describes a semiclassical limit (because it appears systematically together with $\hbar$ in $S/\hbar$, as argued above), the leading contribution turns out to come from evaluating Feynman graphs with no loops[3] (*i.e.* tree graphs).

[3] A connected graph with no loops (or a 'tree' graph) is one which can be broken into two disconnected parts by cutting any internal line. Precisely how to count the number of loops and why this is related to powers of the small coupling λ is the topic of §3.

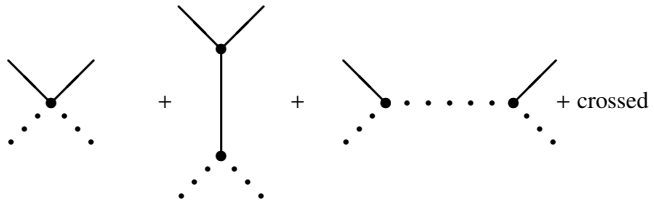


Fig. 1.2 The tree graphs that dominate $\tilde{\phi}_R\,\tilde{\phi}_I$ scattering. Solid (dotted) lines represent $\tilde{\phi}_R$ ($\tilde{\phi}_I$), and 'crossed' graphs are those with external lines interchanged relative to those displayed.

Consider the reaction $\tilde{\phi}_R(p) + \tilde{\phi}_I(q) \to \tilde{\phi}_R(p') + \tilde{\phi}_I(q')$, where $p^\mu = \{p^0, \mathbf{p}\}$ and $q^\mu = \{q^0, \mathbf{q}\}$ respectively denote the 4-momenta of the initial $\tilde{\phi}_R$ and $\tilde{\phi}_I$ particle, while p'^μ and q'^μ are 4-momenta of the final $\tilde{\phi}_R$ and $\tilde{\phi}_I$ states. The Feynman graphs of Fig. 1.2 give a scattering amplitude proportional to[4] $\mathcal{A}_{RI\to RI}\delta^4(p+q-p'-q')$, where the Dirac delta function, $\delta^4(p+q-p'-q')$, expresses energy-momentum conservation, and

$$\begin{aligned}\mathcal{A}_{RI\to RI} &= 4\mathrm{i}\left(-\frac{\lambda}{8}\right) + \left(\frac{\mathrm{i}^2}{2}\right)\left(-\frac{\lambda v}{2\sqrt{2}}\right)^2\left[\frac{24(-\mathrm{i})}{(p-p')^2+m_R^2} + \frac{8(-\mathrm{i})}{(p+q)^2} + \frac{8(-\mathrm{i})}{(p-q')^2}\right] \\ &= -\frac{\mathrm{i}\lambda}{2} + \frac{\mathrm{i}(\lambda v)^2}{2m_R^2}\left[\frac{3}{1-2q\cdot q'/m_R^2} - \frac{1}{1-2p\cdot q/m_R^2} - \frac{1}{1+2p\cdot q'/m_R^2}\right]. \quad (1.7)\end{aligned}$$

Here the factors like 4, 24 and 8 in front of various terms count the combinatorics of how many ways each particular graph can contribute to the amplitude. The second line uses energy-momentum conservation, $(p-p')^\mu = (q'-q)^\mu$, as well as the kinematic conditions $p^2 = -(p^0)^2 + \mathbf{p}^2 = -m_R^2$ and $(q')^2 = q^2 = -(q^0)^2 + \mathbf{q}^2 = 0$, as appropriate for relativistic particles whose energy and momenta are related by $E = p^0 = \sqrt{\mathbf{p}^2 + m^2}$.

Notice that the terms involving the square bracket arise at the same order in λ as the first term, despite nominally involving two powers of $S_{\rm int}$ rather than one (provided that the square bracket itself is order unity). To see this keep in mind $m_R^2 = \lambda v^2$ so that $(\lambda v/m_R)^2 = \lambda$.

For future purposes it is useful also to have the corresponding result for the reaction $\tilde{\phi}_I(p) + \tilde{\phi}_I(q) \to \tilde{\phi}_I(p') + \tilde{\phi}_I(q')$. A similar calculation, using instead the Feynman graphs of Fig. 1.3, gives the scattering amplitude

$$\begin{aligned}\mathcal{A}_{II\to II} &= 24\mathrm{i}\left(-\frac{\lambda}{16}\right) + 8\left(\frac{\mathrm{i}^2}{2}\right)\left(-\frac{\lambda v}{2\sqrt{2}}\right)^2\left[\frac{-\mathrm{i}}{(p+q)^2+m_R^2} + \frac{-\mathrm{i}}{(p-p')^2+m_R^2} + \frac{-\mathrm{i}}{(p-q')^2+m_R^2}\right] \\ &= -\frac{3\mathrm{i}\lambda}{2} + \frac{\mathrm{i}(\lambda v)^2}{2m_R^2}\left[\frac{1}{1+2p\cdot q/m_R^2} + \frac{1}{1-2q\cdot q'/m_R^2} + \frac{1}{1-2p\cdot q'/m_R^2}\right]. \quad (1.8)\end{aligned}$$

[4] See Exercise 1.1 and Appendix B for the proportionality factors.

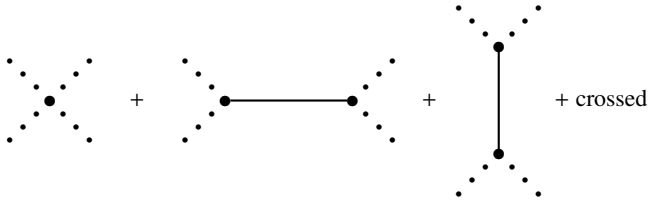


Fig. 1.3 The tree graphs that dominate the $\tilde{\phi}_I\,\tilde{\phi}_I$ scattering amplitude. Solid (dotted) lines represent $\tilde{\phi}_R$ and $\tilde{\phi}_I$ particles.

1.1.3 The low-energy limit

For the present purposes it is the low-energy regime that is of most interest: when the centre-of-mass kinetic energy and momentum transfers during scattering are very small compared with the mass of the heavy particle. This limit is obtained from the above expressions by taking $|p\cdot q|$, $|p\cdot q'|$ and $|q\cdot q'|$ all to be small compared with m_R^2.

Taylor expanding the above expressions shows that both $\mathcal{A}_{RI\to RI}$ and $\mathcal{A}_{II\to II}$ are suppressed in this limit by powers of (q or q')/m_R, in addition to the generic small perturbative factor λ:

$$\mathcal{A}_{RI\to RI} \simeq 2\mathrm{i}\lambda\left(\frac{q\cdot q'}{m_R^2}\right) + O\left(m_R^{-4}\right), \tag{1.9}$$

while

$$\mathcal{A}_{II\to II} \simeq 2\mathrm{i}\lambda\left[\frac{(p\cdot q)^2+(p\cdot q')^2+(q\cdot q')^2}{m_R^4}\right] + O\left(m_R^{-6}\right). \tag{1.10}$$

Both of these expressions use 4-momentum conservation, and kinematic conditions like $q^2=0$ *etc.* to simplify the result, and both expressions end up being suppressed by powers of q/m_R and/or q'/m_R once this is done.

The basic simplicity of physics at low energies arises because physical quantities typically simplify when Taylor expanded in powers of any small energy ratios (like scattering energy/m_R in the example above). It is this simplicity that ultimately underlies the phenomenon of decoupling: in the toy model the low-energy implications of the very energetic $\tilde{\phi}_R$ states ultimately can be organized into a sequence in powers of m_R^{-2}, with only the first few terms relevant at very low energies.

1.2 The simplicity of the low-energy limit ♢

Now imagine that your task is to build an experiment to test the above theory by measuring the cross section for scattering $\tilde{\phi}_I$ particles from various targets, using only accelerators whose energies, E, do not reach anywhere near as high as the mass m_R. Since the

experiment is more difficult if the scattering is rare, the suppression of the order-λ cross sections by powers of q/m_R and/or q'/m_R at low energies presents a potential problem. But maybe this suppression is an accident of the leading, $O(\lambda)$, prediction? If the $O(\lambda^2)$ result is not similarly suppressed, then it might happen that $\mathcal{A} \simeq \lambda^2$ is measurable even if $\mathcal{A} \simeq \lambda(E/m_R)^2$ is not.

It turns out that the suppression of $\tilde{\phi}_I$ scattering at low energies persists order-by-order in the λ expansion, so any hope of evading it by working to higher orders would be in vain. But the hard way to see this is to directly compute the $O(\lambda^n)$ amplitude as a complete function of energy, and then take the low-energy limit. It would be much more efficient if it were possible to zero in directly on the low-energy part of the result *before* investing great effort into calculating the complete answer. Any simplicity that might emerge in the low-energy limit then would be much easier to see.

Indeed, a formalism exists precisely for efficiently identifying the nature of physical quantities in the low-energy limit — *effective field theories* — and it is this formalism that is the topic of this book. This formalism exists and is so useful because one is often in the situation of being faced with a comparatively simple low-energy limit of some, often poorly understood, more complicated system.

The main idea behind this formalism is to take advantage of the low-energy approximation as early as possible in a calculation, and the best way to do so is directly, once and for all, in the action (or Hamiltonian or Lagrangian), rather than doing it separately for each independent observable. But how can the low-energy expansion be performed directly in the action?

1.2.1 Low-energy effective actions

To make this concrete for the toy model discussed above, a starting point is the recognition that the low-energy limit, eq. (1.10), of $\mathcal{A}_{II\to II}$ has precisely the form that would be expected (at leading order of perturbation theory) if the $\tilde{\phi}_I$ particles scattered only through an *effective* interaction of the form $S_{\rm eff} = S_{\rm eff\,0} + S_{\rm eff\,int}$, with

$$S_{\rm eff\,0} = -\frac{1}{2}\int \mathrm{d}^4x\, \partial_\mu\tilde{\phi}_I\, \partial^\mu\tilde{\phi}_I\,, \tag{1.11}$$

and

$$S_{\rm eff\,int} = \frac{\lambda}{4m_R^4}\int \mathrm{d}^4x\, (\partial_\mu\tilde{\phi}_I\, \partial^\mu\tilde{\phi}_I)(\partial_\nu\tilde{\phi}_I\, \partial^\nu\tilde{\phi}_I)\,, \tag{1.12}$$

up to terms of order λ^2 and/or m_R^{-6}.

What is less obvious at this point, but nonetheless true (and argued in detail in the chapters that follow), is that this same effective interaction, eqs. (1.11) and (1.12), also correctly captures the leading low-energy limit of other scattering processes, such as for $\tilde{\phi}_I\tilde{\phi}_I \to \tilde{\phi}_I\tilde{\phi}_I\tilde{\phi}_I\tilde{\phi}_I$ and reactions involving still more $\tilde{\phi}_I$ particles. That is, *all* amplitudes obtained from the full action, eqs. (1.5) and (1.6), precisely agree with those obtained from the effective action, eqs. (1.11) and (1.12), provided that the predictions of both theories are expanded only to leading order in λ and m_R^{-2} [2].

Given that a low-energy action like $S_{\rm eff}$ exists, it is clear that it is much easier to study

the system's low-energy limit by first computing S_{eff} and then using S_{eff} to work out any observable of interest, than it is to calculate all observables using $S_0 + S_{\text{int}}$ of eqs. (1.5) and (1.6), and only then expanding them to find their low-energy form.

As an example of this relative simplicity, because each factor of $\tilde{\phi}_I$ appears differentiated in eq. (1.12), it is obvious that the amplitudes for more complicated scattering processes computed with it are also suppressed by high powers of the low-energy scattering scale. For instance, the amplitude for $\tilde{\phi}_I\tilde{\phi}_I \to N\tilde{\phi}_I$ (into N final particles) computed using tree graphs built using just the quartic interaction $S_{\text{eff int}}$ would be expected to give an amplitude proportional to at least

$$\mathcal{A}_{II\to I\cdots I} \propto \lambda^{N/2}\left(\frac{\text{scattering energy}}{m_R}\right)^{N+2}, \tag{1.13}$$

in the low-energy limit. Needless to say, this type of low-energy suppression is much harder to see when using the full action, eqs. (1.5) and (1.6).

It may seem remarkable that an interaction like S_{eff} exists that completely captures the leading low-energy limit of the full theory in this way. But what is even more remarkable is that a similar effective action also exists that reproduces the predictions of the full theory to *any* fixed higher order in λ and m_R^{-2}. This more general effective action replaces eq. (1.12) by

$$S_{\text{eff int}} = \int \mathrm{d}^4x\, \mathfrak{L}_{\text{eff int}}\,, \tag{1.14}$$

where

$$\begin{aligned}\mathfrak{L}_{\text{eff int}} = {}& a\,(\partial_\mu\tilde{\phi}_I\partial^\mu\tilde{\phi}_I)(\partial_\nu\tilde{\phi}_I\partial^\nu\tilde{\phi}_I)\\ &+b\,(\partial_\mu\tilde{\phi}_I\partial^\mu\tilde{\phi}_I)(\partial_\nu\tilde{\phi}_I\partial^\nu\tilde{\phi}_I)(\partial_\rho\tilde{\phi}_I\partial^\rho\tilde{\phi}_I)+\cdots\,,\end{aligned} \tag{1.15}$$

where the ellipses represent terms involving additional powers of $\partial_\mu\hat{\phi}_I$ and/or its derivatives, though only a finite number of such terms is required in order to reproduce the full theory to a fixed order in λ and m_R^{-2}.

In principle the coefficients a and b in eq. (1.15) are given as a series in λ once the appropriate power of m_R is extracted on dimensional grounds,

$$a = \frac{1}{m_R^4}\left[\frac{\lambda}{4} + a_2\lambda^2 + O(\lambda^3)\right] \quad \text{and} \quad b = \frac{1}{m_R^8}\left[b_1\lambda + b_2\lambda^2 + b_3\lambda^3 + O(\lambda^4)\right]. \tag{1.16}$$

which displays explicitly the order-λ value for a found above that reproduces low-energy scattering in the full theory. Explicit calculations in later sections also show $b_1 = 0$. More generally, to the extent that the leading (classical, or tree-level) part of the action should be proportional to $1/\lambda$ once m_R is eliminated for v using $m_R^2 = \lambda v^2$ (as is argued above, and in more detail in eq. (2.24) and §3), it must also be true that b_2 vanishes.

1.2.2 Why it works

Why is it possible to find an effective action capturing the low-energy limit of a theory, along the lines described above? The basic idea goes as follows.

It is not in itself surprising that there is some sort of Hamiltonian describing the time

evolution of low-energy states. After all, in the full theory time evolution is given by a unitary operation

$$|\psi_f(t)\rangle = U(t,t')\,|\psi_i(t')\rangle\,, \tag{1.17}$$

where $U(t,t') = \exp[-iH(t-t')]$ with a Hamiltonian[5] $H = H(\hat{\phi}_R, \hat{\phi}_l)$ depending on both the heavy and light fields. But if the initial state has an energy $E_i < m_R$ it cannot contain any $\hat{\phi}_R$ particles, and energy conservation then precludes $\hat{\phi}_R$ particles from ever being produced by subsequent time evolution.

This means that time evolution remains a linear and unitary transformation even when it is restricted to low-energy states. That is, suppose we define

$$U_{\rm eff}(t,t') := P_\Lambda\, U(t,t')\, P_\Lambda := \exp\left[-iH_{\rm eff}(t-t')\right]\,, \tag{1.18}$$

with $P_\Lambda^2 = P_\Lambda$ being the projection operator onto states with low energy $E < \Lambda \ll m_R$. P_Λ commutes with H and so also with time evolution. Because $H_{\rm eff} = P_\Lambda H P_\Lambda$ if H is hermitian then so must be $H_{\rm eff}$ and so if $U(t,t')$ is unitary then so must be $U_{\rm eff}(t,t')$ when acting on low-energy states.

Furthermore, because the action of $H_{\rm eff}$ is well-defined for states having energy $E < \Lambda$, it can be written as a linear combination of products of creation and annihilation operators for the $\hat{\phi}_l$ field only (since these form a basis for operators that transform among only low-energy states).[6] As a consequence, it must be possible to write $H_{\rm eff} = H_{\rm eff}[\hat{\phi}_l]$, without making any reference to the heavy field $\hat{\phi}_R$ at all.

But there is no guarantee that the expression for $H_{\rm eff}[\hat{\phi}_l]$ obtained in this way is anywhere as simple as is $H[\hat{\phi}_R, \hat{\phi}_l]$. So the real puzzle is why the effective interaction found above is so simple. In particular, why is it local:

$$H_{\rm eff}[\hat{\phi}_l] = \int \mathrm{d}^3x\, \mathcal{H}_{\rm eff}(x)\,, \tag{1.19}$$

with $\mathcal{H}_{\rm eff}(x)$ a simple polynomial in $\hat{\phi}_l(x)$ and its derivatives, all evaluated at the same spacetime point?

Ultimately, the simplicity of this local form can be traced to the uncertainty principle. Interactions, like eq. (1.12), in $H_{\rm eff}$ not already present in H describe the influence on low-energy $\hat{\phi}_l$ particles of virtual processes involving heavy $\hat{\phi}_R$ particles. These virtual processes are not ruled out by energy conservation even though the production of real $\hat{\phi}_R$ particles is forbidden. One way to understand why they are possible is because the uncertainty principle effectively allows energy conservation to be violated,[7] $E_f = E_i + \Delta E$, but only over time intervals that are sufficiently short, $\Delta t \lesssim \hbar/\Delta E$. The effects of virtual $\hat{\phi}_R$ particles are necessarily localized in time over intervals that are of order $1/m_R$, which

[5] The convention here is to use $\tilde{\phi}$ to denote the fluctuation when this is a non-operator field (appearing within a path integral, say) and instead use $\hat{\phi}$ for the quantum operator fluctuation field.

[6] See the discussion around eq. (C.9) of Appendix C for details.

[7] More precisely, energy need not be conserved at each vertex when organized in old-fashioned Rayleigh-Schrödinger perturbation theory from undergraduate quantum mechanics classes. Once reorganized into manifestly relativistic Feynman-Schwinger-Dyson perturbation theory energy actually *is* preserved at each vertex, but internal particles are not on-shell: $E \neq \sqrt{\mathbf{p}^2 + m^2}$. Either way the locality consequences are the same.

are unobservably short for observers restricted to energies $E \ll m_R$. Consequently they are described at these energies by operators all evaluated at effectively the same time.

In relativistic theories, large momenta necessarily involve large energies and since the uncertainty principle relates large momenta to short spatial distances, a similar argument can be made that the effect of large virtual momentum transfers on the low-energy theory can also be captured by effective interactions localized at a single spatial point. Together with the localization in time just described, this shows that the effects of very massive particles are local in both space and time, as found in the toy model above.

Locality arises explicitly in relativistic calculations when expanding the propagators of massive particles in inverse powers of m_R, after which they become local in spacetime since

$$G(x,y) := \langle 0|T\hat{\phi}_R(x)\hat{\phi}_R(y)|0\rangle = -\mathrm{i}\int \frac{\mathrm{d}^4p}{(2\pi)^4}\,\frac{e^{ip(x-y)}}{p^2+m_R^2} \tag{1.20}$$
$$\simeq -\frac{\mathrm{i}}{m_R^2}\sum_{k=0}^{\infty}\int \frac{\mathrm{d}^4p}{(2\pi)^4}\left(-\frac{p^2}{m_R^2}\right)^k e^{ip(x-y)} = -\frac{\mathrm{i}}{m_R^2}\sum_{k=0}^{\infty}\left(\frac{\Box}{m_R^2}\right)^k \delta^4(x-y)\,,$$

where the 'T' denotes time ordering, $p(x-y) := p\cdot(x-y) = p_\mu(x-y)^\mu$ and $\Box = \partial_\mu\partial^\mu = -\partial_t^2 + \nabla^2$ is the covariant d'Alembertian operator.

The upshot is this: to any fixed order in $1/m_R$ the full theory usually can be described by a local effective lagrangian.[8] The next sections develop tools for its efficient calculation and use.

1.2.3 Symmetries: linear vs nonlinear realization

Before turning to the nitty gritty of how the effective action is calculated and used, it is worth first pausing to extract one more useful lesson from the toy model considered above. The lesson is about symmetries and their low-energy realization, and starts by asking why it is that the self-interactions among the light $\hat{\phi}_I$ particles — such as the amplitudes of eqs. (1.9) and (1.10) — are so strongly suppressed at low energies by powers of $1/m_R^2$.

That is, although it is natural to expect some generic suppression of low-energy interactions by powers of $1/m_R^2$, as argued above, why does nothing at all arise at zeroeth order in $1/m_R$ despite the appearance of terms like $\lambda\hat{\phi}_I^4$ in the full toy-model potential? And why are there so very many powers of $1/m_R$ in the case of $2\hat{\phi}_I \to N\hat{\phi}_I$ scattering in the toy model? (Specifically, why is the amplitude for two $\hat{\phi}_I$ particles scattering to N $\hat{\phi}_I$ particles suppressed by $(1/m_R)^{N+2}$?)

This suppression has a very general origin, and can be traced to a symmetry of the underlying theory [3, 4, 5]. The symmetry in question is invariance under the $U(1)$ phase rotation, $\phi \to e^{\mathrm{i}\omega}\phi$, of eqs. (1.1) and (1.2). In terms of the real and imaginary parts this acts as

$$\begin{pmatrix}\phi_R\\ \phi_I\end{pmatrix} \to \begin{pmatrix}\cos\omega & -\sin\omega\\ \sin\omega & \cos\omega\end{pmatrix}\begin{pmatrix}\phi_R\\ \phi_I\end{pmatrix}. \tag{1.21}$$

[8] For non-relativistic systems locality sometimes breaks down in space (*e.g.* when large momenta coexist with low energy). It can also happen that the very existence of a Hamiltonian (without expanding the number of degrees of freedom) breaks down for open systems — the topic of §16.

A symmetry such as this that acts linearly on the fields is said to be *linearly realized*. As summarized in Appendix C.4, if the symmetry is also linearly realized on particle states then these states come in multiplets of the symmetry, all elements of which share the same couplings and masses. However (as is also argued in Appendix C.4) linear transformations of the fields — such as (1.21) — are insufficient to infer that the symmetry also acts linearly for particle states, $|\mathbf{p}\rangle = a^*_{\mathbf{p}}|0\rangle$, unless the ground-state, $|0\rangle$, is also invariant. If a symmetry of the action does not leave the ground state invariant it is said to be *spontaneously broken*.

For instance, in the toy model the ground state satisfies $\langle 0|\phi(x)|0\rangle = v$, and so the ground state is only invariant under $\phi \to e^{i\omega}\phi$ when $v = 0$. Indeed for the toy model if $v = 0$ both particle masses are indeed equal: $m_R = m_I = 0$, as are all of their self-couplings. By contrast, when $v \neq 0$ the masses of the two types of particles differ, as does the strength of their cubic self-couplings. Although $\phi \to e^{i\omega}\phi$ always transforms linearly, the symmetry acts inhomogeneously on the deviation $\hat{\phi} = \phi - v = \frac{1}{\sqrt{2}}(\hat{\phi}_R + i\hat{\phi}_I)$ that creates and destroys the particle states. It is because the deviation does not transform linearly (and homogeneously) that the arguments in Appendix C.4 no longer imply that particle states need have the same couplings and masses when $v \neq 0$.

To see why this symmetry should suppress low-energy $\hat{\phi}_I$ interactions, consider how it acts within the low-energy theory. Even though ϕ transforms linearly in the full theory, because the low-energy theory involves only the single real field $\hat{\phi}_I$, the symmetry cannot act on it in a linear and homogeneous way. To see what the action of the symmetry becomes purely within the low-energy theory, it is useful to change variables to a more convenient set of fields than $\hat{\phi}_R$ and $\hat{\phi}_I$.

To this end, define the two real fields χ and ξ by[9]

$$\phi = \left(v + \frac{\chi}{\sqrt{2}}\right) e^{i\xi/\sqrt{2}\,v}\,. \tag{1.22}$$

These have the advantage that the action of the $U(1)$ symmetry, $\phi \to e^{i\omega}\phi$ takes a particularly simple form,

$$\xi \to \xi + \sqrt{2}\,v\,\omega\,, \tag{1.23}$$

with χ unchanged, so ξ carries the complete burden of symmetry transformation.

In terms of these fields the action, eq. (1.1), becomes

$$S = -\int \mathrm{d}^4x \left[\frac{1}{2}\,\partial_\mu\chi\partial^\mu\chi + \frac{1}{2}\left(1 + \frac{\chi}{\sqrt{2}\,v}\right)^2 \partial_\mu\xi\partial^\mu\xi + V(\chi)\right], \tag{1.24}$$

with

$$V(\chi) = \frac{\lambda}{4}\left(\sqrt{2}\,v\chi + \frac{\chi^2}{2}\right)^2\,, \tag{1.25}$$

Expanding this action in powers of χ and ξ gives the perturbative action $S = S_0 + S_{\rm int}$, with unperturbed contribution

$$S_0 = -\frac{1}{2}\int \mathrm{d}^4x \left[\partial_\mu\chi\partial^\mu\chi + \partial_\mu\xi\partial^\mu\xi + \lambda v^2\chi^2\right]. \tag{1.26}$$

[9] Numerical factors are chosen here to ensure fields are canonically normalized.

This shows that χ is an alternative field representation for the heavy particle, with $m_\chi^2 = m_R^2 = \lambda v^2$. ξ similarly represents the massless field.

It also shows the symmetry is purely realized on the massless state, as an inhomogeneous shift (1.23) rather than a linear, homogeneous transformation. Such a transformation — often called a *nonlinear realization* of the symmetry (both to distinguish it from the linear realization discussed above, and because the transformations turn out in general to be nonlinear when applied to non-abelian symmetries) — is a characteristic symmetry realization in the low-energy limit of a system which spontaneously breaks a symmetry.

The interactions in this representation are given by

$$S_{\rm int} = -\int \mathrm{d}^4x \left[\left(\frac{\chi}{\sqrt{2}\,v} + \frac{\chi^2}{4\,v^2}\right)\partial_\mu\xi\partial^\mu\xi + \frac{\lambda v}{2\sqrt{2}}\chi^3 + \frac{\lambda}{16}\chi^4\right]. \tag{1.27}$$

For the present purposes, what is important about these expressions is that ξ always appears differentiated. This is a direct consequence of the symmetry transformation, eq. (1.23), which requires invariance under constant shifts: $\xi \to \xi + \text{constant}$. Since this symmetry forbids a ξ mass term, which would be $\propto m_I^2\,\xi^2$, it ensures ξ remains exactly massless to all orders in the small expansion parameters. ξ is what is called a *Goldstone boson* for the spontaneously broken $U(1)$ symmetry: it is the massless scalar that is guaranteed to exist for spontaneously broken (global) symmetries. Because ξ appears always differentiated it is immediately obvious that an amplitude describing N_i ξ particles scattering into N_f ξ particles must be proportional to at least $N_i + N_f$ powers of their energy, explaining the low-energy suppression of light-particle scattering amplitudes in this toy model.

For instance, explicitly re-evaluating the Feynman graphs of Fig. 1.3, using the interactions of eq. (1.27) instead of (1.6), gives the case $N_i = N_f = 2$ as

$$\begin{aligned}\mathcal{A}_{\xi\xi\to\xi\xi} &= 0 + 8\left(\frac{\mathrm{i}^2}{2}\right)\left(-\frac{1}{\sqrt{2}\,v}\right)^2\left[\frac{-\mathrm{i}(p\cdot q)(p'\cdot q')}{(p+q)^2+m_R^2} + \frac{-\mathrm{i}(p\cdot p')(q\cdot q')}{(p-p')^2+m_R^2} + \frac{-\mathrm{i}(p\cdot q')(q\cdot p')}{(p-q')^2+m_R^2}\right] \\ &= \frac{2\mathrm{i}\lambda}{m_R^4}\left[\frac{(p\cdot q)^2}{1+2p\cdot q/m_R^2} + \frac{(q\cdot q')^2}{1-2q\cdot q'/m_R^2} + \frac{(p\cdot q')^2}{1-2p\cdot q'/m_R^2}\right],\end{aligned} \tag{1.28}$$

in precise agreement with eq. (1.8) — as may be seen explicitly using the identity $(1 + x)^{-1} = 1 - x + x^2/(1 + x)$ — but with the leading low-energy limit much more explicit.

This representation of the toy model teaches several things. First, it shows that scattering amplitudes (and, more generally, arbitrary physical observables) do not depend on which choice of field variables are used to describe a calculation [8, 9, 10]. Some kinds of calculations (like loops and renormalization) are more convenient using the variables $\hat\phi_R$ and $\hat\phi_I$, while others (like extracting consequences of symmetries) are easier using χ and ξ.

Second, this example shows that it is worthwhile to use the freedom to perform field redefinitions to choose those fields that make life as simple as possible. In particular, it is often very useful to make symmetries of the high-energy theory as explicit as possible in the low-energy theory as well.

Third, this example shows that once restricted to the low-energy theory it need not be true that a symmetry remains linearly realized by the fields [11, 12, 13], even if this were true for the full underlying theory including the heavy particles. The necessity of realizing symmetries nonlinearly arises once the scales defining the low-energy theory (*e.g.* $E \ll m_R$)

are smaller than the mass difference (*e.g.* m_R) between particles that are related by the symmetry in the full theory, since in this case some of the states required to fill out a linear multiplet are removed as part of the high-energy theory.

1.3 Summary

This first chapter defines a toy model, in which a complex scalar field, ϕ, self-interacts *via* a potential $V = \frac{\lambda}{4}(\phi^*\phi - v^2)^2$ that preserves a $U(1)$ symmetry: $\phi \to e^{i\omega}\phi$. Predictions for particle masses and scattering amplitudes are made as a function of the model's two parameters, λ and v, in the semiclassical regime $\lambda \ll 1$. This model is used throughout the remaining chapters of Part I as a vehicle for illustrating how the formalism of effective field theories works in a concrete particular case.

The semiclassical spectrum of the model has two phases. If $v = 0$ the $U(1)$ symmetry is preserved by the semiclassical ground state and there are two particles whose couplings and masses are the same because of the symmetry. When $v \neq 0$ the symmetry is spontaneously broken, and one particle is massless while the other gets a nonzero mass $m = \sqrt{\lambda}\, v$.

The model's symmetry-breaking phase has a low-energy regime, $E \ll m$, that provides a useful illustration of low-energy methods. In particular, the massive particle decouples at low energies in the precise sense that its virtual effects only play a limited role for the low-energy interactions of the massless particles. In particular, explicit calculation shows the scattering of massless particles at low energies in the full theory to be well-described to leading order in λ and E/m in terms of a simple local 'effective' interaction with lagrangian density $\mathfrak{L}_{\rm eff} = a_{\rm eff}(\partial_\mu \xi\, \partial^\mu \xi)^2$, with effective coupling: $a_{\rm eff} = \lambda/(4m^4)$. The $U(1)$ symmetry of the full theory appears in the low-energy theory as a shift symmetry $\xi \to \xi +$ constant.

Exercises

Exercise 1.1 Use the Feynman rules coming from the action $S = S_0 + S_{\rm int}$ given in eqs. (1.5) and (1.6) to evaluate the graphs of Fig. 1.2. Show from your result that the corresponding S-matrix element is given by

$$\langle \hat\phi_R(p'), \hat\phi_I(q') | \mathcal{S} \,|\, \hat\phi_R(p), \hat\phi_I(q)\rangle = -i(2\pi)^4 \mathcal{A}_{RI\to RI}\, \delta^4(p+q-p'-q')\,,$$

with $\mathcal{A}_{RI\to RI}$ given by eq. (1.7). Taylor expand your result for small q, q' to verify the low-energy limit given in eq. (1.9). [Besides showing the low-energy decoupling of Goldstone particles, getting right the cancellation that provides this suppression in

these variables is a good test of — and a way to develop faith in — your understanding of Feynman rules.]

Exercise 1.2 Using the Feynman rules coming from the action $S = S_0 + S_{\rm int}$ given in eqs. (1.5) and (1.6) evaluate the graphs of Fig. 1.3 to show

$$\langle \hat\phi_I(p'), \hat\phi_I(q') | \mathcal{S} \, | \, \hat\phi_I(p), \hat\phi_I(q) \rangle = -{\rm i}(2\pi)^4 \mathcal{A}_{II\to II} \, \delta^4(p+q-p'-q') \,,$$

with $\mathcal{A}_{II\to II}$ given by eq. (1.8). Taylor expand your result for small q, q' to verify the low-energy limit given in eq. (1.10).

Exercise 1.3 Using the toy model's leading effective interaction $S = S_{\rm eff\,0} + S_{\rm eff\,int}$, with Feynman rules drawn from (1.11) (1.12), draw the graphs that produce the dominant contributions — *i.e.* carry the fewest factors of λ and (external energy)/m_R — to the scattering process $\hat\phi_I + \hat\phi_I \to 4\hat\phi_I$. Show that these agree with the estimate (1.13) in their prediction for the leading power of λ and of external energy.

Exercise 1.4 Using the Feynman rules coming from the action $S = S_0 + S_{\rm int}$ given in eqs. (1.26) and (1.27) evaluate the graphs of Fig. 1.3 to show

$$\langle \hat\xi_I(p'), \hat\xi(q') | \mathcal{S} \, | \, \hat\xi_I(p), \hat\xi_I(q) \rangle = -{\rm i}(2\pi)^4 \mathcal{A}_{\xi\xi\to\xi\xi} \, \delta^4(p+q-p'-q') \,,$$

with $\mathcal{A}_{\xi\xi\to\xi\xi}$ given by eq. (1.28). [Comparing this result to the result in Exercise 1.2 provides an illustration of Borcher's theorem [8, 9, 10], that states that scattering amplitudes remain unchanged by a broad class of local field redefinitions.]

2 Effective actions

Having seen in the previous chapter how the low-energy limit works in a specific example, this chapter gets down to the business of defining the low-energy effective theory more explicitly. The next chapter is then devoted to calculational issues of how to use and compute with an effective theory. The first two sections start with a brief review of the formalism of generating functionals, introducing in particular the generator of one-particle irreducible correlation functions. These can be skipped by field theory aficionados interested in cutting immediately to the low-energy chase.

2.1 Generating functionals - a review ♡

The starting point is a formalism convenient for describing the properties of generic observables in a general quantum field theory. Consider a field theory involving N quantum fields $\hat{\phi}^a(x)$, $a = 1, ..., N$, governed by a classical action $S[\phi]$, and imagine computing the theory's 'in-out' correlation functions,[1]

$$G^{a_1 \cdots a_n}(x_1, \cdots, x_n) := {}_o\langle \Omega | T[\hat{\phi}^{a_1}(x_1) \cdots \hat{\phi}^{a_n}(x_n)] | \Omega \rangle_i \,, \tag{2.1}$$

where $|\Omega\rangle_i$ (or $|\Omega\rangle_o$) denotes the system's ground state in the remote past (or future) and T denotes time ordering.[2] These correlation functions are useful to consider because general observables including (but not limited to) scattering amplitudes can be extracted from them using standard procedures.

All such correlation functions can be dealt with at once by working with the generating functional, $Z[J]$, defined by

$$Z[J] := \sum_{n=0}^{\infty} \frac{\mathrm{i}^n}{n!} \int \mathrm{d}^4 x_1 \cdots \mathrm{d}^4 x_n \, G^{a_1 \cdots a_n}(x_1, \cdots, x_n) J_{a_1}(x_1) \cdots J_{a_n}(x_n) \,, \tag{2.2}$$

from which each correlation function can be obtained by functional differentiation

$$G^{a_1 \cdots a_n}(x_1, \cdots, x_n) = (-\mathrm{i})^n \left(\frac{\delta^n Z[J]}{\delta J_{a_1}(x_1) \cdots \delta J_{a_n}(x_n)} \right)_{J=0} . \tag{2.3}$$

A useful property of $Z[J]$ is that it has a straightforward expression in terms of path integrals, which in principle could be imagined to be computed numerically, but in practice

[1] In what follows these are assumed to be bosonic fields though a similar treatment goes through for fermions.

[2] Strictly speaking, for relativistic theories T denotes the T^* ordering, which includes certain additional equal-time 'seagull' contributions required to maintain Lorentz covariance [14].

usually means that it is relatively simple to calculate perturbatively in terms of Feynman graphs.

The basic connection between operator correlation functions and path integrals is the expression

$$G^{a_1\cdots a_n}(x_1,\cdots,x_n) = \int \mathcal{D}\phi \left[\phi^{a_1}(x_1)\cdots\phi^{a_n}(x_n)\right] \exp\left\{\mathrm{i}S[\phi]\right\}, \tag{2.4}$$

where $\mathcal{D}\phi = \mathcal{D}\phi^{a_1}\cdots\mathcal{D}\phi^{a_n}$ denotes the functional measure for the sum over all field configurations, $\phi^a(x)$, with initial and final times weighted by the wave functional, $\Psi_i[\phi]$ and $\Psi_o^*[\phi]$, appropriate for the initial and final states, ${}_o\langle\Omega|$ and $|\Omega\rangle_i$. The special case $n = 0$ is the example most frequently encountered in elementary treatments, for which

$${}_o\langle\Omega|\Omega\rangle_i = \int \mathcal{D}\phi \ \exp\left\{\mathrm{i}S[\phi]\right\}. \tag{2.5}$$

Direct use of the definitions then leads to the following expression for $Z[J]$:

$$Z[J] = \int \mathcal{D}\phi \ \exp\left\{\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^4x\, \phi^a(x)J_a(x)\right\}, \tag{2.6}$$

and so $Z[J=0] = {}_o\langle\Omega|\Omega\rangle_i$.

Semiclassical evaluation

Semiclassical perturbation theory can be formulated by expanding the action within the path integral about a classical background:[3] $\phi^a(x) = \varphi^a_{\mathrm{cl}}(x) + \tilde{\phi}^a(x)$, where φ^a_{cl} satisfies

$$\left(\frac{\delta S}{\delta\phi^a}\right)_{\phi=\varphi_{\mathrm{cl}}} + J_a = 0\,. \tag{2.7}$$

The idea is to write the action, $S_J[\phi] := S[\phi] + \int \mathrm{d}^4x\,(\phi^a J_a)$, as

$$S_J[\varphi_{\mathrm{cl}} + \tilde{\phi}] = S_J[\varphi_{\mathrm{cl}}] + S_2[\varphi_{\mathrm{cl}}, \tilde{\phi}] + S_{\mathrm{int}}[\varphi_{\mathrm{cl}}, \tilde{\phi}], \tag{2.8}$$

with

$$S_2 = -\int \mathrm{d}^4x\, \tilde{\phi}^a\, \Delta_{ab}(\varphi_{\mathrm{cl}})\tilde{\phi}^b, \tag{2.9}$$

being the quadratic part in the expansion (for some differential operator Δ_{ab}). The 'interaction' term, S_{int}, contains all terms cubic and higher order in $\tilde{\phi}^a$; no linear terms appear because the background field satisfies eq. (2.7).

The relevant path integrals can then be evaluated by expanding

$$\exp\left\{\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^4x\,\phi^a J_a\right\} = \exp\left\{\mathrm{i}S_J[\varphi_{\mathrm{cl}}] + \mathrm{i}S_2[\varphi_{\mathrm{cl}}, \tilde{\phi}]\right\} \sum_{r=0}^{\infty} \frac{1}{r!}\left(\mathrm{i}S_{\mathrm{int}}[\varphi_{\mathrm{cl}}, \tilde{\phi}]\right)^r, \tag{2.10}$$

in the path integral (2.6) and explicitly computing the resulting gaussian functional integrals.

[3] It is sometimes useful to make a more complicated, nonlinear, split $\phi = \phi(\varphi_{\mathrm{cl}}, \tilde{\phi})$ in order to make explicit convenient properties (such as symmetries) of the action.

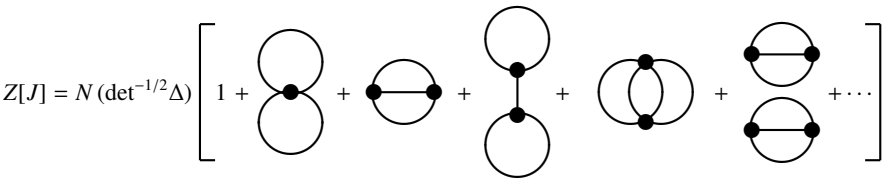


Fig. 2.1 A sampling of some leading perturbative contributions to the generating functional $Z[J]$ expressed using eq. (2.11) as Feynman graphs. Solid lines are propagators (Δ^{-1}) and solid circles represent interactions that appear in $S_{\rm int}$. 1-particle reducible and 1PI graphs are both shown as examples at two loops and a disconnected graph is shown at four loops. The graphs shown use only quartic and cubic interactions in $S_{\rm int}$.

This process leads in the usual way to the graphical representation of any correlation function. Gaussian integrals ultimately involve integrands that are powers of fields, leading to integrals of the schematic form[4]

$$\int \mathcal{D}\tilde{\phi}\, e^{i\tilde{\phi}^a \Delta_{ab} \tilde{\phi}^b} \tilde{\phi}^{c_1}(x_1) \cdots \tilde{\phi}^{c_n}(x_n) \propto \left(\det{}^{-1/2}\Delta\right) \left[(\Delta^{-1})^{c_1 c_2} \cdots (\Delta^{-1})^{c_{n-1} c_n} + (\text{perms})\right], \quad (2.11)$$

if n is even, while the integral vanishes if n is odd. Here the evaluation ignores a proportionality constant that is background-field independent (and so isn't important in what follows). The interpretation in terms of Feynman graphs comes because the combinatorics of such an integral correspond to the combinatorics of all possible ways of drawing graphs whose internal lines represent factors of Δ^{-1} and whose vertices correspond to interactions within $S_{\rm int}$.

Within this type of graphical expression $Z[J]$ is given as the sum over all vacuum graphs, with no external lines. All of the dependence on J appears through the dependence of the result on $\varphi_{\rm cl}$, which depends on J because of (2.7). The graphs involving the fewest interactions (vertices) first arise with two loops, a sampling of which are shown in Fig. 2.1 that can be built using cubic and quartic interactions within $S_{\rm int}$.

As mentioned earlier this includes *all* graphs, including those that are disconnected, like the right-most four-loop graph involving four cubic vertices in Fig. 2.1. Graphs like this are disconnected in the sense that it is not possible to get between any pair of vertices along some sequence of contiguous internal lines.

Although simple to state, the perturbation expansion outlined above in terms of vacuum graphs is not yet completely practical for explicit calculations. The problem is the appearance of the background field φ in the propagator $(\Delta^{-1})_{ab}$. Although $\Delta_{ab}(x,y) = -\delta^2 S/\delta\phi^a(x)\delta\phi^b(y)$ itself is easy to compute, it is often difficult to invert explicitly for generic background fields. For instance, for a single scalar field interacting through a scalar potential $U(\phi)$ one has $\Delta(x,y) = [-\Box + U''(\varphi)]\,\delta^4(x-y)$ and although this is easily inverted in momentum space when φ is constant, it is more difficult to invert for arbitrary $\varphi(x)$.

[4] This expression assumes a bosonic field, but a similar expression holds for fermions.

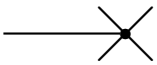


Fig. 2.2 The Feynman rule for the vertex coming from the linear term, $S_{\rm lin}$, in the expansion of the action. The cross represents the sum $\delta S/\delta\varphi^a + J_a$.

This problem is usually addressed by expanding in powers of $J_a(x)$, so that the path integral is evaluated as a semiclassical expansion about a simple background configuration, $\varphi^a_{\rm cl}$, that satisfies $(\delta S/\delta\phi^a)(\phi = \varphi_{\rm cl}) = 0$ instead of eq. (2.7). The Feynman graphs for this modified expansion differ in two ways from the expansion described above: (*i*) the propagators Δ^{-1} now are evaluated at a J-independent configuration, $\varphi^a_{\rm cl}$, which can be explicitly evaluated if this configuration is simple enough (such as, for instance, if $\varphi^a_{\rm cl} = 0$); and (*ii*) the term $\phi^a J_a$ in the exponent of the integrand in (2.6) is now treated as an interaction. Since this interaction is linear in ϕ^a it corresponds graphically to a 'tadpole' contribution (as in Fig. 2.2) with the line ending in a cross whose Feynman rule is $J_a(x)$.

Within this framework the Feynman graphs giving $Z[J]$ are obtained from those given in Fig. 2.1 by inserting external lines in all possible ways (both to propagators and vertices), with the understanding that the end of each external line represents a factor of $J_a(x)$. This kind of modified expansion gives $Z[J]$ explicitly as a Taylor expansion in powers of J.

2.1.1 Connected correlations

As Fig. 2.1 shows, the graphical expansion for $Z[J]$ in perturbation theory includes both connected and disconnected Feynman graphs. It is often useful to work instead with a generating functional, $W[J]$, whose graphical expansion contains only connected graphs. As shown in Exercise 2.4, this is accomplished simply by defining $Z[J] := \exp\{{\rm i}W[J]\}$ [5, 15], since taking the logarithm has the effect of subtracting out the disconnected graphs. This implies the path-integral representation

$$\exp\{{\rm i}W[J]\} = \int \mathcal{D}\phi\ \exp\left\{{\rm i}S[\phi] + {\rm i}\int {\rm d}^4x\,\phi^a J_a\right\}. \tag{2.12}$$

The connected, time-ordered correlation functions are then given by functional differentiation:

$$\langle T[\phi^{a_1}(x_1)\cdots\phi^{a_n}(x_n)]\rangle_c := (-{\rm i})^{n-1}\left(\frac{\delta^n W[J]}{\delta J_{a_1}(x_1)\cdots\delta J_{a_n}(x_n)}\right)_{J=0}. \tag{2.13}$$

For example

$$\langle\phi^a(x)\rangle_c = \left(\frac{\delta W[J]}{\delta J_a(x)}\right)_{J=0} = -{\rm i}\left(\frac{1}{Z[J]}\frac{\delta Z[J]}{\delta J_b(y)}\right)_{J=0} = \frac{{}_o\langle\Omega|\phi^a(x)|\Omega\rangle_i}{{}_o\langle\Omega|\Omega\rangle_i}, \tag{2.14}$$

while

$$\langle T[\phi^a(x)\,\phi^b(y)]\rangle_c = -\mathrm{i}\left(\frac{\delta^2 W[J]}{\delta J_a(x)\delta J_b(y)}\right)_{J=0} \tag{2.15}$$
$$= \frac{{}_o\langle\Omega|T[\phi^a(x)\,\phi^b(y)]|\Omega\rangle_i}{{}_o\langle\Omega|\Omega\rangle_i} - \left(\frac{{}_o\langle\Omega|\phi^a(x)|\Omega\rangle_i}{{}_o\langle\Omega|\Omega\rangle_i}\right)\left(\frac{{}_o\langle\Omega|\phi^b(y)|\Omega\rangle_i}{{}_o\langle\Omega|\Omega\rangle_i}\right),$$

and so on. As is easily verified, the graphical expansion of the factor ${}_o\langle\Omega|T[\phi^a(x)\,\phi^b(y)]|\Omega\rangle_i$ in this last expression corresponds to the sum over all Feynman graphs with precisely two external lines, corresponding to the fields $\phi^a(x)$ and $\phi^b(y)$. The graphical representation of a term like ${}_o\langle\Omega|\phi^a(x)|\Omega\rangle_i$ is similarly given by the sum over all Feynman graphs (called tadpole graphs) with precisely one external line, corresponding to $\phi^a(x)$.

Dividing all terms by the factors of ${}_o\langle\Omega|\Omega\rangle_i$ in the denominator is precisely what is needed to cancel disconnected vacuum graphs (*i.e.* those disconnected subgraphs having no external lines). But this does not remove graphs in ${}_o\langle\Omega|T[\phi^a(x)\,\phi^b(y)]|\Omega\rangle_i$ corresponding to a pair of disconnected 'tadpole' graphs, each of which has a single external line. These disconnected graphs precisely correspond to the product ${}_o\langle\Omega|\phi^a(x)|\Omega\rangle_i\,{}_o\langle\Omega|\phi^b(y)|\Omega\rangle_i$ in (2.15), whose subtraction therefore cancels the remaining disconnected component from $\langle T[\phi^a(x)\,\phi^b(y)]\rangle_c$.

A similar story goes through for the higher functional derivatives, and shows how correlations obtained by differentiating W have their disconnected parts systematically subtracted off. Indeed eqs. (2.13) and (2.12) can be used as non-perturbative definitions of what is meant by connected correlations functions and their generators [15].

2.1.2 The 1PI (or quantum) action ♠

As eqs. (2.12) and (2.4) show, the functional $Z[J] = \exp\{\mathrm{i}W[J]\}$ can be physically interpreted as the 'in-out' vacuum amplitude in the presence of an applied current $J_a(x)$. Furthermore, the applied current can be regarded as being responsible for changing the expectation value of the field, since not evaluating eq. (2.14) at $J_a = 0$ gives

$$\varphi^a(x) := \langle\phi^a(x)\rangle_J = \frac{\delta W}{\delta J_a(x)}\,, \tag{2.16}$$

as a functional of the current $J_a(x)$. However, it is often more useful to have the vacuum-to-vacuum amplitude expressed directly as a functional of the expectation value, $\varphi^a(x)$, itself, rather than $J_a(x)$. This is accomplished by performing a Legendre transform, as follows.

Legendre transform

To perform a Legendre transform, define [15]

$$\Gamma[\varphi\,] := W[J] - \int \mathrm{d}^4x\,\varphi^a J_a\,, \tag{2.17}$$

with $J_a(x)$ regarded as a functional of $\varphi^a(x)$, implicitly obtained by solving eq. (2.16). Once $\Gamma[\varphi\,]$ is known, the current required to obtain the given $\varphi^a(x)$ is found by directly

differentiating the definition, eq. (2.17), using the chain rule together with eq. (2.16) to evaluate the functional derivative of $W[J]$:

$$\frac{\delta\Gamma}{\delta\varphi^a(x)} = \int \mathrm{d}^4y\, \frac{\delta J_b(y)}{\delta\varphi^a(x)}\,\frac{\delta W}{\delta J_b(y)} - J_a(x) - \int \mathrm{d}^4y\, \varphi^b(y)\,\frac{\delta J_a(y)}{\delta\varphi^a(x)} = -J_a(x)\,. \tag{2.18}$$

In particular, this last equation shows that the expectation value for the 'real' system with $J_a = 0$ is a stationary point of $\Gamma[\varphi\,]$. In this sense $\Gamma[\varphi\,]$ is related to $\langle\phi^a\rangle$ in the same way that the classical action, $S[\phi\,]$, is related to a classical background configuration, $\varphi^a_{\rm cl}$. For this reason $\Gamma[\varphi\,]$ is sometimes thought of as the quantum generalization of the classical action, and known as the theory's *quantum action*.

This similarity between $\Gamma[\varphi\,]$ and the classical action is also reinforced by other considerations. For instance, because the classical action is usually the difference, $S = K - V$, between kinetic and potential energies, for time-independent configurations (for which the kinetic energy is $K = 0$) the classical ground state actually minimizes $V = -S$. It can be shown that for time-independent systems — *i.e.* those where the ground state $|\Omega\rangle$ is well-described in the adiabatic approximation — the configuration $\varphi^a = \langle\phi^a\rangle$ similarly minimizes the quantity $-\Gamma$. In particular, for configurations φ^a independent of spacetime position the configuration minimizes the quantum 'effective potential' $V_{\rm q}(\varphi) = -\Gamma[\varphi]/(\mathrm{Vol})$ where 'Vol' is the overall volume of spactime.

One way to prove this [16, 17] is to show that, for any static configuration, φ^a, the quantity $-\Gamma[\varphi\,]$ can be interpreted as the minimum value of the energy, $\langle\Psi|H|\Psi\rangle$, extremized over all normalized states, $|\Psi\rangle$, that satisfy the condition $\langle\Psi|\phi^a(x)|\Psi\rangle = \varphi^a(x)$. The global minimum to $-\Gamma[\varphi\,]$ then comes once φ^a is itself varied over all possible values.

Semiclassical expansion

How is $\Gamma[\varphi\,]$ computed within perturbation theory? To find out, multiply the path-integral representation for $W[J]$, eq. (2.12), on both sides by $\exp\left\{-\mathrm{i}\int \mathrm{d}^4x\,(\varphi^a J_a)\right\}$. Since neither φ^a nor J_a are integration variables, this factor may be taken inside the path integral, giving

$$\begin{aligned}\exp\Bigl\{\mathrm{i}\Gamma[\varphi\,]\Bigr\} &= \exp\left\{\mathrm{i}W[J] - \mathrm{i}\int \mathrm{d}^4x\,\varphi^a J_a\right\}\\ &= \int \mathcal{D}\phi\ \exp\left\{\mathrm{i}S[\phi\,] + \mathrm{i}\int \mathrm{d}^4x\,(\phi^a-\varphi^a)J_a\right\}\\ &= \int \mathcal{D}\tilde\phi\ \exp\left\{\mathrm{i}S[\varphi+\tilde\phi\,] + \mathrm{i}\int \mathrm{d}^4x\,\tilde\phi^a J_a\right\}.\end{aligned} \tag{2.19}$$

The last line uses the change of integration variable $\phi^a \to \tilde\phi^a := \phi^a - \varphi^a$.

At face value eq. (2.19) doesn't seem so useful in practice, since the dependence on J_a inside the integral is to be regarded as a functional of φ^a, using eq. (2.18). This means that $\Gamma[\varphi\,]$ is only given implicitly, since it appears on both sides. But on closer inspection the situation is much better than this, because the implicit appearance of Γ through J_a on the right-hand-side is actually very easy to implement in perturbation theory.

To see how this works, imagine evaluating eq. (2.19) perturbatively by expanding the

action inside the path integral about the configuration $\phi^a = \varphi^a$, using

$$S[\varphi + \tilde{\phi}] = S[\varphi] + S_2[\varphi, \tilde{\phi}] + S_{\rm lin}[\varphi, \tilde{\phi}] + S_{\rm int}[\varphi, \tilde{\phi}]\,. \tag{2.20}$$

This is very similar to the expansion in eq. (2.10), apart from the term linear in $\tilde{\phi}^a$,

$$S_{\rm lin}[\varphi, \tilde{\phi}] = \int \mathrm{d}^4x \left[\left(\frac{\delta S}{\delta \phi^a(x)} \right)_{\phi=\varphi} + J_a(x) \right] \tilde{\phi}^a(x)\,, \tag{2.21}$$

which does not vanish as it did before because $\delta\varphi^a := \varphi^a - \varphi^a_{\rm cl} \neq 0$. However, because the difference $\delta\varphi^a$ is perturbatively small, it may be grouped with the terms in $S_{\rm int}$ and expanded within the integrand, and not kept in the exponential.

The resulting expansion for $\Gamma[\varphi]$ then becomes

$$e^{\mathrm{i}\Gamma[\varphi]} = e^{\mathrm{i}S[\varphi]} \int \mathcal{D}\tilde{\phi}\; e^{\mathrm{i}S_2[\varphi,\tilde{\phi}]} \sum_{r=0}^{\infty} \frac{1}{r!} \big(\mathrm{i}S_{\rm int} + \mathrm{i}S_{\rm lin} \big)^r\,, \tag{2.22}$$

and so

$$\Gamma[\varphi] = S[\varphi] + \frac{\mathrm{i}}{2} \ln\det\Delta + (\text{2-loops and higher})\,, \tag{2.23}$$

where $\Delta(\varphi)$ is the operator appearing in $S_2 = -\int \mathrm{d}^4x\, \tilde{\phi}^a \Delta_{ab}\, \tilde{\phi}^b$, and the contribution called '2-loops and higher' denotes the sum of all Feynman graphs involving two or more loops (and no external lines) built with internal lines representing the propagator Δ^{-1}, and vertices built using $S_{\rm int} + S_{\rm lin}$.

Expression (2.23) shows why the perturbative expansion of $\Gamma[\phi]$ is often related to the semiclassical approximation and loop expansion. As argued above eq. (1.3), whenever a large dimensionless parameter pre-multiplies the classical action — such as if $S = \tilde{S}/\lambda$ for $\lambda \ll 1$ — then each additional loop costs a factor of λ. In detail this is true because each vertex in a Feynman graph carries a factor $1/\lambda$ while each propagator is proportional to $\Delta^{-1} \propto \lambda$. Consequently a graph with $\mathcal{I}$ internal lines and $\mathcal{V}$ vertices is proportional to λ^x with

$$x = \mathcal{I} - \mathcal{V} = \mathcal{L} - 1\,, \tag{2.24}$$

where $\mathcal{L}$ is the total number of loops[5] in the graph (more about which below).

