

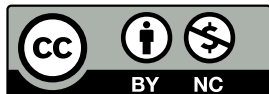
arXiv:2404.14518v1 [hep-th] 22 Apr 2024
Tomas Brauner

Effective Field Theory for Spontaneously Broken Symmetry

April 24, 2024

Preprint of a book published by Springer Cham, available
open-access at <https://doi.org/10.1007/978-3-031-48378-3>.

This work is licensed under a [Creative Commons “Attribution-NonCommercial 4.0 International”](https://creativecommons.org/licenses/by-nc/4.0/) license.



To Laura and Julie

Preface

Symmetry has accompanied our attempts to understand the inner workings of nature ever since the birth of science. The powerful concept of *spontaneous symmetry breaking* (SSB) makes it possible to reconcile the complex experimental phenomenology with the simplicity of underlying natural laws. It is therefore a necessary ingredient of virtually any textbook on quantum field theory. However, a typical textbook-level treatment offers a mere simplified version of the state of the art of the subject in its early days of the 1960s. This hardly does justice to the richness and beauty of SSB. Moreover, the subject has undergone development in the last two decades that has substantially sharpened our understanding of SSB in quantum many-body systems.

This book is an attempt to fill the gap in the literature, and to provide an up-to-date survey of the landscape of SSB from the perspective of *effective field theory* (EFT). In writing the book, I was aiming at learners of all seniority levels who are familiar with the basic notions of quantum field theory. The style of the text and the mathematical background required will probably be a best match for graduate students of theoretical physics. However, the introductory parts of the book might be accessible even to advanced undergraduates. Those will benefit from studying first Part I of the book, which introduces the concepts of SSB and the EFT description thereof at an elementary level. More experienced readers might proceed directly to Part II. This offers an overview of the physics of SSB at a standard textbook level. However, it already contains numerous details that are hard to find in a student-friendly form elsewhere. The core techniques of EFT for SSB are developed in Parts III and IV. These provide a comprehensive account of the subject, from the pioneering works of the 1960s to some recent developments that are otherwise only found scattered in the research literature.

The most important recent results included are, as of writing this book, about a decade old. This applies in particular to the completed classification of *Nambu–Goldstone* (NG) *bosons* (Chap. 6), the EFT formalism for spontaneously broken internal symmetries in quantum many-body systems (Chap. 8), and some applications of EFT to systems with spontaneously broken spacetime symmetry (Chaps. 13 and 14). However, some of the material included also reflects new works that appeared while I was already writing the book. This is the case especially for the discussion

of scattering of NG bosons in Chap. 10. Finally, Chap. 12 puts forward a novel formalism for implementation of spontaneously broken spacetime symmetry that, to the best of my knowledge, has not appeared in this form before. This significantly influences the narrative surrounding the EFT for broken spacetime symmetries in Chaps. 13 and 14. My hope is that the change is for the better.

The one thing I have learned while writing this book is that one's understanding of any topic is a never-ending process of evolution. I would therefore like to extend my gratitude to those who helped me along the way. First and foremost, I am indebted to Angelo Esposito, Yoshimasa Hidaka and Jiří Novotný, who have read critically parts of the manuscript and given me constructive feedback, helping me improve the text on multiple fronts. My personal view of the subject of this book and of effective field theory in general has been shaped by interactions with numerous mentors and collaborators throughout the years, including Jens O. Andersen, Christoph P. Hofmann, Jiří Hošek, Martin Mojahed, Sergej Moroz, Hitoshi Murayama, Jiří Novotný, Riccardo Penco, Aleksi Vuorinen, Haruki Watanabe and Naoki Yamamoto. Last but not least, I would like to thank Stefan Theisen for encouraging me to write this book.

Stavanger, Norway
September 2023

Tomáš Brauner

Contents

Part I Prologue

1	Introduction	3
1.1	What Is Effective Field Theory?	3
1.2	Broken Symmetry Zoo	4
1.3	Structure of This Book	6
1.3.1	Further Reading	7
	References	8
2	Our First Model	9
2.1	Spontaneous Symmetry Breaking	10
2.2	Nambu–Goldstone Boson and Its Interactions	12
2.2.1	Linear Parameterization	12
2.2.2	Scattering of Nambu–Goldstone Bosons	13
2.2.3	Nonlinear Parameterization	16
2.3	Low-Energy Effective Field Theory	19
2.3.1	Matching	19
2.3.2	Eliminating the Heavy Modes	20
2.4	Moral Lessons	22
	References	22
3	Generalizations of the Model	23
3.1	Relativistic Models with Non-Abelian Symmetry	24
3.1.1	Spectrum of Nambu–Goldstone Bosons	25
3.1.2	Low-Energy Effective Field Theory	27
3.2	Going Nonrelativistic	29
3.2.1	Single Schrödinger Field	29
3.2.2	Multiple Nambu–Goldstone Fields	32
3.3	Moral Lessons	35
	References	36

Part II Foundations

4	Symmetry and Conservation Laws	39
4.1	What Is Symmetry?	39
4.1.1	Symmetry Transformations	40
4.1.2	Object of Symmetry	41
4.2	Lagrangian Approach to Symmetry	42
4.2.1	Noether's Theorem	42
4.2.2	Tensor Conservation Laws	45
4.3	Symmetry and Conservation Laws in Hamiltonian Formalism	46
4.3.1	Symplectic Formulation of Hamiltonian Dynamics	47
4.3.2	Symmetry in Quantum Physics	51
	References	52
5	Spontaneous Symmetry Breaking	53
5.1	Physical State and Its Symmetry	54
5.1.1	Broken and Unbroken Symmetry	54
5.1.2	Symmetrization by Group Averaging	56
5.2	Effect of External Perturbations	59
5.2.1	Taking the Thermodynamic Limit	60
5.2.2	Order Parameter and the Vacuum Manifold	63
5.2.3	Intermediate Summary	65
5.3	Some Subtle Features of Spontaneous Symmetry Breaking	66
5.3.1	Free Schrödinger Field in Finite Volume	67
5.3.2	Pathologies of the Infinite-Volume Limit	69
5.3.3	Uniqueness of the Finite-Volume Ground State	71
	References	74
6	Nambu–Goldstone Bosons	75
6.1	Intuitive Picture	75
6.1.1	Redundancy of Order Parameter Fluctuations	76
6.1.2	Canonical Conjugation of Nambu–Goldstone Fields	77
6.1.3	The Big Picture	79
6.2	Goldstone Theorem	80
6.2.1	Operator Proof	80
6.2.2	Effective Action Proof	83
6.3	Classification and Counting	84
6.3.1	Independent Order Parameter Fluctuations	84
6.3.2	Type-A and Type-B Nambu–Goldstone Bosons	87
6.4	Nambu–Goldstone-Like Modes	90
6.4.1	Pseudo-Nambu–Goldstone Bosons	90
6.4.2	Quasi-Nambu–Goldstone Bosons	93
6.4.3	Massive Nambu–Goldstone Bosons	93
	References	96

Part III Spontaneously Broken Internal Symmetry

7	Nonlinear Realization of Symmetry	99
7.1	Group Action on Manifolds	101
7.2	Classification of Nonlinear Realizations	104
7.2.1	Linearization of Group Action	105
7.2.2	From Linear Representation to Nonlinear Realization	108
7.3	Standard Realization of Symmetry	110
7.3.1	Nonlinear Realization on Coset Spaces	112
7.3.2	Symmetric Coset Spaces	116
7.4	Geometry of the Coset Space	119
7.4.1	Canonical and Torsion-Free Connection	121
7.4.2	Riemannian Metric	122
	References	124
8	Low-Energy Effective Field Theory	125
8.1	Structure of the Effective Lagrangian	125
8.1.1	Reminder of the Standard Nonlinear Realization	127
8.1.2	Lagrangians with Two Spatial or Two Temporal Derivatives	128
8.1.3	Lagrangians with One Temporal Derivative	131
8.1.4	Overview of the Lowest-Order Effective Lagrangian	133
8.2	Effective Lagrangians from Background Gauge Invariance	136
8.2.1	Methodology of Construction of Effective Actions	138
8.2.2	Lagrangians up to Order Two in Derivative Expansion	140
8.2.3	Effects of Explicit Symmetry Breaking	142
8.2.4	Coupling to Matter Fields	144
8.3	Equation of Motion	146
8.3.1	Spectrum of Nambu–Goldstone Bosons Revisited	148
8.3.2	More on the Geometry of the Coset Space	150
	References	152
9	Applications to Particle and Condensed-Matter Physics	153
9.1	Chiral Perturbation Theory of Mesons	153
9.1.1	Power Counting	155
9.1.2	Effective Lagrangian	157
9.1.3	Interaction with External Fields	161
9.1.4	Effects of the Chiral Anomaly	165
9.2	Spin Waves in Ferro- and Antiferromagnets	169
9.2.1	Power Counting and Effective Lagrangian	170
9.2.2	Equation of Motion and Magnon Spectrum	173
9.2.3	Effects of Symmetry-Breaking Perturbations	175
9.2.4	Some Topological Aspects of Ferromagnets	180
	References	184

10 Scattering of Nambu–Goldstone Bosons	185
10.1 Adler Zero Revisited	186
10.1.1 Generalized Soft Theorem	189
10.1.2 Application to Coset Effective Theories	191
10.2 Geometric Framework for Scattering Amplitudes	193
10.2.1 Geometric Soft Theorem for Nambu–Goldstone Bosons	195
10.2.2 Adler Zero or Not?	197
10.2.3 Symmetric Coset Spaces	199
10.3 Beyond Adler Zero	200
10.3.1 Dirac–Born–Infeld Theory	202
10.3.2 Galileon and Special Galileon Theory	203
10.3.3 Effective Theories with Enhanced Soft Limit from Symmetry	206
10.4 Soft Recursion	207
10.4.1 Complexified Kinematics	208
10.4.2 Recursion Relation for On-Shell Amplitudes	209
10.4.3 Soft Bootstrap	213
References	214
 Part IV Spontaneously Broken Spacetime Symmetry	
11 Locally Equivalent Symmetries	217
11.1 Relations Between Noether Currents	217
11.2 Noether Currents from Background Gauging	219
11.3 Examples	221
11.3.1 Galileon Symmetry	221
11.3.2 Spacetime Translations and Rotations	223
11.3.3 Galilei Invariance	227
11.3.4 Changing the Background: Magnetic Translations	229
11.4 Application to Scattering of Nambu–Goldstone Bosons	231
11.4.1 Galileon Theory	232
11.4.2 Theories with Generalized Shift Symmetry	233
References	234
 12 Nonlinear Realization of Spacetime Symmetry	235
12.1 Reminder of Nonlinear Realization of Internal Symmetry	236
12.2 Spacetime Symmetry as a Point Transformation	237
12.3 Standard Nonlinear Realization	238
12.3.1 Summary of the Construction	239
12.3.2 Relation to Physics of Broken Spacetime Symmetry	242
12.4 Examples	246
12.4.1 Lorentz Scalars with Internal Symmetry	246
12.4.2 Lorentz Scalars with Scale Invariance	247
12.4.3 Lorentz Vector with(out) Lorentz Scalar	248
12.4.4 Schrödinger Scalars with Galilei Symmetry	250
References	252

13	Broken Spacetime Symmetry in Quantum Matter	253
13.1	Building Blocks for Construction of Effective Actions	254
13.1.1	Maurer–Cartan Form	254
13.1.2	Covariant Derivatives of Fields	257
13.2	Twisting Order Parameter for Internal Symmetry	260
13.2.1	New Features of the Old Setup	261
13.2.2	Case Study: Relativistic Superfluids	263
13.3	Vector Modes: the Relevant, the Irrelevant and the Unphysical	266
13.3.1	The Relevant: Helimagnets	267
13.3.2	The Irrelevant: Smectic Liquid Crystals	271
13.3.3	The Unphysical: Nonrelativistic Superfluids	275
13.3.4	Inverse Higgs Constraints	280
13.4	Genuine Breaking of Translation Invariance	285
13.4.1	One-Dimensional Modulation of the Order Parameter	285
13.4.2	Case Study: Fluctuations of a Domain Wall	287
13.4.3	Effective Action from Background Gauge Invariance	292
13.4.4	Further Possible Applications	296
	References	297
14	Broken Spacetime Symmetry in Classical Matter	299
14.1	Emergent Symmetry of Classical Matter	299
14.1.1	Introduction: Spring Model of Elasticity	300
14.1.2	Emergent Symmetries of Solids and Fluids	303
14.2	Nonlinear Realization of Emergent Symmetry	307
14.2.1	Field Variables and Unbroken Symmetry	307
14.2.2	Building Blocks for Construction of Effective Actions	309
14.3	Effective Field Theory of Classical Matter	310
14.3.1	Relativistic Solids	310
14.3.2	Nonrelativistic Supersolids	311
14.3.3	Nonrelativistic Solids	313
14.3.4	Perfect Fluids	316
14.4	Coupling Nambu–Goldstone Bosons to Classical Matter	320
	References	322
Part V Epilogue		
15	Topics Not Covered in This Book	325
15.1	Effects of Nonzero Temperature	325
15.2	No-Go Theorems for Spontaneous Symmetry Breaking	326
15.3	Topological Aspects of Spontaneous Symmetry Breaking	329
15.4	Generalized Symmetries	332
	References	334
16	Some Open Questions	335
	References	336

A	Elements of Differential Geometry	337
A.1	Smooth Manifolds	337
A.2	Linear Structures on Manifolds	340
A.2.1	Tangent Vectors and Vector Fields	340
A.2.2	Tensors and Tensor Fields	342
A.3	Maps Between and on Manifolds	345
A.3.1	Push-Forward and Pull-Back	345
A.3.2	Flow of Vector Fields	348
A.3.3	Lie Derivative	350
A.4	Exterior Derivative	352
A.5	Affine Connection	355
A.5.1	Covariant Derivative	356
A.5.2	Curvature and Torsion	358
A.6	Riemannian Geometry	360
A.6.1	Riemannian Metric	360
A.6.2	Isometries of Riemannian Metric	364
A.6.3	Symmetries of Curvature Tensor	366
A.6.4	Geodesic Normal Coordinates	367
A.6.5	Hodge Star	369
A.7	Integration on Manifolds	372
A.7.1	Orientable Manifolds	373
A.7.2	Riemannian Manifolds	377
A.8	Homology and Cohomology	378
A.8.1	Singular Homology	378
A.8.2	De Rham Cohomology	380
	References	385

Notation and Conventions

List of Acronyms

Acronyms that appear in the book and their use in established collocations:

ChPT	Chiral perturbation theory
CS	Chern–Simons (Lagrangian, theory)
DBI	Dirac–Born–Infeld (Lagrangian, theory)
DM	Dzyaloshinskii–Moriya (interaction, term)
EFT	Effective field theory
EM	Energy–momentum (tensor)
EoM	Equation of motion
GW	Goldstone–Wilczek (current)
IHC	Inverse Higgs constraint
LC	Levi-Civita (connection, symbol, tensor)
LO	Leading order
MC	Maurer–Cartan (equation, form)
NG	Nambu–Goldstone (boson, field, mode)
NLO	Next-to-leading order
QCD	Quantum chromodynamics
SSB	Spontaneous symmetry breaking
VEV	Vacuum expectation value
VPD	Volume-preserving diffeomorphism
WZ	Wess–Zumino (action, Lagrangian, term)

For the reader’s convenience, each of these acronyms is also defined in the text the first time it appears in a particular chapter.

Mathematical Conventions

This book aims at audiences from diverse areas of physics, which makes it virtually impossible to maintain coherent notation without alienating some of the readers. I therefore try to adopt conventions that are as close as possible to practice, at the cost of being somewhat dependent on the context.

Thus, a “relativistic” notation is used to discuss physical systems whose dynamics is Lorentz-invariant. I assume a flat Minkowski spacetime unless (exceptionally) stated otherwise. Much of the content of the book is valid for spacetimes of any dimension $D \geq 3$; see Sect. 15.2 for a justification of this constraint. The timelike Minkowski metric,

$$g_{\mu\nu} \equiv \text{diag}(+1, -1, -1, \dots) ,$$

is implicitly assumed when needed. Spacetime vectors (or, in $D = 4$ dimensions, four-vectors) are denoted using italics: p, q, \dots . An inner product of spacetime vectors, implied by the Minkowski metric, is indicated with a dot,

$$p \cdot q \equiv g_{\mu\nu} p^\mu q^\nu = p^\mu q_\mu = p_\mu q^\mu .$$

Here the Einstein summation convention is applied to any pair of repeated indices, one covariant and one contravariant. The Levi-Civita tensor $\varepsilon_{\mu\nu\lambda\dots}$ in Minkowski spacetime is normalized so that $\varepsilon^{012\dots} = 1$. Finally, I consistently use the “natural units” of high-energy physics where both the speed of light c and the reduced Planck constant \hbar are set to one. Thus, the energy and momentum of a relativistic particle are identified respectively with the frequency and wave vector of the corresponding quantum-mechanical plane wave. Moreover, the Minkowski square of the energy–momentum of the particle is $p^2 \equiv p \cdot p = m^2$, m being the rest mass of the particle.

On the contrary, in systems lacking manifest Lorentz invariance, typical for applications to condensed-matter physics, I adopt a “nonrelativistic” notation. Here, spatial vectors are denoted using boldface italics: $\mathbf{a}, \mathbf{b}, \dots$. The Einstein summation convention is still used, but indices are no longer raised or lowered with the Minkowski metric. Instead, upper and lower indices are used interchangeably just for visual clarity and are treated as equivalent. Formally, this amounts to using the flat Euclidean metric, $g_{rs} \equiv \delta_{rs}$. The dot product of two spatial vectors can then be written in a number of equivalent forms such as

$$\mathbf{a} \cdot \mathbf{b} = \delta_{rs} a^r b^s = a_r b_r = a^r b^r = a^r b_r = a_r b^r .$$

The symbol ∇ is used for the gradient operator, that is a spatial-vector differential operator with components $\partial_r \equiv \partial/\partial x^r$. Wherever needed explicitly, the number of spatial dimensions is denoted as $d \equiv D - 1$.

Part I
Prologue

Chapter 1

Introduction

1.1 What Is Effective Field Theory?

The laws of nature have a hierarchical structure, stratified by the resolution scale of our observations. When Isaac Newton published in 1687 his gravitational law explaining Kepler's laws of planetary motion, he did not need to know the details of the planets' inner structure. His theory gives an accurate description of the dynamics of the solar system while treating the planets as point-like objects.

The phenomenological understanding of the basic material properties of solids, liquids and gases culminated in the mid-19th century through the work of Clausius, Gibbs, Kelvin, Maxwell and others. Much about these different phases of matter and the transitions between them can be, and was, learned without insight into the molecular structure of matter. It was only about a half-century later that quantum mechanics provided an adequate framework for the description of atoms and molecules.

Quantum mechanics itself is extremely successful in explaining the chemical properties of different substances and mechanical properties of solids, liquids and gases. In fact, virtually all natural phenomena observed at macroscopic scales boil down either to gravity or to Maxwell's electromagnetism combined with quantum mechanics. Up to rare exceptions, for instance radioactivity, one does not need to understand the structure of atomic nuclei. Nuclear physics itself only advanced later, in the 1930s. And again, much about nuclear structure and reactions was understood before the discovery in the 1960s that individual nucleons are composed of yet smaller particles, the quarks.

The smallest currently known constituents of matter are quarks, bound in atomic nuclei, and leptons, of which the electron is but one example. How do we know that the chain of successive divisions into smaller and smaller constituents ends here? We do not. Perhaps, one day, even smaller building blocks of nature will be discovered. Nevertheless, we do have a successful theory that is able to account for experimental observations at the current resolution frontier: the Standard Model of particle physics. This theory does not depend on the details of as yet undiscovered

microscopic physics. Our ignorance of such microscopic structure currently out of reach is subsumed into the values of the input parameters of the Standard Model.

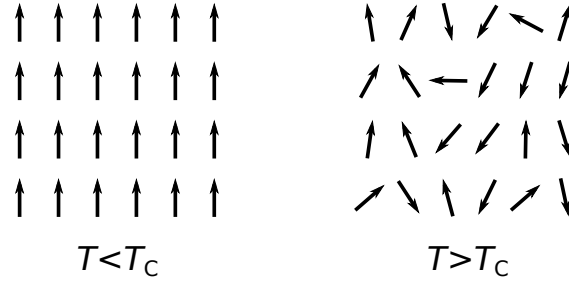
This is the essence of *effective field theory* (EFT). Going back to where we started, the information about the detailed structure of planets, unknown to Newton, could be subsumed into the values of the planets' masses. Likewise, the quantum-mechanical nature of molecular interactions only affects macroscopic thermodynamics through material constants such as specific heats or elastic moduli. All that quantum mechanics in turn needs to know about atomic nuclei are their masses and electric charges. Finally, the quark structure of nucleons only manifests itself through binding forces between nucleons that hold the nuclei together. These nuclear forces were studied thoroughly already in the mid-1930s by Fermi, Yukawa and others, three decades before the existence of quarks was conceived by Gell-Mann and Zweig.

Generally speaking, EFT is a framework for a quantitative description of nature at certain level of resolution. Whatever physics might exist at shorter, unresolved length scales can be taken into account through the values of the parameters of the EFT. In the absence of further information about the microscopic physics, these parameters have to be determined by experiment. This line of reasoning has implicitly accompanied the evolution of physics from its early stages. In the second half of the 20th century, it was however developed into a full-fledged quantitative formalism with tremendous impact, from concrete physical applications to the philosophy of science. One of the founders of modern EFT, Steven Weinberg, offers a nice account of the early history of EFT in [1]. The array of successful applications of the EFT program has grown so long in the last decades that I cannot even list them all here. Instead, I point the reader to the literature. References to introductory-level expositions of EFT with applications are given below in Sect. 1.3.1.

1.2 Broken Symmetry Zoo

Most macroscopic phenomena observed in nature are dominated by the collective dynamics of the fundamental constituents of matter. In hindsight, this is the reason why 19th-century physics was so successful in describing nature. Consider, for instance, a single-component fluid in thermodynamic equilibrium. All the thermodynamic properties of the fluid are determined by the values of two independent observables such as temperature and pressure. If we now disturb the fluid locally, the interactions among the fluid molecules will cause the perturbation to propagate. This propagation manifests macroscopically as a sound wave, possibly accompanied by a background flow. Such small perturbations from thermodynamic equilibrium are described by hydrodynamics, which is an early example of an EFT of the modern type. The microscopic input required here is the equation of state of the fluid along with a set of transport coefficients. Other than that, hydrodynamics is based only on a set of local conservation laws including energy, momentum and whatever other conserved charge the fluid might carry. We know from Noether's theorem that local conservation laws are a consequence of continuous symmetries. One can therefore

Fig. 1.1 Schematic picture of the ordered state of a ferromagnet below the Curie temperature T_C and of the disordered state above the Curie temperature T_C



say that the dynamics of fluids near equilibrium is largely dictated by symmetry. That is typical of *spontaneous symmetry breaking* (SSB).

Let me illustrate this on another example. Ferromagnets are materials which, below certain temperature called the Curie temperature, exhibit spontaneous alignment of magnetic moments of their atoms. A solid ferromagnet can be pictured as a lattice, each of whose nodes carries a single spin degree of freedom, see Fig. 1.1. The mutual interactions between the spins may be perfectly isotropic, that is invariant under spatial rotations. Yet, the ferromagnetic state below the Curie temperature obviously possesses a preferred direction of alignment of the spins. This direction is a priori arbitrary and may be set for instance by boundary conditions, not by the internal dynamics of the ferromagnet itself. The emergence of an equilibrium state that breaks the intrinsic symmetry of the system is a hallmark of SSB.

A local perturbation of the ferromagnetic state will be propagated by spin–spin interactions and manifest itself macroscopically as a spin wave. The existence of wave-like excitations which dominate the physics at long distances is another general feature of SSB; these are called *Nambu–Goldstone* (NG) *bosons*. Just like for sound waves in fluids, the dynamics of ferromagnetic spin waves is controlled by symmetry. A quantitatively accurate EFT of spin waves can be constructed solely based on the underlying spacetime translation invariance, spatial rotation invariance, and the fundamental commutation relations of angular momentum (spin).

The above two examples hint at a particular type of EFT that governs the long-distance physics of systems with SSB. The relevant degrees of freedom of this EFT are the NG bosons whereas the details of the EFT are dictated by symmetry. It should therefore not come as a surprise that many of the techniques developed in this book rely on the theory of Lie groups, and to a lesser extent on differential geometry. The precise mathematical structure of the EFT depends on the kind of symmetry in question. It is common to use the term *internal symmetry* for symmetries that act directly on the dynamical degrees of freedom of a given physical system. SSB of an internal symmetry lies for instance behind the phenomena of (anti)ferromagnetism or superfluidity. On the other hand, a *spacetime symmetry* is a geometric property of space and time. It affects the dynamical degrees of freedom indirectly as fields living in the physical spacetime. SSB of spacetime symmetries is more subtle yet ubiquitous and relevant for *all* phases of matter. This is the reason why the form of hydrodynamics is largely fixed by the local conservation laws of energy and momen-

tum. A different pattern of SSB gives rise to the elasticity theory of solids. There are also examples of phenomena that exhibit a combination of SSB of internal and spacetime symmetry. These include for instance vortex lattices in rotating superfluids or type-II superconductors, anisotropic Cooper pairing in p -wave superconductors, or the helical order in chiral magnets.

1.3 Structure of This Book

The subject of this book is of central importance to quantum field and many-body theory. As such, it connects several branches of physics including, but not limited to, high-energy physics, condensed-matter physics, astrophysics and cosmology. With the explicit aim to cater to these different communities, I have tried my best to make the text relevant and comprehensible to audiences with different backgrounds. It is up to the reader to judge to what extent this effort has been successful.

The book aims primarily at graduate students of theoretical physics regardless of their concrete specialization. The text should in principle be accessible to anyone who has taken a first course on quantum field theory. With this target group in mind, I have interspersed the text with two didactic elements. On the one hand, there are numbered examples, graphically separated from the main text, that mostly serve to illustrate newly introduced theoretical concepts or arguments. Occasional more advanced examples mention interesting applications that would disrupt the line of discussion if included in the text. On the other hand, gray blocks such as this one point to important facts or subtleties that might otherwise be missed. They are meant to encourage the reader to take a critical look at the presented material.

The mathematics background required to benefit from this book likewise corresponds to that of a typical student of theoretical physics. It consists mostly of linear algebra, advanced calculus including complex calculus and the calculus of variations, and the theory of partial differential equations. The only piece of background I assume that might go beyond the curriculum of some graduate programs is certain familiarity with group theory and its applications to physics. The level and extent as covered by [2] is fully sufficient; Chaps. 14 and 15 of [3] are a good start that will get the reader very far. Some more advanced parts of the book rely on basic background in differential geometry. Everything that is needed in this regard (and presumably more) is covered in a self-contained manner in Appendix A.

In order to make the book potentially useful also for more experienced researchers, the content is composed of several layers. In the following Chaps. 2 and 3, I introduce the concepts of SSB, NG bosons and the EFT description thereof through the case study of a simple toy model. This is intended for an uninitiated reader. Anybody else can proceed directly to Part II. Chapters 5 and 6 of this part give a thorough

introduction to the physics of SSB. This serves as a supplement to advanced courses on field theory and may thus be suitable for graduate students or junior researchers. An expert reader can skip this part as well.

Parts **III** and **IV** constitute the core of this book that will bring the reader close to the research frontier. Part **III** is dedicated to spontaneously broken internal symmetries. Here the EFT methodology is by now well-settled, and is developed in detail in Chaps. **7** and **8**. Chapters **9** and **10** work out some concrete applications of the general formalism. In contrast to internal symmetries, spontaneously broken spacetime symmetries remain a rather active area of research with a number of open questions left to answer. The choice and organization of material in Part **IV** is therefore necessarily more subjective. I however try to draw on the analogy with broken internal symmetries as much as possible. This largely fixes the content of Chap. **12**, which details the necessary modifications to the techniques of Chap. **7**. The following Chaps. **13** and **14** then develop the EFT for spontaneously broken spacetime symmetry using numerous examples, roughly sorted by increasing complexity.

In spite of the rather large extent of the book, I was forced to make compromises regarding the choice of material. Sadly, some very exciting aspects of SSB had to be left out altogether. Covering them at the same level of detail as in the rest of the book would have required a substantial amount of additional background. Some of these topics are mentioned at least briefly in Part **V**.

Let me conclude this introduction with a remark on the chosen style of the book. In writing the text, I have not tried to give a review of all existing research literature on the subject. My intention was instead to create a unified narrative that would give the reader the tools necessary to critically assess existing results and approaches, and to launch a research program of their own. In line with this philosophy, I have deliberately adopted a very restrictive policy regarding bibliography. The purpose of references in this book is primarily to aid the reader. Generally, I thus only include references to point the reader to specified supplementary information, or occasionally to fill in a gap in the argumentation. Only when I explicitly borrow an idea or example from a single identifiable source, do I credit this with a citation. I therefore apologize to all those who might be missing their name in the reference list.

1.3.1 Further Reading

This book revolves around the application of EFT to physical systems with SSB. Most of the graduate-level physics background that is needed to follow the advanced parts of the book is covered in Part **II**. This however does not mean that the text is comprehensive. Here is a (biased) list of suggestions for readers who wish to learn more details than what this book can cover.

A monograph fully devoted to EFT was missing on the market for years. This gap has now been filled by the new opus by Burgess [4]. A reader seeking a more concise introduction to EFT will find a variety of excellent resources with slightly different balance of topics included. Thus, for instance, [5] and [6] cover a spectrum of

applications to particle phenomenology and nuclear physics. Reference [7] includes a discussion of physics beyond the Standard Model and of the application of EFT to fluid dynamics, which is hard to find in a student-friendly form elsewhere. Cohen's lectures [8] are a comprehensive source on EFT with a very useful list of further references including annotations. There are also great texts that mostly focus on EFT for NG bosons. This is true especially of [9, 10] and to some extent of [11].

SSB is rarely a subject of a dedicated monograph. An exception is the classic by Strocchi [12], which will satisfy a mathematically oriented reader desiring a rigorous treatment of SSB. The comprehensive early review [13] remains a valuable source of insight, including details that seem to have been forgotten by practitioners. The lecture notes [14] give a modern introduction to SSB that covers recent developments. They omit a discussion of EFT for NG bosons though, and are thus nicely complemented by [15] which focuses largely on the latter. Finally, the specific topic of classification of NG bosons is addressed in detail by [16].

References

1. S. Weinberg, Eur. Phys. J. **H46**, 6 (2021). DOI [10.1140/epjh/s13129-021-00004-x](https://doi.org/10.1140/epjh/s13129-021-00004-x)
2. H. Georgi, *Lie Algebras in Particle Physics*. Frontiers in Physics (Perseus Books, Reading, MA, USA, 1999)
3. M. Stone, P. Goldbart, *Mathematics for Physics: A Guided Tour for Graduate Students* (Cambridge University Press, Cambridge, UK, 2009)
4. C.P. Burgess, *Introduction to Effective Field Theory* (Cambridge University Press, Cambridge, UK, 2021). DOI [10.1017/9781139048040](https://doi.org/10.1017/9781139048040)
5. D.B. Kaplan, [arXiv:nucl-th/0510023](https://arxiv.org/abs/nucl-th/0510023)
6. A.V. Manohar, [arXiv:1804.05863](https://arxiv.org/abs/1804.05863)
7. B. Gripaios, [arXiv:1506.05039](https://arxiv.org/abs/1506.05039)
8. T. Cohen, PoS **TASI2018**, 011 (2019)
9. S. Scherer, M.R. Schindler, *A Primer for Chiral Perturbation Theory, Lecture Notes in Physics*, vol. 830 (Springer, 2012). DOI [10.1007/978-3-642-19254-8](https://doi.org/10.1007/978-3-642-19254-8)
10. A. Pich, [arXiv:1804.05664](https://arxiv.org/abs/1804.05664)
11. R. Penco, [arXiv:2006.16285](https://arxiv.org/abs/2006.16285)
12. F. Strocchi, *Symmetry Breaking*. Theoretical and Mathematical Physics (Springer Nature Switzerland AG, 2021). DOI [10.1007/978-3-662-62166-0](https://doi.org/10.1007/978-3-662-62166-0)
13. G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, in *Advances in Particle Physics*, vol. II, ed. by R.L. Cool, R.E. Marshak (Wiley, New York, NY, USA, 1968), pp. 567–708
14. A.J. Beekman, L. Rademaker, J. van Wezel, SciPost Phys. Lect. Notes **11** (2019). DOI [10.21468/SciPostPhysLectNotes.11](https://doi.org/10.21468/SciPostPhysLectNotes.11)
15. D. Naegels, [arXiv:2110.14504](https://arxiv.org/abs/2110.14504)
16. H. Watanabe, Ann. Rev. Cond. Mat. Phys. **11**, 169 (2020). DOI [10.1146/annurev-conmatphys-031119-050644](https://doi.org/10.1146/annurev-conmatphys-031119-050644)

Chapter 2

Our First Model

The purpose of this chapter is to introduce the uninitiated reader to the key concepts underlying the book. I assume that the reader has taken a first course on quantum field theory, but has not necessarily heard about *spontaneous symmetry breaking* (SSB). The field theory tools we are going to need include identification of a continuous symmetry and the corresponding conserved current via Noether's theorem, extraction of interaction vertices from the Lagrangian and their representation by Feynman diagrams, and their use in a perturbative calculation of scattering amplitudes.

I will start straight away by introducing a simple toy model for a complex scalar field ϕ and a Dirac field Ψ , defined by the Lagrangian density

$$\begin{aligned} \mathcal{L} = & \partial_\mu \phi^* \partial^\mu \phi + m^2 \phi^* \phi - \lambda (\phi^* \phi)^2 + (\epsilon^* \phi + \epsilon \phi^*) \\ & + \bar{\Psi} i \not{\partial} \Psi - g (\bar{\Psi}_L \Psi_R \phi + \bar{\Psi}_R \Psi_L \phi^*) . \end{aligned} \quad (2.1)$$

Here m , λ and g are parameters that are assumed to be real and positive but otherwise arbitrary. The parameter ϵ is complex and should be sufficiently small in a sense made more precise below. Finally, the usual slash notation indicates contraction of a Lorentz vector with Dirac γ -matrices, thus $\not{\partial} \equiv \gamma^\mu \partial_\mu$. I am not going to explain the motivation for the choice of this particular model. Let us rather take this as an invitation to explore with open mind its interesting physical properties.

I have chosen the relativistic notation just to make the mathematical analysis of the model as simple as possible. It is not essential for qualitative understanding of the results; much of the calculation could be repeated with nonrelativistic (Schrödinger) fields. This would of course require some minor modifications of the Lagrangian (2.1). First, the kinetic terms for ϕ and Ψ would have to be changed. Second, the chiral components $\Psi_{L,R}$ of the Dirac field Ψ would have to be replaced with two species of nonrelativistic spin-1/2 fermions.

Apart from Poincaré symmetry, guaranteed by the manifestly relativistic setup, the model (2.1) has two natural symmetries,

$$\begin{aligned}
\phi &\rightarrow \phi, & \Psi &\rightarrow \exp(i\epsilon_V)\Psi & (\text{exact}), \\
\phi &\rightarrow \exp(-2i\epsilon_A)\phi, & \Psi &\rightarrow \exp(i\epsilon_A\gamma_5)\Psi & (\text{approximate}).
\end{aligned} \tag{2.2}$$

In the context of relativistic theories with fermionic degrees of freedom, the transformations with the parameters $\epsilon_{V,A}$ are known respectively as *vector* and *axial*. The axial transformation only becomes a true symmetry of the Lagrangian (2.1) in the limit $\epsilon \rightarrow 0$. For small nonzero ϵ , it is therefore sensible to think of it as an approximate symmetry. By means of Noether's theorem, the vector and axial symmetries imply the existence of the following currents and conservation laws,

$$\begin{aligned}
J_V^\mu &= \bar{\Psi}\gamma^\mu\Psi, & J_A^\mu &= \bar{\Psi}\gamma^\mu\gamma_5\Psi - 2i(\phi^*\partial^\mu\phi - \partial^\mu\phi^*\phi), \\
\partial_\mu J_V^\mu &= 0, & \partial_\mu J_A^\mu &= 2i(\epsilon^*\phi - \epsilon\phi^*).
\end{aligned} \tag{2.3}$$

The expression for $\partial_\mu J_A^\mu$ confirms the observation that the axial symmetry becomes exact in the limit $\epsilon \rightarrow 0$.

In the rest of this chapter, we will analyze the physical properties of our toy model. To that end, we will utilize the basic workflow of perturbative quantum field theory, rooted in the theory of oscillations of mechanical systems [1]:

- Find the ground state (Sect. 2.1). This is the mandatory first step for any quantum system and is carried out by minimizing the (classical) Hamiltonian of the model.
- Identify the spectrum of excitations above the ground state (Sect. 2.2). This is based on the part of the Lagrangian, bilinear in the fluctuations of the fields ϕ, Ψ around the ground state.
- Work out the physical consequences of interactions of the excitations (Sect. 2.2). This follows from the part of the Lagrangian of higher orders in the fluctuations.

Finally, in Sect. 2.3 we will see how the most distinguishing features of the model can be captured by a low-energy *effective field theory* (EFT).

2.1 Spontaneous Symmetry Breaking

As the first step, we would like to find the (classical) ground state of the model (2.1). To that end, we need the classical Hamiltonian density,

$$\mathcal{H} = \partial_0\phi^*\partial_0\phi + \nabla\phi^* \cdot \nabla\phi + V(\phi, \phi^*) - \bar{\Psi}i\gamma \cdot \nabla\Psi + g(\bar{\Psi}_L\Psi_R\phi + \bar{\Psi}_R\Psi_L\phi^*), \tag{2.4}$$

where the scalar potential $V(\phi, \phi^*)$ is given by

$$V(\phi, \phi^*) \equiv -m^2\phi^*\phi + \lambda(\phi^*\phi)^2 - (\epsilon^*\phi + \epsilon\phi^*). \tag{2.5}$$

In the ground state of any quantum system, the *vacuum expectation value* (VEV) of a fermionic field must be zero. We can therefore focus on the scalar sector of the Hamiltonian (2.4). The part thereof containing derivatives of ϕ, ϕ^* is positive-semidefinite. Hence, the lowest-energy state will be such that the VEV of $\phi, \langle\phi\rangle$, is

a coordinate-independent constant. The value of this constant is determined by minimizing the potential $V(\phi, \phi^*)$. It is rather obvious that the ground state is nontrivial in that $\langle \phi \rangle \neq 0$. First, for any nonzero ϵ , the first partial derivatives of $V(\phi, \phi^*)$ at $\phi = \phi^* = 0$ are nonzero as well. Second, even in the limit $\epsilon \rightarrow 0$, the Hessian matrix of $V(\phi, \phi^*)$ at $\phi = \phi^* = 0$ is negative-definite. In the jargon of high-energy physics, this is a consequence of the “mass term” $m^2 \phi^* \phi$ in (2.1) having a “wrong sign.”

What is the ground state then? The VEV $\langle \phi \rangle$ must be a solution to the condition $0 = \partial V / \partial \phi^* = -m^2 \phi + 2\lambda \phi (\phi^* \phi) - \epsilon$. It must therefore have the same complex phase as ϵ , possibly up to an overall minus sign. Let us introduce the notation

$$\epsilon = \frac{w}{\sqrt{2}} e^{i\theta}, \quad \langle \phi \rangle = \frac{v}{\sqrt{2}} e^{i\theta}, \quad (2.6)$$

where θ is the common complex phase and v, w are real. The value of v characterizing the scalar VEV is then related to the parameter w by

$$v(\lambda v^2 - m^2) = w. \quad (2.7)$$

Being a cubic equation, (2.7) has up to three real solutions for v . In the limit $w \rightarrow 0$, these are $v = 0, \pm m/\sqrt{\lambda}$. Note that for $\epsilon = 0$, the potential $V(\phi, \phi^*)$ only depends on $\phi^* \phi$ and thus is insensitive to the phase θ . There is then a continuum of states of lowest energy, having $v = \pm m/\sqrt{\lambda}$ and arbitrary θ . This can be traced to the fact that for $\epsilon = 0$, the axial transformation in (2.2), which changes the phase of ϕ , is an exact symmetry of the Lagrangian (2.1). For any field configuration $\phi(x)$, the field $e^{i\theta} \phi(x)$ with constant θ therefore has the same energy as $\phi(x)$ itself. The existence of degenerate ground states is a hallmark of SSB. The defining property of SSB, in the context of our toy model, is that the VEV $\langle \phi \rangle$ is not invariant under the axial transformation (2.2). The ground state has a lower symmetry than the Lagrangian.

The reader might be concerned that having multiple degenerate ground states could lead to subtleties. (It does.) In any case, this was not anticipated by our algorithmic workflow outlined below (2.3). A simple workaround is to keep nonzero ϵ , which ensures the existence of a unique state of lowest energy. For positive w , this is the solution of (2.7) with the highest value of v . Using the closed formula for the solutions of a cubic equation would be impractical. It is however easy to find an approximate expression for v as long as w is small enough. One just needs to rewrite (2.7) in a form suitable for iteration,

$$v = \frac{m}{\sqrt{\lambda}} \sqrt{1 + \frac{w}{m^2 v}} = \frac{m}{\sqrt{\lambda}} \left(1 + \frac{\sqrt{\lambda} w}{2m^3} + \dots \right) = \frac{m}{\sqrt{\lambda}} + \frac{w}{2m^2} + \dots. \quad (2.8)$$

We end up with an infinite series with the expansion parameter $\sqrt{\lambda} w / m^3$. This gives us a precise mathematical condition for w being “small enough,” $w \ll m^3 / \sqrt{\lambda}$. The condition guarantees that the value of v in the ground state differs very little from $m/\sqrt{\lambda}$. In other words, ϵ now acts merely as a perturbation that enables us to pin down a unique ground state, without modifying this state appreciably. In the following, I will implicitly assume that the condition $w \ll m^3 / \sqrt{\lambda}$ is satisfied.

2.2 Nambu–Goldstone Boson and Its Interactions

The next step is to identify the spectrum of excitations above our ground state, and to work out the consequences of their interactions. To that end, we need to find field variables with vanishing VEV that make the bilinear part of the Lagrangian diagonal. Such variables correspond to the *normal modes*, well-known from the classical theory of small oscillations [1]. Upon expansion in powers of the new fields, the bilinear part of the Lagrangian determines the spectrum of normal modes, and the higher-order parts their interactions.

The change of variables to the normal modes can be thought of as a choice of parameterization of ϕ . (The Dirac field Ψ already has a vanishing VEV, so we do not have to do anything about it.) There is no a priori preferred choice of parameterization. Below, I will illustrate two different options, one of which is intuitively natural, whereas the other is more physical and practically convenient.

2.2.1 Linear Parameterization

One obvious possibility how to parameterize ϕ is to shift it by its VEV, and represent the resulting complex field in terms of its real and imaginary parts. This is still not completely unambiguous, but a smart choice is for instance¹

$$\phi(x) = \frac{e^{i\theta}}{\sqrt{2}} [v + \chi(x) + i\pi(x)] . \quad (2.9)$$

Taking out an overall factor $e^{i\theta}$ ensures that the phase θ drops out of the Lagrangian. Inserting this parameterization into (2.1), we get a Lagrangian in terms of χ, π, Ψ that can be organized by the dependence on the scalar fields,

$$\mathcal{L} = \mathcal{L}_{\text{vac}} + \mathcal{L}_{\text{bilin}} + \mathcal{L}_{\text{int}} + \mathcal{L}_{\Psi} . \quad (2.10)$$

The piece independent of the fields, $\mathcal{L}_{\text{vac}} = (3/4)\lambda v^4 - (1/2)m^2 v^2$, determines up to a minus sign the energy density of the ground state. There are no terms linear in χ, π ; this is ensured by their definition (2.9) through removing from ϕ its VEV. The bilinear part of the Lagrangian is of greatest interest to us,

$$\mathcal{L}_{\text{bilin}} = \frac{1}{2}(\partial_\mu \chi)^2 - \frac{1}{2}m_\chi^2 \chi^2 + \frac{1}{2}(\partial_\mu \pi)^2 - \frac{1}{2}m_\pi^2 \pi^2 , \quad (2.11)$$

where the mass parameters are given by

$$m_\chi^2 = 2m^2 + \frac{3w}{v} , \quad m_\pi^2 = \frac{w}{v} . \quad (2.12)$$

¹ The somewhat unusual notation is required for compatibility with the later discussion in Chap. 7.

An obvious corollary is that in the limit of exact axial symmetry ($\epsilon = 0$), the field π becomes massless. This is our first *Nambu–Goldstone* (NG) *boson*. It is easy to see the origin of the massless mode in the spectrum. With a canonically normalized kinetic term for ϕ in the Lagrangian, the mass spectrum is determined by the eigenvalues of the Hessian matrix of the potential $V(\phi, \phi^*)$ in the ground state. But I have already argued that for any $\phi(x)$, changing its overall phase will give a configuration of the same energy. Hence for any $\phi \neq 0$, there is a direction in the field space in which the potential does not change. This guarantees that the Hessian has an eigenvector with zero eigenvalue. The existence of the NG boson is therefore a direct consequence of the axial symmetry and its spontaneous breakdown, independent of the specific Lagrangian. This is the essence of *Goldstone’s theorem* [2, 3].

Equation (2.12) likewise tells us that in the limit of exact axial symmetry, that is $\epsilon = 0$, the mass of χ approaches $m\sqrt{2}$. The massive counterpart of the NG boson is usually referred to as the *Higgs boson* (or *Higgs mode*). The interaction part of the scalar Lagrangian in (2.10) consists of cubic and quartic terms,

$$\mathcal{L}_{\text{int}} = -\lambda v \chi (\chi^2 + \pi^2) - \frac{\lambda}{4} (\chi^2 + \pi^2)^2 . \quad (2.13)$$

Finally, the fermion part of the Lagrangian reads

$$\mathcal{L}_\Psi = \bar{\psi} (i\not{\partial} - m_\psi) \psi - \frac{g}{\sqrt{2}} \bar{\psi} (\chi + i\gamma_5 \pi) \psi , \quad (2.14)$$

where I have redefined the Dirac field by $\Psi = \exp [-(i/2)\theta\gamma_5] \psi$ in order to bring the kinetic term into the standard Dirac form. We can see that the originally massless fermion has acquired mass, $m_\psi = gv/\sqrt{2}$. This is an example of generation of fermion mass by spontaneous breaking of chiral symmetry, which plays an important role in the Standard Model of particle physics.

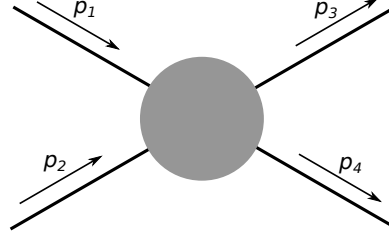
2.2.2 Scattering of Nambu–Goldstone Bosons

Now that we know the mass spectrum of our model, let us see how the different particles interact with each other. We are going to need Feynman rules for the interaction vertices, encoded in the Lagrangians (2.13) and (2.14),

$$\begin{array}{llll} \text{---} \bullet \text{---} & = -6i\lambda v , & \text{---} \bullet \text{---} & = -2i\lambda v , & \text{---} \bullet \text{---} & = -\frac{ig}{\sqrt{2}} , & \text{---} \bullet \text{---} & = \frac{g\gamma_5}{\sqrt{2}} , \\ \text{---} \bullet \text{---} & = -6i\lambda , & \text{---} \bullet \text{---} & = -2i\lambda , & \text{---} \bullet \text{---} & = -6i\lambda . & & \end{array} \quad (2.15)$$

The undirected solid and dashed lines represent respectively χ and π . The oriented solid lines stand for the fermion ψ . These Feynman rules for the interaction vertices

Fig. 2.1 Kinematics of the four-particle scattering processes discussed in this chapter. The arrows indicate flow of energy–momentum. In other words, the initial state includes two particles with momenta $\mathbf{p}_1, \mathbf{p}_2$, whereas the final state two particles with momenta $\mathbf{p}_3, \mathbf{p}_4$



must be augmented with standard relativistic propagators for χ, π, ψ with respective masses m_χ, m_π, m_ψ .