The first two contributions to (2.23) are the classical and one-loop results, while the last term turns out to consist of the sum over all Feynman graphs with two or more loops. This can be seen because the first two terms of (2.23) are proportional to λ^{-1} and λ^0 respectively, while all terms built using $S_{\rm int}$ start at order λ, with two loop graphs being order λ; 3-loop graphs are order λ^2 and so on. The connection between the loop and semiclassical expansions comes because the semiclassical expansion is normally regarded as a series in powers of $\hbar$. In ordinary units it is $S/\hbar$ that appears in the path integral, so the argument just given shows that powers of $\hbar$ also count loops. Counting powers of small dimensionless

[5] When restricted to graphs that can be drawn on a plane this identity agrees with an intuitive definition of what the number of loops in a graph should be (because the combination $\mathcal{L} - \mathcal{I} + \mathcal{V} = 1$ is a topological invariant for all such graphs). For graphs that cannot be drawn on a plane this expression defines the number of loops.

quantities is more informative than counting powers of $\hbar$ because the semiclassical expansion is really an expansion in powers of the dimensionless ratio $\hbar/S$ and this ultimately is small *because* of its proportionality to small dimensionless parameters.

Now comes the main point. Because $S_{\rm lin}$ is linear in $\tilde{\phi}^a$, its Feynman rule is as given in Fig. (2.2), which inserts a 'tadpole' contribution proportional to $(\delta S/\delta\varphi^a) + J_a$. But something wonderful happens once this is evaluated at $J_a = -\delta\Gamma/\delta\varphi^a$, since eq. (2.23) implies

$$\begin{aligned}\frac{\delta S}{\delta\varphi^a(x)} + J_a(x) &= \frac{\delta}{\delta\varphi^a(x)}\Big(S[\varphi] - \Gamma[\varphi]\Big) \\ &= -\frac{\delta}{\delta\varphi^a(x)}\left[\frac{\mathrm{i}}{2}\ln\det\Delta + (\text{2-loops and higher})\right] \\ &= -(\text{sum of tadpole graphs})\,, \end{aligned} \tag{2.25}$$

where 'tadpole graphs' mean all those graphs involving one or more loops having one dangling unconnected internal propagator that ends at the point x.

What is wonderful about this condition is it is easy to implement without having a detailed expression for $\Gamma[\varphi]$. The condition $J_a = -\delta\Gamma/\delta\varphi^a$ simply ensures that all graphs involving explicit dependence on J_a precisely cancel all graphs without J_a that are *1-particle reducible*: that is, all graphs that can be cut into two disconnected pieces by breaking a single internal line. (A graph that cannot be broken into two in this way is called *1-particle irreducible*, or 1PI.)

The upshot is very simple: $\Gamma[\varphi]$ is computed by calculating graphs that do not involve the vertex $S_{\rm lin}$ at all, but evaluating only those graphs that are 1-particle irreducible. For this reason $\Gamma[\varphi]$ is often called the generator of 1-particle irreducible correlations, or the *1PI action* for short. In the semiclassical expansion about an arbitrary configuration φ^a_j eq. (2.23) is evaluated as a sum of 1PI connected vacuum graphs (*i.e.* 1PI connected graphs having no external lines), obtained by dropping all disconnected and one-particle reducible graphs from the sum sketched in Fig. (2.1).

Recognizing that $\Gamma[\varphi]$ involves only 1PI graphs gives another way in which the quantity $\Gamma[\varphi]$ generalizes the classical action [17]. In a perturbative expansion the leading, classical, approximation corresponds to using the leading term of eq. (2.23): $\Gamma[\varphi] \simeq S[\varphi]$. As discussed in Appendix C, when applied to scattering amplitudes this amounts to just summing the tree graphs built from vertices coming from the classical interactions in $S_{\rm int}$. By contrast, imagine instead computing Feynman graphs using the expansion of $\Gamma[\varphi + \tilde{\phi}]$, rather than $S[\varphi + \tilde{\phi}]$, to generate the propagators and vertices. In this case the all-loops result for any correlation function constructed with Feynman rules built from $S[\varphi]$ is precisely the same as the quantity obtained by summing just tree graphs constructed from the Feynman rules built from $\Gamma[\varphi]$. In this sense, the full quantum amplitude is obtained by calculating using $\Gamma[\varphi]$ but working within the classical approximation. (It is this property that makes $\Gamma[\varphi]$ useful for discussing ultraviolet (UV) divergences, since the absence of new UV divergences in tree graphs makes it sufficient to renormalize divergences in $\Gamma[\varphi]$ in order to ensure all amplitudes are finite.)

Similar to the discussion for $Z[J]$, an important practical issue arises when φ is too complicated to allow an explicit calculation of the propagators $\Delta^{-1}_{ab} = -(\delta^2 S/\delta\phi^a\delta\phi^b)_{\phi=\varphi}$

used in the Feynman rules. This is often dealt with by expanding the result in powers of $\varphi^a(x)$ (or, more generally, in powers of the displacement of φ^a away from a sufficiently simple background for which Δ^{-1}_{ab} can be evaluated).

In this case using

$$\left(\frac{\delta^2 S}{\delta\phi^a(x)\,\delta\phi^b(y)}\right)_\varphi = \left(\frac{\delta^2 S}{\delta\phi^a(x)\,\delta\phi^b(y)}\right)_0 + \int \mathrm{d}^4z \left(\frac{\delta^3 S}{\delta\phi^a(x)\,\delta\phi^b(y)\,\delta\phi^c(z)}\right)_0 \varphi^c(z) + \cdots\,, \quad (2.26)$$

and $(\Delta_0 - \delta\Delta)^{-1} = \Delta_0^{-1}\sum_{n=0}^{\infty}(\delta\Delta\,\Delta_0^{-1})^n$ shows that the required Feynman graphs are obtained by inserting external lines in all possible ways (both to the internal lines and the vertices) in the 1PI vacuum graphs of Fig. 2.1, with the external lines representing the Feynman rule $\varphi^a(x)$.

2.2 The high-energy/low-energy split ◇

So far so good, but how can the above formalism be used to compute and use low-energy effective actions? The rest of this chapter specializes to theories having two very different intrinsic mass scales — like $m_l \ll m_h$ of the toy model in Chapter 1 — in order to address this question. After formalizing the split into low- and high-energy theory in this section, the following two sections identify two useful ways of defining a low-energy effective action.

2.2.1 Projecting onto low-energy states

The starting point, in this section, is to define more explicitly the split between low- and high-energy degrees of freedom. There are a variety of ways to achieve this split. Most directly, imagine dividing the quantum field ϕ^a into a low-energy and high-energy part: $\phi^a(x) = l^a(x) + \mathfrak{h}^a(x)$, where

$$l^a(x) := P_\Lambda \phi^a(x) P_\Lambda\,, \quad (2.27)$$

where $P_\Lambda = P_\Lambda^2$ denotes the projector onto states having energy $E < \Lambda$. To be of practical use, the scale Λ should lie somewhere between the two scales (such as m_l and m_h) that define the underlying hierarchy ($m_l \ll m_h$) in terms of which the low-energy limit is defined for the theory of interest.

This can be made more explicit in semiclassical perturbation theory, where $\phi^a = \varphi^a_{\rm cl} + \hat\phi^a$. Since in the interaction representation the quantum field satisfies the linearized field equation, $\Delta_{ab}\hat\phi^b = 0$, one can decompose $\hat\phi^a(x)$ in terms of a basis of eigenmodes, $u_p(x)$,

$$\hat\phi^a(x) = \sum_p \Big[\mathfrak{a}_p u_p^a(x) + \mathfrak{a}_p^* u_p^{a*}(x)\Big]\,. \quad (2.28)$$

For time-independent backgrounds these eigenmodes can be chosen to simultaneously diagonalize the energy, $i\partial_t u_p = \varepsilon_p\, u_p$, and so the low-energy part of the field is that part of

the sum for which the mode energies are smaller than the reference scale Λ. That is,

$$\hat{l}^a(x) := \sum_{\varepsilon_p<\Lambda} \left[\mathfrak{a}_p u_p^a(x) + \mathfrak{a}_p^* u_p^{a*}(x) \right], \tag{2.29}$$

and so

$$\hat{\mathfrak{h}}^a := \hat{\phi}^a - \hat{l}^a = \sum_{\varepsilon_p>\Lambda} \left[\mathfrak{a}_p u_p^a(x) + \mathfrak{a}_p^* u_p^{a*}(x) \right]. \tag{2.30}$$

It is natural at this point to worry that a division into high- and low-energy modes introduces an explicit frame-dependence into the problem. After all, a collision that appears to involve only low energies to one observer would appear to involve very high energies to another observer who moves very rapidly relative to the first one. Although this is true in principle, in practice frame-independent physical quantities (like the scattering amplitudes examined for the toy model in Chapter 1) only depend on invariant quantities like centre-of-mass energies, and all observers agree when these are large or small. For scattering calculations the natural split between low- and high-energies is therefore made in the centre-of-mass frame. The point is that in order to profit from the simplification of physics at low energies, it suffices that there exist *some* observers who see a process to be at low energies; it is not required that *all* observers do so.

Notice that correlation functions of low-energy states necessarily do not vary very quickly with time. This may be seen by inserting a complete set of intermediate energy eigenstates between any two pairs of low-energy fields, such as

$$\begin{aligned} {}_o\langle\Omega|\, \hat{l}^a(x)\, \hat{l}^b(y)\, |\Omega\rangle_i &= {}_o\langle\Omega|P_\Lambda \hat{\phi}^a(x) P_\Lambda^2 \hat{\phi}^b(y) P_\Lambda|\Omega\rangle_i \\ &= \sum_{\varepsilon_p<\Lambda} {}_o\langle\Omega|\hat{\phi}^a(x)|p\rangle\langle p\,|\hat{\phi}^b(y)|\Omega\rangle_i\,, \end{aligned} \tag{2.31}$$

which uses $P_\Lambda|\Omega\rangle = |\Omega\rangle$ and $P_\Lambda|p\rangle = |p\rangle$ for low-energy states while $P_\Lambda|p\rangle = 0$ for high-energy states. This result clearly has support only for frequencies $\omega = \varepsilon_p < \Lambda$. In relativistic and translation-invariant theories, for which low energy also means low momentum, the same argument shows that correlations also have slow spatial variation. Of course one might also implement a cutoff more smoothly, by weighting high-energy states in amplitudes by some suitably decreasing function of energy rather than completely cutting them off above Λ.

Example: The Toy Model

To make this concrete, consider the toy model of Chapter 1. In this case there are two quantum fields, $\hat{\phi}_l$ and $\hat{\phi}_R$ (or equivalently, $\hat{\xi}$ and $\hat{\chi}$) and the two intrinsic mass scales are $m_l = 0 \ll m_R$. The energy eigenmodes for these are labeled by 4-momentum, $u_p(x) \propto e^{ipx}$, and the linearized field equation $(-\Box + m^2)\,\hat{\phi} = 0$, relates the energy to the momentum by $q^0 = \epsilon_q = q$ for $\hat{\phi}_l$ and $p^0 = \varepsilon_p = \sqrt{p^2 + m_R^2}$ for $\hat{\phi}_R$.

In this case the useful choice is $m_l \ll \Lambda \ll m_R$, which is possible because of the hierarchy $m_l \ll m_R$. With this choice the light fields consist only of the long-wavelength modes of $\hat{\phi}_l$,

$$\hat{l}(x) = \sum_{\epsilon_q<\Lambda} \left[\mathfrak{c}_q\, e^{\mathrm{i}qx} + \mathfrak{c}_q^*\, e^{-\mathrm{i}qx} \right], \tag{2.32}$$

while the heavy fields contain all of the $\hat{\phi}_R$ modes together with the short-wavelength modes of $\hat{\phi}_l$:

$$\hat{\mathfrak{h}}_l(x) = \sum_{\epsilon_q > \Lambda} \left[\mathfrak{c}_q \, e^{\mathrm{i}qx} + \mathfrak{c}_q^* \, e^{-\mathrm{i}qx} \right] \quad \text{and} \quad \hat{\mathfrak{h}}_R(x) = \sum_p \left[\mathfrak{b}_p \, e^{\mathrm{i}px} + \mathfrak{b}_p^* \, e^{-\mathrm{i}px} \right]. \tag{2.33}$$

2.2.2 Generators of low-energy correlations ♠

The next step is to seek generating functionals specifically for low-energy correlation functions, and to investigate their properties. The key tool for this purpose is the observation made above that the correlation functions themselves can vary only over time and length scales larger than Λ^{-1}.

Imagine now defining the generating functional, $Z_{\text{le}}[J]$, for the time-ordered in-out correlations of only the low-energy fields, $\hat{l}^a(x)$. This can be done simply by restricting the definition, eq. (2.2), of $Z[J]$, to include only correlation functions that vary slowly in space and time (*i.e.* only over scales larger than Λ^{-1}), leading to the result

$$Z_{\text{le}}[J] := \sum_{n=0}^{\infty} \frac{\mathrm{i}^n}{n!} \int \mathrm{d}^4x_1 \cdots \mathrm{d}^4x_n G_{\text{le}}^{a_1 \cdots a_n}(x_1, \cdots, x_n) J_{a_1}(x_1) \cdots J_{a_n}(x_n) \,, \tag{2.34}$$

Because the low-energy correlation functions only vary slowly in space and time, the same is true of any currents, $J_a(x)$, appearing in $Z_{\text{le}}[J]$. That is, if the current is split into long- and short-wavelength Fourier modes, $J_a(x) = j_a(x) + \mathcal{J}_a(x)$, with

$$j_a(x) = \sum_{\text{slowly varying}} j_a(p) \, e^{\mathrm{i}px} \,, \tag{2.35}$$

then the generating functional for low-energy correlations, $Z_{\text{le}}[J]$, is simply the restriction of the full generating functional to slowly varying currents:

$$Z_{\text{le}}[j] = Z[j, \mathcal{J} = 0] \,. \tag{2.36}$$

Here the precise definition of 'slowly varying' in eq. (2.35) depends on the details of the particle masses and the way the cutoff Λ is implemented — *c.f.* eq. (2.29) for example — for the quantum states.

It might seem bothersome that the generating functionals for low-energy correlation functions depend explicitly on the value of Λ, as well as on all of the details of precisely how the high-energy modes are cut off. One of the tasks of later chapters is to show how this dependence can be absorbed into appropriate renormalizations of effective couplings, so that predictions for physical processes (like scattering amplitudes) only depend on physical mass scales like m_R (rather than Λ or other definitional details).

For relativistic, translationally invariant theories a slightly more convenient way to break Fourier modes into slowly and quickly varying parts is to Wick rotate [18] to euclidean signature, $\{x^0, \mathbf{x}\} \to \{\mathrm{i}x^4, \mathbf{x}\}$. In this case the time-components of any 4-vector must be similarly rotated, so the invariant inner product of two 4-vectors becomes

$$p \cdot q = p_\mu q^\mu = -p^0 q^0 + \mathbf{p} \cdot \mathbf{q} \to +p^4 q^4 + \mathbf{p} \cdot \mathbf{q} = (p \cdot q)_E \,. \tag{2.37}$$

This ensures that the invariant condition $p_\mu p^\mu = (p^4)^2 + \mathbf{p}^2 < \Lambda^2$ excludes large values of

both $|\mathbf{p}|$ and p^4 (unlike for Minkowski signature, where $p_\mu p^\mu = -(p^0)^2 + \mathbf{p}^2 < \Lambda^2$ allows both $|\mathbf{p}|$ and p^0 to be arbitrarily large but close to the light cone).

The generator, $W_{\rm le}[j]$, of low-energy connected correlations can be defined as before, by taking the logarithm of $Z_{\rm le}[j]$, leading to the path integral representation

$$\exp\Big\{\mathrm{i}W_{\rm le}[j]\Big\} = \int \mathcal{D}\phi\ \exp\left\{\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^4x\,\phi^a j_a\right\}. \tag{2.38}$$

The main difference between this and the expression for $W[J]$ is the absence of currents coupled to the high-frequency components of ϕ^a. That is, if $\phi^a = l^a + \mathfrak{h}^a$ is split into slowly varying ('light', l^a) and rapidly varying ('heavy', $\mathfrak{h}^a$) parts, along the same lines as eq. (2.35) for J_a, then eq. (2.38) becomes

$$\exp\Big\{\mathrm{i}W_{\rm le}[j]\Big\} = \int \mathcal{D}l\mathcal{D}\mathfrak{h}\ \exp\left\{\mathrm{i}S[l+\mathfrak{h}] + \mathrm{i}\int \mathrm{d}^4x\,l^a j_a\right\}. \tag{2.39}$$

Physically, this states that a restriction to low-energy correlations can be obtained simply by restricting oneself only to probing the system with slowly varying currents.

2.2.3 The 1LPI action

At this point it is hard to stop from performing a Legendre transformation to obtain the generating functional, $\Gamma_{\rm le}[\ell]$, directly in terms of the low-energy field configurations, ℓ^a, rather than j_a. To this end define

$$\Gamma_{\rm le}[\ell] := W_{\rm le}[j] - \int \mathrm{d}^4x\,\ell^a j_a\,, \tag{2.40}$$

with $j_a = j_a[\ell]$ regarded as a functional of ℓ^a found by inverting the relation $\ell^a = \ell^a[j]$ obtained from

$$\ell^a := \frac{\delta W_{\rm le}}{\delta j_a}\,, \tag{2.41}$$

with the result — *c.f.* eq. (2.18)

$$j_a = -\frac{\delta\Gamma_{\rm le}}{\delta\ell^a}\,. \tag{2.42}$$

It is important to realize that although $\Gamma_{\rm le}[\ell]$ obtained in this way only has support on slowly varying field configurations, $\ell^a(x)$, it is *not* simply the restriction of $\Gamma[\varphi\,] = \Gamma[\ell, h]$ to long-wavelength configurations: $h^a := \delta W/\delta\mathcal{J}_a = 0$. To see why not, consider its path-integral representation:

$$\begin{aligned}\exp\Big\{\mathrm{i}\Gamma_{\rm le}[\ell]\Big\} &= \int \mathcal{D}l\,\mathcal{D}\mathfrak{h}\ \exp\left\{\mathrm{i}S[l,\mathfrak{h}] + \mathrm{i}\int \mathrm{d}^4x\,(l^a-\ell^a)j_a\right\}\\ &= \int \mathcal{D}\tilde{l}\,\mathcal{D}\mathfrak{h}\ \exp\left\{\mathrm{i}S[\ell+\tilde{l},\mathfrak{h}] + \mathrm{i}\int \mathrm{d}^4x\,\tilde{l}^a j_a\right\}.\end{aligned} \tag{2.43}$$

For comparison, the earlier result, eq. (2.4), for $\Gamma[\varphi\,] = \Gamma[\ell, h]$ is

$$\exp\Big\{\mathrm{i}\Gamma_{\rm le}[\ell,h]\Big\} = \int \mathcal{D}\tilde{l}\,\mathcal{D}\tilde{\mathfrak{h}}\ \exp\left\{\mathrm{i}S[\ell+\tilde{l},h+\tilde{\mathfrak{h}}] + \mathrm{i}\int \mathrm{d}^4x\,(\tilde{l}^a j_a + \tilde{\mathfrak{h}}^a\mathcal{J}_a)\right\}, \tag{2.44}$$

The key point is that the condition $h = 0$ is *not* generically equivalent to the condition $\mathcal{J}_a(x) = 0$ that relates $\Gamma_{\rm le}[\ell]$ to $\Gamma[\ell, h]$. Instead, the condition $\mathcal{J}_a = 0$ states that the short-wavelength part of the field should be chosen as that configuration, $h^a = h^a_{\rm le}(\ell)$, that satisfies

$$\left(\frac{\delta\Gamma}{\delta h^a}\right)_{h=h_{\rm le}(\ell)} = 0\,. \tag{2.45}$$

In particular, the vanishing of $\mathcal{J}_a$ means that the short-wavelength components of the current are not available to take the values $\mathcal{J}_a = -\delta\Gamma/\delta h^a$ they would have taken in the Legendre transform of $W[J] = W[j, \mathcal{J}]$. They are therefore not able to cancel the one-particle-reducible graphs that can be broken in two by cutting a single $\hat{\mathfrak{h}}^a$ line. The quantity $\Gamma_{\rm le}[\varphi]$ is therefore given as sum of *one-light -particle irreducible* (or 1LPI) graphs, that are only irreducible in the sense that they cannot be broken into two disconnected pieces by cutting a *light* -particle, $\hat{l}^a$, line.

Example: The Toy Model

How does all this look in the toy model of Chapter 1? In this case, with Λ chosen to satisfy $m_l \ll \Lambda \ll m_h$, the 'light' fields consist only of the low-energy modes of the massless field, ξ (or $\hat{\phi}_l$), and the 'heavy' fields consist of both the high-energy modes of ξ together with all of the modes of the massive field χ (or $\hat{\phi}_h$). The 1LPI generator of low-energy connected correlation functions then is

$$\exp\Big\{\mathrm{i}\Gamma_{\rm le}[\xi]\Big\} = \int \mathcal{D}\tilde{\xi}\mathcal{D}\tilde{\chi}\ \exp\left\{\mathrm{i}S\,[\xi+\tilde{\xi},\tilde{\chi}] + \mathrm{i}\int \mathrm{d}^4x\,\tilde{\xi}^a\,j_a\right\}\,, \tag{2.46}$$

with ξ^a and $j_a = -\delta\Gamma_{\rm le}/\delta\xi^a$ only varying over times and distances longer than Λ^{-1}.

Recall that small λ controls a semiclassical expansion, and imagine computing $\Gamma_{\rm le}[\xi]$ in the leading, classical approximation. As argued earlier (and elaborated in §3 below), in this limit the full 1PI generator reduces to the classical action: $\Gamma[\xi,\chi] \simeq S\,[\xi,\chi]$, explicitly given in eq. (1.24),

$$S\,[\xi,\chi] = -\int \mathrm{d}^4x\,\left[\frac{1}{2}\,\partial_\mu\chi\partial^\mu\chi + \frac{1}{2}\left(1+\frac{\chi}{\sqrt{2}\,v}\right)^2\partial_\mu\xi\partial^\mu\xi + V(\chi)\right]\,, \tag{2.47}$$

with

$$V(\chi) = \frac{m_h^2}{2}\,\chi^2 + \frac{\lambda\,v}{2\,\sqrt{2}}\,\chi^3 + \frac{\lambda}{16}\,\chi^4\,. \tag{2.48}$$

In general the above arguments say that $\Gamma_{\rm le}[\xi] = \Gamma_{\rm le}[\xi,\chi_{\rm le}(\xi)]$, where $\chi_{\rm le}(\xi)$ is obtained by solving $\delta\Gamma[\xi,\chi]/\delta\chi = 0$ (*c.f.* eq. (2.45)). So in the classical approximation $\Gamma_{\rm le}[\xi] \simeq S\,[\xi,\chi_{\rm le}(\xi)]$, where eq. (2.45) in the classical approximation says $\chi_{\rm le}(\xi)$ is found by solving the classical field equation

$$\left(-\Box + m_h^2 + \frac{1}{2v^2}\,\partial_\mu\xi\partial^\mu\xi\right)\chi_{\rm le} = -\frac{1}{\sqrt{2}\,v}\,\partial_\mu\xi\partial^\mu\xi - \frac{3\lambda v}{2\,\sqrt{2}}\,\chi_{\rm le}^2 - \frac{\lambda}{4}\,\chi_{\rm le}^3\,. \tag{2.49}$$

Using this in the classical action leads (after an integration by parts) to

$$S[\xi,\chi_{\rm le}(\xi)] = \int {\rm d}^4x \left[-\frac{1}{2}\left(1+\frac{\chi_{\rm le}}{\sqrt{2}\,v}\right)\partial_\mu\xi\partial^\mu\xi + \frac{\lambda v}{4\sqrt{2}}\chi_{\rm le}^3 + \frac{\lambda}{16}\chi_{\rm le}^4\right], \tag{2.50}$$

where $\chi_{\rm le}(\xi)$ is to be interpreted as the function of ξ obtained by solving (2.49).

To proceed further expand the solution $\chi_{\rm le}$ in powers of $\partial_\mu\xi\partial^\mu\xi$ and $\Box$, using

$$(-\Box + m_R^2 + X)^{-1} \simeq \frac{1}{m_R^2} - \frac{(-\Box + X)}{m_R^4} + \cdots, \tag{2.51}$$

and so on. This gives

$$\begin{aligned}\chi_{\rm le} \simeq & -\frac{1}{\sqrt{2}\,v m_R^2}\left(1+\frac{\Box}{m_R^2}+\frac{\Box^2}{m_R^4}+\cdots\right)(\partial_\mu\xi\partial^\mu\xi) \\ & -\frac{1}{4\sqrt{2}\,v^3 m_R^4}\left(1+\frac{\Box}{m_R^2}+\cdots\right)(\partial_\mu\xi\partial^\mu\xi)^2+\cdots,\end{aligned} \tag{2.52}$$

where the ellipses involve terms with more powers of $\Box$ and/or more powers of $(\partial_\mu\xi\partial^\mu\xi)$.

Combining results then gives

$$\begin{aligned}\Gamma_{\rm le}[\xi] \simeq \int {\rm d}^4x \, \Big[& -\frac{1}{2}\,\partial_\mu\xi\partial^\mu\xi + a\,\big(\partial_\mu\xi\partial^\mu\xi\big)\Big(1+\frac{\Box}{m_R^2}\Big)\big(\partial_\mu\xi\partial^\mu\xi\big) \\ & + b\,(\partial_\mu\xi\partial^\mu\xi)^3+\cdots\Big],\end{aligned} \tag{2.53}$$

where again ellipses include terms with more powers of $\Box$ and/or $\partial_\mu\xi\partial^\mu\xi$ than those shown. In the classical approximation used here, the coefficients a and b evaluate to

$$a = \frac{\lambda}{4m_R^4} = \frac{1}{4\lambda v^4} \quad \text{and} \quad b = \left(\frac{1}{16}-\frac{1}{16}\right)\frac{\lambda^2}{m_R^8} = 0\,. \tag{2.54}$$

In particular, b vanishes due to a cancelation between the $\chi_{\rm le}(\partial_\mu\xi\partial^\mu\xi)$ and $\chi_{\rm le}^3$ terms, while $a = \lambda/4m_R^4$ precisely agrees with eq. (1.12) (and is proportional to λ^{-1} when expressed in terms of v, as expected for a classical contribution). As shown earlier, when used in the classical – no loop – approximation, this value ensures that the interactions quadratic in $\partial_\mu\xi\partial^\mu\xi$ accurately reproduce the first two terms of the low-energy limit for $\xi\xi \to \xi\xi$ scattering found earlier.

Of course Feynman graphs greatly simplify calculations such as these, particularly once one progresses beyond the leading classical approximation, and it is useful to see how these reproduce the above calculation. To this end imagine working with the Feynman rules described around eq. (2.26), in which one perturbs in powers of the background field (in this case ξ). As shown in the next chapter, at the classical level one seeks tree (*i.e.* no loops) graphs whose external lines correspond to the fields appearing in the appropriate effective interaction. Since $\Gamma_{\rm le}[\xi]$ is 1LPI at tree level only heavy states can appear as internal lines.

For instance, for the calculation performed above the term quadratic in $\partial_\mu\xi\,\partial^\mu\xi$ arises from the Feynman graph shown in panel (a) of Fig. 2.3,[6] once the heavy-particle propagator

[6] Note that no combinatorial factors are in this case necessary for external lines – in contrast to the scattering evaluation of Fig. 1.3 – because external lines here represent factors of $\partial_\mu\xi$ all evaluated at the same position, rather than scattering states with distinguishable momenta.

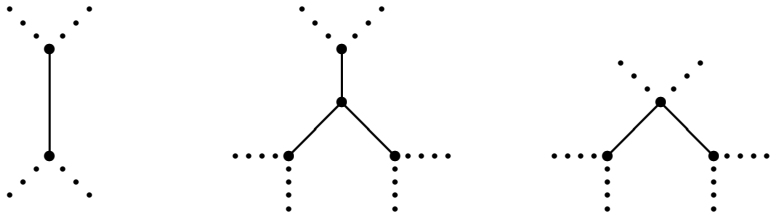


Fig. 2.3 The tree graphs that dominate the $(\partial_\mu \xi\, \partial^\mu \xi)^2$ (panel a) and the $(\partial_\mu \xi\, \partial^\mu \xi)^3$ (panels b and c) effective interactions. Solid lines represent χ propagators while dotted lines denote external ξ fields.

is expanded in powers of $\Box/m_R^2$. The result found in this way for the parameter a is

$$\mathrm{i}a_{\text{graph (a)}} = \left[\left(\frac{\mathrm{i}^2}{2!}\right)\right]\left(-\frac{1}{\sqrt{2}\,v}\right)^2\left(\frac{-\mathrm{i}}{m_R^2}\right) = \frac{\mathrm{i}}{4v^2 m_R^2}\,, \tag{2.55}$$

where the square bracket contains the numerical factors from the expansion of $\exp[\mathrm{i}S_{\text{int}}]$, the next factor is the coupling constant Feynman rule for the vertices and the final factor is the leading contribution from the propagator. The left-hand side is obtained by expanding $\exp\{\mathrm{i}\Gamma[\xi]\}$ using (2.53) and identifying the coefficient of the $(\partial_\mu \xi\, \partial^\mu \xi)^2$ term. This result for a agrees with the one found in (2.54) by eliminating $\chi_{\text{le}}(\xi)$ from the action, and in (1.16) by demanding the correct result for low-energy ξ-particle scattering.

The contribution cubic in $\partial_\mu \xi\, \partial^\mu \xi$ is similarly obtained from the graphs in panels (b) and (c) of Fig. 2.3, which contribute the following to the effective coupling b:

$$\begin{aligned}
\mathrm{i}b_{\text{graph (b)}} &= 6\left[4\left(\frac{\mathrm{i}^4}{4!}\right)\right]\left(-\frac{1}{\sqrt{2}\,v}\right)^3\left(-\frac{\lambda v}{2\sqrt{2}}\right)\left(\frac{-\mathrm{i}}{m_R^2}\right)^3 = \frac{\mathrm{i}\lambda}{8v^2 m_R^6} \\
\mathrm{i}b_{\text{graph (c)}} &= 2\left[3\left(\frac{\mathrm{i}^3}{3!}\right)\right]\left(-\frac{1}{\sqrt{2}\,v}\right)^2\left(-\frac{1}{4v^2}\right)\left(\frac{-\mathrm{i}}{m_R^2}\right)^2 = -\frac{\mathrm{i}}{8v^4 m_R^4}\,.
\end{aligned} \tag{2.56}$$

Starting with the square bracket, each factor here has the same origin as its counterpart in eq. (2.55) for panel (a), and the first number arises as the number of ways of connecting the relevant external lines and vertices into the given graph. These again cancel once the relation $m_R^2 = \lambda v^2$ is used, giving the tree-level prediction $b = 0$.

The above arguments make clearer why — if low-energy observables are the only things of interest — the physics of the two fields in the toy model can be traded for a collection of effective interactions involving only the light particle: it suffices to know $\Gamma_{\text{le}}[\ell]$ to capture all low-energy physical applications. What remains is to find a way to do so more efficiently than by first computing the full generating functional $\Gamma[\ell, h\,]$, and then eliminating $h = h_{\text{le}}(\ell)$ to obtain $\Gamma_{\text{le}}[\ell] = \Gamma[\ell, h_{\text{le}}(\ell)]$.

2.3 The Wilson action ◇

Although the low-energy 1LPI generator captures all of a theory's low-energy observables, what remains elusive is an efficient means for capturing, *as early in a calculation as possible*, how high-energy physics appears in low-energy processes. The main tool for doing this — the *Wilson action*[7] — is topic of this section.

2.3.1 Definitions

A good starting point for describing the Wilson action is the path-integral expression for the 1LPI generator, eq. (2.43):

$$\exp\Big\{\mathrm{i}\Gamma_{\rm le}[\ell]\Big\} = \int \mathcal{D}\tilde{l}\,\mathcal{D}\mathfrak{h}\;\exp\left\{\mathrm{i}S\left[\ell+\tilde{l}+\mathfrak{h}\right]+\mathrm{i}\int \mathrm{d}^4x\,\tilde{l}^a j_a\right\}. \tag{2.57}$$

What is noteworthy about this expression is that — because the currents are chosen to explore only low-energy quantities — the heavy field, $\mathfrak{h}^a$, appears only in the classical action and not in the current term. As a consequence, all possible low-energy influences of the heavy field must be captured in the quantity

$$\exp\Big\{\mathrm{i}S_w[l]\Big\} := \int \mathcal{D}\mathfrak{h}\;\exp\Big\{\mathrm{i}S\left[l+\mathfrak{h}\right]\Big\}, \tag{2.58}$$

in terms of which the full 1LPI action is given by

$$\exp\Big\{\mathrm{i}\Gamma_{\rm le}[\ell]\Big\} = \int \mathcal{D}\tilde{l}\;\exp\left\{\mathrm{i}S_w[\ell+\tilde{l}]+\mathrm{i}\int \mathrm{d}^4x\,\tilde{l}^a j_a\right\}. \tag{2.59}$$

Eq. (2.58) defines the Wilson action, obtained by *integrating out* all heavy degrees of freedom having energies above the scale Λ. It has several noteworthy features, that are explored in detail throughout the rest of the book.

- As the definition shows, the Wilson action provides the earliest place in a calculation to systematically identify, once and for all, the low-energy influence of the heavy degrees of freedom $\mathfrak{h}$. Best of all, this can be done in one fell swoop, before choosing precisely which observable or correlation function is of interest in a particular application.
- For practical applications, most real interest is on obtaining the Wilson action as a series expansion in inverse powers of the heavy mass scales in the problem of interest. As shall be seen in some detail, at any fixed order in this expansion the Wilson action is a local functional, $S_w = \int \mathrm{d}^4x\;\mathfrak{L}_w(x)$, with $\mathfrak{L}_w(x)$ being a function of the light fields and their derivatives all evaluated at the same spacetime point.
- What is striking about eq. (2.59) is that the Wilson action, S_w, appears in the expression for the generator, $\Gamma_{\rm le}$, of low-energy correlators, in precisely the way that the classical action, S, appears in the expression, eq. (2.19), for the generator, Γ, of generic correlators. This suggests that the classical action of the full theory might itself be better regarded as the Wilson action from some even higher-energy theory.

[7] Named for Ken Wilson, a pioneer in the development of renormalization techniques (see the brief historical notes in the Bibliography).

- Eq. (2.58) shows that S_w depends in detail on things like Λ and precisely how the split is made between the high- and low-energy sectors, since these are buried in the definitions of the split between $\mathfrak{h}^a$ and l^a. So it is misleading to speak about 'the' Wilson action, rather than 'a' Wilson action. Yet we know that Λ cannot appear in any physical observables, because it is just an arbitrary artificial scale that is introduced for calculational convenience. Part of the story to follow must therefore be why all these calculational details in S_w drop out of observables. The outlines of this argument are already clear in eq. (2.59), which shows that the Λ dependence introduced by performing the integration over $\mathcal{D}\mathfrak{h}$ is later canceled when integrating over the rest of the fields, $\mathcal{D}l$, since the total measure $\mathcal{D}\phi = \mathcal{D}l\,\mathcal{D}\mathfrak{h}$ is Λ-independent.

In semiclassical perturbation theory, the arguments of earlier sections show that eq. (2.58) gives S_w as the sum over all connected vacuum graphs — not just 1PI graphs, say — using Feynman rules computed for the 'high-energy' fields with the 'low-energy' fields regarded as fixed background values. (Recall in this split that high-energy fields can include the high-energy modes of particles with small masses.) Eq. (2.59) then says to construct Feynman graphs using propagators and vertices for the light fields defined from S_w, with $\Gamma_{\rm le}$ then obtained by computing all 1PI graphs in this low-energy theory. This combination reproduces the set of 1LPI graphs using the Feynman rules of the full theory.

In particular, since any tree graph with an internal line is one-particle reducible, this means that $\Gamma_{\rm le}[\ell] \simeq S_w[\ell]$ within the classical approximation (no loops). Furthermore, in the same approximation both are related to the classical action of the full theory by

$$\Gamma_{\rm le}[\ell] \simeq S_w[\ell] \simeq S[\ell, h_{\rm le}(\ell)] \qquad \text{(classical approximation)}\,, \tag{2.60}$$

where $h^a = h^a_{\rm le}(\ell)$ is obtained (in the classical approximation) by solving $(\delta S/\delta h^a)_{h=h_{\rm le}(\ell)} = 0$. But — as is seen more explicitly below — $\Gamma_{\rm le}[\ell]$ and $S_w[\ell]$ need no longer agree once loops are included.

It is the Wilson action that is the main tool used in the rest of this book. But why bother with S_w, given that $\Gamma_{\rm le}$ also captures all of the information relevant for low-energy predictions? As later examples show in more detail, in real applications it is often the Wilson action that is the easier to use, since it exploits the simplicity of the low-energy limit as early as possible. It plays such a central role because it has two very attractive properties.

First, it contains enough information to be useful. That is, any low-energy observable can be constructed from low-energy correlation functions (and so also from $\Gamma_{\rm le}$), and because $\Gamma_{\rm le}$ can be computed from S_w using only low-energy degrees of freedom, it follows that S_w carries all of the information necessary to extract the predictions for any low-energy observable.

But it is the second property that makes it such a practical tool: it doesn't contain too much information. That is, the Wilson action is the bare-bones quantity that contains all of the information about the system's high-energy degrees of freedom without polluting it with any low-energy details. Unlike $\Gamma_{\rm le}$, the Wilson action is constructed by integrating only over the high-energy degrees of freedom. This means that there is maximal labour savings in exploiting any simplicity in S_w, since this simplicity is present *before* performing the rest of the low-energy part of the calculation.

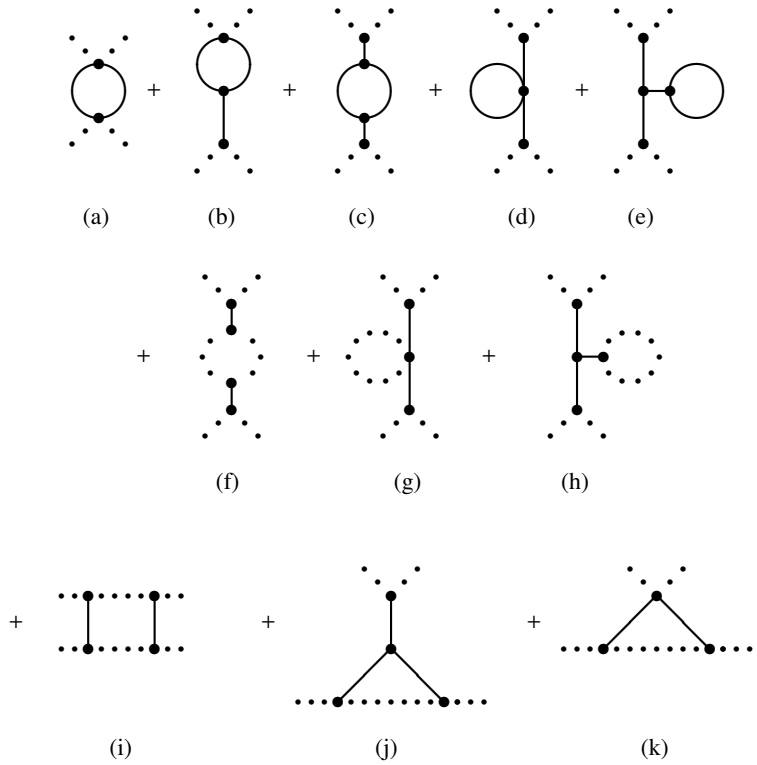


Fig. 2.4 One-loop graphs that contribute to the $(\partial_\mu \xi \partial^\mu \xi)^2$ interaction in the Wilson and 1LPI actions using the interactions of eqs. (1.24) and (1.25). Solid (dotted) lines represent χ (and ξ) fields. Graphs involving wave-function renormalizations of ξ are not included in this list.

Example: the Toy Model

To better understand how the Wilson action is defined, and is related to the low-energy 1LPI generator, it is useful to have a concrete example to examine in detail. Once again the toy model of Chapter 1 provides a useful place to start.

Since $\Gamma_{\rm le}$ and S_w only begin to differ beyond the classical approximation, imagine computing both $\Gamma_{\rm le}$ and S_w at one loop. According to its definition, the Feynman graphs contributing to S_w can involve only the high-energy degrees of freedom in the internal lines, while those contributing to $\Gamma_{\rm le}$ also involve virtual low-energy states. For both S_w and $\Gamma_{\rm le}$ the graphs can be one-particle reducible when cutting heavy-particle lines, but for $\Gamma_{\rm le}$ the graphs must be one-particle irreducible when light-particle lines are cut.

For concreteness' sake, for the toy model consider the one-loop contributions to the effective interaction

$$a \int \mathrm{d}^4 x \, (\partial_\mu \xi \partial^\mu \xi)^2 \,, \tag{2.61}$$

in both $\Gamma_{\rm le}$ and $S_{\scriptscriptstyle W}$. The relevant Feynman graphs are shown in Fig. 2.4, using Feynman rules appropriate for the χ and ξ fields using the interactions given in eqs. (1.24) and (1.25). (An equivalent set of graphs could also be written for the variables $\hat\phi_{\scriptscriptstyle R}$ and $\hat\phi_{\scriptscriptstyle I}$. Although both ultimately give the same physical results, they can differ in intermediate steps and which is more useful depends on the application one has in mind.)

Since all of the internal lines for Feynman graphs (a) through (e) involve only the field χ, and since all modes of this field are classified as 'high-energy' — *c.f.* the discussion in §2.2.1 — all five of these graphs contribute to both $S_{\scriptscriptstyle W}$ and $\Gamma_{\rm le}$. In order of magnitude, each contributes to the effective interaction an amount

$$a_{\rm graph(a)} \propto \left(-\frac{1}{4v^2}\right)^2 L_1 = \frac{L_1}{16\,v^4}\,, \tag{2.62}$$

$$a_{\rm graph(b)} \propto \left(-\frac{1}{4v^2}\right)\left(-\frac{\lambda v}{2\sqrt{2}}\right)\left(-\frac{1}{\sqrt{2}\,v}\right)\frac{L_1}{m_{\scriptscriptstyle R}^2} = -\frac{L_1}{16\,v^4}\,, \tag{2.63}$$

$$a_{\rm graph(c)} \propto \left(-\frac{\lambda v}{2\sqrt{2}}\right)^2\left(-\frac{1}{\sqrt{2}\,v}\right)^2\frac{L_1}{m_{\scriptscriptstyle R}^4} = \frac{L_1}{16\,v^4}\,, \tag{2.64}$$

$$a_{\rm graph(d)} \propto \left(-\frac{\lambda}{16}\right)\left(-\frac{1}{\sqrt{2}\,v}\right)^2\frac{L_2}{m_{\scriptscriptstyle R}^4} = -\frac{L_2}{32\lambda v^6}\,, \tag{2.65}$$

and

$$a_{\rm graph(e)} \propto \left(-\frac{\lambda v}{2\sqrt{2}}\right)^2\left(-\frac{1}{\sqrt{2}\,v}\right)^2\frac{L_2}{m_{\scriptscriptstyle R}^6} = \frac{L_2}{16\lambda v^6}\,, \tag{2.66}$$

where the powers of $1/m_{\scriptscriptstyle R}^2$ come from the internal lines that do not appear within a loop, since these are evaluated at momenta much smaller than $m_{\scriptscriptstyle R}$. The contribution of the loop integrals themselves are of order

$$\begin{aligned} L_1 &= \int^{\Omega}\frac{{\rm d}^4p}{(2\pi)^4}\left(\frac{1}{p^2+m_{\scriptscriptstyle R}^2}\right)^2 \propto \frac{1}{16\pi^2}\,\ln\left(\frac{\Omega}{m_{\scriptscriptstyle R}}\right) \\ \text{and}\quad L_2 &= \int^{\Omega}\frac{{\rm d}^4p}{(2\pi)^4}\left(\frac{1}{p^2+m_{\scriptscriptstyle R}^2}\right) \propto \frac{\Omega^2}{16\pi^2}\,. \end{aligned} \tag{2.67}$$

Here $\Omega \gg m_{\scriptscriptstyle R}$ is a cutoff that is introduced because the loop in the full theory is UV divergent. This divergence is ultimately dealt with by renormalizing the couplings of the microscopic theory; a point to be returned to in more detail shortly.

For the present purposes — keeping in mind that $m_{\scriptscriptstyle R}^2 = \lambda\, v^2$ — what is important is that graphs (a) through (c) clearly contribute to the coupling a (in both $S_{\scriptscriptstyle W}$ and $\Gamma_{\rm le}$) an amount of order $a_{1-\rm loop} \propto L_1/v^4 \propto (1/4\pi v^2)^2 \ln\left(\Omega^2/m_{\scriptscriptstyle R}^2\right)$. Graphs (d) and (e) instead contribute an amount of order $a_{1-\rm loop} \propto (1/4\pi v^2)^2\left(\Omega^2/m_{\scriptscriptstyle R}^2\right)$. Once the UV divergent function of $\Omega/m_{\scriptscriptstyle R}$ is renormalized into an appropriate coupling, the remaining coefficient for each of these loop contributions is suppressed by a factor of $\lambda/16\pi^2$ relative to the tree-level result, $a_{\rm tree} = 1/(4\lambda\, v^4)$, in agreement with the discussion surrounding eq. (2.24). As such they all contribute to the coefficient a_2 of eq. (1.16).

The difference between S_w and $\Gamma_{\rm le}$ arises in graphs (f) through (k), with S_w only receiving contributions where the momentum in the internal ξ propagators is larger than Λ, whereas there is no such a restriction for $\Gamma_{\rm le}$. The contribution from graphs (i) through (k) when all loop momenta are large is of order

$$\begin{aligned} a_{\text{graph(i)}} &\propto \left(-\frac{1}{\sqrt{2}\,v}\right)^4 L_1 = \frac{L_1}{4v^4} \\ a_{\text{graph(j)}} &\propto \left(-\frac{1}{\sqrt{2}\,v}\right)^3 \left(-\frac{\lambda v}{2\sqrt{2}}\right) \frac{L_1}{m_R^2} = \frac{L_1}{8v^4} \\ a_{\text{graph(k)}} &\propto \left(-\frac{1}{\sqrt{2}\,v}\right)^2 \left(-\frac{1}{4v^2}\right) L_1 = -\frac{L_1}{8v^4}\,, \end{aligned} \tag{2.68}$$

where the new loop integrals are also logarithmically divergent in the UV, and so up to numerical factors are again or order L_1 in size. These contribute to a an amount comparable to the size of graphs (a) through (c). By contrast, graphs (f) through (h) give the results,

$$\begin{aligned} a_{\text{graph(f)}} &\propto \left(-\frac{1}{\sqrt{2}\,v}\right)^4 \frac{L_3}{m_R^4} = \frac{L_3}{4v^4 m_R^4} \\ a_{\text{graph(g)}} &\propto \left(-\frac{1}{\sqrt{2}\,v}\right)^2 \left(-\frac{1}{4v^2}\right) \frac{L_3}{m_R^4} = -\frac{L_3}{8v^4 m_R^4} \\ a_{\text{graph(h)}} &\propto \left(-\frac{1}{\sqrt{2}\,v}\right)^3 \left(-\frac{\lambda v}{2\sqrt{2}}\right) \frac{L_3}{m_R^6} = \frac{L_3}{8v^4 m_R^4}\,, \end{aligned} \tag{2.69}$$

and so are of order $(1/v^4)(L_3/m_R^4)$ where the ξ loop has the ultraviolet behaviour

$$L_3 = \int^{\Omega} \frac{{\rm d}^4 p}{(2\pi)^4} \left(\frac{p^2}{p^2}\right)^k \propto \frac{\Omega^4}{16\pi^2}\,, \tag{2.70}$$

where $k = 2$ for graph (f) and $k = 1$ for graphs (g) and (h).

All of these graphs are dominated by large momenta (small wavelengths), which is why they diverge for large Ω. Although a more systematic treatment of these UV divergences (in particular how to treat them using dimensional regularization) is given in chapter 3, there is a conceptual point to be made concerning their *lower* limit of integration. The point is that for graphs (f) through (k) this lower limit differs when computing $\Gamma_{\rm le}$ and S_w. For $\Gamma_{\rm le}$ the contributions to the effective interaction

$$a_{\rm le} \int {\rm d}^4 x\, (\partial_\mu \xi \partial^\mu \xi)^2 \subset \Gamma_{\rm le}\,, \tag{2.71}$$

integrate over all momenta. But for S_w, in the contribution to

$$a_w \int {\rm d}^4 x\, (\partial_\mu \xi \partial^\mu \xi)^2 \subset S_w\,, \tag{2.72}$$

the integrations exclude momenta smaller than Λ (for which the internal ξ propagators are then 'light' degrees of freedom) that are not integrated out in the path-integral representation of S_w.

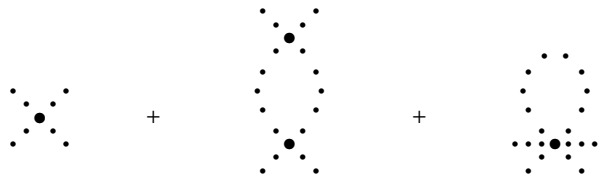


Fig. 2.5 The tree and one-loop graphs that contribute to the $(\partial_\mu \xi\, \partial^\mu \xi)^2$ interaction in the 1LPI action, using Feynman rules built from the Wilson action. All dotted lines represent ξ particles, and the 'crossed' versions of (b) are not drawn explicitly.

Take, for instance, graphs (i) through (k) of Fig. (2.4). Since $\Lambda \ll m_R$, the predicted coefficient for $\Gamma_{\rm le}$ differs from the coefficient in S_W by an amount of order

$$a_{\rm le} - a_W(\Lambda) \simeq \frac{1}{v^4} \int_0^\Lambda \frac{{\rm d}^4 p}{(2\pi)^4} \left(\frac{1}{p^2 + m_R^2}\right)^2 \left(\frac{p^2}{p^2}\right)^k \propto \frac{1}{16\pi^2 v^4} \left(\frac{\Lambda^4}{m_R^4}\right). \tag{2.73}$$

The suppression by powers of Λ/m_R ensures this difference is numerically small, although that turns out to be an artefact of this particular example. It is nonetheless conceptually important. In particular, the Λ-dependence of the right-hand side is associated with the Wilsonian coupling $a_W(\Lambda)$ because the scale Λ does not appear at all in the definition of $a_{\rm le}$ (which, after all, is defined in terms of integrations over modes at *all* scales).

How can these different values, $a_{\rm le} \neq a_W$, lead to the same physical predictions for observables? The answer lies in eq. (2.59), which states that $\Gamma_{\rm le}$ is obtained from S_W by integrating over the light degrees of freedom, using S_W rather than S as the action. It is this that fills in those parts of $\Gamma_{\rm le}$ not produced through loops involving heavy degrees of freedom. The relevant one-particle-irreducible one-loop graphs for generating the 4-point interaction in $\Gamma_{\rm le}$ using the interactions of S_W are shown in Fig. (2.5).

To see how things work in detail, consider each of the graphs in Fig. (2.5) in turn. Graph (a) just contributes an amount

$$a_{\rm graph(a)} = a_W(\Lambda)\,, \tag{2.74}$$

where a_W is computed up to one loop order (using only short-distance scales in the loop). This is the way that the high-energy parts of graphs (a) through (k) of Fig. (2.4) contribute when using the Wilson action. The dependence on the scale Λ is emphasized, since (as described above) this appears once a_W is computed at the loop level.

Next consider the contribution of the one-loop graph (b) of Fig. (2.5). This corresponds to the low-energy contribution to $a_{\rm le}$ from graphs (f) and (i) of Fig. (2.4), as can be seen if the χ internal lines in these graphs are shrunk down to a point, since the position-space version of the χ propagator is $G(x,y) \propto m_R^{-2}\delta^4(x-y)$. The vertex appearing at both ends in graph (b) is again the effective 4-point interaction, $\mathfrak{L}_{\rm eff} = a_W\,(\partial_\mu \xi \partial^\mu \xi)^2$, but because it is used in a loop its effective coupling should only be kept to tree-level accuracy: $a_{\rm tree} =$

$\lambda/(4m_R^4)$. Evaluating graph (b) of Fig. (2.5) gives a contribution to $a_{\rm le}$ of order

$$a_{\rm graph(b)} \propto a_{\rm tree}^2 \int_0^\Lambda \frac{{\rm d}^4p}{(2\pi)^4}\,\frac{p^4}{p^4} \propto \left(\frac{\lambda}{m_R^4}\right)^2 \frac{\Lambda^4}{16\pi^2} = \frac{1}{16\pi^2 v^4}\left(\frac{\Lambda^4}{m_R^4}\right). \tag{2.75}$$

Schematically, graph (b) of Fig. (2.5) is obtained from the low-frequency part of graphs (f) and (i) of Fig. (2.4) by contracting the heavy internal χ lines down into the effective point-like quartic self-interaction in the Wilson action. Notice that eq. (2.75) depends on the parameters Λ, m_R and v in precisely the way that is required to capture the low-energy part of graph (i), as estimated in eq. (2.73). It also captures the low-energy part of graph (f), since in this case the contribution of the lower limit of integration in L_3, eq. (2.70), is also quartic in Λ.

Similarly, graph (c) of Fig. (2.5), using the tree-level 6-point ξ interaction of S_W as the vertex, captures the low-energy contributions to $a_{\rm le}$ of graphs (g), (h), (j) and (k) of Fig. (2.4). The relevant 6-point coupling, b, arises (in principle) at tree level due to the graphs of Fig. (2.3), and so the result for graph (c) scales with the parameters v, m_R and Λ in precisely the way required,

$$a_{\rm graph(c)} \propto b_{\rm tree} \int_0^\Lambda \frac{{\rm d}^4p}{(2\pi)^4}\,\frac{p^2}{p^2} \propto \left(\frac{1}{v^4 m_R^4}\right) \frac{\Lambda^4}{16\pi^2} = \frac{1}{16\pi^2 v^4}\left(\frac{\Lambda^4}{m_R^4}\right), \tag{2.76}$$

to reproduce the missing low-energy contributions of graphs (g), (h), (j) and (k). When calculated earlier — *c.f.* §2.2.3 — it transpired that the two tree graphs contributing to b cancel, so $b_{\rm tree} = 0$. This corresponds to a similar cancelation in the low-energy limits of graphs (g), (h), (j) and (k) of Fig. (2.4).

When the dust settles, graphs (b) and (c) of Fig. 2.5 fill in those low-frequency parts of $a_{\rm le}$ that are missing from the contribution, a_W, of graph (a). The final result obtained for $a_{\rm le}$ using the Wilson action in Fig. 2.5 agrees with the one obtained directly from the full theory using Fig. 2.4.

In this particular example the difference $a_{\rm le} - a_W$ vanishes as $\Lambda \to 0$, because the quantity being computed is largely insensitive to long-wavelength modes. As a result the explicit Λ-dependence appearing in a_W cancels with the Λ-dependence implicit as cutoffs to the loop momenta in graphs (b) and (c) of Fig. 2.5. As shall be seen, in less simple examples loop contributions from long-wavelength modes can be important, and this cancellation of Λ in physical quantities is more subtle (involving renormalizations of low-energy effective couplings).

2.4 Dimensional analysis and scaling ♢

Although the discussion and examples described to this point contain the basic definitions of the Wilson action, many details remain to be filled in about its systematic use. Before developing the tools required for more systematic heavy lifting in the next chapter, this section first highlights some useful general properties of effective interactions, illustrated along the way using the toy model of Chapter 1.

2.4.1 Dimensional analysis

Dimensional analysis plays an important role in what follows, since it can be used to track the appearance of large mass scales in physical predictions. This section reviews the dimensions of the various ingredients from which the Wilson action is built.

To this end imagine writing down the Wilson action, $S_{\rm w} = \int \mathrm{d}^4x\, \mathfrak{L}_{\rm w}$, for some theory, describing physics below some high-energy mass scale: $E < M$. Suppose also this low-energy action comes as a functional of some field, ϕ, and its derivatives. As discussed in previous chapters, once organized into powers of the inverse of the heavy scale, $1/M$, the effective interactions in $\mathfrak{L}_{\rm w}$ must be local. These conditions require the action to have the general form

$$S_{\rm w}[\phi] = \int \mathrm{d}^4x\; \mathfrak{L}_{\rm w}\,, \tag{2.77}$$

with

$$\mathfrak{L}_{\rm w} = \sum_n c_n\, \mathcal{O}_n(\phi, \partial\phi, \cdots)\,, \tag{2.78}$$

a sum of powers of ϕ and its derivatives all evaluated at the same point, and (for relativistic systems) built to transform like a Lorentz scalar. If the low-energy theory is unitary then $\mathfrak{L}_{\rm w}$ should also be real. The goal is to use dimensional analysis to identify the power of M appearing in each effective coupling, c_n.

This book uses 'fundamental' units for which[8] $\hbar = c = 1$, and so the (engineering) dimension of any quantity can be regarded as a power of energy or mass (see Appendix A for conversions between these and more conventional units). In these units the action, $S_{\rm w}$, itself is dimensionless — *i.e.* has dimension (energy)^0 — or, more precisely, $S_{\rm w}/\hbar$ is dimensionless. Similarly time and space coordinates, t and $\mathbf{x}$, have dimension (energy)$^{-1}$ while derivatives like ∂_μ have dimension of energy. It is common to use the notation $[A] = p$ as a short form for the statement 'quantity A has dimension (energy)^p in units where $\hbar = c = 1$,' and in this notation $[S_{\rm w}] = 0$, $[x^\mu] = -1$ and $[\partial_\mu] = 1$.