I will work out a few sample scattering processes including the NG boson π . All of these processes will have the simple four-particle kinematics displayed in Fig. 2.1. In all cases, I will use the same notation p_1, p_2 for the energy–momentum of the incoming particles and p_3, p_4 for the energy–momentum of the outgoing particles. For the sake of brevity, I will also use the Lorentz-invariant *Mandelstam variables* $s \equiv (p_1 + p_2)^2$, $t \equiv (p_1 - p_3)^2$ and $u \equiv (p_1 - p_4)^2$. As a consequence of energy and momentum conservation, the Mandelstam variables satisfy the constraint $s + t + u = m_1^2 + m_2^2 + m_3^2 + m_4^2$, where m_1, m_2, m_3, m_4 are the masses of the particles participating in the scattering process.

Let us first inspect $\pi\pi \rightarrow \pi\pi$ scattering. The invariant amplitude for this process is given in terms of Feynman diagrams by

$$-i\mathcal{A}_{\pi\pi \rightarrow \pi\pi} = \text{diagram 1} + \text{diagram 2} + \text{diagram 3} + \text{diagram 4} \quad (2.16)$$

Using the Feynman rules (2.15), this evaluates to

$$\mathcal{A}_{\pi\pi \rightarrow \pi\pi} = 6\lambda + 4\lambda^2 v^2 \left(\frac{1}{s - m_\chi^2} + \frac{1}{t - m_\chi^2} + \frac{1}{u - m_\chi^2} \right). \quad (2.17)$$

Of particular interest is the behavior of this amplitude at low energies. Given that the NG boson π becomes massless in the limit $\epsilon \rightarrow 0$, we can certainly find a kinematical regime such that $s, t, u \ll m_\chi^2$ for sufficiently small w . It then makes sense to expand the amplitude in powers of the Mandelstam variables. Upon a bit of manipulation using (2.7) and (2.12), one finds

$$\mathcal{A}_{\pi\pi \rightarrow \pi\pi} = -\frac{4\lambda w}{m_\chi^4 v} \left(m^2 - \frac{w}{2v} \right) - \frac{4\lambda^2 v^2}{m_\chi^6} (s^2 + t^2 + u^2) + \mathcal{O}(s^3, t^3, u^3). \quad (2.18)$$

Interestingly, the leading term vanishes in the limit $\epsilon \rightarrow 0$. But that is not all. In the same limit, we find that $s = 2p_1 \cdot p_2 = 2p_3 \cdot p_4$, $t = -2p_1 \cdot p_3 = -2p_2 \cdot p_4$ and $u = -2p_1 \cdot p_4 = -2p_2 \cdot p_3$. If we now take the additional limit in which the momentum

of any of the four particles goes to zero, then all the Mandelstam variables, and thus the amplitude $\mathcal{A}_{\pi\pi\rightarrow\pi\pi}$, will vanish. This property is concealed in (2.17), where all the four individual contributions coming from the diagrams in (2.16) have a nonzero limit. The algebraic cancellation leading to the eventual vanishing of the amplitude requires an interplay of the cubic and quartic interactions in the Lagrangian. This seems to be too much of a coincidence. In fact, it is our first example of the *Adler zero* principle: scattering amplitudes of a NG boson vanish in the limit that its momentum goes to zero. In plain terms, low-energy NG bosons interact weakly.

I have just made a much stronger statement than what the single amplitude worked out so far would seem to justify. We need more examples to check whether the Adler zero principle actually holds. To keep things simple, I will from now on take the limit of exact axial symmetry, $\epsilon = 0$, in which the NG boson π is exactly massless. The next example we will look at is $\chi\pi \rightarrow \chi\pi$ scattering, with the amplitude

$$-i\mathcal{A}_{\chi\pi\rightarrow\chi\pi} = \text{[Four Feynman diagrams]} . \quad (2.19)$$

Here a straightforward application of the Feynman rules gives

$$\mathcal{A}_{\chi\pi\rightarrow\chi\pi} = 2\lambda + 4\lambda^2 v^2 \left(\frac{1}{s} + \frac{3}{t - m_\chi^2} + \frac{1}{u} \right) . \quad (2.20)$$

Suppose that we again take the momentum of one of the NG bosons, p_2 or p_4 , to zero. In this limit, $s, u \rightarrow m_\chi^2 = 2m^2$ whereas $t \rightarrow 0$. Using finally the fact that $v = m/\sqrt{\lambda}$, it is easy to see that the whole amplitude $\mathcal{A}_{\chi\pi\rightarrow\chi\pi}$ vanishes.

We are starting to see a pattern. We have now checked two different scattering processes, which together involve all the scalar interaction vertices in (2.15) except for the χ^4 one. The Adler zero principle seems to hold, but it emerges out of a nontrivial interplay of different interaction terms and a cancellation among different Feynman diagrams.

As the last example, consider the scattering of a NG boson off a fermion, $\psi\pi \rightarrow \psi\pi$. The invariant amplitude for this process is

$$-i\mathcal{A}_{\psi\pi\rightarrow\psi\pi} = \text{[Three Feynman diagrams]} . \quad (2.21)$$

The Feynman rules (2.15) now give us

$$\begin{aligned} \mathcal{A}_{\psi\pi\rightarrow\psi\pi} = & \frac{g^2}{2} \bar{u}(p_3) \left(\frac{1}{\not{p}_1 + \not{p}_2 + m_\psi} + \frac{1}{\not{p}_3 - \not{p}_2 + m_\psi} \right) u(p_1) \\ & + \frac{\sqrt{2}\lambda g v}{t - m_\chi^2} \bar{u}(p_3) u(p_1) , \end{aligned} \quad (2.22)$$

where $u(p_1)$ and $u(p_3)$ are the usual plane-wave Dirac spinors with polarization indices suppressed for brevity. In order to further simplify the expression for the amplitude, we invert the Dirac propagators and use the Dirac equation for the Dirac spinors, $(\not{p}_1 - m_\psi)u(p_1) = \bar{u}(p_3)(\not{p}_3 - m_\psi) = 0$. Using finally the fact that $s - m_\psi^2 = 2p_1 \cdot p_2$ and $u - m_\psi^2 = -2p_2 \cdot p_3$, the amplitude (2.22) can be cast as

$$\mathcal{A}_{\psi\pi\rightarrow\psi\pi} = \frac{g^2}{4} \bar{u}(p_3) \left(\frac{\not{p}_2}{p_1 \cdot p_2} + \frac{\not{p}_2}{p_2 \cdot p_3} \right) u(p_1) + \frac{\sqrt{2}\lambda g v}{t - m_\chi^2} \bar{u}(p_3) u(p_1) . \quad (2.23)$$

This form is suitable for checking the limit $p_2 \rightarrow 0$. Should one prefer to take the momentum of the other NG boson, p_4 , to zero, it is possible to replace the expression in the large parentheses with $\not{p}_4/(p_3 \cdot p_4) + \not{p}_4/(p_1 \cdot p_4)$. Either way, the amplitude does not go to zero in the limit of vanishing momentum of the NG boson, in contrast to the previously discussed scalar amplitudes. The problem lies in the first term in (2.23), which does not even have a well-defined limit when p_2 or p_4 is taken to zero. That is because the fermion propagators become singular in this limit. I will only be able to offer a proper discussion of this issue in Chap. 10. For the moment, it suffices to keep in mind that the Adler zero principle has exceptions.

2.2.3 Nonlinear Parameterization

The whole preceding discussion of scattering amplitudes in our toy model was based on the linear parameterization (2.9). However, as already pointed out, this is not the only choice. Let us try something else. In elementary complex calculus, one learns about two different representations of complex numbers: the linear (“Cartesian”) and exponential (“polar”). I will thus replace (2.9) with

$$\phi(x) = \frac{e^{i[\theta + \pi(x)/v]}}{\sqrt{2}} [v + \chi(x)] . \quad (2.24)$$

With this parameterization, it is natural to think of χ as the fluctuation of the magnitude of ϕ , and of π as the fluctuation of its phase.

We can now follow the same steps as in Sect. 2.2.1. Using the partition (2.10) of the Lagrangian, we find that the constant part \mathcal{L}_{vac} and the bilinear part $\mathcal{L}_{\text{bilin}}$ remain unchanged. This follows from the fact that we are expanding around the same ground state as before, and that the parameterizations (2.9) and (2.24) agree to first order in χ, π . The scalar interaction Lagrangian is now different though,²

$$\mathcal{L}_{\text{int}} = -\lambda v \chi^3 - \frac{\lambda}{4} \chi^4 + \left(\frac{\chi}{v} + \frac{\chi^2}{2v^2} \right) (\partial_\mu \pi)^2 . \quad (2.25)$$

² Remember that I have already set $\epsilon = 0$. Otherwise, the interaction Lagrangian would include additional terms proportional to w . This is in contrast to (2.13), which is correct for any $\epsilon \in \mathbb{C}$.

Likewise, the fermionic part of the Lagrangian (2.14) turns into

$$\mathcal{L}_\Psi = \bar{\psi}(i\not{\partial} - m_\psi)\psi - \frac{g}{\sqrt{2}}\chi\bar{\psi}\psi + \frac{\partial_\mu\pi}{2v}\bar{\psi}\gamma^\mu\gamma_5\psi, \quad (2.26)$$

where I have redefined the fermion field as $\Psi = \exp[-(i/2)(\theta + \pi/v)\gamma_5]\psi$.

Note that the NG field π now enters the Lagrangian only through its derivatives. This remarkable feature is a direct consequence of the exponential parameterization (2.24). Namely, since $\pi(x)$ appears therein together with the constant phase θ , any constant field π must drop out of the Lagrangian as a consequence of the axial symmetry. Another way to look at this is that the axial transformation (2.2) acts on the fields χ, π, ψ via

$$\chi \rightarrow \chi, \quad \pi \rightarrow \pi - 2v\epsilon_A, \quad \psi \rightarrow \psi. \quad (2.27)$$

While χ and ψ are left intact by the axial transformation, π is shifted by a constant. Hence, there is no other way to construct a Lagrangian invariant under the axial symmetry than to forbid operators containing π without any derivatives. This is typical for NG bosons, and it gives us insight into the origin of the Adler zero principle. If the NG field only interacts via its derivatives, it should not be surprising that its interactions are suppressed at low energies.

Let us check how this works explicitly. We will need the Feynman rules for the interaction vertices contained in (2.25) and (2.26),

$$\begin{aligned} \text{---} \bullet \text{---} &= -6i\lambda v, & \text{---} \bullet \text{---} &= -\frac{2i}{v}p \cdot q, & \text{---} \bullet \text{---} &= -\frac{ig}{\sqrt{2}}, \\ \text{---} \bullet \text{---} &= \frac{1}{2v}\not{p}\gamma_5, & \text{---} \bullet \text{---} &= -6i\lambda, & \text{---} \bullet \text{---} &= -\frac{2i}{v^2}p \cdot q. \end{aligned} \quad (2.28)$$

The notation is such that p, q are energy–momenta carried by the NG legs, oriented towards the vertex. A comparison with (2.15) shows that, unsurprisingly, the interaction vertices involving only χ or ψ remain unchanged. The vertices of π now depend on energy–momentum though.

Since there is no π^4 vertex, the $\pi\pi \rightarrow \pi\pi$ amplitude is now given by the last three diagrams in (2.16). A brief calculation leads to

$$\mathcal{A}_{\pi\pi \rightarrow \pi\pi} = \frac{1}{v^2} \left(\frac{s^2}{s - m_\chi^2} + \frac{t^2}{t - m_\chi^2} + \frac{u^2}{u - m_\chi^2} \right), \quad (2.29)$$

which is easily seen to be equivalent to our previous result (2.17) in the limit $\epsilon \rightarrow 0$. Note, however, that in the exponential parameterization (2.24), the Adler zero property is manifest. Each of the three contributions to (2.29) vanishes separately when the momentum of any of the four particles is taken to zero. There is no need for cancellation between different Feynman diagrams. This is a direct consequence of the fact that the NG field π is now derivatively coupled. An important lesson

of this exercise is that while the scattering amplitude can be calculated in any parameterization of ϕ , not all parameterizations are equal. Indeed, some make the physical properties of the amplitude manifest.

I have glossed over the fact that the scattering amplitude $\mathcal{A}_{\pi\pi\rightarrow\pi\pi}$, and in fact the whole S -matrix, is independent of the choice of parameterization of ϕ around its VEV. This may seem intuitively obvious. After all, it is common lore that physical predictions of a theory should not depend on arbitrary choices such as a reference frame or a coordinate system. Yet, the mathematical proof of such *reparameterization invariance* of the S -matrix is nontrivial [4, 5]. Moreover, the independence on the choice of parameterization really only applies to physical observables. Quantities that are in principle unobservable, such as off-shell Green's functions of fields, may differ in different parameterizations. Finally, verifying the reparameterization invariance of the S -matrix explicitly beyond the tree-level (classical) approximation may require taking carefully account of the Jacobian of the functional integral measure [6].

The next amplitude to check is that for the $\chi\pi \rightarrow \chi\pi$ process. In this case, the same four diagrams as in (2.19) contribute, although the individual graphs of course take different values in the two parameterizations (2.9) and (2.24). A straightforward application of the Feynman rules (2.28) leads to

$$\mathcal{A}_{\chi\pi\rightarrow\chi\pi} = \frac{1}{v^2} \left[t + \frac{(s - m_\chi^2)^2}{s} + \frac{3m_\chi^2 t}{t - m_\chi^2} + \frac{(u - m_\chi^2)^2}{u} \right], \quad (2.30)$$

which is equivalent to the previous result (2.20). In its present form, the amplitude however vanishes manifestly, diagram by diagram, in the limit where the momentum of one of the NG bosons is taken to zero.

The last amplitude we are interested in is for $\psi\pi \rightarrow \psi\pi$. Again the same diagrams as in (2.21) contribute, but again they take individually different values than before due to the different Feynman rules (2.28). The result can be written in a number of different forms; one that is not too far from a direct application of Feynman rules is

$$\begin{aligned} \mathcal{A}_{\psi\pi\rightarrow\psi\pi} = & \frac{1}{8v^2} \bar{u}(p_3) \left[\frac{\not{p}_4(\not{p}_1 - m_\psi)\not{p}_2}{p_1 \cdot p_2} - \frac{\not{p}_2(\not{p}_1 - m_\psi)\not{p}_4}{p_2 \cdot p_3} \right] u(p_1) \\ & + \frac{g}{\sqrt{2}v} \frac{t}{t - m_\chi^2} \bar{u}(p_3) u(p_1). \end{aligned} \quad (2.31)$$

This time it takes more effort to prove equivalence with our previous result (2.23). I will spare the reader of the details, making use of the Dirac equation and the properties of the Dirac γ -matrices. What matters is that the contribution of every single diagram now carries a factor of p_2 and p_4 in the numerator, coming from the derivative couplings of the NG field π . Yet, this is not sufficient to make the amplitude vanish if either p_2 or p_4 goes to zero.

2.3 Low-Energy Effective Field Theory

The spectrum of our model (2.1) contains the massless NG boson π , the Higgs mode χ with mass $m_\chi = m\sqrt{2}$, and the Dirac fermion ψ with mass $m_\psi = gv/\sqrt{2} = mg/\sqrt{2\lambda}$. At energies well below the mass scales of m_χ and m_ψ , these massive modes will not be excited. As a consequence, the physics of the toy model will reduce to that of the NG boson π . Yet, in both parameterizations (2.9) and (2.24), the Higgs field is needed even at low energies since it mediates interactions between NG bosons. In fact, in the exponential parameterization (2.24), the Lagrangian does not contain any direct self-interaction of π at all. This sounds like an overkill. We should not need any other fields just to describe self-interactions of the NG bosons. In the spirit of EFT, introduced in Chap. 1, we should be able to describe the low-energy physics of NG bosons using the π field alone.

Such a low-energy EFT should respect the symmetries of the underlying theory defined by (2.1). With the transformation rules (2.27), this does not place any constraints on χ, ψ , which can thus be safely dropped. The shift transformation of π , on the other hand, forbids operators without derivatives on π . It follows that the effective Lagrangian for the NG boson must be described by some, a priori unknown, function of derivatives of π .

We also observed above that derivatives in the interaction vertices suppress their contribution to the S -matrix at low energies. More derivatives imply stronger suppression. This suggests an organization principle, whereby operators (with the same number of π fields) are hierarchically ordered according to the number of derivatives they contain. The dominant contributions to the S -matrix will come from operators with the fewest derivatives possible. Since each field π must still carry at least one derivative, the EFT will be dominated by interaction operators with *exactly* one derivative per each π . Lorentz invariance then constrains the effective Lagrangian to

$$\mathcal{L}_{\text{eff}} = \frac{1}{2}(\partial_\mu \pi)^2 + \sum_{n=2}^{\infty} c_{2n} [(\partial_\mu \pi)^2]^n + \dots, \quad (2.32)$$

where the ellipsis stands for operators with more than one derivative per π .

The as yet undetermined couplings c_{2n} govern the low-energy properties of scattering amplitudes of NG bosons. It remains to find out what these couplings are. In a concrete physical system, this could be done by performing a set of scattering experiments. It is however also possible to fix the values of c_{2n} theoretically from the underlying model (2.1). I will outline two strategies for doing so.

2.3.1 Matching

One possibility is to evaluate a set of scattering amplitudes, or other observables, in both the underlying model (2.1) and its low-energy EFT (2.32). Comparing the

predictions allows one to fix c_{2n} in terms of the parameters m, λ, g . In the jargon of EFT, this is called *matching*.

Let us see how it works in practice on the example of the $\pi\pi \rightarrow \pi\pi$ scattering. Within the EFT, this is described by the c_4 coupling. There is a single Feynman diagram, namely the first diagram in (2.16). The corresponding invariant amplitude reads, in terms of the Mandelstam variables, $\mathcal{A}_{\pi\pi \rightarrow \pi\pi} = -2c_4(s^2 + t^2 + u^2)$. Note how the dependence on the particle momenta exactly copies the leading contribution to the previously calculated amplitude (2.18) in the limit of exact axial symmetry. Upon comparing the coefficients of the kinematical invariant $s^2 + t^2 + u^2$, we find

$$c_4 = \frac{\lambda}{4m^4} . \quad (2.33)$$

The same procedure could in principle be followed to determine c_6, c_8 and so on. That would however be a tedious task, since the number of Feynman diagrams contributing to an n -particle process grows rapidly with n . Sometimes, an alternative approach is feasible whereby the EFT is deduced from the underlying theory directly at the level of the Lagrangian. I will now demonstrate how to do this in the case of our toy model.

A cautious reader might be wondering why I have made no mention of Feynman diagrams containing loops in all the discussion above. Within the toy model (2.1), restricting to tree level amounts to considering the leading contribution to scattering amplitudes in a power expansion in the small couplings λ, g . The result (2.33) should therefore likewise be interpreted as the leading contribution to c_4 , induced by the presence of the heavy modes χ and ψ . See Sect. 2.3 of [7] for more details.

2.3.2 Eliminating the Heavy Modes

In the exponential parameterization (2.24), all tree-level scattering amplitudes of NG bosons in the model (2.1) arise from Feynman diagrams that include virtual χ modes in the propagators. Recall now that quantum field theory at tree level is equivalent to its classical limit. Including the effect of virtual χ quanta is then equivalent to solving the classical *equation of motion* (EoM) for χ and inserting the result back to the Lagrangian. In this way, one can obtain an EFT for π alone that takes into account all the interactions of the original model (2.1).

Let us see this program through to its end. We start by dropping the fermion field and putting the scalar interactions (2.25) together with the kinetic term into a complete scalar Lagrangian,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \chi)^2 - m^2 \chi^2 - \lambda v \chi^3 - \frac{\lambda}{4} \chi^4 + \frac{1}{2} \left(1 + \frac{\chi}{v}\right)^2 (\partial_\mu \pi)^2 . \quad (2.34)$$

The parameters m, λ, v are related by $v = m/\sqrt{\lambda}$. The EoM for χ , which we would like to solve for χ in terms of π , reads

$$\square\chi + 2m^2\chi + 3\lambda v\chi^2 + \lambda\chi^3 - \frac{1}{v}\left(1 + \frac{\chi}{v}\right)(\partial_\mu\pi)^2 = 0. \quad (2.35)$$

It is of course not possible to solve this equation in a closed form, we can however still extract useful information from it. At low energies, we expect χ to be very small. (The probability to excite virtual χ quanta far away from their mass shell is tiny.) In the first approximation, the terms in (2.35) quadratic and cubic in χ can therefore be neglected. We can likewise drop the $\square\chi$ term, suppressed by two derivatives. Equation (2.35) then becomes a linear algebraic equation for χ with the solution $\chi \approx (\partial_\mu\pi)^2/(2m^2v)$. In order to go beyond this approximation, we rewrite (2.35) in a form suitable for iteration,

$$\chi = \frac{1}{v} \left[2m^2 + \square - \frac{1}{v^2}(\partial_\mu\pi)^2 + 3\lambda v\chi + \lambda\chi^2 \right]^{-1} (\partial_\mu\pi)^2. \quad (2.36)$$

All terms but $2m^2$ in the square brackets are small due to containing either derivatives or extra factors of χ . This makes it possible to iteratively expand (2.36) in inverse powers of m^2 up to any desired order in π and its derivatives. Up to and including the first subleading contribution to χ , we thus get

$$\chi = \frac{(\partial_\mu\pi)^2}{2m^2v} - \frac{[(\partial_\mu\pi)^2]^2}{8m^4v^3} + \dots. \quad (2.37)$$

When inserted back into (2.34), this is enough to generate the couplings c_4 and c_6 in the Lagrangian (2.32). When the dust settles, we find $c_4 = \lambda/(4m^4)$ and $c_6 = 0$. The result for c_4 agrees with our previous calculation relying on direct matching of scattering amplitudes. It may however come as a surprise that c_6 vanishes.

To get further insight, we multiply the EoM (2.35) by $\chi/2$ and add to the Lagrangian (2.34), giving

$$\mathcal{L} \simeq \frac{\lambda v}{2}\chi^3 + \frac{\lambda}{4}\chi^4 + \frac{1}{2}\left(1 + \frac{\chi}{v}\right)(\partial_\mu\pi)^2. \quad (2.38)$$

The symbol \simeq indicates that I have dropped a surface term. The obtained Lagrangian can be further manipulated by rewriting the EoM (2.35) as

$$\lambda\chi(\chi + 2v) = \frac{(\partial_\mu\pi)^2}{v^2} - \frac{\square\chi}{\chi + v}. \quad (2.39)$$

Using this twice in sequence, it is possible to remove from (2.38) all terms that only contain χ without any derivatives. One thus arrives at an equivalent Lagrangian,

$$\mathcal{L} \simeq \frac{1}{2}(\partial_\mu\pi)^2 + \frac{\lambda}{4m^4}[(\partial_\mu\pi)^2]^2 - \left[\frac{(\partial_\mu\pi)^2}{m^2} + \chi^2 \right] \frac{\square\chi}{4(\chi + v)}. \quad (2.40)$$

The first two terms here are just the kinetic term for π and the c_4 coupling. The last term, upon expansion in π using (2.36), contains only operators with more than one derivative per π . This allows one to make the strong conclusion that *all* the couplings in the effective Lagrangian (2.32) except for c_4 vanish.

2.4 Moral Lessons

In this introductory chapter, we have analyzed a simple toy model, introducing along the way some concepts that will play a key role throughout the rest of the book. Let me briefly summarize what we have found, and draw a few morals.

Lesson #1 Some physical systems have multiple degenerate ground states. Barring accidental degeneracy, this happens typically when the system possesses a symmetry under which the ground states are not invariant. This is the defining property of SSB.

Lesson #2 When the spontaneously broken symmetry is continuous, the spectrum of the system contains a massless particle: the NG boson. This is Goldstone's theorem.

Lesson #3 The scattering amplitude for a process involving a NG boson vanishes in the limit where its momentum is taken to zero. That is, NG bosons interact weakly at low energy. This is the Adler zero principle. The principle may be violated in case taking the NG boson momentum to zero brings some of the virtual particles in the process on the mass shell.

Lesson #4 It is convenient to choose a nonlinear field parameterization in which a constant NG field can be eliminated by a symmetry transformation. In such a parameterization, the NG field is derivatively coupled, making its physical properties of zero mass and weak interactions at low energy manifest.

Lesson #5 The nonlinear field parameterization allows one to construct a low-energy EFT in terms of the NG field alone. This captures the physics of NG bosons at energies well below the mass scale of other, massive particles present in the system.

References

1. H. Goldstein, C.P. Poole, J.L. Safko, *Classical Mechanics* (Pearson Education Limited, Harlow, UK, 2013)
2. J. Goldstone, *Nuovo Cim.* **19**, 154 (1961). DOI [10.1007/BF02812722](https://doi.org/10.1007/BF02812722)
3. J. Goldstone, A. Salam, S. Weinberg, *Phys. Rev.* **127**, 965 (1962). DOI [10.1103/PhysRev.127.965](https://doi.org/10.1103/PhysRev.127.965)
4. J.S.R. Chisholm, *Nucl. Phys.* **26**, 469 (1961). DOI [10.1016/0029-5582\(61\)90106-7](https://doi.org/10.1016/0029-5582(61)90106-7)
5. S. Kamefuchi, L. O'Raifeartaigh, A. Salam, *Nucl. Phys.* **28**, 529 (1961). DOI [10.1016/0029-5582\(61\)90056-6](https://doi.org/10.1016/0029-5582(61)90056-6)
6. J. Honerkamp, K. Meetz, *Phys. Rev.* **D3**, 1996 (1971). DOI [10.1103/PhysRevD.3.1996](https://doi.org/10.1103/PhysRevD.3.1996)
7. C.P. Burgess, *Introduction to Effective Field Theory* (Cambridge University Press, Cambridge, UK, 2021). DOI [10.1017/9781139048040](https://doi.org/10.1017/9781139048040)

Chapter 3

Generalizations of the Model

In the previous chapter, I worked out in detail a model of a single complex scalar field, in which a continuous symmetry is spontaneously broken. The corresponding axial transformations span the Abelian group $U(1)$. Due to its simplicity, such a model is suitable for a first encounter with *spontaneous symmetry breaking* (SSB). However, it falls short of exhibiting the full spectrum of possible realizations of SSB. In this chapter, I will therefore generalize the model in two aspects. First, I will allow for multiple scalar fields carrying a nontrivial representation of a possibly non-Abelian symmetry group. This will take us to the level of standard expositions of SSB in textbooks on quantum field theory, oriented towards high-energy physics. Second, I will pay off my debt to readers with other backgrounds and show how the story changes if one gives up relativistic (Lorentz) invariance. This is particularly relevant for applications to condensed-matter physics, but also concerns dense relativistic matter that one deals with, for instance, in astrophysics and cosmology.

The primary goal of this chapter is to introduce the reader to the intricate interplay between the pattern of SSB and the spectrum of the associated *Nambu–Goldstone* (NG) *bosons*. This goes a long way towards a broad qualitative understanding of the behavior of physical systems with SSB at low energies. Namely, as explained in the previous chapter, NG bosons tend to interact weakly. In the absence of other massless particles, the low-temperature thermodynamics of the system will then be accurately described in terms of a free gas of NG bosons. The latter is completely characterized by the number of NG bosons in the spectrum and their dispersion relations.

Within the landscape of relativistic field theory, a NG boson is always a massless particle whose dispersion relation is fixed by Lorentz invariance. Section 3.1 illustrates how, in this case, the number of NG bosons can be determined solely from group theory. As soon as we give up Lorentz invariance, however, interesting things start to happen. Importantly, the NG state in the spectrum requires more data to specify than mere rest mass. Section 3.2 shows that the asymptotic behavior of the dispersion relation of NG bosons at low momentum is intimately related to their number. Within this chapter, I will not discuss interactions of NG bosons any further, since there are no qualitative changes compared to the toy model of Chap. 2.

3.1 Relativistic Models with Non-Abelian Symmetry

Suppose we are given a set of real scalar fields, ϕ^i . Similarly to the previous chapter, I will introduce at once a class of toy model Lagrangians with a fixed kinetic term and a generic potential,

$$\mathcal{L} = \frac{1}{2} \delta_{ij} \partial_\mu \phi^i \partial^\mu \phi^j - V(\phi) . \quad (3.1)$$

Here $V(\phi)$ is an a priori arbitrary function of ϕ^i , only restricted by the requirements that it is bounded from below and its Taylor expansion around $\phi^i = 0$ starts at the quadratic order. Since any complex scalar field can always be represented by two real ones, the class of Lagrangians (3.1) includes our original toy model (2.1). I have however dropped the fermion sector of (2.1). I have also set to zero the terms proportional to ϵ and ϵ^* , as they only serve to select a unique ground state. With the experience gathered in Chap. 2, the reader should by now be comfortable with degenerate ground states.

The symmetry of the theory defined by (3.1) will play a key role throughout this chapter. I will assume that the fields ϕ^i carry a linear representation of some Lie group G . Invariance of the whole Lagrangian (3.1) under transformations from G requires that the kinetic term and the potential are invariant separately. Let the total number of scalar fields be n . Then the kinetic term alone is invariant under continuous rotations of ϕ^i , forming the orthogonal group $\text{SO}(n)$. We therefore expect that G is a subgroup of $\text{SO}(n)$, depending on the concrete choice of potential. With this in mind, I will focus in the following on the symmetry of the potential $V(\phi)$ itself.

Suppose that the fields ϕ^i transform in some (real) representation \mathcal{R} of G . The invariance of the potential $V(\phi)$ under G is expressed by the condition

$$V(\mathcal{R}(g)\phi) = V(\phi) \quad \text{for any } g \in G . \quad (3.2)$$

For continuous symmetries, it is usually more convenient to work with infinitesimal transformations. Denoting the generators of G as $Q_{A,B,\dots}$, the condition (3.2) then boils down to

$$\frac{\partial V(\phi)}{\partial \phi^i} \mathcal{R}(Q_A)^i{}_j \phi^j = 0 \quad \text{for any } Q_A \in \mathfrak{g} , \quad (3.3)$$

where \mathfrak{g} is the Lie algebra of G . This will be the starting point for the analysis of the spectrum of NG bosons below.

Example 3.1

It is straightforward to construct a potential that is invariant under the complete group of rotations, $\text{SO}(n)$. Indeed, any function of the quadratic invariant, $\Phi^2 \equiv \delta_{ij} \phi^i \phi^j$, will do. For instance, a generic $\text{SO}(n)$ -invariant quartic polynomial potential will take the form $V(\phi) = -m^2 \Phi^2 + \lambda \Phi^4$, which generalizes (2.5) from $n = 2$ to any n . The fields ϕ^i transform under the vector representation of $\text{SO}(n)$, in which $\mathcal{R}(g)$ are

$n \times n$ orthogonal matrices. The generators of $SO(n)$ are labeled by a pair of vector indices; up to overall normalization, $\mathcal{R}(Q_{kl})^i_j$ equals $\delta_k^i \delta_{jl} - \delta_l^i \delta_{jk}$.

3.1.1 Spectrum of Nambu–Goldstone Bosons

Let us choose a potential $V(\phi)$ so that there is a ground state in which some of the fields ϕ^i have a nonzero *vacuum expectation value* (VEV), $\langle \phi^i \rangle$. The elements of G that leave the VEVs unchanged form a group H , referred to as the *unbroken subgroup* of G ,

$$H \equiv \{h \in G \mid \mathcal{R}(h)^i_j \langle \phi^j \rangle = \langle \phi^i \rangle\} . \quad (3.4)$$

The generators of H will be labeled as $Q_{\alpha, \beta, \dots}$. As an immediate consequence of (3.4), such unbroken generators satisfy

$$\mathcal{R}(Q_{\alpha})^i_j \langle \phi^j \rangle = 0 \quad \text{for any } Q_{\alpha} \in \mathfrak{h} , \quad (3.5)$$

where \mathfrak{h} is the Lie algebra of H . In many physical systems, H is a proper subgroup of G . That is, there are transformations from G that do change some of the VEVs $\langle \phi^i \rangle$, hence also the ground state itself. Representing a symmetry of the Lagrangian (3.1) but not of the ground state, these are said to be spontaneously broken.

By definition of symmetry, acting with an element of G on any field configuration gives a (possibly different) field configuration with the same energy. Hence, as already observed in the previous chapter, the existence of symmetry transformations that do not leave the ground state invariant implies the existence of degenerate ground states. We also previously found that this is closely related to the presence of massless particles in the spectrum.

Let us see how the NG bosons emerge in the present general setting. We start by choosing a basis $\{Q_{\alpha}, Q_a\}$ of the Lie algebra \mathfrak{g} such that $\{Q_{\alpha}\}$ is a basis of \mathfrak{h} and $Q_a \notin \mathfrak{h}$, that is $\mathcal{R}(Q_a)^i_j \langle \phi^j \rangle \neq 0$. Next, we take a derivative of (3.3) with respect to ϕ . Upon some relabeling of indices, this becomes

$$\frac{\partial^2 V(\phi)}{\partial \phi^i \partial \phi^j} \mathcal{R}(Q_A)^j_k \phi^k + \frac{\partial V(\phi)}{\partial \phi^j} \mathcal{R}(Q_A)^j_i = 0 . \quad (3.6)$$

Finally, we evaluate this condition in the chosen ground state and use the fact that by definition, the first derivatives of $V(\phi)$ in the ground state vanish,

$$\left. \frac{\partial^2 V(\phi)}{\partial \phi^i \partial \phi^j} \right|_{\phi=\langle \phi \rangle} \mathcal{R}(Q_A)^j_k \langle \phi^k \rangle = 0 . \quad (3.7)$$

This does not say anything about the unbroken generators Q_{α} , for which the left-hand side identically vanishes. However, for the broken generators Q_a , this implies that $\mathcal{R}(Q_a)^i_j \langle \phi^j \rangle$ is an eigenvector of the Hessian matrix of $V(\phi)$ in the ground state

with zero eigenvalue (*zero mode*). Moreover, the set of eigenvectors $\mathcal{R}(Q_a)^i_j \langle \phi^j \rangle$ with all the different choices of Q_a is linearly independent. To see why, suppose that there are coefficients c^a such that $c^a \mathcal{R}(Q_a)^i_j \langle \phi^j \rangle = 0$. Then by (3.5), $c^a Q_a \in \mathfrak{h}$. But we have assumed that the set $\{Q_a, Q_a\}$ is linearly independent, hence $c_a = 0$.

This proves that for every broken symmetry generator Q_a , the Hessian matrix of $V(\phi)$ in the ground state has a corresponding independent zero mode. Now recall that with a canonically normalized kinetic term as in (3.1), the Hessian of $V(\phi)$ represents the mass matrix of the theory. That is, its eigenvalues give the squared masses of elementary one-particle excitations in the spectrum. Thus, there is one NG boson in the spectrum for each broken symmetry generator. The total number of NG bosons in the toy model (3.1) is $\dim G - \dim H$.

What I have just presented is (a simplified version of) one of the original proofs of Goldstone's theorem [1]. It is worth stressing that the number of zero modes of the Hessian matrix of the potential may be larger than $\dim G - \dim H$. First, I have not proven that the Hessian cannot have zero modes entirely unrelated to symmetry. (It can.) Second, it is possible that the symmetry group of the potential $V(\phi)$ alone is larger than that of the kinetic term, and thus of the whole Lagrangian (3.1). In both cases, a classical analysis as worked out in this and the preceding chapter may yield “fake” NG bosons. These are scalar fields that have no classical mass term in the Lagrangian, but acquire mass solely due to quantum corrections [2, 3]. As I will explain in Chap. 6, the vanishing of the mass of true NG bosons is guaranteed by SSB. This is an exact result that must also hold within any approximation that respects the symmetry of the given theory.

Let me stress that everything said so far relies on the kinetic term in the Lagrangian being a positive-definite quadratic form in the time derivatives of ϕ^i . Finding the mass spectrum of a scalar theory such as (3.1) is then equivalent to finding the normal modes using the theory of small oscillations of classical mechanical systems [4]. Section 3.2 revolves largely around the subtleties that arise when this requirement is not satisfied, which may happen once we depart from the landscape of Lorentz-invariant field theory.

Example 3.2

Consider a model of an n -plet of scalar fields ϕ^i whose Lagrangian is invariant under the group $G \simeq \text{SO}(n)$. Any ground state in which at least one of the VEVs $\langle \phi^i \rangle$ is nonzero will be invariant under the subgroup $H \simeq \text{SO}(n-1)$ of transformations that leave $\langle \phi^i \rangle$ fixed. In other words, H consists of field rotations in the $n-1$ directions orthogonal to $\langle \phi^i \rangle$. By the general argument developed above, we expect the spectrum of such a model to contain $\dim \text{SO}(n) - \dim \text{SO}(n-1) = n-1$ NG bosons. Note that the special case of $n=2$ correctly recovers the single NG boson we found in Chap. 2.

3.1.2 Low-Energy Effective Field Theory

On general grounds, we expect the dynamics of NG bosons to be captured by a low-energy *effective field theory* (EFT). In Sect. 2.3.2, I showed how to derive such an EFT explicitly from the underlying Lagrangian. Here we do not know the precise form of the potential $V(\phi)$ or its symmetry. It is however still possible to gain useful insight by following, if only schematically, the same argument as in Sect. 2.3.2.

The first step towards the EFT for NG bosons is a suitable choice of parameterization of the fields ϕ^i . The exponential parameterization (2.24) for a single complex field can be generalized to the set of real fields ϕ^i as

$$\phi^i(x) = U^i_j(\pi(x)) [\langle \phi^j \rangle + \chi^j(x)] . \quad (3.8)$$

Here $U(\pi)$ is a matrix taking values in the representation \mathcal{R} of G . It encodes a set of NG fields π^a , one for each broken generator Q_a . One can for instance choose the generators as real antisymmetric matrices and imagine that $U(\pi)$ is the orthogonal matrix $\exp[i\pi^a \mathcal{R}(Q_a)]$. But a precise form of $U(\pi)$ is not important. All that I will need is that when expanded in powers of π^a , the leading term is $U(0) = \mathbb{1}$ and the linear term is proportional to $\pi^a \mathcal{R}(Q_a)$. Finally, χ^i encodes a set of *Higgs modes* whose masses are expected to be nonzero.

In spite of the suggestive notation, the different components of χ^i are not linearly independent. By mere counting of degrees of freedom, the number of independent Higgs fields should be $n - \dim G + \dim H$. What exactly these independent linear combinations of χ^i are, can be identified using group theory. One just has to decompose the space of ϕ^i into irreducible representations of H and drop the representations corresponding to the NG bosons.

The advantage of the parameterization (3.8) is that the matrix $U(\pi)$ drops out of the potential $V(\phi)$, because the latter is by construction G -invariant, cf. (3.2). The fact that $U(\pi)$ depends implicitly on spacetime coordinates through the fields $\pi^a(x)$ does not need to bother us since the potential does not contain any derivatives. In this parameterization, the Lagrangian (3.1) will therefore have a similar structure to that in (2.34). The NG fields π^a will only enter through the kinetic term, that is together with derivatives. The nonderivative potential, on the other hand, will only depend on the Higgs fields χ^i . Using the shorthand notation ϕ, χ for the vectors ϕ^i, χ^i , the Lagrangian reads explicitly

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \partial_\mu \chi \cdot \partial^\mu \chi + \partial_\mu \chi \cdot U^T \partial^\mu U \cdot (\langle \phi \rangle + \chi) \\ & + \frac{1}{2} (\langle \phi \rangle + \chi) \cdot \partial_\mu U^T \partial^\mu U \cdot (\langle \phi \rangle + \chi) - V(\chi) . \end{aligned} \quad (3.9)$$

We would now like, at least in principle, to eliminate χ^i by using its *equation of motion* (EoM). It is important to make sure that the expansion of χ^i in powers of π^a generalizing (2.37) starts with a term with (at least) two derivatives and two powers of π^a . This in turn requires that the Lagrangian (3.9) does not contain any mixing term, linear in both ϕ^i and χ^i . The vanishing of such mixing terms can be achieved by choosing the vector χ to be orthogonal to all $\mathcal{R}(Q_a) \cdot \langle \phi \rangle$.

The rest of the argument is simple. Inserting the solution for χ^i in the Lagrangian (3.9) gives terms that contain at least four derivatives. The effective Lagrangian for the NG fields will however be dominated by operators with only two derivatives, since a higher number of derivatives means stronger suppression at lower energies. Such two-derivative terms are obtained by simply setting $\chi^i \rightarrow 0$ in (3.9),

$$\mathcal{L}_{\text{eff}} = \frac{1}{2} \langle \phi \rangle \cdot \partial_\mu U^T \partial^\mu U \cdot \langle \phi \rangle + \dots, \quad (3.10)$$

where the ellipsis stands for contributions with more than two derivatives. Remarkably, this leading contribution to the EFT arises solely from the kinetic term in (3.1). The concrete choice of potential $V(\phi)$ does not matter. All we need to know are the VEVs $\langle \phi^i \rangle$ and the ensuing *symmetry-breaking pattern* $G \rightarrow H$. These together fix the matrices $U(\pi)$ up to a choice of parameterization in terms of the NG fields π^a .

Example 3.3

Let me illustrate this on the case of spontaneous breaking $\text{SO}(n) \rightarrow \text{SO}(n-1)$ by an n -vector field ϕ , introduced in Example 3.2. Here $U(\pi)$ takes values in the defining (vector) representation of $\text{SO}(n)$. Thus, $U(\pi) \cdot \langle \phi \rangle$ is a vector of the same length as $\langle \phi \rangle$. By a suitable choice of units, we can make this length whatever we want. It is conventional to set $U(\pi) \cdot \langle \phi \rangle \equiv v \mathbf{n}(\pi)$, where \mathbf{n} is a unit vector field and v is a dimensionful constant. Then the effective Lagrangian (3.10) acquires the form

$$\mathcal{L}_{\text{eff}} = \frac{v^2}{2} \partial_\mu \mathbf{n} \cdot \partial^\mu \mathbf{n} + \dots, \quad (3.11)$$

which is known as the *nonlinear sigma model*. Note that in spite of the appearance, this is not a noninteracting field theory. The unit vector field \mathbf{n} takes values from the unit $(n-1)$ -sphere, S^{n-1} . The NG fields π^a can be thought of as $n-1$ independent coordinates on the sphere. Once a choice of coordinates is made and the Lagrangian (3.11) is expanded in powers of π^a , it is going to contain operators with two derivatives and an arbitrarily high number of NG fields.

This example underlines an important distinction between the broader class of theories (3.1) and the special case analyzed in Chap. 2, where the symmetry group $G \simeq \text{U}(1)$ was Abelian. In the latter case, the leading contribution to the effective Lagrangian contains operators with exactly one derivative on each NG field π . The part of the Lagrangian with two derivatives is then just the kinetic term, and any interactions necessarily come with four or more derivatives. Once G is allowed to be

non-Abelian, the two-derivative part of the Lagrangian (3.10) contains interactions bringing together an arbitrarily high number of NG bosons.

Equation (3.10) gives us the first hint that the form of the EFT for NG bosons is controlled by symmetry, regardless of the details of the underlying model. The Lagrangian can be cast solely in terms of the NG fields π^a ,

$$\begin{aligned}\mathcal{L}_{\text{eff}} &= \frac{1}{2} g_{ab}(\pi) \partial_\mu \pi^a \partial^\mu \pi^b + \dots, \\ g_{ab}(\pi) &\equiv \langle \phi \rangle \cdot \frac{\partial U(\pi)^T}{\partial \pi^a} \frac{\partial U(\pi)}{\partial \pi^b} \cdot \langle \phi \rangle.\end{aligned}\tag{3.12}$$

I will show in Chap. 8 that the matrix function $g_{ab}(\pi)$ can be interpreted in terms of the geometry of the Lie groups G and H . The dependence of $g_{ab}(\pi)$ on the NG fields is fixed by this geometry. All that is left of the concrete model leading to the EFT (3.12) is the numerical value of the constant matrix $g_{ab}(0)$.

3.2 Going Nonrelativistic

In all the examples worked out in the previous and this chapter so far, I restricted myself to relativistic, Lorentz-invariant field theories. This choice was based on technical simplicity. The notation using Lorentz indices is more compact, and the calculations of scattering amplitudes in terms of the relativistic Mandelstam variables are more transparent. There are good reasons to step out of the box, though. First, the great majority of natural phenomena occurs at energies well below the scale at which relativity starts to matter. In the spirit of EFT, such phenomena should therefore have an accurate description in terms of nonrelativistic field theory. Second, as we shall see in this section, relaxing the requirement of Lorentz invariance leads to rich phenomenology. The examples developed below will prepare the reader for the more thorough discussion of SSB that comes in Part II of the book.

Operationally, what I will do is to modify the previously discussed toy models in a way that respects invariance under spatial rotations, and spatial and temporal translations. While this does not place any restrictions on the potential $V(\phi)$, the kinetic term may now assume a more general form. The Lagrangian may then contain independent terms with either two spatial derivatives, or one or two temporal derivatives. The very possibility of adding terms with a single time derivative constitutes the main qualitative difference compared to relativistic field theory.

3.2.1 Single Schrödinger Field

In order to copy as closely as possible the previous discussion of relativistic field theory, let us start with the following model,

$$\mathcal{L} = 2iM\phi^*\partial_0\phi - \nabla\phi^* \cdot \nabla\phi + m^2\phi^*\phi - \lambda(\phi^*\phi)^2. \quad (3.13)$$

Here ϕ is a complex scalar field and ∂_0 stands for a time derivative. This Lagrangian is almost identical to the scalar sector of (2.1), except for the temporal part of the kinetic term. For $m = \lambda = 0$, the corresponding EoM is the Schrödinger equation for a free particle of mass M . It is therefore natural to refer to the field ϕ as a “Schrödinger field.”

The analysis of the model (3.13) now follows the same steps as in Chap. 2. The classical Hamiltonian density reads

$$\mathcal{H} = \nabla\phi^* \cdot \nabla\phi - m^2\phi^*\phi + \lambda(\phi^*\phi)^2. \quad (3.14)$$

The lowest energy is achieved by a coordinate-independent state such that

$$\langle\phi\rangle = \frac{v}{\sqrt{2}}e^{i\theta}, \quad v \equiv \frac{m}{\sqrt{\lambda}}, \quad (3.15)$$

where θ is an arbitrary phase, reflecting the ground state degeneracy. Making use of the exponential parameterization (2.24) brings the Lagrangian to the form

$$\mathcal{L} = \mathcal{L}_{\text{vac}} + \mathcal{L}_{\text{bilin}} + \mathcal{L}_{\text{int}}. \quad (3.16)$$

The vacuum piece equals $\mathcal{L}_{\text{vac}} = (1/2)m^2v^2 - (\lambda/4)v^4 = m^4/(4\lambda)$. The bilinear and interaction parts of the Lagrangian read, respectively,

$$\begin{aligned} \mathcal{L}_{\text{bilin}} &= -2M\chi\partial_0\pi - \frac{1}{2}(\nabla\chi)^2 - m^2\chi^2 - \frac{1}{2}(\nabla\pi)^2, \\ \mathcal{L}_{\text{int}} &= -\lambda v\chi^3 - \frac{\lambda}{4}\chi^4 - \frac{M}{v}\chi^2\partial_0\pi - \left(\frac{\chi}{v} + \frac{\chi^2}{2v^2}\right)(\nabla\pi)^2, \end{aligned} \quad (3.17)$$

where I have discarded some contributions that amount to a total time derivative.

Just like for the relativistic model (2.1), the NG field π enters the Lagrangian only with derivatives. This is an immediate consequence of the choice of parameterization (2.24). Here is where the similarity ends, though. It would be a mistake to conclude that π represents a NG boson and χ describes a massive, Higgs mode. The two fields are mixed by the bilinear term $-2M\chi\partial_0\pi$, which makes χ, π canonically conjugated to each other. We therefore expect that the spectrum of the model (3.13) contains just one type of excitation, which should then be a NG boson.