Because the action is related to the lagrangian density by eq. (2.77), in four spacetime dimensions it follows that $\mathfrak{L}_{\rm w}$ has dimension (energy)^4 — *i.e.* $[\mathfrak{L}_{\rm w}] = 4$ — because the measure, d^4x, satisfies $\left[\mathrm{d}^4x\right] = -4$. If M is the only relevant mass scale in the problem and if a particular interaction, $\mathcal{O}_n$, has dimension $[\mathcal{O}_n] = \Delta_n$, then because $[c_n\mathcal{O}_n] = 4$ it follows that $[c_n] = 4 - \Delta_n$ and so one expects

$$[\mathcal{O}_n] = \Delta_n \quad \Rightarrow \quad c_n = \frac{a_n}{M^{p_n}} \quad \text{with} \quad p_n = \Delta_n - 4\,, \tag{2.79}$$

where a_n is dimensionless. To the extent that it is M that sets the scale of c_n in this way (and much of the next chapter is devoted to showing that the low-energy theory often can be set up so that it is), higher-dimensional interactions in $S_{\rm w}$ should be expected to be more suppressed at low energies by higher powers of M.

Further progress requires a way to compute the dimension, Δ_n, of a given operator, $\mathcal{O}_n$. For weakly interacting systems dimensions can be computed in perturbation theory. To see

[8] When temperatures are considered, units are also chosen with Boltzmann's constant satisfying $k_B = 1$, so temperature can also be measured in units of energy.

how this works, consider a real scalar field, ϕ, and suppose the regime of interest is one where it is relativistic and very weakly interacting. This means the action $S_w = S_0 + S_{\rm int}$ is dominated by its kinetic term

$$S_0 = -\frac{1}{2}\int \mathrm{d}^4x\, \partial_\mu\phi\, \partial^\mu\phi\,, \tag{2.80}$$

while all remaining terms, lumped together into $S_{\rm int} = \int \mathrm{d}^4x\, \mathfrak{L}_{\rm int}$ with

$$\mathfrak{L}_{\rm int} = -\frac{m^2}{2}\,\phi^2 + c_{4,0}\,\phi^4 + \frac{c_{6,0}}{\Lambda^2}\,\phi^6 + \frac{c_{4,2}}{\Lambda^2}\,\phi^2\,\partial_\mu\phi\,\partial^\mu\phi + \frac{c_{4,4}}{\Lambda^4}(\partial_\mu\phi\,\partial^\mu\phi)^2 + \cdots\,, \tag{2.81}$$

are assumed to be small. In this expression a symmetry of the form $\phi \to -\phi$ is imposed (for simplicity) so that only terms involving an even power of ϕ need be considered. Furthermore, appropriate powers of the cutoff, Λ, for the Wilsonian EFT are included explicitly in the coefficients for each effective interaction for reasons now to be explained.

Any effective coupling premultiplying S_0 is imagined to be removed by appropriately rescaling ϕ, with the choice of $\frac{1}{2}$ in eq. (2.80) called canonical normalization. (The reasons for using this normalization are elaborated below, and in Appendix C.3.) Given this normalization the dimension, $[\phi\,]$, of the (scalar) field ϕ is then determined by demanding that $[\mathfrak{L}_0] = 4$ and so

$$4 = [\partial_\mu\phi\,\partial^\mu\phi\,] = 2 + 2[\phi\,]\,, \tag{2.82}$$

which implies $[\phi\,] = 1$ (or ϕ has dimensions of energy). With this choice an identical argument shows $\left[m^2\right] = 2$ and so m also has dimensions of energy (as expected, since m is interpreted as the ϕ-particle's mass).

A similar story applies to all of the other terms in $S_{\rm int}$, and shows that the factors of Λ in (2.81) are extracted so that the remaining effective couplings are dimensionless: $[c_{n,d}] = 0$. For later purposes notice that a term in $S_{\rm int}$ involving n powers of ϕ and d derivatives comes premultiplied by Λ^p with $p = 4 - n - d$ and so the infinite number of local interactions not written explicitly in (2.81) have effective couplings with only negative powers of Λ.

The goal is to identify the domain of validity of the assumed perturbative hierarchy between S_0 and $S_{\rm int}$. The next few paragraphs argue that perturbative arguments are appropriate when the dimensionless couplings are assumed to be small: $|c_{n,d}| \ll 1$, following arguments made in [19]. To this end consider evaluating $S_w[\phi_k]$ at a wave-packet configuration $\phi_k(x) = f_k(x)\,e^{ikx}$, where $f_k(x)$ is a smooth envelope that is order unity for a spacetime region of linear size $2\pi/k$ in all four rectangular spacetime directions. For such a configuration spacetime derivatives are of order $\partial_\mu\phi_k \sim k_\mu\phi_k$ and the spacetime volume integral is of order $\int \mathrm{d}^4x \sim (2\pi/k)^4$, and so

$$\begin{aligned} S_w[\phi_k] &\sim \left(\frac{2\pi}{k}\right)^4 \left[k^2\,\phi_k^2 + m^2\phi_k^2 + c_{4,0}\,\phi_k^4 + \frac{c_{6,0}}{\Lambda^2}\,\phi_k^6 + c_{4,2}\left(\frac{k^2}{\Lambda^2}\right)\phi_k^4 + c_{4,4}\left(\frac{k^4}{\Lambda^4}\right)\phi_k^4 + \cdots\right] \\ &\sim (2\pi)^4\left[\varphi_k^2 + \frac{m^2}{k^2}\,\varphi_k^2 + c_{4,0}\,\varphi_k^4 + c_{6,0}\left(\frac{k^2}{\Lambda^2}\right)\varphi_k^6 + c_{4,2}\left(\frac{k^2}{\Lambda^2}\right)\varphi_k^4 + c_{4,4}\left(\frac{k^4}{\Lambda^4}\right)\varphi_k^4 + \cdots\right], \end{aligned} \tag{2.83}$$

where $\varphi_k := \phi_k/k$ is a new dimensionless variable. What is key about φ_k is that the path

integral over φ_k would be dominated by $\varphi_k \lesssim O(1)$ in the absence of interactions[9] (*i.e.* when $S_w = S_0$). This conclusion that dominant configurations for φ_k are order unity is contingent on the coefficient of $(\partial\phi)^2$ in S_0 being order unity due to the choice of canonical normalization.

Perturbation theory in $S_{\rm int}$ requires $|S_{\rm int}| \ll |S_0|$ throughout the regime from which the path integral receives significant contributions, which from the above considerations is the regime $\varphi_k \lesssim O(1)$. Consider first choosing k as large as possible: near the UV cutoff $|k^2| \sim \Lambda^2$. Since $\Lambda \gg m$ it follows that $|k^2| \gg m^2$ and so perturbation theory in this regime requires $|c_{n,d}| \ll 1$.

How does this conclusion change if k is now dialled down to smaller values? Since all interactions except the ϕ^2 and ϕ^4 interactions come pre-multiplied by positive powers of k^2/Λ^2, they become less and less important for smaller k. Interactions like this that are less important at lower energies are called *irrelevant*. Irrelevant interactions are also often called 'non-renormalizable'.[10]

By contrast, the ϕ^4 interaction is k-independent and so has strength controlled by $c_{4,0}$ for all k. Interactions like this whose strength does not vary with k are called *marginal*.

Finally, the mass (or ϕ^2) term is the only interaction that grows in importance for smaller k, the defining property of a *relevant* interaction. Once $|k^2| \lesssim m^2$ the mass term competes with S_0 and so changes the nature of the dominant path-integral configurations. This non-relativistic regime is of course important to many applications, and so is returned to in some detail as the topic of Part III below. Relevant interactions are sometimes also called 'super-renormalizable' while marginal and relevant interactions taken together are called 'renormalizable'.

A similar story goes through for fields representing other spins at weak coupling. For instance, a field, ψ, describing a free relativistic spin-half particle with lagrangian density

$$S_{1/2} = -\int \mathrm{d}^4x\, \overline{\psi}\, \partial\!\!\!/\, \psi\,, \tag{2.84}$$

with $\partial\!\!\!/ = \gamma^\mu\partial_\mu$ for dimensionless Dirac matrices, γ^μ (see Appendices A.2.3 and C.3.2) must have dimension $[\psi] = 3/2$. The kinetic term for an electromagnetic potential, A_μ, is

$$S_1 = -\frac{1}{4}\int \mathrm{d}^4x\, F_{\mu\nu}F^{\mu\nu} = -\frac{1}{2}\int \mathrm{d}^4x\, \partial_\mu A_\nu(\partial^\mu A^\nu - \partial^\nu A^\mu)\,, \tag{2.85}$$

and so the potential has dimension $[A_\mu] = 1$ while the field-strength satisfies $[F_{\mu\nu}] = 2$.

It is an important fact that all of the weakly interacting fields most commonly dealt with — such as ϕ, ψ and A_μ, as well as the derivative ∂_μ — have positive dimension, so more complicated interactions involving more powers of fields and/or derivatives always have higher and higher dimension. The corresponding effective couplings must therefore be proportional to more and more powers of $1/\Lambda$ (and so be less and less important at low energies). This is what ensures all but a handful of effective interactions are irrelevant at

[9] This is clearest if the problem is Wick-rotated to euclidean signature by going to imaginary time, so that $e^{iS} \to e^{-S}$. This estimate also ignores factors of 2π, though their inclusion somewhat broadens the domain of validity of perturbative methods.

[10] The recognition that it is useful to classify interactions according to their dimension came early [6], as did the connection to renormalizable and non-renormalizable interactions [7].

low energies, in the sense defined above. Precisely how irrelevant they are for any given k depends on the power of k^2/Λ^2 involved, so it makes sense to organize any list of potential interactions in order of increasing operator dimension, since the leading terms on the list are likely to be the most important at low energies.

From this point of view it is clear that the limited number of renormalizable (marginal and relevant) interactions with dimension $[O_n] \leq 4$ are special, since their importance is not diminished (and can be enhanced) at low energies.

For concreteness' sake the introductory discussion given above is phrased in perturbation theory, so it is worth mentioning in passing that this is not in principle necessary. That is, although EFT methods always exploit expansions in ratios of energy scales (like E/m_R for the toy model) it is *not* a requirement of principle that the dimensionless couplings of the underlying UV theory (like λ for the toy model) be perturbatively small.[11] Although it goes beyond the scope of this chapter to show in detail, strong underlying couplings can change some of the detailed statements used above, such as by changing the dimension of the field to be $[\phi\,] = 1 + \delta$, with $|\delta| \to 0$ as these couplings are taken to zero. (Examples along these lines where δ is perturbatively small are considered in later sections.) Differences like δ are called 'anomalous dimensions' for the quantities involved. What counts in the dimensional arguments to follow is that the full scaling dimensions (including these anomalous contributions) are used, rather than the lowest-order 'naive' scaling dimension.

Example: the toy model

As applied to the toy model, because the kinetic terms for the two fields ξ and χ have the form of eq. (2.80), the dimensions of both are $[\chi] = [\xi] = 1$. Using this with the full classical action, eqs. (1.24) and (1.25), shows that $[\lambda\,] = 0$ and $[v] = [m_R] = 1$, as expected.

Applied to the Wilson action, the effective coupling a appearing in the interaction $\mathfrak{L}_W \supset a\,(\partial_\mu \xi \partial^\mu \xi)^2$ must have dimension (energy)$^{-4}$, consistent with its computed tree-level value $a_{\rm tree} = \lambda/4m_R^4$. This shows that at leading order it is explicitly the heavy scale m_R that plays the role of the dimensional parameter of the general discussion. Powers of $\Lambda \ll m_R$ also arise once loops are included, and subsequent sections are devoted to identifying which scale is important in any particular application.

2.4.2 Scaling

It is worth rephrasing the above discussion more formally in terms of a scaling transformation. This is useful for several reasons: because it sets up the use of renormalization-group methods; and because it provides a framework that is more easy to generalize in more complicated settings, such as in the non-relativistic limit considered in Part III.

[11] In an unfortunate use of language the breakdown of the low-energy (*e.g.* E/m_R) expansion has in some quarters come to be called 'strong coupling'. This is misleading because it can happen that the physics appropriate to these energies is weakly coupled, in the sense that it involves dimensionless couplings that are small. For this reason in this book 'strong coupling' never means 'breakdown of the low-energy limit', and is reserved for situations where underlying dimensionless couplings (like λ in the toy model) are not small.

To this end consider again the scalar-field Wilson action of eqs. (2.80) and (2.81):

$$S_w[\phi(x); m, c_{4,0}, c_{6,0}, \cdots] = S_0[\phi(x)] + S_{\rm int}[\phi(x); m, c_{4,0}, \cdots], \tag{2.86}$$

where the notation explicitly highlights the dependence on the effective couplings as well as on the field ϕ. Now perform the scale transformation, $x^\mu \to sx^\mu$ and $\phi(x) \to \phi(sx)$, where s is a real parameter. For a configuration like $\phi_k(x) \propto e^{ikx}$ this becomes $\phi_k(sx) \propto e^{iskx} \propto \phi_{sk}(x)$ and so taking $s \to 0$ corresponds to taking the infrared limit where $k \to 0$.

Inserting these definitions into S_0 gives

$$\begin{aligned} S_0[\phi(sx)] &= -\frac{1}{2}\int \mathrm{d}^4x\, \partial_\mu\phi(sx)\, \partial^\mu\phi(sx) \\ &= -\frac{1}{2}\int \frac{\mathrm{d}^4x'}{s^4}\left[s^2\partial_{\mu'}\phi(x')\partial^{\mu'}\phi(x')\right], \end{aligned} \tag{2.87}$$

in which the spacetime integration variable is changed from x^μ to $x'^\mu = sx^\mu$. This shows that S_0 remains unchanged if the scalar field variable is also rescaled according to $\phi(x) \to \phi_s(x) := \phi(x)/s$. Requiring $S_0[\phi(sx)] = S_0[\phi_s(x)]$ is natural in the weak-coupling limit since this keeps fixed the configurations that dominate in the path integral over ϕ and ϕ_s.

With these choices the effects of rescaling on the interaction terms can be read off, giving

$$\begin{aligned} S_{\rm int}[\phi(sx); m, c_{4,0}, c_{6,0}, c_{4,2}, \cdots] &= \int \mathrm{d}^4x' \left[-\frac{m^2}{2s^2}\,\phi_s^2 + c_{4,0}\phi_s^4 + \frac{s^2c_{6,0}}{\Lambda^2}\phi_s^6 + \right. \\ &\qquad \left. + \frac{s^2c_{4,2}}{\Lambda^2}\,\phi_s^2(\partial_{\mu'}\phi_s\partial^{\mu'}\phi_s) + \cdots\right] \\ &= S_{\rm int}\left[\phi_s(x); \frac{m}{s}, c_{4,0}, s^2c_{6,0}, s^2c_{4,2}\ \cdots\right]. \end{aligned} \tag{2.88}$$

This shows that changes of scale can be compensated by appropriately rescaling all effective couplings. For instance, for an interaction $S_n[\phi; a_n] = \int \mathrm{d}^4x\, a_n\mathcal{O}_n[\phi] \in S_{\rm int}$, where $\mathcal{O}_n$ has engineering dimension $[\mathcal{O}_n] = \Delta_n$, the required scaling is

$$S_n[\phi(sx); a_n] = S_n[\phi_s(x); s^{p_n}a_n], \tag{2.89}$$

where $p_n = -[a_n] = \Delta_n - 4 =$ (so that $a_n = c_n/\Lambda^{p_n}$ for dimensionless c_n).

Rescalings can be regarded as motions within the space of coupling constants. This provides an alternate way to define relevant, marginal and irrelevant interactions. Since low energies correspond to $s \to 0$, an effective interaction is irrelevant if $p_n > 0$, it is marginal if $p_n = 0$ and it is relevant if $p_n < 0$. This definition clearly agrees with the one presented earlier.

2.5 Redundant interactions ◇

It is generally useful to have in mind what are the most general possible kinds of effective interactions that can arise in a Wilson action at any given dimension, and it is tempting to think that this means simply listing all possible combinations of local interactions involving

the given fields and their derivatives. In practice such a list over-counts the number of possible interactions because there is considerable freedom to rewrite the effective action in superficially different, but equivalent, ways.

Because of this freedom some combinations of interactions can turn out not to influence observables at all (or only do so in special situations), and it is important to identify these to avoid over-counting the couplings that are possible. Ignorable interactions like this that have no physical effects are called *redundant* interactions, and this section describes two generic kinds of redundancy that commonly arise.

Total derivatives

The first category of often-ignorable interactions are total derivatives — such as, for the generic low-energy field ϕ of the last section

$$S_{\rm int} = -g \int \mathrm{d}^4x\, \partial_\mu\Big(\phi^2 \partial^\mu \phi\Big)\,. \tag{2.90}$$

Stokes' theorem allows this kind of interaction to be written as a function only of boundary data

$$\int_M \mathrm{d}^4x\, \partial_\mu\Big(\phi^2 \partial^\mu \phi\Big) = \int_{\partial M} \mathrm{d}^3x\, n_\mu\Big(\phi^2 \partial^\mu \phi\Big)\,, \tag{2.91}$$

where n_μ denotes the outward-directed unit normal to the boundary ∂M of the initial integration region M. To the extent that the physics of interest does not depend on this boundary data (such as if there are no boundaries, or if spatial infinity is the 'boundary' and all fields fall off sufficiently quickly at infinity) such total derivatives can be dropped.

Of course, no mistakes are actually made by keeping redundant interactions in a calculation, one just works unnecessarily hard. This is because couplings like g in the example above do not in any case appear in physical observables. To see how this can happen in detail, consider the example of the momentum-space Feynman rule computed for the 3-point vertex described by eq. (2.90):

$$g(p_1 + p_2 + p_3)\cdot p_3 \left[\mathrm{i}(2\pi)^4 \delta^4(p_1 + p_2 + p_3)\right]. \tag{2.92}$$

Here $p_i \cdot p_j = \eta_{\mu\nu} p_i^\mu p_j^\nu$ where p_i^μ denote the inward-pointing 4-momenta for each of the lines attached to the vertex. Due to the presence of the energy-momentum conserving delta function this has the form $x\,\delta(x) = 0$, and so identically vanishes. Consequently this kind of interaction cannot contribute to any order in a perturbative expansion organized in terms of Feynman graphs.

The neglect of total-derivative terms must be re-examined in situations with boundaries or where the asymptotic behaviour of fields cannot be ignored. This can happen either when there really are physical boundaries or when there are fields that are sensitive to nontrivial topology.[12] When boundaries are present the above discussion continues to apply provided that the appropriate boundary interactions are also included (see §5 and §7.4 for examples along these lines).

[12] Topology enters if there are terms in the lagrangian that are locally total derivatives, but cannot be globally written this way throughout all of field space.

Field redefinitions

A second type of ignorable interaction is one that can be removed by performing a local field redefinition. Since physical quantities cannot depend on the particular choice of variables used by specific physicists for their description,[13] anything that can be removed by a change of variables cannot contribute to any physical observables.

But how to decide if a particular interaction can be removed in this way? It turns out there is a very simple criterion that works for any situation where the action is given as a series in a small quantity, ϵ:

$$S[\phi] = S_0[\phi] + \epsilon S_1[\phi] + \epsilon^2 S_2[\phi] + \cdots . \tag{2.93}$$

This form always applies for the Wilson action in particular, where the corresponding expansion would be the low-energy approximation.

Imagine performing a generic infinitesimal field redefinition of the form

$$\delta\phi = \epsilon f_1[\phi] + \epsilon^2 f_2[\phi] + \cdots , \tag{2.94}$$

for some arbitrary local functions, $f_i = f_i(\phi, \partial\phi, \cdots)$, of the fields and their derivatives. The change in eq. (2.93) is

$$\begin{aligned} \delta S &= \int \mathrm{d}^4x \left\{ \frac{\delta S_0}{\delta\phi(x)} + \epsilon \frac{\delta S_1}{\delta\phi(x)} + \epsilon^2 \frac{\delta S_2}{\delta\phi(x)} + \cdots \right\} \delta\phi \\ &= \int \mathrm{d}^4x \left\{ \epsilon \left[\frac{\delta S_0}{\delta\phi(x)} f_1 \right] + \epsilon^2 \left[\frac{\delta S_0}{\delta\phi(x)} f_2 + \frac{\delta S_1}{\delta\phi(x)} f_1 \right] + \cdots \right\} . \end{aligned} \tag{2.95}$$

This shows that the function f_1 can be used to remove any interaction in S_1 that is proportional to $\delta S_0/\delta\phi$ — and so vanishes when its lowest-order equations of motion are used[14] — up to terms that are $O\left(\epsilon^2\right)$. For instance, a term in S_i of the form

$$S_i \supset \int \mathrm{d}^4x \, \frac{\delta S_0}{\delta\phi(x)} \, B[\phi] , \tag{2.96}$$

(for some local function $B[\phi]$) is removed using the choice $f_i = -B$. That is, the quantities f_1 through f_{n+1} can be used to remove any interaction in $S - S_0$ that vanishes when $\delta S_0/\delta\phi = 0$, order-by-order in ϵ. Notice also that (2.95) also shows that the redefinition that removes terms from S_i can do so at the expense of also introducing new terms into S_j for $j > i$.

Example: the toy model

As usual, it is useful to make things concrete with an explicit example, for which the toy model of chapter 1 is again pressed into service. To apply these ideas to the toy model, recall what has been found for its Wilson action so far. The calculations performed in the

[13] This is a theorem for scattering amplitudes [8], but it also applies to more general observables.

[14] This observation seems to have been general knowledge back into the mists of time, but the earliest explicit mention of it in the literature I have found is [20] (see also footnote 9 of [2]).

previous sections show it to have the form

$$S_w[\xi] = -\int \mathrm{d}^4x \left[\frac{1}{2}\,\partial_\mu\xi\partial^\mu\xi - a(\partial_\mu\xi\partial^\mu\xi)^2 - a'(\partial_\mu\xi\partial^\mu\xi)\Box(\partial_\mu\xi\partial^\mu\xi) - b\,(\partial_\mu\xi\partial^\mu\xi)^3 - b'(\partial_\mu\xi\partial^\mu\xi)\Box(\partial_\mu\xi\partial^\mu\xi)^3 + \cdots\right], \tag{2.97}$$

where

$$a = \frac{1}{m_R^4}\left[\frac{\lambda}{4} + O(\lambda^2)\right], \quad a' = \frac{1}{m_R^6}\left[\frac{\lambda}{4} + O(\lambda^2)\right], \quad b = \frac{1}{m_R^8}\left[0 + O(\lambda^3)\right] \tag{2.98}$$

and so on. The question is: are these the most general kinds of interactions possible? In particular, are there terms suppressed by only two powers of $1/m_R$? If not, why not?

Of course, symmetries restrict the form of S_w, and for the toy model symmetry under the shift $\xi \to \xi + \sqrt{2}\,v\,\omega$ requires ξ always to appear in S_w differentiated, so all interactions must involve at least as many derivatives as powers of ξ. Furthermore, to be a Lorentz scalar it must involve an even number of derivatives, so that all Lorentz indices can be contracted. But these conditions allow interactions that do not appear in eq. (2.97). For instance, they allow the following effective interactions with dimension (energy)^6,

$$\begin{aligned}\mathfrak{L}_6 &= \frac{a_1}{m_R^2}(\partial_\mu\xi\Box\partial^\mu\xi) + \frac{a_2}{m_R^2}(\partial_\mu\partial_\nu\xi\,\partial^\mu\partial^\nu\xi)\\ &= \frac{(a_1-a_2)}{m_R^2}(\partial_\mu\xi\Box\partial^\mu\xi) + \frac{a_2}{m_R^2}\partial_\nu(\partial_\mu\xi\,\partial^\mu\partial^\nu\xi)\,,\end{aligned} \tag{2.99}$$

where (given the explicit dimensional factor of m_R^{-2}) a_1 and a_2 are dimensionless effective couplings.

The point is that both of these interactions are redundant, in the sense outlined above. The second line shows that one combination can be regarded as a total derivative, and so is redundant to the extent boundaries (or topology) do not play an important role in the physics of interest. The remaining term, involving the combination $\partial_\mu\xi\Box\partial^\mu\xi$, vanishes once evaluated using the equations of motion, $\Box\,\xi = 0$, for the lowest-order action. It can therefore be removed using the field redefinition

$$\xi \to \xi + \frac{a_2 - a_1}{m_R^2}\,\Box\,\xi\,, \tag{2.100}$$

since in this case

$$-\frac{1}{2}\,\partial_\mu\xi\partial^\mu\xi \to -\frac{1}{2}\,\partial_\mu\xi\partial^\mu\xi + \frac{a_2 - a_1}{m_R^2}\,(\partial_\mu\xi\Box\partial^\mu\xi)\,, \tag{2.101}$$

up to terms of order $1/m_R^4$. This shows that (in the absence of boundaries) the first low-energy effects of virtual heavy particles arise at order $1/m_R^4$ rather than $1/m_R^2$.

What about interactions with dimension (energy)^8: is $(\partial_\mu\xi\,\partial^\mu\xi)^2$ the only allowed dimension-8 interaction? Since total derivatives are dropped, integration by parts can be used freely to simplify any candidate interactions. The most general possible Lorentz-scalar interactions invariant under constant shifts of ξ then are

$$\mathfrak{L}_8 = -a(\partial_\mu\xi\partial^\mu\xi)^2 - \frac{a_3}{m_R^4}\,(\partial_\mu\xi\Box^2\partial^\mu\xi)\,, \tag{2.102}$$

where a_3 is a new dimensionless coupling and the freedom to integrate by parts is used (for the terms quadratic in ξ) to ensure all of the derivatives but one act on only one of the fields. The second term in (2.102) can be removed using the field redefinition

$$\xi \to \xi + \frac{a_3}{m_R^4}\,\Box^2\,\xi\,, \tag{2.103}$$

without changing the coefficient a (or coefficients of any lower-dimension interactions), showing that a captures all of the effects that can arise at order $1/m_R^4$.

2.6 Summary

This chapter lays one of the cornerstones for the rest of the book; laying out how effective lagrangians fit into the broader context of generating functionals and the quantum (1PI) action.[15] By doing so it provides a constructive framework for defining and explicitly building effective actions for a broad class of physical systems.

The 1PI action is a useful starting point for this purpose because it already plays a central role in quantum field theory. It does so partly because it is related to the full correlation functions and the energetics of field expectation values in the same way that the classical action is related to the classical correlation functions and the energetics that fixes the values of classical background fields. The low-energy 1LPI action is the natural generalization of the 1PI action because it is constructed in precisely the same way, but with the proviso that it only samples slowly varying field configurations. As such it contains all the information needed to construct any observable that involves only low-energy degrees of freedom.

In this chapter the Wilson action, S_w, emerges as the minimal object for capturing the implications of high-energy degrees of freedom for the low-energy theory. The Wilson action is related to the 1LPI action in the low-energy theory in precisely the same way that the classical action is related to the 1PI generator in the full theory. Because S_w is obtained by integrating out only high-energy states, its interactions efficiently encode their low-energy implications. And because knowledge of S_w allows the calculation of the 1LPI action it contains all of the information required to compute any low-energy observable.

The chapter concludes with a few tools that prove useful in later chapters when computing and using the Wilson action. The first tool is simply dimensional analysis, which classifies effective interactions based on their operator dimension (in powers of energy). More and more interactions exist with larger operator dimension, but it is the relatively few lower-dimension interactions in this classification that are more important at low energies. This chapter also describes the related

[15] In some of the earlier literature the quantum action is also called the effective action, unlike the modern usage where effective action usually means the Wilson action.

renormalization-group scaling satisfied by the effective couplings. These express how the effective couplings differentially adjust as more and more modes are integrated out, lowering the energy scale Λ that differentiates low energies from high. (The next chapter also has more to say about this dimensional scaling and its utility for identifying which interactions are important at low energies.)

The second tool described in this chapter identifies classes of effective interactions that are redundant in the sense that they do not contribute at all to physical processes. They do not contribute for one of two reasons: either they are total derivatives and so are only sensitive to physics that depends in detail on the information at the system's boundaries; or because there exists a change of variables that allows them to be completely removed.

Exercises

Exercise 2.1 Prove eq. (2.6) starting from eqs. (2.2) and (2.4).

Exercise 2.2 Draw all possible two-loop vacuum diagrams that contribute to $Z[J]$ in a theory involving both cubic and quartic interactions (such as a scalar potential $V(\phi) = g\phi^3 + \lambda\phi^4$ for scalar-field self-interactions). Which of these diagrams contribute to $W[J]$ and to $\Gamma[\varphi]$?

Exercise 2.3 For a scalar field self-interacting through the potential $V = g\phi^3 + \lambda\phi^4$ express eq. (2.15) as a sum of Feynman graphs with two external lines. Draw all graphs that contribute out to two-loop order. Show how the disconnected graphs cancel in the result.

Exercise 2.4 Prove that the graphical expansion of $W[J]$, defined by $Z[J] = \exp\{iW[J]\}$, is obtained by simply omitting any disconnected graphs that contribute to $Z[J]$. Do so by showing that the exponential of the sum of all connected graphs reproduces all of the combinatorial factors in the sum over all (connected and disconnected) graphs. For this argument it is not necessary to assume that only cubic or quartic interactions arise in $S_{\rm int}$.

Exercise 2.5 Consider a single scalar field, φ, self-interacting through a scalar potential $U(\varphi)$. Evaluate eq. (2.23) in one-loop approximation for φ specialized to a constant spacetime-independent configuration. To do so use the identity $\ln\det\Delta = \mathrm{Tr}\,\ln\Delta$ and work in momentum space, for which $\Delta(p,p') = (p^2 + m^2 - i\epsilon)\,\delta^4(p-p')$, where $m^2 := U'' := \partial^2 U/\partial\varphi^2$. Evaluate the trace explicitly and Wick rotate to Euclidean signature ($p^0 = ip_E^4$) to derive the following expression

$$V_{\rm q}(\varphi) = U(\varphi) + \frac{1}{2}\int \frac{d^4p_E}{(2\pi)^4}\,\ln(p_E^2 + m^2) = U_\infty(\varphi) + \frac{1}{64\pi^2}\,m^4\ln\left(\frac{m^2}{\mu^2}\right),$$

for the quantum effective potential. Regulate the UV divergences using dimensional regularization (for which μ is the arbitrary scale – see Appendix A.2.4), and show that $U_\infty(\varphi) = U(\varphi) + A + B\,m^2(\varphi) + Cm^4(\varphi)$, where A, B and C are divergent constants

in 4 spacetime dimensions. Show that if $U(\varphi) = U_0 + U_2\varphi^2 + U_4\varphi^4$ is quartic (and so renormalizable) then all divergences can be absorbed into the constants U_0, U_2 and U_4.

Exercise 2.6 Prove that the quantum effective potential is always convex [16, 21] when constructed about a stable vacuum. That is, show that for $0 \leq s \leq 1$

$$V_{\rm q}[s\varphi_1 + (1-s)\varphi_2] \leq sV_{\rm q}(\varphi_1) + (1-s)V_{\rm q}(\varphi_2)\,.$$

Exercise 2.7 Suppose the action $S[\phi]$ for a field theory is invariant under a symmetry transformation of the form $\delta\phi^a = \omega\,\zeta^a(\phi)$, where $\zeta^a(\phi)$ is a possibly nonlinear function and ω is an infinitesimal symmetry parameter. Show that the 1PI generator, $\Gamma[\varphi]$, is invariant under the symmetry $\delta\varphi^a = \omega\,\langle\zeta^a\rangle_J$, where the matrix element is taken in the adiabatic vacuum in the presence of the current $J_a(\varphi)$ defined by eq. (2.16). In the special case of a linear transformation, with $\zeta^a(\phi) = M^a{}_b\,\phi^b$ the condition $\langle\phi^a\rangle_J = \varphi^a$ implies both $\Gamma[\varphi]$ and $S[\phi]$ share a symmetry with the same functional form.

The invariance condition $\delta\Gamma = 0$ for this transformation can be expressed as

$$\int {\rm d}^4x\,\langle\zeta^a(x)\rangle_J\,\frac{\delta\Gamma}{\delta\varphi^a(x)} = 0\,.$$

Since this is true for all $\varphi^a(x)$ repeated functional differentiation leads to a sequence of relations — called Taylor-Slavnov identities [22, 23] — amongst the 1PI correlation functions obtained by differentiating $\Gamma[\varphi]$.

Exercise 2.8 For the toy model of §1 draw all tree-level (no loops) Feynman graphs that can contribute to the effective interaction $\mathfrak{L}_w \supset c\,(\partial_\mu\xi\,\partial^\mu\xi)^4$ within the Wilson action. Evaluate these graphs and compute the effective coupling c at tree level.

Exercise 2.9 Construct the most general possible renormalizable relativistic interactions for a single real scalar field ϕ in $D = 4$ spacetime dimensions. Repeat this exercise for $D = 6$ spacetime dimensions. For $D = 4$ find the most general possible renormalizable relativistic interactions for a real scalar field coupled to a spin-half Dirac field ψ.

Exercise 2.10 For a real scalar field, ξ, subject to a shift symmetry, $\xi \to \xi +$ constant, every appearance of ξ in the Wilson action must be differentiated at least once. Show that the most general effective interactions possible for such a field involving six or fewer derivatives is

$$\mathfrak{L}_w = -\frac{1}{2}\,(\partial_\mu\xi\,\partial^\mu\xi) + a\,(\partial_\mu\xi\,\partial^\mu\xi)^2 + b\,(\partial_\mu\xi\,\partial^\mu\xi)^3 + c\,(\partial_\mu\xi\,\partial^\mu\xi)(\partial_\lambda\partial_\rho\xi)\,(\partial^\lambda\partial^\rho\xi)\,,$$

up to redundant interactions, for effective couplings a, b and c.

3 Power counting and matching

The previous chapter argues that the Wilson action captures the influence of virtual high-energy states on all low-energy observables, but a good number of questions remain to be addressed before it becomes a tool of practical utility. In particular, the Wilson action in principle contains an infinite number of interactions of various types, and although these are local (once expanded in inverse powers of the heavy scale) they ultimately involve arbitrarily many powers of the low-energy fields and their derivatives. What is missing is a simple way to identify systematically which interactions are required to calculate any given observable to a given order in the low-energy (and any other) expansions.

In principle, as argued in §2.4, what makes the Wilson action useful is dimensional analysis, which shows that more complicated interactions (with more derivatives or powers of fields) have coupling constants more suppressed by inverse powers of the physical heavy mass scale, M (like m_R in the toy model). This suggests that a dimension-Δ interaction can be ignored at low energies, E, provided effects of order $(E/M)^p$, with $p = \Delta - 4$, are negligible.

Sounds simple. Unfortunately, there is a confounding factor that complicates the simple dimensional argument. Although each use of an effective interaction within a Feynman graph costs inverse powers of a heavy scale like M, it is also true that the 4-momenta of virtual particles circulating within loops can include energies that are not small. This means that heavy scales can appear in numerators of calculations as well as in denominators, making it trickier to quantify the size of higher-order effects. Power-counting — the main subject of this chapter — makes this argument more precise, and is the tool with which to identify which effective interactions are relevant to any particular order in the low-energy expansion.

To see how scales appear in calculations, for some purposes it is useful to track explicitly cutoffs, like Λ, that label the highest energies allowed to circulate within loops. Depending on the relative size of scales like M and Λ it can happen that loop effects can cause effective couplings to acquire coefficients $c_n \propto \Lambda^{-p}$ rather than $c_n \propto M^{-p}$. For $p > 0$ these naively dominate because $\Lambda \ll M$. Estimates of the size of such corrections are discussed in this chapter in the section devoted to the 'exact renormalization group' (or 'exact RG').

But it is also true that Λ ultimately drops out of physical quantities, making its presence an unnecessary complication when formulating dimensional arguments. (Λ drops out of physical quantities because the precise split between low- and high-energy quantities is ultimately a book-keeping device for making calculations convenient, so Λ is not a physical scale. As this chapter shows, the disappearance of Λ in physical predictions happens in

detail because the explicit Λ-dependence of the effective couplings in S_w cancels the Λ-dependence implicit in the definition of the low-energy path integral in which S_w is used.)

These observations suggest it is likely to be more efficient to formulate low-energy quantities in a cutoff-independent way. In particular, power-counting is most efficiently formulated when dimensional regularization is used to define the Wilson action, rather than a floating cutoff like Λ. Ultimately, power-counting turns the Wilson action from a chainsaw into a scalpel, making it a tool for making precision calculations. Along the way it shows why UV divergences are only nuisances that are not fatal complications to the formulation of the low-energy theory, and provides deep conceptual insights into the physical meaning of renormalizability.

3.1 Loops, cutoffs and the exact RG ♠

The goal of this section is to track explicitly how the scales appearing in the Wilson action propagate into low-energy observables.[1] To this end, suppose the Wilson action computed from a specific underlying theory has the form

$$S_w = S_{w,0} + S_{w,\text{int}}\,, \tag{3.1}$$

with

$$\begin{aligned} S_{w,0} &= -\frac{\mathfrak{f}^4}{M^2 v^2}\int \mathrm{d}^4x\,\left[\partial_\mu\phi\,\partial^\mu\phi + m^2\phi^2\right] \\ S_{w,\text{int}} &= -\mathfrak{f}^4\sum_n \frac{\hat{c}_n}{M^{d_n}v^{f_n}}\int \mathrm{d}^4x\,\mathcal{O}_n(\phi)\,, \end{aligned} \tag{3.2}$$

where ϕ denotes a generic low-energy field which the form for S_0 assumes to be bosonic, with $[\phi\,] = 1$. For simplicity only one field is kept here, though the dimensional arguments to follow remain unchanged if ϕ instead represents a collection of fields. The index n runs over a complete set of labels for all possible interactions, for each of which there are two non-negative integers, d_n and f_n, that respectively count the number of powers of derivatives and fields that appear in the interaction $\mathcal{O}_n$. For example, for an effective interaction like $\mathcal{O} = \phi^2\partial_\mu\phi\,\partial^\mu\phi$ these constants are $d_n = 2$ and $f_n = 4$.

The three dimensionful quantities $\mathfrak{f}$, v and M are all energy scales much larger than the light-particle mass, m, that can be (but need not be) independent of one another or of Λ. Roughly speaking, this writes the action as an expansion in powers of fields, ϕ/v, and derivatives, ∂/M, with no *a-priori* requirement that the two comparison scales v and M be similar. The scale $\mathfrak{f}^4$ gives the rough energy density associated with $\phi \sim v$ and $\partial \sim M$. The goal is to track how these scales appear in physical quantities once S_w is used to compute them.

The kinetic term coming from $S_{w,0}$ carries the factor $\mathfrak{f}^4/M^2v^2$ so that it scales with these parameters in the same way as do the interaction terms. Although this means ϕ is not

[1] This and later sections broadly follow the logic of [2], though details and notation used follow that of [24].

canonically normalized (unless $\mathfrak{f}^4 = \frac{1}{2} M^2 v^2$), the discussion is nonetheless kept general by working with Feynman rules that do not assume canonical normalization.

With these definitions, to leading order the total dimension of an interaction having d_n derivatives and f_n powers of ϕ is therefore

$$\Delta_n := [O_n] = d_n + f_n \,, \tag{3.3}$$

so the powers of $\mathfrak{f}$, M and v ensure that the coefficients $\hat{c}_n$ are dimensionless to leading order (assuming $[\phi\,] = 1$).

3.1.1 Low-energy amplitudes

Imagine now using these effective interactions in a path integral to compute a low-energy observable. Working perturbatively in $S_{w,\rm int}$ amounts to evaluating various Feynman graphs using $S_{w,0}$ to define the internal lines and $S_{w,\rm int}$ to define the effective interaction vertices.

Suppose $A_{\mathcal{E}}(q)$, denotes the result of evaluating an amplitude involving $\mathcal{E}$ external lines, regarded as a function of a collection of external kinetic variables, q, sharing a common low-energy scale $E \ll \mathfrak{f}, M, v, \Lambda$. $A_{\mathcal{E}}$ could be a scattering amplitude among low-energy particles, or might be a contribution to the generating functional[2] $\Gamma_{\rm le}$. The goal is to determine the systematics of how $A_{\mathcal{E}}(q)$ depends on the various energy scales as a function of $\mathcal{E}$ and q. But in general different Feynman graphs involving different numbers of internal lines and vertices can depend on these variables differently, so it is also worth tracking the dependence on other quantities like the number, $\mathcal{I}$, of internal lines and the number, $\mathcal{V}_n$, of vertices coming from the interaction O_n. Since O_n involves f_n fields there are f_n lines that converge at the corresponding vertex, while d_n counts the number of derivatives appearing in the corresponding Feynman rule.

Some useful identities:

The first observation is that the positive integers, $\mathcal{I}$, $\mathcal{E}$ and $\mathcal{V}_n$, that characterize any particular graph are not all independent. Rather they they are related by the rules for constructing graphs from lines and vertices.

One such a relation is obtained by equating the two equivalent ways of counting the number of ends of internal and external lines in a graph:

- On one hand, since all lines end at a vertex, the number of ends is given by summing over all of the ends appearing in all of the vertices: $\sum_n f_n \mathcal{V}_n$;
- On the other hand, there are two ends for each internal line, and one end for each external line in the graph: making a total of $2\mathcal{I} + \mathcal{E}$ ends.

Equating these two ways of counting gives the identity expressing the 'conservation of

[2] For applications to scattering amplitudes and generating functionals external lines are amputated from Feynman graphs, which matters when applying dimensional analysis to the result. For this reason a minor modification is required to apply the power-counting results here to correlation functions – see *e.g.* the discussion of §10.2.1.

ends':

$$2\mathcal{I} + \mathcal{E} = \sum_n f_n \mathcal{V}_n, \qquad \text{(conservation of ends)}. \tag{3.4}$$

A second useful identity defines of the number of loops, $\mathcal{L}$, for each (connected) graph:

$$\mathcal{L} = 1 + \mathcal{I} - \sum_n \mathcal{V}_n, \qquad \text{(definition of } \mathcal{L}\text{)}. \tag{3.5}$$

As mentioned around eq. (2.24), this definition is motivated by the topological identity that applies to any graph that can be drawn on a plane, that states that $\mathcal{L} - \mathcal{I} + \sum_n \mathcal{V}_n = 1$ (which is the Euler number of a disc). In what follows eqs. (3.4) and (3.5) are used to eliminate $\mathcal{I}$ and $\sum_n f_n \mathcal{V}_n$.

Feynman rules

The next step is to use the action of eqs. (3.2) and (3.55) to construct the Feynman rules for the graph of interest. This is done here in momentum space, but since the argument to be made is in essence a dimensional one it could equally well be made in position space.

Schematically, in momentum space the product of all of the vertices contributes the following factor to the amplitude:

$$\text{(Vertices)} = \prod_n \left[\mathrm{i}(2\pi)^4 \delta^4(p)\, \mathfrak{f}^4 \left(\frac{p}{M}\right)^{d_n} \left(\frac{1}{v}\right)^{f_n} \right]^{\mathcal{V}_n}, \tag{3.6}$$

where p generically denotes the various momenta running through the vertex. The product of all of the internal line factors gives the additional contribution:

$$\text{(Internal Lines)} = \left[-\mathrm{i} \int \frac{\mathrm{d}^4 p}{(2\pi)^4} \left(\frac{M^2 v^2}{\mathfrak{f}^4}\right) \frac{1}{p^2 + m^2} \right]^{\mathcal{I}}, \tag{3.7}$$

where p again denotes the generic momentum flowing through the lines. m is the mass of the light particle (or their generic order of magnitude — for simplicity taken to be similar in size — should there be more than one light field) coming from the unperturbed term, eq. (3.2). For the 'amputated' Feynman graphs relevant to scattering amplitudes and contributions to effective couplings in Γ_{le} the external lines are removed, and so no similar factors are included for external lines.

The momentum-conserving delta functions appearing in (3.6) can be used to perform many of the integrals appearing in (3.7) in the usual way. Once this is done, one delta function remains that depends only on external momenta, $\delta^4(q)$, and so cannot be used to perform additional integrals. This is the delta function that enforces the overall conservation of energy and momentum for the amplitude. It is useful to extract this factor once and for all, by defining the reduced amplitude, $\mathcal{A}_\mathcal{E}$, by

$$A_\mathcal{E}(q) = \mathrm{i}(2\pi)^4 \delta^4(q)\, \mathcal{A}_\mathcal{E}(q)\,. \tag{3.8}$$

The total number of integrations that survive after having used all of the momentum-conserving delta functions is then $\mathcal{I} - \sum_n \mathcal{V}_n + 1 = \mathcal{L}$. This last equality uses the definition, eq. (3.5), of the number of loops, $\mathcal{L}$.

3.1.2 Power counting using cutoffs

The hard part in computing $\mathcal{A}_{\mathcal{E}}(q)$ is to evaluate the remaining multi-dimensional integrals. Things are not so bad if the only goal is to track how the result depends on the scales ϱ, M and v, however, since then it suffices to use dimensional arguments to estimate the size of the result. Since the integrals typically diverge in the ultraviolet, they are most sensitive to the largest momenta in the loop, and according to eq. (2.59) this is set by the cutoff Λ. (The contributions of loops having momenta higher than Λ are the ones used when computing $S_{\scriptscriptstyle W}$ itself from the underlying theory.)

This leads to the following dimensional estimate for the result of the integration

$$\int \cdots \int \left[\frac{\mathrm{d}^4 p}{(2\pi)^4}\right]^A \frac{p^B}{(p^2+m^2)^C} \sim \left(\frac{1}{4\pi}\right)^{2A} \Lambda^{4A+B-2C} . \tag{3.9}$$

For the purposes of counting 2π's, a factor of π^2 is included for each $\mathrm{d}^4 p$ integration corresponding to the result of performing the three angular integrations.[3]

The idea is to Taylor expand the amplitude $\mathcal{A}_{\mathcal{E}}(q)$ in powers of external momentum, q, using eq. (3.9) to estimate the size of the coefficients. Schematically,

$$\mathcal{A}_{\mathcal{E}}(q) \simeq \sum_{\mathcal{D}} \mathcal{A}_{\mathcal{E}\mathcal{D}}\, q^{\mathcal{D}} , \tag{3.10}$$

where the coefficients require an estimate for the following integral

$$\begin{aligned} \mathcal{A}_{\mathcal{E}\mathcal{D}}\, q^{\mathcal{D}} &\propto \int \cdots \int \left[\frac{\mathrm{d}^4 p}{(2\pi)^4}\right]^{\mathcal{L}} \frac{1}{(p^2+m^2)^{\mathcal{I}}} \left(\frac{q}{p}\right)^{\mathcal{D}} \prod_n p^{d_n \mathcal{V}_n} \\ &\sim \left(\frac{1}{4\pi}\right)^{2\mathcal{L}} \left(\frac{q}{\Lambda}\right)^{\mathcal{D}} \Lambda^{4\mathcal{L}-2\mathcal{I}+\sum_n d_n \mathcal{V}_n} . \end{aligned} \tag{3.11}$$

Combining this with the powers of ϱ, M and v given by the Feynman rules then gives, after using identities (3.4) and (3.5),

$$\mathcal{A}_{\mathcal{E}\mathcal{D}}\, q^{\mathcal{D}} \sim \varrho^4 \left(\frac{1}{v}\right)^{\mathcal{E}} \left(\frac{q}{\Lambda}\right)^{\mathcal{D}} \left(\frac{M\Lambda}{4\pi\,\varrho^2}\right)^{2\mathcal{L}} \left(\frac{\Lambda}{M}\right)^{2+\sum_n (d_n-2)\mathcal{V}_n} . \tag{3.12}$$

This is the main result of this section, whose properties are now explored.

A reality check for this formula comes if it is applied to the simplest graph of all (see Fig. 3.1): one including no internal lines (so $\mathcal{L}=0$) and only a single vertex, $n=n_0$, with $f_{n_0}=\mathcal{E}$ external lines and $d_{n_0}=\mathcal{D}$ derivatives (so $\sum_n \mathcal{V}_n = 1$ and $\sum_n d_n \mathcal{V}_n = \mathcal{D}$). In this case (3.12) implies the amplitude depends on the scales M, Λ and ϱ in precisely the same way as does the starting lagrangian (3.2): $\mathcal{A}_{\mathcal{E}\mathcal{D}}\, q^{\mathcal{D}} \sim \varrho^4 (1/v)^{\mathcal{E}} (q/M)^{\mathcal{D}}$.

A second reality check applies (3.12) to the special case where all scales are set by Λ (*i.e.* $\varrho = M = v = \Lambda$) since this corresponds to the choices made in the dimensional arguments of §2.4.1. In this limit (3.12) becomes

$$\mathcal{A}_{\mathcal{E}\mathcal{D}}\, q^{\mathcal{D}} \sim \Lambda^4 \left(\frac{1}{\Lambda}\right)^{\mathcal{E}} \left(\frac{q}{\Lambda}\right)^{\mathcal{D}} \left(\frac{1}{4\pi}\right)^{2\mathcal{L}} . \tag{3.13}$$

[3] The factor of π^2 is clearest to see if momenta are Wick rotated to euclidean signature, since it there represents the volume of the unit 3-sphere corresponding to the integration over the three directions taken by a 4-vector.

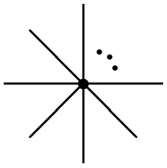


Fig. 3.1 The graph describing the insertion of a single effective vertex with $\mathcal{E}$ external lines and no internal lines.

Since $\mathcal{E}$ and $\mathcal{D}$ are fixed by external characteristics (the total number of external legs and power of q in the final answer) the last factor is the only part of (3.13) that changes for more and more complicated diagrams that share these external properties. This factor simply says that $1/(4\pi)^2$ is the price for each additional loop; that is to say, all graphs with a fixed number of loops are similar in size assuming the (unwritten) dimensionless couplings — *i.e.* the $\hat{c}_n$ of (3.2) — are also similar in size. Furthermore, perturbation theory in this regime is ultimately controlled by the ratio of the dimensionless $\hat{c}_n$ compared with $16\pi^2$. The statement that perturbation theory applies for small enough $\hat{c}_n$ agrees with (and refines) the simple estimate of §2.4.1.

Validity of the perturbative expansion

More broadly, eq. (3.12) outlines the domain of validity of the perturbative expansion itself. If the contribution estimated in eq. (3.12) is small for all choices of $\mathcal{D}$, $\mathcal{E}$, $\mathcal{L}$ and $\mathcal{V}_n$, then this ensures that the perturbative expansion used in its derivation is a good approximation, particularly if more complicated graphs (higher $\mathcal{L}$ and $\mathcal{V}_n$) are more suppressed than less complicated ones. Conversely, if there is a choice for $\mathcal{D}$, $\mathcal{E}$, $\mathcal{L}$ and $\mathcal{V}_n$ for which eq. (3.12) is not small, then the perturbative expansion fails unless some other small parameter — such as the dimensionless couplings $\hat{c}_n$ of (3.2) — can be found that can systematically suppress more complicated graphs. Furthermore, since the semiclassical expansion is an expansion in loops,[4] the perturbative expansion becomes a semiclassical expansion when it is the $\mathcal{L}$-dependent factor that controls perturbation theory.

Eq. (3.12) shows that there are three small quantities whose size can help control perturbative corrections: q/Λ, Λ/M and $\Lambda M/4\pi\dagger^2$. Some remarks are in order for each of these.

Derivative expansion

Consider first the factor q/Λ, which controls the suppression of higher powers of external momenta. There is no question that $q/\Lambda \ll 1$ since the entire construction of the low-energy theory presupposes Λ can be chosen much smaller than the scale of the heavy

[4] The connection between loops and the semiclassical expansion is established in the discussion surrounding (2.24). In essence the semiclassical expansion counts loops because it is an expansion in powers of $\hbar$, which appears as an overall factor in the quantity $S/\hbar$ within the path integral.

physics that is being integrated out, but much larger than the low energies, $q \simeq E$ of applications. But when $\Lambda \ll M$ the ratio q/Λ is much bigger than q/M, even if both are separately very small. Eq. (3.12) therefore shows that it could happen that the derivative expansion in physical quantities (like scattering amplitudes) and in quantities like S_w ends up being controlled by powers of q/Λ rather than the powers of q/M assumed for the original action, (3.2). This point is returned to in §3.1 below, but it means that (all other things being equal) derivatives like to be suppressed by the lowest possible UV scale available: in this case Λ. The case $M \simeq \Lambda$ is one that should be taken seriously in what follows.

Field expansion

Notice that the same thing does *not* happen for the expansion in powers of ϕ/v, since the factor $(1/v)^{\mathcal{E}}$ assumed to appear in S_w does not get converted into a power like $(1/\Lambda)^{\mathcal{E}}$ or $(1/M)^{\mathcal{E}}$ in $\mathcal{A}_{\mathcal{E}}$. This means that it is consistent to have the scale v that controls the field expansion be systematically different from the scales Λ or M that control the derivative expansion. These scales are logically independent because in general large fields need not imply large energies, so small-field expansions are not necessarily required in a low-energy limit.

Loop expansion

Next consider the suppressions coming from loops and vertices. Notice first that if $M \simeq \Lambda$ then the only systematic perturbative suppression in (3.12) comes from loops, due to the factor $(\Lambda^2/4\pi\mathfrak{f}^2)^{2\mathcal{L}}$. If all dimensionless couplings are order unity then perturbation theory in this limit is revealed to be a semiclassical expansion (*i.e.* controlled purely by the number of loops) whose validity rests on the assumption $4\pi\mathfrak{f}^2 \gg \Lambda^2$.

This condition is automatically satisfied in the regime $\Lambda \ll M$ provided that $\mathfrak{f} \gtrsim M$ is also true. It is a much stronger condition on Λ, however, if $\mathfrak{f}$ should be much smaller than M. In the particularly interesting case where $\mathfrak{f}^2 \simeq Mv$ (corresponding to canonical normalization in (3.2)) the loop-suppression factor becomes $(M\Lambda/4\pi\mathfrak{f}^2)^{2\mathcal{L}} \simeq (\Lambda/4\pi v)^{2\mathcal{L}}$.

Dangerous non-derivative interactions

Finally consider the final factor in (3.12). If $\Lambda \lesssim M$ the power of $(\Lambda/M)^{\mathcal{P}}$ appearing in eq. (3.12) represents a suppression rather than an enhancement provided the power

$$\mathcal{P} := 2 + \sum_n (d_n - 2)\mathcal{V}_n \,, \tag{3.14}$$

is non-negative. It is this factor that expresses the suppression of the effects of interactions involving three or more derivatives.

Lorentz invariance often requires d_n to be even (*e.g.* for scalar fields), and in this case it is only interactions with no derivatives at all ($d_n = 0$) for which $\mathcal{P}$ can be negative. These interactions are potentially dangerous in that they can in principle allow an *enhancement* in $\mathcal{A}_{\mathcal{E}}$ when $\Lambda \ll M$. When such interactions exist a more detailed estimate is required to see whether higher-order effects really are suppressed.

As an example of non-derivative interactions, imagine the low-energy field, ϕ, self-

interacts through a scalar potential,

$$S_{\scriptscriptstyle W} \supset -\int \mathrm{d}^4x\; V(\phi)\,, \tag{3.15}$$

where

$$V(\phi) := \mathfrak{f}_{\scriptscriptstyle V}^4 \sum_r g_r \left(\frac{\phi}{v}\right)^r . \tag{3.16}$$

Here g_r are dimensionless couplings and $\mathfrak{f}_{\scriptscriptstyle V}^4$ is the typical potential energy density associated with fields of order $\phi \simeq v$. If $\mathfrak{f}_{\scriptscriptstyle V} \neq \mathfrak{f}$ then repeating the above power-counting argument shows that each appearance of a vertex drawn from $V(\phi)$ contributes an additional factor $g_r(\mathfrak{f}_{\scriptscriptstyle V}/\mathfrak{f})^4$, modifying eq. (3.12) to

$$\begin{aligned}\mathcal{A}_{\mathcal{E}\mathcal{D}}\, q^{\mathcal{D}} \sim \frac{\Lambda^2\mathfrak{f}^4}{M^2}\left(\frac{1}{v}\right)^{\mathcal{E}}\left(\frac{q}{\Lambda}\right)^{\mathcal{D}}\left(\frac{M\Lambda}{4\pi\,\mathfrak{f}^2}\right)^{2\mathcal{L}} \\ \times\left\{\prod_r\left[g_r\left(\frac{\mathfrak{f}_{\scriptscriptstyle V}^4M^2}{\mathfrak{f}^4\Lambda^2}\right)\right]^{\mathcal{V}_{0,r}}\right\}\left\{\prod_{d\geq 2}\prod_{i_d}\left(\frac{\Lambda}{M}\right)^{(d-2)\mathcal{V}_{d,i_d}}\right\},\end{aligned} \tag{3.17}$$

where the product over vertex labels is now subdivided into groups involving precisely d derivatives: $\{n\} = \{d, i_d\}$, with $i_0 = r$.

This last expression shows that the potentially hazardous enhancement factor, $(M/\Lambda)^{2\mathcal{V}_{0,r}}$, need not be dangerous if the potential energy density in the low-energy theory is sufficiently small relative to the generic energy density, $\mathfrak{f}_{\scriptscriptstyle V}^4/\mathfrak{f}^4 \lesssim \Lambda^2/M^2$. But if this is not so, generic non-derivative interactions can be legitimate obstructions to having a well-behaved low-energy limit, a point that must be checked on a case-by-case basis.

Example: the toy model

The Wilson action for the toy model, eq. (2.97), is a special case of the general form assumed in eqs. (3.2), with $M = m_{\scriptscriptstyle R}$ and $\mathfrak{f}^2 = m_{\scriptscriptstyle R}v$ and no zero-derivative interactions for the low-energy field ξ. For this special case the estimate eq. (3.12) becomes

$$\mathcal{A}_{\mathcal{E}\mathcal{D}}\, q^{\mathcal{D}} \sim v^2\Lambda^2\left(\frac{1}{v}\right)^{\mathcal{E}}\left(\frac{q}{\Lambda}\right)^{\mathcal{D}}\left(\frac{\Lambda}{4\pi\,v}\right)^{2\mathcal{L}}\left(\frac{\Lambda}{m_{\scriptscriptstyle R}}\right)^{\sum_n(d_n-2)\mathcal{V}_n}, \tag{3.18}$$

which neglects dimensionless factors, that come as a series in powers of the coupling λ.

Notice that eq. (3.18) agrees with the calculations of the previous chapter for the size of tree and loop contributions to the effective vertex, $a_{\rm le}(\partial_\mu\xi\,\partial^\mu\xi)^2$ appearing in $\Gamma_{\rm le}$, for which $\mathcal{E} = \mathcal{D} = 4$. For instance, consider the three contributions of Fig. 2.5. Figure (a) has $\mathcal{L} = 0$ and $\mathcal{V}_{4,4} = 1$, and so eq. (3.18) gives

$$\begin{aligned}\delta\, a_{\rm le} &\sim v^2\Lambda^2(1/v)^4(1/\Lambda)^4(\Lambda/m_{\scriptscriptstyle R})^2 \\ &\sim 1/(v^2m_{\scriptscriptstyle R}^2) = \lambda/m_{\scriptscriptstyle R}^4\,;\end{aligned} \tag{3.19}$$

Figure (b) has $\mathcal{L} = 1$ and $\mathcal{V}_{4,4} = 2$, and so eq. (3.18) gives

$$\begin{aligned}\delta\, a_{\rm le} &\sim v^2\Lambda^2(1/v)^4(1/\Lambda)^4(\Lambda/4\pi v)^2(\Lambda/m_{\scriptscriptstyle R})^4 \\ &\sim [1/(16\pi^2v^4)](\Lambda/m_{\scriptscriptstyle R})^4\,;\end{aligned} \tag{3.20}$$

Figure (c) has $\mathcal{L} = 1$ and $\mathcal{V}_{6,6} = 1$, and so eq. (3.18) gives

$$\begin{aligned} \delta a_{\rm le} &\sim v^2\Lambda^2(1/v)^4(1/\Lambda)^4(\Lambda/4\pi v)^2(\Lambda/m_R)^4 \\ &\sim [1/(16\pi^2 v^4)](\Lambda/m_R)^4 \,. \end{aligned} \tag{3.21}$$

These all agree with the estimates performed in §2.3, above.

But eq. (3.18) contains much more information than just this. Most importantly, since the symmetry $\xi \to \xi +$ constant implies there are no interactions with $d < 4$ it follows that higher-order graphs are always suppressed by positive powers of the small ratios $\Lambda/4\pi v$ or Λ/m_R. The small size of these ratios is what is responsible for weak coupling in the low-energy theory, showing that it is the derivative coupling of Goldstone bosons at low energies that allows S_W to be treated perturbatively (and *not* the size of the coupling λ in the underlying theory).

The toy model also provides insight into the relationship between the scales M and v that respectively control the derivative and field expansions in the Wilson action. To see why recall that $m_R^2 = \lambda v^2$, so the perturbative semiclassical regime $\lambda \ll 1$ is where the scales $M = m_R$ and v are very different from one another. Yet even so, eq. (3.18) demonstrates that each external line (which, for calculations of $\Gamma_{\rm le}$, counts the power of ϕ) is accompanied by at least one power of $1/v$. Quantum corrections do not change the fact that it is always the size of ϕ/v that controls the field expansion, regardless of the scale appearing in the low-energy expansion.

3.1.3 The exact renormalization group

Another instructive use of (3.12) is to estimate how the couplings in S_W themselves evolve as Λ changes. At first sight this might seem surprising, since the evolution of couplings in S_W are obtained by starting from the underlying high-energy theory and integrating out all physics at energy scales above Λ. For this calculation, however, Λ is the *lowest* scale in the integration rather than the highest scale, so naively the dimensional arguments leading to (3.12) (which take Λ to be the *largest* scale in all integrals) might seem not to apply.

The estimate (3.12) is nevertheless useful because any Λ-dependence acquired by S_W when integrating out modes with energies larger than Λ must ultimately be cancelled when S_W is used to integrate out the remaining modes with energies smaller than Λ — to which (3.12) does apply.

This can be formalized by comparing the result obtained when computing something like the 1LPI generator, $\Gamma_{\rm le}[\phi;\Lambda]$, using the Wilson action defined at a cutoff scale Λ with that obtained with the Wilson action defined at a slightly lower cutoff $S_W[\phi,\Lambda']$, with $\Lambda' = \Lambda - {\rm d}\Lambda$. Either of these is an equally good starting point for computing $\Gamma_{\rm le}$, since this is Λ-independent. (It could, after all, have been computed for the full theory without ever dividing the problem into a contribution from above and below the scale Λ).

For example, for scalar fields when used with (2.59) this implies

$$0 = \Lambda\frac{\rm d}{{\rm d}\Lambda}e^{{\rm i}\Gamma_{\rm le}[\varphi]} = \Lambda\frac{\rm d}{{\rm d}\Lambda}\int \mathcal{D}\phi\, e^{{\rm i}S_W[\varphi+\phi;\Lambda]+\int {\rm d}^4x\, j_a\phi^a}\,, \tag{3.22}$$

where $j_a = -\delta\Gamma_{\rm le}[\varphi]/\delta\varphi^a$ has support only for modes well below Λ. In detail the full result

is independent of Λ because the Λ-dependence of S_w cancels the Λ-dependence implicit in the measure $\mathcal{D}\phi$, since this includes a functional integration only over modes with energies below Λ.

In practice it is useful to implement the cutoff excluding modes larger than Λ from the path integral by suitably modifying the Wilson action.[5] For instance, in perturbation theory high-energy modes can be suppressed in internal lines by Wick rotating to imaginary time and introducing a cutoff function into the unperturbed action. Writing $S_w = S_{w0} + S_{w,\rm int}$ one takes

$$\begin{aligned} S_{w0} &= -\frac{1}{2}\int \mathrm{d}^4x\,\mathrm{d}^4x'\;\mathcal{K}_\Lambda(x-x')\Big[\partial_\mu\phi(x)\,\partial'^\mu\phi(x') + m^2\phi(x)\,\phi(x')\Big] \\ &= -\frac{1}{2}\int \frac{\mathrm{d}^4p}{(2\pi)^4}\,\phi(p)\phi(-p)(p^2+m^2)K^{-1}(p^2/\Lambda^2)\,, \end{aligned} \tag{3.23}$$

where the kernel $\mathcal{K}_\Lambda(x-x')$ is defined by its Fourier transform

$$\mathcal{K}_\Lambda(x-x') = \int \frac{\mathrm{d}^4p}{(2\pi)^4}\;K^{-1}(p^2/\Lambda^2)\;e^{\mathrm{i}p\cdot(x-x')}\,. \tag{3.24}$$

$K(p^2/\Lambda^2)$ is a smooth step-like cutoff function that satisfies $K(u) = 1$ for $u \ll 1$ and $K(u) \to 0$ for $u \gg 1$.

Once this is done eq. (3.22) can be read as a differential equation governing how $S_w[\phi;\Lambda]$ depends on Λ, called the 'exact renormalization group'. When the derivative hits the factor $K(p^2/\Lambda^2)$ in a propagator $K(p^2/\Lambda^2)/(p^2+m^2)$ the result has support only for $p^2 \simeq \Lambda^2$, effectively removing the corresponding internal line from the given Feynman graph.

The change wrought by this in the path integral must be compensated by appropriately modifying the interactions, and this is what defines the flow of effective couplings with Λ (along the lines illustrated in Fig. 3.2). This turns out to imply an evolution equation for the interaction lagrangian density, $S_{w,\rm int}$, of the form [25, 26, 27]

$$\Lambda\,\frac{\mathrm{d}S_{w,\rm int}}{\mathrm{d}\Lambda} = -\frac{1}{2}\int \mathrm{d}^4p\left[\frac{(2\pi)^4}{p^2+m^2}\right]\Lambda\,\frac{\partial K}{\partial\Lambda}\left[\frac{\delta S_{w,\rm int}}{\delta\phi(p)}\;\frac{\delta S_{w,\rm int}}{\delta\phi(-p)} + \frac{\delta^2 S_{w,\rm int}}{\delta\phi(p)\,\delta\phi(-p)}\right]. \tag{3.25}$$

The size of the resulting changes to the effective couplings in $S_{w,\rm int}$ can be estimated using (3.12). To this end suppose the interaction terms can be written as

$$\begin{aligned} S_{w,\rm int} &= -\Lambda^2 v^2 \sum_n \frac{\hat{c}_n}{\Lambda^{d_n} v^{f_n}} \int \mathrm{d}^4x\; \mathcal{O}_n^{(d_n,f_n)}(\phi) \\ &= -\int \mathrm{d}^4x\left[\frac{\hat{c}_{4,0}\Lambda^2}{v^2}\,\phi^4 + \frac{\hat{c}_{4,2}}{v^2}\,\phi^2(\partial\phi)^2 + \cdots\right], \end{aligned} \tag{3.26}$$

where the $\mathcal{O}_n^{(d_n,f_n)}$ describe all possible local interactions involving f_n powers of the fields and d_n derivatives, and the last line specializes to a single scalar field for concreteness' sake.

The power-counting result, (3.12), provides an estimate of the size of amputated Feynman graphs built using these interactions, involving fields defined below the cutoff Λ. But

[5] If cutoffs are instead implemented directly for the integrations in Feynman graphs the results can depend on the way virtual momentum is routed through the graph.

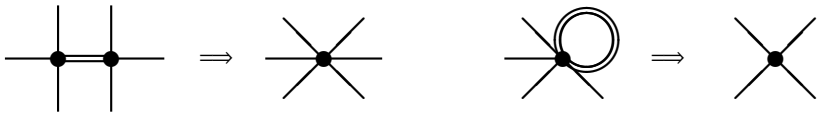


Fig. 3.2 Graphs illustrating the two effects that occur when an internal line is contracted to a point, depending on whether or not the propagator connects distinct vertices (left two figures) or ties off a loop on a single vertex (right two figures). In both cases a double line represents the differentiated propagator. The two options respectively correspond to the terms $[\delta S_{w,\,\rm int}/\delta\phi(p)][\delta S_{w,\,\rm int}/\delta\phi(-p)]$ and $\delta^2 S_{w,\,\rm int}/\delta\phi(p)\delta\phi(-p)$ appearing in the Wilson-Polchinski relation, eq. (3.25) of the text.

this also determines the Λ-dependence of perturbative corrections to the couplings in $S_{w\rm int}$ because the direct contribution of interactions represented by graphs like Fig. 3.1 must precisely cancel the Λ-dependence coming from loop graphs, as estimated by eq. (3.12).

For the action of (3.26) the estimate (3.12) gives the contribution to the effective coupling of a term in S_w involving $\mathcal{E}$ powers of ϕ and $\mathcal{D}$ derivatives to be

$$\delta\left[\hat{c}_n v^2\Lambda^2\left(\frac{1}{v}\right)^{\mathcal{E}}\left(\frac{1}{\Lambda}\right)^{\mathcal{D}}\right] \sim v^2\Lambda^2\,\left(\frac{1}{v}\right)^{\mathcal{E}}\left(\frac{1}{\Lambda}\right)^{\mathcal{D}}\left(\frac{\Lambda}{4\pi\, v}\right)^{2\mathcal{L}}\,, \tag{3.27}$$

and so $\delta\hat{c}_n$ acquires corrections from $\mathcal{L}$-loop graphs that are of order

$$\delta\,\hat{c}_n \sim \left(\frac{\Lambda}{4\pi\, v}\right)^{2\mathcal{L}} \times \quad \text{(combinations of other } \hat{c}_n\text{'s)}\,. \tag{3.28}$$

If $v \gtrsim \Lambda$ this shows that it is consistent to have the $\hat{c}_n$'s all be generically $\lesssim 1$ for all Λ. Some couplings can be much smaller than this if $v \gg \Lambda$ (or other hierarchies like powers of Λ/M or small dimensionless couplings are buried in the $\hat{c}_n$'s), provided these additional suppressions preserve any initially small values.

Log running vs power-law running

The exact cutoff-dependent renormalization group is not pursued further in this book, since the focus here is instead on more practical methods of approximate calculation. Before leaving the subject, though, it is useful to address a conceptual question and by so doing contrast the implications of logarithmic and power-law running of effective couplings, $\hat{c}_n(\Lambda)$, as Λ is varied.

The conceptual question is this: why does one care how couplings run with Λ if Λ itself ultimately does not appear in any physical results? It is emphasized many times in this book that Λ enters calculations purely as a convenient book-keeping device: it is useful to organize calculations by scales and integrate out physics one scale at a time. But in the

end physical quantities are obtained only after *all* scales are integrated out, after which the arbitrary separations between these scales disappear.

This section follows [28] to argue that understanding the running of couplings in S_W is useful to the extent that it helps track the dependence of physical quantities on large physical ratios of scale, M/m. In particular, there is often a precise connection between logarithmic dependence of low-energy quantities on the cutoff Λ and a logarithmic dependence on physical scales. The analogous connection is only qualitative for power-law dependence, however, and so is usually less useful.

To see why this is so imagine a system characterized by two very different scales, $m \ll M$, such as the masses of two different particles. Further imagine that there is a physical quantity, A, whose dependence on M/m happens to be logarithmic, so

$$A = a_0 \, \ln\left(\frac{M}{m}\right) + a_1 \,, \tag{3.29}$$

for some calculable constants a_0 and a_1. If both of these constants are similar in size then the value of a_0 can be important in practice since the large logarithm can make it dominate numerically in the total result.

Next suppose a Wilsonian calculation is performed that divides the contributions to A coming from physics above and below the scale Λ, with $m \ll \Lambda \ll M$. How does the large logarithm get into the low-energy part of the theory, given that it depends on scales that lie on opposite sides of Λ? Typically this happens as follows:

$$\begin{aligned} A_{\rm le} &= a_{0\,\rm le} \, \ln\left(\frac{\Lambda}{m}\right) + a_{1\,\rm le} \\ A_{\rm he} &= a_{0\,\rm he} \, \ln\left(\frac{M}{\Lambda}\right) + a_{1\,\rm he} \\ \text{so that} \quad A &= A_{\rm le} + A_{\rm he} = a_0 \, \ln\left(\frac{M}{m}\right) + a_1 \,. \end{aligned} \tag{3.30}$$

The requirement that Λ cancels implies $a_{0\,\rm le} = a_{0\,\rm he}$ and then having the results agree with the full theory implies $a_{0\,\rm le} = a_{0\,\rm he} = a_0$ and $a_1 = a_{1\,\rm le} + a_{1\,\rm he}$. What is significant is that the coefficient, a_0, of the logarithm in the full answer is calculable purely within the low-energy theory because Λ-cancellation dictates that $a_{0\,\rm le} = a_0$.

The same is not so for power-law dependence. Suppose for example that another observable, B, is computed that depends quadratically on masses, so

$$B = b_0 \, M^2 + b_1 \, m^2 \,. \tag{3.31}$$

Again the coefficient b_0 is of practical interest since the large size of M can make this term dominate numerically. In this case the low- and high-energy parts of the calculations instead are

$$\begin{aligned} B_{\rm le} &= b_{0\,\rm le} \, \Lambda^2 + b_{1\,\rm le} \, m^2 + \cdots \\ B_{\rm he} &= b_{0\,\rm he} \, M^2 + b_{1\,\rm he} \, \Lambda^2 + \cdots \\ \text{so that} \quad B &= B_{\rm le} + B_{\rm he} = b_0 \, M^2 + b_1 \, m^2 \,, \end{aligned} \tag{3.32}$$

with $b_0 = b_{0\,\rm he}$ and $b_1 = b_{1\,\rm le}$ while Λ-cancellation requires $b_{0\,\rm le} + b_{1\,\rm he} = 0$.

Evidently the b_0 term cannot be computed purely within the low-energy part of a Wilsonian calculation simply by tracking the dependence on Λ^2, unlike the way the $\ln\Lambda$ terms reproduce the value for a_0. This is a fairly generic result: quantitative predictions for quantities like b_0 really require detailed knowledge of the UV theory and cannot be computed using the low-energy Wilsonian theory alone. But logarithms can often be inferred purely from within the low-energy Wilsonian perspective. For this reason considerable attention is given to renormalization-group methods that allow efficient extraction of large logarithms using Wilsonian EFTs.

Method of regions I

The cancellations of powers of cutoff and the utility of logarithms can be promoted to a useful tool – sometimes called the *method of regions* [29] – for estimating Feynman integrals in situations where different integration regions compete in their contributions to the final result.

To illustrate the method, follow [30] and consider the integral

$$\int_0^\infty \frac{k\,\mathrm{d}k}{(k^2+m^2)(k^2+M^2)} = \frac{\ln(M/m)}{M^2-m^2} \simeq \frac{\ln(M/m)}{M^2}\left[1+\frac{m^2}{M^2}+\cdots\right], \tag{3.33}$$

and imagine trying to extract the dominant small-m/M expansion without first evaluating the full integral. Naively one simply Taylor expands the integrand in powers of m and integrates term-by-term, but this has the problem that each term involves an integral that diverges in the infrared

$$\int_0^\infty \frac{k\,\mathrm{d}k}{k^2(k^2+M^2)}\left[1-\frac{m^2}{k^2}+\cdots\right], \tag{3.34}$$

as might be expected given that the full result (3.33) is not analytic at $m=0$.

A better procedure instead separates the integral into two regions: an IR region $0<k<\Lambda$ and a UV region $k>\Lambda$ and expands the integrand differently in each. For the IR the integrand is expanded with $k^2 \sim m^2 \ll M^2$ while in the UV one takes instead $m \ll k \sim M$. This leads to the result $\mathcal{I}(m,M) = \mathcal{I}^{IR}(m,M,\Lambda) + \mathcal{I}^{UV}(m,M,\Lambda)$ with

$$\mathcal{I}^{IR} = \int_0^\Lambda \frac{k\,\mathrm{d}k}{(k^2+m^2)M^2}\left[1-\frac{k^2}{M^2}+\cdots\right] \simeq \frac{\ln(1+\Lambda^2/m^2)}{2M^2} - \frac{\Lambda^2}{2M^4}+\cdots, \tag{3.35}$$

and

$$\mathcal{I}^{UV} = \int_\Lambda^\infty \frac{k\,\mathrm{d}k}{k^2(k^2+M^2)}\left[1-\frac{m^2}{k^2}+\cdots\right] \simeq \frac{\ln(1+M^2/\Lambda^2)}{2M^2} + \mathcal{O}(m^2). \tag{3.36}$$

Once these last two formulae are summed all Λ-dependence cancels (as it must), leaving residual logarithms of M/m as outlined above that reproduce the expansion of (3.33). Large logarithms like $\ln(M/m)$ are ultimately leftovers from the cancellation between the IR divergence of $\mathcal{I}^{UV}$ and the UV divergence of $\mathcal{I}^{IR}$.

3.1.4 Rationale behind renormalization ♢

The above discussion about integrating out high-energies also provides physical insight into the entire framework of renormalization. This is because a central message is that the scale Λ is ultimately a calculational convenience that drops out of all physical quantities. In detail Λ drops out because of a cancellation between: (*i*) the explicit Λ-dependence of the cutoff on the limits of integration for virtual low-energy states in loops, and (*ii*) the cutoff-dependence that is implicitly contained within the effective couplings of $\mathfrak{L}_w$.

But this cancellation is eerily reminiscent of how UV divergences are traditionally handled within *any* renormalizable theory, and in particular for the underlying UV theory from which S_w is calculated. The entire renormalization program relies on any UV-divergent cutoff-dependence arising from regulated loop integrals being cancelled by the regularization dependence of the counterterms of the renormalized lagrangian. There are, however, the following important differences.

1. The cancellations in the effective theory occur even though Λ is not sent to infinity, and even though $\mathfrak{L}_w$ contains arbitrarily many terms that are not renormalizable in the traditional sense.
2. The cancellation of regularization dependence in the traditional picture of renormalization appears completely ad-hoc and implausible, while the cancellation of Λ from observables within the effective theory is essentially obvious. It is obvious due to the fact that Λ only was introduced as an intermediate step in a calculation, and so cannot survive uncancelled in the answer.

This resemblance is likely not accidental. It suggests that rather than considering a model's classical lagrangian as something pristine or fundamental, it is better regarded as an effective lagrangian obtained by integrating out still-more-microscopic degrees of freedom. The cancellation of ultraviolet divergences within the renormalization program is within this interpretation simply the usual removal of an intermediate step in a calculation to whose microscopic part we are not privy.

This is the modern picture of what renormalization really means. When discovering successful theories what is found is not a 'classical' action, to be quantized and compared with experiment. What is found is really a Wilsonian action describing the low-energy limit obtained by integrating out high-energy degrees of freedom in some more fundamental theory that describes what is really going on at much, much higher energies.

It is this Wilsonian theory, itself potentially already containing many high-energy quantum effects, whose low-energy states are quantized and compared with observations. Physics progresses by successively peeling back layer after layer of structure in nature, and our mathematics describes this through a succession of Wilsonian descriptions with ever-increasing accuracy.

This is how real progress often happens in science. Efforts to solve concrete practical questions — in this case a desire to exploit hierarchies of scale as efficiently as possible — can ultimately provide deep insights about foundational issues — in this case about what it is that is really achieved when new fundamental theories (be it Maxwell's equations, General Relativity or the Standard Model) are discovered.

3.2 Power counting and dimensional regularization ◊

As previous sections make clear, there is a lot of freedom of definition when setting up a Wilson action: besides the freedom to make field redefinitions, there are also all the details of precisely how the differentiate between scales above and below Λ. Physical results do not depend on any of these choices at all since observables are independent of field redefinitions and are blind to the details of a regularization scheme. This freedom should be exploited to make the Wilson action as useful as possible for practical calculations. In particular it should be used to optimize the efficiency with which effective interactions and Feynman graphs can be identified that completely capture the contributions to low-energy processes at any fixed order in low-energy expansion parameters like q/M.

Though instructive, the power-counting analysis of the previous section does not yet do this, due to the appearance in all estimates of the cutoff Λ. Since Λ ultimately cancels in all physical quantities, it is inconvenient to have to rely on it when estimating the size of contributions from different interactions in the low-energy Wilson action. For this reason most practical applications (and most of the rest of this book) define the Wilson action using dimensional regularization rather than cutoffs [31, 32]. Dimensional regularization is useful because it is both simple to use and preserves more symmetries than do other regularization schemes. This section explores how this can be done.

3.2.1 EFTs in dimensional regularization

At first sight it seems impossible to define a Wilson action in terms of dimensional regularization at all. After all, the entire purpose of the Wilson action is to encapsulate efficiently the high-energy part of a calculation, for later use in a variety of low-energy applications. This seems to *require* something like Λ to distinguish high energies from low energies. By contrast, although dimensional regularization is designed to regulate UV-divergent integrals, it does not do so by cutting them off at large momenta and energies. The regularization is instead provided by defining the integral (including contributions from arbitrarily large momenta) for complex dimension, D, taking advantage of the fact that the integral converges in the ultraviolet if D is sufficiently small or negative. The result still diverges in the limit $D \to 4$, but usually as a pole or other type of isolated singularity when D is a positive integer. The limit $D \to 4$ is taken at the end of a calculation, after any singularities are absorbed into the renormalization of the appropriate couplings.

This section describes how dimensional regularization can nonetheless be used to define a Wilson action, despite it not seeming to explicitly separate high from low energies. This is done first by briefly describing dimensional regularization itself, followed by a presentation of the logic of constructing an effective theory using it. (See also Appendix §A.2.4 for more details about dimensional regularization.)

What is dimensional regularization?

Consider the following integral over D-dimensional Euclidean momentum p^μ, where $p^2 = \delta_{\mu\nu}p^\mu p^\nu$ (and similarly for q^2),[6] [33, 34]

$$I_D^{(A,B)}(q) := \int \frac{\mathrm{d}^D p_E}{(2\pi)^D}\left[\frac{p^{2A}}{(p^2+q^2)^B}\right] = \frac{1}{(4\pi)^{D/2}}\left[\frac{\Gamma\left(A+\frac{D}{2}\right)\Gamma\left(B-A-\frac{D}{2}\right)}{\Gamma(B)\,\Gamma\left(\frac{D}{2}\right)}\right]\left(q^2\right)^{A-B+D/2}, \tag{3.37}$$

where $\Gamma(z)$ is Euler's Gamma function, defined to satisfy $z\,\Gamma(z) = \Gamma(z+1)$ with $\Gamma(n+1) = n!$ when restricted to positive integers, n. This integral converges in the UV if the real part of $2(B-A)$ is bigger than D, and the right-hand side is the result of explicit evaluation.

Dimensional regularization uses the right-hand side to *define* this integral for values of A, B and D for which it does not converge, even when D is not an integer. In dimensional regularization D is regarded as complex during intermediate steps, with $D \to 4$ taken at the end of the calculation. It happens that $\Gamma(z)$ is analytic for all complex z apart from poles at the non-positive integers, and so when A and B are positive integers eq. (3.37) provides a definition of $I_D(q)$ that is finite for all complex D apart from possible poles when D is a positive even integer.

In particular, integrals defined in this way typically have divergences that arise as poles at $D = 4$. For instance, a useful example encountered earlier is the case $A = 0$ and $B = 2$,

$$\int \frac{\mathrm{d}^D p}{(2\pi\mu)^D}\left[\frac{\mu^4}{(p^2+q^2)^2}\right] = \frac{\Gamma\left(2-\frac{D}{2}\right)}{(4\pi)^{D/2}}\left(\frac{q^2}{\mu^2}\right)^{(D-4)/2} = \frac{1}{16\pi^2}\left[\frac{2}{4-D} - \gamma + \frac{1}{2}\ln\left(\frac{q^2}{4\pi\mu^2}\right) + \cdots\right], \tag{3.38}$$

where the ellipses represent terms that vanish when $D \to 4$; μ is an arbitrary scale included on dimensional grounds and $\gamma \simeq 0.577215664901532\ldots$ is the Euler-Mascheroni constant. The pole as $D \to 4$ reflects the logarithmic divergence that would have been present if the integral were to be defined with $D = 4$ from the get-go.

Similarly, the integral with $A = B = k$ gives

$$\begin{aligned}\int \frac{\mathrm{d}^D p}{(2\pi\mu)^D}\left[\frac{p^{2k}}{(p^2+q^2)^k}\right] &= \frac{\Gamma\left(k+\frac{D}{2}\right)\Gamma\left(-\frac{D}{2}\right)}{\Gamma(k)\,\Gamma\left(\frac{D}{2}\right)}\left(\frac{q^2}{4\pi\mu^2}\right)^{D/2} \\ &= \left(\frac{q^2}{4\pi\mu^2}\right)^2\left[\frac{2}{4-D} - \gamma_k + \frac{1}{2}\ln\left(\frac{q^2}{\mu^2}\right) + \cdots\right],\end{aligned} \tag{3.39}$$

where γ_k is a k-dependent pure number, whose precise value is not important for later purposes. This is an example of an integral that would diverge like a power of the cutoff if directly evaluated at $D = 4$. Notice that the dimensionally regularized version vanishes as $q^2 \to 0$, because the integrand has no other scale to which the result can be proportional.[7]

Because momenta get integrated from $-\infty$ to ∞ in dimensional regularization, both high

[6] Such Euclidean expressions are obtained by Wick rotating with $\mathrm{d}^4 p = \mathrm{i}\mathrm{d}^4 p_E$, meaning that the Minkowski-signature result has an additional factor of i. Notice that there are no additional explicit signs in continuing positive q^2 from Euclidean to Minkowski signature because of the wisdom of using conventions with a $(-+++)$ metric.

[7] The result cannot be proportional to μ when $D = 4$ since this scale is introduced in such a way as to ensure that the integral is proportional to μ^{D-4}.

and low energies are explicitly included. This makes its use seem contrary to the entire philosophy of defining a low-energy effective theory. But the utility of effective field theories is founded on the observation that *any* contribution of generic high-energy dynamics to low-energy amplitudes can be captured within the low-energy sector by an appropriate collection of local effective interactions. Since the error made in dimensional regularization by keeping all modes up to infinite energies is itself a particular choice of high-energy physics, any damage done can also be undone using an appropriate choice of effective couplings in the low-energy theory.

To see how the Wilson action is defined in dimensional regularization, consider a theory containing a light field, ϕ, of mass m, and a heavy field, ψ, of mass $M \gg m$. For the purposes of argument the full theory can be imagined to be a renormalizable theory coupling ϕ to ψ, $S = S[\phi, \psi]$, regularized using dimensional regularization and then renormalized in any convenient way (such as with the modified minimal subtraction – or $\overline{\text{MS}}$ – scheme).[8]

The low-energy applications of interest in this model are to $E \ll M$. The effective Wilsonian action in this regime contains only the light field, $S_w = S_w[\phi]$. Just like the full theory this effective theory is also dimensionally regularized (and renormalized in a way specified below). In practice this means that the dimensionally-regularized effective theory is not obtained by explicitly integrating out successively higher-energy modes of all the fields in the underlying theory. Instead the dimensionally regularized effective theory simply omits the heavy field ψ.

A convenient renormalization choice for the effective couplings in S_w demands that the predictions of $S_w[\phi]$ agree with the low-energy predictions of $S[\phi, \psi]$ order-by-order in $1/M$. That is, the renormalized effective couplings of the low-energy theory are obtained by performing a *matching* calculation, whereby the couplings of the low-energy effective theory are chosen to reproduce scattering amplitudes or Greens functions of the underlying theory order-by-order in powers of the inverse heavy scale, $1/M$. Once the couplings of the effective theory are determined in this way in terms of those of the underlying fundamental theory, they may be used to compute any other purely low-energy observable.

Method of regions II

Before seeing how this all works in detail, it is worth first pausing here to see how dimensional regularization works when exploring how integrals depend on large hierarchies of parameters. To this end consider again the sample integral, (3.33), examined above when discussing the method of regions. In dimensional regularization this integral would have the form

$$\mathcal{I}_\epsilon(m, M) := \int_0^\infty \left(\frac{\mu}{k}\right)^\epsilon \frac{k\,\mathrm{d}k}{(k^2+m^2)(k^2+M^2)} = \Gamma\left(1-\frac{\epsilon}{2}\right)\Gamma\left(\frac{\epsilon}{2}\right)\frac{(m/\mu)^{-\epsilon}-(M/\mu)^{-\epsilon}}{M^2-m^2}\,, \tag{3.40}$$

where $\epsilon \propto D-4$ is a parameter that is taken to zero at the end. The limit $\epsilon \to 0$ is nonsingular in this case because the integral of (3.33) converges when $D = 4$, with $\mathcal{I}_\epsilon = \ln(M/m)/(M^2-m^2) + O(\epsilon)$ near $\epsilon = 0$.

[8] As described in Appendix A.2.4, 'minimal subtraction' simply drops the $(4-D)^{-1}$ term in divergent quantities, while 'modified minimal subtraction' drops both the $(4-D)^{-1}$ and the constants γ and $\ln(4\pi)$ [35, 36, 37].

In the spirit of the method of regions consider regarding $\mathcal{I}_\epsilon$ again as the sum of an IR contribution, for which $k^2 \sim m^2 \ll M^2$, plus a UV contribution with $m \ll k \sim M$:

$$\mathcal{I}^{IR}_\epsilon := \int_0^\infty \left(\frac{\mu}{k}\right)^\epsilon \frac{k\,\mathrm{d}k}{(k^2+m^2)M^2}\left[1 - \frac{k^2}{M^2} + \cdots\right] \tag{3.41}$$
$$\simeq \frac{(m/\mu)^{-\epsilon}}{M^2}\left[\frac{1}{\epsilon} + O(\epsilon)\right] = \frac{1}{M^2}\left[\frac{1}{\epsilon} - \ln\left(\frac{m}{\mu}\right) + O(\epsilon)\right],$$

and

$$\mathcal{I}^{UV}_\epsilon := \int_0^\infty \left(\frac{\mu}{k}\right)^\epsilon \frac{k\,\mathrm{d}k}{k^2(k^2+M^2)}\left[1 - \frac{m^2}{k^2} + \cdots\right] \tag{3.42}$$
$$\simeq \frac{(M/\mu)^{-\epsilon}}{M^2}\left[-\frac{1}{\epsilon} + O(\epsilon)\right] = \frac{1}{M^2}\left[-\frac{1}{\epsilon} + \ln\left(\frac{M}{\mu}\right) + O(\epsilon)\right].$$

The approximate equality starting the second line for both of these formulae drops all subdominant terms in powers of m/M. Notice that the pole $1/\epsilon$ arises due to a UV divergence in $\mathcal{I}^{IR}$ as $\epsilon \to 0$ while it corresponds to the IR divergence in $\mathcal{I}^{UV}$ as $\epsilon \to 0$. Notice also that these poles cancel in the sum $\mathcal{I} = \mathcal{I}^{UV} + \mathcal{I}^{IR}$, after which the limit $\epsilon \to 0$ can be taken, revealing agreement with (3.40).

The surprise in this exercise was that it was not important to explicitly separate out the region with $k < \Lambda$ and $k > \Lambda$ when defining $\mathcal{I}^{UV}$ and $\mathcal{I}^{IR}$, which nevertheless reproduce the correct dependence on m/M once summed to give the full integral. This works because any cutoff-dependence in the definition of these integrals is guaranteed to cancel in any case, and so although including the cutoff-dependence could be done, it is wasted effort.

Several concrete examples of the use of dimensional regularization when matching between underlying and effective theories are examined in more detail (for relativistic theories) in §7, which also explores a modification to minimal subtraction called 'decoupling subtraction', that proves useful when matching is done at or above one-loop accuracy. Non-relativistic examples of beyond-leading-order matching are similarly studied in §12 and §15.

3.2.2 Matching vs integrating out

Matching — the fixing of low-energy couplings by comparing the predictions of the full theory with the predictions of its low-energy Wilsonian approximation — is often much easier to carry out than is the process of explicitly integrating out a heavy state using a cut-off path integral. This is partly because the comparison can be made for *any* physical quantity, and in particular this quantity can be chosen to make the comparison as simple as possible. Furthermore, because the comparison is made at the level of renormalized interactions, for both the full and low-energy theories, there are no UV divergences to worry about.

Example: the toy model

As usual, the toy model helps make the above statements more concrete. For the toy model the heavy mass scale is m_R and the full theory describing the physics of the two fields χ and ξ above this scale is given by the action $S[\chi,\xi]$ of eq. (1.24) and (1.25) (repeated for convenience here):

$$S = -\int \mathrm{d}^4x \left[\frac{1}{2}\,\partial_\mu\chi\partial^\mu\chi + \frac{1}{2}\left(1+\frac{\chi}{\sqrt{2}\,v}\right)^2 \partial_\mu\xi\partial^\mu\xi + V(\chi)\right], \tag{3.43}$$

with

$$V(\chi) = \frac{\lambda\, v^2}{2}\,\chi^2 + \frac{\lambda\, v}{2\sqrt{2}}\,\chi^3 + \frac{\lambda}{16}\,\chi^4\,, \tag{3.44}$$

One could equivalently use the fields $\hat\phi_R$ and $\hat\phi_I$ with the action $S[\hat\phi_R,\hat\phi_I]$ of eqs. (1.1) and (1.2), and renormalization is actually easier using these variables. This chapter sticks to χ and ξ to keep the symmetries of the problem more manifest. UV divergences in this theory are handled using dimensional regularization, and where necessary divergences are renormalized using modified minimal subtraction (see Appendix A.2.4 for details).

The low-energy Wilson action in this case is $S_W[\xi]$ (or $S_W[\hat\phi_I]$), depending only on the single light field, with UV divergent integrals again defined using dimensional regularization. Renormalization is again based on minimal subtraction, though with the difference that the finite part of the coupling is fixed by matching to predictions of the full theory (rather than again using modified minimal subtraction).

For present purposes the first step when matching is to write down all possible interactions in S_W up to some order in $1/m_R$, since this identifies the effective couplings whose value matching is meant to determine. For the toy model we know from §2.5 that all $1/m_R^2$ interactions are redundant, and the most general interactions (consistent with Lorentz invariance and the symmetry under constant shifts in ξ) are given to order $1/m_R^4$ by:

$$S_W = -\int \mathrm{d}^4x \left[\frac{z_W}{2}\,\partial_\mu\xi\,\partial^\mu\xi - a_W\,(\partial_\mu\xi\,\partial^\mu\xi)^2 + \cdots\right], \tag{3.45}$$

where on dimensional grounds z_W is dimensionless while $a_W \propto 1/m_R^4$ and terms not explicitly written are suppressed by at least $1/m_R^6$ (see §2.5).

Whereas earlier sections use the freedom to rescale ξ to set $z_W = 1$ (*i.e.* to canonically normalize ξ), writing (3.45) recognizes this has only been done at the classical level and not exactly, so at one loop $z_W = 1 + z_W^{(1)}$ with $z_W^{(1)} \simeq O(\lambda/16\pi^2)$ due to graphs like those of Fig. 3.3.

The contributions to $z_W^{(1)}$ found by Wick rotating these graphs, evaluating them in dimensional regularization and matching them to the contribution of (3.45) are

$$\mathrm{i}z_{W\,3.3(\mathrm{a})}^{(1)} = \mathrm{i}\left(-\frac{1}{4v^2}\right)\int \frac{\mathrm{i}\,\mathrm{d}^4p_E}{(2\pi)^4}\,\frac{-\mathrm{i}}{p^2+m_R^2} = -\frac{\mathrm{i}}{4v^2}\;I_D^{(0,1)}(m_R) \tag{3.46}$$

and

$$\mathrm{i}z_{W\,3.3(\mathrm{b})}^{(1)} \simeq 3\left[2\times\frac{\mathrm{i}^2}{2!}\right]\left(-\frac{1}{\sqrt{2}\,v}\right)\left(-\frac{\lambda v}{2\sqrt{2}}\right)\left(-\frac{\mathrm{i}}{m_R^2}\right)\int \frac{\mathrm{i}\,\mathrm{d}^4p_E}{(2\pi)^4}\,\frac{-\mathrm{i}}{p^2+m_R^2} = \frac{3\mathrm{i}\lambda}{4m_R^2}\;I_D^{(0,1)}(m_R) \tag{3.47}$$

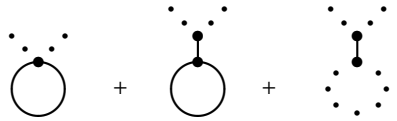


Fig. 3.3 One-loop graphs that contribute to the $\partial_\mu\xi\partial^\mu\xi$ kinetic term in the Wilson and 1LPI actions using the interactions of eqs. (3.43) and (3.44). Solid (dotted) lines represent χ (and ξ) fields.

while $z^{(1)}_{W\,3.3(c)} = 0$ because its dimensionally regularized loop evaluates to $I^{(1,1)}_D(0)$ and so vanishes. Here the integrals $I^{(0,1)}_D(m)$ and $I^{(1,1)}_D(m)$ are defined in (3.37), and the nonzero result evaluates to

$$\begin{aligned} I^{(0,1)}_D(m) &= \frac{m^2}{(4\pi)^{D/2}}\Gamma\left(1-\frac{D}{2}\right)\left(\frac{m^2}{\mu^2}\right)^{(D-4)/2} \\ &= \frac{m^2}{16\pi^2}\left[\frac{1}{(D/2)-2} + \gamma - 1 + \ln\left(\frac{m^2}{4\pi\mu^2}\right) + O(D-4)\right]. \end{aligned} \tag{3.48}$$

Summing these contributions (using $m_R^2 = \lambda v^2$) gives the one-loop prediction

$$z^{(1)}_W = \frac{I^{(0,1)}_D(m_R)}{2v^2} = \frac{\lambda}{16\pi^2}\left[\frac{1}{D-4} + \frac{1}{2}(\gamma-1) + \frac{1}{2}\,\ln\left(\frac{m_R^2}{4\pi\mu^2}\right) + O(D-4)\right]. \tag{3.49}$$

These corrections to z_W can again be absorbed into a rescaling of ξ — *i.e.* ξ is 're-normalized' by defining $\xi \to z_W^{-1/2}\xi$ — leading to the following rescaled version of (3.45):

$$S_W = -\int \mathrm{d}^4x\,\left[\frac{1}{2}\,\partial_\mu\xi\,\partial^\mu\xi - \frac{a_W}{z_W^2}\,(\partial_\mu\xi\,\partial^\mu\xi)^2 + \cdots\right]. \tag{3.50}$$

The remainder of the matching calculation computes another observable using (3.50) and (3.49) and compares with the result to the calculation of the same quantity in the full theory to read off the coefficient a_W. It is relatively simple to do this with the same quantity as used at lowest order in §1.2.1: the amplitude for $\xi(p) + \xi(q) \to \xi(p') + \xi(q')$ scattering, keeping only terms up to order $1/m_R^4$. Since this calculation is already performed in chapter 1 at leading order in λ, it suffices here to sketch how things change once subdominant contributions are included.

To this end write the coefficients in S_W as a series in λ,

$$a_W = a_W^{(0)} + a_W^{(1)} + \cdots \qquad \text{and} \qquad z_W = 1 + z_W^{(1)} + \cdots, \tag{3.51}$$

for which $z_W^{(1)}$ is given in (3.49). Starting on the Wilson side of the calculation, for $\xi - \xi$ scattering the required graphs up to one loop order are given by (a), (b) and (c) of Fig. 2.5. Of these, graphs (b) and (c) and their crossed counterparts both evaluate to give a contribution to $\xi\xi \to \xi\xi$ scattering that is suppressed by more than just four powers of $1/m_R$.

This is easy to see in dimensional regularization because the coefficients of the interactions are suppressed by more than $1/m_R^4$ and the loop integrals only involve massless states and so cannot introduce compensating factors of m_R into the numerator. If one wishes to work only to lowest order in $1/m_R$ but to higher order in λ, it suffices to work with the tree contribution, graph (a), within the Wilsonian theory, but with λ-corrected effective coefficients $a_W^{(1)}$ and $z_W^{(1)}$.

Evaluating graph (a) using the Wilsonian coupling a_W/z_W^2 expanded out to subdominant order in λ, $a_W^{(1)} - 2a_W^{(0)}z_W^{(1)}$, then gives:

$$\mathcal{A}^{W(a)}_{\xi\xi\to\xi\xi} = 8i\Big(a_W^{(0)} + a_W^{(1)} - 2a_W^{(0)}z_W^{(1)}\Big)\Big[(p\cdot q)(p'\cdot q') + (q\cdot q')(p\cdot p') + (p\cdot q')(p'\cdot q)\Big] + \cdots . \quad (3.52)$$

Field-theory aficionados will recognize the $z_W^{(1)}$ term as the wave-function renormalization counter-term, that cancels UV divergences due to the loops of Fig. (3.3) inserted into the external lines of the tree-level scattering graphs. (These graphs are not drawn explicitly in Figure (2.4)).)

This is to be compared with the one-loop contributions computed within the UV theory, working out to subdominant order in λ. The leading contribution comes from the tree graphs of Fig. 1.3, that evaluate to the result given in eq. (1.28):

$$\mathcal{A}^{\text{full,tree}}_{\xi\xi\to\xi\xi} = \frac{2i\lambda}{m_R^4}\Big[(p\cdot q)(p'\cdot q') + (q\cdot q')(p\cdot p') + (p\cdot q')(p'\cdot q)\Big] + \cdots , \quad (3.53)$$

where the ellipses contain terms of higher order than $1/m_R^4$. Equating this to the lowest-order part of (3.52) then gives the previously obtained tree-level result: $a_W^{(0)} = \lambda/(4m_R^4) = 1/(4\lambda v^4)$.

Repeating this procedure including one-loop $O(\lambda/16\pi^2)$ corrections in the UV theory is less trivial, but in principle proceeds in precisely the same manner. This involves evaluating the graphs of Fig. 2.4, plus the 'wave-function renormalization' graphs obtained by inserting Fig. 3.3 into the external lines of the tree-level scattering graphs of Fig. 1.3. The $z_W^{(1)}$ contributions of (3.52) are important for reproducing the effects of these latter graphs in the full theory. The final result is a prediction for $a_W^{(1)}$ that is of order $\lambda^2/(16\pi^2 m_R^4) = 1/(4\pi v^2)^2$ in size.

This example shows how loops in the Wilsonian theory are not counted in the same way as are loops in the UV theory. Loops in the Wilsonian theory necessarily involve higher powers of E/m_R (more about this below), while loops in the UV theory are suppressed by factors of $\lambda/(16\pi^2)$ only, with all powers of E/m_R appearing at each loop order.

3.2.3 Power counting using dimensional regularization

The previous section made assertions about the size of the contributions of loop graphs — like graphs (b) and (c) of Fig. 2.5 in the toy model — which this section explores more systematically. More generally, this section's goal is to track how a generic Feynman graph computed using the Wilsonian action depends on a heavy scale like $1/m_R$, given that this scale does not appear in the same way for all interactions within $\mathfrak{L}_W$.

The logic here is much as used in §3.1.1 where dimensional analysis was employed to track how the cutoff Λ appears in amplitudes. The only difference now is to regulate the UV

divergences in these graphs with dimensional regularization, since the size of a dimensionally regulated integral is set by the physical scales (light masses or external momenta) that appear in the integrand (rather than Λ). The power-counting rules obtained in this way are much more useful since they directly track how amplitudes depend on physical variables, rather than unphysical quantities like Λ that in any case cancel from physical quantities.

The basic observation is that dimensional analysis applied to a dimensionally-regulated integral estimates its size as

$$\int \cdots \int \left(\frac{d^D p}{(2\pi)^D}\right)^A \frac{p^B}{(p^2+q^2)^C} \sim \left(\frac{1}{4\pi}\right)^{DA} q^{DA+B-2C}, \tag{3.54}$$

with a dimensionless prefactor that depends on the dimension, D, of spacetime, and which may well be singular in the limit that $D \to 4$. Here q represents the dominant scale appearing in the integrand of the momentum integrations. If the light particles appearing as external states in $A_E(q)$ should be massless, or highly relativistic, then the typical external momenta are much larger than their masses and q in the above expression represents these momenta.[9] If all masses and momenta are comparable then q is their common value. The important assumption is that there is only one low-energy scale (the more complicated case of multiple hierarchies is examined in later chapters, in particular in the non-relativistic applications of Part III for which small speed, $v \sim E_{\rm kin}/p$, can be regarded as a ratio of two separate low-energy scales).

With this in mind, the idea is to repeat the steps of §3.1.1 and use the effective action, $S_w = S_{w,0} + S_{w,\rm int}$, of (3.2) — repeated here for ease of reference:

$$\begin{aligned} S_{w,0} &= -\frac{\tilde{\mathfrak{f}}^4}{M^2 v^2} \int \mathrm{d}^4x \left[\partial_\mu\phi\,\partial^\mu\phi + m^2\phi^2\right] \\ S_{w,\rm int} &= -\tilde{\mathfrak{f}}^4 \sum_n \frac{\hat{c}_n}{M^{d_n} v^{f_n}} \int \mathrm{d}^4x\, \mathcal{O}_n(\phi)\,, \end{aligned} \tag{3.55}$$

to compute amputated Feynman amplitudes, $\mathcal{A}_\mathcal{E}(q)$, having $\mathcal{E}$ external lines, $\mathcal{I}$ internal lines, $\mathcal{L}$ loops and $\mathcal{V}_n$ vertices coming from the effective interaction with label 'n.' Respectively denoting (as before) the number of derivatives and fields appearing in this interaction as d_n and f_n, the amplitude becomes proportional to the following multiple integral:

$$\int \cdots \int \left(\frac{d^D p}{(2\pi)^D}\right)^{\mathcal{L}} \frac{p^{\mathcal{R}}}{(p^2+q^2)^{\mathcal{I}}} \sim \left(\frac{1}{4\pi}\right)^{2\mathcal{L}} q^{4\mathcal{L}-2\mathcal{I}+\mathcal{R}}, \tag{3.56}$$

where $\mathcal{R} = \sum_n d_n \mathcal{V}_n$ and the final estimate takes $D \to 4$. Liberally using the identities (3.4) and (3.5) then gives the following order of magnitude for $\mathcal{A}_\mathcal{E}(q)$:

$$\mathcal{A}_\mathcal{E}(q) \sim \tilde{\mathfrak{f}}^4 \left(\frac{1}{v}\right)^{\mathcal{E}} \left(\frac{Mq}{4\pi\tilde{\mathfrak{f}}^2}\right)^{2\mathcal{L}} \left(\frac{q}{M}\right)^{2+\sum_n (d_n-2)\mathcal{V}_n}. \tag{3.57}$$

This last formula is the main result, used extensively in many applications considered later. Its utility lies in the fact that it links the contributions of the various effective interactions in the effective lagrangian, (3.55), with the dependence of observables on small

[9] Any logarithmic dependence on q and infrared mass singularities that might arise in this limit are ignored here (for now), since the main interest is in following *powers* of ratios of the light and heavy mass scales.

ratios of physical scales such as q/M. Notice in particular that more and more complicated graphs – for which $\mathcal{L}$ and $\mathcal{V}_n$ become larger and larger – are generically suppressed in their contributions to the graphical expansion if q is much smaller than the other scales M and $\mathfrak{f}^2/M$. This suppression assumes only that the powers appearing in (3.57) are all non-negative, and this is true so long as $d_n \geq 2$. The special cases where $d_n = 0, 1$ are potentially dangerous in this context, and require examination on a case-by-case basis.

Example: the toy model

The toy model Wilsonian action has the form of (3.55) provided we take $M = m_R$ and $\mathfrak{f}^2 = m_R v$, in which case (3.57) becomes

$$\mathcal{A}_{\mathcal{E}}(q) \sim q^2 v^2 \left(\frac{1}{v}\right)^{\mathcal{E}} \left(\frac{q}{4\pi v}\right)^{2\mathcal{L}} \left(\frac{q}{m_R}\right)^{\sum_n (d_n-2)\mathcal{V}_n} , \tag{3.58}$$

underlining that low-energy amplitudes come as a double expansion, in powers of both q/m_R and $q/4\pi v$. Furthermore, the symmetry $\xi \to \xi +$ constant implies all interactions in $S_{W,\,\text{int}}$ have $d_n \geq 4$ and so interactions involving larger powers of $\mathcal{L}$ and $\mathcal{V}_n$ are always suppressed by one or both of these small parameters.

Recalling that the validity of the UV theory's loop expansion — $\lambda/(16\pi^2) \ll 1$ — implies $m_R = \sqrt{\lambda}\, v \ll 4\pi v$, it follows that the expansion in q/m_R converges more slowly in this regime than does the expansion in $q/4\pi v$. Both expansion parameters in the low-energy theory become similar in size just at the border of the domain of perturbative validity in the UV theory. In no way does the effective theory require small λ in order to be predictive, and the diagnostic for weak coupling in the underlying UV theory is the existence of two separate scales, m_R and $4\pi v$, against which q becomes compared.

Applying the estimate of (3.58) to the graphs of Fig. (2.5) requires specializing to $\mathcal{E} = 4$, with graph (a) having $\mathcal{L} = 0$ and $\sum_n (d_n - 2)\mathcal{V}_n = 2$ and so

$$\mathcal{A}_4^{(a)}(q) \sim \frac{q^2}{v^2} \left(\frac{q}{m_R}\right)^2 , \tag{3.59}$$

in agreement with the explicit tree-level formulae found earlier.

The one-loop contributions coming from graphs (b) and (c) similarly have $\mathcal{L} = 1$ and $\sum_n (d_n - 2)\mathcal{V}_n = 4$ and so both satisfy

$$\mathcal{A}_4^{(b)}(q) \sim \mathcal{A}_4^{(c)}(q) \sim \frac{q^2}{v^2} \left(\frac{q}{4\pi v}\right)^2 \left(\frac{q}{m_R}\right)^4 . \tag{3.60}$$

Both are similar in size and are suppressed relative to the tree level result by the small factor $q^4/(4\pi v\, m_R)^2$.

Notice that the above power-counting estimates apply equally well for *any* theory of an abelian Goldstone boson subject to the shift symmetry $\xi \to \xi +$ constant, provided that scales like v and M are regarded as both being large compared with q and that dimensionless parameters $\hat{a}$ and $\hat{b}$ in effective couplings like $a = \hat{a}\, \mathfrak{f}^4/(Mv)^4$ and $b = \hat{b}\, \mathfrak{f}^4/(Mv)^6$ in eq. (2.97) are regarded as being independent and not systematically large. The low-energy limit for the specific toy model of §1.1 is more predictive than is the generic low-energy

Goldstone-boson theory, because it predicts all of these parameters in terms of two fundamental ones: λ and v, say. The predictiveness enters both because of the relationship implied amongst the generic scales — *e.g.* $\mathfrak{f}^2 = Mv$ and $M^2/v^2 = \lambda$ — and the inferences it allows for the values of coefficients like $\hat{a}$ and $\hat{b}$, given as a series in powers of λ. What is usually informative in any application of EFT methods is therefore the comparison between the generic expectations for the assumed low-energy field content and the more detailed predictions allowed by specific UV theories that can give rise to this field content.

3.2.4 Power-counting with fermions

It is straightforward to extend these results to include light fermions in the effective theory. The complication here is not really the statistics of the fields, it is the different momentum dependence of the propagators and the related canonical dimension of the fields.

To sort this out, first generalize the starting form assumed for the lagrangian to include fermion fields, ψ, in addition to boson fields, ϕ:

$$\mathfrak{L}_{\text{eff}} = \mathfrak{f}^4 \sum_n \frac{c_n}{M^{d_n}} \, O_n\left(\frac{\phi}{v_B}, \frac{\psi}{v_F^{3/2}}\right). \tag{3.61}$$

Fermions and bosons come with different powers of the scales v_B and v_F because their kinetic terms (that dominate the unperturbed action in a weak-coupling perturbative analysis) involve different numbers of derivatives: only one for fermions but two for bosons. This also implies fermion and boson propagators fall off differently for large momenta, with bosonic propagators varying like $1/p^2$ for large p and fermion propagators only falling like $1/p$. The resulting differences in contributions to the power counting of Feynman graphs makes it important to keep separate track of the number of fermion and boson lines.

To this end it is useful to choose to label vertices using *three* indices: d_n, b_n and f_n. As before, d_n labels the numbers of derivatives in the interaction, but now b_n and f_n separately count the number of bose and fermi lines terminating at the vertex of interest. The number of vertices in a graph carrying a given value for d_n, b_n and f_n is, as before, labeled by $\mathcal{V}_n$.

Consider now computing an amplitude with $\mathcal{E}_B$ external bosonic lines, $\mathcal{E}_F$ external fermion lines and $\mathcal{I}_B$ and $\mathcal{I}_F$ internal bose and fermi lines. The constraints of graph-making relate these in three ways. First, the definition of the number of loops generalizes (3.5) to

$$\mathcal{L} = 1 + \mathcal{I}_B + \mathcal{I}_F - \sum_n \mathcal{V}_n \,. \tag{3.62}$$

Similarly 'conservation of ends' now holds separately for both bosonic and for fermionic lines and so implies (3.4) is replaced by the two separate conditions

$$2\mathcal{I}_B + \mathcal{E}_B = \sum_n b_n \mathcal{V}_n \qquad \text{and} \qquad 2\mathcal{I}_F + \mathcal{E}_F = \sum_n f_n \mathcal{V}_n \,. \tag{3.63}$$

Repeating, with the lagrangian of eq. (3.61), the power counting argument which led (using dimensional regularization) to eq. (3.57) now gives instead the following result:

$$\mathcal{A}_{\mathcal{E}_B \mathcal{E}_F}(q) \sim \mathfrak{f}^4 \left(\frac{1}{v_B}\right)^{\mathcal{E}_B} \left(\frac{1}{v_F}\right)^{3\mathcal{E}_F/2} \left(\frac{Mq}{4\pi\mathfrak{f}^2}\right)^{2\mathcal{L}} \left(\frac{q}{M}\right)^{\mathcal{P}}, \tag{3.64}$$

where the power $\mathcal{P}$ can be written

$$\mathcal{P} = 2 + \mathcal{I}_F + \sum_n (d_n - 2)\,\mathcal{V}_n = 2 - \frac{1}{2}\,\mathcal{E}_F + \sum_n \left(d_n + \frac{1}{2}\,f_n - 2\right)\mathcal{V}_n. \tag{3.65}$$

Since $\mathcal{I}_F \geq 0$ the first equality shows (3.64) is suppressed relative to the corresponding term in the purely bosonic result (3.57), as makes sense since each fermionic propagator is order $1/q$, and so is suppressed by q relative to the bosonic propagator $1/q^2$. The second equality trades the dependence on $\mathcal{I}_F$ for $\mathcal{E}_F$ using (3.63).