In order to describe this excitation more accurately, let us rewrite the bilinear part of the Lagrangian in a matrix form,

$$\mathcal{L}_{\text{bilin}} \simeq \frac{1}{2} \begin{pmatrix} \chi & \pi \end{pmatrix} \begin{pmatrix} \nabla^2 - 2m^2 & -2M\partial_0 \\ 2M\partial_0 & \nabla^2 \end{pmatrix} \begin{pmatrix} \chi \\ \pi \end{pmatrix}. \quad (3.18)$$

The symbol \simeq again indicates that I have dropped a surface term. What is entirely new compared to relativistic field theory is that the bilinear Lagrangian cannot be

diagonalized by any local field redefinition. This is just a minor nuisance when it comes to finding the spectrum of one-particle states. It is however a major complication if one wants to study scattering of the particles. Calculating a cross-section for a scattering process now requires a careful mapping of one-particle states to fields in the Lagrangian, and a revision of rules for phase space integration. Some discussion of these issues can be found for instance in [5–7].

In order to find the dispersion relation of the NG mode, we Fourier-transform and set the determinant of the matrix in (3.18) to zero. This gives the energy E of the NG boson as a function of its momentum \mathbf{p} ,

$$E(\mathbf{p}) = \frac{|\mathbf{p}|}{2M} \sqrt{\mathbf{p}^2 + 2m^2} . \quad (3.19)$$

Note that in the limit $m \rightarrow 0$, this recovers the conventional dispersion relation of a nonrelativistic particle, $E(\mathbf{p}) = \mathbf{p}^2/(2M)$.

Once the dispersion relation is no longer fixed by Lorentz invariance and the rest mass, it makes little sense to refer to a NG boson as a massless particle. Instead, it is common to say that the NG boson is a *gapless* mode or excitation. This is however a bit of a misnomer. The gap in the excitation spectrum is defined as $\min_{\mathbf{p}} E(\mathbf{p})$. On the other hand, the characteristic property of NG bosons is, as I will explain in detail in Chap. 6, that

$$\lim_{\mathbf{p} \rightarrow \mathbf{0}} E(\mathbf{p}) = 0 . \quad (3.20)$$

Since the excitation energy above a stable ground state is by definition positive, a NG boson is always gapless. However, a gapless excitation may not necessarily be a NG boson. There are physical systems where the dispersion relation of a one-particle excitation develops a local minimum due to dynamical effects. This is the case for instance for the “roton” mode in superfluid helium [8]. I am however not aware of any example where the energy of the excitation at the local minimum could be tuned to zero without breaking some symmetry.

One might hope that it is possible to get rid of the annoying mixing between the χ, π fields by eliminating the Higgs field χ as in Sect. 2.3.2. Let us see how the derivation of the low-energy EFT for π alone goes. We start by writing the bilinear and interaction parts of the Lagrangian (3.17) together as

$$\mathcal{L} \simeq -\frac{1}{2}(\nabla\chi)^2 - m^2\chi^2 - \lambda v\chi^3 - \frac{\lambda}{4}\chi^4 - \frac{1}{2}\left(1 + \frac{\chi}{v}\right)^2 \Xi , \quad (3.21)$$

up to terms that are a total time derivative. I have introduced a shorthand notation,

$$\Xi \equiv 2Mv\partial_0\pi + (\nabla\pi)^2 . \quad (3.22)$$

Note that the Lagrangian (3.21) is nearly identical to (2.34) except for the replacement $(\partial_\mu \chi)^2 \rightarrow -(\nabla \chi)^2$ and a different identification of Ξ ; in (2.34) it is simply $-(\partial_\mu \pi)^2$. We can therefore follow the same steps as in Sect. 2.3.2 to eliminate χ . All we have to do is to take the results thereof and substitute $(\partial_\mu \pi)^2 \rightarrow -\Xi$ and $\square \rightarrow -\nabla^2$ wherever appropriate. Thus, an iterative solution of the EoM for χ is obtained at once from (2.36),

$$\begin{aligned} \chi &= -\frac{1}{v} \left(2m^2 - \nabla^2 + \frac{\Xi}{v^2} + 3\lambda v \chi + \lambda \chi^2 \right)^{-1} \Xi \\ &= -\left(1 - \frac{\nabla^2}{2m^2} \right)^{-1} \frac{\Xi}{2m^2 v} + O(\Xi^2) . \end{aligned} \quad (3.23)$$

By using this in (2.40), we then readily determine the first two terms of the effective Lagrangian expanded in powers of Ξ ,

$$\mathcal{L}_{\text{eff}} \simeq -\frac{\Xi}{2} + \frac{\Xi}{4m^2 v^2} \left(1 - \frac{\nabla^2}{2m^2} \right)^{-1} \Xi + O(\Xi^3) . \quad (3.24)$$

While restricted to the lowest orders in Ξ and hence π , this Lagrangian still contains terms with an arbitrarily high number of derivatives. The price for having eliminated the kinetic mixing between χ and π therefore is an introduction of nonlocal operators. This is necessary to get the dispersion relation of the NG boson right. Indeed, by substituting from (3.22), we find

$$\mathcal{L}_{\text{eff}} \simeq 2M^2 \partial_0 \pi \frac{\partial_0}{2m^2 - \nabla^2} \pi - \frac{1}{2} (\nabla \pi)^2 + O(\pi^3) . \quad (3.25)$$

Fourier transformation to the energy–momentum space then recovers our previous result (3.19).

3.2.2 Multiple Nambu–Goldstone Fields

We would of course like to understand also what happens in models that contain multiple nonrelativistic scalar fields. Are there any further surprises awaiting us? Instead of developing a general picture akin to the relativistic framework of Sect. 3.1, I will focus here on some instructive examples. A more complete analysis will follow in Chap. 8 using the powerful machinery of EFT.

Let us first contemplate what we expect to find. Consider a model with multiple (real) scalar fields ϕ^i and a generic potential $V(\phi)$ as in (3.1). The argument of Sect. 3.1.1 then goes through without change. The conclusion that there is a well-defined injective mapping from broken symmetry generators to zero modes of the Hessian matrix of $V(\phi)$ in the ground state remains. One can say that the number of linearly independent NG *fields* still equals $\dim G - \dim H$, in one-to-one correspondence with the broken generators. However, with a generalized, nonrelativistic

kinetic term, we are no longer guaranteed the existence of an independent NG *mode* for each such NG field. In presence of terms in the Lagrangian with a single time derivative, some of the NG fields may be canonically conjugated to each other. We then expect one NG mode in the spectrum to be associated with a pair of NG fields.

Let me conclude this chapter and the whole Part I of the book with a couple of illustrative examples.

Example 3.4

Consider the limit $m = \lambda = 0$ of (3.13), already briefly mentioned previously,

$$\mathcal{L} = 2iM\phi^*\partial_0\phi - \nabla\phi^* \cdot \nabla\phi. \quad (3.26)$$

Here the potential is vanishing, hence the real and imaginary parts of ϕ are both trivially zero modes. There is a single one-particle excitation in the spectrum with dispersion relation $E(\mathbf{p}) = \mathbf{p}^2/(2M)$. This is a theory of a noninteracting Schrödinger field, which seems to be rather uninspiring. Things get more interesting, though, once we look at the theory from the point of view of symmetry.

There is a $U(1) \simeq SO(2)$ symmetry under which the field transforms as $\phi \rightarrow e^{i\epsilon}\phi$. The natural choice of ground state with $\langle\phi\rangle = 0$ leaves this symmetry unbroken. However, the free Schrödinger theory (3.26) has another symmetry, namely $\phi \rightarrow \phi + \epsilon_1 + i\epsilon_2$, where both $\epsilon_{1,2}$ are real parameters. This shifts the Lagrangian by a total time derivative and thus leaves the action invariant. Such a shift symmetry is necessarily spontaneously broken no matter how the VEV $\langle\phi\rangle$ is chosen. At the end of the day, we have two spontaneously broken generators but only one NG mode with a quadratic dispersion relation. In some aspects, the free Schrödinger theory (3.26) constitutes a rather nontrivial realization of a nonrelativistic NG boson. A detailed discussion of SSB in the quantized free Schrödinger theory is offered in Sect. 5.3.

For an example of an interacting field theory, replace the single complex field ϕ in (3.13) with a two-component complex field (doublet) Φ ,

$$\mathcal{L} = 2iM\Phi^\dagger\partial_0\Phi - \nabla\Phi^\dagger \cdot \nabla\Phi + m^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2. \quad (3.27)$$

Here the dagger indicates Hermitian conjugation. This Lagrangian possesses a $U(2)$ symmetry. It has an Abelian subgroup, $U(1)$, under which the doublet Φ changes its overall phase, $\Phi \rightarrow e^{i\epsilon}\Phi$. What is new is the non-Abelian $SU(2)$ subgroup of $U(2)$ under which $\Phi \rightarrow \mathcal{R}(g)\Phi$, where $g \in SU(2)$ and \mathcal{R} stands for the fundamental representation thereof.

The analysis of the model (3.27) proceeds as before. The classical Hamiltonian is minimized by any constant field Φ such that $\langle\Phi^\dagger\Phi\rangle = v^2/2 \equiv m^2/(2\lambda)$. Geometrically, the set of all degenerate ground states corresponds to a 3-sphere, S^3 . This is easy to see if one thinks of Φ as a four-component real vector whose length is fixed by minimization of the Hamiltonian. While any of the different degenerate ground

states is equally good, the conventional choice that simplifies notation is¹

$$\langle \Phi \rangle = \frac{v}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.28)$$

Out of all the generators of $U(2)$, the only one (up to normalization) that leaves this VEV unchanged is $\mathbb{1} + \tau_3$, where τ_3 is the third Pauli matrix. The remaining three linearly independent generators are spontaneously broken. We therefore expect three of the four degrees of freedom contained in Φ to represent NG fields.

The low-energy spectrum of the model (3.27) is most easily addressed within an EFT for the three NG fields. To that end, we need to choose a parameterization for Φ . Inspired by (3.8), let us set

$$\Phi(x) = \frac{1}{\sqrt{2}} U(\pi(x)) \begin{pmatrix} 0 \\ v + \chi(x) \end{pmatrix}, \quad (3.29)$$

where $U(\pi)$ is a unitary matrix encoding the three NG degrees of freedom and χ is a (real) Higgs field. A straightforward manipulation then leads to a Lagrangian of the type (3.21), except that we now have to set

$$\Xi = -2iMv^2(U^\dagger \partial_0 U)_{22} + v^2(\nabla U^\dagger \cdot \nabla U)_{22}. \quad (3.30)$$

The leading contributions to the effective Lagrangian are again given by (3.24). To complete the analysis, we parameterize $U(\pi)$ in terms of a triplet of NG fields, $\boldsymbol{\pi}(x)$,

$$U(\pi) = \exp\left(\frac{i}{v} \boldsymbol{\tau} \cdot \boldsymbol{\pi}\right). \quad (3.31)$$

Up to second order in the NG fields, which is needed to pin down the bilinear part of the effective Lagrangian, one finds

$$\Xi = -2Mv\partial_0\pi^3 - 2M(\boldsymbol{\pi} \times \partial_0\boldsymbol{\pi})^3 + \delta_{ab}\nabla\pi^a \cdot \nabla\pi^b + \mathcal{O}(\pi^3). \quad (3.32)$$

Upon using this in (3.24), one thus obtains

$$\begin{aligned} \mathcal{L} \simeq & 2M^2\partial_0\pi^3 \frac{\partial_0}{2m^2 - \nabla^2} \pi^3 - \frac{1}{2}(\nabla\pi^3)^2 \\ & + M(\pi^1\partial_0\pi^2 - \pi^2\partial_0\pi^1) - \frac{1}{2}(\nabla\pi^1)^2 - \frac{1}{2}(\nabla\pi^2)^2 + \mathcal{O}(\pi^3). \end{aligned} \quad (3.33)$$

The dispersion relation of π^3 is the same as for the model (3.13) with a single complex field, namely (3.19). The π^1, π^2 sector is however very different. The second line of (3.33) is just the free Schrödinger theory in disguise; π^1, π^2 are the real and imaginary parts of the complex Schrödinger field. The corresponding dispersion

¹ Any other choice of ground state can be obtained from our $\langle \Phi \rangle$ by multiplication with a constant $U(2)$ matrix. This generalizes the arbitrary phase $e^{i\theta}$ in (3.15).

relation is $E(\mathbf{p}) = \mathbf{p}^2/(2M)$. It is interesting to note that the bilinear Lagrangian in the π^1, π^2 sector comes entirely from the term in the effective Lagrangian (3.24) linear in Ξ . It is therefore insensitive to the choice of potential in the underlying Lagrangian (3.27).

The above discussion has an immediate generalization to a class of models described by the same Lagrangian (3.27), in which Φ is an n -component complex field. The symmetry is then $G \simeq \text{U}(n)$. The ground state can still be chosen as any constant field such that $\langle \Phi^\dagger \Phi \rangle = v^2/2$. It is convenient to pick $\langle \Phi \rangle$ so that, analogously to (3.28), only its n -th component is nonzero and equals $v/\sqrt{2}$. The symmetry group G is then spontaneously broken down to the subgroup $H \simeq \text{U}(n-1)$, acting on the first $n-1$ components of Φ . Accordingly, there are $\dim \text{U}(n) - \dim \text{U}(n-1) = 2n-1$ broken symmetry generators as well as NG fields. The analysis of the excitation spectrum also proceeds following the same steps, except for the necessary complications due to the presence of unitary $n \times n$ matrices. At the end of the day, one finds one NG boson with dispersion relation (3.19). The remaining $2n-2$ NG fields pair up into $n-1$ modes with the Schrödinger-like dispersion relation $E(\mathbf{p}) = \mathbf{p}^2/(2M)$.

3.3 Moral Lessons

In this chapter, I have refined the discussion of toy models for SSB in two ways. First, I have allowed for the possibility of multiple scalar fields and multiple broken symmetries. I have however remained within the territory of *internal* symmetries, that is symmetries only acting on the fields, independently of spacetime coordinates. Second, we explored the consequences of giving up Lorentz invariance, in particular by adding to the Lagrangian terms with a single time derivative. With the newly gathered experience, let me briefly revisit the four lessons drawn in Sect. 2.4.

Lesson #1 Nothing to change. The statement made in Sect. 2.4 is generally valid.

Lesson #2 Suppose that the given system possesses a continuous symmetry group G which is spontaneously broken to its subgroup H . If the action of the system is Lorentz-invariant, its spectrum contains $\dim G - \dim H$ massless particles: the NG bosons. If the system is not Lorentz-invariant, NG bosons are characterized as one-particle excitations with dispersion relation $E(\mathbf{p})$ such that $\lim_{\mathbf{p} \rightarrow 0} E(\mathbf{p}) = 0$. The low-energy dynamics of the system can still be captured in terms of $\dim G - \dim H$ NG fields. However, the actual number of NG modes in the spectrum may be lower than $\dim G - \dim H$. This happens when two NG fields are canonically conjugated by a term with a single time derivative, leading to a single NG mode. The dispersion relation of such a NG mode is typically quadratic in momentum.

Lesson #3 Nothing to change. The statement made in Sect. 2.4 is generally valid.

Lesson #4 It is convenient to choose a field parameterization in which a set of constant NG fields can be eliminated by a symmetry transformation. In such a

parameterization, every operator in the Lagrangian containing NG fields carries at least one derivative. If the symmetry group G is non-Abelian, it is however not necessary that every NG field carries a derivative. The nonlinear constraints imposed by the broken symmetry nevertheless ensure that the Adler zero property of scattering amplitudes is preserved, modulo the exceptions alluded to in Lesson #3.

Lesson #5 Nothing to change. The statement made in Sect. 2.4 is generally valid.

References

1. J. Goldstone, A. Salam, S. Weinberg, Phys. Rev. **127**, 965 (1962). DOI [10.1103/PhysRev.127.965](https://doi.org/10.1103/PhysRev.127.965)
2. S. Weinberg, Phys. Rev. Lett. **29**, 1698 (1972). DOI [10.1103/PhysRevLett.29.1698](https://doi.org/10.1103/PhysRevLett.29.1698)
3. S.R. Coleman, E.J. Weinberg, Phys. Rev. **D7**, 1888 (1973). DOI [10.1103/PhysRevD.7.1888](https://doi.org/10.1103/PhysRevD.7.1888)
4. H. Goldstein, C.P. Poole, J.L. Safko, *Classical Mechanics* (Pearson Education Limited, Harlow, UK, 2013)
5. T. Brauner, Phys. Rev. **D74**, 085010 (2006). DOI [10.1103/PhysRevD.74.085010](https://doi.org/10.1103/PhysRevD.74.085010)
6. S. Gongyo, Y. Kikuchi, T. Hyodo, T. Kunihiro, PTEP **2016**, 083B01 (2016). DOI [10.1093/ptep/ptw095](https://doi.org/10.1093/ptep/ptw095)
7. T. Brauner, M.F. Jakobsen, Phys. Rev. **D97**, 025021 (2018). DOI [10.1103/PhysRevD.97.025021](https://doi.org/10.1103/PhysRevD.97.025021)
8. A.J. Leggett, *Quantum Liquids* (Oxford University Press, Oxford, UK, 2006)

Part II

Foundations

Chapter 4

Symmetry and Conservation Laws

Anybody who has taken a first course on field theory has been exposed to the correspondence between symmetry and conservation laws via Noether's theorem. Indeed, I already assumed basic familiarity with this correspondence in Part I. The semiclassical analysis of the toy models therein would have hardly been possible without the concepts of symmetry of a Lagrangian and Noether current.

In this chapter, I will however restart the discussion of symmetries largely from scratch. With the rich spectrum of modern applications that emerged in the last decades, it appears appropriate to start by carefully defining the basic notions. I will not attempt to give the most general definition of symmetry. In particular, most of this chapter is phrased in the language of classical field theory, tailored to the needs of Parts III and IV of the book. Some remarks on a further generalization of the concept of symmetry developed here are postponed to the concluding Sect. 15.4. Nevertheless, within the conservative exposition offered here, I will devote more space to stressing exceptions rather than to repeating ad nauseam familiar concepts.

4.1 What Is Symmetry?

In very general terms, any definition of symmetry must include two ingredients. The first of these is an object that we wish to declare to be symmetric. The second is the operation, or transformation, that should constitute the desired symmetry. In classical physics, one usually defines symmetry by its action on a set of local fields. I will use the notation ψ^i for a set of fields treated as functions $\psi^i : x^\mu \rightarrow \psi^i(x)$. Here x^μ denotes collectively a set of coordinates, which may for the time being include space, time, or both. The case where x^μ includes only time t corresponds to mechanics with $\psi^i(t)$ as the dynamical variables. In mathematics, it is common to refer to x^μ and ψ^i respectively as the *independent* and *dependent variables*. These (in)dependent variables may take values from some linear space or from a more general mathematical structure such as a manifold. In the latter case, x^μ and ψ^i are identified with the corresponding local coordinates in the sense of Appendix A.1.

4.1.1 Symmetry Transformations

Let us initially focus on the second ingredient of symmetry, that is the operation. This book revolves largely around the consequences of continuous symmetries, for which it is sufficient to consider infinitesimal transformations. We shall deal with the following generic class of simultaneous transformations of the fields and coordinates,

$$\begin{aligned}\delta\psi^i(x) &\equiv \psi'^i(x') - \psi^i(x) = \epsilon F^i[\psi, x](x) , \\ \delta x^\mu &\equiv x'^\mu - x^\mu = \epsilon X^\mu[\psi, x](x) .\end{aligned}\quad (4.1)$$

Here ϵ is an infinitesimal parameter of the transformation. The square bracket notation indicates that F^i and X^μ are local functions of the fields and (a finite number of) their derivatives, possibly depending explicitly on the coordinates. Transformations of the type (4.1) are known in mathematical literature as *generalized local transformations*. The terminology is historical. Namely, the class (4.1) generalizes so-called *point transformations*, first studied by Sophus Lie in the 1860s. For those, F^i and X^μ are only allowed to depend on fields and coordinates, not on field derivatives.

Indicating that the fields and coordinates should be transformed simultaneously is, while conventional, a red herring. The coordinates are independent variables that can be chosen at will. Any nonsingular transformation of coordinates can be undone by a change of variables. The content of (4.1) can therefore be equivalently encoded in a transformation of the fields alone,

$$\psi'^i(x) - \psi^i(x) = \epsilon F^i[\psi, x](x) - \epsilon X^\mu[\psi, x](x) \partial_\mu \psi^i(x) . \quad (4.2)$$

This is the *evolutionary form* of the transformation (4.1). More generally, the representation (4.1) of the transformation is ambiguous with respect to the redefinition $X^\mu \rightarrow X^\mu + \tilde{X}^\mu$ and $F^i \rightarrow F^i + \tilde{X}^\mu \partial_\mu \psi^i$, where $\tilde{X}^\mu[\psi, x]$ is any local function of the coordinates, fields and their derivatives. The evolutionary form can thus be viewed as fixing the ambiguity by setting $X^\mu = 0$.

The class of transformations (4.1) is so broad that it is often convenient to consider special cases. Throughout this book, I will use the term *internal symmetry* for a point transformation with $X^\mu = 0$ and a coordinate-independent F^i , that is

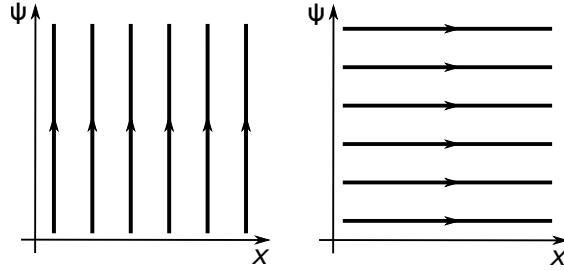
$$\delta\psi^i(x) = \epsilon F^i(\psi(x)) , \quad \delta x^\mu = 0 \quad (\text{internal symmetry}) . \quad (4.3)$$

A point transformation for which X^μ is nonzero but only depends on the spacetime coordinates will be referred to as a *spacetime symmetry*,

$$\delta\psi^i(x) = \epsilon F^i(\psi(x), x) , \quad \delta x^\mu = \epsilon X^\mu(x) \neq 0 \quad (\text{spacetime symmetry}) . \quad (4.4)$$

These definitions of internal and spacetime symmetries define the agenda for Parts III and IV of the book. The reader should however be warned that these are not standard

Fig. 4.1 Schematic illustration of internal (left panel) and spacetime (right panel) symmetries. The oriented curves indicate the flow in the space of fields and coordinates (see Appendix A.3), generated respectively by the infinitesimal transformations (4.3) and (4.4). To stress the difference between the two cases, the right panel shows a “purely spacetime” transformation, for which $F^i = 0$



definitions aligned with published literature. Of all the reasons, this is because a precise definition of internal and spacetime symmetries is rarely found.

Example 4.1

A spacetime translation in the direction of coordinate x^ν is most naturally implemented by setting $\delta_\nu \psi^i(x) = 0$ and $\delta_\nu x^\mu = \epsilon \delta_\nu^\mu$. According to (4.2), this can be equivalently encoded as $\delta_\nu \psi^i(x) = -\epsilon \partial_\nu \psi^i(x)$ without any change of coordinates. Translations are an example of a “purely spacetime” transformation, illustrated in the right panel of Fig. 4.1. As the figure shows, one can intuitively think of internal and spacetime symmetries as point transformations, generating flows along different subspaces of the space of fields and coordinates.

4.1.2 Object of Symmetry

I will assume throughout the book that the set of all symmetry transformations, generated by the infinitesimal motions (4.1), constitutes a group. Historically, the mathematical field of Lie groups arose from the study of symmetries of *differential equations*. This is now a mature subject with a large body of literature devoted to it. An interested reader is referred to [1–3] for more details. Very briefly, the major applications of symmetry methods to differential equations include:

- Finding new solutions of a given differential equation from already known ones.
- Finding solutions respecting the symmetry of the differential equation.
- Classifying differential equations with given symmetry.

In physics, it is much more common to define symmetry by invariance of the *action* of a system under some transformation. It is worth stressing that the two notions of symmetry are not equivalent. On the one hand, there are differential equations that

do not originate from any variational principle, yet may have nontrivial symmetries. On the other hand, differential equations that do descend from an action functional may have a larger symmetry group than the action itself.

The chief advantage of defining symmetry through an action functional is that this gives us a direct link to conservation laws. The link is supplied by the celebrated Noether theorem, which is the subject of most of Sect. 4.2. It is in principle also possible to define conservation laws directly on the level of a differential equation. This allows one to deduce a restricted correspondence between symmetries and conservation laws that generalizes the Noether theorem [3]. However, as far as I know, such a generalized notion of symmetry is of limited use in physics, and I will therefore not pursue it any further.

4.2 Lagrangian Approach to Symmetry

The correspondence between continuous symmetries and conservation laws has a fascinating history. A nice overview of the developments following Noether's groundbreaking discovery, including an account of the contributions of various authors to the subject, can be found in [4]. Noether's theorem is covered to some extent in virtually any textbook on classical mechanics or (classical or quantum) field theory. Unfortunately, this is frequently done under unnecessary restrictions on the Lagrangian or on the type of symmetry transformation. Here I will present a fairly general version of Noether's theorem following a trick, which to the best of my knowledge goes back to Gell-Mann and Lévy [5].

4.2.1 Noether's Theorem

I consider a class of field theories defined by a Lagrangian density \mathcal{L} , which is a local function of a set of fields ψ^i and their derivatives,

$$S = \int d^D x \mathcal{L}[\psi, x](x) . \quad (4.5)$$

Here D is the dimension of the space of independent variables, which may still include space, time or both. The Lagrangian density may depend explicitly on the coordinates, and there is no restriction on the order of field derivatives it contains. Suppose that the action S is invariant under the infinitesimal transformation (4.1).¹ Then there is a local vector field $J^\mu[\psi, x]$, called *Noether current*, which is divergence-less,

$$\partial_\mu J^\mu[\psi, x] = 0 \quad (\text{on-shell}) . \quad (4.6)$$

¹ When evaluating the variation of the action under (4.1), it is allowed to drop whatever boundary terms might appear. Thus, the “invariance of the action” should be more accurately interpreted as invariance of the Lagrangian density up to the divergence of a vector field.

The qualifier *on-shell* reminds us that the conservation law (4.6) only holds for fields satisfying the Euler–Lagrange *equation of motion* (EoM).

To prove this claim, we evaluate the variation of the action δS under a modification of (4.1) where the parameter $\epsilon(x)$ is allowed to depend on the coordinates. With the assumption of invariance under the original transformation (4.1), δS may only depend on derivatives of $\epsilon(x)$. Using integration by parts, it can then always be brought to the form

$$\delta S = \int d^D x J^\mu[\psi, x](x) \partial_\mu \epsilon(x) \quad (\text{off-shell}) . \quad (4.7)$$

Here the qualifier *off-shell* indicates that I have not yet imposed the EoM on the fields. We can see that the Noether current can be extracted as the coefficient of $\partial_\mu \epsilon$ in the variation of the action. Once the EoM is applied, δS must vanish by the definition of the variational principle. This guarantees the on-shell conservation law (4.6).

Example 4.2

For a simple example, consider a free massless relativistic scalar field ϕ with the Lagrangian density $\mathcal{L}[\phi] = (1/2)(\partial_\mu \phi)^2$. This is obviously invariant under a shift of the field, $\phi \rightarrow \phi + \epsilon$. Making the shift coordinate-dependent, we find the variation of the action $\delta S = \int d^D x \partial^\mu \phi(x) \partial_\mu \epsilon(x)$. Comparison with (4.7) tells us that the Noether current is $J^\mu[\phi] = \partial^\mu \phi$. Conservation of this current is equivalent to the EoM for ϕ , which is the massless Klein–Gordon equation. As an aside, such an equivalence holds for any theory of a real scalar ϕ , invariant under the shift $\phi \rightarrow \phi + \epsilon$. This follows immediately from the variation of the action under the localized transformation $\phi(x) \rightarrow \phi(x) + \epsilon(x)$,

$$\delta S = \int d^D x \frac{\delta S}{\delta \phi(x)} \epsilon(x) = - \int d^D x \partial_\mu J^\mu[\phi, x](x) \epsilon(x) , \quad (4.8)$$

which implies that $\partial_\mu J^\mu[\phi, x](x) = -\delta S/\delta \phi(x)$ off-shell.

The statement of Noether’s theorem is well-known. Instead of collecting numerous examples, I will thus focus on a few comments and illustrations that go somewhat off the beaten track. To start with, one should keep in mind that the Noether current is not uniquely determined by the assumed symmetry. The variation (4.7) only allows us to extract $J^\mu[\psi, x]$ up to addition of a vector field whose divergence vanishes off-shell. Such a modification of the current does not affect the conservation law (4.6).

There is another, somewhat more subtle ambiguity in the Noether current, related to the definition of the transformation used to produce (4.7). Namely, it would be tempting to simply take (4.1) and make it local by replacing $\epsilon \rightarrow \epsilon(x)$. But there is a more general possibility that the localized transformation also depends on the derivatives of $\epsilon(x)$. All that is required is that for constant ϵ , the transformation reduces to (4.1). To see how this ambiguity affects the Noether current, consider a

local transformation of the form

$$\delta\psi^i(x) = \epsilon(x)F^i[\psi, x](x) + \partial_\mu\epsilon(x)K^{i\mu}[\psi, x](x) , \quad (4.9)$$

where $K^{i\mu}[\psi, x]$ is an arbitrary local function of the fields and their derivatives. To keep things simple, I have used the evolutionary form of the transformation where the coordinates x^μ do not change. Applying (4.7), one finds that the new term in the transformation rule for ψ^i shifts the Noether current by a term linear in $K^{i\mu}[\psi, x]$,

$$J^\mu[\psi, x](x) = J^\mu[\psi, x](x) \Big|_{K=0} + K^{i\mu}[\psi, x](x) \frac{\delta S}{\delta\psi^i(x)} . \quad (4.10)$$

The argument is easily generalized to transformations depending on arbitrarily high derivatives of $\epsilon(x)$. All the ensuing corrections to the current are proportional to $\delta S/\delta\psi^i$, which defines the EoM for the fields. The moral is that the ambiguity in the definition of the localized symmetry transformation leads to new contributions to the Noether current that vanish on-shell [6]. Neither this ambiguity does therefore affect the conservation law (4.6).

Example 4.3

By the naive replacement $\epsilon \rightarrow \epsilon(x)$, we can localize the action of spacetime translation in the x^ν -direction as $\delta_\nu\psi^i(x) = -\epsilon(x)\partial_\nu\psi^i(x)$. Assuming that the Lagrangian density does not depend explicitly on the coordinates and on higher than first derivatives of ψ^i , (4.7) gives us the set of currents

$$T^\mu_\nu = \delta^\mu_\nu \mathcal{L} - \frac{\partial \mathcal{L}}{\partial(\partial_\mu\psi^i)} \partial_\nu\psi^i . \quad (4.11)$$

This is the familiar *canonical energy–momentum (EM) tensor*. (The unusual overall sign is a consequence of the conventions used here.) This EM tensor is known to possess some undesired features that have inspired various “improvements” with a history as long as Noether’s theorem itself. For an illustration, consider a relativistic theory of a scalar field ϕ and a vector field A_μ , defined by

$$\mathcal{L}[\phi, A] = A^\mu \partial_\mu \phi - \frac{1}{2} A^\mu A_\mu . \quad (4.12)$$

The canonical EM tensor of this theory is $T^{\mu\nu} = g^{\mu\nu} \mathcal{L} - A^\mu \partial^\nu \phi$; I have raised the second index with the flat Minkowski metric $g^{\mu\nu}$. This EM tensor is notably not symmetric, which is a general trait shared by theories that contain nonscalar fields. This is quite troubling, if only for the fact that (4.12) is actually the free massless scalar theory in disguise. Indeed, using the EoM for the vector field, $A_\mu = \partial_\mu \phi$, turns the Lagrangian into $\mathcal{L}[\phi] = (1/2)(\partial_\mu \phi)^2$.

One way to solve this problem is to note that the naive local translation, $\delta_\nu A_\mu(x) = -\epsilon(x)\partial_\nu A_\mu(x)$, is not compatible with the EoM for A_μ . Under a local translation, A_μ should transform as a covariant vector field. Let us therefore try

$$\delta_\nu \phi(x) = -\epsilon(x) \partial_\nu \phi(x) , \quad \delta_\nu A_\mu(x) = -\epsilon(x) \partial_\nu A_\mu(x) - A_\nu(x) \partial_\mu \epsilon(x) . \quad (4.13)$$

This is a generalized local transformation of the type (4.9), which leads to the correspondingly modified EM tensor, $\tilde{T}^{\mu\nu} = g^{\mu\nu} \mathcal{L} + A^\mu A^\nu - (A^\mu \partial^\nu \phi + A^\nu \partial^\mu \phi)$. This is symmetric off-shell, which is ultimately because the transformation (4.13) has a well-defined geometric meaning [6]. In the language of Appendix A.3, (4.13) represents the Lie derivative of ϕ and A_μ along the vector field $-\epsilon(x) \partial_\nu$.

The above example shows that the ambiguity in the definition of the local transformation used to produce the Noether current is not just a nuisance. It may be used as a tool to construct “improved” Noether currents with desired properties. Quite recently, this idea was exploited to systematically construct EM tensors tailored to Lorentz, scale and conformal symmetry [7]. For further background and references on Noether’s theorem and improvement of Noether currents, see for instance [8].

4.2.2 Tensor Conservation Laws

Before closing the discussion of Noether’s theorem, let me stress that (4.6), while generic, is not the only form the ensuing conservation law may take. It may happen that the variation of the action (4.7), or its parts, only depends on higher than first derivatives of $\epsilon(x)$. This leads to conservation laws with likewise higher than first derivatives of a generalized tensor current.

Example 4.4

Consider the theory of a free real *Lifshitz scalar* field ϕ [9],

$$\mathcal{L}[\phi] = \frac{1}{2} (\partial_0 \phi)^2 - \frac{1}{2} (\partial_r \partial_s \phi)^2 . \quad (4.14)$$

This Lagrangian is, just like that of the massless relativistic scalar theory, invariant under the shift $\phi \rightarrow \phi + \epsilon$. Making the replacement $\epsilon \rightarrow \epsilon(x)$ leads to the variation

$$\delta S = \int d^D x \{ J^0[\phi](x) \partial_0 \epsilon(x) - J^{rs}[\phi](x) \partial_r \partial_s \epsilon(x) \} , \quad (4.15)$$

where $J^0[\phi] \equiv \partial_0 \phi$ and $J^{rs}[\phi] \equiv \partial_r \partial_s \phi$. The corresponding on-shell conservation law takes the form

$$\partial_0 J^0[\phi] + \partial_r \partial_s J^{rs}[\phi] = 0 . \quad (4.16)$$

This is, unsurprisingly, equivalent to the EoM for ϕ .

From now on I will always assume that the coordinates x^μ include both space and time. We can then integrate the local conservation law (4.6) over *space*. Assuming asymptotic behavior of the fields at infinity such that the boundary term produced by integrating $\partial_r J^r[\psi, x]$ vanishes, we conclude that the integral charge

$$Q \equiv \int d^d \mathbf{x} J^0[\psi, x](x) \quad (4.17)$$

is time-independent. Curiously, a local conservation law of the type (4.16) is stronger. Namely, apart from conservation of Q , it also implies time-independence of

$$Q^r \equiv \int d^d \mathbf{x} x^r J^0[\psi, x](x) . \quad (4.18)$$

This looks like the dipole moment of the charge distribution defined by the density $J^0[\psi, x]$. Conservation laws of the dipole type (4.16) have recently attracted considerable attention due to their relevance for so-called *fracton* phases of matter. The reader is referred to [10, 11] for more details about this intriguing subject.

4.3 Symmetry and Conservation Laws in Hamiltonian Formalism

The Lagrangian formalism provides a natural framework for the discussion of symmetries and conservation laws, yet it also offers numerous other benefits. For applications to *effective field theory* (EFT), it is particularly important that the formalism works without change for Lagrangians depending on higher field derivatives. Also, it makes the transition from classical to quantum theory straightforward within the path-integral approach to quantization. However, we will see in Chap. 5 that the most striking manifestations of *spontaneous symmetry breaking* (SSB) include peculiar properties of the quantum ground state and the spectrum of excitations above it. Such features are not easily addressed using path integrals; here it is more natural to use the operator approach to quantization.

With this in mind, I will now devote some space to the Hamiltonian formalism, which is the classical counterpart of the operator language of quantum field theory. This comes at the cost of having to restrict the discussion to theories whose action functional depends only on the first time derivatives of fields. In return, we gain the reverse of Noether's theorem, allowing us to extract the corresponding symmetry from a given conservation law. What follows is a brief survey of the *symplectic* formulation of the Hamiltonian formalism, adapted to local field theory. The background developed here will prove useful in Chap. 8. For a further generalization of the approach outlined below, see for instance [12]. A reader interested in modern developments of this approach is advised to consult [13].

4.3.1 Symplectic Formulation of Hamiltonian Dynamics

The starting point is a manifold called the *target space* of the given theory. This manifold carries a geometric structure fixed by a (locally defined) 1-form ω called the *symplectic potential*. I will use the notation $\omega \equiv \omega_i(\xi) d\xi^i$ where ξ^i is a set of (local) coordinates on the target space. The *phase space* of the theory consists of all time-independent fields taking values in the target space. With some abuse of notation, I will denote such fields as $\xi^i(\mathbf{x})$. The Hamiltonian of the theory is a local functional on the phase space, $H = \int d^d\mathbf{x} \mathcal{H}[\xi, \mathbf{x}](\mathbf{x})$. As indicated by the square bracket notation, the Hamiltonian density \mathcal{H} is a local function of the fields and their spatial derivatives, possibly also depending explicitly on spatial coordinates.

The action is now a functional of trajectories on the phase space, which I with some further abuse of notation denote as $\xi^i(\mathbf{x}, t) \equiv \xi^i(\mathbf{x})$. It is fixed by the choice of symplectic potential and Hamiltonian,

$$S = \int d^Dx \left\{ \omega_i(\xi(x)) \dot{\xi}^i(x) - \mathcal{H}[\xi, \mathbf{x}](x) \right\}. \quad (4.19)$$

The EoM for the variational principle based on (4.19) takes the form

$$\Omega_{ij}(\xi(x)) \dot{\xi}^j(x) = \frac{\delta H}{\delta \xi^i(x)}, \quad \Omega_{ij}(\xi) \equiv \frac{\partial \omega_j(\xi)}{\partial \xi^i} - \frac{\partial \omega_i(\xi)}{\partial \xi^j}. \quad (4.20)$$

Here $\delta/\delta \xi^i(x)$ indicates taking the variational derivative of a functional on the phase space with respect to $\xi^i(\mathbf{x})$, and substituting the trajectory $\xi^i(x)$ in the result. The antisymmetric matrix $\Omega_{ij}(\xi)$ collects the components of the *symplectic 2-form*, $\Omega \equiv (1/2)\Omega_{ij}(\xi) d\xi^i \wedge d\xi^j$. This is an object of central importance for the Hamiltonian approach to mechanics and field theory. In the language of Appendix A.8, it is a closed 2-form, for it is related to the symplectic potential via $\Omega = d\omega$. Moreover, the matrix $\Omega_{ij}(\xi)$ is assumed to be nonsingular so that (4.20) constitutes a complete set of evolution equations for the fields ξ^i . One can then rewrite (4.20) as

$$\dot{\xi}^i(x) = \Omega^{ij}(\xi(x)) \frac{\delta H}{\delta \xi^j(x)}, \quad (4.21)$$

where $\Omega^{ij}(\xi)$ is the matrix inverse of $\Omega_{ij}(\xi)$.

Example 4.5

Consider a Lagrangian field theory of n scalar fields ϕ^i , taking values in \mathbb{R}^n . The target space of the Hamiltonian description of this theory is $\mathbb{R}^n \times \mathbb{R}^n$, spanned on the fields ϕ^i and their conjugate momenta π_i . The Lagrangian and Hamiltonian densities are related by the Legendre transform, $\mathcal{L} = \pi_i \dot{\phi}^i - \mathcal{H}$. Matching this to (4.19) allows one to identify the symplectic potential, $\omega(\phi, \pi) = \pi_i d\phi^i$. The symplectic 2-form in turn becomes $\Omega(\phi, \pi) = d\pi_i \wedge d\phi^i$, and the EoM (4.20) reduces to the familiar form

$$\phi^i(x) = \frac{\delta H}{\delta \pi_i(x)} , \quad \pi_i(x) = -\frac{\delta H}{\delta \phi^i(x)} . \quad (4.22)$$

In general, coordinates ϕ^i, π_i on the target space in which the symplectic 2-form acquires the simple form $\Omega(\phi, \pi) = d\pi_i \wedge d\phi^i$ are called *Darboux coordinates*. By the *Darboux theorem*, such coordinates exist at least locally on any manifold endowed with a symplectic 2-form (see for instance Sect. 43 of [14]). The global existence of Darboux coordinates on the target space is however not guaranteed. In fact, it is ruled out whenever the symplectic 2-form is closed but not exact. Example 4.7 below provides a nontrivial illustration of this possibility. The geometry of manifolds carrying a symplectic structure is therefore locally identical to that of $\mathbb{R}^n \times \mathbb{R}^n$, but may be globally nontrivial.

With the symplectic 2-form at hand, one can define the *Poisson bracket* of any two local functionals F, G on the phase space,

$$\{F, G\} \equiv \int d^d \mathbf{x} \, \Omega^{ij}(\xi(\mathbf{x})) \frac{\delta F}{\delta \xi^i(\mathbf{x})} \frac{\delta G}{\delta \xi^j(\mathbf{x})} . \quad (4.23)$$

A special case is the fundamental Poisson bracket of the phase space coordinates,

$$\{\xi^i(\mathbf{x}), \xi^j(\mathbf{y})\} = \Omega^{ij}(\xi(\mathbf{x})) \delta^d(\mathbf{x} - \mathbf{y}) . \quad (4.24)$$

In local Darboux coordinates, (4.23) reduces to the textbook definition of Poisson bracket in terms of derivatives with respect to canonical coordinates $\psi^i(\mathbf{x})$ and momenta $\pi_i(\mathbf{x})$. Likewise, (4.24) generalizes the fundamental brackets $\{\psi^i(\mathbf{x}), \psi^j(\mathbf{y})\} = \{\pi_i(\mathbf{x}), \pi_j(\mathbf{y})\} = 0$ and $\{\psi^i(\mathbf{x}), \pi_j(\mathbf{y})\} = \delta_j^i \delta^d(\mathbf{x} - \mathbf{y})$ in a way independent of the choice of coordinates on the target space. It is easy to check that (4.23) has the following properties. First, $\{F, G\}$ is obviously linear in both arguments and anti-symmetric. Moreover, it satisfies the Leibniz (product) rule. Finally, it satisfies the Jacobi identity; this takes more effort to prove and requires using the closedness of the symplectic 2-form.

The Poisson bracket offers a very compact expression for time evolution. For any local functional F on the phase space that does not depend explicitly on time,

$$\dot{F} = \int d^d \mathbf{x} \, \frac{\delta F}{\delta \xi^i(x)} \dot{\xi}^i(x) = \{F, H\} , \quad (4.25)$$

where I used (4.21). The latter is itself equivalent to $\dot{\xi}^i = \{\xi^i, H\}$. In this form, it is natural to think of time evolution as a flow on the phase space, generated by H .

This completes the preparation required to address the main objective of this survey of the Hamiltonian formalism: the connection of symmetries and conservation laws. First, everything said in Sect. 4.2 about Noether's theorem remains valid in the present Hamiltonian setting. For the sake of identifying the Noether current, there is no difference between the action functionals (4.5) and (4.19). We can therefore

conclude at once that a continuous symmetry of the action (4.19) implies the existence of a current $J^\mu[\xi, x]$ conserved on-shell. This current can be identified by localizing the symmetry transformation and using (4.7).

Example 4.6

The Hamiltonian density of a free massless relativistic scalar field ϕ is $\mathcal{H}[\phi, \pi] = (1/2)\pi^2 + (1/2)(\nabla\phi)^2$. The symplectic potential takes the Darboux form, $\omega(\phi, \pi) = \pi d\phi$. The action functional

$$S = \int d^Dx \left\{ \pi(x)\dot{\phi}(x) - \frac{1}{2}\pi(x)^2 - \frac{1}{2}[\nabla\phi(x)]^2 \right\} \quad (4.26)$$

is invariant under the constant shift $\phi \rightarrow \phi + \epsilon$ with π kept unchanged. Making the shift coordinate-dependent, the action varies by $\delta S = \int d^Dx [\pi(x)\partial_0\epsilon(x) - \nabla\phi(x) \cdot \nabla\epsilon(x)]$. This leads to the identification of the temporal and spatial components of the Noether current as $J^0[\phi, \pi] = \pi$, $J^r[\phi, \pi] = -\partial_r\phi = +\partial^r\phi$. Upon using the EoM for ϕ , $\dot{\phi} = \pi$, the current is seen to coincide with that derived in Example 4.2, $J^\mu[\phi] = \partial^\mu\phi$.

What is new in the Hamiltonian framework is that we can now reverse Noether's theorem and reconstruct the symmetry from a given conservation law. By (4.25), a local functional F that does not depend explicitly on time defines an integral of motion (conserved charge) if and only if $\{F, H\} = 0$. Moreover, if F and G are both conserved, then so is $\{F, G\}$ thanks to the Jacobi identity. Thus, all conserved charges of the theory furnish a Lie algebra with respect to the Poisson bracket. Finally, every conserved charge generates a flow on the phase space, defined by the infinitesimal transformation

$$\delta\xi^i(x) = \epsilon\{\xi^i(x), F\} = \epsilon\Omega^{ij}(\xi(x))\frac{\delta F}{\delta\xi^j(x)}. \quad (4.27)$$

We thus have an explicit realization of a geometric transformation, associated with a given conserved charge. It remains to demonstrate that this is in fact a symmetry of the action (4.19). To that end, a short calculation using the definition of the symplectic 2-form shows that under the transformation (4.27), the action varies by

$$\delta S = -\epsilon \int d^Dx \frac{\delta F}{\delta\xi^i(x)}\dot{\xi}^i(x) - \epsilon \int dt \{H, F\}(t). \quad (4.28)$$

The first term vanishes upon integration, being a total time derivative. The second term vanishes by the assumption that F is conserved. This completes the proof that every conserved charge (that does not depend explicitly on time) generates a symmetry of the action.

Example 4.7

The low-energy physics of ferromagnets can be encoded in a continuous field theory whose basic degree of freedom is the local spin (magnetization) density. In the first approximation, fluctuations of the ferromagnetic equilibrium correspond to local changes in spin orientation, while the magnitude M of spin density remains constant. The target space is thus equivalent to the sphere S^2 . The local spin variable $n^i(\mathbf{x})$ is a unit three-vector field that satisfies the fundamental Poisson bracket

$$\{n^i(\mathbf{x}), n^j(\mathbf{y})\} = \frac{1}{M} \varepsilon^{ij}_k n^k(\mathbf{x}) \delta^d(\mathbf{x} - \mathbf{y}) , \quad (4.29)$$

reflecting the Lie algebra of angular momentum. This is a prominent example of a system with a phase space whose symplectic structure is largely fixed by symmetry alone. Other examples can be found in [15].

For any Hamiltonian that is a local functional of n^i , the ensuing EoM is $\dot{n}^i = \{n^i, H\}$. Using the chain rule, this can be put into a neat vector form²

$$\dot{\mathbf{n}}(x) = \frac{1}{M} \frac{\delta H}{\delta \mathbf{n}(x)} \times \mathbf{n}(x) , \quad (4.30)$$

known as the *Landau–Lifshitz equation*. The concrete form of the Hamiltonian depends on the physical system. For isotropic ferromagnets, the simplest choice is quadratic in gradients of $\mathbf{n}(x)$, as suggested by the Landau theory of phase transitions,

$$H = \int d^d \mathbf{x} \frac{\varrho_s}{2} \delta_{ij} \nabla n^i(x) \cdot \nabla n^j(x) . \quad (4.31)$$

Here ϱ_s is the so-called spin stiffness. The EoM then reduces to

$$\dot{\mathbf{n}}(x) = \frac{\varrho_s}{M} \mathbf{n}(x) \times \nabla^2 \mathbf{n}(x) . \quad (4.32)$$

Matching (4.30) to the general EoM (4.20) allows one to extract the symplectic 2-form on S^2 , corresponding to the Poisson bracket (4.29),

$$\Omega = -\frac{M}{2} \varepsilon_{ijk} n^i dn^j \wedge dn^k . \quad (4.33)$$

This is, up to normalization, the area 2-form on S^2 , see Example A.7 in Appendix A.3. The area 2-form is closed but not exact. Hence the sphere S^2 is an example of a symplectic manifold on which globally well-defined Darboux coordinates do not exist. This fact has deep consequences for the topological properties of ferromagnets. As to symmetry, the symplectic 2-form (4.33) and the Hamiltonian (4.31) are both manifestly invariant under $SO(3)$ rotations of the spin variable n^i . This symmetry corresponds to the conservation of total spin, $\mathbf{S} = M \int d^d \mathbf{x} \mathbf{n}(x)$.