The factor $(q/M)^{-\mathcal{E}_F/2}$ in (3.64) might also seem problematic at low energies, indicating as it does that more external lines necessarily imply more factors of the large ratio M/q. However, because such factors are fixed for all graphs contributing to any explicit process with a given number of external legs they usually do not in themselves undermine the validity of a perturbative expansion.

Furthermore, for the specific case where $\mathcal{A}_{\mathcal{E}_B\mathcal{E}_F}(q)$ represents a scattering amplitude each external fermion line corresponds to an initial-state or final-state spinor — $u_{q\sigma}$ or $\overline{u}_{q\sigma}$ — or the corresponding antiparticle spinor — $v_{q\sigma}$ or $\overline{v}_{q\sigma}$ — labelled by the corresponding state's momentum and spin. But each of these is itself proportional to an external particle — and so low-energy, $O(q)$ — scale, as can be seen from their appearance in spin-averaged expressions like $\sum_\sigma u_{q\sigma}\,\overline{u}_{q\sigma} = m - \mathrm{i}\not{q}$ and $\sum_\sigma v_{q\sigma}\,\overline{v}_{q\sigma} = -m - \mathrm{i}\not{q}$. This $q^{1/2}$ scaling of each external fermion line systematically cancels the factor $q^{-\mathcal{E}_F/2}$ in the amplitude $\mathcal{A}_{\mathcal{E}_B\mathcal{E}_F}(q)$, leading to non-singular predictions for scattering process at low energies. The same is true for effective couplings in S_W if these are obtained by matching to scattering processes.

Dangerous interactions

As usual, interactions with the same number of fields and derivatives as the kinetic terms — either $f_n = 0$ and $d_n = 2$ (for bosons) or $f_n = 2$ and $d_n = 1$ (for fermions) — are unsuppressed by powers of q/M, beyond the usual loop factor. Interactions with more fields or derivatives than the kinetic terms additionally suppress a graph each time they are used. But interactions with no derivatives and two or fewer fermions can be potentially dangerous at low energies, introducing as they do negative powers of the small ratio q/M.

The kinds of interactions that are dangerous in this way are terms in a scalar potential ($d_n = f_n = 0$) and Yukawa couplings ($d_n = 0$ and $f_n = 2$). In principle these kinds of interactions can be genuine threats to the consistency of the low-energy expansion, and whether such interactions are consistent with low-energy physics depends on the details.

What can make these interactions benign at low energies is if they do not carry too much energy for generic field configurations, $\phi \sim v_B$ and $\psi \sim v_F^{3/2}$. For instance suppose, following the discussion around eq. (3.17), suppose the scalar potential only carries energy density $\mathfrak{f}_V^4 \ll \mathfrak{f}^4$ when fields are order $\phi \sim v_B$ in size, such as if

$$V(\phi) \sim \mathfrak{f}_V^4 \sum_r g_r \left(\frac{\phi}{v_B}\right)^r . \tag{3.66}$$

In particular the $r = 2$ term represents a mass for the field ϕ of order $m_B^2 \sim \mathfrak{f}_V^4/v_B^2$, so a

natural criterion for ϕ to survive into the low energy theory might be that $\mathfrak{f}_V^4 \sim m_B^2 v_B^2$ with $m_B \lesssim q$ for q a typical (possibly relativistic) momentum in the low-energy sector.

If this is the case then — assuming the couplings g_r are order unity — all the dimensionless couplings c_n of eq. (3.61) for these particular interactions are secretly suppressed, with $c_n(d_n = 0) \sim g_r\,(\mathfrak{f}_V^4/\mathfrak{f}^4) \sim g_r\,(m_B^2 v_B^2/\mathfrak{f}^4)$. The contributions of these particular $d_n = f_n = 0$ interactions to (3.64) then become

$$\prod_{d_n=f_n=0}\left[c_n\left(\frac{q}{M}\right)^{-2}\right]^{\mathcal{V}_n} \sim \prod_r\left[g_r\left(\frac{\mathfrak{f}_V^4 M^2}{\mathfrak{f}^4 q^2}\right)\right]^{\mathcal{V}_r} \sim \prod_r\left[g_r\left(\frac{m_B^2}{q^2}\right)\left(\frac{v_B^2 M^2}{\mathfrak{f}^4}\right)\right]^{\mathcal{V}_r}, \tag{3.67}$$

which is no longer enhanced because $m_B \lesssim q$.

A similar story goes through for Yukawa interactions for which an interaction like

$$\mathfrak{L}_{\rm yuk} \sim \mathfrak{f}_Y^4 \sum_r h_r\left(\frac{\phi}{v_B}\right)^r\left(\frac{\overline{\psi}\psi}{v_F^3}\right), \tag{3.68}$$

would contribute a fermion mass of order $m_F \sim \mathfrak{f}_Y^4/v_F^3$ for fields $\phi \sim v_B$ and couplings $h_r \sim O(1)$. This can remain a genuine low-energy mass satisfying $m_F \lesssim q$ if $\mathfrak{f}_Y^4 \sim m_F v_F^3$ is systematically small relative to $\mathfrak{f}^4$. In this case the contribution of these $d_n = 0$ but $f_n = 2$ terms to (3.64) (assuming the h_r are order unity) then become

$$\prod_{d_n=0,\,f_n=2}\left[c_n\left(\frac{q}{M}\right)^{-1}\right]^{\mathcal{V}_n} \sim \prod_r\left[h_r\left(\frac{\mathfrak{f}_Y^4 M}{\mathfrak{f}^4 q}\right)\right]^{\mathcal{V}_r} \sim \prod_r\left[h_r\left(\frac{m_F}{q}\right)\left(\frac{v_F^3 M}{\mathfrak{f}^4}\right)\right]^{\mathcal{V}_r}, \tag{3.69}$$

which again would not be enhanced.

3.3 The Big Picture ◊

So far this chapter has examined a variety of types of effective actions: 1PI, 1LPI and Wilsonian, and explored how dimensional analysis constraints how their effective couplings contribute to low-energy observables (both using a cutoff and in dimensional regularization). It is easy at this point to lose sight of the forest for the trees, so this section aims to reiterate some broader features of the overall EFT program.

3.3.1 Low-energy theorems

Given the huge number of effective couplings available in the Wilson action, it is tempting to conclude that there is nothing robust that can be said about the low-energy world. But this is misleading, because there are some predictions that can be made very robustly without running into renormalization ambiguities. These predictions are generally known as an EFT's 'low-energy theorems', and consist of those predictions that depend only on the low-energy theory's leading non-renormalizable interaction and the number and type of low-energy states present. In particular, they are insensitive to the values of the long list of higher-dimension effective couplings in the Wilsonian EFT.

The main observation underlying the construction of these theorems is that locality implies the higher-dimensional effective interactions all depend only on polynomials of momenta. They do so because they are built using only powers of fields and their derivatives. As a result they do not in themselves give rise to non-analytic contributions to scattering amplitudes at vanishing external momenta.

This observation may not seem so remarkable given that the power-counting arguments of the previous sections also only involve powers of the external energy scale q. But (as remarked in footnote [9]) these power-counting arguments focus exclusively on powers of q and explicitly ignore quantities like $\ln(q^2/\mu^2)$ that could make the full result non-analytic at $q = 0$. Indeed such non-analytic contributions do arise in scattering amplitudes, with their presence being guaranteed by general principles [38, 39].

In fact, direct evaluation of the scattering amplitude for $\xi\xi \to \xi\xi$ scattering at one-loop order in the Toy model (in particular graph (b) of Fig. 2.5) gives precisely this kind of contribution, behaving as $q^8 \ln(q^2/m_R^2)$. The coefficient of this dependence is proportional to $a_{\rm eff}^2/(4\pi)^2$ (consistent with (3.60)), with $a_{\rm eff} = \lambda/(4m_R^4) = 1/(4v^2m_R^2)$ the lowest-order effective coupling found from tree-level scattering in eqs. (1.15) and (1.16).

Here is the point: although loop graphs like this arise potentially at the same order as tree graphs built from higher-derivative terms (whose coefficients renormalize the UV divergences in the loops) the higher-derivative interactions only contribute to the analytic part of the amplitude at $q = 0$ and so *cannot* contribute to the coefficient of the non-analytic $q^8 \ln q^2$ term. The leading coefficients of such non-analytic terms are therefore absolute predictions, given only the lowest-order non-renormalizable effective coupling (in the Toy Model case the coupling $a_{\rm eff}$). These predictions, collectively with the other predictions depending only on $a_{\rm eff}$, are called the EFT's low-energy theorems. (The applications of EFT methods in later chapters give practical examples of these theorems for the strong interactions §8 and for General Relativity §10.)

High-energy physics cannot contribute to non-analytic behaviour for q in a low-energy regime because the general arguments of [38, 39] imply this non-analytic behaviour arises only at the threshold for the particle whose circulation within loops generates the non-analytic behaviour. Non-analytic behaviour at $q^2 = 0$ comes from massless particles and so does not arise when heavy particles are integrated out. (This is also why such nonanalytic dependence on derivatives does not arise in the Wilson action itself, for which the light states responsible for singularities at $q^2 = 0$ are not yet integrated out.)

3.3.2 The effective-action logic ◇

Historically, what made theories with non-renormalizable interactions daunting was the seeming necessity to include an infinite number of interactions. This is partly because there are only a finite number of renormalizable (and super-renormalizable) interactions, but an infinite number of non-renormalizable ones. But it is also true that one usually cannot cherry-pick amongst non-renormalizable interactions because they are all needed to absorb the UV divergences appearing at higher loop orders.[10]

[10] It is this need for a nominally infinite number of couplings to renormalize UV divergences that underlies the name 'non-renormalizable'.

The utility of power-counting formulae like (3.57) lies in their ability to cut through the conundrum of how to deal with so many interactions, and so to organize how to calculate predictively (including quantum effects). The key observation is that non-renormalizable theories always implicitly involve a low-energy expansion, so predictivity must be assessed order-by-order in this expansion. The logic for doing so unfolds with the following steps:

[1] Choose the accuracy desired in the answer. (For instance an accuracy of 1% might be required for a particular scattering amplitude.)

[2] Determine the order in the small ratio of scales q/M (*e.g.* q/m_R in the toy model) required to achieve the desired accuracy. (For instance if $q/M = 0.1$ then order $(q/M)^2$ would be required to achieve 1% accuracy.)

[3] Use the power-counting results to identify which terms in $\mathfrak{L}_W$ can contribute to the observable of interest to the desired order in q/M. At any fixed order only a finite number (say, N) of terms in $\mathfrak{L}_W$ can contribute.

[4a] If the underlying theory is known and is calculable, then compute the coefficients of the N required effective interactions to the needed accuracy.

[4b] If the underlying theory is unknown, or is too complicated to permit explicit *ab initio* calculation of $\mathfrak{L}_W$, then treat the N required coefficients as free parameters. This is nonetheless predictive if more than N observables can be identified whose predictions depend only on these parameters.

It is in the spirit of step [4a] that EFT methods are developed for the toy model in previous sections: the full theory is explicitly known and parameters are within a calculable regime. (For the toy model this corresponds to using the full theory within the perturbative small-λ regime.) In this case EFT methods simply provide an efficient means to identify and calculate the combination of parameters on which low-energy observables depend.

It is option [4b], however, that is responsible for the great versatility of EFT methods, because it completely divorces the utility of the low-energy theory from the issue of whether or not the underlying theory is understood. It allows EFT methods to be used for *any* low-energy situation, regardless of whether the underlying theory is completely unknown or is known but too complicated to allow reliable predictions. (Examples of both of these cases are considered in subsequent sections.)

From this point of view the conundrum of dealing with an infinite number of interactions is really only a problem in the limit where infinite precision is required of the answer, since it is only then one must work to all orders in the low-energy expansion.

Within this point of view traditional renormalizable theories are simply the special case where the above logic is invoked, but with accuracy that only requires working at zeroeth order in q/M. This is equivalent to renormalizability because in this case *all* interactions suppressed by $1/M$ can be dropped, which amounts to dropping all interactions whose couplings have dimensions of negative power of mass (*i.e.* all non-renormalizable interactions).

Renormalizable theories are revealed in this way to be the ones that should always dominate in the limit that the light scales q and the heavy scales M are so widely separated as to allow the complete neglect of heavy-particle effects. This is the beginnings of an explanation of *why* renormalizable interactions turn out to play such ubiquitous roles through physics: the message behind their success is that any 'new' physics not included in them involves scales too high to be relevant in practice.

3.4 Summary

The main topic of this chapter is power counting: the tool that makes the Wilson action a precise instrument. The purpose of this tool is to identify the effective interactions and Feynman graphs that are required to make physical predictions to any given order in the low-energy expansion, E/M.

At face value this counting seems like it should be easy: the power of $1/M$ for any graph is found simply by collecting all such factors from the coefficients of a graph's effective interactions. What complicates this argument is the extreme sensitivity of some loop integrals to short wavelengths, which leads to the appearance of large scales like M in numerators rather than denominators. Successful tracking of high-energy scales in loops therefore requires handling their ultraviolet divergences.

This chapter provides two kinds of power-counting estimates. One of these regulates UV divergences with explicit cutoffs, and because short wavelengths often dominate in loops what is mostly learned is how a graph depends on this cutoff regulator. This can be useful, particularly when asking how effective couplings flow as successive scales are integrated out. This flow is described by exact renormalization-group methods, culminating with Wilson-Polchinski type evolution equations.

The second kind of power-counting estimate this chapter provides regulates divergences using dimensional regularization. This is usually more convenient for practical applications, partly because dimensional regularization allows useful symmetries to be kept explicit. For EFTs dimensional regularization also proves useful because the absence of an explicit cutoff scale simplifies dimensional power-counting arguments revealing how graphs depend on low- and high-energy scales. These arguments culminate in the very useful expressions (3.64) and (3.65), appropriate for low-energy theories dominated by a single low-energy scale.

Power-counting has all of the utility and glamour of accounting: it is both crucial and not that exciting to do. The payoff for understanding it in detail is the power of the overall perspective it provides. One such insight is about what the cancellation of divergences during renormalization really means. In this new picture renormalization stops being a miraculous cancellation between divergences and countert-

erms and starts being an obvious cancellation of a scale, Λ, that is not really in the original problem.

A second insight concerns the utility of both renormalizable theories and non-renormalizable theories. Non-renormalizable theories are not daunting in themselves once it is recognized that they can be predictive to the extent that they only hope to capture low orders in a low-energy expansion. The enormous predictive success of renormalizable theories similarly emerges as a special case of this general low-energy predictiveness when the UV scale M is so high that it suffices to work to zeroth order in the ratio E/M.

Exercises

Exercise 3.1 Derive the cut-off dependent power-counting result of eq. (3.12) starting from the effective lagrangian of eq. (3.2).

Exercise 3.2 Extend the result of Exercise 3.1 to see how eq. (3.12) changes when graphs are built using fermions as well as bosons.

Exercise 3.3 For Quantum Electrodynamics there is only a single type of interaction, $\mathfrak{L}_{\rm int} = ieA_\mu\overline{\psi}\gamma^\mu\psi$, for which $d = 0$ (no derivatives), $f = 2$ (two fermion fields) and $b = 1$ (one bosonic field). Show that any Feynman graph built with only this vertex (and the Dirac and electromagnetic propagators) satisfies

$$\mathcal{I}_{_B} = \frac{\mathcal{E}_{_F}}{2} + \mathcal{L} - 1\,, \quad \mathcal{I}_{_F} = \frac{\mathcal{E}_{_F}}{2} + \mathcal{E}_{_B} + 2(\mathcal{L}-1) \quad \text{and} \quad \mathcal{V} = \mathcal{E}_{_F} + \mathcal{E}_{_B} + 2(\mathcal{L}-1)\,,$$

for the number of internal fermionic lines, bosonic lines and vertices. Here $\mathcal{E}_{_F}$ and $\mathcal{E}_{_B}$ are the number of external fermionic and bosonic lines and $\mathcal{L}$ is the number of loops, defined by (3.62).

Apply the power counting arguments leading to (3.12) to QED and show that they predict that a generic amputated Feynman graph evaluated at zero external momentum varies as $\mathcal{A}_{\mathcal{E}_F,\mathcal{E}_B} \propto \Lambda^{\mathcal{S}}$ where $\mathcal{S} = 4 - \frac{3}{2}\,\mathcal{E}_{_F} - \mathcal{E}_{_B}$ is called the graph's superficial degree of divergence. Show that $\mathcal{S}$ is only non-negative for configurations for which there exist tree-level vertices in the action.

Show that the power of electromagnetic coupling appearing in any graph is

$$\mathcal{A}_{\mathcal{E}_F\mathcal{E}_B} \propto e^{\mathcal{E}_F+\mathcal{E}_B-2}\left(\frac{e^2}{16\pi^2}\right)^{\mathcal{L}}\,.$$

and thereby that it is $\alpha/4\pi$, where $\alpha = e^2/4\pi$ is the fine-structure constant, that controls the loop expansion.

Exercise 3.4 The Fermi theory of weak interactions involves only fermions and has only a single interaction, for which $d = 0$ (no derivatives) and $f = 4$ (four fermion fields). Show that any Feynman graph built using only this vertex (and the Dirac propagators) satisfies $\mathcal{I} = 2(\mathcal{L}-1) + \frac{1}{2}\,\mathcal{E}$ and $\mathcal{V} = (\mathcal{L}-1) + \frac{1}{2}\,\mathcal{E}$, where $\mathcal{E}$ denotes the

number of external lines and $\mathcal{L}$ is the number of loops, defined by (3.5). Show that power-counting predicts a generic amputated Feynman graph evaluated at zero external momentum varies as $\mathcal{A}_\mathcal{E} \propto \Lambda^\mathcal{S}$ where $\mathcal{S} = 2(\mathcal{L}+1) - \frac{1}{2}\mathcal{E}$. Unlike for QED (see exercise 3.3) this eventually becomes positive (and so the graph becomes UV divergent) for large enough $\mathcal{L}$, regardless of the number of external lines involved.

Exercise 3.5 Complete the calculation of this chapter and use the toy model of §1.1 to compute the order λ^2/m_R^4 contribution to the effective coupling a appearing in the interaction $a\,(\partial_\mu\xi\,\partial^\mu\xi)^2 \in \mathfrak{L}_W$.

Exercise 3.6 Derive the central power-counting result, eqs. (3.64) and (3.65), starting from the lagrangian (3.61) and regulating UV divergences using dimensional regularization.

Exercise 3.7 Consider a Goldstone boson (or axion) with shift symmetry $\xi \to \xi +$ constant coupled to a fermion ψ at low energies. What are the lowest-dimension couplings possible involving these fields within the Wilsonian effective lagrangian? Assume the couplings are such that the lagrangian has the form (3.61) with two independent scales, M and v, with $v_B = v$, $v_F = M$ and $\dagger^2 = Mv$. Use the power-counting result of (3.64) to derive the leading dependence on these scales of the reduced amplitude $\mathcal{A}_4(q)$ for $2 \to 2$ fermion scattering in this theory. Draw the Feynman graphs that provide this leading contribution. Identify and draw the Feynman graphs that provide the next-to-leading contribution, both in the case where $M \sim 4\pi v$ and when $M \ll 4\pi v$. How suppressed are these subleading contributions relative to the leading order contribution? Repeat these leading and subleading power-counting estimates for $2 \to 2$ axion-fermion scattering rather than fermion-fermion scattering.

Exercise 3.8 Compute the graphs of Fig. 2.5 using the Toy Model's low-energy Wilsonian action to evaluate the one-loop contribution to the $\xi(q_1)\xi(q_2) \to \xi(q_3)\xi(q_4)$ scattering amplitude, $\mathcal{A}_4$, as a function of the Mandelstam variables $s = -(q_1+q_2)^2$, $t = (q_1-q_3)^2$ and $u = (q_1-q_4)^2$. Use dimensional regularization to evaluate any UV divergences you encounter. Perform the same calculation for $\phi_I\phi_I \to \phi_I\phi_I$ using the renormalizable Feynman rules of the underlying full theory, also to one-loop order using dimensional regularization. Show that these calculations agree on the leading low-energy limit of the amplitude, and in particular on the form and coefficient of the logarithmic term (which in the full theory requires a cancellation of the first few powers of q^2). Which Feynman graphs are responsible for the cancellations in the full theory?

Exercise 3.9 For the discussion of the 'method of regions' introduce a cutoff into the integrals $\mathcal{I}_\epsilon^{IR}$ and $\mathcal{I}_\epsilon^{UV}$ by defining

$$\mathcal{I}_\epsilon^{IR} := \int_0^\Lambda \left(\frac{\mu}{k}\right)^\epsilon f(k,M,m) \quad \text{and} \quad \mathcal{I}_\epsilon^{UV} := \int_\Lambda^\infty \left(\frac{\mu}{k}\right)^\epsilon f(k,M,m)\,,$$

with integrands as given in (3.41) and (3.42). Evaluate the first few terms in the $1/M$ expansions explicitly as functions of m, M and Λ. Show that the sum $\mathcal{I}_\epsilon = \mathcal{I}_\epsilon^{IR} + \mathcal{I}_\epsilon^{UV}$ is given by the same result as in (3.40). This shows that there is no loss in removing the cutoff when defining $\mathcal{I}_\epsilon^{IR}$ and $\mathcal{I}_\epsilon^{UV}$, as done in the main text.

4 Symmetries

The above sections involving the toy model show that the low-energy implications of an EFT — such as the suppression of scattering amplitudes by powers of energy — can be much more transparent when some fields (*e.g.* the field ξ in the toy model) are used to represent the light particles in the Wilsonian theory, than they are when expressed in terms of others (such as ϕ_I in the toy model). Why should this be?

Notice that the issue here is *not* that different variables imply different predictions, since the calculations of §1 reveal the low-energy suppression of scattering amplitudes to be precisely the same when computed using either ξ or ϕ_I. The issue instead is why this suppression is manifest at every step when using ξ, whereas with ϕ_I the suppression emerges quite mysteriously only at the end of the calculation due to cancellations amongst the contributing Feynman graphs.

This section argues that the main difference between these variables is the way they realize the symmetries of the system. How they are realized is relevant because symmetries (by way of Goldstone's theorem - see below) are ultimately the origin of the low-energy suppression seen in scattering amplitudes. As is so often the case the lesson is: although predictions for physical quantities can be made using any variables you like, if you use the wrong ones you will be sorry.[1]

To make this point it is first good to step back and summarize some implications of symmetries more generally. In particular, from an EFT perspective the discussion divides into two cases, depending on whether or not particles related to one another by a symmetry all lie within the low-energy effective theory or if some symmetries relate low-energy states to high-energy states. This latter situation can happen in particular when the relevant symmetry is spontaneously broken. From the point of view of EFTs the main observation is that the nature of the description necessarily changes if the energy scale, v, associated with symmetry breaking becomes much larger than the scale, M, associated with any heavy states that have been integrated out.

4.1 Symmetries in field theory ♡

The first step is to review how symmetries act within quantum field theory, and more generally within quantum mechanics.

As reviewed in Appendix C.4, a symmetry is described in quantum mechanics by a

[1] This paraphrases one of Steven Weinberg's three laws of theoretical physics [40].

unitary transformation,[2] $|\psi\rangle \to |\psi'\rangle = U|\psi\rangle$ with $U^*U = UU^* = I$, within Hilbert space, that leaves the system's hamiltonian unchanged:

$$H \to H' = UHU^* = H\,. \tag{4.1}$$

Such transformations are important for (at least) two reasons:[3]

- *Spectral degeneracy:* Because (4.1) implies $HU = UH$, a symmetry can only relate distinct energy eigenstates to one another if they share exactly the same energy eigenvalue. That is, if $H|\psi_i\rangle = E_i|\psi_i\rangle$ and $H|\psi_j\rangle = E_j|\psi_j\rangle$ and $|\psi_j\rangle = U|\psi_i\rangle$ are all true for some i and j, then $E_i = E_j$. This ensures that energy eigenspaces can all be organized into linear representations of the symmetry group: $U|\psi_i\rangle = \mathcal{U}^j{}_i|\psi_j\rangle$ for some coefficients $\mathcal{U}^j{}_i$, where $U_1U_2 = U_3$ implies $(\mathcal{U}_1)^i{}_j(\mathcal{U}_2)^j{}_k = (\mathcal{U}_3)^i{}_k$.
- *Conservation laws:* If a symmetry is labelled by a continuous parameter, θ, (such as is true for rotations in space, or the symmetry (1.21) of the toy model), then it can be written $U(\theta) = \exp[\mathrm{i}\theta\, Q]$, for some hermitian operator Q (unitarity of U implies Q is hermitian). Because Q is hermitian it is an observable, and because $[U, H] = 0$ implies $[Q, H] = 0$ the quantity Q is conserved. That is, if $Q|\psi(t=0)\rangle = \mathfrak{q}\,|\psi(t=0)\rangle$ for some real eigenvalue $\mathfrak{q}$ at a particular time, then $Q|\psi(t)\rangle = \mathfrak{q}\,|\psi(t)\rangle$ for all t. This follows because $[Q, H] = 0$ implies $Q|\psi(t)\rangle = Q\exp[-\mathrm{i}Ht]|\psi(0)\rangle = \exp[-\mathrm{i}Ht]\,Q|\psi(0)\rangle$.

Similar implications also apply in quantum field theory, though with important qualifications. The difference arises because in essence field theory makes quantum mechanics local by assigning different operators at different spacetime points. As a result symmetries in field theory are usually defined in terms of local transformations amongst fields that leave the action invariant — such as, for the toy model, the transformation (1.21), whose action leaves the lagrangian density $\mathfrak{L}$ of (1.1) unchanged.

This difference in starting point sometimes leads to symmetries being realized differently in field theory than in ordinary quantum mechanics, as this and the following sections describe. In particular, for reasons explained below, it turns out that having a symmetry act linearly on fields need not imply that particles fall into linear representations of the symmetry with identical energies.

A second issue raised by the local framework of field theory is the possibility that continuous symmetries might have position-dependent symmetry parameters: $\theta = \theta(x)$. Spacetime dependent symmetry transformation rules are called *local* or *gauge* symmetries, in contrast with *global* symmetries — for which θ and $U(\theta)$ do not depend on spacetime position. The focus of this chapter is mostly on global symmetries, though some of the issues arising for gauge symmetries are also discussed (both here and in Appendices C.3.3 and C.5).

A further distinction amongst symmetries is between internal symmetries and spacetime symmetries. These differ by whether or not they act on spacetime position, with for

[2] Except for time-reversal, which is described by an anti-unitary transformation. The discussion of symmetries in quantum mechanics goes back to [41].

[3] Here and throughout the Einstein summation convention is used, for which repeated indices are implicitly summed over their entire range: $\mathcal{U}^j{}_i|\psi_j\rangle := \sum_j \mathcal{U}^j{}_i|\psi_j\rangle$.

example

$$\phi^a(x) \to U\phi^a(x)U^* = \mathcal{U}_b{}^a\,\phi^b(x)\,, \tag{4.2}$$

being an example of an internal symmetry (because the spacetime coordinate x is unchanged) while a Lorentz transformation like

$$V^\mu(x) \to UV^\mu(x)U^* = \Lambda_\nu{}^\mu V^\nu(x') \quad \text{with} \quad x'^\mu = \Lambda_\nu{}^\mu x^\nu\,, \tag{4.3}$$

is a representative spacetime symmetry. Both internal and spacetime symmetries can arise in global or gauged varieties, with the local spacetime symmetries leading to the diffeomorphism invariance of general-covariant theories like General Relativity. Unless otherwise stated most of this chapter restricts to internal symmetries, which is not too restrictive in practice because it includes the majority of the symmetries of practical interest in later applications.

4.1.1 Unbroken continuous symmetries

Suppose, then, that a field theory enjoys a symmetry defined as some action-preserving infinitesimal continuous global transformation of the fields of the form $\phi^i \to \phi^i + \delta\phi^i$ with

$$\delta\phi^i := \omega^a \Sigma_a^i(\phi)\,, \tag{4.4}$$

where ϕ^i denote a generic collection of fields and ω^a denote a set of independent and spacetime-independent symmetry parameters while $\Sigma_a^i(\phi)$ are a given (possibly nonlinear) collection of functions of the fields at a specific spacetime point.

What makes this a symmetry is the requirement that it leaves invariant the system's action: $\delta S = \int \mathrm{d}^4x\, \delta\mathfrak{L} = 0$. For internal symmetries the transformation satisfies the stronger condition $\delta\mathfrak{L} = 0$ separately at each point in spacetime, while for spacetime transformations the invariance of the action only requires the weaker condition

$$\delta\mathfrak{L} = \partial_\mu \left(\omega^a\, V_a^\mu\right) = \omega^a \partial_\mu V_a^\mu\,, \tag{4.5}$$

for some quantities $V_a^\mu(\phi)$, since in this case $\delta S = \int \mathrm{d}^4x\, \delta\mathfrak{L}$ can still vanish.

Noether's Theorem

In field theory the existence of a continuous class of action-preserving field transformations guarantees the existence of a conserved current, j^μ; with there typically being one current for every global continuous symmetry of the action [42]. To low orders in the derivative expansion it is usually enough to work with actions that depend only on the fields and their first derivatives, so for simplicity the rest of the argument deriving these currents is restricted to this case.

Consider therefore an action $S = \int d^4x\, \mathfrak{L}(\phi, \partial_\mu\phi)$, that by assumption is invariant under the transformations (4.4). Including global spacetime symmetries in the discussion, this invariance implies $\mathfrak{L}$ must vary at most into a total derivative, so (4.5) is satisfied with $\delta\mathfrak{L}$

on the left-hand side found by directly varying the fields and their derivatives. Equating to zero the coefficient of the arbitrary constant ω^a in the result then gives:

$$\begin{aligned}\partial_\mu V^\mu_a &= \frac{\partial \mathfrak{L}}{\partial \phi^i}\,\Sigma^i_a + \frac{\partial \mathfrak{L}}{\partial(\partial_\mu \phi^i)}\,\partial_\mu \Sigma^i_a \\ &= \left[\frac{\partial \mathfrak{L}}{\partial \phi^i} - \partial_\mu\left(\frac{\partial \mathfrak{L}}{\partial(\partial_\mu \phi^i)}\right)\right]\Sigma^i_a + \partial_\mu\left(\frac{\partial \mathfrak{L}}{\partial(\partial_\mu \phi^i)}\,\Sigma^i_a\right).\end{aligned} \tag{4.6}$$

This equation holds as an identity, both for arbitrary field configurations, ϕ^i, and for arbitrary (though spacetime-independent) symmetry parameters, ω^a.

The theorem follows from this last equation which says that the definitions

$$j^\mu_a := \frac{\partial \mathfrak{L}}{\partial(\partial_\mu \phi^i)}\,\Sigma^i_a - V^\mu_a, \tag{4.7}$$

automatically satisfy the property

$$\partial_\mu j^\mu_a = 0\,, \tag{4.8}$$

whenever they are evaluated at any solution to the equations of motion for ϕ^i — *i.e.* on fields satisfying $\delta S = 0$, which is equivalent to the field equation

$$\frac{\partial \mathfrak{L}}{\partial \phi^i} - \partial_\mu\left(\frac{\partial \mathfrak{L}}{\partial(\partial_\mu \phi^i)}\right) = 0\,. \tag{4.9}$$

The conclusion, eq. (4.8), is more general than the relativistic notation being used here seems to suggest, since it holds also for nonrelativistic systems. For these systems write $\rho_a = j^0_a$ for the time component of j^μ_a, and denote its spatial components by the three-vector $\mathbf{j}_a$. Then current conservation — eq. (4.8) — becomes the continuity equation

$$\frac{\partial \rho_a}{\partial t} + \nabla \cdot \mathbf{j}_a = 0\,. \tag{4.10}$$

Eqs. (4.8) and (4.10) are conservation laws because they guarantee that the *charges*, Q_a, defined by

$$Q_a(t) = \int_{\text{fixed } t} \mathrm{d}^3x\; \rho_a(\mathbf{r}, t) = \int_{\text{fixed } t} \mathrm{d}^3x\; j^0_a(x), \tag{4.11}$$

are conserved in the sense that they are independent of t. This t-independence may be seen by using Stokes' theorem to infer

$$\partial_t Q_a = \int \mathrm{d}^3x\; \partial_t \rho_a = -\int \mathrm{d}^3x\; \nabla \cdot \mathbf{j}_a = -\oint_{r\to\infty} \mathrm{d}^2\Omega\, \mathbf{e}_r \cdot \mathbf{j}_a = 0\,, \tag{4.12}$$

where $\mathbf{e}_r = \mathbf{r}/r$ is the unit vector in the radial direction, and the last equality assumes boundary conditions are such that the net flux of the current $\mathbf{j}_a$ through a sphere at spatial infinity vanishes.

Representation on particle states

The charges Q_a provide the link back to the usual description of symmetries in quantum mechanics because their commutator with the fields gives the symmetry transformation

itself

$$i\omega^a \left[Q_a, \phi^i(x)\right] = \mathrm{i}\omega^a \int_{y^0=x^0} \mathrm{d}^3y \left[\rho_a(y), \phi^i(x)\right] = \omega^a \Sigma_a^i[\phi(x)] = \delta\phi^i(x)\,. \tag{4.13}$$

The first equality here uses conservation of Q_a to choose the time at which $j_a^0(x)$ is evaluated to agree with the time appearing in $\phi^i(x)$. The second equality then uses the definition (4.7), written as

$$\rho_a = j_a^0 = \frac{\partial \mathfrak{L}}{\partial \dot\phi^i}\, \Sigma_a^i - V_a^0 = \Pi_j\, \Sigma_a^i + V_a^0\,, \tag{4.14}$$

where overdots denote differentiation with respect to time and $\Pi_j(x) = \delta S/\delta\dot\phi^j(x)$ is the canonical momentum for the field $\phi^i(x)$. The second equality of (4.13) then follows from these definitions, together with the equal-time canonical commutation relations

$$\left[\Pi_j(\mathbf{x},t), \phi^i(\mathbf{y},t)\right]_{x^0=y^0} = -\mathrm{i}\,\delta^3(\mathbf{x}-\mathbf{y})\delta^i_j \quad \text{and} \quad \left[\phi^i(\mathbf{x},t), \phi^j(\mathbf{y},t)\right] = 0\,. \tag{4.15}$$

This derivation also assumes both $\Sigma_a^i(\phi)$ and V_a^0 depend only on ϕ^j and not also on the canonical momenta.

Eq. (4.13) makes a connection to the usual story of symmetries in quantum mechanics because it ensures that $U = \exp[\mathrm{i}\omega^a Q_a]$ (if well-defined – see below for when it is not) would be the unitary operator that implements the action of a finite symmetry transformation within the Hilbert space:

$$\phi^i \to \tilde\phi^i = U\phi^i\, U^*\,. \tag{4.16}$$

To see how this works, consider a weakly coupled system of particles, for which interactions can be treated perturbatively. Working within the interaction picture means that the fields satisfy the free-field equations and so can be expanded in a complete set of free single-particle modes, $u_n^i(x)$, (see Appendix C for details)

$$\phi^i(x) = \sum_n \Big[u_n^i(x)\mathfrak{a}_n + \text{c.c.}\Big]\,, \tag{4.17}$$

where $\mathfrak{a}_n$ is the destruction operator of a particle with label 'n,' $\mathfrak{a}_n|m\rangle = \delta_{mn}|0\rangle$, whose adjoint is the particle creation operator: $|n\rangle = \mathfrak{a}_n^*|0\rangle$. Here $|0\rangle$ denotes the usual no-particle ground state defined by $\mathfrak{a}_n|0\rangle = 0$.

If the no-particle state is invariant under the symmetry (as would be automatic if it were non-degenerate and separated from all other energy eigenstates by an energy gap — as is often true for simple systems) then $U|0\rangle = |0\rangle = U^*|0\rangle$, and so single-particle states are related to one another by the symmetry just like in ordinary quantum mechanics. That is, because (4.16) implies $\tilde{\mathfrak{a}}_n = U\mathfrak{a}_n U^*$ (and its adjoint) it follows that

$$|\tilde n\rangle = \tilde{\mathfrak{a}}_n^*|0\rangle = (U\mathfrak{a}_n^* U^*)|0\rangle = U\,\mathfrak{a}_n{}^*|0\rangle = U|n\rangle\,. \tag{4.18}$$

In particular, this ensures all the usual consequences of symmetries within quantum mechanics; in particular that the single-particle states $|\tilde n\rangle$ and $|n\rangle$ must have the same energy, and so on. Since internal symmetries commute with spacetime translations, they do not change particle momenta and so particles related by a symmetry must also have the same rest mass.

Now comes the main point. When the ground state is invariant under a symmetry then all of the usual implications of the symmetry in quantum mechanics go through as usual. In particular there is no loss of generality in representing the symmetry linearly on single-particle states and so also using linear transformations amongst the creation and annihilation operators and fields: $\Sigma_a^i(\phi) = \mathcal{S}_j{}^i\,\phi^j$ with $\mathcal{S}_j{}^i$ field-independent. Such a symmetry is said to be *linearly realized*.

The toy model provides an explicit example of this type of symmetry, but only in the special case that $v = 0$ since this choice makes the classical ground state $\phi = v$ invariant under the symmetry $\phi \to e^{i\omega}\phi$ — or (1.21). As expected, in this case both fields ϕ_R and ϕ_I represent particles with exactly the same mass (both are massless in this limit). More generally, if the toy model were instead to have a scalar potential with the sign of the $\phi^*\phi$ term reversed, as in $V = V_0 + m^2\phi^*\phi + \frac{1}{4}\lambda(\phi^*\phi)^2$, then the invariant configuration $\phi = 0$ remains the ground state also for nonzero masses, and in this case both ϕ_R and ϕ_I share the common nonzero mass m.

4.1.2 Spontaneous symmetry breaking

The key assumption in the previous section is that the ground state is invariant, and this is absolutely crucial for the above arguments to go through. Furthermore, this is not an empty exception: for field theories the ground state can fail to be invariant.[4] When the ground state of a system is not invariant under a symmetry of its action the symmetry is said to be *spontaneously broken*.

If a symmetry is spontaneously broken then (by assumption) another state is produced once a transformation is applied to the ground state. Since the transformation is a symmetry, this new state must have the same energy and so also be a candidate ground state. For a continuous symmetry one expects a continuous family of vacua, all sharing the same energy. This is indeed what happens for the toy model, for which the semiclassical vacuum corresponds to any spacetime-independent configuration satisfying $\phi^*\phi = v^2$. The one-parameter family of ground states is parameterized by $\phi = v\,e^{i\xi}$ for any constant ξ. From here on a ground state of such a system is denoted by $|\Omega\rangle$ rather than $|0\rangle$, to emphasize the fact that it can be more complicated than the single no-particle state of a simple harmonic Fock space.

Whenever $U|\Omega\rangle \neq |\Omega\rangle$ the line of argument given above that says single-particle states, $|n\rangle = a_n^*|\Omega\rangle$, must be linearly related by the symmetry, $|\tilde{n}\rangle = U|n\rangle$, also fails.[5] It fails because the symmetry changes the ground state, $|\Omega\rangle \to |\widetilde{\Omega}\rangle$, in addition to acting on the single-particle states that are built from them. As a result fields related by the symmetry (like ϕ_R and ϕ_I of the toy model) need no longer correspond to particles with equal masses, unlike what happens when the vacuum is invariant). This is seen explicitly in the toy model spectrum when $v \neq 0$, since in this case the fields ϕ_R and ϕ_I represent particles with masses $m_R = \sqrt{\lambda}\,v$ and $m_I = 0$ respectively, despite being linearly related by the symmetry (1.21).

[4] This is unlike what happens for the quantum mechanics of a small number of degrees of freedom, for which the ground state tends to be unique (and so therefore is also invariant under a symmetry transformation).

[5] For field theories there can be problems even defining operators like U for spontaneously broken symmetries [43] (see also Appendix C.5.1).

Although traditional implications (like equality of particle masses) can break down for spontaneously broken symmetries, these symmetries nonetheless do come with consequences, as is now described [3, 4, 5].

Goldstone's Theorem

Whenever the ground state of a system does not respect one of the system's global continuous symmetries, there are very general implications for the low-energy theory that are summarized by *Goldstone's theorem* [4].

Goldstone's theorem states that any system for which a continuous, global symmetry is spontaneously broken must contain a state, $|G\rangle$ — called a *Goldstone mode*, or *Nambu-Goldstone boson* [6] — with the defining property that it is created from the ground state by a spacetime-dependent symmetry transformation.

In equations, $|G\rangle$ is defined by the condition that the following matrix element cannot vanish:

$$\langle G|\,\rho(\mathbf{r},t)|\Omega\rangle \neq 0\,, \tag{4.19}$$

where $|\Omega\rangle$ represents the ground state and[7] $\rho = j^0$ is the density for the symmetry's conserved charge, as is guaranteed to exist by Noether's theorem.

All of the properties of a Goldstone state follow from the definition (4.19), but before turning to them it is worth first sketching why this equation is true. The starting point is the assumption of the existence of a *local order parameter*. This is a field, $\phi(x)$, in the problem satisfying two defining conditions:

1. ϕ must transform nontrivially under the symmetry in question: *i.e.* there is another field, $\psi(x)$, for which:

   $$\delta\psi \equiv \mathrm{i}[Q,\psi(x)] = \phi(x)\,, \tag{4.20}$$

   where Q is the conserved charge defined by integrating the current density, $\rho(\mathbf{r},t)$, throughout all of space.
2. The field ϕ must have a nonzero expectation in the ground state:

   $$\langle\phi(x)\rangle := \langle\Omega|\phi(x)|\Omega\rangle = v \neq 0. \tag{4.21}$$

   This last condition would be inconsistent with eq. (4.20) if the ground state were invariant under the symmetry of interest, since invariance means $Q|\Omega\rangle = 0$, and this would mean the expectation value of eq. (4.20) must vanish.

To see why (4.19) follows from (4.20) and (4.21) use the following steps. First substitute (4.20) into eq. (4.21). Second, use $Q = \int \rho\, \mathrm{d}^3x$ in the result, as is guaranteed to be possible by Noether's theorem. Third, insert a partition of unity, $1 = \sum_n |n\rangle\langle n|$, between the operators ρ and ψ. The resulting expression shows that if no state exists satisfying the defining

[6] For internal symmetries this state must be a boson, but for graded symmetries like supersymmetry spontaneously breakdown ensures the existence of a Goldstone fermion, the goldstino.

[7] The nonrelativistic notation $\rho = j^0$ is used to emphasize that the conclusions presented are not specific to relativistic systems.

condition, eq. (4.19), then the right-hand-side of eq. (4.21) must vanish, in contradiction with the starting assumptions.

Goldstone's theorem states that the consistency of the matrix element, eq. (4.19), with the conservation law, eq. (4.10), requires the state $|G\rangle$ to have a number of important properties. Two of these are the state's spin and statistics. Because $\rho(x)$ transforms under rotations as a scalar and because $|\Omega\rangle$ is rotationally invariant, it follows that $|G\rangle$ must have spin zero and so (from the spin-statistics theorem) must also be a boson.

Furthermore, the state $|G\rangle$ must also be *gapless*, in that its energy must vanish in the limit that its (three-) momentum vanishes:

$$\lim_{p\to 0} E(p) = 0. \tag{4.22}$$

In relativistic systems, for which $E(p) = \sqrt{p^2 + m^2}$ with m being the particle's rest mass, the gapless condition is equivalent to the masslessness of the Goldstone particle. This gaplessness follows by using the fact that $|G\rangle$ and $|\Omega\rangle$ are both energy and momentum eigenstates (with $E_\Omega = \mathbf{p}_\Omega = 0$) to write (see e.g. Appendix C.5.1)

$$\langle G|\rho(\mathbf{r},t)|\Omega\rangle = e^{\mathrm{i}E_G t - \mathrm{i}\mathbf{p}_G\cdot\mathbf{r}}\langle G|\rho(0)|\Omega\rangle\,, \tag{4.23}$$

Using this (and a similar expression for $\langle G|\mathbf{j}(\mathbf{r},t)|\Omega\rangle$) when taking the matrix element of the conservation equation, (4.10), between $\langle G|$ and $|\Omega\rangle$, leads to

$$0 = \langle G|\left(\partial_t\rho + \nabla\cdot\mathbf{j}\right)|\Omega\rangle = \mathrm{i}\Big[E_G\langle G|\rho(\mathbf{r},t)|\Omega\rangle - \mathbf{p}_G\cdot\langle G|\mathbf{j}(\mathbf{r},t)|\Omega\rangle\Big]. \tag{4.24}$$

Since the last term vanishes as $\mathbf{p}_G \to 0$ and we know $\langle G|\rho(\mathbf{r},t)|\Omega\rangle \neq 0$ it follows that E_G also vanishes in this limit.

More generally, the Goldstone boson must completely decouple from all of its interactions in the limit that its momentum vanishes. This is because eq. (4.19) states that in the zero-momentum limit the Goldstone state literally is a symmetry transformation of the ground state. As a result it is *completely indistinguishable* from the vacuum in this limit.

These properties say a lot about the low-energy behaviour of any system that satisfies the assumptions of the theorem. The first guarantees that the Goldstone boson must itself be one of the light states of the theory, and so it must be included in any effective lagrangian analysis of this low energy behaviour. The second property ensures that the Goldstone mode must be very weakly coupled in the low-energy limit, and strongly limits the possible form its interactions can take.

The toy model again provides a simple example of all of these consequences. The symmetry of this model is spontaneously broken whenever $v \neq 0$, and there is certainly a gapless state in the spectrum whenever this is true: the state represented by the field ϕ_I or ξ. To see that this state satisfies (4.19) requires constructing the Noether current for the symmetry (1.21). For the toy model this is

$$\begin{aligned} j_\mu = \mathrm{i}(\phi\,\partial_\mu\phi^* - \phi^*\partial_\mu\phi) &= \sqrt{2}\left(v + \frac{\hat\phi_R}{\sqrt{2}}\right)\partial_\mu\hat\phi_I - \hat\phi_I\,\partial_\mu\hat\phi_R \\ &= \sqrt{2}\left(v + \frac{\chi}{\sqrt{2}}\right)^2\frac{\partial_\mu\xi}{v} \end{aligned} \tag{4.25}$$

and so its single-particle matrix element is $\langle \xi(p)|j^\mu(x)|\Omega\rangle \propto \sqrt{2}\, v p^\mu\, e^{-ip\cdot x}$, which is nonzero whenever $v \neq 0$. Furthermore, conservation of the Noether current also implies this particle is massless because

$$0 = \langle \xi(p)|\partial_\mu j^\mu(x)|\Omega\rangle \propto \partial_\mu\Big(\sqrt{2}\, v p^\mu\, e^{-ip\cdot x}\Big) = -i\sqrt{2}\, v\, p_\mu p^\mu e^{-ip\cdot x} = i\sqrt{2}\, v\, m^2 e^{-ip\cdot x}\,. \quad (4.26)$$

The more general statement that interactions for the Goldstone particle must turn off at zero momentum is also clear for the toy model, since this is the property (much discussed in §1) that the scattering amplitudes for toy-model massless states approach zero as the scattering energy goes to zero.

4.2 Linear vs nonlinear realizations ◇

With the above discussion in mind, we are in a position to formalize more explicitly how symmetries are realized within a Wilsonian low-energy effective theory. The most basic statement is that the low-energy theory must share the symmetry properties of the full UV theory, both for the symmetries of the action and the symmetries of the ground state. That is, it should be possible to read off the symmetry properties of the system directly from the EFT at any scale one chooses.[8]

The discussion of the previous section shows that how this is done depends on whether or not the symmetry of interest is spontaneously broken. If the symmetry is unbroken then it can without loss of generality be realized to act linearly on the fields. Since in this case all particles related by the symmetry share the same mass, if any of them is light enough to be in the low-energy theory then all of them are. But the same need not be true if the symmetry is spontaneously broken.

Whether a linear realization is possible for a spontaneously broken symmetry depends on the relative size of two important scales: the scale M of the UV physics whose integrating out led to the EFT in question, and the scale v of the expectation value responsible for the symmetry's spontaneous breaking (see Fig. 4.1). This section argues that a linear realization can continue to be useful when $v \ll M$, but necessarily breaks down if M is too small relative to v. If the symmetry cannot be realized linearly, there turns out to be an alternative standard realization that is always possible for a broad class of symmetry-breaking patterns.

The toy model provides an example where linear realization remains useful even though a symmetry spontaneously breaks. That is, suppose the entire toy model were itself regarded as being the low-energy limit of some larger UV completion; being obtained by integrating out some new states, ψ^i, with masses $M_\psi \gg v \gtrsim m_R = \sqrt{\lambda}\, v$. In this case the toy model (plus all possible higher-dimensional interactions built from ϕ) is the EFT for energies in the regime $v \ll E \ll M_\psi$, and includes both fields ϕ_R and ϕ_I, with the $U(1)$ toy model symmetry realized linearly as is done in (1.21).

[8] Anomalies – the failure of a classical symmetry to survive quantization – can complicate this statement slightly, inasmuch as what can look like an anomalous symmetry at some scales can look like a classical breaking of the symmetry at other scales. More about this in §4.3

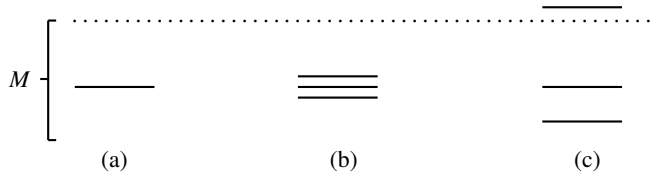


Fig. 4.1 A sketch of energy levels in the low-energy theory relative to the high-energy scale, M, and the relative splitting, v, within a global 'symmetry' multiplet. Three cases are pictured: panel (a) unbroken symmetry (with unsplit multiplets); panel (b) low-energy breaking ($v \ll M$) and panel (c) high-energy breaking (with $v \gtrsim M$). Symmetries are linearly realized in cases (a) and (b) but not (c). If spontaneously broken, symmetries in case (c) are nonlinearly realized in the EFT below M. (If explicitly broken in case (c) there is little sense in which the effective theory has approximate symmetry at all.)

But this linear realization is no longer possible for an EFT aimed at the regime $E \ll m_R$. Linear realization is not possible in this regime because one of the two fields required for its existence is now integrated out. Notice that the possibility of integrating out part of a symmetry multiplet only arises if the symmetry is spontaneously broken because only then can some particles within a symmetry multiplet differ in mass. In the toy model what is required in this regime is a way to realize the symmetry using the EFT's only low-energy field. The required realization is provided by the symmetry under which the field ξ shifts, as in (1.23). The inhomogeneous nature of this symmetry is characteristic of a Goldstone boson, since the symmetry is necessarily inconsistent with choosing a specific vacuum value for the field ξ.

4.2.1 Linearly realized symmetries

The simplest situation is when $v \ll M$, which includes the case $v = 0$ where the symmetry is not broken at all. In this case the low-energy theory contains the right number of particles to fill out linear representations of the symmetry, and so fields can be chosen in such a way as to represent the symmetry linearly. For an internal symmetry this means we can take:[9]

$$\phi^i(x) \to \tilde{\phi}^i(x) = \mathcal{M}^i{}_j\, \phi^j(x)\,, \tag{4.27}$$

for some choice of matrices $\mathcal{M}^i{}_j$.

Not much need be said in this case, which is the one most commonly used in practice. Because all of the fields needed for a linear realization are present, the potential order parameters are also present as fields within the low-energy theory. The only difference be-

[9] The attentive reader may notice a difference in the index ordering between this equation and (4.2). These representations are conjugates of one another, with the choice of (4.2) designed to ensure that if $U_3 = U_1 U_2$ then $(\mathcal{U}_3)_i{}^j = (\mathcal{U}_1)_i{}^k (\mathcal{U}_2)_k{}^j$.

tween unbroken symmetry and spontaneous breaking therefore lies in the choice of action and the energy it assigns to a nonzero value for the order parameters.

Explicit symmetry breaking

An important variation on this chapter's theme is the situation where symmetries are only approximate rather than exact. In this case the action for the system at any scale is assumed to have the form

$$S = S_{\rm inv} + \epsilon S_1 + \epsilon^2 S_2 + \cdots , \tag{4.28}$$

where $S_{\rm inv}$ is invariant under some group of symmetries while the S_i are not. Some small dimensionless parameter ϵ is assumed to be present to quantify the notion that the symmetry breaking is 'small'. In particular, if an expansion like (4.28) is possible in the full high-energy theory, it must also be possible in the low-energy theory (which can be a useful observation when explicit *ab-initio* calculations are not possible in the high-energy theory – an important example of which is described in §8).

An explicit example helps make things concrete. The toy model's $U(1)$ symmetry $\phi \to e^{i\theta}\phi$ is responsible for many of its predictions, such as the equality of the two particle masses when $v = 0$ and the elimination of the Goldstone boson from the scalar potential when $v \neq 0$. Both of these properties are easily seen to fail once interactions breaking the $U(1)$ symmetry are included. Approximate symmetries remain useful, however, because any failure of symmetry relations tends to zero as the symmetry breaking turns off (*i.e.* as $\epsilon \to 0$), allowing predictions to be given perturbatively, in powers of ϵ.

The simplest example of explicit symmetry breaking in this model is to add a term linear in ϕ that tilts the potential, such as by adding to (1.1) the symmetry-breaking term

$$\Omega_{\rm sb} = \mu^3 (\phi + \phi^*) , \tag{4.29}$$

where μ is a symmetry-breaking parameter with dimensions of mass. Writing $\phi = \frac{1}{\sqrt{2}} (\phi_R + i\phi_I)$ the modified potential becomes

$$V(\phi_R, \phi_I) = \frac{\lambda}{16} \left(\phi_R^2 + \phi_I^2 - 2v^2\right)^2 - \sqrt{2}\mu^3 \phi_R , \tag{4.30}$$

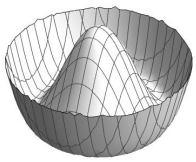
which has extrema that are given to leading order in $0 \leq \mu^3/(\lambda v^3) \ll 1$ by

$$\phi_0 \simeq -\frac{2\mu^3}{\lambda v^2} \quad \text{and} \quad \phi_\pm \simeq \pm v + \frac{\mu^3}{\lambda v^2} . \tag{4.31}$$

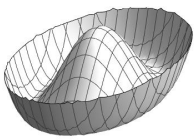
The degenerate circular minimum present for the sombrero-shaped potential when $\mu = 0$ is now tilted, leaving a unique real minimum at ϕ_+ (assuming $\mu > 0$) with a saddle point on the opposite side of the circle at ϕ_-. ϕ_0 turns out to be the shifted local maximum at the centre of the Mexican hat. The mass eigenvalues for the two fields, found as before by evaluating the potential's second derivatives at $\phi_\pm$, then become

$$m_{R\pm}^2 = \lambda v^2 \pm \frac{3\mu^3}{v} \quad \text{and} \quad m_{I\pm}^2 = \pm \frac{\mu^3}{v} . \tag{4.32}$$

In particular this shows how explicit symmetry breaking gives the would-be Goldstone



(a) The shape of the potential $V(\phi_R, \phi_I)$, in the limit of no explicit symmetry breaking, showing its sombrero shape and the circular line of minima.



(b) The potential $V(\phi_R, \phi_I)$, with an explicit symmetry-breaking term added that is linear in ϕ_R, showing how the degeneracy of the minimum gets lifted.

boson state — commonly called a 'pseudo-Goldstone' boson [44] — a nonzero mass $m_G^2 = m_{I+}^2 \simeq \mu^3/v$, that vanishes (as it must) when $\mu \to 0$.

Explicit symmetry breaking also interferes with the other low-energy Goldstone properties, as can be seen by expressing the toy model in terms of the fields χ and ξ using (1.22). With these variables the symmetry-breaking term (4.29) modifies the potential from (1.25) to

$$V(\chi, \xi) = \frac{\lambda}{4}\left(\sqrt{2}\, v\chi + \frac{\chi^2}{2}\right)^2 - 2\mu^3\left(v + \frac{\chi}{\sqrt{2}}\right)\cos\left(\frac{\xi}{\sqrt{2}\, v}\right). \tag{4.33}$$

The would-be Goldstone particle, ξ, no longer drops out of the potential and as a result its scattering amplitudes no longer need vanish at low energies, by a calculable amount in powers of μ^3. The low-energy Wilson action for ξ then also acquires a scalar potential whose leading contribution is proportional to μ^3, in order to capture the low-energy limit of the full theory's symmetry-breaking behaviour,

$$V_W(\xi) \simeq -2\mu^3 v \cos\left(\frac{\xi}{\sqrt{2}\, v}\right) + \cdots. \tag{4.34}$$

Here the ellipses represent terms suppressed by additional powers of $1/m_R$ and/or μ^3.