² This is a slight abuse of notation. I am using boldface italics both for spatial vectors living in \mathbb{R}^d with arbitrary d and for the spin vector living on $S^2 \subset \mathbb{R}^3$.

Ferromagnets are fascinating materials that exhibit many of the nontrivial features of SSB. This example is just a taster; in Sect. 9.2, I will serve the reader a much more thorough discussion of the low-energy physics of spin systems.

4.3.2 Symmetry in Quantum Physics

Our discussion of symmetries in field theory has been strictly classical so far. The implementation of symmetries in quantum field theory comes with numerous subtleties. Some of these are related to SSB and I will return to them in the next chapter. Here I will therefore just briefly outline the transition from classical to quantum physics. As is well known, this transition is streamlined in the Hamiltonian formalism. Local functionals on the phase space F, G, \dots are replaced with operators \hat{F}, \hat{G}, \dots on the Hilbert space of physical states. The commutator of these operators is then obtained from the Poisson bracket of their classical counterparts, roughly speaking, by the replacement $\{F, G\} \rightarrow -i[\hat{F}, \hat{G}]$. In the following, I will drop the hat on operators, but otherwise closely follow the Hamiltonian representation of symmetries outlined above. For the sake of simplicity, I will only consider conserved charges that do not explicitly depend on time.

With this qualification, a Hermitian operator Q represents a conserved charge of a quantum system if and only if it commutes with the Hamiltonian, $[Q, H] = 0$. By analogy with (4.27), the conserved charge generates a flow on the algebra of observables. For a given Hermitian operator A , the shift induced by Q is

$$\delta A = -i\epsilon[A, Q] . \quad (4.34)$$

This can be extended to a transformation with a finite parameter ϵ . One can think of such a transformation as a map $A \rightarrow A(\epsilon)$ where $A(\epsilon)$ satisfies the “flow equation”

$$\frac{dA(\epsilon)}{d\epsilon} = -i[A(\epsilon), Q] = i[Q, A(\epsilon)] . \quad (4.35)$$

This has the formal solution

$$A(\epsilon) = e^{i\epsilon Q} A(0) e^{-i\epsilon Q} . \quad (4.36)$$

The duality between the Schrödinger and Heisenberg pictures of quantum mechanics suggests an interpretation of (4.36) in terms of a formal unitary operator $U(\epsilon) \equiv e^{-i\epsilon Q}$, acting on the Hilbert space of states. Instead of transforming operators via (4.36), we could then equivalently transform physical states by $U(\epsilon)$. In either case, the fact that Q is a conserved charge is reflected by the invariance of the Hamiltonian under the transformation generated by Q , $H(\epsilon) = H$ or $[U(\epsilon), H] = 0$.

As I will demonstrate in Sect. 5.3, when the symmetry in question is spontaneously broken, the operator $U(\epsilon)$ may in fact not exist. This is one of the quirks of SSB. The symmetry may not be realized by unitary operators on the Hilbert space. As we will see, this feature is related to the nontrivial structure of the Hilbert space in presence of spontaneously broken symmetries. Even then, it makes sense to ask how the symmetry affects results of measurements. The transformation of physical observables under the symmetry remains well-defined and is still expressed by the flow equation (4.35).

The starting point of the discussion in this chapter was that a definition of symmetry requires a transformation and an object. All the objects we have worked with so far—actions, Hamiltonians, and EoM—capture the dynamics of the entire physical system. It is however no less interesting and useful to study the symmetries of a particular *state* of the system, whether classical or quantum. This brings us to the realm of SSB, which will be addressed in detail in the next chapter.

References

1. P.J. Olver, *Applications of Lie Groups to Differential Equations* (Springer-Verlag, New York, 1986)
2. G.W. Bluman, S.C. Anco, *Symmetry and Integration Methods for Differential Equations* (Springer-Verlag, New York, 2002)
3. G.W. Bluman, A.F. Cheviakov, S.C. Anco, *Applications of Symmetry Methods to Partial Differential Equations* (Springer-Verlag, New York, 2010)
4. Y. Kosmann-Schwarzbach, *The Noether Theorems: Invariance and Conservation Laws in the Twentieth Century* (Springer-Verlag, New York, 2011)
5. M. Gell-Mann, M. Lévy, *Nuovo Cim.* **16**, 705 (1960). DOI 10.1007/BF02859738
6. T. Brauner, *Phys. Scr.* **95**, 035004 (2020). DOI 10.1088/1402-4896/ab50a5
7. I. Kourkoulou, A. Nicolis, G. Sun, *Phys. Rev.* **D106**, 125005 (2022). DOI 10.1103/PhysRevD.106.125005
8. F. Gieres, *Fortschr. Phys.* **70**, 2200078 (2022). DOI 10.1002/prop.202200078
9. T. Griffin, K.T. Grosvenor, P. Hořava, Z. Yan, *Phys. Rev.* **D88**, 101701 (2013). DOI 10.1103/PhysRevD.88.101701
10. N. Seiberg, *SciPost Phys.* **8**, 050 (2020). DOI 10.21468/SciPostPhys.8.4.050
11. K.T. Grosvenor, C. Hoyos, F. Peña-Benítez, P. Surówka, *Front. Phys.* **9**, 792621 (2022). DOI 10.3389/fphy.2021.792621
12. J.V. José, E.J. Saletan, *Classical Dynamics: A Contemporary Approach* (Cambridge University Press, Cambridge, UK, 1998)
13. F. Gieres, *SciPost Phys. Lect. Notes* **77** (2023). DOI 10.21468/SciPostPhysLectNotes.77
14. V.I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer-Verlag, New York, 1989)
15. I.E. Dzyaloshinskii, G.E. Volovick, *Ann. Phys.* **125**, 67 (1980). DOI 10.1016/0003-4916(80)90119-0

Chapter 5

Spontaneous Symmetry Breaking

In this chapter, I will finally spell out a formal definition of *spontaneous symmetry breaking* (SSB). While the previous chapter was largely phrased in terms of classical field theory, I will now switch entirely to the quantum-theoretic language. This is more than justified by Sect. 5.3, which gives the reader a flavor of the many subtleties, associated with SSB in quantum systems. Some further details on the physical aspects of SSB can be found for instance in the recent lecture notes [1]. A mathematically oriented reader will find the classic book [2] a unique source of additional information.

To motivate what comes below, let me briefly return to the roots of Lie group theory in the study of symmetries of differential equations. Suppose that we have a set of (partial) differential equations invariant under some group G of transformations. A given solution to the equations can fall into one of two classes:

- (i) The solution is not invariant under G . This is the generic case, in which one can use the symmetry to obtain new solutions by the action of transformations from G on the solution already known.
- (ii) The solution is invariant under G . Such solutions are special, and are usually easier to find than a general solution to the same equations. The reason for this is that assuming a priori G -invariance of the solution reduces the number of independent variables of the differential equations.

The situation in quantum physics maps closely to what one does in the context of the theory of differential equations. The major difference is that one tends to focus on the eigenstates of the quantum Hamiltonian as the primary tool to investigate physical observables. For this reason, it is common to start with the ground state, or at least a metastable equilibrium state, and study the excitations above it. This is the quantum counterpart of the theory of small oscillations in classical mechanics [3]. We will see that quantum physics offers natural analogs to both of the above-mentioned cases (i) and (ii) one encounters in differential equations. In both cases, symmetry has profound consequences for the structure of the excitation spectrum. For most of this chapter, I will focus on the symmetry properties of the equilibrium state. I will return to the excitation spectrum in Sect. 5.3, and then again in detail in Chap. 6.

5.1 Physical State and Its Symmetry

A proper discussion of SSB requires a number of new concepts. In order to make the narrative as natural as possible, I will augment it with some simple examples from condensed-matter and statistical physics. This will highlight the close connection between SSB and phase transitions. To parallel the language of statistical mechanics, I will represent the state of a quantum system with a *density operator* (or *density matrix*) ϱ . Readers who need to refresh their memory of the density operator formalism will find an undergraduate-level introduction for instance in [4–6].

We already know that symmetries of quantum systems are represented, if only formally, by unitary operators. An operator U is said to constitute a symmetry of the (pure or mixed) state ϱ if the density operators ϱ and $U\varrho U^\dagger$ are indistinguishable by any measurement. The latter means equal probabilities for any specific outcome of any measurement, and by extension equal averages of all observables. Insofar as our description of the state of the system is free of redundancies, this implies that the density operators must be equal,

$$U\varrho U^\dagger = \varrho \quad (\text{symmetry of a state}) . \quad (5.1)$$

Suppose that ϱ is a pure state, that is $\varrho = |\psi\rangle\langle\psi|$ for some normalized ket-vector $|\psi\rangle$ in the Hilbert space. Then a unitary operator U represents a symmetry of ϱ if and only if $|\psi\rangle$ is an eigenstate of U . One side of the equivalence is obvious: if $|\psi\rangle$ is an eigenstate of U , then $U\varrho U^\dagger = \varrho$. To prove the opposite implication, write $|\langle\psi|U|\psi\rangle|^2$ as $\langle\psi|U|\psi\rangle\langle\psi|U^\dagger|\psi\rangle = \langle\psi|U\varrho U^\dagger|\psi\rangle = 1$. Thus, the states $|\psi\rangle$ and $U|\psi\rangle$ are both normalized to unity and have a unit overlap, which is only possible if they are equal up to a phase. It follows that $|\psi\rangle$ is an eigenstate of U .

In the following, I will frequently use a shorthand notation for the average of an observable A in the state ϱ ,

$$\langle A \rangle_\varrho \equiv \text{tr}(\varrho A) . \quad (5.2)$$

It is a simple corollary of (5.1) that a symmetry transformation of observables, $A \rightarrow U^\dagger A U$, does not affect their average in symmetric states. To see this, write $\langle U^\dagger A U \rangle_\varrho$ as $\text{tr}(\varrho U^\dagger A U) = \text{tr}(U\varrho U^\dagger A) = \text{tr}(\varrho A)$, which is just $\langle A \rangle_\varrho$. This conclusion holds for discrete and continuous symmetries alike.

5.1.1 Broken and Unbroken Symmetry

Suppose now that the Hamiltonian of a system is invariant under a set of transformations that span a group G . These are represented, even if just formally, by a set of unitary operators $U(g)$, $g \in G$. A randomly chosen state ϱ will not be symmetric under all of these. Those transformations that are symmetries of ϱ will span a subgroup of G called the *unbroken subgroup* of ϱ , denoted by

$$H_\varrho \equiv \{g \in G \mid U(g)\varrho U(g)^\dagger = \varrho\} . \quad (5.3)$$

All other elements of G are said to be *spontaneously broken* in the state ϱ . This is our formal definition of SSB.

Example 5.1

Pure spin states of a spin-1/2 particle comprise a two-dimensional Hilbert space isomorphic to \mathbb{C}^2 . This Hilbert space carries a representation of the spin group, $G \simeq \text{SU}(2)$, generated by the spin operator \mathbf{S} . Consider the projection of spin to a direction, defined by the unit vector $\mathbf{n} = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$,

$$\mathbf{n} \cdot \mathbf{S} = \frac{\hbar}{2} \mathbf{n} \cdot \boldsymbol{\tau} = \frac{\hbar}{2} \begin{pmatrix} \cos \theta & e^{-i\varphi} \sin \theta \\ e^{i\varphi} \sin \theta & -\cos \theta \end{pmatrix}. \quad (5.4)$$

Here θ, φ are the standard spherical angles and $\boldsymbol{\tau}$ is the vector of Pauli matrices. The projected spin operator has an eigenvector $|\mathbf{n}, +\rangle$ with eigenvalue $+\hbar/2$,

$$\mathbf{n} \cdot \mathbf{S} |\mathbf{n}, +\rangle = \frac{\hbar}{2} |\mathbf{n}, +\rangle, \quad |\mathbf{n}, +\rangle = \begin{pmatrix} \cos \theta/2 \\ e^{i\varphi} \sin \theta/2 \end{pmatrix}. \quad (5.5)$$

Any nonzero vector in \mathbb{C}^2 can, up to an overall factor, be cast in this form with a suitable choice of θ, φ . Hence, any nonzero vector in this Hilbert space is an eigenstate of a projection of the spin operator to some direction. We conclude that for any pure state of a spin-1/2 particle, the unbroken subgroup is isomorphic to $\text{U}(1)$, and consists of spin rotations generated by $\mathbf{n} \cdot \mathbf{S}$.

How do we check whether a given symmetry is spontaneously broken in a given state ϱ ? For the density operators ϱ and $U\varrho U^\dagger$ to be different, there must be at least one observable A whose average in the two states differs. This observable must itself not be invariant under the symmetry, as $U^\dagger A U = A$ would imply

$$\langle A \rangle_{U\varrho U^\dagger} = \text{tr}(U\varrho U^\dagger A) = \text{tr}(\varrho U^\dagger A U) = \text{tr}(\varrho A) = \langle A \rangle_\varrho. \quad (5.6)$$

In order to detect SSB, one must therefore use an observable that itself breaks the symmetry. This fairly simple observation forms a conceptual foundation of the approach to SSB developed in Sect. 5.2.

One observable that is by definition invariant under all the operators $U(g)$ from the symmetry group is the Hamiltonian itself. Hence, for those $g \in G$ that are spontaneously broken, the state $U(g)\varrho U(g)^\dagger$ differs from ϱ but has the same average energy. This presents a serious conundrum. Suppose that ϱ is a purported ground state of a quantum system. From quantum mechanics, we are used to thinking about the ground state as a unique vector in the Hilbert space. But now $U(g)\varrho U(g)^\dagger$ for any $g \notin H_\varrho$ is a different state of the same energy. How do we know which of the different states having the same energy to choose as *the* ground state?

We noticed already in Chap. 2 that SSB implies the existence of degenerate ground states. This property is so universal that it earned the privilege of being my first “moral lesson” in Sect. 2.4. But now our problem is even worse. In the classical analysis of Chap. 2, the various candidate ground states could be uniquely labeled by the expectation value of a scalar field. In the present fully quantum setting, we have not only the set of density operators $U(g)\varrho U(g)^\dagger$ with $g \in G$. We can even make linear combinations of these states. In the desire to identify a unique ground state, we might wish to take a “democratic average” of all the would-be ground states, connected by transformations from G . This is an idea that certainly deserves at least consideration, even if it will eventually turn out to be physically incorrect. However, the related mathematical procedure of averaging over different elements of the symmetry group will actually prove quite useful.

5.1.2 Symmetrization by Group Averaging

When performing an average over the symmetry group, I will repeatedly make use of a simple statement that I will formulate as a standalone lemma. Let G be a finite group or a compact Lie group, and $U(g)$ be its unitary representation on a finite-dimensional vector space. Then

$$\mathcal{P} \equiv \frac{1}{|G|} \sum_{g \in G} U(g) \quad (5.7)$$

is a projector to the subspace of vectors invariant under G . In particular, if the representation of G by $U(g)$ does not contain in its decomposition any singlet (that is, a trivial one-dimensional representation of G), then the action of \mathcal{P} on any vector gives zero. The proof of the lemma is very simple and I will therefore leave it up to the reader. Let me just add a comment on the sum over group elements. Equation (5.7) as it stands is only valid for finite groups; $|G|$ then denotes the number of group elements. For compact Lie groups, the sum has to be replaced by group integration with an invariant measure on G ; $|G|$ then stands for the volume of G with respect to this measure. An interested reader will find more details in Chap. 3 of [7]. In the following, I will pragmatically use a sum over $g \in G$ as in (5.7) without repeating this disclaimer.

Let us now get back to density operators. The action of the group G on these is defined by $g : \varrho \rightarrow U(g)\varrho U(g)^\dagger$. This suggests that we consider the following statistical average,

$$\bar{\varrho} \equiv \frac{1}{|G|} \sum_{g \in G} U(g)\varrho U(g)^\dagger. \quad (5.8)$$

This average still satisfies the requirements on a density operator: it is Hermitian, positive-semidefinite and has a unit trace. Moreover, it is manifestly invariant under all transformations from G . It therefore represents a (mixed) physical state of the

system which is perfectly symmetric under the whole group G , yet has the same average energy as ϱ .

In case the original state is pure, $\varrho = |\psi\rangle\langle\psi|$, we have the alternative choice to symmetrize the ket-vector $|\psi\rangle$ rather than the density operator,

$$|\psi\rangle \rightarrow \mathcal{P}|\psi\rangle = \frac{1}{|G|} \sum_{g \in G} U(g) |\psi\rangle . \quad (5.9)$$

This averaged ket-vector, if well-defined, represents a pure state of the same energy as $|\psi\rangle$ that is symmetric under the whole group G . The catch is that the ket-vector may be ill-defined. By our symmetrization lemma, this happens whenever $|\psi\rangle$ belongs to a representation of G that does not contain any singlet. Recall, for instance, Example 5.1. There, the whole Hilbert space \mathbb{C}^2 constitutes an irreducible representation of the spin group $G \simeq \text{SU}(2)$. The naive average (5.9) therefore gives zero no matter which pure state we start from. On the other hand, the average density operator (5.8) must describe *some* physical state of the spin-1/2 particle. By Schur's lemma, $\bar{\varrho}$ must be proportional to the identity operator on \mathbb{C}^2 . The requirement that $\text{tr } \bar{\varrho} = 1$ then fixes the normalization so that $\bar{\varrho} = (1/2) \text{id}$, for any choice of the initial state ϱ . This mixed state describes an unpolarized distribution on the Hilbert space.

It seems that we managed to have our cake and eat it. By the simple averaging procedure, we can make any state G -invariant without changing its energy. Have we done away with the whole notion of SSB? To see that the group averaging is not so innocuous, we need to work out the consequences. Consider a set of observables A^i such that under the action of the operators $U(g)$, they transform by some (real and orthogonal) matrix representation \mathcal{R} of G ,

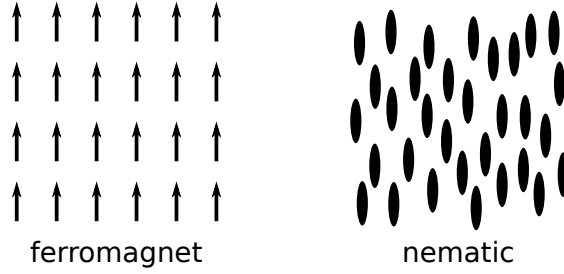
$$U(g)^\dagger A^i U(g) = \mathcal{R}(g)^i_j A^j , \quad g \in G . \quad (5.10)$$

Suppose that all these observables and their nontrivial linear combinations break (part of) the symmetry under G . In other words, the representation \mathcal{R} does not contain any singlet of G in its decomposition. It then follows as a simple consequence of the definition of the density operator (5.8) and the symmetrization lemma that

$$\begin{aligned} \langle A^i \rangle_{\bar{\varrho}} &= \text{tr}(\bar{\varrho} A^i) = \frac{1}{|G|} \sum_{g \in G} \text{tr}[U(g) \varrho U(g)^\dagger A^i] \\ &= \frac{1}{|G|} \sum_{g \in G} \text{tr}[\varrho U(g)^\dagger A^i U(g)] = \frac{1}{|G|} \sum_{g \in G} \mathcal{R}(g)^i_j \langle A^j \rangle_{\varrho} = 0 . \end{aligned} \quad (5.11)$$

We could have anticipated this: in the G -invariant state $\bar{\varrho}$, the average of any observable breaking the symmetry vanishes.

Fig. 5.1 Comparison of the symmetry-breaking states of ferromagnets and nematic liquid crystals. The symmetry of the former corresponds to that of a regular lattice of oriented, mutually aligned spins. Nematics, on the other hand, can be represented by mutually aligned undirected rods with random positions



Why should this be bad? Any quantum system possesses a distinguished state invariant under its symmetry group G : the thermodynamic equilibrium. This is represented by the canonical density operator

$$\varrho_\beta \equiv \frac{1}{Z} e^{-\beta H}, \quad (5.12)$$

where β is the inverse temperature and $Z \equiv \text{tr} \exp(-\beta H)$ the canonical partition function. The claim therefore is that in thermodynamic equilibrium, any observable breaking the symmetry of the system must average to zero. There cannot be any SSB in thermodynamic equilibrium. This would be very pretty, were it not in a blatant contradiction with empirical evidence.

Example 5.2

Solid ferromagnets can be visualized in terms of a lattice of mutually aligned spins (see the left panel of Fig. 5.1). The ferromagnetic state obviously breaks rotational symmetry. Moreover, the presence of the crystal lattice also breaks spatial translations. Both rotations and translations are fundamental symmetries that are expected to be preserved by the microscopic dynamics of any material. Hence ferromagnetism certainly is a macroscopic manifestation of SSB.

In nematic liquid crystals (right panel of Fig. 5.1), long organic molecules are aligned along the same direction. However, the positions of the molecules are random as in ordinary liquids. As far as symmetry is concerned, the molecules can be represented by (undirected) rigid rods or ellipsoids with randomized positions. Hence the nematic state breaks rotational symmetry but not spatial translations. See Sect. 2.7 of [8] for basic phenomenology of liquid crystals.

One of the greatest accomplishments of physics is the discovery that the fundamental laws of nature possess a high degree of symmetry. How is it then possible at all that we observe so many manifestly asymmetric macroscopic states of matter? There seem to be two logical possibilities to avoid a contradiction:

- The observed asymmetric states are not in thermodynamic equilibrium, and thus may be at most metastable.

- The asymmetry of the observed states is a result of perturbations breaking the symmetry of the microscopic interactions.

Both of these possibilities are relevant. In order to understand their role, we have to dive deeper into the thermodynamics of physical systems with SSB. In the next section, I will focus solely on the canonical density operator ϱ_β as the state of greatest importance for macroscopic observations.

5.2 Effect of External Perturbations

In Sect. 5.1.2, I introduced the set of operators A^i as mere observables. Let us now see what happens if we allow them to actually affect the dynamics of the system. To that end, I couple each A^i to an external field λ_i and perturb the Hamiltonian of the system by a term linear in both,

$$H(\lambda) \equiv H - \lambda_i A^i . \quad (5.13)$$

In the case of a spin-1/2 particle discussed in Example 5.1, A^i can be chosen for instance as the components of the spin operator. The perturbations λ_i can then be interpreted as an external magnetic field. In general, one can think of λ_i as a set of “chemical potentials” associated with the observables A^i . In presence of the perturbations, the canonical density operator is deformed to

$$\varrho_{\beta,\lambda} = \frac{1}{Z(\lambda)} e^{-\beta H(\lambda)} , \quad (5.14)$$

and the modified partition function reads $Z(\lambda) = \text{tr} e^{-\beta H(\lambda)} = \text{tr} \exp[-\beta(H - \lambda_i A^i)]$. The symmetry of the Hamiltonian H under the group G is reflected in the invariance of the partition function under the transformation $\lambda_i \xrightarrow{g} \lambda_j \mathcal{R}(g)^j_i$ of its variables. This follows from a simple manipulation,

$$\begin{aligned} Z(\lambda) &= \text{tr} [U(g)^\dagger e^{-\beta H(\lambda)} U(g)] = \text{tr} \exp[-\beta U(g)^\dagger H(\lambda) U(g)] \\ &= \text{tr} \exp\{-\beta [H - \lambda_i \mathcal{R}(g)^i_j A^j]\} = Z(\lambda \mathcal{R}(g)) . \end{aligned} \quad (5.15)$$

It is convenient to trade the partition function for the free energy $F(\lambda)$ by $Z(\lambda) \equiv e^{-\beta F(\lambda)}$. The free energy inherits the symmetry of the partition function under G -transformations of λ_i . Moreover, it makes it easy to evaluate the statistical average of A^i in presence of the perturbations,

$$\langle A^i \rangle_{\varrho_{\beta,\lambda}} = \frac{1}{Z(\lambda)} \text{tr} \{ A^i \exp[-\beta(H - \lambda_j A^j)] \} = -\frac{\partial F(\lambda)}{\partial \lambda_i} . \quad (5.16)$$

One of the distinguishing features of ferromagnets is that once magnetized by an external magnetic field, the *macroscopic* magnetization persists even after the external field is turned off. With this in mind, we will be particularly interested in the

behavior of $\langle A^i \rangle_{\mathcal{Q}_{\beta,\lambda}}$ when the volume V of the system is large and simultaneously λ_i is small. In other words, we want to take the double limit $V \rightarrow \infty$ and $\lambda_i \rightarrow 0$. This is where interesting things start to happen.

5.2.1 Taking the Thermodynamic Limit

In a finite volume and for any nonzero temperature, the partition function, and thus the free energy, is analytic in λ_i . When all the λ_i are small, one can therefore perform a Taylor expansion,

$$F(\lambda) = F(0) + \left. \frac{\partial F(\lambda)}{\partial \lambda_i} \right|_{\lambda=0} \lambda_i + \mathcal{O}(\lambda^2) . \quad (5.17)$$

But the linear term in the expansion is forbidden by the G -invariance of the free energy combined with the symmetrization lemma. Thus, in a finite volume and at nonzero temperature, one has a well-defined limit

$$\lim_{\lambda \rightarrow 0} \langle A^i \rangle_{\mathcal{Q}_{\beta,\lambda}} = 0 . \quad (5.18)$$

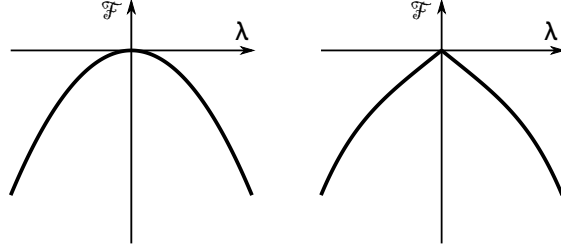
This should not be particularly surprising.

How does the behavior of free energy as a function of λ_i change when we take the limit of infinite volume or zero temperature? Some common mathematical features of these two limits may be discussed jointly. It is however the large-volume (thermodynamic) limit that is responsible for the existence of macroscopically distinct phases of matter. Only in special cases may the zero-temperature limit lead to nontrivial physics even in a finite volume; I will give an explicit example below. The general discussion of the effects of external perturbations will however be phrased in terms of the limit $V \rightarrow \infty$. The temperature will be assumed to be fixed and nonzero unless explicitly stated otherwise.

It is convenient to trade the free energy for its density, $\mathcal{F}(\lambda) \equiv F(\lambda)/V$, which is expected to be well-defined for any volume, including the limit $V \rightarrow \infty$. It is sensible to assume that even in this limit, the free energy density remains continuous as a function of λ_i . However, the analyticity may be lost. With regard to the behavior of the average of A^i around the point $\lambda_i = 0$, there are three possible scenarios:

- The free energy density remains analytic in the limit $V \rightarrow \infty$. This automatically guarantees the validity of (5.18). There is no SSB once the external field λ_i is turned off.
- The free energy density becomes nonanalytic in the limit $V \rightarrow \infty$, yet (5.18) remains valid. This is an interesting borderline case that is relevant for the physics of low-dimensional systems.
- The free energy density becomes nonanalytic in the limit $V \rightarrow \infty$. The average $\langle A^i \rangle_{\mathcal{Q}_{\beta,\lambda}}$ remains nonzero even if the limit $\lambda_i \rightarrow 0$ is subsequently taken. This is the paradigm of SSB, illustrated by the right panel of Fig. 5.2.

Fig. 5.2 Schematic illustration of scenarios in which the free energy density \mathcal{F} is (left panel) and is not (right panel) analytic at $\lambda_i = 0$. In the latter case, the observables A^i may retain a nonzero average even in the limit $\lambda \rightarrow 0$. Its value may however depend on the way the limit is taken



An alert reader might wonder why I focused on possible nonanalytic behavior of \mathcal{F} at $\lambda_i = 0$, yet tacitly assumed that analyticity survives the thermodynamic limit for nonzero λ_i . This is intimately related to the degeneracy of the ground (equilibrium) state. What I really assumed was that adding the external fields λ_i breaks the symmetry of the system sufficiently so as to make the ground state unique. I will further elaborate on this assumption in Sect. 5.2.2. The point is that the free energy density typically becomes nonanalytic as a consequence of a competition of degenerate ground states with different values of $\langle A^i \rangle_{\mathcal{Q}_{\beta, \lambda}}$. This is something one expects only at the G -invariant point $\lambda_i = 0$.

It is instructive to work out at least one concrete example to verify our prediction how SSB may emerge when a particular limit of the equilibrium state is taken. Ideally, this example should be exactly solvable in order to leave no doubt about the validity of the conclusions. It is therefore time to meet the *Ising model*, the most well-known and best-understood model of ferromagnetism in statistical physics. The price for all its benefits is that the limit in which the free energy density becomes nonanalytic is realized by taking the temperature to zero; the $V \rightarrow \infty$ limit itself is nearly trivial. I will largely follow the treatment in Sect. 2.2 of [9].

Example 5.3

Consider a one-dimensional chain of N spin variables σ_i , allowed to take values ± 1 . The Hamiltonian is given by

$$H = -J \sum_{i=1}^N (\sigma_i \sigma_{i+1} - 1) - B \sum_{i=1}^N \sigma_i, \quad (5.19)$$

where σ_{N+1} is identified with σ_1 to ensure a periodic boundary condition. The (positive) coupling J represents interaction of nearest-neighbor spins, and B an external magnetic field. In the absence of the external field, the Hamiltonian (5.19) has a $G \simeq \mathbb{Z}_2$ symmetry under the spin flip $\sigma_i \rightarrow -\sigma_i$.

The partition function of the model is easily calculable thanks to the fact that the exponential of a sum equals a product of exponentials. Thus, the contribution of a

spin configuration $\{\sigma_i\}_{i=1}^N$ to the partition function equals $\prod_{i=1}^N V(\sigma_i, \sigma_{i+1})$, where

$$V(\sigma, \sigma') \equiv \exp \left[\beta J (\sigma \sigma' - 1) + \frac{\beta B}{2} (\sigma + \sigma') \right]. \quad (5.20)$$

The latter can be thought of as a matrix element of the so-called *transfer matrix*,

$$V = \begin{pmatrix} e^{\beta B} & e^{-2\beta J} \\ e^{-2\beta J} & e^{-\beta B} \end{pmatrix}. \quad (5.21)$$

With the periodic boundary condition on the spins, the partition function of the Ising model is now $Z = \text{tr } V^N = \lambda_+^N + \lambda_-^N$, where $\lambda_{\pm} = \cosh \beta B \pm \sqrt{\sinh^2 \beta B + e^{-4\beta J}}$ are the eigenvalues of the transfer matrix. The average spin at each site of the chain is obtained by applying (5.16) to $\mathcal{F}(B) \equiv F(B)/N$,

$$\langle \sigma_i \rangle_{\mathcal{Q}_{\beta, B}} = \frac{\sinh \beta B}{\sqrt{\sinh^2 \beta B + e^{-4\beta J}}} \frac{\lambda_+^N - \lambda_-^N}{\lambda_+^N + \lambda_-^N} \xrightarrow{N \rightarrow \infty} \frac{\sinh \beta B}{\sqrt{\sinh^2 \beta B + e^{-4\beta J}}}. \quad (5.22)$$

The total number of spins N plays the role of volume, and sending it to infinity amounts to the thermodynamic limit.

It is obvious from (5.22) that the Ising model does not feature SSB at any nonzero temperature. Indeed, whether N is finite or not, the average spin necessarily vanishes in the limit $B \rightarrow 0$ whenever the temperature is nonzero. On the other hand, taking the temperature to zero gives

$$\lim_{\beta \rightarrow \infty} \langle \sigma_i \rangle_{\mathcal{Q}_{\beta, B}} = \text{sgn } B, \quad (5.23)$$

again regardless of whether N is finite or not. Taking subsequently a (one-sided) limit of vanishing magnetic field will render the average spin nonzero, with a sign aligned with that of the magnetic field. This can be traced to the properties of the canonical density operator, which in the zero-temperature limit becomes a projector to the subspace of states with lowest energy. For the Ising Hamiltonian (5.19),

$$\begin{aligned} B > 0 : \quad & \lim_{\beta \rightarrow \infty} \mathcal{Q}_{\beta, B} = |++ \cdots\rangle \langle ++ \cdots|, \\ B = 0 : \quad & \lim_{\beta \rightarrow \infty} \mathcal{Q}_{\beta, B} = \frac{1}{2} (|++ \cdots\rangle \langle ++ \cdots| + |-- \cdots\rangle \langle -- \cdots|), \\ B < 0 : \quad & \lim_{\beta \rightarrow \infty} \mathcal{Q}_{\beta, B} = |-- \cdots\rangle \langle -- \cdots|, \end{aligned} \quad (5.24)$$

where $|++ \cdots\rangle$ and $|-- \cdots\rangle$ denote respectively the normalized states with all spins being positive and negative. Note that the \mathbb{Z}_2 -symmetric state on the second line of (5.24) is unstable: an infinitesimally weak external field will project it onto one of the two fully polarized states.

The example demonstrates that taking an appropriate (infinite-volume or zero-temperature) limit in the presence of a perturbation may lead to an equilibrium state violating the symmetry of the system. The system then remains in an asymmetric state even if the perturbation is subsequently removed. It is very important that the limits of infinite volume and vanishing external field are performed in this order. Indeed, taking $\lambda_i \rightarrow 0$ first would delete the average of any symmetry-breaking observable in line with (5.18). Colloquially, one often speaks of the “noncommutativity of limits,”

$$\lim_{\lambda \rightarrow 0} \lim_{V \rightarrow \infty} \langle A^i \rangle_{\varrho_{\beta, \lambda}} \neq \lim_{V \rightarrow \infty} \lim_{\lambda \rightarrow 0} \langle A^i \rangle_{\varrho_{\beta, \lambda}}, \quad (5.25)$$

as a smoking gun of SSB.

This concludes our excursion to thermodynamics. We have seen why asymmetric macroscopic states of matter can—and why symmetric macroscopic states may not—be stable. Identifying a (meta)stable equilibrium is a necessary first step if we ever want to understand the spectrum of a quantum system. Our next goal is therefore to establish a general operational procedure for finding a stable ground state.

5.2.2 Order Parameter and the Vacuum Manifold

In the Ising model, it was enough to distinguish the two possible asymmetric ground states by the sign of the external field. This is eventually because the model possesses a discrete \mathbb{Z}_2 symmetry. In systems with continuous symmetry we expect a continuum of degenerate equilibrium states. An already familiar example is the isotropic ferromagnet, where the spontaneous magnetization can take an arbitrary direction in space. Thus, we must be careful when taking the limit of vanishing external fields.

Suppose that we are given a specific set of fields, $\hat{\lambda}_i$. I will define the limit of vanishing perturbations by changing the magnitude but keeping the “direction” of these fields. This can be implemented precisely by using a positive scaling parameter ϵ . We can thus construct a macroscopic thermal state $\varrho_{\beta, \hat{\lambda}}^\infty$ by first taking the thermodynamic limit and then removing the external fields via

$$\varrho_{\beta, \hat{\lambda}}^\infty \equiv \lim_{\epsilon \rightarrow 0^+} \lim_{V \rightarrow \infty} \varrho_{\beta, \epsilon \hat{\lambda}}. \quad (5.26)$$

In presence of SSB, the average

$$\langle A^i \rangle_{\varrho_{\beta, \hat{\lambda}}^\infty} = \lim_{\epsilon \rightarrow 0^+} \lim_{V \rightarrow \infty} \langle A^i \rangle_{\varrho_{\beta, \epsilon \hat{\lambda}}} \quad (5.27)$$

is expected to be nonzero. This is an important proxy of SSB, known as the *order parameter*. It does exactly what its name suggests: parameterize the appearance of order in the thermodynamic state of matter described by the density operator $\varrho_{\beta, \hat{\lambda}}^\infty$.

Is the state (5.26) stable? This question is closely related to the choice of observables A^i . Namely, the set A^i should be “complete” in the sense that the values of the order parameters $\langle A^i \rangle_{\varrho_{\beta, \hat{\lambda}}^\infty}$ determine the thermal state $\varrho_{\beta, \hat{\lambda}}^\infty$ uniquely. In the limit

of zero temperature, this means that $\varrho_{\beta,\lambda}^\infty$ should be a pure state, analogously to the $B \neq 0$ states in (5.24) for the Ising model.

Example 5.4

In the Ising model, the role of the order parameter is played by the average spin $\langle \sigma_i \rangle$. According to (5.23), in the zero-temperature limit in the presence of a nonzero magnetic field, $\langle \sigma_i \rangle$ acquires one of two possible values. These uniquely specify the corresponding symmetry-breaking pure ground states via (5.24).

In an isotropic ferromagnet, the order parameter can be taken as the average of the operator of total spin (magnetization) \mathbf{S} . Specifying the vector $\langle \mathbf{S} \rangle$ is then sufficient to determine the stable macroscopic equilibrium uniquely. This is however not necessarily the only possible choice of order parameter. Instead of the three components of the vector S_i , we might try to consider for instance the traceless symmetric tensor $T_{ij} \equiv S_i S_j + S_j S_i - (2/3) \delta_{ij} S^2$. This set of operators transforms under a nontrivial irreducible representation of the spin group $G \simeq \text{SU}(2)$, hence it satisfies the requirements we have imposed on the operators A^i . The average $\langle T_{ij} \rangle$, however, does not specify a unique equilibrium state. Namely, this order parameter cannot distinguish between the two states that only differ by the overall sign of $\langle \mathbf{S} \rangle$.

The order parameter is usually much easier to determine, or at least estimate, than the actual equilibrium state, $\varrho_{\beta,\lambda}^\infty$. Its real value lies in the fact that it carries the same information about the symmetry of the equilibrium state as the density operator $\varrho_{\beta,\lambda}^\infty$ itself. To see this, just note that

$$\langle A^i \rangle_{U(g)\varrho_{\beta,\lambda}^\infty U(g)^\dagger} = \text{tr}[U(g)\varrho_{\beta,\lambda}^\infty U(g)^\dagger A^i] = \mathcal{R}(g)^i_j \langle A^j \rangle_{\varrho_{\beta,\lambda}^\infty}. \quad (5.28)$$

This shows that for any $g \in G$ that is a symmetry of $\varrho_{\beta,\lambda}^\infty$, the matrix $\mathcal{R}(g)$ leaves the order parameter unchanged. But the opposite is true as well. For any $g \in G$ such that $\mathcal{R}(g)^i_j \langle A^j \rangle_{\varrho_{\beta,\lambda}^\infty} = \langle A^i \rangle_{\varrho_{\beta,\lambda}^\infty}$, the assumed completeness of the set of operators A^i guarantees that the states $\varrho_{\beta,\lambda}^\infty$ and $U(g)\varrho_{\beta,\lambda}^\infty U(g)^\dagger$ are the same. The pattern of SSB in a macroscopic equilibrium state is therefore completely determined by the associated order parameter(s).

Finally, let me briefly return to the correspondence between the order parameter and the pure ground state (*vacuum*) obtained by taking the zero-temperature limit of (5.26). It is in principle possible that the physical system possesses two completely unrelated ground states of equal energy. (This happens for instance at a first-order phase transition.) Barring such accidental degeneracy, however, one expects the possible candidate vacua to be mutually related by (broken) symmetry transformations. It is therefore sufficient to know just one of the vacuum states; it should be possible to reconstruct all the others by the action of the operators $U(g)$, $g \in G$. We will see in Sect. 5.3 that this procedure is subtle in the limit of infinite volume. What one can however do safely is to trade the chosen vacuum state ϱ for the corresponding

order parameter $\langle A^i \rangle_{\mathcal{O}}$. All other possible values of the order parameter can then be generated by acting on $\langle A^i \rangle_{\mathcal{O}}$ with the matrices $\mathcal{R}(g)$.

In case G is a Lie group, the possible vacuum values of the order parameter span a manifold of dimension $\dim G - \dim H_{\mathcal{O}}$; see Appendix A.1 for the mathematical background. This is known as the *vacuum manifold*. The central tenet of this book is that the physics of systems with spontaneously broken continuous symmetry can be captured by a low-energy *effective field theory* (EFT). The form of this EFT is largely fixed by the geometry of the vacuum manifold. The latter can in turn be understood in terms of G and its unbroken subgroup $H_{\mathcal{O}}$. To develop the EFT framework based on the pattern of SSB is the goal of Parts III and IV of the book.

Example 5.5

Quantum chromodynamics (QCD) is a gauge theory of the strong nuclear interaction. Strongly-interacting matter is represented by the quark field Ψ_i^α . This is a Dirac spinor where α is an index of the fundamental representation of the SU(3) gauge group of QCD. The index i represents quark flavor and belongs to the fundamental representation of SU(n_f). In applications, the number of relevant (light) quark flavors n_f is usually two or three.

In the limit of vanishing quark masses, QCD possesses a $G \simeq \text{SU}(n_f)_L \times \text{SU}(n_f)_R$ symmetry, consisting of independent flavor transformations of the left- and right-handed components of Ψ_i^α . In the ground state of QCD, this symmetry is spontaneously broken by the complex order parameter $\langle \Sigma_{ij} \rangle \equiv \langle \delta_{\alpha\beta} \bar{\Psi}_{iL}^\alpha \Psi_{jR}^\beta \rangle$. Note that $\text{tr}(\Sigma + \Sigma^\dagger)$ is the usual mass term for a Dirac spinor, here with equal masses for all quark flavors. We can therefore think of “switching on” small equal quark masses as an analog of the external field λ_i that selects a particular vacuum state. In this vacuum state, $\langle \Sigma_{ij} \rangle = \sigma \delta_{ij}$, where the constant σ is an intrinsic scale of QCD. The unbroken, “vector” subgroup of G , $H \simeq \text{SU}(n_f)_V$, consists of identical unitary transformations of left- and right-handed quarks. Broken symmetry transformations convert this vacuum to other vacuum states, in which $\langle \Sigma_{ij} \rangle / \sigma$ is unitary. The vacuum manifold is therefore diffeomorphic to the Lie group SU(n_f).

5.2.3 Intermediate Summary

We have taken quite a conceptual tour to arrive at the notions of order parameter and vacuum manifold, essential for understanding the physics of SSB. Let us therefore make a little break and review what we have done in this chapter so far.

I started Sect. 5.1 with a basic definition of a state of a quantum system in terms of a density operator, and of its symmetry. Section 5.1.1 makes the distinction between broken and unbroken symmetries, and highlights energetic degeneracy of different states as a universal feature of SSB. A naive way out of the puzzle this poses might be to consider only states respecting the full symmetry G of the system.

In Sect. 5.1.2, I motivated the consideration of such G -symmetric states by recalling the thermal (canonical) density operator. The latter however renders the average of any observable breaking the symmetry zero. This is in a striking contradiction with the existence of asymmetric, macroscopically stable states of matter in nature. A careful consideration of the effects of small external perturbations in Sect. 5.2.1 shows that these may drive the system towards a symmetry-breaking state. In fact, in the large-volume limit, even an infinitesimally small perturbation may be sufficient to destabilize the G -invariant thermal state.

For most applications, it is mandatory to build a physical description of a given system upon a stable ground state. Such vacuum states are constructed in Sect. 5.2.2 by taking the thermodynamic limit in presence of a perturbation, before the latter may be switched off. The different, degenerate vacuum states can now be labeled uniquely by different values of an order parameter, corresponding to the average of a suitably chosen set of observables. The choice of the order parameter itself is not unique. Depending on this choice, the same vacuum state may be represented by a point in different order parameter spaces. However, the set of all degenerate vacuum states spans a vacuum manifold, whose geometric structure is determined solely by the symmetry group G and its unbroken subgroup.

This completes the background needed to understand the following chapters. When talking about a vacuum or ground state, I will always implicitly have in mind a pure state, obtained by the limiting procedure in (5.26). Also, I will from now on always assume vanishing thermodynamic temperature. Nonzero temperature was only used in this chapter as a tool to avoid singularities in the partition function.

The story of SSB, however, does not end here. A closer look reveals a number of intriguing aspects that I have skipped so far in the desire to provide a minimal self-contained introduction to SSB. While not of direct relevance for the rest of the book, it would be a pity to omit these aspects altogether. I therefore mention some of them at least briefly in the next section, in the form of a case study.

5.3 Some Subtle Features of Spontaneous Symmetry Breaking

In the previous section, the thermodynamic limit played a crucial role in selecting a unique symmetry-breaking ground state. This begs the question of what one can expect in volumes that are large yet finite. After all, real macroscopic systems in nature certainly are finite, albeit possibly large from the point of view of microscopic physics. To what extent are then symmetry-breaking states stabilized by an external perturbation? Besides, how does the finite-volume quantum ground state look in the absence of perturbations?

These are some of the questions I will address here. I am not aware of any formal framework that would allow one to tackle these questions in full generality. In fact, the answer to some of them may depend on the specific physical system. Instead of trying to be general, I will therefore work out in detail one concrete example, just to

illustrate what is at stake. Some rigorous results on the spectrum of quantum systems with SSB in a finite volume can be found in [10] or in Part I of [11].

5.3.1 Free Schrödinger Field in Finite Volume

Following Sect. 3 of [12], consider the theory of a free Schrödinger field ψ , defined by the Lagrangian density

$$\mathcal{L} = i\psi^\dagger \partial_0 \psi - \frac{1}{2m} \nabla \psi^\dagger \cdot \nabla \psi , \quad (5.29)$$

where m is a mass parameter. I will quantize this theory in a d -dimensional box of size L , with volume $V = L^d$, endowed with a periodic boundary condition. The boundary condition admits only discrete values of momentum \mathbf{p} , given component-wise by

$$p_r = \frac{2\pi n_r}{L} , \quad n_r \in \mathbb{Z} . \quad (5.30)$$

The field operator can be Fourier-expanded in terms of plane-wave solutions to the Schrödinger equation. In the Schrödinger picture,

$$\psi(\mathbf{x}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{p}} a_{\mathbf{p}} e^{i\mathbf{p} \cdot \mathbf{x}} , \quad (5.31)$$

where $a_{\mathbf{p}}$ is the annihilation operator for a single-particle state of momentum \mathbf{p} . This is normalized so that together with the creation operator $a_{\mathbf{p}}^\dagger$, it satisfies $[a_{\mathbf{p}}, a_{\mathbf{q}}^\dagger] = \delta_{\mathbf{p}\mathbf{q}}$. The normalization of the plane-wave expansion by the factor $1/\sqrt{V}$ then guarantees that the canonical coordinate $\psi(\mathbf{x})$ and its conjugate momentum $i\psi^\dagger(\mathbf{x})$ obey the canonical commutation relation $[\psi(\mathbf{x}), i\psi^\dagger(\mathbf{y})] = i\delta^d(\mathbf{x} - \mathbf{y})$.

Within second quantization, all these operators act on a Hilbert space, built using the Fock construction. The starting point is the Fock vacuum $|0\rangle$, defined by the condition $a_{\mathbf{p}} |0\rangle = 0$ for all allowed values of \mathbf{p} . An orthogonal basis of the space is then obtained by acting on $|0\rangle$ with a finite number of creation operators in all possible ways. The Schrödinger Hamiltonian has a simple expression in terms of the annihilation and creation operators,

$$H = \frac{1}{2m} \int d^d \mathbf{x} \nabla \psi^\dagger(\mathbf{x}) \cdot \nabla \psi(\mathbf{x}) = \sum_{\mathbf{p}} \frac{\mathbf{p}^2}{2m} a_{\mathbf{p}}^\dagger a_{\mathbf{p}} . \quad (5.32)$$

This is a positive-semidefinite operator. The lowest (zero) energy is reached by any state $|\Omega\rangle$ in the Hilbert space that satisfies

$$a_{\mathbf{p}} |\Omega\rangle = 0 \quad \text{for all } \mathbf{p} \neq \mathbf{0} \quad (\text{ground state}) . \quad (5.33)$$

There are infinitely many such states in the Hilbert space, in particular any state obtained from $|0\rangle$ using the zero-momentum creation operator a_0^\dagger . Hence, the subspace of states of zero energy is identical to the Fock space of a linear harmonic oscillator.