The toy model illustrates what is also true in the general case: what were exact symmetry implications survive in an approximate form in the limit that the symmetry-breaking physics is small. Deviations from these predictions are then obtained by expanding systematically in powers of the small symmetry-breaking parameter ϵ. In particular, although the light particle ξ is no longer massless, it remains much lighter than the heavy χ particle provided the symmetry-breaking physics is small: $m_G^2/m_R^2 = \mu^3/(\lambda v^3)$. Although nonzero, its low-energy couplings at zero momentum are similarly suppressed. Chapter 8 includes a very practical example along these lines that was very influential for the development of effective field theories.

4.2.2 Nonlinearly realized symmetries

An important constraint on any low-energy EFT is the requirement that it share all the symmetry properties of its underlying UV completion. This is fairly straightforward to do when the symmetry is linearly realized since the symmetry then groups particles into multiplets with similar masses and couplings, both at high and low energies.

How spontaneously broken symmetries get expressed in the low-energy theory is more subtle, particularly when the symmetry-breaking scale, v, is bigger than the UV scale M — *i.e.* panel (c) of Fig. 4.1. In the regime $E \ll M \ll v$ the low-energy field content can be inconsistent with realizing the symmetry linearly, since some members of multiplets in the UV theory might be heavy enough to have been integrated out while others remain light enough to appear in the effective theory.

This section explores how the symmetries of the underlying UV theory manifest themselves at low energies in this case.

Abelian case

Once again the toy model is informative, since it enjoys an abelian $U(1)$ global symmetry and its best-understood parameter regime (that of small λ) satisfies $m_R \lesssim v$. The low-energy theory described in earlier sections for $E \ll m_R$ therefore falls precisely into the regime of interest. With only one low-energy field present at these energies it is impossible to realize the $U(1)$ linearly, and it is instead realized as an inhomogeneous shift symmetry on the one low-energy field: $\xi \to \xi + \sqrt{2}\,\omega\, v$ (where ω is the symmetry parameter).

In §1 this formulation was found by redefining the fields $\phi = F(\chi)\, e^{i\xi/f}$ — with the specific form $F(\chi) = v + (\chi/\sqrt{2})$ not important and $f = \sqrt{2}\, v$ if ξ is to be canonically normalized. The significance of this choice is that ξ appears as would the parameter of a symmetry transformation, $\phi \to \phi\, e^{i\theta}$, with θ replaced by $\xi(x)/(\sqrt{2}\ v)$. Because ξ, when spacetime-independent, appears as does a symmetry parameter, constant configurations of ξ must drop out of the action (which is after all invariant under the symmetry). Consequently the action can only depend on ξ through its derivative, $\partial_\mu \xi$. The heavy field χ is then found by identifying that part of ϕ that is in some sense orthogonal to this symmetry direction.

Two things are instructive about this construction. First, as the next section shows, it can be extended to more general (and in particular nonabelian) symmetries. Second, the rules for constructing the most general $U(1)$-invariant lagrangian in the EFT for the toy model are fairly simple, and also generalize to more complicated groups.

If written out pedantically, the instructions for building a general interaction for a $U(1)$ Goldstone boson (such as ξ in the toy model) go as follows. First build a generic local action using a vector field V_μ and its derivatives, plus any other fields ψ^m that happen to be in the low-energy theory, constrained only by Lorentz-invariance (or whatever other spacetime and unbroken internal symmetries are relevant). In particular, do so without making any reference at all to a global $U(1)$ symmetry. The terms involving the fewest V_μ fields and

the fewest derivatives then have the general form

$$\mathfrak{L}(\psi, V_\mu) = \mathfrak{L}_0(\psi) - j^\mu(\psi)V_\mu - \frac{1}{2}\, m^{\mu\nu}(\psi)\, V_\mu V_\nu - n^{\mu\nu}(\psi)\, \partial_\mu V_\nu + \cdots , \tag{4.35}$$

where $\mathfrak{L}_0$, j^μ, $m^{\mu\nu}$ and $n^{\mu\nu}$ are functions of the ψ^m and their derivatives. Then *any* such a lagrangian is automatically promoted to a $U(1)$-invariant one simply by replacing everywhere $V_\mu \to \partial_\mu \xi$.

The lesson is this: whereas a linearly realized symmetry restricts the kinds of terms that are allowed in a lagrangian (such as by forbidding interactions with unequal powers of ϕ and ϕ^* in the toy model lagrangian), the spontaneously broken $U(1)$ symmetry does not directly restrict the kinds of terms that can be written in (4.35). What the spontaneously broken $U(1)$ symmetry instead does is dictate how the Goldstone boson must couple to other fields at low energies, given a generic lagrangian, like (4.35), constrained only by the *unbroken* symmetries. It is only through the couplings of the Goldstone field that the low-energy EFT 'learns' about the existence of the broken $U(1)$ theory in its UV completion.

In particular, it is ultimately the shift-symmetry realization of the $U(1)$ symmetry in the low-energy EFT (with $v \gg M$) that forces the Goldstone boson to couple to other fields only through derivatives. In this way the implications of Goldstone's theorem emerge as automatic symmetry consequences when constructing the Wilson action. Properly realizing symmetries in low-energy effective actions means not having to be clever in order to extract their low-energy consequences, with no need for fancy 'current algebra' operator arguments (*e.g.* compare [11] with [45, 181, 180] or [48] to [49]).

Nonabelian case ♠

This section sketches how nonlinear realizations work for more general patterns of spontaneous symmetry breaking, for which the action has a symmetry group G but the ground state is invariant only under a subgroup $H \subset G$. The treatment here is meant mainly to summarize the main results, but since this is an important EFT topic more details are presented about the motivation for and derivation of the results given here – the so-called 'standard realization' – as well as a lightning review of the main properties of Lie groups and Lie algebras that are needed to do so.

There are many ways to represent the Goldstone bosons for a generic symmetry-breaking pattern, $G \to H$, but these are all related by field redefinitions to (and so are equivalent to) a standard one, whose properties are summarized here. (Geometrically, the Goldstone boson fields can be regarded as being coordinates on the coset space G/H; see Appendix C.6.) Before describing this, though, there are a few group-theoretical facts worth collecting.

Group-theoretic aside

As described in Appendix C.6, it is useful to work with the Lie algebra of G rather than in terms of the group itself. Any group element connected to the identity element, $g = 1$, can be written as a matrix exponential:

$$g = \exp\left[\mathrm{i}\omega^a T_a\right] , \tag{4.36}$$

where for a p-parameter group the p generators, T_a, $a = 1, ..., p$, form a basis of the Lie algebra of G. As is often the case in physics, in the first instance the interest is often in a specific matrix representation of the group elements for $g \in G$, perhaps acting on the fields of interest in the UV theory for which the symmetries are linearly realized, rather than something more abstract. For N fields the matrices g would be $N \times N$. As discussed earlier, these specific matrix representations are unitary and they can be also chosen to be real. (Complex fields are, like ϕ in the toy model, can always be broken into their real and imaginary parts, like ϕ_R and ϕ_I.) When this is the done the matrices T_a are hermitian and imaginary:

$$T_a = T_a^\dagger = -T_a^* = -T_a^T \,, \tag{4.37}$$

where the superscript 'T' denotes transpose.

The properties of the group G are encoded in its group multiplication law, and this also has implications for the matrices T_a. In particular, the closure property of the group multiplication law for G implies the generators satisfy commutation relations

$$T_a\, T_b - T_b\, T_a = \mathrm{i}\, c_{abd}\, T_d \,, \tag{4.38}$$

with an implied sum on the index 'd'. Like the group multiplication law, the constant coefficients, c_{abd}, are characteristic of the group involved. (For more about Lie algebras see Appendix C.4.1.)

When describing a symmetry-breaking pattern where G breaks to H it is convenient to choose the basis of generators to include the generators of H as a subset:

$$\{T_a\} = \{t_i, X_\alpha\} \,, \tag{4.39}$$

where the t_i's generate the Lie algebra of H and the X_α's constitute the rest. The broken generators X_α typically do not also generate a group, since the commutator of two X_α's need not involve only X_α's with no t_i's. They instead can be regarded as generating[10] the space of 'cosets', G/H.

The closure of H under multiplication ensures that

$$t_i\, t_j - t_j\, t_i = \mathrm{i}\, c_{ijk}\; t_k \,, \tag{4.40}$$

with no X_α's on the right-hand-side, so $c_{ij\alpha} = 0$. Under broad assumptions it is also possible to choose a basis of generators so that $c_{i\alpha j} = 0$, so

$$t_i\, X_\alpha - X_\alpha\, t_i = \mathrm{i}\, c_{i\alpha\beta} X_\beta \,, \tag{4.41}$$

with no t_j's on the right-hand-side. This implies the X_α's fall into a (possibly reducible) representation of H, which when exponentiated to a finite transformation implies

$$h\, X_\alpha\, h^{-1} = L_\alpha{}^\beta X_\beta \tag{4.42}$$

for some coefficients, $L_\alpha{}^\beta$ and for any $h = \exp[\mathrm{i}\omega^i t_i] \in H$.

[10] Formally, a coset G/H is an equivalence class wherein two elements $g_1, g_2 \in G$ are regarded as equivalent if g_1 can be obtained from g_2 by multiplying by some $h \in H$.

The realization

For internal symmetries, with the symmetry-breaking pattern $G \to H$, the low-energy theory contains a Goldstone boson, ξ^α, for each broken generator, X_α (the counting can be different for other situations, like spacetime symmetries – see §14.3). The low-energy theory might also contain a collection of other non-Goldstone low-energy fields, χ^n, depending on the particular system of interest. The rest of this section provides an explicit nonlinear realization of the symmetry group G on the collection $\{\xi^\alpha\}$ and separately on the collection $\{\chi^n\}$.

As motivated in Appendix C.6 it is always possible to perform a field redefinition so that the fields ξ^α and χ^n transform according to [12, 13]

$$\xi^\alpha \to \tilde{\xi}^\alpha(\xi, g) \qquad \text{and} \qquad \chi^n \to \tilde{\chi}^n(\xi, g, \chi), \tag{4.43}$$

where $\tilde{\xi}^\alpha$ and $\tilde{\chi}^n$ are defined by the relations

$$g\, e^{i\xi^\alpha X_\alpha} = e^{i\tilde{\xi}^\alpha X_\alpha}\, e^{iu^i t_i} \quad \text{and} \quad \widetilde{\mathcal{X}} = e^{iu^i t_i}\, \mathcal{X}\,. \tag{4.44}$$

Here $\mathcal{X}$ (or $\widetilde{\mathcal{X}}$) denotes a column vector whose entries are the fields χ^n (or $\tilde{\chi}^n$).

The first of eqs. (4.44) should be read as defining the nonlinear functions $\tilde{\xi}^\alpha(\xi, g)$ and $u^i(\xi, g)$. Starting with $e^{i\xi\cdot X}$ one multiplies through on the left by $g \in G$ to construct a new element of G: $g\, e^{i\xi\cdot X}$. The functions $\tilde{\xi}^\alpha$ and u^i are then defined by decomposing this new matrix into the product of a factor, $e^{i\tilde{\xi}\cdot X}$, lying in G/H times an element, $e^{iu\cdot t}$, in H. The second of eqs. (4.44) then defines the transformation rule for the non-Goldstone fields, χ^n.

These transformations simplify in the special case where $g = h$ lies in H, in which case both χ^n and ξ^α turn out to transform *linearly* under the unbroken symmetry transformations of H. The simplification happens because the above definitions in this case reduce to:

$$\begin{aligned} \xi^\alpha X_\alpha \to \tilde{\xi}^\alpha X_\alpha = h(\xi^\alpha X_\alpha)h^{-1} = \xi^\alpha L_\alpha{}^\beta X_\beta\,, \\ \mathcal{X} \to \widetilde{\mathcal{X}} = h\mathcal{X}\,, \end{aligned} \tag{4.45}$$

where the last equality in the first line uses (4.42).

More generally the transformation laws are both inhomogeneous and nonlinear in the Goldstone fields, ξ^α. Explicit closed-form expressions can be found for infinitesimal transformations — *c.f* eqs. (C.123) and (C.124) — which when expanded in powers of fields give[11]

$$u^i \approx -c^i{}_{\alpha\beta}\omega^\alpha\xi^\beta + O(\omega\xi^2) \qquad \text{and} \qquad \delta\xi^\alpha = \omega^\alpha - c^\alpha{}_{\beta\gamma}\omega^\beta\xi^\gamma + O(\omega\xi^2)\,. \tag{4.46}$$

Because these transformations are nonlinear they are effectively spacetime-dependent due to their dependence on the field $\xi^\alpha(x)$. This complicates the algorithm for finding the general form for invariant lagrangians, as is now briefly described.

Invariant actions

The field-dependent matrix-valued quantity $U(\xi) := \exp[i\xi^\alpha(x)X_\alpha]$ provides the starting point for constructing actions that are invariant under transformations (4.43) and (4.44). To

[11] Indices on structure constants are raised and lowered using the Killing metric defined in Appendix C.4.1.

see why, notice that U transforms under the above rules as $U(\xi) \to U(\tilde{\xi})$ where $gU(\xi) = U(\tilde{\xi})\,\mathtt{h}$, with

$$\mathtt{h} := \exp\Big[\mathrm{i}u^i(\xi, g)t_i\Big] \in H\,, \tag{4.47}$$

being the matrix appearing in (4.44). The key observation is that this implies the combination[12] $U^{-1}\partial_\mu U$ transforms like a gauge-potential (compare this with eq. (C.71), keeping in mind the footnote immediately after (C.129))

$$U^{-1}\partial_\mu U \to \tilde{U}^{-1}\partial_\mu \tilde{U} = \mathtt{h}\,(U^{-1}\partial_\mu U)\,\mathtt{h}^{-1} - \partial_\mu \mathtt{h}\,\mathtt{h}^{-1}\,. \tag{4.48}$$

Separating $U^{-1}\partial_\mu U$ into a piece proportional to X_α plus one proportional to t_i defines two important quantities, $\mathcal{A}^i_\mu(\xi)$ and $e^\alpha_\mu(\xi)$, according to

$$U^{-1}\partial_\mu U = -\mathrm{i}\mathcal{A}^i_\mu t_i + \mathrm{i}e^\alpha_\mu X_\alpha\,. \tag{4.49}$$

Extracting an overall factor of $\partial_\mu\xi^\alpha$, so that $\mathcal{A}^i_\mu = \mathcal{A}^i_\alpha(\xi)\ \partial_\mu\xi^\alpha$ and $e^\alpha_\mu = e^\alpha{}_\beta(\xi)\ \partial_\mu\xi^\beta$, then the explicit expressions for the small-ξ expansion of these quantities become (see Appendix C.6.3)

$$\mathcal{A}^i_\alpha(\xi) = -\int_0^1 ds\ \ \mathrm{Tr}\Big[t^i e^{-\mathrm{i}s\,\xi\cdot X} X_\alpha\, e^{\mathrm{i}s\,\xi\cdot X}\Big] \simeq \frac{1}{2}\,c^i{}_{\alpha\beta}\xi^\beta + O(\xi^2), \tag{4.50}$$

and

$$e^\alpha{}_\beta(\xi) = \int_0^1 ds\ \ \mathrm{Tr}\Big[X^\alpha e^{-\mathrm{i}s\,\xi\cdot X} X_\beta\, e^{\mathrm{i}s\,\xi\cdot X}\Big] \simeq \delta^\alpha{}_\beta - \frac{1}{2}\,c^\alpha{}_{\beta\gamma}\xi^\gamma + O(\xi^2). \tag{4.51}$$

Their infinitesimal transformation rules similarly are

$$\delta\mathcal{A}^i_\mu(\xi) = \partial_\mu u^i(\xi,\omega) - c^i{}_{jk}\,u^j(\xi,\omega)\mathcal{A}^k_\mu(\xi)\,, \tag{4.52}$$

and

$$\delta e^\alpha_\mu(\xi) = -c^\alpha{}_{i\beta}\,u^i(\xi,\omega)\ e^\beta_\mu(\xi). \tag{4.53}$$

In this last expression, the structure constants themselves define representation matrices, $(\mathcal{T}_i)^\alpha{}_\beta = c^\alpha{}_{i\beta}$, of the Lie algebra of H, whose exponentials appear in (4.42).

To build self-interactions for the Goldstone bosons using these tools one combines the covariant quantity, $e^\alpha_\mu = e^\alpha{}_\beta\,\partial_\mu\xi^\beta$ in all possible H-invariant ways. This is simple to do since this quantity transforms very simply under G: $e_\mu \cdot X \to \mathtt{h}\,(e_\mu \cdot X)\,\mathtt{h}^{-1}$. Derivatives of e^α_μ are then included by using the covariant derivative constructed from $\mathcal{A}^i_\mu t_i$:

$$(D_\mu e_\nu)^\alpha = \partial_\mu e^\alpha_\nu + c^\alpha{}_{i\beta}\mathcal{A}^i_\mu\, e^\beta_\nu, \tag{4.54}$$

which transforms in the same way as does e^α_μ: $\delta(D_\mu e_\nu)^\alpha = -c^\alpha{}_{i\beta}u^i(D_\mu e_\nu)^\beta$.

The invariant lagrangian then is $\mathfrak{L}(e_\mu, D_\mu e_\nu, \ldots)$, where the ellipses denote terms involving higher covariant derivatives and the lagrangian is constrained to be globally H invariant:

$$\mathfrak{L}(he_\mu h^{-1}, hD_\mu e_\nu h^{-1}, \ldots) \equiv \mathfrak{L}(e_\mu, D_\mu e_\nu, \ldots)\,. \tag{4.55}$$

[12] This useful combination is called a Maurer-Cartan form [50].

Whenever $\mathfrak{L}$ satisfies (4.55) for constant h, the definitions of $e^{\alpha}{}_{\beta}$ and $\mathcal{A}^{i}_{\alpha}$ ensure it is also *automatically* invariant under global G transformations.

For a Poincaré-invariant system, this leads to the following terms involving the fewest derivatives

$$\mathfrak{L}_{GB} = -\frac{1}{2}\, g_{\alpha\beta}(\xi)\, \partial^{\mu}\xi^{\alpha}\partial_{\mu}\xi^{\beta} + (\text{higher-derivative terms})\,, \tag{4.56}$$

with $g_{\alpha\beta}(\xi) = f_{\gamma\delta}\, e^{\gamma}{}_{\alpha}\, e^{\delta}{}_{\beta}$ where $f_{\alpha\beta}$ is a constant positive-definite matrix that must satisfy

$$f_{\lambda\beta}c^{\lambda}{}_{i\alpha} + f_{\alpha\lambda}c^{\lambda}{}_{i\beta} = 0\,, \tag{4.57}$$

in order for the lagrangian of (4.56) to be G-invariant. In many situations the representation matrices $(\mathcal{T}_i)^{\alpha}{}_{\beta}$ form an irreducible representation, in which case Schur's lemma implies $f_{\alpha\beta}$ must be proportional to the unit matrix $f_{\alpha\beta} = F^2\delta_{\alpha\beta}$ where F is a constant parameter.

The action for the other matter fields is similarly constructed by using $\mathcal{A}^{i}_{\mu}(\xi)$ to build covariant derivatives for the χ^n: $D_\mu\mathcal{X} = \partial_\mu\mathcal{X} - \mathrm{i}\mathcal{A}^i t_i\,\mathcal{X}$. Because the symmetry H is unbroken, these fields must all transform linearly under H: $\mathcal{X} \to h\mathcal{X}$, for some representation matrices, h, of H. This gets promoted to a nonlinearly realized G-transformation because the transformation law for the χ^n is $\mathcal{X} \to \mathtt{h}\,\mathcal{X}$, with $\mathtt{h} = \mathtt{h}\,(\xi, g) \in H$, as defined in (4.47). The covariant derivative $D_\mu\mathcal{X}$ is defined so that it also transforms in the same way as does $\mathcal{X}$ itself, $D_\mu\mathcal{X} \to \mathtt{h}(\xi, g)\, D_\mu\mathcal{X}$, under the nonlinearly realized G-transformations.

With these rules, any old globally H-invariant lagrangian for $\mathcal{X}$ automatically becomes promoted to a G invariant lagrangian once all derivatives are replaced by the ξ-dependent covariant derivatives. This works because global H-invariance of the original lagrangian means it satisfies

$$\mathfrak{L}(he_\mu h^{-1}, h\,\mathcal{X}, h\partial_\mu e_\nu h^{-1}, h\,\partial_\mu\mathcal{X}, \dots) \equiv \mathfrak{L}(e_\mu, \mathcal{X}, \partial_\mu e_\nu, \partial_\mu\mathcal{X}, \dots)\,, \tag{4.58}$$

for any $h \in H$. But the above constructions are designed to ensure that each covariant quantity transforms under G as $e_\mu \to \mathtt{h}\, e_\mu\, \mathtt{h}^{-1}$, $\mathcal{X} \to \mathtt{h}\,\mathcal{X}$, $\mathcal{D}_\mu\mathcal{X} \to \mathtt{h}\,\mathcal{D}_\mu\mathcal{X}$ and so on, so the condition for G invariance

$$\mathfrak{L}(\mathtt{h}e_\mu\mathtt{h}^{-1}, \mathtt{h}\,\mathcal{X}, \mathtt{h}\,D_\mu e_\nu\mathtt{h}^{-1}, \mathtt{h}\,D_\mu\mathcal{X}, \dots) \equiv \mathfrak{L}(e_\mu, \mathcal{X}, D_\mu e_\nu, D_\mu\mathcal{X}, \dots)\,, \tag{4.59}$$

becomes an automatic consequence of (4.58). As shown in the Appendix C.6, this construction of the invariant lagrangian is also unique, given the assumed transformation rules for the fields.

4.2.3 Gauge symmetries

To this point the discussion has been completely aimed at global symmetries, for which the symmetry parameter is (by definition) independent of spacetime position. At this point a brief aside is warranted on how local (or gauge) symmetries are nonlinearly realized within the low-energy Wilsonian EFT, where the parameters ω^a are no longer required to be constants.

The motivation for doing so is because this situation is not hypothetical. The twin constraints of Lorentz invariance and unitarity in quantum mechanics dictate that the couplings of *any* massless spin-one particle must be gauge invariant [51, 54] (see Appendix C.3.3), to

the extent that their couplings to other matter at very low energies (*i.e.* their renormalizable interactions) is only possible if this other matter enjoys some sort of gauge symmetries. By extension, low-energy couplings of massive spin-one particles (whose masses are nonzero but very small compared with other scales) are only possible to matter that enjoys a spontaneously broken gauge symmetry [55, 56, 57, 58, 59, 60]. This framework includes in particular all presently known fundamental[13] massive spin-one particles in nature [61].

Just like for global symmetries, the discussion naturally breaks up into linearly realized and nonlinearly realized symmetries, so each is considered in turn. For simplicity this section restricts itself to abelian symmetries, though the conclusions drawn apply more generally (see Appendix C.5).

At face value it is simple to construct lagrangians invariant under linearly realized gauge symmetries along standard lines. One starts with a lagrangian that is invariant under a global symmetry and promotes the global symmetry to a gauge symmetry by combining all derivatives of the fields (like $\partial_\mu \psi^m$) with a gauge potential, A_μ, to make gauge-covariant derivatives (denoted $D_\mu \psi^m$). For abelian symmetries the gauge potential transforms as $A_\mu \to A_\mu + \partial_\mu \omega$. These covariant derivatives are designed so that $D_\mu \psi^m$ transforms under the position-dependent symmetry in precisely the same way as $\partial_\mu \psi^m$ did when the symmetry parameter was constant.

For example, in the UV version of the toy model, the symmetry transformation is $\phi \to e^{i\omega}\phi$, and so if ω were not a constant then $\partial_\mu \phi \to e^{i\omega}(\partial_\mu \phi + i\partial_\mu \omega\, \phi)$. The corresponding covariant derivative then is $D_\mu \phi = \partial_\mu \phi - iA_\mu \phi$ since this transforms like $D_\mu \phi \to e^{i\omega} D_\mu \phi$ even for spacetime-dependent ω. No additional symmetry restrictions are placed on the lagrangian itself beyond the requirements already imposed by global invariance.

The gauge-invariant version of the theory is then found by replacing $\partial_\mu \psi^m \to D_\mu \psi^m$ everywhere, and supplementing the result with a dependence on derivatives of A_μ, which appear through the gauge-invariant field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\mu A_\nu$. For instance, the gauge-invariant version of the toy model replaces (1.1) by

$$S := -\int \mathrm{d}^4 x \left[\frac{1}{4g^2} F^{\mu\nu} F_{\mu\nu} + D_\mu \phi^* D^\mu \phi + V(\phi^* \phi) \right], \tag{4.60}$$

where $D_\mu \phi = \partial_\mu \phi - iA_\mu \phi$ as before, and only renormalizable interactions are kept (*i.e.* those with couplings having non-negative dimension in powers of mass).

This describes the particles of the toy model interacting with a spin-one particle described by the field A_μ. The field A_μ can be canonically normalized (see Appendix C.3.3) by rescaling $A_\mu \to gA_\mu$. When this is done the covariant derivative becomes $D_\mu \phi = \partial_\mu \phi - igA_\mu \phi$, revealing g to be the gauge coupling whose value controls how strongly A_μ couples to ϕ. If the scalar potential is minimized at $\phi^* \phi = v^2$, expanding about $\phi = v$ shows that the quadratic terms in A_μ become

$$-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - D_\mu \phi^* D^\mu \phi \supset -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - g^2 v^2 A_\mu A^\mu \,, \tag{4.61}$$

and so spontaneous symmetry breaking gives the spin-one particle a mass $M_A^2 = 2g^2 v^2$.

[13] As this book makes clear, fundamental here simply means point-like in the best effective description known to date, and does not exclude the possibility of their being found to have substructure as knowledge improves.

So far so standard. The next step is to ask how a gauge symmetry is manifest in a low-energy EFT in the limit $v \gg M$ and so for which the symmetry is nonlinearly realized.

Explicit vs spontaneous breaking for gauge symmetries

In principle the procedure for gauging a nonlinearly realized symmetry remains the same as before: combine the derivatives of the Goldstone field, $\partial_\mu \xi$, (which is the only one that transforms in the EFT under the spontaneously broken global abelian symmetry) with the gauge potential A_μ to build a covariant derivative. The only difference is that ξ transforms as a shift under the symmetry, $\xi \to \xi + \sqrt{2}\, v\, \omega$, rather than being multiplied by a phase or a matrix. (A field that shifts as ξ does under a gauge transformation is often called a 'Stueckelberg' field [63].) Consequently the required covariant derivative this time is $D_\mu \xi = \partial_\mu \xi - \sqrt{2}\, v A_\mu$, if $A_\mu \to A_\mu + \partial_\mu \omega$. For canonically normalized A_μ the covariant derivative is instead $D_\mu = \partial_\mu - \sqrt{2}\, g v A_\mu$.

Notice in particular that because ω is an arbitrary function it can be used to completely remove ξ by setting it to zero everywhere. This choice is called 'unitary gauge' [64, 65] and in this gauge $D_\mu \xi = -\sqrt{2}\, g v A_\mu$, revealing how ξ can be completely absorbed into the spin-one field A_μ. In this gauge the canonically normalized kinetic terms for ξ and A_μ become

$$-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2}\, D_\mu \xi D^\mu \xi = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - g^2 v^2 A_\mu A^\mu \,, \tag{4.62}$$

revealing the spin-one particle to have mass $M_A^2 = 2 g^2 v^2$, in agreement with the UV theory (*c.f.* Appendix C.3.3). This absorption of ξ to give the spin-one field a mass is the usual Higgs mechanism in action, with ξ providing the missing degrees of freedom required to convert the two spin states of a massless spin-one particle to the three spin states of massive spin one.

What is interesting is that there are two ways to regard the resulting low-energy EFT. The first way is to think of it is the way just described: it is a massless spin-one field coupled to a nonlinearly realized abelian symmetry. The other way to think of it is simply as a generic theory of a massive spin-one vector field, V_μ, coupled to other particles in an arbitrary way with no reference made to gauge invariance at all; that is to say, the gauge symmetry is explicitly broken.

These two ways of thinking lead to precisely the same lagrangian, since the discussion surrounding eq. (4.35) shows that the most general nonlinearly realized lagrangian is built using arbitrary combinations of a vector field V_μ — with no constraints on the lagrangian coming from the abelian symmetry — followed by the replacement $V_\mu \to \partial_\mu \xi$. In the gauge-invariant construction this vector is simply $V_\mu = D_\mu \xi / (\sqrt{2}\, v)$, and so is simply A_μ written in unitary gauge.

The upshot is this: from a low-energy perspective there is operationally no difference between a nonlinearly realized gauge symmetry and the complete absence of a gauge symmetry (or an explicitly broken gauge symmetry).[14] There is no royal road that allows an observer to learn about broken high-energy gauge symmetries using only low-energy meth-

[14] A similar statement applies for nonabelian symmetries [28].

ods. The same is not true for global symmetries because for these the physical Goldstone mode is always present at low energies to bring the news about the UV theory's broken symmetries.

Of course this doesn't mean that there is no utility in sometimes using unitary gauge (and thereby ignoring the symmetries) and sometimes using a more general covariant gauge (for which the Stueckelberg field is kept and the gauge symmetry is nonlinearly realized). Unitary gauge is usually more useful at tree level, since it makes the physical particle spectrum more transparent. Covariant gauges are more convenient when computing loops (or power-counting in general), for reasons that become clear once the massive spin-one propagator is written.

For a one-parameter family of covariant gauges [66] the massive spin-one propagator has the form

$$G_{\mu\nu}(x-y) = -\mathrm{i}\int \frac{\mathrm{d}^4 p}{(2\pi)^4}\ \frac{1}{p^2+m^2-\mathrm{i}\epsilon}\left[\eta_{\mu\nu} + (\zeta-1)\,\frac{p_\mu p_\nu}{p^2+\zeta m^2}\right] e^{\mathrm{i}p\cdot(x-y)}\,, \tag{4.63}$$

where the real parameter ζ labels the gauge choice. Popular gauge choices within this class are Feynman gauge ($\zeta = 1$), Landau gauge ($\zeta = 0$) and unitary gauge, which corresponds to the limit $\zeta \to \infty$. What is inconvenient about loops in unitary gauge is the propagator's large-momentum limit, since it does not fall off quadratically with momentum as the components of p_μ get large. Naive use of unitary gauge in the power-counting estimates of §3 leads to completely misleading results.[15]

Unitarity bound: the gauged toy model

How can describing a spin-one particle without using gauge invariance be consistent with the statement made at the beginning of this section that relativity and unitarity in quantum mechanics require a massless spin-one particle to be associated with a gauge symmetry? To explore this it is again instructive to ask the question for the toy model of §1; or rather for the gauged version of the toy model for which a gauge potential A_μ 'gauges' the toy model's $U(1)$ symmetry (*i.e* promotes it from a global symmetry to a local one). The renormalizable lagrangian for the UV version of this model is given by (4.60) instead of (1.1), though with scalar potential still given by (1.2).

For nonzero v both the scalar ϕ_R (or χ) and the spin-one field A_μ acquire a mass, with $m_R^2 = \lambda v^2$ and $m_A^2 = 2g^2v^2$. In the regime where the gauge coupling, g, and the scalar self-coupling, λ, are both small (to justify semiclassical methods) but with $g^2 \ll \lambda$ these satisfy $m_A^2 \ll m_R^2$. In this case the low-energy spectrum for energies below m_R consists only of the massive spin-one particle, whose lagrangian is not constrained by gauge invariance apart from the observation that it can be built using arbitrary powers of the invariant field $V_\mu = \partial_\mu\xi - \sqrt{2}\, gvA_\mu$.

With these choices (and in a covariant gauge[16]) the power-counting arguments of §3 go

[15] This is another example where if you use the wrong variables you will be sorry. In principle once everything is included all gauges give the same answer to physical (and so gauge-invariant) quantities. But results that are manifest at every step in covariant gauges emerge in unitary gauge only after obscure cancellations.

[16] For aficianados: including 'ghosts,' though these do not play an important role for the present purposes.

through in the Wilsonian theory of V_μ interactions with only minor modifications due to the presence of the gauge field A_μ. Writing the basic action in the form

$$\mathcal{L}_W = \mathfrak{f}^4 \sum_n c_n \, O\left(\frac{\partial}{M}, \frac{\xi}{v}, \frac{A}{v_A}\right), \tag{4.64}$$

for dimensionless couplings c_n, and repeating the dimensional power-counting arguments of §3.2.3 leads to the following minor generalization of (3.57),

$$\mathcal{A}_{\mathcal{E}_\xi \mathcal{E}_A}(q) \sim \frac{q^2 \mathfrak{f}^4}{M^2} \left(\frac{1}{v}\right)^{\mathcal{E}_\xi} \left(\frac{1}{v_A}\right)^{\mathcal{E}_A} \left(\frac{Mq}{4\pi \mathfrak{f}^2}\right)^{2\mathcal{L}} \prod_n \left[c_n \left(\frac{q}{M}\right)^{d_n - 2}\right]^{\mathcal{V}_n}, \tag{4.65}$$

as an estimate for a graph with $\mathcal{E}_\xi$ and $\mathcal{E}_A$ external ξ and A_μ legs, $\mathcal{L}$ loops and $\mathcal{V}_n$ vertices involving d_n derivatives each. As usual q denotes here the generic size of the external energies flowing through the graph and assumes this is the only scale relevant when making a dimensional estimate of the result's size.

From §3.2.3 it is known that using the choices $\mathfrak{f}^2 = m_R v$ and $M = m_R$ in (4.64) correctly captures the dependence of $\mathcal{L}_W$ on v and m_R, at least for the ξ-dependent terms. With these choices choosing $v_A = M/g = m_R/g$ in (4.64) also ensures that both ∂_μ and gA_μ enter with the same dimensional factor of $1/M$ in $\mathfrak{L}_W$, thereby ensuring that all appearances of the covariant derivative appear in the combination

$$\frac{D_\mu \xi}{vM} = \frac{\partial_\mu \xi - \sqrt{2}\, g v A_\mu}{vM} = \frac{\partial_\mu \xi}{vM} - \sqrt{2}\, \frac{A_\mu}{v_A}, \tag{4.66}$$

appear consistent with the assumption made in (4.64).

What remains is to track the remaining factors of the gauge coupling, g, and to see whether or not there are any systematic powers of m_R and v hidden within the dimensionless coefficients, c_n. The above choices ensure there are none in any terms that explicitly involve the field ξ. What they do not get right are the terms that involve A_μ without ξ, such as terms built using the covariant field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

In particular, using the above rules in (4.64) would predict such terms appear in $\mathcal{L}_W$ proportional to

$$c_n \, \mathfrak{f}^4 \, O_n\left(\frac{\partial}{M}, \frac{F_{\mu\nu}}{M v_A}\right) = c_n \, m_R^2 v^2 \, O_n\left(\frac{\partial}{m_R}, \frac{g F_{\mu\nu}}{m_R^2}\right). \tag{4.67}$$

This gets two things wrong when compared with what is obtained by integrating out the heavy χ scalar from the UV theory using (4.60). First it predicts a kinetic Maxwell action for the gauge field of size

$$\frac{\mathfrak{f}^4}{M^2 v_A^2} \, F_{\mu\nu} F^{\mu\nu} = \frac{g^2 v^2}{M^2} \, F_{\mu\nu} F^{\mu\nu}, \tag{4.68}$$

rather than the standard Maxwell action inherited from (4.60). This shows that the estimate (4.65) misses a factor of $g^2 v^2 / M^2$ from each internal gauge field line, requiring it to be corrected by the factor

$$\text{Correction factor} = \left(\frac{g^2 v^2}{M^2}\right)^{\mathcal{I}_A} = \left(\frac{gv}{M}\right)^{-\mathcal{E}_A + \sum_n f_n^A \mathcal{V}_n}, \tag{4.69}$$

where $\mathcal{I}_A$ is the number of internal gauge-field lines in the graph, and f_n^A is the number of gauge lines that meet at the vertex labelled by interaction 'n'. This expression uses 'conservation of ends' for gauge field lines, $\mathcal{E}_A + 2\mathcal{I}_A = \sum_n f_n^A \mathcal{V}_n$. This correction factor says that each internal gauge-boson line brings an additional suppression by a power of $g^2v^2/M^2 \simeq m_A^2/m_R^2 = g^2/\lambda$, which is small because of the assumption that the spin-one particle is light enough to be in the low-energy theory.

Similarly, integrating out the χ field in the full theory also gives higher powers of $F_{\mu\nu}$ (and its derivatives) that come with a factor of g for each $F_{\mu\nu}$ and with dimensions set by powers of m_R, as in

$$h_n\, m_R^4\, \mathcal{O}\left(\frac{\partial}{m_R}\,,\, \frac{gF}{m_R^2}\right), \tag{4.70}$$

where h_n is a pure number (that contains a factor of $1/(16\pi^2)$ for each loop in the full theory required to generate the operator in question). Comparing this with the (4.67) shows that the dimensionless couplings c_n and h_n are related by

$$c_n = h_n \left(\frac{m_R^4}{\hbar^4}\right) = h_n \left(\frac{m_R^2}{v^2}\right) = \lambda h_n\,, \tag{4.71}$$

(where the last equality uses $m_R^2 = \lambda v^2$), but only for those interactions that involve only $F_{\mu\nu}$ and not ξ.

Physically, interactions in $\mathfrak{L}_W$ that are gauge-invariant and independent of ξ involve only the transverse polarizations of the spin-one particle, while ξ itself plays the role of the massive spin-one particle's longitudinal spin state. The above estimates show that (when $m_R^2 \ll v^2$) the interactions of the transverse states are over-estimated at low-energies when using (4.64) and (4.65), though these estimates do get the size of the contribution of the longitudinal state right.

Now comes the main point: concentrating exclusively on the interactions of the longitudinal state, the dominant size of an $\mathcal{E}$-point amplitude is obtained from the estimate (4.65) using only ξ external and internal lines, leading to the estimate (3.58) (or, equivalently, the estimate (3.57)) for a graph with $\mathcal{E}$ external legs, $\mathcal{L}$ loops and $\mathcal{V}_n$ vertices, each of which involves d_n derivatives:

$$\mathcal{A}_\mathcal{E}(q) \sim q^2 v^2 \left(\frac{1}{v}\right)^\mathcal{E} \left(\frac{q}{4\pi v}\right)^{2\mathcal{L}} \prod_n \left[\left(\frac{q}{m_R}\right)^{(d_n-2)}\right]^{\mathcal{V}_n}. \tag{4.72}$$

Not surprisingly this says that the low-energy expansion is the key to the validity of the loop expansion in the Wilsonian theory. For $q \ll m_R \ll 4\pi v$ all graphs are suppressed and the expansion is controlled. When $q \sim m_R \ll 4\pi v$ multiple insertions of vertices at fixed loop order can become unsuppressed, but need not represent loss of control provided there are fixed numbers of graphs possible at any loop order. But the low-energy expansion underlying the Wilsonian lagrangian necessarily requires

$$q \ll 4\pi v \sim \frac{4\pi m_A}{g}\,. \tag{4.73}$$

This kind of restriction for an EFT is often called a 'unitarity bound' [67], because it is often identified by asking when a cross section computed within the low-energy theory

becomes inconsistent with the energy dependence required at high energy by unitarity [68]. It would of course be misleading to regard inconsistency as a *bona fide* loss of unitarity in the low-energy theory (which nobody does), since it is hard to see how unitarity can be lost if the hamiltonian remains hermitian (as is typically the case). What is really failing is the validity of the low-energy expansion used to infer the perturbative cross-section in the low-energy regime, and this failure is even more systematically revealed through power-counting estimates like (4.72).

It is now possible to circle back to the question that started this section: how can the necessity for gauge invariance for massless spin-one particles be consistent with the observation that nonlinearly realized gauge symmetry is operationally the same as explicit breaking of the gauge symmetry in the EFT for massive spin-one states?

As is seen above, the low-energy description of a massive spin-one particle (either without gauge invariance or with a nonlinearly-realized gauge symmetry) always breaks down at energies of order $q \sim 4\pi m_A/g$, by which point some new UV description must necessarily intervene. (Usually what intervenes is a description involving linearly realized gauge invariance.) Although a non-gauge invariant description of a massive spin-one particle can make sense within an EFT, this description cannot work up to energies that are hierarchically large compared to its mass, and cannot work at all for nonzero g if the particle is massless.[17] Similar restrictions do *not* apply for linearly-realized gauge symmetries, since these can be renormalizable and so be valid up to energies much higher than any of those that appear explicitly in the low-energy theory itself (indeed this is one way that they can be derived [69]).

4.3 Anomaly matching ♠

The previous sections discuss symmetries as if their existence is established by showing the invariance of the *classical* action and so ignores the possibility that classical symmetries might not survive quantization. Traditionally when a classical symmetry fails to survive quantization it is known as an 'anomalous' symmetry [70, 71, 72, 73, 75].

This kind of separate treatment of the classical action and its quantum corrections is a bit too old-school within an EFT framework, because what one naively calls the 'classical' action is really better understood as the Wilsonian action obtained by integrating out higher-energy degrees of freedom. As this section now argues, it is more useful to organize one's thinking in terms of the scales involved than to divide the world artificially into a quantum and classical part. From this point of view an anomalous symmetry is a particu-

[17] A loophole to this argument arises for massive abelian gauge bosons, since for these interactions for the would-be Goldstone field ξ need not exist at all (in which case the massive Stueckelberg description can be renormalizable). (This is only an option for abelian bosons because – as §4.2.2 shows – non-renormalizable self-interactions are compulsory for nonabelian Goldstone bosons.) As the toy-model example shows, the absence of interactions for ξ is a strong condition even for abelian theories, and is not generic in any particular UV completion.

lar instance of a transformation that is simply not a symmetry, but under which the action transforms in a specific way.

4.3.1 Anomalies♡

One way to characterize the failure of a classical symmetry at the quantum level is if the system's 1PI action is not invariant under the transformations in question, even though the classical (or Wilson) action is. To see how this might happen recall the relation between the 1PI and classical actions, given by (2.19), reproduced here for convenience of reference:

$$\exp\Big\{\mathrm{i}\Gamma[\varphi\,]\Big\} = \int \mathcal{D}\hat{\phi}\ \exp\left\{\mathrm{i}S\,[\varphi+\hat{\phi}\,] + \mathrm{i}\int \mathrm{d}^4x\,\hat{\phi}^a J_a(\varphi)\right\}, \tag{4.74}$$

where $J_a(\varphi) = -\delta\Gamma/\delta\varphi^a$.

In this expression consider transforming the argument of $\Gamma[\varphi]$ under a symmetry transformation, which for simplicity's sake[18] is taken to act linearly on the fields: $\varphi^a \to (U\varphi)^a = U^a{}_b\,\varphi^b$. This gives

$$\begin{aligned}\exp\Big\{\mathrm{i}\Gamma[U\varphi\,]\Big\} &= \int \mathcal{D}\hat{\phi}\ \exp\left\{\mathrm{i}S\,[U\varphi+\hat{\phi}\,] + \mathrm{i}\int \mathrm{d}^4x\,\hat{\phi}^a J_a\right\} \\ &= \int \mathcal{D}\hat{\phi}_u\,\mathfrak{J}(\varphi,\hat{\phi}_u)\ \exp\left\{\mathrm{i}S\,[\varphi+\hat{\phi}_u\,] + \mathrm{i}\int \mathrm{d}^4x\,\hat{\phi}^a_u\mathcal{J}_a\right\},\end{aligned} \tag{4.75}$$

where the second line performs the change of integration variable $\hat{\phi} \to \hat{\phi}_u$ where $\hat{\phi} = U\hat{\phi}_u$ (for which $\mathfrak{J}(\varphi,\hat{\varphi}_u)$ is the Jacobian — more about which below) and uses the invariance of the classical action to write $S[U(\varphi+\hat{\phi}_u)] = S[\varphi+\hat{\phi}_u]$. Also used is the definition of the current, $J_a(\varphi) = -\delta\Gamma[\varphi]/\delta\varphi^a$, which implies $J_a(U\varphi) = -(\delta\Gamma[\psi]/\delta\psi^a)_{\psi=U\varphi}$ while $\mathcal{J}_a := -\delta\Gamma[U\varphi]/\delta\varphi^a = U^b{}_a J_b(U\varphi)$.

This manipulation shows that it is consistent to have $\Gamma[U\varphi] = \Gamma[\varphi]$ if both the classical action is invariant *and* the path-integral measure is invariant — *i.e.* the Jacobian is trivial: $\mathfrak{J} = 1$. One way to think about anomalies is they are the situation where there is an obstruction to constructing this type of invariant measure for the path integral [76]. Although it goes beyond the scope of this book to derive the conditions for anomalies in great detail (see the bibliography, §D.2, for further reading) suffice it say that obstructions arise when a system is 'chiral' in the sense that its interactions treat left- and right-handed particles differently. In four dimensions this boils down to systems with chiral fermions.

A concrete way to identify when there is an anomaly is to evaluate the matrix elements of the conserved Noether current, regarded as a quantum operator. Recall that for each classical symmetry Noether's theorem ensures the existence of a current, J^μ, that is locally conserved inasmuch as the field equations imply $\partial_\mu J^\mu = 0$. Explicit calculations of matrix elements like $\langle f|J^\mu|\Omega\rangle$, where $|\Omega\rangle$ is the ground state and $|f\rangle = |A(k),A(q)\rangle$ is a state involving two spin-one particles, show that the matrix element $\langle f|\partial_\mu J^\mu|\Omega\rangle$ cannot be zero, so local conservation fails as an operator statement.

[18] A similar argument for nonlinearly realized symmetries couples the current to a combination $\sigma^a(\varphi,\hat{\phi})$ that transforms more covariantly under field redefinitions. The functional form of the transformation rule can also evolve with scale, and so differ between the microscopic fields ϕ^a and $\varphi^a = \langle\phi^a\rangle$ (see also Exercise 2.7).

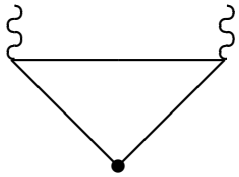


Fig. 4.3 The triangle graph that is responsible for anomalous symmetries (in four spacetime dimensions). The dot represents the operator J^μ and the external lines represent gauge bosons in the matrix element $\langle gg|J^\mu|\Omega\rangle$, where $|\Omega\rangle$ is the ground state.

Evaluating the graph of Fig. 4.3 gives the following result for the conservation of the operator current [70, 71, 72, 73][19]

$$\partial_\lambda J^\lambda_a = A_{abc}\,\frac{g_a g_b}{64\pi^2}\,\epsilon^{\mu\nu\alpha\beta}F^b_{\mu\nu}F^c_{\alpha\beta} \tag{4.76}$$

with $F^b_{\mu\nu}$ the field strength corresponding to a gauge symmetry generator T_b, and g_b its associated gauge coupling. The quantities A_{abc} are called *anomaly coefficients*, and they are given in terms of the symmetry generators (acting on left-handed spin-half fields) by

$$A_{abc} = \mathrm{tr}\left(T_a\left\{T_b, T_c\right\}\right) \tag{4.77}$$

where the curly brackets denote the anticommutator, $\left\{T_b, T_c\right\} := T_bT_c + T_cT_b$. As defined A_{abc} is completely symmetric under the interchange of any pair of indices, and it is real because the generators T_a are hermitian. The trace is over the reducible representation of the symmetry acting on the complete set of left-handed fermions.

A classical gauge symmetry survives quantization — and is said to be 'anomaly free' — if $A_{abc} = 0$ for all T_a, T_b and T_c in the symmetry's Lie algebra. Because gauge invariance ultimately is required by the interplay between Lorentz invariance and unitarity, gauge symmetries must be anomaly-free to be consistent. (In §9.1.2 it is shown that anomalies cancel in an interesting way for the symmetry group $SU_c(3) \times SU_L(2) \times U_Y(1)$ using a single 'generation' of fermion content from the Standard Model.)

An important sufficient condition for the absence of anomalies is simple to state. Any symmetry group must be anomaly-free if its representation on left-handed fermions is *real* (or pseudoreal). A representation is real if its group-representation matrices $\exp(\mathrm{i}\omega^a T_a)$ are real and so the matrices T_a are imaginary. It is pseudo-real if the matrices T_a are imaginary up to a similarity transformation: $T^*_a = -S\,T_aS^{-1}$ for some invertible matrix S.

To see why pseudo-reality ensures freedom from anomalies, notice that because the generators T_a are in general hermitian, it follows that $T^T_a = T^*_a$. Because the trace of a

[19] Non-invariance of $\Gamma[\phi]$ is related to the failure of the Noether current to be conserved, as can be seen by performing a *local* symmetry transformation, under which (for a global symmetry) the classical action (4.81) is not invariant. Evaluating explicitly how it transforms shows: $\delta\Omega_W = -\int \mathrm{d}^4x\,\partial_\mu\omega^a\,J^\mu_a$ which after integration by parts gives $\delta\Omega_W = \int \mathrm{d}^4x\,\omega^a\,\partial_\mu J^\mu_a$.

matrix equals the trace of its transpose it follows that

$$\begin{aligned} A_{abc} &= \mathrm{tr}\,[(T_a\{T_b, T_c\})^T] = \mathrm{tr}\,(\{T_c^T, T_b^T\}T_a^T) \\ &= \mathrm{tr}\,(\{T_c^*, T_b^*\}T_a^*) = -\,\mathrm{tr}\,(S\{T_c, T_b\}T_a S^{-1}) \\ &= -\mathrm{tr}\,(\{T_c, T_b\}T_a) = -A_{abc}\,, \end{aligned} \tag{4.78}$$

and so $A_{abc} = 0$.

A special case of this last result shows why only chiral symmetries are anomalous. To see why, imagine that fermion number is conserved (so that fermions and antifermions are distinguishable) and further assume the symmetry is *not* chiral, in that left- and right-handed fermions (as opposed to antifermions) transform in the same representation, t_a say, of the group. In this case the group generators acting on *all* the left-handed fermions — for fermions *and* antifermions — can be written in the block-diagonal form

$$T_a = \begin{pmatrix} t_a & 0 \\ 0 & -t_a^* \end{pmatrix} \tag{4.79}$$

where the upper-left block gives the action on fermions and the lower-right block on antifermions. This representation is manifestly pseudoreal since $T_a^* = -S\,T_a S^{-1}$, where $S = \tau_1 \otimes I$ — with τ_1 the first Pauli matrix — is the matrix that swaps the upper-left and lower-right blocks. It follows that any symmetry that is left-right symmetric in this way must also be anomaly-free. This is in particular why anomalies are not an issue for either Quantum Electrodynamics or Quantum Chromodynamics.

An important property of the definition of an anomaly is the inability to remove it (and so to restore the symmetry) simply by appropriately adding non-invariant local counter-terms to the lagrangian density. This in itself means that anomalies must have their origins in the low-energy part of the theory, rather than the high-energy part. (After all, the EFT program argues that *any* high-energy physics can be described by some choice for local interactions within an effective theory.) It is for this reason that anomalies are relevant when setting up the Wilsonian description of the low-energy sector [74].

The observation that anomalies cannot (by definition) be canceled by local counter-terms also reveals the difference between an honest-to-God anomaly and just regularizing in a silly way. Any damage done by using an ill-conceived regularization procedure — such as one that does not preserve a system's symmetries — can be undone by renormalizing parameters appropriately, but this is *not* possible if a symmetry is anomalous. Many simple regularization schemes (like explicit cut-offs in momentum integrals or point-splitting techniques) break symmetries (like Lorentz invariance or gauge invariance), but their use does not mean that the physics being described must break these symmetries. This is why it can make sense (though is not normally convenient) to define effective Wilson actions with cutoffs, even for Lorentz-invariant systems or systems with gauge symmetries. The implicit choice made in such cases is to undo any regulatory damage by appropriately renormalizing the theory to restore these symmetries.

A sufficient condition for there to exist a renormalization scheme that preserves a symmetry is the existence of a regularization scheme that explicitly preserves it (such as is often true with dimensional regularization, for example). Since the difference between any

two regularization schemes lies purely at high energies, it must be captured by some choice of effective local couplings. But while this shows that invariant regularizations are not possible for anomalous symmetries,[20] the absence of a known invariant regularization does not necessarily imply the existence of an anomaly.

4.3.2 Anomalies and EFTs

This path-integral way of formulating things shows how the classical/quantum split is more subtle when phrased in terms of the Wilsonian effective action. Given a hierarchy of scales with light, l, and heavy, $\mathfrak{h}$, degrees of freedom, the definition (2.58) of the Wilsonian action (reproduced here)

$$\exp\Big\{\mathrm{i}S_{\scriptscriptstyle W}[l]\Big\} := \int \mathcal{D}\mathfrak{h}\ \exp\Big\{\mathrm{i}S\,[l+\mathfrak{h}]\Big\}\,, \tag{4.80}$$

shows that a nontrivial Jacobian potentially moves the high-energy part of the anomaly from the measure into the (Wilsonian) action itself.

As ever, the basic symmetry statement is — at *any* scale — that the transformation properties of the Wilsonian action are whatever they must be to reproduce the correct transformation properties of quantities like the 1PI action for the full microscopic theory. This means invariance at all scales for honest-to-God symmetries, and it means reproducing the nontrivial transformation properties of $\delta\Gamma[\varphi]$, for broken symmetries. (For anomalous symmetries, this condition that the Wilsonian action reproduce the anomalies of the full theory is called 'anomaly matching', and its power lies in the fact that for anomalies $\delta\Gamma[\varphi]$ takes a restricted form [78, 80].)

Gauge symmetries

For these purposes it is important to distinguish between local (or gauge) symmetries and global symmetries. Linearly realized gauge symmetries are central to the consistency of the coupling of light spin-one particles, since their interactions are only Lorentz invariant and unitary if they are also invariant under local gauge transformations. Consequently gauge symmetries cannot be anomalous, and this must be true for the Wilsonian theory at any energy for which one cares to ask the question.

Absence of anomalies usually means invariance for both $S_{\scriptscriptstyle W}$ and the Jacobian $\mathfrak{J}$, and this is how things turn out to transpire for the Standard Model (see §9.1.2). This need not necessarily be so, however, since in principle both $\delta S_{\scriptscriptstyle W}$ and $\mathfrak{J}-1$ can be nonzero, so long as the total combination $\mathcal{D}\hat{\phi}\, e^{\mathrm{i}S_{\scriptscriptstyle W}[\varphi+\hat{\phi}]}$ is invariant. This is not just an academic observation because this is the way gauge symmetries in many string theories (see §10.3) turn out to be anomaly free. In this context the cancellation between the variation of the action and the Jacobian is called 'Green-Schwarz' anomaly cancellation [77].

What makes a cancellation between the variation of $\mathcal{D}\hat{\phi}$ and $e^{\mathrm{i}S_{\scriptscriptstyle W}}$ tricky is the fact that

[20] Sometimes regularizations superficially appear to preserve an anomalous symmetry (such as the anomaly in Weyl invariance when regularized in $D \neq 4$ dimensions), but when this happens the regularization scheme introduces new light degrees of freedom (such as the $D-4$ components of D-dimensional tensor fields) [75].

$S_W = \int \mathrm{d}^4x\, \mathfrak{L}_W$ is local. In fact, the locality of the Wilson action means the idea that having terms in S_W cancel an anomaly (regarded as the variation of $\mathcal{D}\hat{\phi}$) needs some clarification. This is because — as stated explicitly above — an anomaly is *defined* as a nontrivial transformation of $\mathcal{D}\hat{\phi}$ that cannot be removed by adding local counter-terms to the action.

The main issue here is semantic. Strictly speaking, an 'anomalous' symmetry is not really anomalous if it can be cancelled by terms in S_W, as in Green-Schwarz anomaly cancellation. It is not anomalous precisely because anomalies are defined modulo the variation of local terms in the action. When one speaks of Green-Schwarz cancellations the anomalies in question arise from a particular sector of the theory, usually chiral fermions. The corresponding symmetries are anomalous in the sense that they cannot be cancelled by local counter-terms *purely within this sector*. Green-Schwarz anomaly cancellation becomes possible once the rest of the fields from other sectors are also included.

An example might be helpful here. Consider therefore Quantum Electrodynamics (QED) in the limit of vanishing fermion mass. In this case the fields involved are the electromagnetic potential, A_μ, and the fermion's spinor field, ψ, and the leading, renormalizable, terms in the (Wilsonian) lagrangian are

$$\mathfrak{L}_W = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} - \overline{\psi}\not{D}\psi \,. \tag{4.81}$$

Here $\not{D}\psi = \gamma^\mu D_\mu\psi = \gamma^\mu(\partial_\mu - iqA_\mu)\psi$, where q is the fermion's charge ($q = -e$ for an electron, say) and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

This lagrangian enjoys a classical $U_V(1) \times U_A(1)$ symmetry where $U_V(1)$ is the electromagnetic gauge symmetry: $\delta A_\mu = \partial_\mu\zeta$ with $\delta\psi = \mathrm{i}q\zeta\,\psi$, where $\zeta(x)$ is an arbitrary infinitesimal real local symmetry parameter. Unlike the gauge symmetry, the global $U_A(1)$ symmetry — $\delta\psi = \mathrm{i}\omega\,\gamma_5\psi$ for constant, real infinitesimal symmetry parameter ω — is only present due to the absence of a mass term.