The infinite degeneracy of the ground state hints at SSB. To see what symmetry might be spontaneously broken here, let us think what conserved charges there are in the first place. There turn out to be infinitely many of them. One can take for instance any operator of the form

$$Q_f \equiv \sum_{\mathbf{p}} f(\mathbf{p}) a_{\mathbf{p}}^\dagger a_{\mathbf{p}} = \int d^d \mathbf{x} \psi^\dagger(\mathbf{x}) f(-i\nabla) \psi(\mathbf{x}) , \quad (5.34)$$

where f is assumed to be a real analytic function. This class of operators includes the Hamiltonian itself, with $f(\mathbf{p}) = \mathbf{p}^2/(2m)$, the momentum operator $\mathbf{P} = \sum_{\mathbf{p}} \mathbf{p} a_{\mathbf{p}}^\dagger a_{\mathbf{p}}$, and the operator of particle number $Q_1 = \sum_{\mathbf{p}} a_{\mathbf{p}}^\dagger a_{\mathbf{p}}$. From (5.34) it is obvious that all the operators Q_f mutually commute with each other, hence all are conserved. By (4.34), they induce a local transformation of $\psi(\mathbf{x})$ with parameter ϵ , $\delta\psi(\mathbf{x}) = -i\epsilon f(-i\nabla)\psi(\mathbf{x})$. In addition to the class (5.34), any Hermitian operator constructed out of a_0 and a_0^\dagger commutes with the Hamiltonian, and thus represents a conserved charge as well. Not all of these, however, descend from a local Noether current. Two notable exceptions are the two possible Hermitian operators linear in a_0 and a_0^\dagger . I will list them alongside the symmetry transformations they generate:

$$\begin{aligned} Q_R &\equiv i \int d^d \mathbf{x} [\psi(\mathbf{x}) - \psi^\dagger(\mathbf{x})] = i\sqrt{V}(a_0 - a_0^\dagger) , & \psi &\rightarrow \psi - \epsilon , \\ Q_I &\equiv \int d^d \mathbf{x} [\psi(\mathbf{x}) + \psi^\dagger(\mathbf{x})] = \sqrt{V}(a_0 + a_0^\dagger) , & \psi &\rightarrow \psi - i\epsilon . \end{aligned} \quad (5.35)$$

With all the symmetry transformations at hand, the associated Noether currents can be extracted from the Lagrangian (5.29) using the technique outlined in Sect. 4.2.1.

The Lie algebra of our conserved charges is defined by the commutators

$$[Q_f, Q_R] = -if(\mathbf{0})Q_I , \quad [Q_f, Q_I] = +if(\mathbf{0})Q_R , \quad [Q_R, Q_I] = 2iV . \quad (5.36)$$

These make it natural to split the space of charges Q_f into the one-dimensional subspace of charges proportional to Q_1 , and a subspace of charges with $f(\mathbf{0}) = 0$. The latter commute with Q_R, Q_I and moreover annihilate any ground state $|\Omega\rangle$ as defined by (5.33). We can therefore drop them and focus on the set Q_1, Q_R, Q_I . By (5.36), this set furnishes a central extension of the Lie algebra $\mathfrak{iso}(2)$, corresponding to Euclidean transformations in the complex plane of ψ . The generator Q_1 induces (phase) rotations, whereas Q_R, Q_I generate respectively translations of the real and imaginary parts of ψ .

The Fock vacuum $|0\rangle$ is annihilated by Q_1 , but not by Q_R, Q_I . This suggests that alternative vacuum states may be obtained from $|0\rangle$ by transformations generated by Q_R, Q_I . We can pick any $z \equiv z_1 + iz_2 \in \mathbb{C}$ and define a new vacuum state by

$$|z\rangle \equiv \exp[i(z_1 Q_R + z_2 Q_I)] |0\rangle = \exp[\sqrt{V}(za_0^\dagger - z^* a_0)] |0\rangle . \quad (5.37)$$

This is a coherent state in the Fock space of zero-energy states. Indeed, it is easily seen to be an eigenstate of the annihilation operator, $a_0 |z\rangle = \sqrt{V}z |z\rangle$. In fact, $|z\rangle$ is simultaneously an eigenstate of the field operator at any point in space, $\psi(\mathbf{x}) |z\rangle = z |z\rangle$. We thus end up with a continuum of vacuum states, labeled by a complex parameter z . The order parameter distinguishing these states from each other is provided by the expectation value of $\psi(\mathbf{x})$. The vacuum manifold is the complex plane \mathbb{C} . For any choice of z , the state $|z\rangle$ possesses an unbroken $\text{SO}(2)$ symmetry, corresponding to rotations of the complex plane around the point z . Its generator is ${}^zQ_1 \equiv Q_1 + z_2 Q_R - z_1 Q_I$. The state $|z\rangle$ is an eigenstate of this operator, ${}^zQ_1 |z\rangle = -V|z|^2 |z\rangle$.

Interestingly, it is not possible to select the state $|z\rangle$ by adding to the Hamiltonian a perturbation linear in ψ as I did in Sect. 5.2. That would make the Hamiltonian unbounded from below. This is the price to pay for the vacuum manifold not being compact. An alternative is to add to the Hamiltonian a chemical potential μ for the conserved charge zQ_1 , so that it becomes $H_\mu \equiv H - \mu {}^zQ_1$. The expectation value of such a perturbed Hamiltonian in the state $|z'\rangle$ with any $z' \in \mathbb{C}$ is

$$\langle z' | H_\mu | z' \rangle = -V\mu [|z - z'|^2 - |z|^2] . \quad (5.38)$$

An infinitesimally small *negative* μ will then select $|z\rangle$ as the unique ground state with energy $V\mu |z|^2$.

5.3.2 Pathologies of the Infinite-Volume Limit

The analysis of the free Schrödinger theory has been perfectly clean so far. Most of the properties of the quantized theory match their classical counterparts, in particular the existence of infinitely many conservation laws, and infinitely many degenerate ground states. What happens when we try to take the limit of infinite volume? Let us evaluate the overlap of two different coherent states of the type (5.37),

$$|\langle z' | z \rangle|^2 = |\langle z' | U(z) | 0 \rangle|^2 = \exp(-V |z - z'|^2) , \quad (5.39)$$

$$U(z) \equiv \exp[i(z_1 Q_R + z_2 Q_I)] .$$

We therefore find that

$$\lim_{V \rightarrow \infty} |\langle z' | z \rangle|^2 = \lim_{V \rightarrow \infty} |\langle z' | U(z) | 0 \rangle|^2 = \delta_{zz'} . \quad (5.40)$$

In the limit $V \rightarrow \infty$, any two states $|z\rangle$ and $|z'\rangle$ with $z \neq z'$ become orthogonal. It is easy to generalize the above calculation to show that, in fact, any vector in the Fock space built above $|z\rangle$ must be orthogonal to $|z'\rangle$ in this limit.¹ In other words, the ket-vector $|z'\rangle$ does not lie in the Fock space built above $|z\rangle$ at all. Moreover, the

¹ Recall that the basis of the Fock space consists of states obtained by acting with a *finite* number of creation operators on the vacuum.

matrix element of any observable constructed using a finite number of annihilation and creation operators between the $|z\rangle$ and $|z'\rangle$ states vanishes. Altogether, the states $|z\rangle$ with $z \in \mathbb{C}$ therefore give rise to a completely disconnected Fock space each. Transitions between Fock spaces with different vacuum states are forbidden. This is certainly not what we expected!

This singular behavior of the $V \rightarrow \infty$ limit is easy to understand. Just remember that $|z\rangle$ is an eigenstate of a_0 with the eigenvalue $\sqrt{V}z$. Thus, if we insist on keeping z fixed, then scaling V towards infinity makes the eigenvalue diverge as well. This is ultimately the reason why $|z'\rangle$ does not lie in the Fock space built above $|z\rangle$. We might, on the other hand, try to keep the eigenvalue of a_0 fixed, which would require scaling z as $1/\sqrt{V}$ as the volume grows. Roughly speaking, this means that in a large but finite volume V , transitions between Fock spaces with different z are allowed, but only within a range $|z - z'| \lesssim 1/\sqrt{V}$.

There is another way of looking at this, emphasizing the properties of the operator $U(z)$ defined by (5.39). Since the states $|0\rangle$ and $|z'\rangle$ do not lie in the same Fock space, the operator $U(z)$ becomes ill-defined in the limit $V \rightarrow \infty$. Spontaneously broken symmetry may not be realized by unitary operators on the Hilbert space of the physical system! That however does not mean that in presence of SSB, it makes no sense to speak of symmetry transformations that are spontaneously broken. The transformation of the field operator by $U(z)$,

$$U(z)^\dagger \psi(\mathbf{x}) U(z) = \psi(\mathbf{x}) + z, \quad (5.41)$$

is perfectly well-defined even in the limit $V \rightarrow \infty$. The same applies to any local operator built out of a finite number of ψ and ψ^\dagger . In each of the Fock spaces built above any $|z\rangle$, we thus have a well-defined notion of broken symmetry transformations in terms of averages of observables. We just have to be careful not to think of broken symmetry as an operator connecting different *states* in the same space.

The distinction between the realizations of broken symmetries by operators on states and by transformations of observables is not a mere mathematical curiosity one can safely ignore. It has observable consequences. One of the earliest applications of group theory to physics was to classify energy levels of a quantum system in terms of representations of its symmetry group. In quantum mechanics, the so-called Wigner theorem guarantees that any symmetry is represented by a unitary or antiunitary operator on the Hilbert space (see Chap. 2 of [13]). Energy levels are therefore classified by irreducible representations of the symmetry group G . In quantum field theory, on the other hand, only the unbroken subgroup H is as a rule realized by unitary operators on the Hilbert space. Thus, energy levels form irreducible multiplets of H . SSB implies degeneracy of the ground state, accompanied by a reduced degeneracy of the excitation spectrum.

For example, the spectrum of hadrons is organized into (approximately degenerate) multiplets of the isospin group $SU(2)$. We know from Example 5.5 that this is the unbroken subgroup of QCD with two light quark flavors. There

is an even higher degeneracy, corresponding to multiplets of $SU(3)$. This is however less accurate due to the large splitting between the masses of the up, down and strange quarks. There are no multiplets of the $SU(n_f)_L \times SU(n_f)_R$ symmetry of QCD. Such multiplets would be easily noticeable. For instance, the pseudoscalar mesons such as pions or kaons would have to appear in the spectrum together with particles of opposite parity. The lack of such particles is a directly observable manifestation of the nonexistence of unitary operators representing spontaneously broken symmetry.

5.3.3 Uniqueness of the Finite-Volume Ground State

The above excursion into the subtleties of SSB was purposefully limited to a single, exactly solvable example. The conclusion that in infinite volume, spontaneously broken symmetry is not realized by unitary operators on the Hilbert space is, however, general [14]. Likewise, it is generally true that different symmetry-breaking vacuum states belong to disconnected Hilbert spaces.

One property of the free Schrödinger theory (5.29) that is not general is the existence of exactly degenerate ground states already in a finite volume. This is typical for systems where the order parameter is constructed using an operator that commutes with the Hamiltonian. In our case, this operator can be taken as the spatial average of the Schrödinger field,

$$\bar{\psi} \equiv \frac{1}{V} \int d^d \mathbf{x} \psi(\mathbf{x}) = \frac{1}{\sqrt{V}} a_0 . \quad (5.42)$$

Another system of this type is the ferromagnet, where the order parameter is supplied by a selected component of total spin. Historically, the very special properties of such systems led to a controversy (involving some famous individuals [15]) as to whether they should count as SSB at all or not.

An alternative is that in any finite volume, the system possesses a unique ground state, which is then necessarily G -invariant. Above this ground state, there is a tower of closely spaced energy levels, which all become exactly degenerate in the limit $V \rightarrow \infty$. A notable, much studied example of such a system are antiferromagnets. The realization of SSB is then quite nontrivial. Although a detailed discussion would be beyond the scope of this book, let me make at least a few remarks. The reader is encouraged to consult Sect. 19.1 of [16] or Sect. 2 of [1] for further information.

If one tries to take the naive limit $V \rightarrow \infty$, the unique, finite-volume G -invariant ground state turns into a likewise G -invariant state in infinite volume. The latter is, however, unstable under infinitesimally small external perturbations. The instability has its precursor already in a finite volume. Namely, the magnitude of a perturbation needed to drive the system away from the G -invariant vacuum scales as the inverse of the system's volume. Such a perturbation may be supplied in particular by the

measurement of any symmetry-breaking observable. This makes the system relax into one of the pure symmetry-breaking states constructed in Sect. 5.2.2.

Another way to understand the distinction between the various would-be ground states is through correlations of spatially separated observables. For any two local observables $A(\mathbf{x})$ and $B(\mathbf{y})$, one may construct the spatial correlation function $\langle A(\mathbf{x})B(\mathbf{y}) \rangle$; the angular brackets indicate expectation value in a chosen ground state. The so-called *cluster decomposition principle* dictates that

$$\lim_{|\mathbf{x}-\mathbf{y}| \rightarrow \infty} [\langle A(\mathbf{x})B(\mathbf{y}) \rangle - \langle A(\mathbf{x}) \rangle \langle B(\mathbf{y}) \rangle] = 0 . \quad (5.43)$$

It encodes the fundamental requirement that the outcomes of distant measurements should not be correlated with each other. The naive G -invariant ground state obtained by taking the $V \rightarrow \infty$ limit violates this principle, and thus displays long-range entanglement. The entanglement is destroyed by any measurement of a local symmetry-breaking observable, which brings the system to a stable, symmetry-breaking ground state satisfying (5.43).

To illustrate some of the above general observations, I will conclude the discussion of SSB by working out one more simple toy model. We return to the free Schrödinger theory, but twist the boundary condition to

$$\psi(\mathbf{x} + \mathbf{e}_r L) = e^{i\theta_r} \psi(\mathbf{x}) . \quad (5.44)$$

Here \mathbf{e}_r is a unit vector along the r -th Cartesian coordinate axis, and $\theta_r \in (-\pi, +\pi]$ is an arbitrarily chosen phase. I assume that at least one of θ_r is nonzero, otherwise (5.44) would be just the periodic boundary condition we used previously.

The quantization of the theory proceeds as before, except for the allowed values of momenta, where (5.30) is replaced with

$$p_r = \frac{\theta_r}{L} + \frac{2\pi n_r}{L} , \quad n_r \in \mathbb{Z} . \quad (5.45)$$

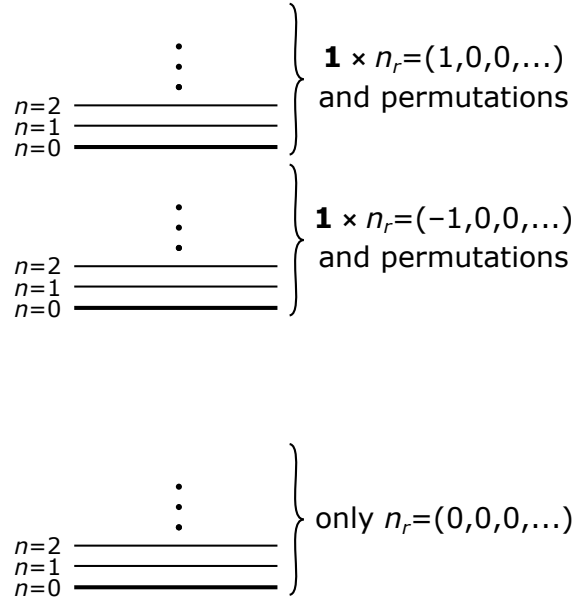
The Hamiltonian (5.32) now has a unique ground state, namely the Fock vacuum $|0\rangle$. The lowest-lying excitation of this ground state corresponds to $n_r = 0$ for all r , that is, $\mathbf{p} = \boldsymbol{\theta}/L$. This is a one-particle state $|\boldsymbol{\theta}/L\rangle \equiv a_{\boldsymbol{\theta}/L}^\dagger |0\rangle$ with excitation energy

$$E_{\boldsymbol{\theta}/L} = \frac{\boldsymbol{\theta}^2}{2mL^2} . \quad (5.46)$$

As long as all the θ_r are sufficiently small, the low end of the excitation spectrum will consist entirely of states with $n \in \mathbb{N}$ quanta of momentum $\boldsymbol{\theta}/L$ and excitation energy $nE_{\boldsymbol{\theta}/L}$. Only above these will one find states for which some of the n_r are nonzero. See Fig. 5.3 for a visualization of the spectrum of energy levels.

How is the Fock vacuum $|0\rangle$ affected by perturbations? Since the $\mathbf{p} = \mathbf{0}$ state is excluded by the boundary condition, we can now add a perturbation linear in the field without running into an instability. Let us therefore add to (5.29) the term

Fig. 5.3 Schematic visualization of the spectrum of the free Schrödinger theory with the twisted boundary condition (5.44), for small (positive) θ_r . Each of the indicated groups of energy levels consists of states with multiple quanta of momentum θ/L ; the number of such quanta is denoted as n . The lowest-lying multiplet includes no other particles. The higher multiplets include one or more particles where some of the n_r in (5.45) are nonzero. For simplicity, all the θ_r are assumed to be equal; otherwise the spectrum would feature additional structure



$$\mathcal{L}_\epsilon[\psi, \mathbf{x}](x) \equiv \frac{\epsilon^*}{\sqrt{V}} \psi(x) e^{-i\theta \cdot \mathbf{x}/L} + \frac{\epsilon}{\sqrt{V}} \psi^\dagger(x) e^{+i\theta \cdot \mathbf{x}/L}, \quad (5.47)$$

where ϵ is a small constant parameter with the dimension of energy. The exponential, coordinate-dependent factors are added in order that \mathcal{L}_ϵ itself satisfies a periodic boundary condition. The corresponding correction to the Hamiltonian (5.32) is

$$H_\epsilon = -(\epsilon^* a_{\theta/L} + \epsilon a_{\theta/L}^\dagger). \quad (5.48)$$

This perturbation will change the energy eigenstates within the multiplets shown in Fig. 5.3. It will however not lead to any mixing between states from different multiplets, that is states with different n_r . Let us focus on the lowest-lying multiplet, which is where the ground state resides. Here the Hamiltonian is projected down to

$$\begin{aligned} H_{\theta/L} &\equiv E_{\theta/L} a_{\theta/L}^\dagger a_{\theta/L} - (\epsilon^* a_{\theta/L} + \epsilon a_{\theta/L}^\dagger) \\ &= E_{\theta/L} [(a_{\theta/L}^\dagger - z_\epsilon^*)(a_{\theta/L} - z_\epsilon) - |z_\epsilon|^2], \end{aligned} \quad (5.49)$$

where $z_\epsilon \equiv \epsilon/E_{\theta/L}$. There is still a unique ground state, but it is now the coherent state $|z_\epsilon\rangle$. The perturbation has lowered the energy compared to the Fock vacuum $|0\rangle$ to $-E_{\theta/L} |z_\epsilon|^2 = -|\epsilon|^2/E_{\theta/L}$.

Let us see what we have found. The original vacuum $|0\rangle$ is nondegenerate. Its separation from the nearest excited states is however only of the order of $E_{\theta/L} \sim 1/L^2$. In the limit of infinite volume, all the different multiplets of states shown in Fig. 5.3 become degenerate. Accordingly, the Fock vacuum becomes unstable under small perturbations. For small fixed θ_r , one can always find ϵ that satisfies the bounds

$$1 \ll \frac{|\epsilon|}{E_{\theta/L}} \ll \frac{2\pi}{|\theta|}. \quad (5.50)$$

The lower bound makes the change in energy caused by the perturbation much larger than the level spacing within a given multiplet. Thus, even a small perturbation can drive the ground state from $|0\rangle$ to a coherent state $|z_\epsilon\rangle$ with large z_ϵ . The upper bound in (5.50) ensures that the shift of energy eigenvalues induced by the perturbation is much smaller than the separation of the different multiplets in the spectrum.

The free Schrödinger theory with the twisted boundary condition (5.44) serves as a good example of the instability of the finite-volume ground state with respect to small perturbations. Likewise, it provides a decent illustration of the structure of the spectrum of systems with SSB in a finite volume. One important feature that it does not illustrate is the G -invariance of the finite-volume ground state. This is because the symmetries (5.35) of the free Schrödinger theory are broken by the boundary condition (5.44) for any nonzero θ_r . In fact, insofar as the expectation value $\langle\psi(x)\rangle$ is well-defined, shift symmetries of the type (5.35) must always be spontaneously broken.

References

1. A.J. Beekman, L. Rademaker, J. van Wezel, *SciPost Phys. Lect. Notes* **11** (2019). DOI [10.21468/SciPostPhysLectNotes.11](https://doi.org/10.21468/SciPostPhysLectNotes.11)
2. F. Strocchi, *Symmetry Breaking*. Theoretical and Mathematical Physics (Springer Nature Switzerland AG, 2021). DOI [10.1007/978-3-662-62166-0](https://doi.org/10.1007/978-3-662-62166-0)
3. H. Goldstein, C.P. Poole, J.L. Safko, *Classical Mechanics* (Pearson Education Limited, Harlow, UK, 2013)
4. L.E. Ballentine, *Quantum Mechanics: A Modern Development* (World Scientific, Singapore, 1998)
5. J.L. Basdevant, J. Dalibard, *Quantum Mechanics* (Springer-Verlag, Berlin, Heidelberg, 2002)
6. R. Balian, *From Microphysics to Macrophysics*, vol. I (Springer, Berlin, Heidelberg, 2007)
7. A.O. Barut, R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific Publishers, Warszawa, Poland, 1977)
8. P.M. Chaikin, T.C. Lubensky, *Principles of condensed matter physics* (Cambridge University Press, Cambridge, UK, 1995)
9. G. Mussardo, *Statistical Field Theory* (Oxford University Press, Oxford, UK, 2010)
10. T. Koma, H. Tasaki, *J. Stat. Phys.* **76**, 745 (1994). DOI [10.1007/BF02188685](https://doi.org/10.1007/BF02188685)
11. H. Tasaki, *Physics and Mathematics of Quantum Many-Body Systems* (Springer Nature Switzerland, 2020). DOI [10.1007/978-3-030-41265-4](https://doi.org/10.1007/978-3-030-41265-4)
12. T. Brauner, *Symmetry* **2**, 609 (2010). DOI [10.3390/sym2020609](https://doi.org/10.3390/sym2020609)
13. S. Weinberg, *The Quantum Theory of Fields*, vol. I (Cambridge University Press, Cambridge, UK, 1995)
14. G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, in *Advances in Particle Physics*, vol. II, ed. by R.L. Cool, R.E. Marshak (Wiley, New York, NY, USA, 1968), pp. 567–708
15. R. Peierls, T.A. Kaplan, P.W. Anderson, *Phys. Today* **44**, 13 (1991). DOI [10.1063/1.2809982](https://doi.org/10.1063/1.2809982)
16. S. Weinberg, *The Quantum Theory of Fields*, vol. II (Cambridge University Press, Cambridge, UK, 1996)

Chapter 6

Nambu–Goldstone Bosons

In Chaps. 2 and 3, we saw examples of how *spontaneous symmetry breaking* (SSB) gives rise to massless particles in the spectrum. Our operational understanding of these *Nambu–Goldstone* (NG) *bosons* was based on a classical (tree-level) analysis of certain scalar field potentials. The purpose of this chapter is to develop a general understanding of NG bosons without the limitations of a particular model or approximation. The basic intuition is built in Sect. 6.1. This explains both where NG bosons come from, and how their spectrum is related to the nature of broken symmetry. The subsequent sections put this intuition on a solid footing. In Sect. 6.2 I prove the so-called Goldstone theorem, which asserts the very existence of a NG boson as a consequence of SSB. Section 6.3 then delves into the question how many NG bosons there are in a given system, and how their number relates to their dispersion relations. This generalizes the observations made in Chap. 3; see Sect. 3.3 for a quick summary. Altogether, this chapter completes the background needed in Parts III and IV, where the *effective field theory* (EFT) formalism for SSB is developed in detail.

6.1 Intuitive Picture

The intuitive understanding of NG bosons relies heavily on the concept of order parameter. Consider for simplicity a system whose ground state is spatially uniform, and imagine that we disturb it by a weak, local perturbation. Locally, it is energetically favorable for the system to remain in one of the degenerate ground states. The order parameter will therefore respond to the perturbation by developing spatial variation while remaining on the vacuum manifold everywhere. The energy cost of creating such a spatially varying excited state must be proportional to *gradients* of the order parameter. This is because changing the order parameter uniformly just amounts to a different choice of ground state. We can then imagine order parameter configurations that vary over progressively longer and longer length scales. In the limit that the scale of spatial variation goes to infinity, the gradient of the order parameter vanishes, and the energy cost must go to zero.

This is a completely general observation, valid regardless of the choice of field variables and dynamics behind SSB. The only assumption that is really necessary is that the spontaneously broken symmetry is continuous. This makes smooth spatial variation of the order parameter possible while keeping it on the vacuum manifold everywhere. We conclude that NG bosons are local, propagating fluctuations of the order parameter whose energy goes to zero in the infinite-wavelength limit.

The identification of NG bosons with fluctuations of the order parameter is not necessarily one-to-one. Eventually, we would like to know the precise spectrum of NG bosons; this among others governs the low-temperature thermodynamics of systems with SSB. What we really need in order to establish a dictionary between broken symmetries and NG bosons is to answer the following questions:

- How many different types of order parameter fluctuations (NG fields) are there in a given system?
- What is the correspondence between the various order parameter fluctuations (NG fields) and NG modes in the spectrum?

I will deal with these two questions in the given order in the next two subsections.

6.1.1 Redundancy of Order Parameter Fluctuations

I will build upon the intuitive picture of a NG boson in terms of a fluctuation of the order parameter that locally remains on the vacuum manifold. One can imagine such a fluctuation as being generated from the ground state by a broken symmetry transformation with a coordinate-dependent parameter. This leads immediately to the important observation that fluctuations induced by different broken symmetry generators may coincide.

Example 6.1

The Lagrangian of a free massless relativistic scalar field ϕ is $\mathcal{L}[\phi] = (1/2)(\partial_\mu \phi)^2$. The action of this theory is invariant under the polynomial shift transformation $\phi(x) \rightarrow \phi(x) + \epsilon_1 + \epsilon_{2\mu} x^\mu$, where $\epsilon_1, \epsilon_{2\mu}$ are constant parameters. This symmetry is necessarily spontaneously broken, and the order parameter can be chosen as $\langle \phi(x) \rangle$. One might expect to obtain $D + 1$ different fluctuations of the order parameter by applying a polynomial shift with coordinate-dependent parameters $\epsilon_1(x), \epsilon_{2\mu}(x)$. However, shifting $\phi(x)$ by $\epsilon_{2\mu}(x)x^\mu$ is identical to shifting it by $\epsilon_1(x)$ if we set $\epsilon_1(x) \equiv \epsilon_{2\mu}(x)x^\mu$. The set of $D + 1$ broken generators associated with our polynomial shift therefore corresponds to a single independent fluctuation, induced by $\phi(x) \rightarrow \phi(x) + \epsilon_1(x)$. The only independent NG field in the theory is ϕ itself.

Symmetries that become equivalent once their parameters are made coordinate-dependent are called *redundant*. Such a redundancy is typical for spacetime symmetries. As Example 6.1 shows, however, it is also possible for point symmetries that

do not affect spacetime coordinates. Here is a less trivial example, relevant for any crystalline phase of matter.

Example 6.2

In relativistic theories of scalar fields, the generators of spacetime rotations $J^{\mu\nu}$ and translations P^μ are known to satisfy the relation $J^{\mu\nu} = x^\mu P^\nu - x^\nu P^\mu$. A local rotation $\exp[(i/2)\epsilon_{2\mu\nu}(x)J^{\mu\nu}]$ with antisymmetric matrix parameter $\epsilon_{2\mu\nu}$ is thus equivalent to a local translation $\exp[i\epsilon_{1\mu}(x)P^\mu]$ with $\epsilon_{1\mu}(x) \equiv -\epsilon_{2\mu\nu}(x)x^\nu$. Imagine now a system where spacetime rotations and translations (or a subset thereof) are both spontaneously broken. For example, in crystalline solids, all the continuous spatial rotations and translations are spontaneously broken. The rotations are clearly redundant; the only independent NG fields are those associated with broken translations. These parameterize the vibrations of the crystal lattice.

The basic moral to remember is that the number of independent order parameter fluctuations may be lower than the number of broken symmetry generators. I will further refine this observation in Sect. 6.3.1. The discussion of redundancy of local symmetry transformations will be of great importance for the development of EFT for broken spacetime symmetries in Part IV.

6.1.2 Canonical Conjugation of Nambu–Goldstone Fields

The number of NG modes in the spectrum may be lower than the number of independent NG fields in case some of the latter are canonically conjugated. We saw this in Sect. 3.2. To make statements independent of a specific model, we need to relate the possibility of canonical conjugation of NG fields directly to the broken symmetry.

Suppose we have two NG fields, π^1 and π^2 , associated with two broken symmetry generators $Q_{1,2}$. We can assume without loss of generality that under the respective transformations generated by $Q_{1,2}$, these fields transform as

$$\pi^a \rightarrow \pi^a + \epsilon^b [\delta_b^a + O(\pi)] , \quad a = 1, 2 . \quad (6.1)$$

Indeed, one can always choose the fields so that the ground state corresponds to $\pi^a = 0$. That the leading, constant piece of the Taylor expansion of the transformation rule for π^a around this ground state is nonzero, follows from the assumption of broken symmetry. The bases of NG fields and broken generators can then be aligned so that this constant piece is simply $\pi^a \rightarrow \pi^a + \epsilon^a$.

Suppose in addition that we can construct a low-energy EFT for the NG fields whose Lagrangian contains a term with a single time derivative,

$$\mathcal{L}_{\text{eff}} = \varrho_{12} \pi^1 \partial_0 \pi^2 + \dots . \quad (6.2)$$

The ellipsis stands for terms with more than one derivative or more than two NG fields. Recall from Sect. 4.2.1 that the Noether currents corresponding to (6.1) can be identified by making the parameters ϵ^a coordinate-dependent. From (6.2) alone, one can then extract the leading contributions to the Noether charge densities,

$$J_1^0[\pi] = -\varrho_{12}\pi^2 + \cdots, \quad J_2^0[\pi] = +\varrho_{12}\pi^1 + \cdots. \quad (6.3)$$

The ellipses now represent terms with derivatives or more than one NG field.

Here comes the key step: I will evaluate the transformation of $J_2^0[\pi]$ under the symmetry generated by Q_1 in two different ways. On the one hand, it is obvious that under (6.1), $J_2^0[\pi] \rightarrow J_2^0[\pi] + \varrho_{12}\epsilon^1 + \cdots$. On the other hand, one may think of the currents as quantum operators and represent the same transformation as

$$\exp(i\epsilon^1 Q_1) J_2^0 \exp(-i\epsilon^1 Q_1) = J_2^0 + i\epsilon^1 [Q_1, J_2^0] + O((\epsilon^1)^2). \quad (6.4)$$

The NG fields vanish in the ground state. Hence, upon taking the *vacuum expectation value* (VEV), a comparison of our two little calculations leads to

$$\varrho_{12} = i\langle [Q_1, J_2^0] \rangle = -i\langle [Q_2, J_1^0] \rangle, \quad (6.5)$$

where the second equality follows by running the same argument on $J_1^0[\pi]$ and Q_2 .

This is a remarkable result that is as model-independent as it gets. In any theory that realizes the same symmetry-breaking pattern, one can construct the Noether currents via Noether's theorem. Evaluating the VEV of the commutator (6.5) then gives us a simple criterion for when to expect two NG fields to be canonically conjugated: whenever (6.5) is nonzero.

Example 6.3

We already saw the relation between canonical conjugation and the charge commutator (6.5) at work in Sect. 5.3.1. Indeed, one can think of the free Schrödinger field ψ therein as corresponding to two real NG fields. These are in turn associated with the invariance of the free Schrödinger theory under independent shifts of the real and imaginary parts of ψ . The commutator of the generators of these shifts, Q_R and Q_I , turns out to be nonzero upon quantization. It is easy to check that the normalization of the commutator as shown in (5.36) agrees with (6.5). The factor of spatial volume V in (5.36) arises from the latter displaying a commutator of two charges rather than a charge and a charge density.

Although noninteracting, the free Schrödinger theory is nontrivial in that it makes the VEV (6.5) arise from a central charge in the symmetry algebra. Let us have a look at one more example, which is mathematically simpler and physically familiar.

Example 6.4

In the ideal (that is spatially uniform and isotropic) approximation, both ferromagnets and antiferromagnets possess a symmetry $G \simeq \text{SU}(2)$, generated by the operator \mathbf{S} of total spin. In the ground state of both, this is spontaneously broken down to $H \simeq \text{U}(1)$, consisting of spin rotations around the axis along which the spins are aligned. Suppose that one chooses the ground state oriented along the third direction in the spin space. Then H is generated by S_3 , whereas the two independent broken generators can be chosen as $S_{1,2}$.

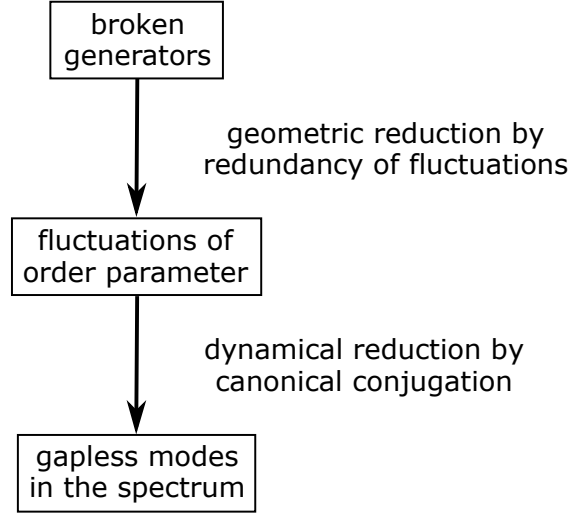
The difference between ferro- and antiferromagnets is that the former feature a nonzero net magnetization. This amounts to a nonzero VEV $\langle S_3 \rangle = -i\langle [S_1, S_2] \rangle$. As a consequence, the two NG fields in a ferromagnet are canonically conjugated. The spectrum only contains a single NG mode: the ferromagnetic *magnon*. Its dispersion relation at long wavelengths is known to be quadratic in momentum. In antiferromagnets, on the other hand, $\langle \mathbf{S} \rangle = \mathbf{0}$. As a consequence, the spectrum of antiferromagnets features two different magnon branches. Their dispersion relation is known to be linear in momentum in the long-wavelength limit.

6.1.3 The Big Picture

Let me summarize what we have learned so far. Spontaneous breaking of a continuous symmetry implies the existence of a gapless mode in the spectrum: the NG boson. This can be viewed as a propagating fluctuation of the order parameter. The number of such NG modes may be lower than the number of broken symmetry generators. There are two basic mechanisms how such a reduction may occur.

First, different broken symmetry generators may induce fluctuations of the order parameter that are indistinguishable from each other. I will refer to this mechanism as “geometric” to keep in mind that it represents a restriction on the configuration space of off-shell NG fields. Second, there may not be a one-to-one correspondence between the independent NG fields and the independent NG modes in the spectrum. This happens whenever a pair of NG fields is canonically conjugated, which in turn requires a nonzero VEV of the charge commutator (6.5). This mechanism I will refer to as “dynamical” since it restricts the space of on-shell fields, or eigenstates of the Hamiltonian. For the reader’s convenience, an outline of the correspondence between broken symmetry, NG fields and NG modes is shown graphically in Fig. 6.1. This scheme is intuitively simple yet somewhat imprecise. I will return to the question of how many NG modes there are in a given system in Sect. 6.3.

Fig. 6.1 Basic mechanisms that may cause the number of independent NG modes in the spectrum to be lower than the number of broken symmetry generators. The geometric reduction arises whenever fluctuations of the order parameter, induced by different generators, are indistinguishable from each other. The dynamical reduction is signaled by a nonzero VEV of the charge commutator (6.5), or by a term in the effective Lagrangian for NG fields with a single time derivative



6.2 Goldstone Theorem

The Goldstone theorem is one of the most profound exact results in quantum field theory that deserves a more careful justification than the intuitive but hand-waving argument of Sect. 6.1. In the next two subsections, I will give two different proofs, following in spirit the original work of Goldstone, Salam and Weinberg [1, 2]. Along the way, I will clarify the technical assumptions on which the theorem is based. A reader seeking a higher level of mathematical rigor than what I can offer here is advised to consult the early review [3] or the book [4].

6.2.1 Operator Proof

I will start with a proof using the operator formalism of quantum field theory in the Heisenberg picture. Before proceeding to the details of the proof, however, I need to spend some time listing its technical assumptions.

First, I assume that in the limit of infinite spatial volume, the system possesses unbroken continuous (spacetime) translation invariance. This ensures the existence of a basis of the Hilbert space consisting of eigenstates of the energy–momentum operator P^μ . It is always possible, and I will implicitly do so, to define this operator so that the energy and momentum of the ground state are both zero. The assumption of continuous spatial translation invariance is, in fact, unnecessarily strong. The minimum requirement is unbroken discrete translation invariance, which is necessary to have a well-defined notion of quasiparticles. The proof then proceeds along the same steps at the cost of a somewhat cluttered notation [5].

Second, I assume the existence of a local operator $J^0(x)$, interpreted as the density of a conserved charge. The charge density should not depend explicitly on spacetime coordinates; this is needed to ensure the translation property

$$J^0(x) = e^{iP \cdot x} J^0(0) e^{-iP \cdot x} . \quad (6.6)$$

Now define the total charge contained in a finite spatial domain Ω ,

$$Q_\Omega(t) \equiv \int_\Omega d^d \mathbf{x} J^0(\mathbf{x}, t) . \quad (6.7)$$

The last assumption we need is the existence of a local, time-independent operator Φ such that the VEV in the chosen ground state $|0\rangle$,

$$\langle 0 | [Q_\Omega(t), \Phi] | 0 \rangle , \quad (6.8)$$

is nonzero and time-independent in the limit $\Omega \rightarrow \infty$.

It is this last assumption that captures the essence of SSB. Think of Q_Ω as the Noether charge associated with a continuous symmetry. One would be tempted to say that in the limit $\Omega \rightarrow \infty$, Q_Ω itself becomes time-independent. But we saw in Sect. 5.3 that in the infinite-volume limit, the operators representing spontaneously broken symmetry may be ill-defined. The limit $\Omega \rightarrow \infty$ is only safe if performed on the commutator in (6.8), not on Q_Ω itself.

Usually, the time-independence of the integral charge is a consequence of a local conservation law, $\partial_\mu J^\mu = 0$. That is however strictly speaking not necessary. Taking this additional step requires the vanishing of the surface integral $\int_\Omega d^d \mathbf{x} \partial_r J^r(\mathbf{x}, t)$ as $\Omega \rightarrow \infty$. This condition, or even its weaker form imposed only on the VEV $\langle 0 | [\partial_r J^r(\mathbf{x}, t), \Phi] | 0 \rangle$, may be compromised in systems with long-range interactions [3]. Here I avoid having to deal with this issue simply by assuming right away the time-independence of (6.8).

Think of $\langle 0 | \Phi | 0 \rangle$ as an order parameter. Then (6.8) is, up to a factor, its infinitesimal variation under the transformation $\Phi \rightarrow \exp(i\epsilon Q_\Omega) \Phi \exp(-i\epsilon Q_\Omega)$. The assumption that (6.8) is nonzero guarantees that $|0\rangle$ cannot be an eigenstate of Q_Ω . Hence the symmetry generated by Q_Ω must be spontaneously broken.

We are now ready to prove the existence of a NG mode in the system. I will initially assume that the system is enclosed in a finite volume V with a periodic boundary condition. The Hilbert space then admits a basis of momentum eigenstates $|n, \mathbf{p}\rangle$, where \mathbf{p} stands for a set of discrete momenta consistent with the boundary condition. In a finite volume, the basis states can be normalized as $\langle m, \mathbf{p} | n, \mathbf{q} \rangle = \delta_{mn} \delta_{\mathbf{p}\mathbf{q}}$. The label n indicates all other quantum numbers the states may possess, such as relative momenta in multiparticle states, or internal degrees of freedom.

Upon inserting the partition of unity in terms of $|n, \mathbf{p}\rangle$ and using the translation property (6.6), the VEV (6.8) can be rewritten as

$$\begin{aligned} \langle 0 | [Q_\Omega(t), \Phi] | 0 \rangle &= \sum_{n, \mathbf{p}} \int_\Omega d^d \mathbf{x} \left[\exp(-i \mathbf{p}_n \cdot \mathbf{x}) \langle 0 | J^0(0) | n, \mathbf{p} \rangle \langle n, \mathbf{p} | \Phi | 0 \rangle \right. \\ &\quad \left. - \exp(+i \mathbf{p}_n \cdot \mathbf{x}) \langle 0 | \Phi | n, \mathbf{p} \rangle \langle n, \mathbf{p} | J^0(0) | 0 \rangle \right]. \end{aligned} \quad (6.9)$$

Here p_n is a shorthand notation for the energy–momentum of $|n, \mathbf{p}\rangle$. Now recall that SSB implies the existence of degenerate ground states. Even if one picks a unique vacuum $|0\rangle$, the other, alternative ground states do not disappear. They contribute to (6.9) among the states $|n, \mathbf{0}\rangle$ with $\mathbf{p} = \mathbf{0}$. Is this something to worry about?

In the free Schrödinger theory, worked out in Sect. 5.3.1, $\langle 0 | J^0(0) | n, \mathbf{0} \rangle$ scales as $1/\sqrt{V}$, cf. (5.35). This observation turns out to be general. With the periodic boundary condition, the integral charge $Q_V(t)$ is translationally invariant. A classic argument [6] then shows that the norm of $Q_V(t) | 0 \rangle$ scales as \sqrt{V} ,

$$\|Q_V(t) | 0 \rangle\|^2 = \int_V d^d \mathbf{x} \langle 0 | J^0(\mathbf{x}, t) Q_V(t) | 0 \rangle = V \langle 0 | J^0(0) Q_V(0) | 0 \rangle. \quad (6.10)$$

This means, roughly, that the norm of $J^0(0) | 0 \rangle$ scales as $1/\sqrt{V}$, and so does therefore $\langle 0 | J^0(0) | n, \mathbf{0} \rangle$. Altogether, the contribution of the alternative ground states to (6.9) at fixed Ω and increasing V is suppressed as $1/\sqrt{V}$. Should (6.8) be nonzero in the infinite-volume limit, it must be dominated by excitations with nonzero momentum.

We can now switch to infinite volume. This requires changing the normalization of the basis states to $\langle m, \mathbf{p} | n, \mathbf{q} \rangle = (2\pi)^d \delta_{mn} \delta^d(\mathbf{p} - \mathbf{q})$ and replacing the sum over \mathbf{p} in (6.9) with an integral. Keeping Ω still finite, the integration over \mathbf{x} then amounts to a Fourier transform of unity, $\int_\Omega d^d \mathbf{x} e^{i \mathbf{p} \cdot \mathbf{x}} \equiv (2\pi)^d \delta_\Omega^d(\mathbf{p})$. The notation underlines that $\delta_\Omega^d(\mathbf{p})$ is a finite-volume approximation of the Dirac δ -function. We then have

$$\begin{aligned} \langle 0 | [Q_\Omega(t), \Phi] | 0 \rangle &= \sum_n \int d^d \mathbf{p} \left\{ \exp[-i E_n(\mathbf{p}) t] \delta_\Omega^d(\mathbf{p}) \langle 0 | J^0(0) | n, \mathbf{p} \rangle \right. \\ &\quad \left. \times \langle n, \mathbf{p} | \Phi | 0 \rangle - \exp[+i E_n(\mathbf{p}) t] \delta_\Omega^d(-\mathbf{p}) \langle 0 | \Phi | n, \mathbf{p} \rangle \langle n, \mathbf{p} | J^0(0) | 0 \rangle \right\}, \end{aligned} \quad (6.11)$$

where $E_n(\mathbf{p}) = p_n^0$ is the dispersion relation of the state $|n, \mathbf{p}\rangle$.

By our assumptions, (6.8) should be nonzero and time-independent in the limit $\Omega \rightarrow \infty$. For large Ω , the function $\delta_\Omega^d(\mathbf{p})$ is smooth but sharply peaked around $\mathbf{p} = \mathbf{0}$. The time-independence thus requires that only states such that

$$\lim_{\mathbf{p} \rightarrow \mathbf{0}} E_n(\mathbf{p}) = 0 \quad (6.12)$$

contribute to (6.11). The assumption that (6.8) should be nonzero guarantees that such states exist. These must be one-particle states, for if the one-particle spectrum had a gap, then so would the multiparticle one. We have thus established the existence of one-particle states $|n, \mathbf{p}\rangle$ with the property (6.12), for which both $\langle 0 | J^0(0) | n, \mathbf{p} \rangle$ and $\langle 0 | \Phi | n, \mathbf{p} \rangle$ are nonzero. This concludes the proof of Goldstone’s theorem.

There is a simple, intuitive way to understand why NG modes must satisfy (6.12). Suppose we know beforehand that there is a one-particle state $|n, \mathbf{p}\rangle$ created by J^0

and that the latter descends from a local conservation law, $\partial_\mu J^\mu = 0$. Extending the translation property (6.6) to J^μ , we get at once $\langle 0 | J^\mu(x) | n, \mathbf{p} \rangle = \exp(-i p_n \cdot x) \langle 0 | J^\mu(0) | n, \mathbf{p} \rangle$. Upon using current conservation and the fact that $\langle 0 | J^0(0) | n, \mathbf{p} \rangle$ is nonzero, it then follows that $E_n(\mathbf{p}) = \mathbf{p} \cdot \langle 0 | \mathbf{J}(0) | n, \mathbf{p} \rangle / \langle 0 | J^0(0) | n, \mathbf{p} \rangle$. This shows that the vanishing of the energy of NG modes in the long-wavelength limit is an immediate consequence of symmetry. The nontrivial part of Goldstone's theorem is that such NG states exist at all.

6.2.2 Effective Action Proof

I will now present an entirely different proof of Goldstone's theorem which will prepare the ground for the discussion of redundancies among NG fields in the next section. In spirit, this proof is not based on current conservation but rather directly on the symmetry of the action. I will follow closely the appendix of [7].

The starting assumption is the existence of an action Γ as a functional of a set of fields ψ^i . The notation indicates that Γ is the quantum effective action, since we want to make exact statements about a quantum field theory. The argument below applies in principle equally well to the classical action S . However, the fields ψ^i need not be the “elementary” fields that appear in S . They may be composite operators, and are chosen so that the VEV $\langle \psi^i \rangle$ is a suitable order parameter for SSB. The action is assumed to be invariant under the infinitesimal transformation

$$\delta \psi^i(x) = \epsilon F^i[\psi, x](x) . \quad (6.13)$$

Unlike in the previous, operator proof of Goldstone's theorem, the symmetry transformation, and thus the Noether current, is allowed to depend explicitly on coordinates.

It is not guaranteed a priori that the quantum effective action Γ inherits verbatim all the symmetries of the classical action S . It turns out that the symmetry transformations on the classical and quantum actions are the same at least if $F^i[\psi, x]$ is linear in the fields; see Sect. 16.4 of [8]. This is something I assumed already back in Sect. 5.2 when I introduced the very concept of order parameter. The only restriction involved in the transition to the quantum effective action therefore is that the symmetry should preserve the functional integral measure. This is needed to exclude quantum anomalies, whose presence would invalidate Goldstone's theorem.