The axial symmetry in this theory proves to have an anomaly[21] under which the 1PI action transforms as

$$\delta\Gamma = \frac{\omega\, q^2}{16\pi^2} \int \mathrm{d}^4x\, \epsilon^{\mu\nu\lambda\rho} F_{\mu\nu}F_{\lambda\rho} \,, \tag{4.82}$$

where indeed the right-hand side cannot be written as the variation, δS, of some local functional of A_μ and ψ. This is not inconsistent with (4.82) because — unlike the Wilson action — the Γ that satisfies (4.82) is not local.

But this anomaly *can* be the variation of a local action (and so amenable to Green-Schwarz anomaly cancellation) once other fields are added. In particular, adding a real scalar field, ϕ, transforming inhomogeneously under $U_A(1)$ as a Goldstone field, $\delta\phi = \omega$, allows (4.82) to be cancelled by a contribution to the Wilson action of the form

$$\mathfrak{L}_{GS} = -\frac{q^2}{16\pi^2} \int \mathrm{d}^4x\, \phi\, \epsilon^{\mu\nu\lambda\rho} F_{\mu\nu}F_{\lambda\rho} \,. \tag{4.83}$$

[21] Strictly speaking, there is an anomaly in the full $U_V(1)\times U_A(1)$ symmetry, but precisely which factor is anomalous can be chosen by adding appropriate counterterms. Requiring Lorentz invariance and unitarity precludes letting the $U_V(1)$ factor from being anomalous and so forces the anomaly onto the axial transformation, $U_A(1)$.

Global symmetries and anomaly matching

Anomaly matching enters for global symmetries, since these can be anomalous within a consistent theory. The presence of anomalies for global symmetries can make the difference between having a theory agree with experiment or not.

A famous practical example of this arises in the Quantum Chromodynamics (QCD) (discussed in more detail in §8), wherein anomalies prove to be crucial for describing the decay rate for π^0 mesons. Unlike most mesons, π^0 mesons are seen to decay electromagnetically through the decay into two photons, $\pi^0 \to \gamma\gamma$. Its decays are well-described by an interaction term involving pions and photons of the form

$$\mathfrak{L}_{\text{decay}} = \frac{e^2}{32\pi^2 F_\pi}\pi^0\,\epsilon^{\mu\nu\lambda\rho}F_{\mu\nu}F_{\lambda\rho}\,, \tag{4.84}$$

where e is the electromagnetic coupling and $F_\pi = 92$ MeV is a parameter discussed in some detail in §8.

The interaction (4.84) was a puzzle in QCD before the role played by anomalies was appreciated [81, 82, 83, 84, 85, 86]. This is because in QCD the π^0 is understood to be a bound state consisting of a quark-antiquark combination, where quarks (and their antiparticles) are bound together by the strong interactions. The quarks involved in the π^0 are the up and down quarks, each of which comes in $N_c = 3$ copies (or colours) and whose charges are respectively $q_u = \frac{2}{3}\,e$ and $q_d = -\frac{1}{3}\,e$. The strong force in QCD couples to colour in much the way that electromagnetism couples to electric charge in QED, and this binds the quarks together into mesons with binding energies of order $4\pi F_\pi \sim 1$ GeV or so. In a Wilsonian picture the EFT appropriate at energies higher than this is built using the quarks, while at energies well below $4\pi F_\pi \sim 1$ GeV the effective action instead directly involves the bound-states like π^0 (for more details see §8).

The puzzle arises because the π^0 proves to be a Goldstone boson for a global symmetry, $\delta\pi^0 = \omega F_\pi$, of the strong and electromagnetic interactions, but this seems at first sight to be inconsistent with its appearing undifferentiated in the sub-GeV EFT in a term like (4.84). Instead of being invariant, eq. (4.84) predicts

$$\delta\mathfrak{L}_{\text{decay}} = \frac{\omega\, e^2}{32\pi^2}\,\epsilon^{\mu\nu\lambda\rho}F_{\mu\nu}F_{\lambda\rho}\,. \tag{4.85}$$

The resolution of the puzzle lies in the observation that QCD predicts there is an anomaly in the underlying symmetry for which π^0 is a Goldstone boson. It is the anomaly that allows terms like (4.84), and it is anomaly matching that predicts the size of its coefficient.

To see how this works, it is convenient to write the action on the u and d quarks of the symmetry for which π^0 is the Goldstone boson action as

$$\delta\begin{pmatrix} u \\ d \end{pmatrix} = \mathrm{i}T_A\gamma_5\begin{pmatrix} u \\ d \end{pmatrix} \quad \text{with} \quad T_A = \begin{pmatrix} \frac{1}{2} & \\ & -\frac{1}{2} \end{pmatrix}. \tag{4.86}$$

In this same notation the electric charge of these quarks has the form

$$Q_{\text{em}} = \begin{pmatrix} \frac{2}{3}\,e & \\ & -\frac{1}{3}\,e \end{pmatrix}. \tag{4.87}$$

Since (4.86) is an axial symmetry, it has an anomaly of the form given in (4.82), which when summed over all the colours of the two types of quarks gives

$$\delta\Gamma = \frac{\omega\,\mathcal{A}}{16\pi^2}\int \mathrm{d}^4x\;\epsilon^{\mu\nu\lambda\rho}F_{\mu\nu}F_{\lambda\rho}\,, \tag{4.88}$$

with anomaly coefficient that counts the number of quarks, weighted by their electric charges

$$\mathcal{A} = \mathrm{tr}\left[T_A Q_{\rm em}^2\right] = \frac{N_{\rm c}}{2}\left[\left(\frac{2}{3}\right)^2 - \left(-\frac{1}{3}\right)^2\right] = \frac{N_{\rm c}}{6}\,. \tag{4.89}$$

The success of (4.84) in describing π^0 decays provides one of the experimental confirmations that $N_c = 3$ is the number of colours in QCD.

This success is a special case of anomaly matching for a larger group of approximate global symmetries in low-energy QCD, which includes an entire $U_L(3)\times U_R(3)$ invariance, associated with separate unitary rotations amongst the left- and right-handed parts of the three lightest quarks: u, d and s. Some of the broader implications of these symmetries are described in §8.

In particular, all evidence indicates that the QCD vacuum spontaneously breaks the axial combination of these symmetries, giving rise to 8 pseudo-Goldstone bosons.[22] These symmetries experience several anomalies when combined with various Standard Model gauge symmetries, and their existence implies the existence in the low-energy meson EFT of a specific kind of self-interaction amongst the 8 would-be Goldstone particles. It goes beyond the scope of the book to work out this anomaly-matching lagrangian — called a *Wess-Zumino* action — for the entire anomalous action, but the leading term it generates once it is expanded in powers of the 3×3 hermitian, traceless meson field, $\mathcal{M}$, has the form

$$\mathcal{L}_{WZW} = \frac{N_{\rm c}}{240\pi^2 F_\pi^5}\epsilon^{\mu\nu\lambda\rho}\,\mathrm{tr}\left[\mathcal{M}\partial_\mu\mathcal{M}\partial_\nu\mathcal{M}\partial_\lambda\mathcal{M}\partial_\rho\mathcal{M}\right] + \cdots \tag{4.90}$$

where the ellipses represent terms involving more powers of $\mathcal{M}$ and $N_{\rm c} = 3$ here denotes the number of quark colours.

As above, the coefficient is fixed by demanding that its transformation under $SU_L(3)\times SU_R(3)$ reproduces the anomalies of the underlying quarks, and the resulting value is successful in describing low-energy meson properties.

Anomaly matching can also provide a powerful constraint for theories where it is different species of chiral fermions that contribute to the anomalies at high energies and in the low-energy theory. Particularly interesting models of this form are those where chiral elementary fermions get bound into composite fermions that are also chiral. Such theories arise when contemplating whether or not quarks or leptons might be built from smaller constituents, much as protons and neutrons (once considered to be fundamental) are built from up and down quarks.

The physics involved in such models is often chiral because the puzzle such theories raise is why the bound-state masses (*i.e.* the mass of the ordinary quark or lepton) should

[22] There are only 8 rather than the 9 Goldstone bosons expected for $U_A(3)$ because the overall rotation of all three quarks by a common axial phase is also anomalous and so turns out is broken by the strong interactions.

be so much smaller than the typical energy $E \sim 1/\ell$ associated with the size of the bound object. Experimental searches for compositeness already tell us that if quarks or leptons are composite then the size of the associated bound state must be extremely small, $m \ll 1/\ell$. Chiral theories can help with this because chiral symmetries can allow such theories to have states whose binding energies are much smaller than their size.

For any such model the total pattern of anomalies carried by the constituent fermions at high energies must also be reflected in the spectrum of particles at lower energies, either by having the composite fermions produce the same anomalies or by having composite Goldstone bosons arise (much like the π^0 meson in the QCD example described above). See Exercise 4.6 for a more explicit example of anomaly matching for with composite fermions.

4.4 Summary

Symmetries play a central role in modern physics and effective theories are no different in this regard. This chapter opens with a section that recaps the various roles that symmetries play in quantum mechanics and in quantum field theory. The main new ingredient that the locality of quantum field theory introduces is the possibility that symmetries can be spontaneously broken: the ground state might not be invariant under some of the symmetries of the action (or equations of motion).

If the symmetry that breaks spontaneously is both a continuous symmetry and a global symmetry then spontaneous breaking requires the existence of gapless (or, in a relativistic context, massless) Goldstone states. This makes them card-carrying members of the low-energy sector, whose properties are largely dictated on symmetry grounds.

The main message of this chapter is that any symmetry properties of the full UV theory must also be reflected in any Wilsonian description of its low-energy sector. Much of the discussion is devoted to identifying Goldstone boson properties as a function of the assumed symmetry breaking pattern. This is done by identifying the general nonlinear realization of the broken global symmetries that the Goldstone bosons carry since that is how the news of these symmetries gets brought to the low-energy theory. Many examples of the structures found here arise in later chapters on applications, such as §8, §13 and §14.

Finally, the latter sections of this chapter examine related issues, such as how a nonlinear realization goes through when the spontaneously broken symmetry is local rather than global (*i.e.* is a gauge symmetry). The main new feature is that spontaneously broken gauge symmetries have no gap, inasmuch as the would-be Goldstone bosons get 'eaten' (through the Higgs mechanism) to provide the longitudinal spin state required for a massive spin-one particle.

As a result the low-energy theory loses the information about the existence of the symmetry in the high-energy sector. For the Wilson action there is operationally no difference at all between a nonlinearly realized gauge symmetry and no gauge symmetry at all. The consistency of this observation with the requirement of gauge symmetries for massless spin-one particles is explored, including the associated breakdown of the low-energy EFT at scales not higher than of order $4\pi m_A/g$, if m_A and g are the spin-one particle's mass and coupling constant.

The final section provides a superficial description of anomalies — the failure of a classical symmetry to survive quantization — as a lead-in to a discussion of anomaly matching. From an EFT perspective anomalous symmetries are not symmetries at all, since for them the 1PI and 1LPI actions are not invariant, $\delta\Gamma[\varphi] \neq 0$, even if the classical action might be. What is special about anomalies is that $\delta\Gamma[\varphi]$ is quite constrained in form, so there can be content in requiring the Wilsonian action to reproduce the transformation properties of the underlying theory.

Exercises

Exercise 4.1 Consider the Goldstone bosons for the symmetry-breaking pattern where the group $G = SU(2)$ breaks down to $H = U(1)$. Take the generators of G to be the 2×2 Pauli matrices $T_a = \frac{1}{2}\tau_a$, with (as usual)

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \quad \text{and} \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and take the generator of H to be T_3. Using the standard realization compute explicit formulae for the two Goldstone fields, ξ^1 and ξ^2, under arbitrary infinitesimal G transformations. Compute the Maurer-Cartan form and the associated quantities $\mathcal{A}_\alpha(\xi)$ and $e^\alpha{}_\beta(\xi)$, and their transformation properties under G. Write down the most general lagrangian up to two derivatives describing the self-couplings of these Goldstone fields, ξ^α, and compute the Noether currents implied by this action for the symmetry group G.

Show that there exists a change of variables $(\xi^1, \xi^2) \to (\vartheta, \varphi)$ that turns your result into the Goldstone fields for a target space that is a 2-sphere:

$$\mathfrak{L}_W = -\frac{F^2}{2}\left[(\partial_\mu\vartheta\,\partial^\mu\vartheta) + \sin^2\vartheta\,(\partial_\mu\varphi\,\partial^\mu\varphi)\right], \tag{4.91}$$

with F^2 an arbitrary positive real constant.

Exercise 4.2 Consider the Goldstone bosons for the symmetry-breaking pattern where the group $G = SU(2) \times SU(2)$ breaks down to $H = SU(2)$ corresponding to the diagonal subgroup (for which both $SU(2)$ factors rotate in the same way rather than independently). How many Goldstone bosons are there for this pattern?

Using the standard realization compute explicit formulae for the Goldstone fields

under arbitrary infinitesimal G transformations. Compute the Maurer-Cartan form and the associated quantities $\mathcal{A}_\alpha(\xi)$ and $e^\alpha{}_\beta(\xi)$, and their transformation properties under G. Write down the most general two-derivative self-couplings for the Goldstone fields ξ^α, and compute its Noether currents for the symmetry group G. This action describes the low-energy interactions of pions.

Show that there is a change of variables that allows your result to be rewritten in the 'nonlinear σ-model' form

$$\mathfrak{L}_w = -\frac{1}{2}\,\frac{\partial_\mu\vec{\pi}\cdot\partial^\mu\vec{\pi}}{(1+\vec{\pi}\cdot\vec{\pi}/F^2)^2}\,. \tag{4.92}$$

Exercise 4.3 Derive the useful identity, eq. (C.132), that is used when proving formulae (4.50) and (4.51) of the main text.

Exercise 4.4 For the symmetry breaking pattern of Exercise 4.1 suppose that the group G is gauged. Show that the low-energy nonlinear realization is equivalent to the theory of a massive charged complex vector field W_μ coupled to a single unbroken $U(1)$ gauge boson, subject only to the constraints of the $U(1)$ invariance. Compute the most general interactions for this theory involving up to four fields and at most two derivatives.

Exercise 4.5 Explicitly evaluate the Feynman graph of Fig. 4.3 and derive the anomaly equation, eq. (4.76).

Exercise 4.6 The strong interactions have a gauge group $SU(3)_c$ (where 'c' stands for 'colour' – for more details see §8). Suppose there are three types of left-handed massless spin-half quarks, q, that transform under $SU(3)_c$ as a triplet (**3**) as well as three types of left-handed spin-half anti-quarks, $\bar{q}$, that transform as an anti-triplet ($\bar{\mathbf{3}}$). It happens that the strong dynamics preserves a global 'flavour' symmetry group $G_f := SU(3)_L \times SU(3)_R \times U(1)$ that commutes with $SU(3)_c$, under which the q's transform as $(\mathbf{3},\mathbf{1})_1$ while the $\bar{q}$ transform as $(\mathbf{1},\bar{\mathbf{3}})_{-1}$ where the subscript gives the charge of the field for the $U(1)$ generator. Evaluate the anomaly coefficients A_{abc} for the generators of G_f using the generators T_a acting on the left-handed quarks and anti-quarks.

It is believed that the strong interactions form bound states that are singlets under $SU(3)_c$. For this quark content these include fermionic bound states (or 'baryons') in the completely antisymmetric colour combination: $B = \epsilon_{abc}q^a q^b q^c$ as well as its conjugate (or 'anti-baryon') $\overline{B} = \epsilon^{abc}\bar{q}_a\bar{q}_b\bar{q}_c$. What are the possible representations that B and $\overline{B}$ can transform in under the flavour group G_f?

Evaluate the anomaly coefficients A_{abc} for the generators of G_f acting on the baryons in each of these representations allowed for the bound-state baryons. Prove that it is impossible to choose the number of types of these representations in the bound-state spectrum in such a way that the A_{abc} for the baryons agree with those obtained from the quarks. The impossibility of doing so provides an argument that for these choices of quantum numbers the strong interactions must spontaneously break the flavour group G_f.

5 Boundaries

The presence of boundaries modifies the above discussion in several ways, such as by removing the freedom to drop total derivative effective interactions in the low-energy action when identifying redundant terms. This short chapter sketches in some of the details of the new features that boundaries bring to low-energy theories. It is short because this is an area for which Wilsonian methods remain relatively poorly developed. It is nonetheless included because it provides a useful starting point for later sections, such as §7.4 and §13, of this book.

In general spacetime is regarded to be a manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$, and the action must be specified on both of these regions to completely specify the problem:

$$S = S_{B}(\phi) + S_{b}(\phi, \psi) = \int_{\mathcal{M}} \mathrm{d}^4x\, \mathfrak{L}_{B} + \int_{\partial\mathcal{M}} \mathrm{d}^3x\, \mathfrak{L}_{b}\,, \tag{5.1}$$

where S_{B} (or the 'bulk' action) describes the dynamics of a collection of fields, $\phi(x)$, in the interior of $\mathcal{M}$ while S_{b} (or the 'boundary' action) describes how these bulk fields couple to the boundary, possibly including to any boundary-localized dynamical degrees of freedom (such as the boundary position, $y^{\mu}(t)$, itself, if it is free to move). Unless stated otherwise the boundary of interest is timelike, consisting of a boundary to space at a given time (in some preferred frame) whose world-volume sweeps out the spacetime boundary as time evolves.

The division of interactions between the bulk and boundary is somewhat fluid since Stokes' theorem can be used to rewrite total derivatives in $\mathfrak{L}_{B}$ as a contribution to $\mathfrak{L}_{b}$. That is, if $\mathfrak{L}_{B} \supset \mathfrak{L}_{\text{tot deriv}} = \partial_{\mu}V^{\mu}$ for some V^{μ}, then the corresponding contribution to the action is

$$\int_{\mathcal{M}} \mathrm{d}^4x\, \mathfrak{L}_{\text{tot deriv}} = \int_{\mathcal{M}} \mathrm{d}^4x\, \partial_{\mu}V^{\mu} = \int_{\partial\mathcal{M}} \mathrm{d}^3x\, n_{\mu}V^{\mu}\,, \tag{5.2}$$

where $n_{\nu} = \{0, \mathbf{n}\}$ is a normal vector on the surface, conventionally chosen to point *out* of the bulk. In the absence of boundaries total derivatives are redundant because they can be simply dropped from the action with no physical consequences. With boundaries these same effective couplings carry consequences, but remain redundant inasmuch as their consequences are not distinct from those of interactions within the boundary action.

5.1 'Induced' boundary conditions

What boundary conditions should be imposed at the boundary $\partial\mathcal{M}$? Since quantum field theory is the business of evaluating path integrals over fields, one way to approach boundary conditions is to imagine formulating the path integral itself to be over a space of fields all of which satisfy some condition on the boundary — perhaps the fields or their derivatives or some combination of these vanish. The classical limit of such a problem then involves expanding about a saddle point defined by solving the classical field equations subject to the assumed boundary conditions on $\partial\mathcal{M}$. What is unsatisfying about such problems is the arbitrariness of the boundary conditions, which are simply handed down by God when formulating the problem.

Less arbitrary are systems for which the boundary conditions on $\partial\mathcal{M}$ are 'induced' inasmuch as they can be derived from the form of the action, with the action itself acquiring a new contribution specifically associated with the boundary. These kinds of boundary conditions typically arise when the path integral runs over arbitrary field configurations, both within the interior of $\mathcal{M}$ and on its boundary, as is very often the case in real systems.

Induced boundary conditions are of physical interest because in real applications the boundary is usually not a physical ending of spacetime on which fields (and/or their derivatives) are specified once and for all. Instead the boundary usually arises as an approximate description of a place where there is a very rapid change of background properties; for electromagnetic applications perhaps it is the edge of a conducting region or a dielectric object beyond which one chooses not to track field behaviour (see for example Exercise 5.4). Deep down, boundary physics is in such cases no different from bulk physics, and one should imagine integrating over all possible values of both bulk and boundary fields when performing the path integral.

At the classical level this means that saddle points are chosen by demanding the action be stationary against variations of the fields both in the bulk and on the boundary. It is stationarity against variation of fields on the boundary that dynamically dictates the classical boundary conditions that hold on $\partial\mathcal{M}$. Most importantly, tying boundary conditions to an action in this way ultimately allows all of the EFT reasoning described in this book to be brought to bear when deciding which boundary conditions should arise in any given situation.

The toy model

To make the issues concrete return to the toy model of §1.1: the self-interactions of a complex field to which is now added a boundary term:

$$S = -\int_{\mathcal{M}} \mathrm{d}^4x \left[\partial_\mu\phi^*\partial^\mu\phi + V(\phi^*\phi)\right] + \int_{\partial\mathcal{M}} \mathrm{d}^3x\, \mu\, \phi^*\phi\,. \tag{5.3}$$

Here μ is a new parameter with dimensions of mass, and the above choice for S_b is the lowest-dimension possibility involving ϕ that is local and invariant under the symmetry

$\phi \to e^{i\theta}\phi$. The path integral over ϕ is unconstrained both throughout the interior and boundary of $\mathcal{M}$.

In the classical approximation the path integral is computed as an expansion about a saddle point, $\phi_c(x)$, defined as the configuration where $\delta S(\phi = \phi_c)$ vanishes. Writing $\phi \to \phi + \delta\phi$ and linearizing the action in $\delta\phi$ then leads to the expression

$$\delta S = \int_{\mathcal{M}} d^4x \left[\left(\Box\phi - \frac{\partial V}{\partial \phi^*}\right)\delta\phi^* + \text{c.c.}\right] + \int_{\partial\mathcal{M}} d^3x\Big[\Big(-\partial_n\phi + \mu\phi\Big)\delta\phi^* + \text{c.c.}\Big], \tag{5.4}$$

where the first term in the second integral comes from an integration by parts in the bulk, with $\partial_n := n_\mu\partial^\mu$ denoting the normal derivative at the boundary.

Because the path integral is over arbitrary fields the saddle point must be stationary against arbitrary variations $\delta\phi$ everywhere within $\mathcal{M}$ and $\partial\mathcal{M}$. Restricting first to those variations that vanish on the boundary shows that $\phi_c(x)$ must satisfy the usual classical field equations throughout $\mathcal{M}$:

$$\left(\Box\phi - \frac{\partial V}{\partial \phi^*}\right)_{\phi=\phi_c} = 0\,. \tag{5.5}$$

Stationarity of the action against arbitrary variations on the boundary then shows the saddle point must satisfy the induced boundary condition

$$\partial_n\phi_c = \mu\,\phi_c \qquad \text{on } \partial\mathcal{M}\,. \tag{5.6}$$

Several things about this boundary-value problem are noteworthy. First, the boundary condition (5.6) is linear. Because the boundary condition is derived from the action this is not automatic, and in this particular example is a consequence of using only the lowest-dimension term (which is quadratic in ϕ) for the boundary action, S_b, in (5.3). As for any Wilsonian action, ultimately the justification for using low-dimension terms in S_b will rely on the low-energy approximation, and in this lies the seeds of an explanation as to why linear boundary conditions so often play a role throughout physics.

Second, notice that the boundary condition (5.6) forbids the vanilla vacuum solution of constant field, $\phi = v$, which minimizes V. In general the coupling to the boundary causes a trade-off between trying to minimize the scalar potential throughout the bulk and paying some gradient energy to satisfy (5.6) on the boundary.

For instance, suppose the boundary is the $x - y$ plane (at $z = 0$) and the bulk is the region $z > 0$. Then neglecting the interactions of the potential that are cubic and quartic in $\psi = \phi - v$ implies a bulk solution of the form $\phi_c = v + \psi_c(z)$ satisfies $\psi_c'' - m_R^2\,\psi_c \simeq 0$ where primes denote d/dz and (as before) $m_R^2 = \lambda v^2$. Requiring $\phi \to v$ as $z \to \infty$ and satisfying the boundary condition (5.6) implies the approximate saddle point solution

$$\phi_c(z) \simeq v\left(1 - \frac{\mu}{\mu + m_R}\, e^{-m_R z}\right). \tag{5.7}$$

Eq. (5.7) is consistent with the neglect of ψ^3 and ψ^4 in V for the regime $\mu \ll m_R$ since in this case $|\phi_c - v| \sim O(\mu v/m_R) \ll v$.

The bulk energy cost of interacting with the boundary can be estimated by evaluating the

classical energy at the solution (5.7). Dropping subdominant powers of μ/m_R, the resulting classical bulk energy-per-unit-area is

$$\frac{E_B}{A} \simeq \int_0^\infty \mathrm{d}z \left[|\phi'|^2 + V(\phi^*\phi)\right] \simeq \frac{\mu^2 v^2}{m_R} = \frac{\mu^2 v}{\sqrt{\lambda}}\,. \tag{5.8}$$

This expression drops the cubic and quartic terms of V, since these are also down relative to (5.8) by at least one power of μ/m_R. This bulk energy cost is more than compensated by the boundary contribution to the energy-per-unit-area, which is

$$\frac{E_b}{A} = -\mu|\phi(0)|^2 \simeq -\mu v^2 + \frac{2\mu^2 v^2}{m_R}\,. \tag{5.9}$$

Semiclassical quantum corrections to this classical result are computed using the same steps as used without a boundary: expand all fields about the background, $\phi = \phi_c(z) + \hat{\phi}$ and quantize the fluctuations $\hat{\phi}$. The main difference is that any expansion in terms of modes $\hat{\phi}(x) = \sum_n [u_n(x)\, \mathfrak{a}_n + \text{c.c.}]$ involve modes defined in the presence of the background. In the present instance this means that they are not eigenstates of the z-component of momentum, due to the breaking of translation invariance in this direction by the background $\phi_c(z)$. See §13.1 for more about semiclassical expansions about position-dependent classical background fields.

5.2 The low-energy perspective

Any boundary physics of the full theory that persists to low energies should be directly describable in terms of a low-energy EFT. For induced boundary conditions derivable from a boundary action this means the Wilsonian action should also have a boundary component from which the low-energy boundary physics can be inferred. As always with a Wilsonian action the form of the low-energy boundary action is obtained by matching, inasmuch as it is defined by the requirement that it reproduce the boundary physics of the full theory order by order in the low-energy expansion.

This is all made more concrete using the toy model example just described. For fields varying slowly compared with the length scale m_R^{-1} the effects of the boundary should also be calculable within the low-energy EFT appropriate below the mass scale m_R, for which only the Goldstone field ξ survives. This EFT is the one encountered in earlier sections, with a shift symmetry $\xi \to \xi + c$, but now including boundary interactions that also respect this symmetry (because the boundary term in (5.3) respects the $U(1)$ symmetry under rephasings of ϕ).

At the classical level the physics of the UV system that the boundary part of the Wilson action captures is the boundary condition satisfied at $\partial\mathcal{M}$ by ξ. In the example above the UV boundary condition is $\partial_n\phi = \mu\phi$, and the implications of this condition for the light fields can be found in the full theory by using $\phi = \varrho \exp\left[\mathrm{i}\xi/\sqrt{2}\,v\right]$, with $\varrho := v + \chi/\sqrt{2}$. For real μ the real and imaginary parts of the boundary condition for ϕ give the two separate

real boundary conditions for χ and ξ (or, equivalently, for ϱ and ξ)

$$\partial_n \varrho = \mu \varrho \qquad \text{and} \qquad \varrho\, \partial_n \xi = 0 \quad \text{on } \partial\mathcal{M}\,. \tag{5.10}$$

Assuming $\varrho \neq 0$ at $\partial\mathcal{M}$, at leading (classical) order in λ the low-energy boundary action at $\partial\mathcal{M}$ should imply $\partial_n \xi = 0$ there. More generally S_b in the effective theory captures the dependence of the ξ boundary physics order by order in the low-energy expansion.

What does this imply explicitly for S_b in the toy model? Consider first the self-interactions of ξ involving the smallest mass dimension. The most general possible local bulk and boundary interactions consistent with the symmetries[1] are

$$\begin{aligned} \mathfrak{L}_{\scriptscriptstyle B} &= -\frac{Z_1}{2}\, \partial_\mu \xi\, \partial^\mu \xi + \frac{Z_2}{2}\, \xi \Box \xi - m_{\scriptscriptstyle B}\, \Box \xi + \cdots \\ \mathfrak{L}_b &= -w^3 - m_b\, \partial_n \xi - \frac{Z_2}{2}\, \xi\, \partial_n \xi - c_1\, \partial_n^2 \xi - c_2\, \partial^a \partial_a \xi - \frac{h_1}{2} (\partial_n \xi)^2 + \cdots\,, \end{aligned} \tag{5.11}$$

where the effective couplings w, $m_{\scriptscriptstyle B}$ and m_b have dimension mass, the couplings, Z_1, Z_2, c_1, c_2 and c_3 are dimensionless and h_1 has dimension $(\text{mass})^{-1}$. (This list does not exhaust the possibilities for the dimensions shown.) Here ∂_n denotes the normal derivative $n_\mu \partial^\mu$, while ∂_a denotes derivatives only along directions parallel to $\partial\mathcal{M}$ (as opposed to ∂_n, which is in the direction perpendicular to $\partial\mathcal{M}$, and ∂_μ, which indicates differentiation in all of the directions within $\mathcal{M}$). The coefficients of the $\xi\Box\xi$ term in $\mathfrak{L}_{\scriptscriptstyle B}$ and the $\xi\partial_n\xi$ term in $\mathfrak{L}_b$ must be related in the way indicated in order to preserve the invariance of the total action, $S = S_{\scriptscriptstyle B} + S_b$, under the shift symmetry.

Some of these interactions are redundant, for both of the reasons discussed in §2.5. An important difference from this earlier discussion is that total derivatives in $\mathfrak{L}_{\scriptscriptstyle B}$ can no longer simply be dropped. Instead Stokes' theorem relates such terms to terms in the boundary action. For example, the $\Box\xi$ term in $\mathfrak{L}_{\scriptscriptstyle B}$ can be converted in this way to the $\partial_n\xi$ term on $\mathfrak{L}_b$, showing that physical quantities can only depend on the combination $\tilde{m} := m_b + m_{\scriptscriptstyle B}$ rather than either m_b or $m_{\scriptscriptstyle B}$ separately. Similarly, integrating by parts either the Z_1 or Z_2 terms in $\mathfrak{L}_{\scriptscriptstyle B}$ shows that these parameters only can contribute as the sum $Z := Z_1 + Z_2$. Furthermore, the combination $\partial^a \partial_a \xi$ is a total derivative within $\partial\mathcal{M}$, and can always be dropped given that the boundary itself has no boundary.

Using this freedom allows the above action to be rewritten as

$$\mathfrak{L}_{\scriptscriptstyle B} = -\frac{Z}{2}\, \partial_\mu \xi\, \partial^\mu \xi \quad \text{and} \quad \mathfrak{L}_b = -w^3 - \tilde{m}\, \partial_n \xi - c_1\, \partial_n^2 \xi - \frac{h_1}{2} (\partial_n \xi)^2\,, \tag{5.12}$$

and rescaling the field, $\xi \to \xi/\sqrt{Z}$, shows the four parameters Z, $\tilde{m}$, c_1 and h_1 only enter physical quantities through the three combinations $\tilde{m}/\sqrt{Z}$, $c_1/\sqrt{Z}$ and h_1/Z. This freedom is now used to set $Z = 1$ (*i.e.* to 'canonically normalize' the field), leaving only three potentially independent parameters $\tilde{m}$, c_1 and h_1 of the terms considered in (5.11).

But even these parameters need not be independent (or present at all). To see why recall this effective theory arises from UV physics where all fields are integrated freely both

[1] In this example $\partial\mathcal{M}$ is chosen to be flat and Poincaré invariant along the directions parallel to the boundaries, although dependence on the boundary's local geometry – such as its curvature – is in general present if the physics of the boundary is more complicated.

within the bulk and on the boundary. Consequently the functional integral over ξ in the low-energy theory is also unconstrained in both the bulk and on the boundary. The classical limit for such a free integration is then found by evaluating the path integral at a classical path chosen to make the action stationary against arbitrary variations of ξ both in the interior of $\mathcal{M}$ *and* throughout $\partial\mathcal{M}$, and this should be consistent with what is found for the UV completion.

To see what this requires write $\xi \to \xi + \delta\xi$ and linearize the action in $\delta\xi$, to find

$$\begin{aligned} \delta S &= -\int_{\mathcal{M}} \mathrm{d}^4x\, \partial_\mu\xi\, \partial^\mu\delta\xi - \int_{\partial\mathcal{M}} \mathrm{d}^3x \left[\tilde{m}\, \partial_n\delta\xi + c_1\, \partial_n^2\delta\xi + h_1\partial_n\xi\, \partial_n\delta\xi\right] \\ &= \int_{\mathcal{M}} \mathrm{d}^4x\, (\Box\xi)\, \delta\xi - \int_{\partial\mathcal{M}} \mathrm{d}^3x \left[\partial_n\xi\, \delta\xi + (\tilde{m} + h_1\partial_n\xi)\partial_n\delta\xi + c_1\, \partial_n^2\delta\xi\right]. \end{aligned} \tag{5.13}$$

Since the action must be stationary against *arbitrary* $\delta\xi$, first choose $\delta\xi$ and its derivatives to be only nonzero away from the boundary. As usual, this implies the saddle-point $\xi_c(x)$ must satisfy the classical field equations $\Box\xi_c = 0$ everywhere in $\mathcal{M}$.

Next demand also $\delta S = 0$ for arbitrary variations of ξ on $\partial\mathcal{M}$. First do so by choosing the variations so that $\partial_n\delta\xi = \partial_n^2\delta\xi = 0$ but with $\delta\xi \neq 0$ on $\partial\mathcal{M}$. Requiring $\delta S = 0$ for all such $\delta\xi$ implies the saddle point must satisfy *Neumann* boundary conditions: $\partial_n\xi_c(x) = 0$ on $\partial\mathcal{M}$. But now requiring $\delta S = 0$ for variations with $\partial_n^2\delta\xi = 0$ but $\partial_n\delta\xi \neq 0$ requires $\tilde{m} + h_1\partial_n\xi_c = 0$ on $\partial\mathcal{M}$; a result inconsistent with Neumann boundary conditions unless $\tilde{m} = 0$. Similarly, variations with $\partial_n^2\xi \neq 0$ imply further conditions ($c_1 = 0$ if only the displayed terms are kept).

The very presence of nonzero couplings $\tilde{m}$ and c_1 presents an obstruction to being able to find a consistent boundary condition at $\partial\mathcal{M}$, and thereby also obstructs there being a saddle point for which $\delta S = 0$ when ξ varies arbitrarily on the boundary. Since this obstruction did not arise in the UV theory the appropriate matching condition must be that $\tilde{m} = c_1 = 0$. The resulting boundary condition at this order is then $\partial_n\xi_c = 0$ on $\partial\mathcal{M}$, agreeing with the result found in the full theory in (5.10). Nontrivial effective coupling can arise in S_b at higher orders in the semiclassical expansion to the extent that they are required in order to reproduce modifications to the UV physics there.

This example shows that terms in $\mathfrak{L}_b$ involving normal derivatives of the fields can (in general) over-determine the boundary conditions obtained by varying the action freely on the boundary. Normal derivatives are special in this way because they cannot be integrated by parts on the boundary, making it impossible to rewrite their variation in terms only of $\delta\xi$ (as opposed to its derivatives). Because of this it is generic that the effective couplings for interactions involving normal derivatives in the boundary action are completely determined by the effective couplings for terms in $\mathfrak{L}_B$. (For example, in the example above the coefficient of $\xi\,\partial_n\xi \in \mathfrak{L}_b$ is not independent of the coefficient of $\xi\,\Box\xi \in \mathfrak{L}_B$.) It is only the couplings for the rest of the effective interactions that represent independent parameters describing low-energy properties of the boundary.

The other way that interactions can be redundant is if they can be removed using a local field redefinition, again following the arguments of §2.5. To see how this works for terms on the boundary suppose the bulk action is dominated by the kinetic term, $\mathfrak{L}_{B0} = -\frac{1}{2}\,(\partial_\mu\xi\,\partial^\mu\xi)$, in the regime of semiclassical perturbation theory (as it is for the toy model example).

Performing a change of variables $\xi \to \xi + \epsilon\zeta(\xi)$ — with ϵ a small perturbation parameter — then use of Stoke's theorem leads to the following change in the bulk action

$$\delta S_{B0} = -\epsilon \int_{\mathcal{M}} \mathrm{d}^4x\, \partial_\mu\xi\, \partial^\mu\zeta = -\epsilon \int_{\partial\mathcal{M}} \mathrm{d}^3x\, \partial_n\xi\, \zeta + \epsilon \int_{\mathcal{M}} \mathrm{d}^4x\, \zeta\, \Box\xi\,. \tag{5.14}$$

For generic $\zeta(\xi)$ it is the last term of (5.14) that was used in §2.5 to argue that terms proportional to $\Box\xi$ can be removed in an order-ϵ term of the bulk action. Eq. (5.14) then shows that the change of variables also removes the order-ϵ terms in S_b whose effective coupling is tied to the removed bulk term. For example a transformation with $\zeta = c_{\rm eff}\partial_\mu\xi\,\partial^\mu\xi$ that removes a term $-\epsilon c_{\rm eff}(\partial_\mu\xi\,\partial^\mu\xi)\Box\xi$ in $\mathfrak{L}_B$ also adds a term $-\epsilon c_{\rm eff}(\partial_\mu\xi\,\partial^\mu\xi)\partial_n\xi$ to $\mathfrak{L}_b$.

A coupling in (5.11) not determined by boundary conditions in this way is the parameter w^3. At first sight this coupling might be thought to be redundant inasmuch as the corresponding term in the effective action does not depend on the low-energy field ξ. This parameter nonetheless carries physical content since it contributes to the low-energy stress energy[2] and so also to the energy-per-unit-area of the boundary. Matching this to the result, (5.8), obtained in the UV theory for the toy model with a boundary at $z = 0$ then implies (to leading order)

$$w^3 \simeq -\mu v^2 + \frac{3\mu^2 v^2}{m_R} = -\mu v^2 + \frac{3\mu^2 v}{\sqrt{\lambda}}\,. \tag{5.15}$$

5.3 Dynamical boundary degrees of freedom

One thing not yet captured by this chapter's discussion is the possible existence of fields, ψ, that appear only in the boundary action, S_b, and not at all in the bulk, S_B, so $S[\phi,\psi] = S_B[\phi] + S_b[\phi,\psi]$. Such fields capture the low-energy physics of states in the UV theory whose mode-functions have support only in the immediate vicinity of the boundary, and so are said to be 'localized' at the boundary. These could be anything from surface charges on a conductor or boundary states at the interface between materials to the motion of p-branes [87] in supergravity and open strings attached to D-branes [88] in string theory.

Boundary-localized fields depend only on the three coordinates, σ^α, that label position on $\partial\mathcal{M}$. For example, for a boundary consisting of the $x-y$ plane at $z = 0$ in a flat cartesian space these three coordinates might be $\{\sigma^\alpha\} = \{t, x, y\}$.

Perhaps the simplest such a localized field describes the position of the boundary itself, $y^\mu(\sigma)$, where the boundary's position in spacetime is denoted $x^\mu = y^\mu(\sigma)$. This is a dynamical field if this boundary position is free to move at low energies.[3] In the presence of such fields the boundary action, S_b, does double duty: it both determines the dynamics of y^μ given the presence of any bulk 'background' fields, $\phi(x)$, and it determines how boundaries source these same bulk fields.

[2] That is to say, it does couple to a low-energy field: the spacetime metric.

[3] §6.3.1, §13.1.2 and §14.3.1 argue why these fields often behave like Goldstone modes for spacetime symmetries, and as such naturally appear in the low-energy sector relevant for EFT methods.

For example, consider an ordinary real scalar field, ϕ, coupled to a dynamical but slowly-moving boundary $y^\mu(\sigma) = \{t, x, y, \mathfrak{z}(x, y, t)\}$, located at $z = \mathfrak{z}(x, y, t)$ with $\mathfrak{z}(x, y, t)$ a single-valued function whose derivatives are small: *e.g.* $|\dot{\mathfrak{z}}| \ll 1$ with over-dots representing $\mathrm{d}/\mathrm{d}t$. Then an expansion of the boundary action in powers of $\dot{\mathfrak{z}}$ might take the form

$$S_b[\phi, y] = -\int \mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\,\left[W[\phi(z=\mathfrak{z})] + \frac{1}{2}\,K[\phi(z=\mathfrak{z})]\,\dot{\mathfrak{z}}^2 + \cdots\right], \tag{5.16}$$

where $W(\phi)$ and $K(\phi)$ are specified functions that are characteristic of the surface. (Spatial derivatives of $\mathfrak{z}$ can also be considered but are dropped here for simplicity.) The evolution of the bulk field ϕ, is for simplicity imagined to have a bulk lagrangian dominated by

$$S_{\scriptscriptstyle B}[\phi] = -\int \mathrm{d}^4x\left[\frac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi + \frac{1}{2}\,m_\phi^2\,\phi^2 + \cdots\right]. \tag{5.17}$$

With these choices the motion of the boundary in the presence of a given bulk field configuration, $\phi(x)$, is found at the classical level by varying S_b with respect to $\mathfrak{z}$, and gives

$$\left[K\,\ddot{\mathfrak{z}} + \left(\frac{1}{2}\,\frac{\partial K}{\partial\phi}\,\dot{\mathfrak{z}}^2 - \frac{\partial W}{\partial\phi}\right)\frac{\partial\phi}{\partial z}\right]_{z=\mathfrak{z}(x,y,t)} = 0\,, \tag{5.18}$$

as the equation of motion governing the time-dependence of $\mathfrak{z}(x, y, t)$. The classical field equations for ϕ obtained by varying $S_{\scriptscriptstyle B} + S_b$ in the bulk similarly give

$$(\Box - m_\phi^2)\phi = 0 \qquad \text{for } z > \mathfrak{z}(x, y, t) \tag{5.19}$$

while variations on the boundary give the condition

$$\partial_n\phi + \left(\frac{\partial W}{\partial\phi} + \frac{1}{2}\,\frac{\partial K}{\partial\phi}\,\dot{\mathfrak{z}}^2 + \cdots\right) = 0 \qquad \text{for } z = \mathfrak{z}(x, y, t)\,. \tag{5.20}$$

A slightly different but related picture can arise if the boundary is instead regarded as a thin surface (*i.e.* membrane — or in its relativistic incarnations 'brane' [87]) with two sides, rather than effectively being the edge of spacetime. In this case the boundary action (5.16) can instead be written more usefully as the two-sided brane action

$$S_b[\phi, y] = -\int \mathrm{d}^4x\,\left[W[\phi(x)] + \frac{1}{2}\,K[\phi(x)]\,\dot{\mathfrak{z}}^2 + \cdots\right]\delta[z - \mathfrak{z}(x, y, t)]\,, \tag{5.21}$$

leading to a ϕ equation of the form

$$(\Box - m_\phi^2)\phi - \left[\frac{\partial W}{\partial\phi} + \frac{1}{2}\,\frac{\partial K}{\partial\phi}\,\dot{\mathfrak{z}}^2 + \cdots\right]\delta[z - \mathfrak{z}(x, y, t)] = 0\,. \tag{5.22}$$

Here the boundary condition becomes a 'jump' condition obtained by integrating (5.22) over an infinitesimal region $\mathfrak{z} - \epsilon < z < \mathfrak{z} + \epsilon$ that includes the delta function:

$$\left[\partial_n\phi\right]_{\mathfrak{z}} + \left(\frac{\partial W}{\partial\phi} + \frac{1}{2}\,\frac{\partial K}{\partial\phi}\,\dot{\mathfrak{z}}^2 + \cdots\right)_{z=\mathfrak{z}(x,y,t)} = 0\,. \tag{5.23}$$

where the square bracket denotes the jump in a quantity across $z = \mathfrak{z}$, as in $\left[\partial_n\phi\right]_{\mathfrak{z}} = \partial_n\phi(z = \mathfrak{z} + \epsilon) - \partial_n\phi(z = \mathfrak{z} - \epsilon)$ with $\epsilon \to 0$ at the end.

5.4 Summary

Boundaries do not change the EFT story in a dramatic way. This chapter maps out the various small ways that boundaries do change low-energy dynamics focussing on 'induced' boundary conditions, defined as those that are obtained by extremizing an action with respect to field variations on the boundary. Such boundary conditions arise naturally in situations where the path integral is over all fields in an unconstrained way, both in the bulk and on the boundary.

Like the devil, the main differences associated with boundaries are in the details. The central new feature is the addition of a local boundary component to the Wilsonian action. Its effective couplings are (as usual) obtained by demanding that the low-energy theory reproduces the full theory's boundary physics order-by-order in the low-energy expansion. The precise ways that redundant interactions are identified in the Wilsonian action are slightly modified due to the ability to swap terms between the bulk and boundary actions by integrating by parts.

Terms in the boundary action involving normal derivatives play a special role because their presence can over-determine the boundary conditions when extremizing against arbitrary field variations. Consequently the matching process often ends up fixing their effective couplings in terms of the values of couplings appearing in the bulk lagrangian (or makes them vanish). When this happens they do not represent independent parameters associated with the physics of the boundary.

An attractive feature about tying boundary conditions to boundary components of a low-energy Wilsonian action is the context it provides for understanding why some boundary conditions arise more often in physical situations than do others. The leading low-energy contributions to boundary terms in the Wilson action are usually not complicated for the same reasons that terms in the bulk are not: low energy often puts a premium on having few fields and derivatives and not many terms are possible involving the fewest of each. This suggests an underlying reason for why relatively simple boundary conditions (like those linear in the fields) arise so frequently in practice.

A final qualitatively new feature that boundaries can introduce are localized degrees of freedom that live only at the boundary. When these exist their interactions with bulk fields are governed by the boundary action, and their semiclassical treatment goes through much the same as for bulk fields alone.

Exercises

Exercise 5.1 Derive the approximate classical solution eq. (5.7) in the bulk for a field satisfying the field equation (5.5) in the regime $z \geq 0$ subject to the boundary condition $\phi \to v$ as $z \to \infty$ and eq. (5.6) at $z = 0$.

Evaluate the classical energy per unit area, E/A, and verify (5.8) holds when the cubic and quartic terms in $(\phi - v)$ are neglected in the potential V. Include the cubic and quartic terms in the energy when evaluating E/A at the solution of (5.7), and thereby quantify how suppressed they are as a function of the small dimensionless parameters μ/m_R and $\sqrt{\lambda} = m_R/v$.

Exercise 5.2 For the same bulk and boundary action as in Exercise 5.1 write the full quantum field as $\phi = \phi_c + \hat{\phi}$ where ϕ_c is the classical solution (5.7). What are the boundary conditions for the quantum fluctuation field $\hat{\phi}$ at the boundary at $z = 0$? Using this boundary condition compute the mode functions, $u_n(z)e^{i(k_x x + k_y y - \omega t)}$, appearing in the expansion of $\hat{\phi}$ in terms of creation and annihilation operators. (Neglect the cubic and quartic interaction terms in the bulk scalar potential when doing so.) Are any of these modes bound states localized near the boundary? (Bound states have energies $\omega^2 < k_x^2 + k_y^2 + m_R^2$ and so have wave-functions that are normalizable in the z-direction.) If so, what is the mode profile $u_n(z)$ and energy ω_n?

Exercise 5.3 Repeat Exercise 5.1, but with a flat mobile brane held stationary at $z = z_b$ with $z_b > 0$. Compute the approximate field profile for $\phi(z)$ (as a function of z, z_b, μ and m_R) on both sides of the mobile brane (*i.e.* for both $0 \leq z < z_b$ and $z_b < z < \infty$), using the same boundary conditions as before at $z = 0$ and $z \to \infty$. Neglect the cubic and quartic interactions in the bulk potential when doing so, and suppose the action for the mobile brane is given by (5.21) with $K = 1$ and $W = g\phi^*\phi$. Use the jump condition (5.23) (as well as continuity of ϕ itself) to evaluate the boundary conditions at $z = z_b$.

Use eq. (5.18) to evaluate the acceleration of the mobile brane, $\ddot{z}_b$ if it were free to move. Which direction does the mobile brane go once it is released from rest at $z = z_b$?

Exercise 5.4 Consider two bulk regions, R_1 and R_2, on either side of an interface, F, with the interface regarded as a common boundary shared by the two regions. Suppose the bulk action for the electromagnetic field in each region is that of a dielectric with differing dielectric constants, corresponding to

$$S_{R_i} = \frac{1}{2}\int_{R_i} \mathrm{d}^4x \left[\varepsilon_i \, \mathbf{E}\cdot\mathbf{E} - \frac{\mathbf{B}\cdot\mathbf{B}}{\mu_i}\right],$$

where ε_i is the dielectric constant and μ_i the magnetic permeability for each bulk region. Take the boundary action for the common interface to be

$$S_F = -\int_F \mathrm{d}^3x \, A_\mu J^\mu$$

where $A^0 = \Phi$ is the electrostatic potential and $\mathbf{A}$ is the vector potential while

$J^0 = \rho$ is the interface's surface charge density and $\mathbf{J}$ is its surface current (satisfying $\partial_t \rho + \nabla \cdot \mathbf{J} = 0$). By varying the electromagnetic potentials Φ and $\mathbf{A}$ derive the dielectric Maxwell equations in each of the bulk regions R_i as well as the boundary conditions obtained by demanding the action $S_{R_1} + S_{R_2} + S_F$ be stationary under arbitrary variations δA_μ on the interface. Show that your results reproduce the standard ones: the jump across of the interface of the normal component of $\mathbf{D} = \varepsilon \mathbf{E}$ is given by the charge density; the jump in the tangential component of $\mathbf{H} = \mathbf{B}/\mu$ is given by the surface current; while the other components of $\mathbf{E}$ and $\mathbf{B}$ are continuous at F.

6 Time dependent systems

To this point the discussion many of the EFT applications have been to scattering problems for low-energy states arising as fluctuations about a stationary ground state. This really only scratches the surface of the utility of effective field theories, as this chapter hopes to convey. This chapter asks how to apply EFT methods to systems involving background fields that evolve in time. This kind of problem arises throughout physics, including (but not limited to) atomic interactions with time-dependent electromagnetic fields, particle motion through inhomogeneous media and the time-varying fields of early-universe cosmology.

For the purposes of argument in this chapter the time-varying background field is taken to be a scalar, in order to make better contact with the toy model. But examples could equally well be considered using background electromagnetic or gravitational fields, some of which are considered amongst the examples examined in later sections.

A number of new conceptual issues arise when setting up an effective description of systems with time-dependent backgrounds. One such asks why 'low energy' and 'high energy' remain useful as criteria for splitting up the space of states given that the breaking of time-translation invariance means fluctuation energy is not strictly conserved. Another asks whether all solutions to the full theory's field equations have counterpart solutions in the effective theory, and vice versa. A third asks what the correct number of initial conditions should be in an effective theory, given that the low-energy field equations can involve more than two time derivatives.

These issues do not arise in simpler static settings, and although none of them need preclude using low-energy techniques the validity of EFT methods sometimes involves additional criteria that must be checked explicitly for any particular application. Most notable among these new conditions is the requirement that any background evolution be slow enough to be adiabatic (in a sense that is further elaborated below).

6.1 Sample time-dependent backgrounds ♢

Just like in earlier sections it is useful to ground a general discussion of issues by having a concrete example in mind. To this end this chapter starts with an example of time-dependent backgrounds within the toy model introduced in §1.1, using it to illustrate how EFT methods work for time-dependent settings and why they can sometimes fail.

First a brief reminder of the main features of the toy model, for ease of reference. Its

lagrangian density is given by

$$\mathfrak{L} = -\partial_\mu \phi^* \partial^\mu \phi - V(\phi^* \phi)\,, \tag{6.1}$$

where the complex field ϕ is written in terms of two real fields using either $\phi = \frac{1}{\sqrt{2}}(\phi_R + \mathrm{i}\phi_I)$ or $\phi = \varrho\, e^{\mathrm{i}\vartheta}$, where these are related to the variables used in previous sections by $\phi_R = \sqrt{2}\, v + \tilde{\phi}_R$, $\phi_I = \tilde{\phi}_I$, $\xi = \sqrt{2}\, v\, \vartheta$ and $\chi = \sqrt{2}(\varrho - v)$. Furthermore, the potential has the explicit 'Mexican hat' or 'wine-bottle' form

$$V(\phi^* \phi) = \frac{\lambda}{4}\left(\phi^* \phi - v^2\right)^2\,, \tag{6.2}$$

which at low energies has a level, circular trough with a bottom at $V = 0$ along the curve $\sqrt{2}\,\varrho = \sqrt{\phi_R^2 + \phi_I^2} = \sqrt{2}\,|\phi| = \sqrt{2}\, v$.

The model's Noether current for the $U(1)$ symmetry is given by (4.25),

$$j_\mu = \mathrm{i}(\phi\, \partial_\mu \phi^* - \phi^* \partial_\mu \phi) = \phi_R \partial_\mu \phi_I - \phi_I\, \partial_\mu \phi_R = 2\varrho^2 \partial_\mu \vartheta\,, \tag{6.3}$$

and Noether's theorem implies this satisfies $\partial_\mu j^\mu = 0$ whenever the field equations,

$$\Box \phi = -\partial_t^2 \phi + \nabla^2 \phi = \frac{\lambda}{2}\left(\phi^* \phi - v^2\right)\phi\,, \tag{6.4}$$

are satisfied.

Slow-roll backgrounds

Until now the only background solution to (6.4) to be considered has been the static vacuum solution $\phi = v$. Consider instead the time-dependent background corresponding to the scalar field homogeneously rolling around the bottom of its potential [89]:

$$\varrho = \varrho_0 \qquad \text{and} \qquad \vartheta(t) = \vartheta_0 + \omega t\,, \tag{6.5}$$

for constants ϱ_0, ϑ_0 and ω. Eq. (6.4) implies these constants must satisfy

$$\left[\frac{\lambda}{2}\left(\varrho_0^2 - v^2\right) - \omega^2\right]\varrho_0 = 0\,, \tag{6.6}$$

so the only solution with $\varrho_0 > 0$ is

$$\varrho_0 = \sqrt{v^2 + \frac{2\omega^2}{\lambda}}\,. \tag{6.7}$$

This shows how the force due to the scalar potential's gradient competes with the centripetal acceleration due to the circular motion to drive the radial field ϱ slightly away from the trough's bottom.

The density of the conserved Noether charge evaluated at this solution is

$$j^\mu = -2\omega\varrho_0^2\, \delta_0^\mu = -2\omega\left(v^2 + \frac{2\omega^2}{\lambda}\right)\delta_0^\mu\,, \tag{6.8}$$

and its energy density is

$$\varepsilon = \dot\phi^*\dot\phi + \nabla\phi^*\cdot\nabla\phi + \frac{\lambda}{4}(\phi^*\phi - v^2)^2 = \omega^2\varrho_0^2 + \frac{\lambda}{4}(\varrho_0^2 - v^2)^2$$
$$= \omega^2\left(v^2 + \frac{3\omega^2}{\lambda}\right), \tag{6.9}$$

where over-dots denote differentiation with respect to t. For later purposes notice that the appearance of a 3 (instead of a 2) in the last term of the last line of this last equation can be traced to the contribution of the scalar potential to ε in the second-last line. The potential contributes because the motion displaces the field away from the potential's minimum by the amount given in (6.7).

6.1.1 View from the EFT

This section now asks how the above rolling solution looks from the point of view of the low-energy EFT appropriate at energies well below m_R, which should be a valid regime for a sufficiently slowly moving background. In particular, how does the EFT know about the energy increase of (6.9) due to the field ϱ climbing part way up the potential if there is no field ϱ left in the effective theory to adjust to balance centrifugal forces, and no scalar potential (or indeed notion of centripetal acceleration) within the EFT.

Previous sections show that the leading approximation to the low-energy EFT for this model is given by (2.97), which in the classical approximation (using $m_R^2 = \lambda v^2$) is

$$S_w[\xi] = -\int \mathrm{d}^4x\,\left[\frac{1}{2}\,\partial_\mu\xi\partial^\mu\xi - \frac{\lambda}{4m_R^4}(\partial_\mu\xi\,\partial^\mu\xi)^2 + \cdots\right]$$
$$= -\int \mathrm{d}^4x\,\left[v^2\partial_\mu\vartheta\partial^\mu\vartheta - \frac{v^2}{m_R^2}(\partial_\mu\vartheta\,\partial^\mu\vartheta)^2 + \cdots\right]. \tag{6.10}$$

The field equations for ϑ predicted by this action are

$$\partial_\mu\left\{\partial^\mu\vartheta\left[1 - \frac{2}{m_R^2}(\partial_\nu\vartheta\,\partial^\nu\vartheta) + \cdots\right]\right\} = 0\,, \tag{6.11}$$

which admits the solution $\vartheta = \vartheta_0 + \omega t$ for which $\partial_\mu\vartheta = \omega\,\delta^0_\mu$ is constant.