The invariance of the action is encoded in the condition

$$\int d^D y \frac{\delta \Gamma}{\delta \psi^j(y)} F^j[\psi, y](y) = 0 . \quad (6.14)$$

Taking one more variational derivative and setting the fields equal to their VEVs, $\langle \psi^i(x) \rangle \equiv \varphi^i(x)$, one gets

$$\int d^D y \left. \frac{\delta^2 \Gamma}{\delta \psi^i(x) \delta \psi^j(y)} \right|_{\psi=\varphi} F^j[\varphi, y](y) = 0 . \quad (6.15)$$

Now we need the assumption of continuous translation invariance of the action and of the ground state. The second variational derivative of Γ can thus be traded for its Fourier transform, which is the inverse propagator of the theory, $G_{ij}^{-1}[\varphi](p)$. Multiplying (6.15) with $e^{ip \cdot x}$ and integrating over x then yields the condition

$$G_{ij}^{-1}[\varphi](p) \int d^D y e^{ip \cdot y} F^j[\varphi, y](y) = 0 . \quad (6.16)$$

Hence, the propagator $G_{ij}[\varphi](p)$ has a singularity at energy–momentum p^μ whenever the integral in (6.16) is nonzero. The spectral representation ensures that such singularities arise solely from states in the spectrum of the Hamiltonian.

By the assumption of broken symmetry, $F^j[\varphi, y]$, which represents the symmetry transformation of the order parameter, is nonzero. This alone implies that the integral in (6.16) must be nonzero for *some* p^μ . For coordinate-independent symmetries, the integral is proportional to $\delta^D(p)$. Moreover, for known examples of coordinate-dependent symmetries, $F^j[\varphi, y]$ is polynomial in coordinates. In such cases, integration over y leads to a linear combination of derivatives of $\delta^D(p)$. One way or another, we conclude that SSB implies the existence of a mode whose energy vanishes in the limit of zero momentum, as expected.

6.3 Classification and Counting

None of the above two proofs of the Goldstone theorem addresses the question how many NG bosons there are. To that end, we have to put in some extra effort. I will follow closely the scheme outlined in Sect. 6.1.3. This means that as the first step, we need to know, given the symmetry-breaking pattern, how many independent NG *fields*, or order parameter fluctuations, there are.

6.3.1 Independent Order Parameter Fluctuations

We add to (6.13) an index A , distinguishing the action of different symmetry generators Q_A , $\delta \psi^i(x) = \epsilon^A F_A^i[\psi, x](x)$. Recall the view of NG fields as local fluctuations of the order parameter. The easiest (albeit not the only; cf. Sect. 4.2.1) way to generate such fluctuations is by replacing $\epsilon^A \rightarrow \epsilon^A(x)$. Redundancies among order parameter fluctuations are then detected by the existence of nonzero functions $\epsilon^A(x)$ such that

$$\epsilon^A(x) F_A^i[\varphi, x](x) = 0. \quad (6.17)$$

This relation between symmetry transformations implies an analogous dependence between the zero modes of the inverse propagator of the quantum theory. Namely, trading the coordinate inside $\epsilon^A(x)$ for a derivative with respect to the momentum variable p^μ , denoted as ∇_μ , a combination of (6.16) and (6.17) gives

$$\epsilon^A(-i\nabla) \int d^D y e^{ip \cdot y} F_A^j[\varphi, y](y) = 0. \quad (6.18)$$

The conclusion is that the number of independent NG fields equals the dimension of the symmetry group G minus the number of linearly independent solutions of (6.17). Note that the linear independence is meant in the functional sense: multiplying all the $\epsilon^A(x)$ by the same function $f(x)$ does not count as a new solution.

The rule for counting the number of independent NG fields does not have an established form in the literature. My formulation is close in spirit to [9] where, however, the functions $\epsilon^A(x)$ were only allowed to depend on coordinates corresponding to unbroken continuous translations. That seems too restrictive for instance in case of crystalline solids where only a discrete translation invariance survives in the ground state.

Another condition for redundancy similar to (6.17) was put forward in [10]; see also the review [11]. Their criterion is however phrased in terms of the vacuum ket-vector and charge density operators. As such, it also includes the redundancy of NG states under canonical conjugation, which I treat separately in Sect. 6.3.2. Finally, in [12] a rule counting independent NG fields is formulated directly in terms of the zero modes of the inverse propagator.

In case of broken spacetime symmetries, the number of independent would-be NG fields identified with the help of (6.17) turns out to depend on the precise choice of order parameter(s). Remarkably, some of these fields may couple to gapped states in the spectrum. This subtlety is missed by the classical counting rule for NG fields based on (6.17). It is best addressed having the full EFT machinery at hand. I will do so in Chap. 13.

As a basic sanity check, consider spontaneously broken internal symmetries in systems with unbroken translation invariance. In that case, $F_A^i[\varphi, x]$ does not depend on coordinates, either explicitly or implicitly through the order parameter. Also, it is an ordinary local function of φ^i , hence the condition (6.17) reduces to $\epsilon^A(x) F_A^i(\varphi) = 0$. The only nontrivial solutions for $\epsilon^A(x)$ are then those corresponding to generators of the unbroken subgroup H . The number of independent NG fields therefore equals $\dim G - \dim H$. For spontaneously broken internal symmetries, there is a one-to-one correspondence between NG fields and broken symmetry generators. This is one of the moral lessons we reached in Chap. 3 by analyzing a set of specific models.

Oftentimes, the solutions to (6.17) descend from redundancy among local symmetry transformations acting on an arbitrary field configuration. This was certainly

the case for Example 6.1 and Example 6.2. The condition (6.17) is however much weaker, as it only demands redundancy of local symmetry transformations of the ground state. Here is a nontrivial example of this latter type.

Example 6.5

Helimagnets are magnetic materials in which the alignment axis of spins gradually changes with position; see [13] for basic phenomenology. The simplest type of helimagnetic state can be modeled by the unit-vector order parameter

$$\langle \mathbf{n}(\mathbf{x}, t) \rangle = (\cos kz, \sin kz, 0) , \quad (6.19)$$

where k is a positive constant. The local axis of alignment lies in the xy plane but rotates around the z axis, forming a helix with the pitch $2\pi/k$.

Suppose for simplicity that the material in which the helical order develops has a cubic crystal lattice. This is the case for instance for the inorganic compounds MnSi and FeGe; see Sect. 2 of [14] for a more comprehensive list of known helimagnetic materials. Then the low-energy physics of the material exhibits the following approximate continuous symmetries (in addition to time translations):

- Spatial translations (T): $\mathbf{n}'(\mathbf{x}', t) = \mathbf{n}(\mathbf{x}, t)$ with $\mathbf{x}' = \mathbf{x} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \in \mathbb{R}^3$.
- Space–spin rotations (R): $\mathbf{n}'(\mathbf{x}', t) = R\mathbf{n}(\mathbf{x}, t)$ with $\mathbf{x}' = R\mathbf{x}$, $R \in \text{SO}(3)$.

Unlike in ordinary (anti)ferromagnets, spin rotations cannot be considered separately from spatial rotations as a consequence of the spin-orbit coupling. This induces the so-called Dzyaloshinskii–Moriya interaction, which is eventually responsible for the development of the helical order (6.19).

In their evolutionary form, the above transformations correspond respectively to the functions ${}^T F_r^i = -\partial_r n^i$ and ${}^R F_r^i = -\varepsilon_{rs}^u x^s \partial_u n^i + \varepsilon_{rj}^i n^j$. Denoting the corresponding infinitesimal parameters as $\epsilon^r(\mathbf{x}, t)$ (translation in the r -th direction) and $\omega^r(\mathbf{x}, t)$ (rotation around the r -th axis), (6.17) turns into

$$\epsilon^r(\mathbf{x}, t) \partial_r \langle n^i(\mathbf{x}, t) \rangle + \omega^r(\mathbf{x}, t) [\varepsilon_{rs}^u x^s \partial_u \langle n^i(\mathbf{x}, t) \rangle - \varepsilon_{rj}^i \langle n^j(\mathbf{x}, t) \rangle] = 0 . \quad (6.20)$$

This boils down to the following constraints on the parameter functions,

$$\begin{aligned} \omega^1(\mathbf{x}, t) \sin kz &= \omega^2(\mathbf{x}, t) \cos kz , \\ \omega^3(\mathbf{x}, t) &= k[\epsilon^3(\mathbf{x}, t) + y\omega^1(\mathbf{x}, t) - x\omega^2(\mathbf{x}, t)] . \end{aligned} \quad (6.21)$$

Some of the solutions of (6.21) are trivial. First of all, ϵ^1 and ϵ^2 can be chosen arbitrarily since they do not appear in (6.21) at all. These represent unbroken translations in the x, y directions. Furthermore, there is another unbroken symmetry, corresponding to simultaneous translation in the z -direction and rotation around the z -axis. This extends to the solution of (6.21) with $\omega^3(\mathbf{x}, t) = k\epsilon^3(\mathbf{x}, t)$ and $\omega^{1,2}(\mathbf{x}, t) = 0$.

In addition to the three trivial solutions that identify unbroken symmetries, there is one solution of (6.21) that represents a genuine redundancy of order parameter fluctuations. This can be thought of for instance as picking ω^1 arbitrarily and ad-

justing ω^2 and ω^3 so that (6.21) is satisfied. Altogether, this makes four independent solutions of (6.17) in case of the order parameter (6.19). With six different symmetry generators, three translations and three rotations, this leaves us with mere two independent NG fields. That is a result we could have guessed: the unit vector field $\mathbf{n}(\mathbf{x}, t)$ contains exactly two degrees of freedom.

This example demonstrates that while possibly tedious, finding the solutions to (6.17) as a set of linear equations for $\epsilon^A(x)$ is completely straightforward. Even though (6.17) does not give an explicit expression for the number of independent NG fields, it therefore offers a streamlined algorithmic procedure how to find them.

6.3.2 Type-A and Type-B Nambu–Goldstone Bosons

Now we finally get to the question how many NG bosons there are in the spectrum. I will start with the simpler case of spontaneously broken internal symmetry in systems with unbroken translation invariance. In this case, there are guaranteed to be $\dim G - \dim H$ NG fields, in a one-to-one correspondence to broken symmetry generators. Finding the number of NG modes requires just a mild generalization of the argument in Sect. 6.1.2.

As the first step, one introduces the commutator matrix

$$\varrho_{AB} \equiv i\langle [Q_A, J_B^0] \rangle. \quad (6.22)$$

Translation invariance allows to rewrite the matrix as $\varrho_{AB} = i \lim_{V \rightarrow \infty} \langle [Q_A, Q_B] \rangle / V$. This makes it clear that $\varrho_{AB} = 0$ whenever Q_A or Q_B is unbroken. The nontrivial part of the commutator matrix therefore resides in its restriction to broken generators $Q_{a,b,\dots}$, that is ϱ_{ab} . Moreover, ϱ_{AB} is real antisymmetric, and as such can be block-diagonalized by an orthogonal change of basis of symmetry generators. With the shorthand notation $r_\varrho \equiv \text{rank } \varrho$, this implies that the part of the effective Lagrangian for NG fields π^a , linear in time derivatives, takes the form

$$\mathcal{L}_{\text{eff}} = \varrho_{12} \pi^1 \partial_0 \pi^2 + \cdots + \varrho_{r_\varrho-1, r_\varrho} \pi^{r_\varrho-1} \partial_0 \pi^{r_\varrho} + \cdots. \quad (6.23)$$

The ellipsis indicates operators with more than one derivative or two NG fields.

This allows a unique categorization of NG fields. The r_ϱ NG fields π^a with $a = 1, \dots, r_\varrho$ are canonically conjugated into $r_\varrho/2$ pairs. Each such pair corresponds to one *type-B NG boson*. Barring accidental cancellations in the spatial gradient part of the effective Lagrangian, the dispersion relation of type-B NG bosons is typically quadratic in momentum. The remaining NG fields π^a with $a = r_\varrho + 1, \dots, \dim G - \dim H$ do not appear in any bilinear operator with a single time derivative. They correspond to one *type-A NG boson* each. As a rule of thumb, free type-A bosons are described by bilinear operators in the Lagrangian with two temporal or two spatial derivatives. Hence, their dispersion relation is linear in momentum.

We have arrived at a simple counting rule for the respective numbers of type-A and type-B NG bosons [15, 16],

$$n_A = \dim G - \dim H - \text{rank } \varrho, \quad n_B = \frac{1}{2} \text{rank } \varrho. \quad (6.24)$$

This characterizes the spectrum of NG bosons based solely on the pattern of symmetry breaking and a bit of additional information about the ground state, encoded in the matrix ϱ_{AB} . The possible values of ϱ_{AB} are strongly restricted by the unbroken symmetry. To start with, in systems with unbroken Lorentz invariance, terms in the Lagrangian with a single time derivative are forbidden, hence $\varrho_{AB} = 0$. This means that $n_A = \dim G - \dim H$ and $n_B = 0$. All relativistic NG bosons are massless particles with a strictly linear dispersion relation.

Let us now relax the requirement of Lorentz invariance, but keep the assumption that only internal symmetry is spontaneously broken and that the system possesses continuous translation invariance. If in addition the internal symmetry group G is compact and semisimple, we can write

$$\varrho_{AB} = i \lim_{V \rightarrow \infty} \frac{\langle [Q_A, Q_B] \rangle}{V} = -f_{AB}^C \lim_{V \rightarrow \infty} \frac{\langle Q_C \rangle}{V}, \quad (6.25)$$

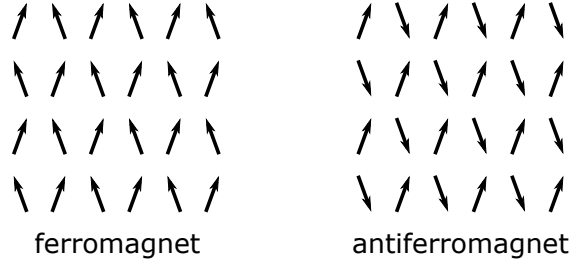
where f_{AB}^C are the structure constants of G . This shows that the VEV of the conserved charges themselves can serve as an order parameter for SSB. The set of VEVs $\langle Q_C \rangle$ furnishes a vector in the adjoint representation of G . At the same time, it must remain unchanged under the action of H . As a consequence, the vector $\langle Q_C \rangle$ must correspond to a trivial one-dimensional representation (singlet) of H . Finding such singlets in the decomposition of the Lie algebra \mathfrak{g} of G into irreducible multiplets of H is a quick way to isolate candidate order parameters.

Focusing on singlets of H in the decomposition of the adjoint representation of G may not be restrictive enough in case H is small or even trivial. We can however do much better than that. Namely, it is always possible to choose the ground state and basis of generators Q_A such that all the charges that have a nonzero VEV mutually commute. A detailed proof of this statement is given in the appendix of [7]. Thus, to identify conserved charge order parameters, it is sufficient to restrict to a Cartan subalgebra of \mathfrak{g} and find singlets of H therein. By the usual root decomposition of Lie algebras, all the generators of G are organized into pairs whose commutator lies in the Cartan subalgebra of \mathfrak{g} . This maps type-B NG bosons uniquely to pairs of mutually conjugate roots of \mathfrak{g} .

Example 6.6

Consider the symmetry-breaking pattern $\text{SU}(n) \rightarrow \text{U}(n-1)$ with integer $n \geq 2$. The unbroken subgroup $H \simeq \text{U}(n-1)$ can be realized explicitly for instance as the set of unitary block-diagonal matrices with blocks of sizes 1 and $n-1$, respectively. The Lie algebra $\mathfrak{g} \simeq \mathfrak{su}(n)$ contains a unique singlet of H , namely the generator of the $\text{U}(1)$ factor of H , $Q \equiv \text{diag}(-n+1, 1, \dots, 1)$.

Fig. 6.2 Schematic illustration of spin order in canted (anti)ferromagnets. In ferromagnets, the individual spins are slightly tilted away from perfect alignment. Antiferromagnets, on the other hand, exhibit overall antialignment of spins, tilted so that a nonzero net magnetization arises. In both cases, the whole $SU(2)$ group of spin rotations is spontaneously broken



The $2n - 2$ broken generators can be chosen as matrices with nonzero elements located in the first row and column. These span a complex $(n - 1)$ -plet of $SU(n - 1) \subset H$. Since all the broken generators reside in a single multiplet of H , we expect all the corresponding NG bosons to be either type-A or type-B. Which of these two options is realized depends on the VEV $\langle Q \rangle$. If $\langle Q \rangle = 0$, we find that $n_A = 2n - 2$ and $n_B = 0$. This happens in Lorentz-invariant theories, but also for instance in antiferromagnets where $n = 2$. If, on the other hand, $\langle Q \rangle \neq 0$ then the broken generators contribute pairwise to ϱ_{AB} . As a result, $\text{rank } \varrho = 2n - 2$. According to (6.24), we then find $n_A = 0$ and $n_B = n - 1$. The special case of $n = 2$ corresponds to ferromagnets.

Example 6.7

For an example of a system where a type-B NG boson arises from the VEV of a broken symmetry generator, consider the symmetry-breaking pattern $SU(2) \rightarrow \{e\}$. Here the unbroken subgroup is trivial and any generator may develop a nonzero VEV. However, the rank of antisymmetric matrices is always even. The only possibilities are therefore $\text{rank } \varrho = 0$ and $\text{rank } \varrho = 2$. In the latter case, we find $n_A = n_B = 1$. This is realized in so-called canted (anti)ferromagnets, schematically shown in Fig. 6.2.

So far I have assumed that the internal symmetry group G is compact and semisimple so that (6.25) holds. When this assumption is not satisfied, the commutator matrix (6.22) may receive contributions from central charges of the Lie algebra \mathfrak{g} . This is the case for instance for the free Schrödinger field. Without going into detail, let me just mention that such central extensions are classified by the so-called second *Lie algebra cohomology* of \mathfrak{g} . See Chap. 6 of [17] for further information.

How does the counting of NG bosons change for symmetries that are not internal, such as spacetime symmetries? In this case, there are no established general results in the literature. However, the simplicity of the argument given in Sect. 6.1.2 suggests that the reduction of independent NG modes based on the charge commutator (6.22) is robust. Suppose that we have already identified a set of independent NG fields

π^a . Let us further assume that we can map these to a set of broken generators Q_a that act on the fields by a shift such as (6.1). We can restrict the definition of the commutator matrix to the generators Q_a . The VEV in (6.22) may now in principle be coordinate-dependent, and it may no longer be appropriate to replace J_B^0 with the spatial average of Q_B . However, the rest of the argument will still go through. We will then get the same identification of type-B NG bosons with pairs of broken generators whose commutator has a nonzero VEV. Nontrivial examples of systems where the counting rule (6.24) for type-B NG bosons was still found to be valid include the centrally extended algebra of spatial translations in an external magnetic field [15], and central extensions of commutators of spatial and internal symmetries in presence of topological defects such as a domain wall [18] or a line defect [19].

6.4 Nambu–Goldstone-Like Modes

The defining property of a NG boson is its relation to symmetry and its spontaneous breaking. In the operator proof of Goldstone’s theorem (Sect. 6.2.1), this manifests through the coupling of the NG state to the broken symmetry current, $\langle 0 | J^0(0) | n, \mathbf{p} \rangle$. In the effective action proof (Sect. 6.2.2), the NG boson appears through a flat direction of the effective potential or action. Many physical systems possess excitations with the above properties that however lack the main attribute of a NG boson, that is being exactly gapless. Due to their similarity to genuine NG modes, I will call such excitations *NG-like bosons*.

The existence of SSB-related modes that are not exactly gapless is of course only possible if some of the assumptions of Goldstone’s theorem are violated. The usual suspect is the symmetry. There are very few *exact* symmetries in nature; such usually only exist in our models, valid to certain, possibly high, accuracy. In this last section of the chapter, I will give a brief survey of different mechanisms how a NG-like boson may arise from a symmetry that is only approximate.

6.4.1 Pseudo-Nambu–Goldstone Bosons

Consider a theory whose Lagrangian can be split into a part invariant under a group G , and a small perturbation invariant only under some proper subgroup of G ,

$$\mathcal{L} = \mathcal{L}_{\text{inv}} + \epsilon \mathcal{L}_{\text{pert}} . \quad (6.26)$$

This is a setup that I used previously to isolate a unique ground state. The difference is that now I want to keep the parameter ϵ small but nonzero. This is called *explicit* symmetry breaking.

Suppose that in the limit $\epsilon \rightarrow 0$, the symmetry under G is broken spontaneously. It is sensible to expect that by restoring small but nonzero ϵ , the spectrum will evolve

adiabatically. The NG mode predicted by Goldstone’s theorem will then survive, but its mass (gap) will no longer be exactly zero. Such “approximate” NG bosons are usually called *pseudo-NG bosons*.

Let us try to make an educated guess at how the gap of a pseudo-NG boson depends parametrically on ϵ . Imagine that we have already derived a low-energy EFT for the (pseudo-)NG bosons in our system. The contributions of the perturbation $\mathcal{L}_{\text{pert}}$ to the effective Lagrangian can be organized by powers of ϵ . The leading contribution to the bilinear part of the effective Lagrangian typically comes from a nonderivative operator linear in ϵ . I will justify this claim more carefully in Chap. 8, where I actually construct the effective Lagrangian.

The rest of the argument depends on the type of the NG boson in the limit of vanishing perturbation. This type is still defined by the presence or absence of a term with a single time derivative, containing the given NG field. For type-A NG bosons, the bilinear Lagrangian contains two time derivatives which translate to squared energy in the Fourier space. For type-B NG bosons, on the other hand, a term with a single time derivative is present and dominates in the low-energy limit. This leads to the following parametric dependence of energy at zero momentum on ϵ ,

$$\begin{aligned} \lim_{p \rightarrow 0} E(p) &\propto \sqrt{\epsilon} && \text{“type-A” pseudo-NG boson} , \\ \lim_{p \rightarrow 0} E(p) &\propto \epsilon && \text{“type-B” pseudo-NG boson} . \end{aligned} \quad (6.27)$$

I have just used quotation marks for a reason. There is no established, unambiguous classification of pseudo-NG bosons. All I did was to rely on the continuity of the spectrum in the limit $\epsilon \rightarrow 0$. Moreover, the parametric scaling in (6.27) is based on the assumption of a rather specific form of perturbation in the effective Lagrangian. It is imaginable that even for an originally type-A NG boson, $\mathcal{L}_{\text{pert}}$ will induce an operator in the effective Lagrangian with a single time derivative and a coefficient proportional to some power of ϵ . That may change the way the gap of the pseudo-NG mode scales with ϵ . This is exactly what happens for the “massive” NG bosons, introduced in Sect. 6.4.3.

It is a simple application of the discussion in Sect. 6.3.2 to find how many pseudo-NG bosons one should expect in a given system. I will denote as G_ϵ and H_ϵ the symmetries of the action and of the ground state in presence of the perturbation. In analogy with (6.22), I will define the commutator matrix ϱ_ϵ by

$$\varrho_{\epsilon, AB} \equiv i \langle [Q_A, J_B^0] \rangle_\epsilon . \quad (6.28)$$

The ϵ in the subscripts reminds us that the indices A, B run over the generators of G_ϵ , and that the VEV is taken in the perturbed vacuum. According to (6.24), there are altogether $\dim G - \dim H - (1/2) \text{rank } \varrho$ NG modes in the limit $\epsilon \rightarrow 0$. By the same formula, $\dim G_\epsilon - \dim H_\epsilon - (1/2) \text{rank } \varrho_\epsilon$ genuine NG bosons survive upon switching on the perturbation. On the assumption of continuity of the spectrum, the

number of pseudo-NG bosons therefore equals

$$n_{\text{pseudoNG}} = (\dim G - \dim G_\epsilon) - (\dim H - \dim H_\epsilon) - \frac{1}{2}(\text{rank } \varrho - \text{rank } \varrho_\epsilon) . \quad (6.29)$$

We saw an example of a pseudo-NG boson already in Chap. 2. In that case, $G \simeq \text{U}(1)_V \times \text{U}(1)_A$ and $H \simeq \text{U}(1)_V$, whereas $G_\epsilon \simeq H_\epsilon \simeq \text{U}(1)_V$. The counting rule (6.29) is trivially satisfied and the prediction (6.27) for the scaling of the gap agrees with our explicit calculation of the mass, cf. (2.12). Next, let us have a look at a couple of less trivial examples.

Example 6.8

We already met *quantum chromodynamics* (QCD) in Chap. 5, see Example 5.5. In this case, the quark mass plays the role of the perturbation ϵ . With n_f flavors of massless quarks, QCD possesses a $G \simeq \text{SU}(n_f)_L \times \text{SU}(n_f)_R$ symmetry, spontaneously broken down to the “vector” subgroup $H \simeq \text{SU}(n_f)_V$. In presence of the perturbation (with equal masses of all quarks), this reduces to $G_\epsilon \simeq H_\epsilon \simeq \text{SU}(n_f)_V$. The vacuum of QCD is Lorentz-invariant, so the matrices ϱ and ϱ_ϵ are both identically zero. According to the counting rule (6.29), this leaves us with $n_f^2 - 1$ pseudo-NG bosons, which can be identified with a multiplet of pseudoscalar mesons. The fact that their mass scales with the square root of the quark mass, as predicted by (6.27), has been verified in numerical lattice simulations of QCD.

Example 6.9

As discussed before, ideal (isotropic) antiferromagnets possess a $G \simeq \text{SU}(2)$ spin symmetry, spontaneously broken down to $H \simeq \text{U}(1)$. The low-energy EFT for antiferromagnetic spin waves (magnons) is a nonrelativistic version of the nonlinear sigma model, introduced in Example 3.3,

$$\mathcal{L}_{\text{eff}} = \frac{\varrho_s}{2v^2} (\delta_{ij} \partial_0 n^i \partial_0 n^j - v^2 \delta_{ij} \nabla n^i \cdot \nabla n^j) + \epsilon (n^3)^2 . \quad (6.30)$$

Here ϱ_s is the so-called spin stiffness, \mathbf{n} a unit-vector variable encoding the two NG fields, and v the magnon phase velocity in the symmetric limit $\epsilon = 0$.

For $\epsilon > 0$, the perturbation proportional to $(n^3)^2$ describes a crystal anisotropy, corresponding to “easy-axis” antiferromagnets; see [20] for a mild introduction to antiferromagnetic magnons. The perturbation breaks the continuous spin symmetry explicitly down to $G_\epsilon \simeq \text{U}(1)$, consisting of rotations around the third spin axis. At the same time, it leaves only two possible ground states, $\langle \mathbf{n} \rangle_\epsilon = (0, 0, \pm 1)$, both of which preserve $H_\epsilon \simeq \text{U}(1)$. By the counting formula (6.29), both magnons become pseudo-NG bosons. Their gap is proportional to $\sqrt{\epsilon}$ in accord with (6.27).

6.4.2 Quasi-Nambu–Goldstone Bosons

The above general discussion of pseudo-NG bosons applies in principle to any NG-like mode, associated with an approximate symmetry. However, the strength of the perturbation, and hence the size of the gap in the spectrum, may depend on the concrete mechanism of explicit symmetry breaking. It is in particular possible to have systems with a well-defined classical limit whose symmetry is explicitly broken only by quantum corrections. The ensuing pseudo-NG modes have been dubbed *quasi-NG bosons* [21]. I refer the reader to [22, 23] for a detailed discussion of the spectrum of quasi-NG modes.

To be more concrete, suppose we are given a Lorentz-invariant Lagrangian whose potential has a higher symmetry than the kinetic term. At the classical level, the spectrum of NG bosons is determined by the zero modes of the Hessian of the potential in the ground state. But some of these zero modes may correspond to symmetries of the potential that do not preserve the kinetic term. The corresponding fluctuations of the order parameter will receive a mass generated by quantum corrections.

The lifting of the mass of the would-be NG modes does not have to be induced by the kinetic term. What matters is that the classical ground state is determined entirely by a part of the Lagrangian that has a higher symmetry than the whole. The reduction of the symmetry may also be caused for instance by other contributions to the classical potential, or by a coupling to gauge fields.

6.4.3 Massive Nambu–Goldstone Bosons

What the pseudo-NG bosons and their subclass, quasi-NG bosons, have in common is that their gap can rarely be determined from first principles. In general, it depends on all parameters a given theory might have. One says that the gap is not protected by symmetry, in contrast to genuine NG bosons whose gap is guaranteed to vanish by SSB. There is however a remarkable special class of pseudo-NG bosons whose gap can be calculated exactly using symmetry alone. Their existence was first pointed out about a decade ago [24] and I will call them *massive NG bosons* following [25].

Consider a system with a “microscopic” Hamiltonian H , possessing an internal symmetry group G . Pick one generator, Q , of G . In statistical physics, the many-body ground state of the system with a fixed average value of Q is determined by minimizing the grandcanonical Hamiltonian $H_\mu \equiv H - \mu Q$. Here μ is the corresponding chemical potential, which is fixed by external conditions imposed on the system. If needed, one can always shift H_μ by a constant so that the many-body ground state $|0\rangle_\mu$ satisfies $H_\mu |0\rangle_\mu = 0$. In this ground state, the symmetry under G may be spontaneously broken. We are particularly interested in the fate of those generators of G that are spontaneously broken but do not commute with Q .

As long as G is compact, we can always choose Q to be part of a Cartan subalgebra of the Lie algebra \mathfrak{g} of G . Those generators of G that do not commute with Q then organize themselves into pairs Q_\pm such that $Q_- = Q_+^\dagger$ and

$$[Q, Q_{\pm}] = \pm q Q_{\pm} . \quad (6.31)$$

The factor q is determined by the corresponding root vector, and is thus fixed by symmetry. We can always choose the generators so that both μ and q are positive. I will now rerun the operator proof of Goldstone’s theorem detailed in Sect. 6.2.1 on the VEV ${}_{\mu}\langle 0|[Q_{+\Omega}(t), J_{-}^0(0)]|0\rangle_{\mu}$.

Of course, (6.31) is only formal since we know that generators of spontaneously broken symmetry are ill-defined as operators on the Hilbert space. I therefore assume that (6.31) also holds for the finite-volume charges $Q_{\pm\Omega}$, possibly up to a correction that vanishes in the limit $\Omega \rightarrow \infty$. This can be ensured for instance if the charge densities satisfy the local commutation relation $[Q, J_{\pm}^0(x)] = \pm q J_{\pm}^0(x)$. One can then define a related pair of charges via “time evolution” with respect to H_{μ} ,

$$\begin{aligned} {}^{\mu}Q_{\pm\Omega}(t) &\equiv \int_{\Omega} d^d \mathbf{x} \exp(iH_{\mu}t - i\mathbf{P} \cdot \mathbf{x}) J_{\pm}^0(0) \exp(-iH_{\mu}t + i\mathbf{P} \cdot \mathbf{x}) \\ &= e^{-i\mu Q t} Q_{\pm\Omega}(t) e^{+i\mu Q t} = e^{\mp i\mu q t} Q_{\pm\Omega}(t) . \end{aligned} \quad (6.32)$$

The point of this redefinition is that the spectrum of the many-body system consists of eigenstates of H_{μ} rather than H . I will keep the notation $E_n(\mathbf{p})$ for the eigenvalue of H_{μ} (excitation energy) in the eigenstate $|n, \mathbf{p}\rangle$. Following the same steps as in Sect. 6.2.1 then leads to an analog of (6.11),

$$\begin{aligned} {}_{\mu}\langle 0|[Q_{+\Omega}(t), J_{-}^0(0)]|0\rangle_{\mu} &= e^{i\mu q t} {}_{\mu}\langle 0|[^{\mu}Q_{+\Omega}(t), J_{-}^0(0)]|0\rangle_{\mu} \\ &= e^{i\mu q t} \sum_n \int d^d \mathbf{p} \left\{ \exp[-iE_n(\mathbf{p})t] \delta_{\Omega}^d(\mathbf{p}) \left| {}_{\mu}\langle 0|J_{+}^0(0)|n, \mathbf{p}\rangle \right|^2 \right. \\ &\quad \left. - \exp[+iE_n(\mathbf{p})t] \delta_{\Omega}^d(-\mathbf{p}) \left| {}_{\mu}\langle 0|J_{-}^0(0)|n, \mathbf{p}\rangle \right|^2 \right\} . \end{aligned} \quad (6.33)$$

With positive μ and q , the assumed time independence of VEVs of commutators of $Q_{+\Omega}(t)$ in the $\Omega \rightarrow \infty$ limit guarantees that $\lim_{\mathbf{p} \rightarrow \mathbf{0}} {}_{\mu}\langle 0|J_{-}^0(0)|n, \mathbf{p}\rangle = 0$ for any eigenstate $|n, \mathbf{p}\rangle$. Assuming that the VEV we started with is nonzero, on the other hand, ensures the existence of a state for which ${}_{\mu}\langle 0|J_{+}^0(0)|n, \mathbf{p}\rangle \neq 0$ and

$$\lim_{\mathbf{p} \rightarrow \mathbf{0}} E_n(\mathbf{p}) = \mu q . \quad (6.34)$$

This is our massive NG boson. With some extra effort, one can also derive a counting rule for the number of massive NG modes in a system. To formulate such a rule, we adapt (6.22) for the present problem by introducing $\mathcal{Q}_{\mu, AB} \equiv i\langle [Q_A, J_B^0] \rangle_{\mu}$ and $\tilde{\mathcal{Q}}_{\mu, AB} \equiv i\langle [\tilde{Q}_A, \tilde{J}_B^0] \rangle_{\mu}$. The matrix $\tilde{\mathcal{Q}}$ is constructed using the generators of the subgroup of G that commutes with Q . The number of massive NG modes then is

$$n_{\text{massiveNG}} = \frac{1}{2} (\text{rank } \mathcal{Q}_{\mu} - \text{rank } \tilde{\mathcal{Q}}_{\mu}) . \quad (6.35)$$

See [25] for a detailed proof.

The counting rule (6.35) may resemble the last term in (6.29) that counts pseudo-NG bosons. It may thus be worthwhile to stress the difference. Namely, the matrices \mathcal{Q}

and ϱ_ϵ in (6.29) are evaluated in different ground states, corresponding respectively to the absence and presence of the perturbation. On the contrary, both matrices ϱ_μ and $\tilde{\varrho}_\mu$ in (6.35) are evaluated in the same ground state, “perturbed” by the presence of the chemical potential. The difference is best illustrated on an example.

Example 6.10

Let us have one more look at antiferromagnets, this time assuming perfect isotropy and spin symmetry, that is $G \simeq \text{SU}(2)$ and $H \simeq \text{U}(1)$. In an external magnetic field \mathbf{B} , the Hamiltonian of the antiferromagnet receives a contribution $-\mu_m \mathbf{B} \cdot \mathbf{S}$, where μ_m is the magnetic moment and \mathbf{S} the operator of total spin. The combination $\mu_m |\mathbf{B}|$ plays the role of a chemical potential for the projection of \mathbf{S} into the direction of \mathbf{B} . The latter can without loss of generality be chosen as, say, S_3 .

Let us first treat the effect of the magnetic field as a perturbation in the sense of Sect. 6.4.1. The field breaks the spin symmetry explicitly down to $G_\epsilon \simeq \text{U}(1)$. This necessarily implies $\text{rank } \varrho_\epsilon = 0$. Somewhat against the intuition, the antiferromagnet responds to the magnetic field by aligning its spins largely in a direction perpendicular to \mathbf{B} . This ground state breaks the residual exact symmetry so that $H_\epsilon \simeq \{e\}$. At the same time, in the unperturbed antiferromagnetic ground state, there is no net magnetization so that $\text{rank } \varrho = 0$. Equations (6.22) and (6.29) then tell us that there is one genuine, type-A NG boson and one pseudo-NG boson. Unlike in the case of an easy-axis anisotropy, one of the two magnons therefore remains gapless.

The fact that the magnetic field acts as a chemical potential allows us to make a stronger statement about the spectrum. The magnetic field does induce net magnetization in the perturbed ground state, given by nonzero $\langle S_3 \rangle_\mu = -i \langle [S_1, S_2] \rangle_\mu$. This implies $\text{rank } \varrho_\mu = 2$. At the same time, $\text{rank } \tilde{\varrho}_\mu = 0$ simply because there are not enough symmetry generators left intact by the magnetic field. In accord with (6.35), we then expect one massive NG mode in the spectrum. In other words, the previously predicted pseudo-NG magnon has a gap exactly fixed by (6.34). The algebraic factor q equals one thanks to the commutator of S_3 with the ladder operators $S_\pm \equiv S_1 \pm iS_2$, $[S_3, S_\pm] = \pm S_\pm$. Hence the gap of the magnon equals $\mu_m |\mathbf{B}|$.

An interested reader will find further examples of massive NG bosons in [25]. Let me just stress that the exact expression (6.34) for the gap is not the only special property of massive NG bosons. The peculiar perturbation by a chemical potential coupled to a conserved charge leaves us with a set of exact, albeit modified, conservation laws. These in turn impose exact constraints on the interactions of massive NG bosons. As a result, the Adler zero principle (cf. Sect. 2.2) remains valid for massive NG modes, although it is generally violated for pseudo-NG bosons [26].

As an aside, note that not every NG boson that acquires a gap due to a chemical potential is a massive NG boson [27]. The simplest example of such an exception is provided by choosing $G \simeq \text{SU}(2)$ and breaking it completely in the ground state. Provided none of the three generators develops a VEV, the spectrum contains three type-A NG bosons. Upon adding a chemical potential μ for one of the generators,

two of the NG modes receive a gap, proportional to μ . Only one of these is however a massive NG boson, as is easily checked using (6.35).

References

1. J. Goldstone, *Nuovo Cim.* **19**, 154 (1961). DOI [10.1007/BF02812722](https://doi.org/10.1007/BF02812722)
2. J. Goldstone, A. Salam, S. Weinberg, *Phys. Rev.* **127**, 965 (1962). DOI [10.1103/PhysRev.127.965](https://doi.org/10.1103/PhysRev.127.965)
3. G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, in *Advances in Particle Physics*, vol. II, ed. by R.L. Cool, R.E. Marshak (Wiley, New York, NY, USA, 1968), pp. 567–708
4. F. Strocchi, *Symmetry Breaking. Theoretical and Mathematical Physics* (Springer Nature Switzerland AG, 2021). DOI [10.1007/978-3-662-62166-0](https://doi.org/10.1007/978-3-662-62166-0)
5. H. Watanabe, T. Brauner, *Phys. Rev.* **D85**, 085010 (2012). DOI [10.1103/PhysRevD.85.085010](https://doi.org/10.1103/PhysRevD.85.085010)
6. E. Fabri, L.E. Picasso, *Phys. Rev. Lett.* **16**, 408 (1966). DOI [10.1103/PhysRevLett.16.408.2](https://doi.org/10.1103/PhysRevLett.16.408.2)
7. H. Watanabe, T. Brauner, *Phys. Rev.* **D84**, 125013 (2011). DOI [10.1103/PhysRevD.84.125013](https://doi.org/10.1103/PhysRevD.84.125013)
8. S. Weinberg, *The Quantum Theory of Fields*, vol. II (Cambridge University Press, Cambridge, UK, 1996)
9. I. Low, A.V. Manohar, *Phys. Rev. Lett.* **88**, 101602 (2002). DOI [10.1103/PhysRevLett.88.101602](https://doi.org/10.1103/PhysRevLett.88.101602)
10. H. Watanabe, H. Murayama, *Phys. Rev. Lett.* **110**, 181601 (2013). DOI [10.1103/PhysRevLett.110.181601](https://doi.org/10.1103/PhysRevLett.110.181601)
11. H. Watanabe, *Ann. Rev. Cond. Mat. Phys.* **11**, 169 (2020). DOI [10.1146/annurev-conmatphys-031119-050644](https://doi.org/10.1146/annurev-conmatphys-031119-050644)
12. T. Hayata, Y. Hidaka, *Phys. Lett.* **B735**, 195 (2014). DOI [10.1016/j.physletb.2014.06.039](https://doi.org/10.1016/j.physletb.2014.06.039)
13. J.i. Kishine, A.S. Ovchinnikov, *Solid State Phys.* **66**, 1 (2015). DOI [10.1016/bs.ssp.2015.05.001](https://doi.org/10.1016/bs.ssp.2015.05.001)
14. Y. Togawa, Y. Kousaka, K. Inoue, J.i. Kishine, *J. Phys. Soc. Jpn.* **85**, 112001 (2016). DOI [10.7566/JPSJ.85.112001](https://doi.org/10.7566/JPSJ.85.112001)
15. H. Watanabe, H. Murayama, *Phys. Rev. Lett.* **108**, 251602 (2012). DOI [10.1103/PhysRevLett.108.251602](https://doi.org/10.1103/PhysRevLett.108.251602)
16. Y. Hidaka, *Phys. Rev. Lett.* **110**, 091601 (2013). DOI [10.1103/PhysRevLett.110.091601](https://doi.org/10.1103/PhysRevLett.110.091601)
17. J.A. de Azcárraga, J.M. Izquierdo, *Lie Groups, Lie Algebras, Cohomology and Some Applications in Physics* (Cambridge University Press, UK, 1995)
18. M. Kobayashi, M. Nitta, *Phys. Rev. Lett.* **113**, 120403 (2014). DOI [10.1103/PhysRevLett.113.120403](https://doi.org/10.1103/PhysRevLett.113.120403)
19. M. Kobayashi, M. Nitta, *Phys. Rev.* **D90**, 025010 (2014). DOI [10.1103/PhysRevD.90.025010](https://doi.org/10.1103/PhysRevD.90.025010)
20. S.M. Rezende, A. Azevedo, R.L. Rodríguez-Suárez, *J. Appl. Phys.* **126**, 151101 (2019). DOI [10.1063/1.5109132](https://doi.org/10.1063/1.5109132)
21. S. Uchino, M. Kobayashi, M. Nitta, M. Ueda, *Phys. Rev. Lett.* **105**, 230406 (2010). DOI [10.1103/PhysRevLett.105.230406](https://doi.org/10.1103/PhysRevLett.105.230406)
22. D.A. Takahashi, M. Nitta, *Ann. Phys.* **354**, 101 (2015). DOI [10.1016/j.aop.2014.12.009](https://doi.org/10.1016/j.aop.2014.12.009)
23. M. Nitta, D.A. Takahashi, *Phys. Rev.* **D91**, 025018 (2015). DOI [10.1103/PhysRevD.91.025018](https://doi.org/10.1103/PhysRevD.91.025018)
24. A. Nícolis, F. Piazza, *Phys. Rev. Lett.* **110**, 011602 (2013). DOI [10.1103/PhysRevLett.110.011602](https://doi.org/10.1103/PhysRevLett.110.011602)
25. H. Watanabe, T. Brauner, H. Murayama, *Phys. Rev. Lett.* **111**, 021601 (2013). DOI [10.1103/PhysRevLett.111.021601](https://doi.org/10.1103/PhysRevLett.111.021601)
26. T. Brauner, M.F. Jakobsen, *Phys. Rev.* **D97**, 025021 (2018). DOI [10.1103/PhysRevD.97.025021](https://doi.org/10.1103/PhysRevD.97.025021)
27. A. Nícolis, R. Penco, F. Piazza, R.A. Rosen, *JHEP* **11**, 055 (2013). DOI [10.1007/JHEP11\(2013\)055](https://doi.org/10.1007/JHEP11(2013)055)

Part III
Spontaneously Broken Internal Symmetry

Chapter 7

Nonlinear Realization of Symmetry

In Parts **I** and **II** of the book, I reviewed the basic physics of *spontaneous symmetry breaking* (SSB) and the corresponding *Nambu–Goldstone* (NG) *bosons*. In order to keep the discussion simple, I relied mostly on general but elementary field-theoretic arguments. Where explicit results were necessary or useful, I resorted to simple models. However, one of the main virtues of SSB is universality: the low-energy physics is largely independent of the microscopic, short-distance details. This agrees with the general spirit of *effective field theory* (EFT); cf. Sect. 1.1. What SSB does for us is ensure a separation of resolution scales, which makes it possible to build a low-energy EFT solely using the NG degrees of freedom.

I will now set out on the main quest of this book: to develop an EFT framework for SSB that is based on symmetry alone. Thanks to the universality of SSB, the predictions of such a framework are guaranteed to match those of any microscopic theory with the same symmetry-breaking pattern. The construction of the EFT requires two main steps. The first step is to realize the action of the symmetry solely in terms of a set of NG fields. This problem is dealt with in the present chapter. The results will be used in the next chapter to construct effective actions for NG bosons.

The celebrated Wigner theorem of quantum mechanics states that any symmetry of a quantum system can be realized by a linear unitary or antilinear antiunitary operator on its Hilbert space (see Chap. 2 of [1] for a detailed discussion). As a consequence, states in the spectrum of a quantum system can be organized into multiplets, corresponding to irreducible representations of its symmetry group. In quantum field theory, the concept of linear realization of symmetry is naturally lifted from states to fields. In a perturbative setting, one usually has a one-to-one correspondence between one-particle states and (elementary) fields. Moreover, the fields span a linear space. One then expects the symmetry to act on fields linearly via some representation just like it does on states in the Hilbert space.

Example 7.1

A field theory of a single complex scalar field ϕ may possess a $U(1)$ symmetry under which the field transforms as $\phi \rightarrow e^{i\epsilon} \phi$, where ϵ is a real parameter. In this case,

the field ϕ belongs to a complex one-dimensional representation of $U(1)$. Similarly, the $SO(n)$ “linear sigma model” includes an n -plet of real scalar fields ϕ^i , subject to symmetry transformations $\phi^i \rightarrow R^i_j \phi^j$, where $R \in SO(n)$. In this case, the fields ϕ^i belong to the vector representation of $SO(n)$.

As we saw in Sect. 5.3, however, a symmetry that is spontaneously broken is not necessarily realized by a set of unitary operators on the Hilbert space of states. Likewise, as illustrated in Chap. 2, it may be convenient to use field variables that do not belong to a linear representation of the symmetry. This may even become a necessity in the low-energy EFT where only the NG bosons of the broken symmetry are present; we simply do not have enough degrees of freedom to fill complete multiplets of the symmetry group.

Example 7.2

The complex scalar field of Example 7.1 can be represented by its real and imaginary parts, $\phi = \phi^1 + i\phi^2$. These span the vector representation of $SO(2) \simeq U(1)$. When the $U(1)$ symmetry is spontaneously broken, it may however be more convenient to use the exponential parameterization of the field, $\phi = \varrho e^{i\theta}$, in terms of its modulus ϱ and phase θ . In the low-energy EFT, the modulus field is integrated out and the only remaining degree of freedom is the NG field θ . The latter transforms under $e^{i\epsilon} \in U(1)$ as $\theta \rightarrow \theta + \epsilon$, which is not a linear representation. This is not merely a matter of bad choice of parameterization; the EFT contains a single real field, yet the group $U(1)$ does not have any nontrivial real one-dimensional representations.

With the above observations in mind, I will develop in this chapter a formalism for nonlinear realization of symmetries. Mathematically, this amounts to generalizing the concept of a linear representation of a symmetry group to that of an *action* of the group. The space on which the group acts need not be linear itself; we can think of it as some manifold. Section 7.1 introduces the necessary mathematical terminology. The main argument, leading to a classification of possible nonlinear realizations of symmetry, is presented in Sect. 7.2. A reader interested mainly in ready-made results of the formalism may want to proceed directly to Sect. 7.3; this collects a number of practically useful formulas that I will refer to in the following chapters. Finally, Sect. 7.4 offers an alternative geometric viewpoint which illuminates some of the mathematical structure used to realize a symmetry nonlinearly. While most of the chapter is phrased in a rather elementary language, this last section relies on some concepts of differential geometry in an extent covered in Appendix A.

7.1 Group Action on Manifolds

To motivate the mathematical language that I need to introduce, suppose that we are given a theory of a set of fields, ψ^i , that possesses a symmetry group, G .¹ We would like to understand how to realize the symmetry group in terms of a set of transformations of the fields,

$$T_g : \psi^i \rightarrow \psi'^i \equiv \mathbb{F}^i(\psi, g), \quad g \in G, \quad (7.1)$$

where the functions \mathbb{F}^i are assumed to be smooth in both of their arguments. The set of transformations T_g is constrained by the requirement that it respects the group structure of G . Thus, the unit element $e \in G$ must be represented by the identity map id , $\mathbb{F}^i(\psi, e) = \psi^i$. Consistency with group multiplication requires that

$$\mathbb{F}^i(\psi, g_1 g_2) = \mathbb{F}^i(\mathbb{F}(\psi, g_2), g_1), \quad g_1, g_2 \in G. \quad (7.2)$$

Finally, the transformation induced by the inverse of an element $g \in G$ has to satisfy

$$\mathbb{F}^i(\mathbb{F}(\psi, g), g^{-1}) = \psi^i, \quad g \in G. \quad (7.3)$$

In the terminology introduced in Chap. 4, (7.1) is an example of a *point transformation* [2, 3]. The class of point transformations is clearly much broader than that of mere linear transformations, induced by a representation of G on the fields. It is therefore worthwhile to recall that even point transformations do not exhaust all conceivable, and physically relevant, realizations of symmetry. First, the field transformation may in principle depend explicitly on the spacetime coordinates. This feature can be included under the umbrella of point symmetries by treating fields and coordinates on the same footing; this will become relevant later when we talk about spacetime symmetries. Moreover, it is perfectly possible that the transformation of the fields also depends on their derivatives; this was dubbed *generalized local transformation* in Sect. 4.1. Such generalized symmetries play a minor role in this book, yet we will see some concrete examples in Chap. 10.

The restriction to point symmetries of the type (7.1), which I will make from now on unless explicitly stated otherwise, is a matter of practical convenience. Namely, it will allow us to disregard the fact that ψ^i actually are fields, that is functions of spacetime coordinates. Instead, I will treat them as independent variables that the group G acts upon. With this important technical assumption, we can now reformulate the problem of finding all possible realizations of a given symmetry group G on the fields ψ^i in geometric terms.

Consider a manifold \mathcal{M} such that each point $x \in \mathcal{M}$ is uniquely specified by a set of values ψ^i . We can think of ψ^i as a set of (possibly only locally defined) coordinates on \mathcal{M} . An *action* of the group G on \mathcal{M} is a set of smooth invertible

¹ Many of the concepts introduced below can be applied without change to any, even finite, group. However, I will always have implicitly in mind a connected Lie group, or the component of a Lie group connected to the unit element.

maps $T_g : \mathcal{M} \rightarrow \mathcal{M}$ that satisfy the group constraints

$$T_e = \text{id} , \quad T_{g_1 g_2} = T_{g_1} \circ T_{g_2} , \quad T_{g^{-1}} = (T_g)^{-1} , \quad g, g_1, g_2 \in G . \quad (7.4)$$

In somewhat more abstract terms, the action of the group on the manifold is defined by a homomorphism from G to the group of diffeomorphisms on \mathcal{M} .

It is important to distinguish actual symmetry transformations on \mathcal{M} from a mere change of coordinates. The action of the group G on \mathcal{M} is defined geometrically by the maps T_g without reference to a particular set of coordinates ψ^i . Depending on the choice of coordinates, the same map T_g may correspond to different functions \mathbb{F}^i as defined by (7.1). The freedom to choose coordinates on \mathcal{M} mirrors the freedom to choose field variables in a given field theory. On the one hand, the independence of geometric properties of manifolds on the choice of local coordinates is a cornerstone of the language of differential geometry. On the other hand, it is an important result of quantum field theory that physical observables such as the S -matrix are invariant under (nearly) arbitrary field redefinitions [4, 5]. See [6] for a recent pedagogical discussion of this issue.

Let us now introduce some further terminology. For a given action of the group G on the manifold \mathcal{M} , the *orbit* of a point $x \in \mathcal{M}$ is the set of all points on \mathcal{M} that can be reached from x by the action of some group element,

$$\mathcal{O}_x \equiv \{T_g x \mid g \in G\} . \quad (7.5)$$

The relation $y \sim x$ if and only if there is a $g \in G$ such that $y = T_g x$ is an equivalence. Orbits of the group G on the manifold \mathcal{M} are the equivalence classes of this relation. As a consequence, \mathcal{M} is a disjoint union of a (possibly infinite) set of orbits.

For a given point $x \in \mathcal{M}$, one defines its *isotropy group* (also called the *stabilizer* or the *little group* of x) as the subgroup of G that maps x to itself,

$$H_x \equiv \{h \in G \mid T_h x = x\} . \quad (7.6)$$

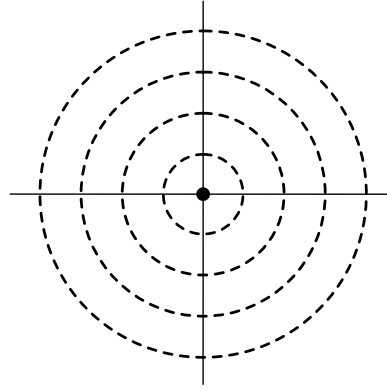
Two points lying on the same orbit of G have isomorphic isotropy groups. Indeed, if $y = T_g x$, then for any $h \in H_x$ we have $T_{ghg^{-1}} y = T_g T_h T_{g^{-1}} y = T_g T_h x = T_g x = y$. Conversely, it is easy to check that $T_h y = y$ implies $g^{-1} h g \in H_x$. Thus, H_x and H_y are conjugate as subgroups,

$$H_{T_g x} = g H_x g^{-1} . \quad (7.7)$$

Example 7.3

The rotation group $\text{SO}(2)$ acts on the Euclidean plane by rotations around the origin, see Fig. 7.1 for an illustration. All points of the plane away from the origin have a

Fig. 7.1 Illustration of the action of the rotation group $SO(2)$ on the Euclidean plane. Some orbits of $SO(2)$ for which the isotropy group is trivial are displayed using dashed lines. The black dot indicates the single orbit whose isotropy group is the whole of $SO(2)$



trivial isotropy group. The corresponding orbits of $SO(2)$ are circles centered at the origin. The origin itself forms an orbit with the isotropy group $SO(2)$.

This example has a straightforward generalization to the action of $SO(n)$ on the Euclidean space \mathbb{R}^n . There, the origin has the isotropy group $SO(n)$. All other points $x \in \mathbb{R}^n$ have the isotropy group $SO(n-1)$, corresponding to $(n-1)$ -dimensional rotations that leave the line connecting x to the origin fixed. The corresponding orbits are $(n-1)$ -dimensional spheres centered at the origin.

In somewhat loose terms, we can say that all points lying on the same orbit of G on the manifold \mathcal{M} have “the same” properties, since they can be related by a symmetry transformation. We thus expect manifolds consisting of a single orbit to have particularly simple geometric properties. In this case, the action of the group G on the manifold \mathcal{M} is called *transitive*; any point on the manifold can be reached from any other point by the action of a suitable group element. A manifold equipped with a transitive action of a group is referred to as a *homogeneous space*.

Homogeneous spaces will play a central role throughout the rest of this book. One of the special properties of a homogeneous space is that its structure is completely determined by the group G and its subgroup H that specifies the isotropy group of the homogeneous space.² To see why, let us introduce a relation between two elements of G : $g_2 \sim g_1$ if and only if there is an $h \in H$ such that $g_2 = g_1 h$. This is an equivalence relation. As a consequence, the group G is partitioned into a disjoint union of the corresponding equivalence classes. The equivalence class

$$gH \equiv \{gh \mid h \in H\}, \quad g \in G, \quad (7.8)$$

is called the *left coset* (or simply *coset*) of H in G . Incidentally, the cosets themselves can be viewed as orbits of a group action if we treat the group G as the manifold on which the subgroup H acts by multiplication from the right.

² Since the isotropy group is the same up to conjugation for all points on the same orbit, and hence all points of the homogeneous space, we can drop the subscript x .

The quotient set G/H , called the *coset space*, is the set of all cosets of H in G . Despite the different appearance, this is just another mathematical realization of a homogeneous space. Indeed, consider a homogeneous space \mathcal{M} along with a single point $x \in \mathcal{M}$ and its isotropy group H . It is easy to see that for any two elements $g_1, g_2 \in G$, $T_{g_1}x = T_{g_2}x$ if and only if $g_1 \sim g_2$, that is if g_1 and g_2 belong to the same coset. Hence there is a one-to-one correspondence between the elements of the coset space G/H and the points of the manifold \mathcal{M} .

Example 7.4

The Euclidean group $\text{ISO}(n)$ consists of proper rotations and translations in \mathbb{R}^n and their combinations. It obviously acts transitively on \mathbb{R}^n for the simple reason that any point in \mathbb{R}^n can be reached from any other point by a suitable translation. Each point in \mathbb{R}^n has as its stabilizer a particular subgroup of $\text{ISO}(n)$, consisting of $\text{SO}(n)$ rotations around that point. Hence \mathbb{R}^n equipped with the action of $\text{ISO}(n)$ is a homogeneous space, equivalent to the coset space $\text{ISO}(n)/\text{SO}(n)$. This view of the Euclidean space in terms of its symmetry group follows the “Erlangen program,” put forward by Felix Klein in 1872.

Example 7.5

The rotation group $\text{SO}(n+1)$ acts naturally on the n -dimensional unit sphere S^n if one thinks of the latter as embedded in \mathbb{R}^{n+1} . We can however think of the sphere S^n in itself as our manifold \mathcal{M} . In that case, the action of $\text{SO}(n+1)$ becomes transitive; any point on the sphere can be reached from any other point by a suitable rotation. The isotropy group of any point on the sphere is a particular $\text{SO}(n)$ subgroup of $\text{SO}(n+1)$. Hence, the sphere S^n equipped with the action of $\text{SO}(n+1)$ is a homogeneous space, equivalent to the coset space $\text{SO}(n+1)/\text{SO}(n)$.

Having set up the necessary mathematical background, I can now give a concise formulation of the main goal of this chapter. In order to understand how a given symmetry group G can act on a given set of fields, we need to classify all possible actions of G on a given manifold \mathcal{M} . This will be accomplished in the following section. A reader seeking further details on the mathematical background reviewed above is recommended to consult Chap. 13 of [7].

7.2 Classification of Nonlinear Realizations

At first sight, the task to classify all actions of an arbitrary group G on an arbitrary manifold \mathcal{M} seems hopeless. We would only know what to do in the case of a linear

Table 7.1 Correspondence between the mathematical terminology used in this chapter and the physics terminology introduced in Chap. 5

Mathematics terminology	Physics terminology
“Origin” $x_0 \in \mathcal{M}$	Vacuum state (order parameter)
Isotropy group H_{x_0}	Unbroken subgroup H
Coset space G/H_{x_0}	Vacuum manifold

action on a vector space, where the problem boils down to the good old representation theory of Lie groups. How can we make use of this?

In this section, I will answer the above question, following the classic work of Coleman, Wess and Zumino [8]. The key step is to observe that the action of a group on a manifold can be *partially* linearized by a suitable choice of coordinates on the manifold. Let us choose a fixed point $x_0 \in \mathcal{M}$. We can always introduce a set of coordinates ψ^i in the neighborhood of x_0 such that x_0 maps to the origin, $\psi^i = 0$. Now any linear transformation leaves the origin intact. On the contrary, only the isotropy group H_{x_0} keeps x_0 fixed; all the other elements of G translate x_0 to some other point on the manifold. Thus, the best we can hope for is that we find a set of coordinates in which the action of H_{x_0} , not of the whole group G , becomes linear.

We will see that provided the isotropy group H_{x_0} is compact, one can indeed construct a set of local coordinates on \mathcal{M} in the neighborhood of x_0 in which the action of H_{x_0} is linear. The good news is that this is in fact sufficient to classify all nonlinear realizations of the whole group G . The details of the argument are the subject of the following two subsections.

While this whole chapter is intended to develop a mathematical formalism that will prove invaluable later, it might be helpful for the reader already now to keep in mind the corresponding concepts pertinent to SSB, introduced in Chap. 5. The fixed point $x_0 \in \mathcal{M}$ corresponds to the selected vacuum state, or the associated order parameter. The subgroup H_{x_0} that leaves x_0 intact is analogous to the subgroup of unbroken symmetries. Finally, the coset space G/H_{x_0} , which will turn out to span a submanifold of \mathcal{M} , is analogous to the vacuum manifold. For the reader’s convenience, this correspondence is summarized in Table 7.1.

7.2.1 Linearization of Group Action

To simplify the notation, I will from now on denote the functions \mathbb{F}^i defined in (7.1) directly as ψ^i whenever possible. Suppose that we have a set of local coordinates ψ^i on \mathcal{M} in which the chosen point x_0 maps to zero. In these coordinates, the isotropy group H_{x_0} will be represented by some nonlinear functions of ψ^i whose Taylor expansion in ψ^i starts at the linear order,

$$\psi'^i(\psi, h) = D(h)^i_j \psi^j + O(\psi^2), \quad (7.9)$$

where $h \in H_{x_0}$ and $D(h)^i_j$ is a set of matrix coefficients. The conditions (7.2) and (7.3) require that the matrices $D(h)$ form a representation of H_{x_0} . We can then change the coordinates in the vicinity of x_0 to

$$\Psi^i(\psi) \equiv \int_{H_{x_0}} dh D(h^{-1})^i_j \psi'^j(\psi, h), \quad (7.10)$$

where dh is an invariant measure on H_{x_0} normalized so that the total volume of H_{x_0} is one.³ The normalization ensures that near the origin, $\Psi^i = \psi^i + O(\psi^2)$. Hence, Ψ^i are a well-defined set of coordinates in some neighborhood of the origin. In these new coordinates, the action of the isotropy group is linear. This can be seen upon a short manipulation,

$$\begin{aligned} \Psi'^i(\Psi, h') &= \int_{H_{x_0}} dh D(h^{-1})^i_j \mathbb{F}^j(\mathbb{F}(\psi, h'), h) = \int_{H_{x_0}} dh D(h^{-1})^i_j \mathbb{F}^j(\psi, hh') \\ &= D(h')^i_j \int_{H_{x_0}} d(hh') [D(hh')^{-1}]^j_k \psi'^k(\psi, hh') = D(h')^i_j \Psi^j, \end{aligned} \quad (7.11)$$

for any $h' \in H_{x_0}$, where I used respectively the group composition law (7.2), the invariance of the integration measure, the representation property of the matrices $D(h)$, and the definition (7.10) of the new coordinates. This concludes the proof of the statement that the action of the isotropy group H_{x_0} can be linearized by a suitable choice of coordinates.

In defining the new coordinates (7.10), I tacitly assumed that H_{x_0} has a finite volume so that the invariant measure on H_{x_0} in fact *can* be normalized to unity. This is a key step in the proof, which requires that the isotropy group H_{x_0} be compact. The technique of nonlinear realization of symmetry developed in this chapter is often introduced straight away under the assumption that the whole group G is compact. That is, however, not necessary.

Equation (7.10) can serve as a useful practical tool to find explicitly the coordinates that linearize the action of the isotropy group. Let us have a look at a simple example, following [10].

Example 7.6

The action of $SO(2)$ on the Euclidean plane, introduced in Example 7.3, can be recast in terms of an action of $U(1) \simeq SO(2)$ on the complex plane \mathbb{C} . Thus, under $e^{i\epsilon} \in U(1)$, the complex coordinate z transforms as $z \rightarrow z' \equiv e^{i\epsilon} z$. This action

³ A reader not familiar with group integration may find more information in Chap. 3 of [9]. I will not dwell on details, since we shall not need group integration again in this book.

is linear. One can however tweak it by changing the coordinate z to w such that $z = f(w)$, where f is a function analytic in the neighborhood of the origin of \mathbb{C} such that $f(w) = w + O(w^2)$. In the new coordinate w , the $U(1)$ group acts via

$$w \rightarrow w' = f^{-1}(z') = f^{-1}(e^{i\epsilon} f(w)) . \quad (7.12)$$

In general, w' will be a nonlinear function of w . The original coordinate z , in which the action of $U(1)$ is linear, can be reconstructed using (7.10). In case of $U(1)$, the group integration in (7.10) amounts to averaging over the phase ϵ of the $U(1)$ rotation. It can be written as an integral over a unit circle in the complex plane,

$$\frac{1}{2\pi} \int_0^{2\pi} d\epsilon e^{-i\epsilon} f^{-1}(e^{i\epsilon} f(w)) = -\frac{i}{2\pi} \oint \frac{dc}{c^2} f^{-1}(cf(w)) , \quad (7.13)$$

where I introduced a new complex integration variable $c \equiv e^{i\epsilon}$. Given the assumptions I made on f , the function $f^{-1}(cf(w))/c^2$ of the complex variable c has a simple pole at the origin with the residue $f(w)$. It then follows at once from the residue theorem that the integral in (7.13) evaluates to $f(w) = z$, as expected.

Note that the coordinate z is not uniquely specified by the requirement that the action of $U(1)$ is linear. We can for instance introduce a new variable w via

$$w = zf(z\bar{z}) , \quad (7.14)$$

where f is a smooth real function such that near the origin, $f(z\bar{z}) = 1 + O(z\bar{z})$. Then w is a well-defined coordinate in some neighborhood of the origin of \mathbb{C} , upon which $e^{i\epsilon} \in U(1)$ acts linearly as $w \rightarrow e^{i\epsilon} w$.

The above example suggests that the local coordinates Ψ^i in which the isotropy group H_{x_0} acts linearly are generally ambiguous. Namely, under the action of H_{x_0} , the manifold \mathcal{M} splits into a disjoint union of orbits. The action of H_{x_0} will remain linear if we rescale the coordinates Ψ^i by an arbitrary H_{x_0} -invariant function on \mathcal{M} , that is a function which takes a constant value on any orbit of H_{x_0} . The only constraint is that such a rescaling leads to a well-defined set of new coordinates. The following example provides a nontrivial illustration of this ambiguity.

Example 7.7

Consider the action of $G \simeq SU(2) \times SU(2)$ on $\mathcal{M} \simeq SU(2)$. As already hinted in Example 5.5 and explained in detail in Sect. 9.1, this is important for a low-energy EFT description of hadron physics. For a given $\mathcal{U} \in \mathcal{M}$ and a given element $(g_L, g_R) \in G$, the action is defined by

$$\mathcal{U} \rightarrow g_L \mathcal{U} g_R^{-1} . \quad (7.15)$$

The isotropy group of $\mathcal{U}_0 \equiv \mathbb{1}$ is the “diagonal” subgroup of G , $H_{\mathcal{U}_0} \simeq \text{SU}(2)$, consisting of elements of the type (g, g) , that is $g_L = g_R = g$. It is easy to guess a triplet of coordinates ψ^i , parameterizing \mathcal{M} in the vicinity of \mathcal{U}_0 , on which $H_{\mathcal{U}_0}$ acts linearly. Some common choices are

$$\mathcal{U} = e^{i\psi \cdot \tau}, \quad \mathcal{U} = \frac{\mathbb{1} + \frac{i}{2}\psi \cdot \tau}{\mathbb{1} - \frac{i}{2}\psi \cdot \tau}, \quad \mathcal{U} = \mathbb{1}\sqrt{1 - \psi^2} + i\psi \cdot \tau, \quad (7.16)$$

where τ is the vector of Pauli matrices. All these choices coincide to linear order when expanded in powers of ψ^i , $\mathcal{U} = \mathbb{1} + i\psi \cdot \tau + \mathcal{O}(\psi^2)$. All of them are mutually connected by coordinate redefinitions of the type $\psi'^i = \psi^i f(\psi^2)$, where f is a suitably chosen function. The isotropy group $H_{\mathcal{U}_0}$ acts on \mathcal{U} in all parameterizations shown in (7.16) via rotations of ψ . Hence ψ^2 is invariant under the action of $H_{\mathcal{U}_0}$ and any function $f(\psi^2)$ is constant on the orbits of $H_{\mathcal{U}_0}$, as expected.

7.2.2 From Linear Representation to Nonlinear Realization

The ambiguity in the choice of coordinates that linearize the action of H_{x_0} around a chosen point $x_0 \in \mathcal{M}$ can be used to complete the classification of group actions. First of all, note that the set $\{T_{g,x_0} | g \in G\}$ defines a submanifold of \mathcal{M} . On this submanifold, G acts transitively; it is thus equivalent to the coset space G/H_{x_0} . We can choose a set of coordinates π^a , $a = 1, \dots, \dim G/H_{x_0}$, on it. Then we add another set of coordinates χ^ϱ , $\varrho = 1, \dots, \dim \mathcal{M} - \dim G/H_{x_0}$, so that (π^a, χ^ϱ) together is a well-defined coordinate system on \mathcal{M} in the vicinity of x_0 . In this coordinate system, the coset space G/H_{x_0} is embedded in \mathcal{M} as the set of points $(\pi^a, 0)$.

As the next step, we subject the coordinates (π^a, χ^ϱ) to the linearization (7.10); with a slight abuse of notation, I will use the same symbols (π^a, χ^ϱ) for the resulting new coordinates. Importantly, the condition $\chi^\varrho = 0$ is preserved by the procedure. We thus end up with a set of coordinates (π^a, χ^ϱ) in which the isotropy group H_{x_0} is represented by linear transformations, and moreover the subset $(\pi^a, 0)$, parameterizing G/H_{x_0} , is invariant under the action of G . The latter implies that the representation of H_{x_0} on the coordinates (π^a, χ^ϱ) has an invariant subspace, that is, it is reducible. I now use once again the assumption that H_{x_0} is compact. This ensures that the representation of H_{x_0} is *completely* reducible. The action of H_{x_0} can then be brought to a block-diagonal form,

$$T_h(\pi, \chi) \equiv (\pi'^a, \chi'^\varrho) = (D^{(\pi)}(h)^a_b \pi^b, D^{(\chi)}(h)^\varrho_\sigma \chi^\sigma), \quad h \in H_{x_0}, \quad (7.17)$$

by orthogonalization that leaves the subspace $(\pi^a, 0)$ intact. The latter feature ensures that we can still use the coordinates π^a to parameterize the submanifold G/H_{x_0} .

A given point $x \in G/H_{x_0}$ with coordinates π^a can be reached from x_0 by the action of any element of G that lies in the coset of x . To proceed, we need to choose

a concrete representative element of the coset. In line with the notation common in theoretical physics, I will denote this coset representative as $U(\pi)$. The concrete choice of the representative can be made fairly arbitrarily. There are however some natural requirements that will make our life easier:

- $U(\pi)$ should be a smooth function of π^a near the origin x_0 .
- The origin x_0 itself, i.e. the trivial coset eH_{x_0} , should be represented by $U(0) = e$.
- The choice of $U(\pi)$ should reflect the linearity of the action of H_{x_0} in the coordinates π^a .

The first two constraints can always be satisfied. As to the third, the linearity of the action of H_{x_0} requires that $T_h x = T_h T_{U(\pi)} x_0 = T_{U(\pi') } x_0$ where π'^a are linear in π^a . Given that $T_h x = T_h T_{U(\pi)} T_{h^{-1}} x_0 = T_{h U(\pi) h^{-1}} x_0$, it is natural to pick $U(\pi)$ so that

$$U(\pi') = h U(\pi) h^{-1} . \quad (7.18)$$

We need to make sure, however, that this can be done consistently.

Consider the Lie algebra \mathfrak{g} of G , which carries an adjoint action of H_{x_0} ,

$$Q \rightarrow h Q h^{-1} , \quad Q \in \mathfrak{g} , \quad h \in H_{x_0} . \quad (7.19)$$

This defines a linear representation of H_{x_0} which has an invariant subspace, namely the Lie algebra \mathfrak{h} of H_{x_0} . Thus, the representation is reducible. Suppose now that it is, in fact, completely reducible. This is certainly the case when H_{x_0} is compact. More generally, coset spaces G/H_{x_0} for which the Lie algebra \mathfrak{g} can be split as $\mathfrak{g} \simeq \mathfrak{h} \oplus \mathfrak{g}/\mathfrak{h}$ where both \mathfrak{h} and $\mathfrak{g}/\mathfrak{h}$ are invariant subspaces under the adjoint action of H_{x_0} (7.19), are called *reductive*. For a reductive coset space, we can always choose $U(\pi)$ so that the linear action of H_{x_0} on G/H_{x_0} is realized by (7.18). We can for instance set $U(\pi) = \exp(i\pi^a Q_a)$, where $Q_{a,b,\dots}$ is a basis of $\mathfrak{g}/\mathfrak{h}$. This is however by far not the only choice, as illustrated by Example 7.7.

Example 7.8

There are many examples of reductive coset spaces for which the isotropy group H is not compact. One can for instance start with a compact H , for which the reductive property is guaranteed, and then switch to a related noncompact Lie group. For a concrete example, take as G the Poincaré group of isometries of D -dimensional Minkowski spacetime and as H its subgroup, the Lorentz group $\text{SO}(d, 1)$. The subspace $\mathfrak{g}/\mathfrak{h}$ can be spanned on the generators of spacetime translations, which carry the vector representation of the Lorentz group. This coset space is a cousin of the Euclidean space $\mathbb{R}^D \simeq \text{ISO}(D)/\text{SO}(D)$, for which $H \simeq \text{SO}(D)$ is compact.

It is easy to promote the action (7.18) of H_{x_0} on G/H_{x_0} to an action of the whole group G . Indeed, the action of any $g \in G$ on an element $x = T_{U(\pi)} x_0$ of G/H_{x_0} is completely characterized by the product $gU(\pi)$. The latter can be, at least in the vicinity of the unit element, uniquely decomposed as $U(\pi')h(\pi, g)$, where π'^a is

defined by $T_g x = T_{U(\pi')} x_0$ and the factor $h(\pi, g) \in H_{x_0}$ ensures that the correct representative of the coset $T_g x$ is used.

In order to lift the action of G from the submanifold G/H_{x_0} to the whole of \mathcal{M} , we now make one last change of coordinates. Namely, we define new coordinates $\tilde{\pi}^a$ and $\tilde{\chi}^e$ by the requirement that the point $(\pi^a, \chi^e) \in \mathcal{M}$ can be expressed as

$$(\pi^a, \chi^e) = T_{U(\tilde{\pi})}(0, \tilde{\chi}) . \quad (7.20)$$

This notation can be intuitively thought of as defining a slicing of \mathcal{M} by orbits of G , starting from the subset of points $(0, \tilde{\chi}^e)$. The submanifold G/H_{x_0} is a special case corresponding to $\tilde{\chi}^e = 0$. In order to see that $\tilde{\pi}^a, \tilde{\chi}^e$ in fact are well-defined coordinates on \mathcal{M} , note that (7.20) defines uniquely π^a, χ^e for given $\tilde{\pi}^a, \tilde{\chi}^e$. It follows from the fact that $U(0) = e$ that if $\tilde{\pi}^a = 0$, then $\pi^a = 0$ and $\chi^e = \tilde{\chi}^e$. Likewise, it follows from the definition of coordinates on the submanifold G/H_{x_0} that if $\tilde{\chi}^e = 0$, then $\chi^e = 0$ and $\pi^a = \tilde{\pi}^a$. Hence the Jacobian matrix $\partial(\pi, \chi)/\partial(\tilde{\pi}, \tilde{\chi})$ equals $\mathbb{1}$ at the origin, and (7.20) defines a valid coordinate system in some neighborhood of x_0 .

The linearity of the action of H_{x_0} is preserved in the new coordinates since

$$T_h T_{U(\tilde{\pi})}(0, \tilde{\chi}) = T_{hU(\tilde{\pi})h^{-1}} T_h(0, \tilde{\chi}) = T_{hU(\tilde{\pi})h^{-1}}(0, D^{(\chi)}(h)\tilde{\chi}) \quad (7.21)$$

for any $h \in H_{x_0}$. It is now a matter of a short manipulation to see that the action of an arbitrary $g \in G$ on (7.20) is already completely fixed,

$$\begin{aligned} T_g T_{U(\tilde{\pi})}(0, \tilde{\chi}) &= T_{gU(\tilde{\pi})}(0, \tilde{\chi}) = T_{U(\tilde{\pi}')h(\tilde{\pi}, g)}(0, \tilde{\chi}) = T_{U(\tilde{\pi}')} T_{h(\tilde{\pi}, g)}(0, \tilde{\chi}) \\ &= T_{U(\tilde{\pi}')} (0, D^{(\chi)}(h(\tilde{\pi}, g))\tilde{\chi}) . \end{aligned} \quad (7.22)$$

This is the end of the line. Equation (7.22) shows that in some neighborhood of a chosen point $x_0 \in \mathcal{M}$, the action of any Lie group G can by a change of coordinates be brought to a “standard form” such that (dropping the tildes)

$$U(\pi) \rightarrow U(\pi') = gU(\pi)h(\pi, g)^{-1} , \quad \chi^e \rightarrow \chi'^e = D^{(\chi)}(h(\pi, g))^\sigma_\sigma \chi^\sigma , \quad (7.23)$$

where $h(\pi, g) \in H_{x_0}$ and $D^{(\chi)}$ is a matrix representation of H_{x_0} . The only technical assumption that was required in the proof was that the isotropy group H_{x_0} is compact. This is the main result of the chapter. We can now harvest the fruits of our labors.

7.3 Standard Realization of Symmetry

Having in mind that the reader might have skipped the last, somewhat technical section, let me give here a brief but self-contained summary of its main result. Consider a manifold \mathcal{M} equipped with an action of a group G , and let us choose a fixed point $x_0 \in \mathcal{M}$. Suppose that the isotropy group H_{x_0} of x_0 is compact. Then it is always possible to redefine coordinates in a neighborhood of x_0 so that the new,

“standard” coordinates (π^a, χ^ϱ) have the following properties. First, the point x_0 itself corresponds to the origin $(0, 0)$. The subset $(\pi^a, 0)$ spans a submanifold of \mathcal{M} , equivalent to the coset space G/H_{x_0} . Every point on the coset space can be uniquely characterized by a choice of a representative element $U(\pi)$ of the corresponding coset. The representative $U(\pi)$ can be chosen so that $U(0) = e$, and that the adjoint action of $h \in H_{x_0}$, $U(\pi) \rightarrow hU(\pi)h^{-1}$, defines a linear transformation of the coordinates π^a . The group G acts on the coset space via left multiplication, which defines implicitly an element $h(\pi, g)$ of H_{x_0} through

$$gU(\pi) = U(\pi'(\pi, g))h(\pi, g), \quad g \in G. \quad (7.24)$$

The last two properties of $U(\pi)$ imply that $h(\pi, g) = g$ for any $g \in H_{x_0}$.

The action of an element $g \in G$ on the whole manifold \mathcal{M} is now defined in terms of the standard coordinates (π^a, χ^ϱ) as

$$\begin{aligned} U(\pi) &\xrightarrow{g} U(\pi'(\pi, g)) = gU(\pi)h(\pi, g)^{-1}, \\ \chi^\varrho &\xrightarrow{g} \chi'^\varrho(\chi, \pi, g) = D(h(\pi, g))^\varrho_{\sigma} \chi^\sigma, \end{aligned} \quad (7.25)$$

where the matrices $D(h)$ define some linear representation of H_{x_0} . Altogether, the action of G is fully specified by the choice of coset representative $U(\pi)$, which fixes the first line of (7.25), and the choice of representation $D(h)$ of H_{x_0} , which fixes the second line thereof.

The above construction of the standard realization of group action goes through without change even if H_{x_0} is noncompact provided the coset space G/H_{x_0} is reductive. In that case, however, the line of argument in Sect. 7.2 fails and it is no longer guaranteed that the standard realization is unique up to a coordinate redefinition. There may then be more mutually inequivalent nonlinear realizations of the group G on the manifold \mathcal{M} , of which the standard realization is but one example.

Finally, one may try to follow the same steps of the construction of the standard realization even when the coset space G/H_{x_0} is not reductive. Then, however, many of the simple features of the standard realization are lost. A concrete example of a nonreductive coset space is worked out in [11].

In the standard nonlinear realization (7.25), the coordinates π^a transform under G on their own, independently of χ^ϱ . The transformation of the latter, on the other hand, is nonlinear in π^a but linear in χ^ϱ themselves. It is therefore possible to set the χ^ϱ s to zero consistently. This is not surprising. In the field theory language, the coordinates π^a correspond to NG bosons and their universal presence therefore mirrors the Goldstone theorem as reviewed in Chap. 6. The remaining coordinates on \mathcal{M} , χ^ϱ , represent other degrees of freedom that are not of NG nature. In the jargon of EFT, they are usually called *matter fields*. Among all the nonlinear realizations of the given symmetry group G with the given subgroup H , there is therefore a “minimal”

nonlinear realization, defined on the coset space G/H , which only includes the NG degrees of freedom. On a general manifold \mathcal{M} , additional matter fields may be present. Throughout this book, I will focus mostly on minimal nonlinear realizations due to their significance for low-energy EFT description of broken symmetries.

Example 7.9

One can gain insight into the standard realization (7.25) of the action of G by looking at some special choices of the isotropy group. If there is a point $x_0 \in \mathcal{M}$ such that $H_{x_0} \simeq G$, then in its vicinity, we do not have any coordinates π^a . The coset space G/H_{x_0} consists of a single point that we can represent with $U = e$, which is consistent with the first line of (7.25) if $h(g) = g$ for all $g \in G$. All coordinates in the neighborhood of x_0 are of the χ^e type. The second line of (7.25) guarantees that the action of G can be completely linearized, $\chi'^e(\chi, g) = D(g)^e_{\sigma} \chi^{\sigma}$. This is the usual textbook realization of symmetry via a linear representation.

The opposite extreme, $H_{x_0} \simeq \{e\}$, is more interesting. Here (7.25) reduces to

$$U(\pi) \xrightarrow{g} gU(\pi), \quad \chi^e \xrightarrow{g} \chi^e. \quad (7.26)$$

The coset space G/H_{x_0} corresponds to the group manifold G and carries an action of G defined by simple left multiplication. Whatever other coordinates χ^e on \mathcal{M} are present can always be chosen to be invariant under G . This is quite surprising. We are used to working with fields spanning linear multiplets of G ; it is not obvious that the same physical content can be encoded in a set of fields that do not transform under G at all. The resolution of this apparent paradox lies in the freedom to choose coordinates at will. Namely, if we start with a set of fields Ψ^e transforming under G as $\Psi^e \xrightarrow{g} D(g)^e_{\sigma} \Psi^{\sigma}$, we can make the redefinition $\chi^e \equiv D(U(\pi)^{-1})^e_{\sigma} \Psi^{\sigma}$. The new variables χ^e are invariant under G in accord with (7.26).

7.3.1 Nonlinear Realization on Coset Spaces

Can we be more explicit about the way that the coordinates (π^a, χ^e) transform under the action of G ? The standard realization (7.25) of the group action requires the knowledge of the nonlinear functions $\pi'^a(\pi, g)$ and $h(\pi, g)$. We would like to be able to compute these, at least for small transformations, that is for $g \in G$ infinitesimally close to unity.⁴ To that end, it is sufficient to consider the action of G on the coset space G/H ; from now on will I drop the subscript x_0 on H unless it is needed to explicitly distinguish the isotropy groups of different points on the coset

⁴ I will only introduce the concept of a metric on a group, and more generally on a homogeneous space, in Sect. 7.4. Statements about infinitesimal distance of group elements should therefore be interpreted within a faithful matrix representation of the group using some standard matrix norm. The same remark applies whenever a sum or difference of group elements is considered below.

space. Once the minimal realization of G on G/H is found, it can be extended to any other manifold by specifying the linear representation $D(h)$ of H under which the additional coordinates χ^e transform.

The general algorithm for calculation of the desired functions $\pi'^a(\pi, g)$ and $h(\pi, g)$ is as follows. Take the first line of (7.25), multiply it with $U(\pi)^{-1}$, and subtract $U(\pi)^{-1}U(\pi) = e$. This gives the master relation

$$U(\pi)^{-1}\delta U(\pi) = U(\pi)^{-1}gU(\pi)h(\pi, g)^{-1} - e, \quad (7.27)$$

where $\delta U(\pi) \equiv U(\pi') - U(\pi)$. For $g \in G$ that is infinitesimally close to unity, both $U(\pi)^{-1}gU(\pi)$ and $h(\pi, g)$ are infinitesimally close to unity as well. By a systematic comparison of the left- and right-hand sides of (7.27), one can then determine $\delta\pi^a(\pi, g) \equiv \pi'^a(\pi, g) - \pi^a$ as well as $h(\pi, g)$.

To make further progress, we first have to establish some notation. In order to be able to discuss different symmetry transformations from G separately, we choose a basis $Q_{A,B,\dots}$ of \mathfrak{g} . A subset of these spans a basis of the Lie subalgebra \mathfrak{h} and will be denoted with Greek indices, $Q_{\alpha,\beta,\dots}$. The rest of the generators spans the subspace $\mathfrak{g}/\mathfrak{h}$ and will be denoted as $Q_{a,b,\dots}$. The structure constants of the Lie algebra \mathfrak{g} will be called f_{AB}^C with a conventional factor of i in the commutation relations, that is, $[Q_A, Q_B] = if_{AB}^C Q_C$. The structure constant does not have to be fully antisymmetric in its three indices even if G is compact. It does have to be antisymmetric under the exchange of its lower two indices though. It also has to satisfy the Jacobi identity,

$$f_{AB}^E f_{EC}^D + f_{BC}^E f_{EA}^D + f_{CA}^E f_{EB}^D = 0. \quad (7.28)$$

The basic commutator of the Lie algebra \mathfrak{g} , $[Q_A, Q_B] = if_{AB}^C Q_C$, can be unfolded into three separate conditions on the subsets of generators $Q_{\alpha,\beta,\dots}$ and $Q_{a,b,\dots}$,

$$\begin{aligned} [Q_\alpha, Q_\beta] &= if_{\alpha\beta}^\gamma Q_\gamma, \\ [Q_\alpha, Q_b] &= if_{\alpha b}^c Q_c, \\ [Q_a, Q_b] &= i(f_{ab}^\gamma Q_\gamma + f_{ab}^c Q_c). \end{aligned} \quad (7.29)$$

The first of these encodes the requirement that the generators $Q_{\alpha,\beta,\dots}$ span a closed Lie algebra (that is $f_{\alpha\beta}^c = 0$). The second of these likewise expresses the assumption that the coset space G/H is reductive (that is $f_{ab}^\gamma = 0$).

Next, we are going to need two simple statements from linear algebra, which I reproduce here for the sake of completeness. The first of these is usually known under the name *Hadamard lemma*,

$$e^A B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]]] + \dots, \quad (7.30)$$

where A, B are arbitrary (square) matrices. The second identity, which to the best of my knowledge does not have an established name, reads

$$e^{-A}de^A = \int_0^1 dt e^{-tA}(dA)e^{tA} . \quad (7.31)$$

Here the symbol d acting on A and e^A can be thought of as a differential, but also as a derivative with respect to whatever parameter A might depend on. Once multiplied from the left with e^A , we can think of (7.31) as a continuous version of the Leibniz (product) rule, applied to e^A . It follows from (7.31) combined with the Hadamard lemma (7.30) that whenever A is a function with values in a Lie algebra, $e^{-A}de^A$ will take values in the same Lie algebra.

Equation (7.31) prepares the ground for the introduction of a concept of central importance for calculus on coset spaces: the *Maurer–Cartan* (MC) form,

$$\omega(\pi) \equiv -iU(\pi)^{-1}dU(\pi) . \quad (7.32)$$

For the time being, the reader may think of the d herein as an ordinary differential of a function. The true geometric significance of the MC form as a differential 1-form will become clear in Sect. 7.4. Recall that any element of a Lie group sufficiently close to unity, in our case $U(\pi)$, can be obtained as the exponential of an element of the corresponding Lie algebra. Equation (7.31) then implies that the MC form takes values in the Lie algebra \mathfrak{g} . We can split it into pieces that belong to \mathfrak{h} and $\mathfrak{g}/\mathfrak{h}$, and represent each of these in terms of their components in a chosen basis of generators,

$$\begin{aligned} \omega &\equiv \omega_{\parallel} + \omega_{\perp} , \\ \omega &\equiv \omega^A Q_A , \quad \omega_{\parallel} \equiv \omega^{\alpha} Q_{\alpha} , \quad \omega_{\perp} \equiv \omega^a Q_a . \end{aligned} \quad (7.33)$$

Finally, by writing $dU(\pi) = [\partial U(\pi)/\partial \pi^a]d\pi^a$, one can introduce explicit components of the MC form in a chosen set of local coordinates π^a on G/H , $\omega^A \equiv \omega_a^A d\pi^a$.

It is instructive to check how the MC form is affected by the action of G on the coset space. This follows directly from (7.25). It is a simple exercise to verify that for given $g \in G$,

$$\begin{aligned} \omega_{\parallel}(\pi) &\xrightarrow{g} \omega_{\parallel}(\pi'(\pi, g)) = h(\pi, g)\omega_{\parallel}(\pi)h(\pi, g)^{-1} - ih(\pi, g)dh(\pi, g)^{-1} , \\ \omega_{\perp}(\pi) &\xrightarrow{g} \omega_{\perp}(\pi'(\pi, g)) = h(\pi, g)\omega_{\perp}(\pi)h(\pi, g)^{-1} . \end{aligned} \quad (7.34)$$

Note how (7.31) guarantees that $\omega_{\parallel}(\pi')$ still takes values in the Lie algebra \mathfrak{h} . While not of direct relevance right here and now, the transformation rules (7.34) will help us understand the geometric meaning of the MC form in Sect. 7.4.

We still need a few last pieces of notation. Conjugation of elements of \mathfrak{g} by $U(\pi)$ will be abbreviated as

$$U(\pi)^{-1}Q_A U(\pi) \equiv v_A^B(\pi)Q_B , \quad (7.35)$$

which defines a set of nonlinear functions $v_A^B(\pi)$ on the coset space. Finally, for the action of an element $g \in G$ infinitesimally close to unity, I will use the notation

$$\begin{aligned}
g &\approx e + i\epsilon^A Q_A , \\
\delta\pi^a(\pi, g) &\approx \epsilon^A \xi_A^a(\pi) , \quad h(\pi, g) \approx e + i\epsilon^A k_A^a(\pi) Q_a .
\end{aligned} \tag{7.36}$$

The \approx symbol indicates that I have expanded all the quantities to linear order in the transformation parameters ϵ^A , defined by the first line of (7.36). The second line thereof introduces notation for the infinitesimal versions of the functions $\pi'^a(\pi, g)$ and $h(\pi, g)$. Of particular interest are the functions $\xi_A^a(\pi)$ that realize the motion induced on the coset space G/H by the group G .

With all the notation at hand, we can now expand (7.27) to linear order in ϵ^A and compare coefficients of the various generators of \mathfrak{g} on the left- and right-hand sides. This leads to the identities

$$\begin{aligned}
v_A^\alpha(\pi) &= \xi_A^a(\pi) \omega_a^\alpha(\pi) + k_A^\alpha(\pi) , \\
v_A^a(\pi) &= \xi_A^b(\pi) \omega_b^a(\pi) .
\end{aligned} \tag{7.37}$$

These are still valid for any choice of the coset representative $U(\pi)$. Once it is fixed, the functions $\omega_a^A(\pi)$ are determined by (7.32). Likewise, the functions $v_A^B(\pi)$ are fixed by (7.35). The identities (7.37) then constitute a set of linear equations for $\xi_A^a(\pi)$ and $k_A^a(\pi)$. At the origin, $v_A^B(0) = \delta_A^B$ as a consequence of the fact that $U(0) = e$. The second line of (7.37) then implies that $\omega_b^a(0)$ is nonsingular. By continuity, it must remain nonsingular in some neighborhood of the origin. This guarantees that a solution of (7.37) for $\xi_A^a(\pi)$ and $k_A^a(\pi)$ exists and it is unique.

I have now achieved the main goal of this subsection: to give an algorithm how, for a chosen set of coordinates π^a , to realize the action of the group G on the coset space G/H . None of the nonlinear functions involved— $\omega_a^A(\pi)$, $v_A^B(\pi)$, $\xi_A^a(\pi)$ and $k_A^a(\pi)$ —can however in general be evaluated in a closed form. For practical applications, it is useful to have explicit expressions for these functions, even if just as a series expansion in a specific set of coordinates π^a . One popular choice of parameterization for which this can easily be done is

$$U(\pi) = \exp(i\pi^a Q_a) . \tag{7.38}$$

The Hadamard lemma (7.30) then tells us at once that

$$v_B^A(\pi) = \delta_B^A - f_{Ba}^A \pi^a + \frac{1}{2} f_{Ba}^C f_{Cb}^A \pi^a \pi^b + O(\pi^3) . \tag{7.39}$$

Likewise, it follows quickly from (7.31) that

$$\omega_a^A(\pi) = \delta_a^A - \frac{1}{2} f_{ab}^A \pi^b + \frac{1}{6} f_{ab}^B f_{Bc}^A \pi^b \pi^c + O(\pi^3) . \tag{7.40}$$

Finally, (7.37) can be solved iteratively for the remaining pieces,

$$\begin{aligned}
\xi_A^\alpha(\pi) &= \delta_A^\alpha - \left(f_{Ab}^a - \frac{1}{2} \delta_A^e f_{eb}^a \right) \pi^b \\
&\quad + \frac{1}{2} \left(f_{Ab}^\alpha f_{\alpha c}^a - \frac{1}{3} \delta_A^e f_{eb}^B f_{Bc}^a + \frac{1}{2} \delta_A^e f_{eb}^d f_{dc}^a \right) \pi^b \pi^c + O(\pi^3), \\
k_A^\alpha(\pi) &= \delta_A^\alpha - \left(f_{Aa}^\alpha - \frac{1}{2} \delta_A^e f_{ea}^\alpha \right) \pi^a \\
&\quad + \frac{1}{2} \left(f_{Aa}^\beta f_{\beta b}^\alpha - \frac{1}{3} \delta_A^e f_{ea}^B f_{Bb}^\alpha + \frac{1}{2} \delta_A^e f_{ea}^d f_{db}^\alpha \right) \pi^a \pi^b + O(\pi^3).
\end{aligned} \tag{7.41}$$

With these explicit expressions at hand, it is easy to illustrate some of the general properties of the standard nonlinear realization. For instance, the action of the isotropy group H reduces to the linear adjoint transformation of the coordinates, $\xi_\alpha^a(\pi) = -f_{\alpha b}^a \pi^b$. Likewise, we find that $k_\beta^\alpha(\pi) = \delta_\beta^\alpha$, which is an infinitesimal version of the relation $h(\pi, g) = g$ for any $g \in H$.

Example 7.10

All the structure introduced above takes a particularly simple form when the group G is Abelian. Then $\xi_A^\alpha(\pi) = \delta_A^\alpha$: the group G acts on the standard coordinates π^a by a mere set of shifts. The MC form reduces to $\omega(\pi) = d\pi^a Q_a$. This suggests a simple interpretation of $\omega_\perp = \omega_b^a d\pi^b Q_a$ in the general situation when G is non-Abelian. Namely, ω^a supplies us with a generalized derivative (or differential) of π^a which, as (7.34) shows, is covariant under the action of G .

7.3.2 Symmetric Coset Spaces

Up to some technical assumptions, we have already accomplished a complete classification of possible nonlinear realizations of symmetry on manifolds. In the language of field theory, this amounts to finding all possible point symmetry transformations of a given set of fields under a given symmetry group G . The price we had to pay for the generality of this result was the rather complicated transformation rule (7.25). This involves nonlinear functions on the coset space G/H that cannot be evaluated in a closed explicit form. There is, however, an important class of coset spaces for which we can do much better.

The coset space G/H is called *symmetric*, and the associated homogeneous space is called a *symmetric space*, if the Lie algebra \mathfrak{g} admits an involutive automorphism under which \mathfrak{h} is even and $\mathfrak{g}/\mathfrak{h}$ is odd. In other words, we require that there is a linear map $\mathcal{R} : \mathfrak{g} \rightarrow \mathfrak{g}$ such that

$$\begin{aligned}
\mathcal{R}([Q_1, Q_2]) &= [\mathcal{R}(Q_1), \mathcal{R}(Q_2)], & Q_1, Q_2 \in \mathfrak{g}, \\
\mathcal{R}(Q) &= Q, & Q \in \mathfrak{h}, \\
\mathcal{R}(Q) &= -Q, & Q \in \mathfrak{g}/\mathfrak{h}.
\end{aligned} \tag{7.42}$$

This property guarantees the vanishing of any structure constant with an odd number of $\mathfrak{g}/\mathfrak{h}$ indices. In particular, $f_{\alpha\beta}^\gamma = 0$: any symmetric coset space is automatically reductive. In addition, $f_{ab}^c = 0$, that is the last term in the commutation relations (7.29) is missing.