Applying Noether's theorem to this action with the low-energy shift symmetry $\vartheta \to \vartheta + c$ ensures the existence of the conserved current,

$$j^\mu_{\rm eff} = 2v^2\partial^\mu\vartheta\left[1 - \frac{2}{m_R^2}(\partial_\nu\vartheta\,\partial^\nu\vartheta) + \cdots\right], \tag{6.12}$$

for which the equations of motion (6.11) clearly imply $\partial_\mu j^\mu_{\rm eff} = 0$. Evaluating this at the rolling solution, $\vartheta = \omega t$ then gives

$$j^\mu_{\rm eff} = -2v^2\omega\left(1 + \frac{2\omega^2}{m_R^2} + \cdots\right)\delta^\mu_0 = -2\omega\left(v^2 + \frac{2\omega^2}{\lambda} + \cdots\right)\delta^\mu_0\,, \tag{6.13}$$

in agreement with (6.8).

To calculate the energy density of this solution in the effective theory compute the effective Hamiltonian density,

$$\mathcal{H}_{\rm eff} = \pi_{\rm eff}\,\dot\vartheta - \mathfrak{L}_{\rm eff}\,, \tag{6.14}$$

where the canonical momentum is defined by

$$\pi_{\rm eff} := \frac{\delta S_{\rm eff}}{\delta\dot\vartheta} = 2v^2\left(\dot\vartheta + \frac{2\dot\vartheta^3}{m_R^2} + \cdots\right). \tag{6.15}$$

Using this the Hamiltonian density becomes

$$\mathcal{H}_{\rm eff} = v^2\dot\vartheta^2 + v^2\nabla\vartheta\cdot\nabla\vartheta + \frac{3\lambda v^4\,\dot\vartheta^4}{m_R^4} + \cdots\,, \tag{6.16}$$

and so the energy density obtained by evaluating this at $\dot\vartheta = \omega$ is

$$\varepsilon_{\rm eff} = v^2\omega^2 + \frac{3\lambda v^4\,\omega^4}{m_R^4} + \cdots = v^2\omega^2 + \frac{3\omega^4}{\lambda} + \cdots\,, \tag{6.17}$$

in agreement with (6.9), including the last term's factor of 3.

These calculations reveal that it is the first subleading term, $(\partial_\mu\xi\,\partial^\mu\xi)^2$, in $\mathfrak{L}_w$ that brings the news to the EFT about the adjustment of ϱ_0 and the centripetal acceleration in the UV theory. Furthermore, the matching performed in previous sections gives precisely the value for the effective coupling needed to get the answer right. This despite the fact that these earlier matching calculations obtain the coupling's value using scattering amplitudes, rather than classical background evolution.

Because it is the higher-derivative terms that carry the information about the ω-dependent response of the system in the EFT, it is also clear that a purely EFT description of this response assumes $\omega \ll m_R$ if it is to neglect the contributions of higher-dimension interactions.

6.2 EFTs and background solutions ◇

The toy-model example just considered shows that the field equations of the Wilsonian effective theory properly capture the time-dependence of slowly evolving classical background solutions of the full UV theory (see §6.3 for the analogous story about fluctuations about such backgrounds). This section asks more generally when the background solutions within an EFT should (and shouldn't) be expected to reproduce the solutions of the underlying UV theory.

The main message is that the space of solutions solving the field equations of an EFT overlaps with (but neither contains nor is contained within) the space of solutions for the UV theory. Not surprisingly, solutions to the EFT's field equations do include those of the full theory that evolve adiabatically but not those that evolve too quickly. But the EFT field equations also have solutions that are not related to those of the full theory (and are typically singular in the limit that the UV scale goes to infinity). To justify these statements (and make them more precise), and to see how to identify which EFT solutions are relevant

to the full theory's low-energy limit (and which are not), it is worth recalling some features of the discussion in §2.1 and §2.2.

6.2.1 Adiabatic equivalence of EFT and full evolution

Why should background solutions for the full theory and low-energy theories agree with one another, and precisely which equations do backgrounds solve?

For the full theory the relation between the field expectation value and the action is given by (2.14) and (2.18), which (in the absence of an external current) state that

$$\varphi^i(x) = \langle \phi^i(x) \rangle = \frac{\langle \text{out}|\phi^i(x)|\text{in}\rangle}{\langle \text{out}|\text{in}\rangle} \tag{6.18}$$

satisfies

$$\left(\frac{\delta\Gamma[\phi]}{\delta\phi^i(x)}\right)_{\phi=\varphi} = 0\,, \tag{6.19}$$

where $\Gamma[\phi]$ is the generator of 1PI graphs, $|\text{in}\rangle$ is the vacuum state in the remote past and $|\text{out}\rangle$ is the vacuum state in the remote future (which need not be the same in the presence of time-dependent fields in between). Crucially, the derivation of this statement assumes $|\text{in}\rangle$ evolves adiabatically[1] as a function of the evolving background quantities as it eventually turns into $|\text{out}\rangle$ [17].

For a system whose fields divide into heavy and light degrees of freedom, $\{\phi^i(x)\} = \{h^a(x), \ell^\alpha(x)\}$, equation (6.19) holds both for $\langle h^a(x)\rangle$ and $\langle \ell^\alpha(x)\rangle$,

$$\left(\frac{\delta\Gamma[h,\ell]}{\delta h^a(x)}\right)_{\langle h\rangle,\langle\ell\rangle} = \left(\frac{\delta\Gamma[h,\ell]}{\delta \ell^\alpha(x)}\right)_{\langle h\rangle,\langle\ell\rangle} = 0\,. \tag{6.20}$$

The closest analogue of $\Gamma[h,\ell]$ for the low-energy part of this theory is the 1LPI generator, $\Gamma_{\text{le}}[\ell]$, introduced in §2.2.3, for which external currents are only turned on for the light fields. Chasing through the definitions implies the relation between $\Gamma[h,\ell]$ and $\Gamma_{\text{le}}[\ell]$ is given by (2.45), which states

$$\Gamma_{\text{le}}[\ell] = \Gamma[h_{\text{le}}(\ell),\ell] \quad \text{where} \quad \left(\frac{\delta\Gamma}{\delta h^a}\right)_{h=h_{\text{le}}(\ell)} = 0\,, \tag{6.21}$$

and so $h_{\text{le}}(\ell) = \langle h\rangle$ is regarded as a function of the specified value for the light field.

Varying this expression with respect to $\ell^\alpha(x)$ — keeping in mind (6.20) — then shows that $\langle \ell^\alpha(x)\rangle$ satisfies the purely low-energy condition

$$\left(\frac{\delta\Gamma_{\text{le}}[\ell]}{\delta\ell^\alpha(x)}\right)_{\langle\ell\rangle} = \left(\frac{\delta\Gamma[h,\ell]}{\delta\ell^\alpha(x)}\right)_{\langle h\rangle,\langle\ell\rangle} = 0\,. \tag{6.22}$$

This shows that $\langle \ell^\alpha(x)\rangle$ can equally well be computed by extremizing $\Gamma[h,\ell]$ in the full theory or by extremizing $\Gamma_{\text{le}}[\ell]$ of the low-energy sector alone. But a central part of this argument is the adiabatic assumption that underpins the starting point, (6.19).

[1] Adiabatic evolution here means the solutions are solved as if they are static functions of external parameters, like the currents J, but these external parameters are themselves allowed to evolve very slowly with time. This is as opposed, for instance, to having levels cross or some other drama between $t = \pm\infty$.

Classical limit

Time-dependent backgrounds are most commonly enountered within a semiclassical approximation, wherein the time-dependent background is the dominant, classical, configuration: $\langle \phi^i(x) \rangle \simeq \phi^i_c(x)$. In this case the above argument goes through order-by-order in the semiclassical expansion, with the leading (classical) contribution being given by

$$\Gamma[h,\ell] \simeq S[h,\ell] \quad \text{and} \quad \Gamma_{\rm le}[\ell] \simeq S_{\rm w}[\ell]\,. \tag{6.23}$$

Here S is the classical action for the full theory and $S_{\rm w}$ is the Wilson action defined in §2.3 and given in the classical approximation by $S_{\rm w}[\ell] \simeq S[h_c(\ell),\ell]$, where $h_c(\ell)$ is found by solving $\delta S[h,\ell]/\delta h = 0$ as a function of a specified light field ℓ (*c.f.* eq. (6.21)). Then (6.22) becomes a statement relating the classical solutions for these two actions:

$$\left(\frac{\delta S_{\rm w}[\ell]}{\delta \ell^\alpha(x)}\right)_{\ell_c} = \left(\frac{\delta S[h,\ell]}{\delta \ell^\alpha(x)}\right)_{h_c,\ell_c} = 0\,, \tag{6.24}$$

Again it seems clear that solutions to the equations of motion for the classical Wilson action reproduce the light-field part of the solutions to the classical equations of the full theory. But how does the adiabatic requirement arise in this purely classical argument? To understand this it helps to consider an example, and our stalwart toy model once more comes in useful.

Example: the toy model

To this end revisit the derivation given in §2.2.3, starting with eq. (2.47) (reproduced here),

$$S[\xi,\chi] = -\int \mathrm{d}^4x \left[\frac{1}{2}\,\partial_\mu\chi\partial^\mu\chi + \frac{1}{2}\left(1+\frac{\chi}{\sqrt{2}\,v}\right)^2 \partial_\mu\xi\partial^\mu\xi + V(\chi)\right], \tag{6.25}$$

in which the heavy field χ is explicitly integrated out within the classical approximation, using the classical potential

$$V(\chi) = \frac{m_R^2}{2}\chi^2 + \frac{\lambda\, v}{2\sqrt{2}}\chi^3 + \frac{\lambda}{16}\chi^4\,. \tag{6.26}$$

To compute $S_{\rm w}[\xi] \simeq S[\xi,\chi_c(\xi)]$ classically requires solving for $\chi_c(\xi)$ using eq. (2.49), which to leading nontrivial order is approximately

$$\left(-\Box + m_R^2\right)\chi_c \simeq -\frac{1}{\sqrt{2}\,v}\,\partial_\mu\xi\partial^\mu\xi + \cdots\,, \tag{6.27}$$

where the higher-order terms are not required to make the point soon to follow. The solution to this equation used in §2.2.3 is

$$\chi_c \simeq -\frac{1}{\sqrt{2}\,v m_R^2}\,(\partial_\mu\xi\partial^\mu\xi) + \cdots\,, \tag{6.28}$$

and substituting this into $S[\chi,\xi]$ leads to the $(\partial_\mu\xi\,\partial^\mu\xi)^2$ interaction found earlier for $S_{\rm w}$.

Now comes the main point: the transition from (6.27) to (6.28) given in §2.2.3 proceeds as if the solution to (6.27) were *unique*. But we know the general solution to (6.27) is

really given by the sum of (6.28) and an arbitrary solution, χ_h, to the homogenous equation $(-\Box + m_R^2)\chi_h = 0$.

It is the adiabatic approximation that dictates choosing the solution $\chi_h = 0$, and it does so because χ_h is generally a sum of modes whose time-dependence is given by e^{-iEt} where $E \geq m_R$. Such modes could be excited if the time evolution of the background were sufficiently rapid, but are not directly excited for slow adiabatic evolution. Of course, interactions can also introduce modes with higher mode energies starting only from those at lower energies, because once interactions are included only the total energy (including interactions) is strictly conserved, and not just the energy of isolated linearized modes (more about this later).

6.2.2 Initial data and higher-derivative instabilities ♣

The previous section's observation that classical solutions to the Wilsonian equations of motion do not precisely overlap with those of the underlying UV theory is actually an EFT feature rather than a bug. It is, of course, reasonable that the UV theory should contain solutions not in the low-energy theory, since the latter cannot capture those solutions of the full theory that evolve rapidly (and so do not exclusively involve low-energy modes). This section argues that there are also solutions to the low-energy equations that do not correspond to solutions of the full UV theory, and that this is also a good thing.

Extra unwanted solutions arise in the low-energy theory because the Wilson action generically contains all possible interactions allowed by the low-energy field content and symmetries. As a result it usually contains terms for which the fields appear multiply differentiated. For instance at the six-derivative level for the toy model one can have

$$\mathfrak{Q}_{\rm hd} = -c_{61}\, X^3 - c_{62}\, X\, \vartheta_{\mu\nu}\, \vartheta^{\mu\nu}\,, \tag{6.29}$$

where

$$X := -\partial_\lambda\vartheta\, \partial^\lambda\vartheta \quad \text{and} \quad \vartheta_{\mu\nu} := \partial_\mu\partial_\nu\vartheta\,, \tag{6.30}$$

while c_{6n} denotes the relevant effective coupling.

What is important for the purposes of counting solutions is that the last term of (6.29) involves more than two time derivatives, as is most easily seen by temporarily ignoring all spatial derivatives. In this case (6.29) contains the term $\mathfrak{Q}_{62} = c_{62}\, \ddot\vartheta^2\dot\vartheta^2$, whose variation is

$$\frac{\delta\mathfrak{Q}_{62}}{2c_{62}} = \Big[\ddot\vartheta\dot\vartheta^2\Big]\delta\ddot\vartheta + \Big[\dot\vartheta\ddot\vartheta^2\Big]\delta\dot\vartheta = \Big[\dddot\vartheta\,\dot\vartheta^2 + 4\,\ddot\vartheta\,\dddot\vartheta\,\dot\vartheta + \ddot\vartheta^3\Big]\delta\vartheta\,, \tag{6.31}$$

and the last equality drops surface terms coming from several integrations by parts. Because the field equation obtained from $\delta S/\delta\vartheta = 0$ involves fourth derivatives of ϑ its integration requires more initial data than usual (it requires initial values for $\ddot\vartheta$ and $\dddot\vartheta$ in addition to the usual initial values for ϑ and $\dot\vartheta$).

Related to the requirement for more initial data is the observation that the general solutions to higher-derivative field equations involve more integration constants and so involve more than the usual two-parameter class of solutions appropriate to second-order field equations.

What is more troubling: these new solutions almost always include unstable runaway solutions. The generic appearance of instability is most easily seen from the canonical formulation [90, 91, 92], for which all field equations are written in terms of single time derivatives by introducing new canonical 'momenta'. The argument is made here for lagrangians of the form $\mathfrak{L} = \mathfrak{L}(\phi, \dot\phi, \ddot\phi)$, but generalizes to the inclusion in $\mathfrak{L}$ of still higher derivatives as well.

The lagrangian $\mathfrak{L} = \mathfrak{L}(\phi, \dot\phi, \ddot\phi)$ has higher-order equations of motion given by

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left[\frac{\partial \mathfrak{L}}{\partial \ddot\phi}\right] - \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial \mathfrak{L}}{\partial \dot\phi}\right] + \frac{\partial \mathfrak{L}}{\partial \phi} = 0\,. \tag{6.32}$$

To set up a canonical formulation for these equations define the new variable $\psi = \dot\phi$ so that $\mathfrak{L} = \mathfrak{L}(\phi, \psi, \dot\psi)$ and define the standard (π) and new (ζ) canonical momenta by

$$\pi(\phi, \psi, \dot\psi) := \frac{\partial \mathfrak{L}}{\partial \dot\phi} = \frac{\partial \mathfrak{L}}{\partial \psi} \quad \text{and} \quad \zeta(\phi, \psi, \dot\psi) := \frac{\partial \mathfrak{L}}{\partial \ddot\phi} = \frac{\partial \mathfrak{L}}{\partial \dot\psi}\,, \tag{6.33}$$

of which it is assumed the defining equation for ζ can be solved for $\dot\psi$ to give an expression of the form $\dot\psi = \dot\psi(\phi, \psi, \zeta)$.

With these choices the hamiltonian density

$$\begin{aligned}\mathcal{H}(\phi, \psi; \pi, \zeta) &:= \pi\,\dot\phi + \zeta\,\dot\psi - \mathfrak{L}(\phi, \psi, \dot\psi) \\ &= \pi\,\psi + \zeta\,\dot\psi(\phi, \psi, \zeta) - \mathfrak{L}[\phi, \psi, \dot\psi(\phi, \psi, \zeta)]\,,\end{aligned} \tag{6.34}$$

generates the equations of motion (6.32) through the first-order system

$$\dot\phi = \frac{\partial \mathcal{H}}{\partial \pi}\,, \quad \dot\psi = \frac{\partial \mathcal{H}}{\partial \zeta}\,, \quad \dot\pi = -\frac{\partial \mathcal{H}}{\partial \phi} \quad \text{and} \quad \dot\zeta = -\frac{\partial \mathcal{H}}{\partial \psi}\,. \tag{6.35}$$

For stability arguments it is crucial that $\mathcal{H}$ also be conserved and bounded from below, since when these are both true the configuration minimizing $\mathcal{H}$ must be stable. In the present case conservation goes through as usual (provided $\mathfrak{L}$ does not itself depend explicitly on t) because

$$\dot{\mathcal{H}} = \frac{\partial \mathcal{H}}{\partial \phi}\,\dot\phi + \frac{\partial \mathcal{H}}{\partial \psi}\,\dot\psi + \frac{\partial \mathcal{H}}{\partial \pi}\,\dot\pi + \frac{\partial \mathcal{H}}{\partial \zeta}\,\dot\zeta = 0 \tag{6.36}$$

with the last equality using (6.35). The generic problem with higher-derivative theories is that $\mathcal{H}$ is not bounded from below, as is seen because (6.34) shows $\mathcal{H}$ is *linear* in the variable π.

To obtain an intuition for how such an instability arises more concretely, consider the following quadratic (but higher-order) toy lagrangian [93]:

$$L = \frac{1}{2}\,\dot\vartheta^2 + \frac{1}{2M^2}\,\ddot\vartheta^2\,, \tag{6.37}$$

whose variation $\delta L = 0$ gives the linear equation of motion

$$-\ddot\vartheta + \frac{\ddddot\vartheta}{M^2} = 0\,. \tag{6.38}$$

The general solution to this equation is

$$\vartheta = A + Bt + Ce^{Mt} + De^{-Mt}\,, \tag{6.39}$$

where A, B, C and D are integration constants. This has an unstable runaway form apart from the special initial condition that chooses $C = 0$. The generic unstable mode encountered for higher-derivative theories is often called the *Ostrogradsky ghost*.

The question of why this issue is not a problem for the low-energy EFT is addressed below, after first a brief detour.

A Galileon aside

Although the above arguments show that the introduction of higher-derivative interactions generically leads to instability, it is also true that not all higher-derivative effective interactions need do so. There are two kinds of relatively benign interactions of this type.

The first type of benign higher-derivative interaction consists of those that are redundant, in the sense made more precise in §2.5. As described there interactions are redundant if they arise as a total derivative or if they can be removed through a local field redefinition. An example of these types of redundancy for the toy model would be a term like $\vartheta^{\mu\nu}\vartheta_{\mu\nu}$ — with $\vartheta_{\mu\nu}$ defined in (6.30) — since this can be rewritten using

$$\vartheta^{\mu\nu}\vartheta_{\mu\nu} = \partial_\mu\Big(\vartheta^{\mu\nu}\partial_\nu\vartheta\Big) - (\partial^\nu\Box\vartheta)\partial_\nu\vartheta = \partial_\mu\Big(\vartheta^{\mu\nu}\partial_\nu\vartheta - \Box\vartheta\,\partial^\mu\vartheta\Big) + (\Box\vartheta)^2\,. \tag{6.40}$$

The first term on the right-hand side is a total derivative and the second term vanishes when the lowest-order field equations, $\Box\vartheta = 0$, are used showing that it can be removed to this order in the derivative expansion by performing a field redefinition of the form $\delta\vartheta \propto \Box\vartheta$.

But there is also a second way that nominally higher-derivative interactions can avoid introducing new solutions and instabilities. This is because there are a handful of higher-derivative lagrangian interactions for which the corresponding higher-derivative terms in the field equations happen to cancel. In four spacetime dimensions, using one scalar field ϕ, the most general such an interaction (up to total derivatives) turns out to be a linear combination of the following *Galileon* interactions [94, 95, 96]

$$\begin{aligned}
\mathfrak{L}_{G2} &:= G_2(\phi, X) \\
\mathfrak{L}_{G3} &:= G_3(\phi, X)\,\Box\phi \\
\mathfrak{L}_{G4} &:= G_4(\phi, X)\Big[(\Box\phi)^2 - \phi_{\mu\nu}\phi^{\mu\nu}\Big] \\
\mathfrak{L}_{G5} &:= G_5(\phi, X)\Big[(\Box\phi)^3 - 3\phi_{\mu\nu}\phi^{\mu\nu}\,\Box\phi + 2\phi^{\mu\nu}\phi_{\nu\lambda}\phi^{\lambda}{}_{\mu}\Big],
\end{aligned} \tag{6.41}$$

where, following earlier notation, these use the definitions $X := -\partial_\lambda\phi\,\partial^\lambda\phi$ and $\phi_{\mu\nu} := \partial_\mu\partial_\nu\phi$. Here $G_i(\phi, X)$ with $i = 2, ...5$ are four arbitrary functions of two arguments. For generic G_i none of these is a total derivative and for all G_i they contribute only terms involving at most two time derivatives to the scalar field equations.

For low enough derivative order it sometimes happens that the most general form for the Wilsonian action is a special case of (6.41) [97]. For instance, for the toy model the most general terms arising out to four-derivative level can be written (up to a total derivative) as a linear combination of X^2, $X\Box\vartheta$ and $(\Box\vartheta)^2$. The last two of these vanish when $\Box\vartheta = 0$, and so can be removed by performing the field redefinition $\delta\vartheta = a\,\Box\vartheta + b\,X$ for appropriate choices for the constants a and b. The remaining term is a special case of the first of (6.41), with $G_2 = \frac{1}{2}\,X + c_4\,X^2$ with c_4 the constant given in eq. (1.12).

Similarly, the most general terms involving six derivatives are given by (6.29), once total derivatives and redundant interactions involving $\Box\vartheta$ are removed. Because these differ from the term in (6.41) just by $\Box\vartheta$ terms, at six-derivative level the shift-symmetric lagrangian also has the form of (6.41), up to field redefinitions. Up to six-derivative level the corresponding terms have $G_2 = \frac{1}{2}X + c_4 X^2 - c_{61} X^3$ and $G_4 = c_{62} X$.

Of course, the terms in (6.41) involve at most six derivatives not involved in a factor of X, and so eventually terms should arise with sufficient numbers of derivatives to preclude their being put into the Galileon form. And for theories with more general field content more structures are possible at each order, so there is no broad expectation that a generic system can always be written, at all orders in the derivative expansion, like (6.41) or its generalizations. How are the instabilities associated with higher-derivative interactions dealt with then?

A more general argument

If EFTs generically involve higher-derivative effective interactions and if these interactions generically produce unstable solutions, how can a generic Wilson action hope to describe the time-evolution of a UV theory that is known to be stable (such as the toy model)?

A key step in the development of the Wilson action was the expansion in powers of $1/M$; in particular it is only after this expansion that the EFT is described by a *local* lagrangian density. Because of this a local Wilsonian action should only be expected to capture the properties of the underlying UV-completion order-by-order in powers of $1/M$. This is true in particular when seeking time-dependent solutions, which should only be trusted to the extent that they fall within the regime of the $1/M$ expansion.[2]

So a crucial feature of the 'new' solutions (including in particular the runaways) associated with the new higher-derivative terms is that they do *not* arise as a series in powers of $1/M$. They do not because they are singular perturbations of the zeroth-order differential equation (because it is the highest-derivative terms of the field equations that are come multiplied with nonzero powers of $1/M$).

This is seen explicitly in the solution (6.39) and field equation (6.38) of the simple higher-derivative action given in (6.37). Only the two-parameter family of these solution obtained using $C = D = 0$ go over to the solutions to the lowest-order field equation, obtained from the $M \to \infty$ lagrangian, $L_0 = \frac{1}{2}\dot{\vartheta}^2$; the other solutions are not captured at any finite order of $1/M$ because for them the $\ddot{\vartheta}^2$ and $\ddot{\vartheta}^2$ terms are comparably large. This is manifest in exponential solutions like $\exp(\pm Mt)$ of eq. (6.39), which have an essential singularity as $M \to \infty$ and are not described at any order by a series in $1/M$.

The lesson is this: a local Wilsonian EFT only aspires to capture the full theory order-by-order in $1/M$, and so any predictions it makes that fall outside of a $1/M$ expansion should be regarded as spurious. Such predictions should not be expected to capture properties of the underlying UV theory.

[2] This is one of those arguments that has been 'in the air' and widely known by those who know for decades, and because of that nowhere written down (almost nowhere; ref. [93] was written to record the argument, which at the time had not percolated into relatively new communities for EFT arguments).

Well-posed evolution

Just having equations of motion that are second-order in time does not mean one can relax, however, since in some circumstances time evolution can nonetheless be difficult to evaluate. This could be because an evolution equation's caustics begin to intersect or because short-wavelength modes grow too quickly even if not initially present. When this happens an initial configuration with small gradients can be driven into a regime of large derivatives, and so beyond the reach of EFT methods.

Sometimes this kind of behaviour is the right answer. The collection and focussing of light by lenses is an example where this is true, as is the phenomenon of gravitational collapse (for which an initially diffuse and low-energy cloud of dust becomes gravitationally compressed, possibly into a singularity with arbitrarily large derivatives). But there are also many other examples of this phenomenon throughout physics, such as the turbulent cascade of fluid energies down to small distances, or the development of caustics for the propagation of light in a medium, or the development of shock fronts within hot materials.[3]

In these situations energy conservation in itself does not prevent moving from long-wavelength initial conditions towards those with larger gradients, and so towards a breakdown of the low-energy description. This need not be a problem of principle for EFT methods (depending on how fast it happens) since nothing says that a system that starts in a long-wavelength regime must remain there. Indeed, if the underlying system moves from smooth configurations towards variations over microscopic scales then the EFT should be able to track the early part of this evolution before showing signs of breaking down.

Two features that lend themselves to this kind of breakdown are nonlinear field equations and the breaking of Lorentz invariance, features that are generic in real applications with time-dependent backgrounds. Both of these undermine the protection energy conservation naively gives against generating short-wavelength modes from long-wavelength initial data. Nonlinearities do so by allowing many low-energy modes to combine into a higher-energy one. Breaking Lorentz invariance can allow large mode momenta to coexist with low mode energies even without nonlinearities, and so can also interfere with the ability to discriminate against short-wavelength modes using only low energy as a criterion.

Studies of nonlinear classical field equations often frame the issue of the growth of small-wavelength modes in terms of the well-posedness of the initial-value problem [100]. An initial-value problem is said to be *locally well-posed* if, given suitable initial data, there exists a unique solution of the equation of motion, and that the space of solutions depends continuously on the initial data. Well-posedness is local inasmuch as the solution is only required to exist for some nonzero, though possibly very small, time.

An example of the kind of thing that would make an initial-value problem ill-posed would be if modes of wave-number $\mathbf{k}$ were to grow in time as quickly as $\exp(+|\mathbf{k}|t)$, say. If the limit $|\mathbf{k}| \to \infty$ were allowed this would represent an arbitrarily fast growth, undermining the continuity of the solutions regarded as functions of their initial data. From an EFT perspective things are never quite this bad, however, since within an effective theory $|\mathbf{k}|$ is bounded to be smaller than some UV scale M. So whereas mode growth can happen, the timeframe for catastrophic growth in an EFT is usually not arbitrarily short. But this

[3] See *e.g.* [98] and [99] for discussions of this issue with applications to gravity and fluids, respectively.

might be cold comfort if it were instead to occur on a UV time-scale like M. As previous sections make clear, evolution over time-scales as short as M^{-1} lies beyond what an EFT can capture.

Well-posedness can also be an important issue even when only asking pragmatic questions well within the EFT regime (for which physical quantities do not evolve on microscopic time scales). This is because in practice evolution is calculated only approximately, perhaps numerically by breaking space and time into a discrete lattice. Such approximations necessarily introduce short-distance errors into the initial conditions and evolution equations, which for all intents and purposes play the role of unknown UV physics at the regulator scale Λ. If these regulation errors were to grow over time-scales as short as Λ^{-1} then this spurious growth could quickly swamp the much slower evolution of the physical system being modelled by the EFT description. It is the desire to integrate effective-field equations in nonlinear settings that makes discussions of well-posedness more than a purely mathematical exercise.

Although not a problem for well-posed evolution, these issues mean that approximate methods typically require some sort of smoothing procedure for ill-posed problems [102] to suppress spurious regulator-scale variations (for a discussion of these issues for the toy model considered here see [103]). Whether such smoothing is necessary requires a diagnosis of the well-posedness of EFT field equations.

Well-posedness for a nonlinear theory is ensured if its field equations are strongly hyperbolic.[4] Sadly, the field equations for many EFTs are known not to be strongly hyperbolic even if the underlying UV theory is. EFTs can run into trouble in this way — even those lying in the Galileon class discussed earlier [101] — because the derivatives appearing within effective interactions modify the character of the second-derivative terms on which hyperbolicity is based, perhaps as a function of the size of (or variation in) a background field.

Since any spurious regulator dependence is a special case of UV physics, in principle it can be absorbed into the values of an EFT's effective couplings. The problem is how to do this in practice, numerically and on the fly. As of this writing (2018) the issue of how to optimally simulate the time-evolution predicted order-by-order in an EFT's low-energy expansion is not yet settled, though is under active study.

6.3 Fluctuations about evolving backgrounds ♠

Earlier sections in this chapter mostly focus on how the background evolves and how this is captured by an effective Wilsonian description. But there can also be interest in the properties of fluctuations about non-static background configurations, and this section explores some of the ways that fluctuations about time-evolving backgrounds differ from those about static vacua. The fluctuations of interest could either describe nearby solutions within a purely classical problem, or be full-on quantum fluctuations. Which is relevant in

[4] A hyperbolic system is strongly hyperbolic if there is a norm for solutions whose behaviour at time t is bounded by the initial value of the same norm multiplied by a function of time that is independent of the initial data.

any particular application can be determined using a power-counting analysis such as that given in §3.

Part of the practical interest in studying EFTs for fluctuations about evolving backgrounds (for relativistic systems) comes from cosmology [104]. As described in more detail in §10.2, quantitative predictions for fluctuations are pressing in cosmology because in the modern understanding the large-scale distribution of matter and radiation throughout the universe arises as the gravitational amplification of small primordial field fluctuations occurring within an expanding spacetime. This allows precise predictions of the properties of these fluctuations to be compared in detail with the wealth of modern cosmological observations.

6.3.1 Symmetries in an evolving background

Time-dependent backgrounds typically preserve fewer symmetries than do static vacua. For instance for the toy model with a homogeneous time-dependent background, $\vartheta(t)$, the background breaks both time-translation invariance and Lorentz invariance, while preserving rotational symmetry and invariance under spatial translations. Since the symmetries are broken by a field configuration the breaking can be considered to be spontaneous, though of spacetime symmetries rather than internal ones (for a recent systematic discussion of the issues see, for example, [105]).

The consequences of this symmetry-breaking pattern for fluctuations follow the general rules outlined earlier for spontaneous symmetry breaking. In particular, fluctuations fall into linear representations only of the unbroken subgroup of symmetries that leave the background invariant. For homogeneous time-dependent backgrounds this means that fluctuations can be labelled by their spin (*i.e.* representation of the field under rotations) and linear momentum (representation under translations). Total momentum and angular momentum are conserved by virtue of the background's invariance under spatial translations and rotations.

Other consequences of Poincaré invariance for static Lorentz-invariant vacua do not carry over to fluctuations about homogeneous time-dependent backgrounds. In principle, the breaking of time-translations by the background means that energy is not strictly conserved for the fluctuations. (That is to say: even if energy is conserved for the whole system – background plus fluctuations – in general there can be energy transfer between the two for time-dependent backgrounds, making the energy of fluctuations themselves not strictly conserved.)

Because Wilsonian actions only capture the time-dependence of adiabatic evolution in the UV theory, when EFT methods are useful it is possible to define a time-dependent energy satisfying

$$H(t)|\mathbf{k},\sigma\rangle = E(k,t)|\mathbf{k},\sigma\rangle\,, \tag{6.42}$$

acting on fluctuation states, where H might parametrically depend on time. Alternatively, the time-evolution of field mode functions can be approximately written as

$$u(t,\mathbf{x}) = v(\mathbf{x})\,\exp\left[-\mathrm{i}\int_{t_0}^{t}\mathrm{d}\tau E(k,\tau)\right]. \tag{6.43}$$

As described earlier, it is this energy that implicitly is used to distinguish low-energy from high-energy states for EFT applications with time-dependent backgrounds. These expressions use rotation invariance, which ensures a mode's energy eigenvalue (or dispersion relation) depends only on the magnitude $k = |\mathbf{k}|$.

Finally, for fluctuations about time-dependent backgrounds the dispersion relation $E(k)$ can differ from the Lorentz-invariant result $\sqrt{k^2 + m^2}$. For instance the explicit calculations to follow for the time-dependent toy model example considered above show the Goldstone mode propagates with dispersion relation $E(k) = kc_s + O(k^2)$ with $1 - c_s = \delta_c$ being a calculable positive function of system parameters (like ω, λ and v in the toy model).

Fluctuations in the toy model

To make the story concrete this section examines how fluctuations around a time-dependent background behave in the toy model of §1.1, both in the full theory and in its low-energy Wilsonian incarnation.

To this end, in the full theory expand $\phi = \varphi(t) + \tilde{\phi}$ where $\varphi(t) = \varrho_0\, e^{i\omega t}$ is the background solution considered above, with (6.7) implying $\varrho_0^2 = v^2 + 2\omega^2/\lambda$. With this choice the lagrangian can be expanded in powers of $\tilde{\phi}$, so $\mathfrak{L} = \mathfrak{L}^{(0)} + \mathfrak{L}^{(1)} + \mathfrak{L}^{(2)} + \mathfrak{L}^{(3)} + \mathfrak{L}^{(4)}$, where

$$\begin{aligned}
\mathfrak{L}^{(0)} &= \omega^2\varrho_0^2 - \frac{\omega^4}{\lambda} = \omega^2 v^2 + \frac{\omega^4}{\lambda} \\
\mathfrak{L}^{(1)} &= \sqrt{2}\,\varrho_0\,\omega\,\frac{\mathrm{d}}{\mathrm{d}t}\Big[-\tilde{\phi}_R s_t + \tilde{\phi}_I c_t\Big] \\
\mathfrak{L}^{(2)} &= -\frac{1}{2}\left(\partial_\mu\tilde{\phi}_R\,\partial^\mu\tilde{\phi}_R + \partial_\mu\tilde{\phi}_I\,\partial^\mu\tilde{\phi}_I\right) - \frac{1}{2}\begin{pmatrix}\tilde{\phi}_R\\ \tilde{\phi}_I\end{pmatrix}^T\begin{pmatrix}\omega^2+\lambda\varrho_0^2\,c_t^2 & \lambda\varrho_0^2\,c_t s_t\\ \lambda\varrho_0^2\,c_t s_t & \omega^2+\lambda\varrho_0^2\,s_t^2\end{pmatrix}\begin{pmatrix}\tilde{\phi}_R\\ \tilde{\phi}_I\end{pmatrix},
\end{aligned} \tag{6.44}$$

and so on for higher powers of $\tilde{\phi}$, where $\tilde{\phi} = \frac{1}{\sqrt{2}}(\tilde{\phi}_R + i\tilde{\phi}_I)$ while $c_t := \cos\omega t$ and $s_t := \sin\omega t$. The term linear in $\tilde{\phi}$ can be dropped for most purposes because it is a total derivative — as is always true when expanding about a classical solution.

Although the quadratic term seems to involve a standard kinetic piece plus lots of oscillatory time-dependence in the mass term, this is actually deceptive since the eigenvalues of the mass matrix are not time-dependent at all:

$$m_+^2 := \omega^2 + \lambda\varrho_0^2 = 3\omega^2 + \lambda v^2 \quad \text{and} \quad m_-^2 := \omega^2\,. \tag{6.45}$$

The oscillatory time-dependence seen in (6.44) and the tempting interpretation of (6.45) as nonzero masses can be misleading if they are used too naively when drawing physical consequences (such as the existence of an energy gap for fluctuations at zero momentum). They are misleading because the shorthand that allows a straightforward inference of physical quantities like masses from quadratic terms in an action breaks down in this particular choice of basis fields, $\tilde{\phi}_R$ and $\tilde{\phi}_I$. The problem arises because these basis fields are fixed in time while the physical basis of mass eigenstates rotates in field space with angular frequency ω.

Drawing inferences using the fields $\tilde{\phi}_R$ and $\tilde{\phi}_I$ might be warranted if there were a physical reason for choosing this basis — *e.g.* if other sectors of the theory were to break the

$U(1)$ symmetry, such as if perhaps only ϕ_I were to couple to observable particles. Otherwise, performing the time-dependent rotation required to reach the mass basis transfers the effects of the time-dependent background into the kinetic part of the fluctuation fields, suggesting very different kinds of observable consequences.

Simpler than performing this time-dependent rotation is to use directly the fluctuation fields $\tilde{\chi}$ and $\tilde{\xi}$ defined by

$$\phi = \left(\varrho_0 + \frac{\tilde{\chi}}{\sqrt{2}}\right) \exp\left[\frac{\mathrm{i}\xi}{\sqrt{2}\varrho_0}\right], \tag{6.46}$$

with $\xi = \sqrt{2}\,\varrho_0\,\omega t + \tilde{\xi}$, since in this case the lagrangian expansion becomes (*c.f.* eqs. (1.24) and (1.25))

$$\mathfrak{L} = -\frac{1}{2}\,\partial_\mu\tilde{\chi}\partial^\mu\tilde{\chi} - \frac{1}{2}\left(1 + \frac{\tilde{\chi}}{\sqrt{2}\,\varrho_0}\right)^2 \partial_\mu\xi\partial^\mu\xi - V(\tilde{\chi})\,, \tag{6.47}$$

with

$$-\partial_\mu\xi\,\partial^\mu\xi = 2\omega^2\varrho_0^2 + 2\,\sqrt{2}\,\omega\,\varrho_0\,\partial_t\tilde{\xi} - \partial_\mu\tilde{\xi}\,\partial^\mu\tilde{\xi}\,, \tag{6.48}$$

and

$$V(\tilde{\chi}) = \frac{\lambda}{4}\left(\frac{2\omega^2}{\lambda} + \sqrt{2}\,\varrho_0\,\tilde{\chi} + \frac{\tilde{\chi}^2}{2}\right)^2 . \tag{6.49}$$

Expanding this lagrangian in powers of $\tilde{\chi}$ and $\tilde{\xi}$ then gives the same expression as before for $\mathfrak{L}^{(0)}$; a total derivative for the linear terms; and the following quadratic term

$$\mathfrak{L}^{(2)} = -\frac{1}{2}\,\partial_\mu\tilde{\chi}\partial^\mu\tilde{\chi} - \frac{1}{2}\partial_\mu\tilde{\xi}\,\partial^\mu\tilde{\xi} + 2\omega\,\tilde{\chi}\,\partial_t\tilde{\xi} - \frac{1}{2}\lambda\varrho_0^2\,\tilde{\chi}^2\,. \tag{6.50}$$

Although not diagonal, this form does not have explicitly time-dependent coefficients.

To identify the dispersion relations of the propagating modes it is convenient to Fourier transform by switching to energy and momentum eigenstates, $\propto e^{\mathrm{i}(-Et+\mathbf{k}\cdot\mathbf{x})}$, leading to a quadratic action proportional to

$$\begin{pmatrix} \tilde{\chi} \\ \tilde{\xi} \end{pmatrix}^\dagger \begin{pmatrix} E^2 - k^2 - \lambda\varrho_0^2 & -2i\omega E \\ 2i\omega E & E^2 - k^2 \end{pmatrix} \begin{pmatrix} \tilde{\chi} \\ \tilde{\xi} \end{pmatrix}, \tag{6.51}$$

where E and $k = |\mathbf{k}|$ are the energy and the magnitude of momentum for the corresponding mode. The dispersion relations, $E(k)$, for the propagating modes correspond to those choices that make the eigenvalues,

$$\Delta_\pm = E^2 - k^2 - \frac{1}{2}\lambda\varrho_0^2\left[1 \pm \sqrt{1 + \frac{16E^2\omega^2}{\lambda^2\varrho_0^4}}\right], \tag{6.52}$$

of this matrix vanish.

For $\omega E \ll \frac{1}{2}\lambda\varrho_0^2$ the corresponding dispersion relations, $E_\pm(k)$, therefore satisfy

$$E_-^2\left(1 + \frac{4\omega^2}{\lambda\varrho_0^2}\right) - k^2 \simeq E_+^2\left(1 - \frac{4\omega^2}{\lambda\varrho_0^2}\right) - k^2 - \lambda\varrho_0^2 \simeq 0\,. \tag{6.53}$$

These show that nonzero ω does not introduce an energy gap at zero momentum for the

Goldstone boson, though such a gap does of course exist for the massive particle (though with v replaced with ϱ_0 when compared with the mass found in §1.1).

The main effect of nonzero ω for the Goldstone boson is to change its 'sound speed', c_s, defined by writing the small-k dispersion relation as $E^2 = k^2 c_s^2$. Comparing with the Goldstone mode relation, $E_-(k)$, implies

$$c_{s-} \simeq \left(1 + \frac{4\omega^2}{\lambda\varrho_0^2}\right)^{-1/2} \simeq 1 - \frac{2\omega^2}{\lambda v^2}\,, \tag{6.54}$$

to leading order in ω^2/m_R^2.

As mentioned earlier, the result $c_s \neq 1$ never arises when expanding about a static background like $\phi = v$ because anything except $c_s = 1$ is in that case forbidden by Lorentz invariance. Nontrivial speed of sound arises for time-dependent backgrounds because these break the underlying Lorentz invariance of the action.

The Wilsonian point of view

This same conclusion about the ω-dependence of the Goldstone-boson dispersion relation also follows directly from the toy-model's Wilsonian EFT, given by (6.10) and repeated here:

$$\mathfrak{L}_w = -v^2 \partial_\mu\vartheta\partial^\mu\vartheta + \frac{v^2}{m_R^2}(\partial_\mu\vartheta\,\partial^\mu\vartheta)^2 + \cdots\,, \tag{6.55}$$

where $\phi = \varrho\, e^{i\vartheta}$. Expanding ϑ about the slowly rolling classical solution, $\vartheta = \omega t + \tilde\vartheta$ then implies $-\partial_\mu\vartheta\,\partial^\mu\vartheta = \omega^2 + 2\omega\,\partial_t\tilde\vartheta + (\partial_t\tilde\vartheta)^2 - \nabla\tilde\vartheta\cdot\nabla\tilde\vartheta$, so the quadratic part of the expanded action becomes

$$\mathfrak{L}_w^{(2)} = v^2\Big[(\partial_t\tilde\vartheta)^2 - \nabla\tilde\vartheta\cdot\nabla\tilde\vartheta\Big] + \frac{\omega^2}{\lambda}\Big[6(\partial_t\tilde\vartheta)^2 - 2\nabla\tilde\vartheta\cdot\nabla\tilde\vartheta\Big] + \cdots\,, \tag{6.56}$$

where ellipses denote terms involving higher powers of ω/m_R.

The field equations for ϑ predicted by this action therefore are

$$-\left(1 + \frac{6\omega^2}{\lambda v^2}\right)\partial_t^2\tilde\vartheta + \left(1 + \frac{2\omega^2}{\lambda v^2}\right)\nabla^2\tilde\vartheta \simeq 0\,, \tag{6.57}$$

which when compared to the wave equation $(-\partial_t^2 + c_s^2\nabla^2)\tilde\vartheta = 0$ leads to a prediction

$$c_s \simeq \sqrt{\frac{1 + 2\omega^2/(\lambda v^2)}{1 + 6\omega^2/(\lambda v^2)}} \simeq 1 - \frac{2\omega^2}{\lambda v^2}\,, \tag{6.58}$$

that agrees to leading nontrivial order in ω^2/m_R^2 with (6.54). Notice that this ω-dependent Goldstone sound speed is less than the speed of light (*i.e.* $c_s < c = 1$) by virtue of the sign of the $(\partial_\mu\vartheta\,\partial^\mu\vartheta)^2$ term.

6.3.2 Counting Goldstone states and currents ♣

Since time-dependent backgrounds spontaneously break spacetime symmetries, one might expect Goldstone's theorem to ensure the existence of new low-energy Goldstone degrees

of freedom. Although it is sometimes true that each new broken symmetry generator implies a new Goldstone particle, the toy model shows that this naive counting of Goldstone states can be misleading, particularly for spacetime symmetries [230]. The discussion given here follows [107] (see also [108]), and is generalized to more general background fields in §14.3.1.

For example, for the toy model expanded about the time-dependent classical background $\vartheta_c = \omega t$, the background breaks both the internal $U(1)$ symmetry — for which $\vartheta \to \vartheta + c$ for constant c — and time-translation invariance: $t \to t + \tau$ for constant τ. Breaking two symmetries naively suggests there should be two Goldstone particles, yet the effective theory only contains the one low-energy state.

For these specific symmetries the real lesson of the toy model is this: with multiple symmetries one must be careful when counting how many symmetries are broken. That is, it is always possible to undo the action of time translation, $t \to t + \tau$, on the background by simultaneously performing a compensating $U(1)$ transformation, $\vartheta \to \vartheta - \omega\tau$, leaving ϑ_c invariant. Only one Goldstone particle arises because the background $\vartheta_c = \omega t$ really breaks only one combination of these two symmetries.

More generally, time-dependent backgrounds also break the six-dimensional group of Lorentz transformations down to the three-dimensional group of rotations. Why doesn't Goldstone's theorem imply there must also be Goldstone modes for these broken symmetries?

To see why, it is worth referring back to the derivation of Goldstone's theorem presented in §4.1.2. What matters for Goldstone theorem is not the number of broken generators of the symmetry group. What matters instead is the number of independent conserved currents, $j^\mu(x)$, implied by the symmetry group, since for each independent current associated with a broken symmetry there should be a Goldstone state $|G\rangle$ satisfying the defining condition that

$$\langle G|j^0(x)|\Omega\rangle \neq 0\,, \tag{6.59}$$

where $|\Omega\rangle$ is the ground state. Furthermore, although Goldstone's theorem establishes the existence of such a state for each broken current, it *doesn't* require that a new state is required for each new current.

As reviewed in Appendix C.5.3, for spacetime symmetries there are only four independent conserved currents regardless of the dimension of the group of spacetime symmetries. This is because spacetime symmetries all have their roots in diffeomorphisms, $\delta x^\mu = V^\mu(x)$, for which the associated conserved current is the *stress-energy tensor*, $T_{\mu\nu}(x) = T_{\nu\mu}(x)$, defined in terms of the matter action by

$$T^{\mu\nu} = \frac{2}{\sqrt{-g}}\,\frac{\delta S_m}{\delta g_{\mu\nu}}\,, \tag{6.60}$$

where the spacetime metric is temporarily introduced for the purpose of performing the variation, before returning to the flat cartesian Minkowski metric of special relativity: $g_{\mu\nu} = \eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$.

As also reviewed in Appendices C.5.2 and C.5.3, spacetime symmetries correspond to those diffeomorphisms that leave the background metric invariant, which for the Minkowski

metric turns out to mean that V^μ must satisfy

$$\delta\eta_{\mu\nu} = \partial_\mu V_\nu + \partial_\nu V_\mu = 0\,. \tag{6.61}$$

This has as solutions $V_\mu = a_\mu + \omega_{\mu\nu}\, x^\nu$, with $\omega_{\mu\nu} = -\omega_{\nu\mu}$, corresponding to the usual translations in spacetime ($\delta x^\mu = a^\mu$) and Lorentz transformations ($\delta x^\mu = \omega^\mu{}_\nu\, x^\nu$).

For each solution to (6.61) a conserved current can be constructed using only the symmetric stress-energy tensor $T^{\mu\nu}$, since conservation $\partial_\mu T^{\mu\nu} = 0$ together with (6.61) implies $\partial_\mu j^\mu_\nu = 0$ where

$$j^\lambda_\nu(x) := T^{\lambda\mu}(x)\, V_\mu(x)\,. \tag{6.62}$$

The corresponding conserved charge (or generator) for this symmetry is constructed by integrating j^0_ν over all of space.

With this in mind the Goldstone states required by spontaneously broken spacetime symmetries are those for which the stress-energy matrix element

$$\langle G|T^{0\mu}(x)|\Omega\rangle\, V_\mu(x) \neq 0\,, \tag{6.63}$$

is nonzero. As seen in §13.1 and §14.3, systems (such as solids or liquids) that spontaneously break Poincaré invariance typically do give rise to Goldstone modes of this type (corresponding to sound waves, or phonons). What is *not* in general guaranteed by Goldstone's theorem is that the state $|G\rangle$ appearing in (6.59) need be different than the state appearing in (6.63). The states appearing in these matrix elements can sometimes be different, but need not always.

The toy model provides an explicit example where both (6.59) and (6.63) are satisfied by the same state: the massless state described by the field ξ. Indeed for weak coupling the low-energy sector only has a single state available to play both roles. To see this explicitly it is instructive to compute explicitly both the Noether current for the internal-$U(1)$ current and the stress energy.

Working to lowest order in the energy expansion the action is simply that of a massless free scalar field,

$$\mathfrak{L}_w \simeq -\frac{1}{2}\,(\partial^\mu\xi\,\partial_\mu\xi) + \frac{\lambda}{4m_R^4}(\partial^\mu\xi\,\partial_\mu\xi)^2 + \cdots\,, \tag{6.64}$$

for which the current predicted by (4.7) for the $U(1)$ symmetry $\xi \to \xi + \sqrt{2}\,cv$ (where c is the constant symmetry parameter) is

$$j_\mu = -\sqrt{2}\,v\,\partial_\mu\xi\left[1 - \frac{\lambda}{m_R^4}(\partial^\nu\xi\,\partial_\nu\xi) + \cdots\right]. \tag{6.65}$$

The stress energy predicted for a minimally coupled scalar is similarly given by

$$T_{\mu\nu} = \partial_\mu\xi\,\partial_\nu\xi - \frac{1}{2}(\partial^\lambda\xi\,\partial_\lambda\xi)\,\eta_{\mu\nu} + \cdots\,, \tag{6.66}$$

where the ellipses in both of these expression denote higher-derivative contributions than those written.

When expanded about a time-dependent solution, $\xi = \sqrt{2}\,v\,\omega t + \tilde\xi$, both j^0 and $T^0{}_\mu$ contain terms linear in the fluctuation $\tilde\xi$. Writing[5] $\langle p|\tilde\xi(x)|\Omega\rangle = F\,e^{ipx}$ (with nonzero F)

[5] As written F contains factors of $\sqrt{E(p)}$ unless $|p\rangle$ is normalized covariantly (see *e.g.* Appendix B.1).

for a single-particle momentum eigenstate $|p\rangle$ shows that the field $\tilde{\xi}$ plays the role of the Goldstone state for all of the broken symmetries, with

$$\begin{aligned} \langle p|j_\mu(x)|\Omega\rangle &= -\mathrm{i}\sqrt{2}\, v\, F\, p_\mu e^{\mathrm{i}px} + \cdots \\ \text{and} \quad \langle p|T^0{}_\mu(x)|\Omega\rangle &= \mathrm{i}\sqrt{2}\, v\, \omega\, F\, p_\mu\, e^{\mathrm{i}px} + \cdots , \end{aligned} \tag{6.67}$$

where ellipses denote terms of relative order ω^2/m_κ^2 (or those suppressed by loop factors). In this sense the low-energy sector of the toy model saturates the requirements of Goldstone's theorem in a minimal way.

6.4 Summary

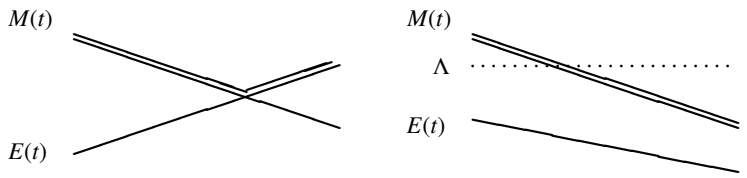
To summarize this chapter, the bottom line is this: time-dependent evolution in the full theory (both of backgrounds and classical and quantum fluctuations about them) can be captured using time-dependent solutions to the low-energy effective theory, but only if the evolution of interest is sufficiently slow.

Generically 'sufficiently slow' means demanding that $\mu_\phi := \dot{\phi}/\phi$ — for all choices of fields $\phi^i(x)$ in the problem — be much smaller than the UV scale M (*i.e.* in the toy model, m_κ). This adiabatic condition is in addition to all the other requirements already needed when formulating a Wilsonian low-energy theory: such as that the energies of all fluctuation modes be much smaller than M.

This points to two kinds of generic new failure modes specific to EFTs applied to time-dependent problems. The first new failure mode arises if the background evolution itself should become too fast. In such a case the transfer of energy between background and fluctuations (such as through particle production using energy extracted from the background) becomes too efficient, destroying the adiabatic approximation (and with it the approximately conserved notion of energy used to discriminate between low- and high-energy fluctuation modes).

A second type of new failure mode is simply the time-dependent version of the old failure mode: a nominally low energy, E, is not small enough to trust the E/M expansion. Time dependent drift of $E(t)$ and $M(t)$ means $E(t)/M(t)$ might eventually become large even if were small initially (such as occurs in level crossing, see panel (a) of Figure 6.1).

Notice that level crossing – for which the EFT expansion in powers of E/M must fail – is different from having UV states simply evolve below some regulator scale Λ (panel (b) of Fig. 6.1). Nothing dramatic need happen as UV levels pass below a cutoff scale, provided the UV states evolve in their adiabatic vacua, since cutoff scales by construction do not appear in any physical quantities.



(a) Level crossing (b) Cutoff crossing

Fig. 6.1 A sketch of the adiabatic time-evolution for the energy, $E(t)$ (solid line), of a nominally low-energy state and the energy, $M(t)$ (double line), for a representative UV state. The left panel shows level crossing where (modulo level repulsion) high- and low-energy states meet so the EFT description fails. In the right panel high-energy states evolve past a cutoff, Λ (dotted line), without level crossing (so EFT methods need not fail).

Exercises

Exercise 6.1 Rederive eq. (6.17) using the stress-energy tensor, $T^{\mu\nu}$, defined in (6.60) applied to the action built from the lagrangian density of eq. (6.64). Doing so requires writing this matter action for a general metric:

$$S_m \simeq \int \mathrm{d}^4x\, \sqrt{-g}\, \left[\frac{1}{2}X + \frac{\lambda}{4m_R^4}X^2\right],$$

where $X := -g^{\mu\nu}\partial_\mu\xi\,\partial_\nu\xi$ and g denotes the determinant of (and $g^{\mu\nu}$ is the matrix inverse of) the covariant components of the metric, $g_{\mu\nu}$ (see Appendices A.2.1 and C.5.2 for more details). Once the stress-energy tensor is computed the energy density is given by $\varepsilon = T_{00}$.

Exercise 6.2 Too-rapid background time-dependence can ruin the low-energy approximation. Consider the toy model of §1.1 in the semiclassical regime, but instead of starting in the vacuum consider the background field configuration describing homogeneous heavy-field oscillations about its minimum: $\chi_c(t) = \chi_0 \cos(m_R t)$, where $\chi_0 \ll v$ so that the cubic and quartic terms in the potential $V(\chi)$ can be neglected. Compute the energies of the ξ particles that are pair-produced by their interactions with this background oscillating field and calculate their production rate. Can the production of these ξ particles be described purely with a low-energy EFT description?

Exercise 6.3 As a toy model of level crossing (and repulsion) consider two real free scalar fields, ϕ_1 and ϕ_2, that mix with one another through the lagrangian density $\mathfrak{L} = -\frac{1}{2}\left[(\partial\phi_1)^2 + (\partial\phi_2)^2\right] + \mathfrak{L}_{\rm mix}$ where

$$\mathfrak{L}_{\rm mix} = -\frac{1}{2}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}^T\begin{pmatrix} gn(t) & \mu^2\\ \mu^2 & m^2\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$$

where μ and m are positive and real mass parameters with $\mu \ll m$, g is a coupling

constant and $n(t)$ is the density of particles in a medium within which the scalars are immersed. $n(t) = n_0\, e^{-t/\tau}$ is assumed to be monotonically decreasing, asymptoting to zero for large t. For any fixed t what are the eigenvalues and eigenvectors for the mass matrix? Assume the system is prepared in a state is that is a ϕ_1 eigenstate with momentum $\mathbf{p}$ at $t = 0$ with $gn_0 > m^2$. After this $n(t)$ falls slowly enough that the evolution is adiabatic — *i.e.* instantaneous energy eigenstates evolve with phase $\exp[-\mathrm{i}\int_{t_0}^{t} \mathrm{d}s\, E(s)]$. What is the likelihood that the state is measured at $t \to \infty$ to be in state ϕ_2?