Example 7.11

The fundamental commutation relation of the $\text{SO}(n)$ group reads

$$[J_{ij}, J_{kl}] = i(\delta_{ik}J_{jl} + \delta_{jl}J_{ik} - \delta_{il}J_{jk} - \delta_{jk}J_{il}) , \quad (7.43)$$

where $i, j, k, l = 1, \dots, n$ and J_{ij} is the antisymmetric tensor of angular momentum in n spatial dimensions. The Lie algebra of $\text{SO}(n)$ possesses an automorphism \mathcal{R} under which $\mathcal{R}(J_{\alpha\beta}) = J_{\alpha\beta}$ and $\mathcal{R}(J_{an}) = -J_{an}$, where $\alpha, \beta = 1, \dots, n-1$. One can think of this automorphism geometrically as an inversion of the n -th coordinate axis. The coset space $\text{SO}(n)/\text{SO}(n-1) \simeq S^{n-1}$ is therefore symmetric.

In a similar vein, the Euclidean space \mathbb{R}^n is a symmetric space. This can be seen most easily by recalling that $\mathbb{R}^n \simeq \text{ISO}(n)/\text{SO}(n)$. The desired automorphism \mathcal{R} is the spatial inversion. Under this, all the generators of $\text{SO}(n)$ (rotations) remain intact, whereas the remaining generators of $\text{ISO}(n)/\text{SO}(n)$ (translations) change sign.

Example 7.12

Consider the “chiral” coset spaces of the type $G_L \times G_R / G_V$, where all the three groups G_L, G_R, G_V are isomorphic to the same Lie group G ; this is a generalization of Example 7.7 where $G \simeq \text{SU}(2)$. The chiral group $G_L \times G_R$ consists of elements (g_L, g_R) where $g_L, g_R \in G$. The “vector” isotropy group G_V consists of elements of the type (g, g) , that is $g_L = g_R = g$. The generators of the chiral group include two copies, $Q_{L,A}$ and $Q_{R,A}$, of the generators of G . The Lie algebra of the chiral group is defined in terms of the structure constants f_{AB}^C of G by

$$\begin{aligned} [Q_{L,A}, Q_{L,B}] &= if_{AB}^C Q_{L,C} , & [Q_{R,A}, Q_{R,B}] &= if_{AB}^C Q_{R,C} , \\ [Q_{L,A}, Q_{R,B}] &= 0 . \end{aligned} \quad (7.44)$$

These commutation relations are invariant under the exchange of $Q_{L,A}$ and $Q_{R,A}$, which defines the desired automorphism,

$$\mathcal{R}(Q_{L,A}) = Q_{R,A} , \quad \mathcal{R}(Q_{R,A}) = Q_{L,A} . \quad (7.45)$$

The generators of G_V , equal to $Q_{L,A} + Q_{R,A}$ up to overall normalization, are even under this automorphism. The generators of the complementary space $\mathfrak{g}/\mathfrak{h}$, which are to be odd under \mathcal{R} , can be taken as $Q_{L,A} - Q_{R,A}$ up to an overall factor.

The automorphism \mathcal{R} of the Lie algebra \mathfrak{g} can be lifted, at least locally near the unit element, to the Lie group G . It is then possible to choose the coset representative $U(\pi)$ so that $\mathcal{R}(U(\pi)) = U(\pi)^{-1}$; one can use for instance the exponential parameterization (7.38). We now take the first line of (7.25) and multiply it with the inverse of its image under \mathcal{R} . The result is a surprise: for symmetric coset spaces, there is a parameterization of G/H in which the whole group G is realized linearly,

$$\Sigma(\pi) \equiv U(\pi)^2, \quad \Sigma(\pi) \xrightarrow{g} \Sigma(\pi'(\pi, g)) = g\Sigma(\pi)\mathcal{R}(g)^{-1}. \quad (7.46)$$

This is of such utility that whenever one deals with a symmetric coset space, one almost always uses the linearly transforming variable Σ instead of working directly with the coordinates π^a . Note, however, that one may need the coordinates π^a if one wishes to extend the coset space G/H to a larger manifold \mathcal{M} . This is because the transformation of χ^e in (7.25) requires the functions $h(\pi, g)$ that depend on π^a .

With the automorphism \mathcal{R} at hand, one may easily project out the ω_{\parallel} and ω_{\perp} components of the MC form,

$$\omega_{\parallel} = \frac{1}{2}[\omega + \mathcal{R}(\omega)], \quad \omega_{\perp} = \frac{1}{2}[\omega - \mathcal{R}(\omega)]. \quad (7.47)$$

The latter has a practically convenient expression in terms of Σ ,

$$\omega_{\perp} = -\frac{i}{2}U^{-1}d\Sigma U^{-1} = \frac{i}{2}Ud\Sigma^{-1}U. \quad (7.48)$$

Example 7.13

Let us see how the variable $\Sigma(\pi)$ is realized on the chiral coset spaces $G_L \times G_R / G_V$ discussed in Example 7.12. Here we can choose the coset representative as $U = (u, u^{-1})$ where $u \in G$; this satisfies the requirement that under the automorphism $\mathcal{R}(g_L, g_R) = (g_R, g_L)$ of the chiral group $G_L \times G_R$, U is turned into its inverse. The general transformation rule as given by the first line of (7.25) then translates to

$$(u, u^{-1}) \xrightarrow{(g_L, g_R)} (g_L, g_R)(u, u^{-1})(h^{-1}, h^{-1}), \quad (7.49)$$

where $h \in G$. The linearly transforming variable (7.46) is given by $\Sigma = U^2 = (u^2, u^{-2})$, where $u^2 \equiv \mathcal{U}$ transforms under $G_L \times G_R$ as $\mathcal{U} \rightarrow g_L \mathcal{U} g_R^{-1}$.

It turns out that (7.46) can be further simplified in case \mathcal{R} is an inner automorphism of G , that is, when there is an element $R \in G$ such that

$$\mathcal{R}(g) = R^{-1}gR, \quad g \in G. \quad (7.50)$$

Then a slight modification of (7.46) leads to a variable that transforms linearly under the adjoint action of G ,

$$N(\pi) \equiv \Sigma(\pi)R = U(\pi)RU(\pi)^{-1}, \quad N(\pi) \xrightarrow{g} N(\pi'(\pi, g)) = gN(\pi)g^{-1}. \quad (7.51)$$

An advantage of trading $\Sigma(\pi)$ for $N(\pi)$ is that $N(\pi)^2$ is a constant independent of π^a , which may be convenient in concrete applications. Indeed, since $\mathcal{R}(\mathcal{R}(g)) = g$ for any $g \in G$, it follows from (7.50) that R^2 belongs to the center of G . Consequently, $N(\pi)^2 = U(\pi)R^2U(\pi)^{-1} = R^2$.

Example 7.14

Consider the coset space $SU(2)/U(1)$, relevant for description of quantum systems with magnetic ordering such as (anti)ferromagnets. In the fundamental representation, the generators of $SU(2)$ are $Q_A = \tau_A/2$, and the generator of the $U(1)$ isotropy group can be taken as $\tau_3/2$. This coset space is symmetric thanks to an inner automorphism (7.50) with $R = i\tau_3$. While the factor of i here is required to make R an element of $SU(2)$, we do not need it for the definition of $N(\pi)$. Let us therefore set

$$N(\pi) \equiv U(\pi)\tau_3U(\pi)^{-1}. \quad (7.52)$$

This matrix variable is unitary and Hermitian. Moreover, it is traceless and it squares to $\mathbb{1}$. It can thus be mapped on a unit vector variable $\mathbf{n}(\pi)$ such that

$$N(\pi) = \boldsymbol{\tau} \cdot \mathbf{n}(\pi), \quad (7.53)$$

which belongs to the vector representation of $SU(2)$. The fact that we ended up describing the coset space in terms of a unit vector is not a coincidence. Thanks to the local isomorphism of $SU(2)$ and $SO(3)$ we also have a local equivalence of coset spaces, $SU(2)/U(1) \simeq SO(3)/SO(2) \simeq S^2$.

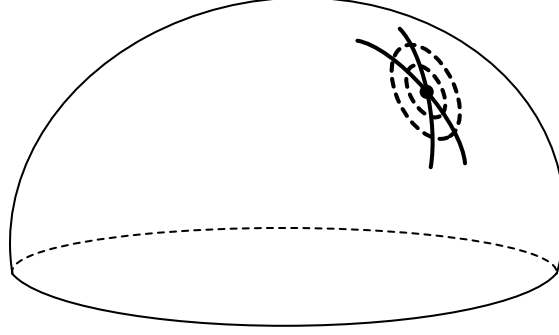
7.4 Geometry of the Coset Space

In this final section of the chapter, I will show that some of the structure that enters the standard realization of symmetry on a coset space can be given a neat geometric interpretation. A reader not familiar with basics of differential geometry is advised to consult Appendix A before proceeding. Further information about the geometry of homogeneous spaces at an easily accessible level can be found in [12].

To start, recall that each point of a coset space G/H corresponds to an entire class of elements of the group G . The approach I have used so far was to parameterize each $x \in G/H$ with coordinates π^a in terms of a fixed coset representative $U(\pi)$. The choice of $U(\pi)$ is however arbitrary and can be changed locally by multiplication from the right; any $\tilde{U}(\pi) = U(\pi)h(\pi)$ with $h(\pi) \in H_{x_0}$ is equally good. One can then promote the basic transformation rule (7.25) for $U(\pi)$ to

$$U(\pi) \xrightarrow{g, h(\pi)} U(\pi'(\pi, g)) = gU(\pi)h(\pi)^{-1}, \quad (7.54)$$

Fig. 7.2 Illustration of the action of G and H_x around a fixed point x (black dot) on the coset space. The thick solid lines indicate orbits of one-parameter subgroups of G passing through x ; these correspond to generators from $\mathfrak{g}/\mathfrak{h}$ and act as translations on the coset space. The thick dashed lines indicate some orbits of H_x in the vicinity of x . The action of H_x naturally induces a set of linear transformations on the tangent space at x



where $h(\pi)$ is independent of g . This realizes an action of the product group $G \times H_{\text{gauge}}$. The group G acts on $U(\pi)$ by left multiplication as usual. The local group H_{gauge} isomorphic to H acts on $U(\pi)$ by right multiplication with $h(\pi)^{-1}$ and encodes the freedom to choose locally the coset representative. This is a typical example of a gauge redundancy; any geometrically or physically well-defined quantity must be independent of the arbitrary choice of coset representative. In theoretical physics, the approach that views the left action of G separately from the right action of the local group H_{gauge} is called “hidden local symmetry” [13].

The different geometric roles of G and its isotropy subgroup can be intuitively understood by looking at Fig. 7.2. One-parameter subgroups of G define a set of flows on G/H that in general translate a given point x to some other point. The isotropy group H_x maps x to itself. It does, however, act nontrivially in the vicinity of x . Based on the figure, we expect that the action of H_x projects to a set of linear maps on the tangent space $\mathcal{T}_x G/H$ at x . By induction, such linear H_x -transformations should exist for any tensor at x .

The local group H_{gauge} gives us the freedom to do independent H -transformations of tangent vectors (and generally tensors) point by point. In order to be able to analyze the properties of vector (tensor) fields on G/H , we therefore need a way to relate tangent vectors (tensors) at different points. This resembles closely the motivation behind the construction of a local frame, or vielbein, and the corresponding connection on a manifold. The only difference to the formalism reviewed in Appendix A.5 is that here the structure group is not $\text{GL}(\dim G/H)$ but H itself. This restriction makes it possible to realize changes in the local frame in terms of a fixed matrix representation of H .

It turns out that we already have both the local frame and the connection: they are granted to us by the MC form (7.32). Specifically, the components ω^a of ω_\perp define a coframe on G/H . By the second line of (7.34), these indeed transform linearly under the local action of H_{gauge} as they should. Likewise, the \mathfrak{h} part of the MC form, ω_\parallel , provides the necessary H -connection. This is confirmed by the transformation rule on the first line of (7.34).

This observation stresses the significance of the MC form as a differential 1-form, taking values in the Lie algebra \mathfrak{g} . Its exterior derivative reflects the structure of G ,

$$\begin{aligned} d\omega &= -i dU^{-1} \wedge dU = iU^{-1} dU \wedge U^{-1} dU = -i\omega \wedge \omega \\ &= -\frac{i}{2} \omega^B \wedge \omega^C [Q_B, Q_C] = \frac{1}{2} f_{BC}^A \omega^B \wedge \omega^C Q_A . \end{aligned} \quad (7.55)$$

This can be split into a pair of equations, one for ω_{\parallel} and one for ω_{\perp} ,

$$\begin{aligned} d\omega^{\alpha} &= \frac{1}{2} f_{\beta\gamma}^{\alpha} \omega^{\beta} \wedge \omega^{\gamma} + \frac{1}{2} f_{bc}^{\alpha} \omega^b \wedge \omega^c , \\ d\omega^a &= f_{\beta c}^a \omega^{\beta} \wedge \omega^c + \frac{1}{2} f_{bc}^a \omega^b \wedge \omega^c , \end{aligned} \quad (7.56)$$

where I used the facts that $f_{\beta\gamma}^a = 0$ (the Lie algebra \mathfrak{h} closes) and that $f_{\beta c}^{\alpha} = 0$ (the coset space is reductive). These so-called *Maurer–Cartan equations* provide a link between the algebraic and geometric properties of the coset space.

7.4.1 Canonical and Torsion-Free Connection

The local basis of 1-forms ω^a transforms under the action of the structure group H as a tangent vector. It follows from (7.34) that upon an infinitesimal transformation by $h \approx e + i\epsilon^{\alpha} Q_{\alpha}$, $\delta\omega^a = -\epsilon^{\alpha} f_{\alpha b}^a \omega^b$. This determines the matrix elements of the action of the H -connection ω_{\parallel} on tangent vectors, in the notation of Appendix A.5,

$$\Omega^a_b = -f_{\alpha b}^a \omega^{\alpha} . \quad (7.57)$$

The MC equations (7.56) encode information about both the torsion and the curvature of this so-called *canonical connection* on G/H . The torsion 2-form follows from the second line of (7.56),

$$T^a \equiv d\omega^a + \Omega^a_b \wedge \omega^b = \frac{1}{2} f_{bc}^a \omega^b \wedge \omega^c . \quad (7.58)$$

Similarly, the curvature 2-form follows from the first line of (7.56),

$$R^a_b \equiv d\Omega^a_b + \Omega^a_c \wedge \Omega^c_b = -\frac{1}{2} f_{\alpha b}^a f_{cd}^{\alpha} \omega^c \wedge \omega^d , \quad (7.59)$$

where I used the Jacobi identity (7.28) to simplify the result. In the language of field theory, this is nothing but the field-strength 2-form of ω_{\parallel} .

The definition (7.57) of the canonical connection arises naturally from the splitting (7.33) of the MC form into the \mathfrak{h} and $\mathfrak{g}/\mathfrak{h}$ subspaces. This is not the only possible connection on G/H though. In fact, there is an infinite class of them. Let us set

$${}^{\lambda}\Omega^a_b \equiv -f_{\alpha b}^a \omega^{\alpha} - \lambda f_{cb}^a \omega^c , \quad \lambda \in \mathbb{R} . \quad (7.60)$$

This time, it takes some manipulation to derive the corresponding torsion and curvature 2-forms. The final result is

$$\begin{aligned}\lambda T^a &= \left(\frac{1}{2} - \lambda\right) f_{bc}^a \omega^b \wedge \omega^c, \\ \lambda R^a{}_b &= -\frac{1}{2} (f_{ab}^a f_{cd}^a + \lambda f_{eb}^a f_{cd}^e - 2\lambda^2 f_{ec}^a f_{bd}^e) \omega^c \wedge \omega^d,\end{aligned}\tag{7.61}$$

which generalizes (7.58) and (7.59) to any nonzero λ .

Within the class of connections (7.60), there is one with vanishing torsion, corresponding to $\lambda = 1/2$. One might expect that it should be possible to recover this connection from a suitable Riemannian metric on the coset space. This is indeed the case under some further technical assumptions, as I will explain below. For symmetric coset spaces, the whole class of connections (7.60) becomes degenerate, and is automatically torsion-free.

7.4.2 Riemannian Metric

Every Lie algebra \mathfrak{g} possesses a bilinear form invariant under the adjoint action of the corresponding Lie group G . To see this, just take any faithful matrix representation of the generators Q_A and set

$$g_{AB} \equiv \text{tr}(Q_A Q_B) .\tag{7.62}$$

The invariance under $Q_A \rightarrow g Q_A g^{-1}$ for any $g \in G$ is manifest. The infinitesimal version of the invariance condition follows by setting $g \approx e + i\epsilon^C Q_C$ and expanding to linear order in ϵ^C ,

$$f_{CA}^D g_{DB} + f_{CB}^D g_{AD} = 0 .\tag{7.63}$$

Since the components ω^a of the MC form define a basis of the space of 1-forms on G/H , any rank-2 covariant tensor can be constructed as a linear combination of $\omega^a \otimes \omega^b$. Given the invariant bilinear form g_{AB} , it is then natural to introduce the following metric on G/H ,

$$g_{G/H} \equiv g_{ab} \omega^a \otimes \omega^b ,\tag{7.64}$$

where g_{ab} is the restriction of g_{AB} to the $\mathfrak{g}/\mathfrak{h}$ subspace.

This construction is not guaranteed to work without further assumptions. A Riemannian metric should be positive-definite. A sufficient, though not necessary, condition for this is that g_{AB} itself is positive-definite, which is generally only true for compact semisimple Lie algebras. For what follows, I will only need the weaker assumption that g_{AB} is nondegenerate and that $g_{a\beta} = 0$, i.e. that the subspace $\mathfrak{g}/\mathfrak{h}$ can be chosen to be “orthogonal” to \mathfrak{h} . Interestingly, this alone already ensures that G/H is a reductive coset space. Namely, a short manipulation using (7.63) gives

$$0 = f_{\alpha\beta}^{\delta} g_{\delta c} = f_{\alpha\beta}^D g_{Dc} = -f_{\alpha c}^D g_{\beta D} = -f_{\alpha c}^{\delta} g_{\beta \delta} , \quad (7.65)$$

which implies that $f_{\alpha c}^{\delta} = 0$ thanks to the fact that $g_{\alpha\beta}$ is nondegenerate. With the assumption that $g_{\alpha\beta} = 0$, the invariance condition (7.63) also splits into two separate conditions on g_{ab} ,

$$f_{\gamma a}^d g_{db} + f_{\gamma b}^d g_{ad} = 0 , \quad f_{ca}^d g_{db} + f_{cb}^d g_{ad} = 0 . \quad (7.66)$$

The metric (7.64) is invariant under the left action of G , since the MC form itself is. It is however also invariant under the right action of H_{gauge} as defined by (7.34). This follows from the first condition in (7.66). The G -invariance of the metric guarantees the existence of a set of Killing vector fields that realize infinitesimal group motions on the coset space. In the local coordinates π^a , these are nothing but the functions $\xi_A^a(\pi)$ introduced in (7.36). According to (7.37), we have $\xi_A^b(0)\omega_b^a(0) = \delta_A^a$ at the origin. This means that the Killing vectors $\xi_a(0)$ corresponding to generators from $\mathfrak{g}/\mathfrak{h}$ define a local frame dual to $\omega^a(0)$, which further illuminates the geometric nature of the MC form. Away from the origin, the duality between ω_{\perp} and the subset of Killing vectors realizing infinitesimal translations on G/H is still expressed by the second line of (7.37). One just has to recall that the isotropy group of $x = T_{U(\pi)x_0}$ is $H_x = U(\pi)H_{x_0}U(\pi)^{-1}$. This conjugation is supplied by the matrix $\nu_A^B(\pi)$ on the left-hand side of (7.37). The local basis of 1-forms $\omega^a(\pi)$ is then dual to the local frame consisting of Killing vectors of the generators $U(\pi)Q_aU(\pi)^{-1}$.

Example 7.15

The metric (7.64) is particularly easy to evaluate explicitly on symmetric coset spaces. Using the expression (7.48) for ω_{\perp} in terms of the linearly transforming variable Σ , we find at once that

$$g_{G/H} = \frac{1}{4} \text{tr}(d\Sigma \otimes d\Sigma^{-1}) = \frac{1}{4} \text{tr}(\partial_a \Sigma \partial_b \Sigma^{-1}) d\pi^a \otimes d\pi^b , \quad (7.67)$$

where I used the abbreviation $\partial_a \equiv \partial/\partial\pi^a$. As an illustration, consider the coset space $\text{SU}(2)/\text{U}(1) \simeq \text{SO}(3)/\text{SO}(2) \simeq S^2$ discussed in Example 7.14. Here we find that up to an overall factor,

$$g_{S^2} \propto d\mathbf{n} \otimes d\mathbf{n} = d\theta \otimes d\theta + \sin^2 \theta d\varphi \otimes d\varphi , \quad (7.68)$$

where I used standard spherical coordinates to parameterize the unit vector \mathbf{n} . This is just an elaborate way to show that the unique $\text{SO}(3)$ -invariant Riemannian metric on S^2 is given by projecting (pulling back) the Euclidean metric on \mathbb{R}^3 to the sphere.

It remains to clarify the relationship between the metric (7.64) and the connections (7.60). The covariant derivative of the metric with respect to these connections, in the direction of an arbitrary vector field ν , is easily calculated in the local frame,

$$(\nabla_{\mathbf{v}} g_{G/H})_{ab} = \mathbf{v}[g_{ab}] - g_{cb} {}^\lambda \Omega^c_a(\mathbf{v}) - g_{ac} {}^\lambda \Omega^c_b(\mathbf{v}) . \quad (7.69)$$

The first term vanishes since g_{ab} is merely a set of constants. The sum of the second and the third term vanishes for any λ as a consequence of the combination of (7.60) and (7.66). Thus, the whole class of connections (7.60) is compatible with the metric (7.64). In case the metric is (pseudo-)Riemannian, it is known that there is a unique metric connection without torsion, called the *Levi-Civita connection*. As shown in the previous subsection, this corresponds to the choice $\lambda = 1/2$.

The metric on the coset space is not uniquely fixed by the requirement of invariance under $G \times H_{\text{gauge}}$. Although it was natural to start the construction with the Cartan–Killing form (7.62) on the whole Lie algebra \mathfrak{g} and then restrict it to the $\mathfrak{g}/\mathfrak{h}$ subspace, we could have as well started from the latter. One then finds that possible metrics on G/H invariant under $G \times H_{\text{gauge}}$ are classified by constant symmetric tensors g_{ab} invariant under the adjoint action of H_{gauge} , that is g_{ab} satisfying the first relation in (7.66). For g_{ab} that does not satisfy the second relation in (7.66), the connections (7.60) with $\lambda \neq 0$ are not metric-compatible. In particular ${}^{1/2}\Omega^a_b$ is not the Levi-Civita connection despite being torsion-free.

References

1. S. Weinberg, *The Quantum Theory of Fields*, vol. I (Cambridge University Press, Cambridge, UK, 1995)
2. P.J. Olver, *Applications of Lie Groups to Differential Equations* (Springer-Verlag, New York, 1986)
3. G.W. Bluman, S.C. Anco, *Symmetry and Integration Methods for Differential Equations* (Springer-Verlag, New York, 2002)
4. J.S.R. Chisholm, Nucl. Phys. **26**, 469 (1961). DOI 10.1016/0029-5582(61)90106-7
5. S. Kamefuchi, L. O’Raifeartaigh, A. Salam, Nucl. Phys. **28**, 529 (1961). DOI 10.1016/0029-5582(61)90056-6
6. J.C. Criado, M. Pérez-Victoria, JHEP **03**, 038 (2019). DOI 10.1007/JHEP03(2019)038
7. M. Fecko, *Differential Geometry and Lie Groups for Physicists* (Cambridge University Press, Cambridge, UK, 2011)
8. S.R. Coleman, J. Wess, B. Zumino, Phys. Rev. **177**, 2239 (1969). DOI 10.1103/PhysRev.177.2239
9. A.O. Barut, R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific Publishers, Warszawa, Poland, 1977)
10. A. Joseph, A.I. Solomon, J. Math. Phys. **11**, 748 (1970). DOI 10.1063/1.1665205
11. R. Alonso, E.E. Jenkins, A.V. Manohar, JHEP **08**, 101 (2016). DOI 10.1007/JHEP08(2016)101
12. A. Arvanitoyeorgos, *An Introduction to Lie Groups and the Geometry of Homogeneous Spaces*, *Student Mathematical Library*, vol. 22 (American Mathematical Society, 2003)
13. M. Bando, T. Kugo, K. Yamawaki, Phys. Rept. **164**, 217 (1988). DOI 10.1016/0370-1573(88)90019-1

Chapter 8

Low-Energy Effective Field Theory

We are now finally in the position to construct a low-energy *effective field theory* (EFT) for systems with a spontaneously broken internal symmetry. We do not need any specific model to describe the dynamics of *spontaneous symmetry breaking* (SSB). All we need is information about the internal symmetry group G and its unbroken subgroup H . These determine the basic degrees of freedom of the EFT: the *Nambu–Goldstone* (NG) fields. With the known symmetry and field content, one must then include in the effective Lagrangian all operators allowed by the symmetry (Chap. 12 of [1]). This ensures that the predictions of the EFT match, to the desired accuracy, those of any microscopic theory with the same pattern of SSB.

The NG fields can be thought of as (local) coordinates on the coset space G/H . In the previous chapter, I showed that the action of the symmetry on these fields can be assumed to take the “standard form,” cf. Sect. 7.3. The task to construct the most general action invariant under such nonlinear realization of symmetry is nontrivial. In Sects. 8.1 and 8.2, I will offer two different approaches to this problem. The first of these is more general but requires substantial background in differential geometry. The second approach is based on certain simplifying assumptions but has the benefit of only requiring elementary field theory. In the concluding Sect. 8.3, I use the EFT to derive the *equation of motion* (EoM) for the NG fields. With the help of the latter, I then reanalyze the spectrum of NG bosons.

8.1 Structure of the Effective Lagrangian

To keep the discussion simple, I will assume unbroken symmetry under continuous spacetime translations and continuous spatial rotations. This is just a matter of practical convenience; the same methodology can be applied, for instance, to systems whose spatial rotation symmetry spans a (possibly finite) subgroup of $SO(d)$. I will also assume that the spatial dimension is $d \geq 2$. On the practical side, this is needed for any continuous rotations to exist. On a more fundamental note, the possibility of SSB in one spatial dimension is severely restricted; see Sect. 15.2.

The action of the low-energy EFT must inherit the symmetries of the underlying microscopic theory. As a rule, the symmetry admits an infinite number of operators in the effective Lagrangian, which require an infinite number of a priori unknown coupling constants. In order that the EFT has any predictive power, we need an organizing principle to tell us which of the allowed operators are “relevant.” Recall that the EFT is expected to be valid at low energies such that the only active degrees of freedom are the NG bosons. The individual operators in the Lagrangian can then be sorted by the number of derivatives they contain. The more derivatives, the smaller effects the operator is expected to produce. Such ordering of operators in the effective Lagrangian is called *derivative expansion*.

Since spatial and temporal derivatives are not related by any of the assumed symmetries, we must count them separately. The low-energy effective Lagrangian can then be organized as a double series,

$$\mathcal{L}_{\text{eff}}[\pi] = \sum_{s,t \geq 0} \mathcal{L}_{\text{eff}}^{(s,t)}[\pi] , \quad (8.1)$$

where s, t denote respectively the numbers of spatial and temporal derivatives. Following loosely [2], I will focus in this section on contributions with at most two derivatives, $s + t \leq 2$. These are needed to pin down the kinetic term for all the NG fields, and thus carry information about the spectrum of NG bosons. At the same time, they encode the dominant interactions of NG bosons at low energies. Operators with a higher number of derivatives are classified more easily using the approach developed in Sect. 8.2.

Within the diagrammatic expansion of quantum field theory, a given observable may receive contributions from graphs with different numbers of loops. The relative importance of various contributions depends on the operators that enter the interaction vertices and the number of loops. I will work out a precise *power-counting* scheme in Chap. 9, where concrete examples of EFTs are discussed at length. For the time being, I will content myself with the rule of thumb that loops lead to parametric suppression just like derivatives in the interaction vertices. This justifies the approach adopted throughout the whole book, whereby NG bosons are treated in the classical (tree-level) approximation.

The analysis of the structure of the effective Lagrangian (8.1) is considerably simplified using the language of differential geometry. Following the presentation below will therefore require some familiarity with the contents of Appendix A. A reader uninterested in the details will find a summary of the leading part of the effective Lagrangian (that is operators with $s + t \leq 2$) in Sect. 8.1.4.

8.1.1 Reminder of the Standard Nonlinear Realization

I will start with a brief reminder of the standard nonlinear realization of symmetry. The intention is to have the basic elements of the formalism at one place for easy reference. Further details can, if needed, be found in Chap. 7.

Every point on the coset space G/H is labeled with a representative $U(\pi) \in G$ of the corresponding coset, where π^a are local coordinates on G/H (NG fields). The action of the group G on the coset space can then be defined via left multiplication,

$$U(\pi) \xrightarrow{g} U(\pi'(\pi, g)) = gU(\pi)h(\pi, g)^{-1}, \quad (8.2)$$

where $h(\pi, g) \in H$. The choice of the representative $U(\pi)$ is not unique, but can be made so that $U(0) = e$ and that any $g \in H$ acts on $U(\pi)$ by conjugation, $U(\pi) \xrightarrow{g} gU(\pi)g^{-1}$. In other words, $h(\pi, g) = g$ for any $g \in H$ independently of π^a .

The generators of G are denoted as $Q_{A,B}, \dots$. A subset of these, $Q_{\alpha,\beta}, \dots$, spans the Lie algebra \mathfrak{h} of the unbroken subgroup H . The remaining, broken generators of G are denoted by $Q_{a,b}, \dots$.¹ The structure of the Lie algebra \mathfrak{g} is fixed by the structure constants f_{AB}^C through $[Q_A, Q_B] = if_{AB}^C Q_C$. Under the action of a group element infinitesimally close to unity, $g \approx e + i\epsilon^A Q_A$, (8.2) reduces to

$$\delta\pi^a(\pi, g) \approx \epsilon^A \xi_A^a(\pi), \quad h(\pi, g) \approx e + i\epsilon^A k_A^\alpha(\pi) Q_\alpha. \quad (8.3)$$

The functions $\xi_A^a(\pi)$ define a set of vector fields on G/H , $\xi_A(\pi) \equiv \xi_A^a(\pi) \partial / \partial \pi^a$, whose Lie bracket reproduces the structure of \mathfrak{g} ,

$$[\xi_A, \xi_B] = f_{AB}^C \xi_C. \quad (8.4)$$

An object of fundamental importance is the \mathfrak{g} -valued *Maurer–Cartan* (MC) form,

$$\omega(\pi) \equiv -iU(\pi)^{-1}dU(\pi) \equiv \omega^A(\pi)Q_A. \quad (8.5)$$

This can be split into a part belonging to the subspaces \mathfrak{h} and $\mathfrak{g}/\mathfrak{h}$,

$$\omega \equiv \omega_{\parallel} + \omega_{\perp}, \quad \omega_{\parallel} \equiv \omega^\alpha Q_\alpha, \quad \omega_{\perp} \equiv \omega^a Q_a. \quad (8.6)$$

The components ω^A of the MC form and the vector fields ξ_A are mutually dual in a well-defined sense. Their precise relationship is expressed by the identities

$$\mathfrak{l}_{\xi_A} \omega^\alpha = \nu_A^\alpha - k_A^\alpha, \quad \mathfrak{l}_{\xi_A} \omega^a = \nu_A^a, \quad (8.7)$$

where the matrix function $\nu_A^B(\pi)$ is defined by the conjugation

$$U(\pi)^{-1}Q_A U(\pi) \equiv \nu_A^B(\pi)Q_B. \quad (8.8)$$

¹ The basis $Q_{a,b}, \dots$ is to be chosen so that it spans a subspace $\mathfrak{g}/\mathfrak{h}$ of the Lie algebra \mathfrak{g} of G carrying a representation of H , hence $\mathfrak{g} \approx \mathfrak{h} \oplus \mathfrak{g}/\mathfrak{h}$.

Under G , the unbroken and broken parts of the MC form transform as

$$\begin{aligned}\omega_{\parallel}(\pi) &\xrightarrow{g} \omega_{\parallel}(\pi'(\pi, g)) = h(\pi, g)\omega_{\parallel}(\pi)h(\pi, g)^{-1} - ih(\pi, g)dh(\pi, g)^{-1}, \\ \omega_{\perp}(\pi) &\xrightarrow{g} \omega_{\perp}(\pi'(\pi, g)) = h(\pi, g)\omega_{\perp}(\pi)h(\pi, g)^{-1}.\end{aligned}\quad (8.9)$$

In the language of differential geometry, infinitesimal symmetry transformations of tensor fields on G/H are given by the Lie derivative along the vector fields ξ_A . Combining (8.3) and (8.9) thus corresponds to

$$\mathcal{L}_{\xi_A}\omega^\alpha = -f_{\beta\gamma}^\alpha k_A^\beta \omega^\gamma - dk_A^\alpha, \quad \mathcal{L}_{\xi_A}\omega^a = -f_{bc}^a k_A^b \omega^c. \quad (8.10)$$

Finally, the exterior derivative of the MC 1-form is expressed by the MC equation $d\omega^A = (1/2)f_{BC}^A \omega^B \wedge \omega^C$. In terms of the ω_{\parallel} and ω_{\perp} components of the MC form, this breaks down into

$$\begin{aligned}d\omega^\alpha &= \frac{1}{2}f_{\beta\gamma}^\alpha \omega^\beta \wedge \omega^\gamma + \frac{1}{2}f_{bc}^\alpha \omega^b \wedge \omega^c, \\ d\omega^a &= f_{bc}^a \omega^b \wedge \omega^c + \frac{1}{2}f_{bc}^a \omega^b \wedge \omega^c.\end{aligned}\quad (8.11)$$

8.1.2 Lagrangians with Two Spatial or Two Temporal Derivatives

Invariance of the effective action under a symmetry group G requires that the Lagrangian (density) be *quasi-invariant*, that is invariant up to a surface term. Internal symmetry transformations, covered in this part of the book, do not involve any derivatives of NG fields. This implies that all the different parts $\mathcal{L}_{\text{eff}}^{(s,t)}$ of the effective Lagrangian (8.1) must be quasi-invariant separately. The piece without any derivatives, $\mathcal{L}_{\text{eff}}^{(0,0)}$, must be strictly G -invariant; its variation under G cannot be a derivative of any local operator.² However, requiring $\mathcal{L}_{\xi_A}\mathcal{L}_{\text{eff}}^{(0,0)} = \xi_A^a \partial \mathcal{L}_{\text{eff}}^{(0,0)} / \partial \pi^a = 0$ leaves a constant, π^a -independent $\mathcal{L}_{\text{eff}}^{(0,0)}$ as the only option. The derivative expansion of the effective Lagrangian thus starts with operators with at least one derivative. Having a sole spatial derivative is forbidden by rotational invariance, hence $\mathcal{L}_{\text{eff}}^{(1,t)} = 0$ for any $t \geq 0$. The only options for (s, t) with $s + t \leq 2$ that are left are therefore $(0, 1)$, $(0, 2)$ and $(2, 0)$. The Lagrangian $\mathcal{L}_{\text{eff}}^{(0,1)}$ with one temporal derivative will be analyzed in Sect. 8.1.3. Here I will focus on $\mathcal{L}_{\text{eff}}^{(2,0)}$ and $\mathcal{L}_{\text{eff}}^{(0,2)}$, which turn out to be much easier to understand and thus constitute a good starting point.

Invariance under spacetime translations and spatial rotations restricts the two-derivative Lagrangians to the generic form

² It is in principle possible that the variation of $\mathcal{L}_{\text{eff}}^{(0,0)}$ under G is a field-independent constant. This can however only happen if $\mathcal{L}_{\text{eff}}^{(0,0)}$ is a tadpole operator, linear in π^a . Such operators must be discarded for the EFT to be perturbatively well-defined.

$$\begin{aligned}\mathcal{L}_{\text{eff}}^{(2,0)} &= -\frac{1}{2}g_{ab}(\pi)\nabla\pi^a\cdot\nabla\pi^b - \frac{1}{2}b_{ab}(\pi)\varepsilon^{rs}\partial_r\pi^a\partial_s\pi^b, \\ \mathcal{L}_{\text{eff}}^{(0,2)} &= \frac{1}{2}\bar{g}_{ab}(\pi)\dot{\pi}^a\dot{\pi}^b,\end{aligned}\tag{8.12}$$

where $b_{ab}(\pi)$, $g_{ab}(\pi)$ and $\bar{g}_{ab}(\pi)$ are some functions on G/H . Note that the b_{ab} term can only exist in $d = 2$ spatial dimensions. Also, I have discarded from the outset operators containing a second derivative of π^a , since those can be brought to the form (8.12) by partial integration.

The g_{ab} and \bar{g}_{ab} terms must be strictly G -invariant. Indeed, should an infinitesimal transformation of, say, the g_{ab} term produce a surface term, it would have to be of the form $\nabla\cdot[f_a(\pi)\nabla\pi^a]$. But that would inevitably contain a second derivative of π^a that was not present in the original Lagrangian. This argument does not apply to the b_{ab} term, which is antisymmetric in spatial derivatives. However, in that case, possible quasi-invariance can be disregarded on physical grounds. Namely, the b_{ab} term contributes to the canonical Hamiltonian density, and its mere quasi-invariance would imply that the energy density of the EFT is ill-defined. All in all, both $\mathcal{L}_{\text{eff}}^{(2,0)}$ and $\mathcal{L}_{\text{eff}}^{(0,2)}$ must be strictly invariant under the symmetry group G .

Let us see what strict G -invariance tells us about the function $g_{ab}(\pi)$. The latter gives rise to a symmetric rank-2 tensor field on G/H ,

$$g(\pi) \equiv g_{ab}(\pi)d\pi^a \otimes d\pi^b.\tag{8.13}$$

We know that the broken components of the MC form, $\omega^a(\pi)$, furnish a basis of the cotangent space (coframe) to G/H ; see Sect. 7.4 for a detailed justification. In this basis, the tensor $g(\pi)$ can be expanded as

$$g(\pi) = \kappa_{ab}(\pi)\omega^a(\pi) \otimes \omega^b(\pi),\tag{8.14}$$

where the components $\kappa_{ab}(\pi)$ are some as yet unknown functions on G/H . Using (8.10), we readily calculate the Lie derivative along ξ_A ,

$$\mathcal{L}_{\xi_A}g = (\mathcal{L}_{\xi_A}\kappa_{ab} - \kappa_{cb}f_{\beta a}^c k_A^\beta - \kappa_{ac}f_{\beta b}^c k_A^\beta)\omega^a \otimes \omega^b.\tag{8.15}$$

The requirement of strict G -invariance then leads to the following condition on κ_{ab} ,

$$k_A^\beta(f_{\beta a}^c \kappa_{cb} + f_{\beta b}^c \kappa_{ac}) = \xi_A^c \frac{\partial \kappa_{ab}}{\partial \pi^c}.\tag{8.16}$$

It is now convenient to resort to the exponential parameterization,

$$U(\pi) = \exp(i\pi^a Q_a),\tag{8.17}$$

in which $\xi_A^a(\pi) = \delta_A^a + \mathcal{O}(\pi)$ and $k_A^\alpha(\pi) = \delta_A^\alpha + \mathcal{O}(\pi)$, cf. (7.41). By choosing the index A in (8.16) in turn as an unbroken and broken index, we thus get the following initial conditions at the origin of G/H ,

$$f_{\alpha a}^c \kappa_{cb} + f_{\alpha b}^c \kappa_{ac} \Big|_{\pi=0} = 0, \quad \frac{\partial \kappa_{ab}}{\partial \pi^c} \Big|_{\pi=0} = 0. \quad (8.18)$$

For a fixed *broken* index A , (8.16) can be viewed as an ordinary differential equation that uniquely determines $\kappa_{ab}(\pi)$ along the integral curve of ξ_A , passing through the origin of G/H . We conclude that κ_{ab} must be constant, subject to the constraint

$$f_{\alpha a}^c \kappa_{cb} + f_{\alpha b}^c \kappa_{ac} = 0. \quad (8.19)$$

This is equivalent to requiring that κ_{ab} is a constant invariant tensor under the adjoint action of the unbroken subgroup H .

The same reasoning applies without change to any G -invariant operator constructed solely out of the broken part of the MC form, ω_\perp . As a consequence, the most general form consistent with G -invariance that the functions $b_{ab}(\pi)$, $g_{ab}(\pi)$ and $\bar{g}_{ab}(\pi)$ can take is

$$\begin{aligned} b_{ab}(\pi) &= \lambda_{cd} \omega_a^c(\pi) \omega_b^d(\pi), \\ g_{ab}(\pi) &= \kappa_{cd} \omega_a^c(\pi) \omega_b^d(\pi), \quad \bar{g}_{ab}(\pi) = \bar{\kappa}_{cd} \omega_a^c(\pi) \omega_b^d(\pi). \end{aligned} \quad (8.20)$$

The constant matrices λ_{ab} , κ_{ab} and $\bar{\kappa}_{ab}$ are all invariant tensors under the adjoint action of H . In the exponential parameterization (8.17), $\omega_b^a(\pi) = \delta_b^a + \mathcal{O}(\pi)$ and thus $\lambda_{ab} = b_{ab}(0)$, $\kappa_{ab} = g_{ab}(0)$ and $\bar{\kappa}_{ab} = \bar{g}_{ab}(0)$.

For the kinetic term of the NG fields to have the correct signature, the symmetric matrices κ_{ab} and $\bar{\kappa}_{ab}$ should be positive-definite. Hence, both $g(\pi)$ and its cousin $\bar{g}(\pi)$ constitute a G -invariant Riemannian metric on G/H . These two metrics are in principle independent from each other. However, in case G is compact and semisimple and all the broken generators Q_a span a single irreducible multiplet of H , Schur's lemma requires that κ_{ab} , $\bar{\kappa}_{ab}$ be proportional to each other. The metrics $g(\pi)$ and $\bar{g}(\pi)$ then coincide up to an overall factor.

Example 8.1

Suppose that the symmetry group G is completely broken. Then the matrices κ_{ab} , $\bar{\kappa}_{ab}$, λ_{ab} remain unconstrained by symmetry, the only physical requirement being the positive-definiteness of κ_{ab} and $\bar{\kappa}_{ab}$. This, however, does not mean that when constructing an EFT for NG bosons on $G/\{e\} \simeq G$, we have to work with arbitrary matrices of couplings of rank $\dim G$. The number of independent parameters can be reduced by a judicious choice of coordinates π^a on G . In the theory of small oscillations of mechanical systems [3], it is known that two quadratic forms, at least one of which is positive-definite, can be simultaneously diagonalized. Thus, we can always assume that, say, $\kappa_{ab} = \delta_{ab}$ and moreover that $\bar{\kappa}_{ab}$ is diagonal and positive-definite. We just have to keep in mind that this simplicity is a consequence of a particular choice of coordinates, not of symmetry. This may restrict the otherwise arbitrary choice of parameterization of the coset space.

8.1.3 Lagrangians with One Temporal Derivative

The part of the effective Lagrangian with one time derivative reads, generically,

$$\mathcal{L}_{\text{eff}}^{(0,1)} = c_a(\pi) \dot{\pi}^a . \quad (8.21)$$

The functions $c_a(\pi)$ define locally a 1-form on G/H , $c(\pi) \equiv c_a(\pi) d\pi^a$. The requirement of quasi-invariance of $\mathcal{L}_{\text{eff}}^{(0,1)}$ under the action of G is then equivalent to

$$\mathcal{L}_{\xi_A} c = d\bar{c}_A , \quad (8.22)$$

where $\bar{c}_A(\pi)$ is a set of local functions (0-forms) on G/H . Using the fact that the Lie derivative commutes with the exterior derivative, one deduces that $\mathcal{L}_{\xi_A}(dc) = d(\mathcal{L}_{\xi_A} c) = d(d\bar{c}_A) = 0$. Hence $dc(\pi)$ is a strictly G -invariant closed 2-form on G/H . Strict G -invariance requires that

$$dc(\pi) = \frac{1}{2} \sigma_{ab} \omega^a(\pi) \wedge \omega^b(\pi) , \quad (8.23)$$

where σ_{ab} is a constant antisymmetric matrix, invariant under the adjoint action of H . This follows by the same argument I used above to derive (8.20).

Not every H -invariant matrix σ_{ab} will do, however, since the 2-form $dc(\pi)$ must also be closed. Upon using (8.11), closedness of $dc(\pi)$ is seen to be equivalent to

$$f_{\alpha a}^c \sigma_{cb} + f_{\alpha b}^c \sigma_{ac} = 0 , \quad f_{ab}^d \sigma_{cd} + f_{bc}^d \sigma_{ad} + f_{ca}^d \sigma_{bd} = 0 . \quad (8.24)$$

The former of these conditions just reasserts the H -invariance of σ_{ab} . The latter constitutes a new constraint that we must now deal with. I will not be able to give a fully general explicit solution to the algebraic conditions on σ_{ab} , and thus a general solution for $\mathcal{L}_{\text{eff}}^{(0,1)}$. We can get very far even without an explicit solution, though. First, note that (8.24) is a set of linear equations for σ_{ab} . In a concrete case where the structure constants are known, the space of solutions to (8.24) can therefore be found using elementary linear algebra. Second, we can do even better and link possible solutions for σ_{ab} to the structure of the Lie groups G , H and their Lie algebras \mathfrak{g} , \mathfrak{h} .

In the exponential parameterization (8.17), $dc(\pi) = (1/2)[\sigma_{ab} + O(\pi)] d\pi^a \wedge d\pi^b$. Thus, the corresponding Lagrangian can be power-expanded in the NG fields as

$$\mathcal{L}_{\text{eff}}^{(0,1)} = \frac{1}{2} \sigma_{ab} \pi^a \partial_0 \pi^b + O(\pi^3) . \quad (8.25)$$

Following the argument of Sect. 6.3.2, the matrix σ_{ab} can be related to the set of *vacuum expectation values* (VEVs) of commutators of broken generators,

$$\sigma_{ab} = i \lim_{V \rightarrow \infty} \frac{\langle [Q_a, Q_b] \rangle}{V} . \quad (8.26)$$

There is a possibility that the representation of the symmetry group G on the Hilbert space of our system features a central extension. Let us write the extended commu-

tation relation of symmetry generators as $[Q_A, Q_B] = i f_{AB}^C Q_C + i V z_{AB}$, where z_{AB} are the densities of the central charges. Then we find that

$$\sigma_{ab} = -f_{ab}^C \lim_{V \rightarrow \infty} \frac{\langle Q_C \rangle}{V} - \langle z_{ab} \rangle . \quad (8.27)$$

We conclude that possible solutions for σ_{ab} directly reflect the Lie algebra of symmetry generators, including possible central extensions. One class of solutions to (8.24) that always exists is

$$\sigma_{ab} = -f_{ab}^C \sigma_C \quad \text{such that} \quad f_{\alpha B}^A \sigma_A = 0 . \quad (8.28)$$

The latter constraint expresses, once again, the invariance of the constant tensor σ_A under the adjoint action of H . This solution corresponds to vanishing (VEVs of) central charges, and σ_A can then be interpreted as the density of Q_A in the ground state. The H -invariance of σ_A descends directly from the H -invariance of the ground state. For this class of solutions, (8.23) is readily integrated using the MC equation (8.11), leading to

$$c(\pi) = -\sigma_A \omega^A(\pi) \quad \text{or} \quad \mathcal{L}_{\text{eff}}^{(0,1)} = -\sigma_A \omega_a^A(\pi) \dot{\pi}^a , \quad (8.29)$$

up to addition of a closed 1-form (surface term). The variation of $\mathcal{L}_{\text{eff}}^{(0,1)}$ under infinitesimal transformations from G can now be computed with the help of (8.10),

$$\mathcal{L}_{\xi_A} c = -\sigma_B \mathcal{L}_{\xi_A} \omega^B = \sigma_B (f_{\alpha C}^B k_A^\alpha \omega^C + \delta_\alpha^B dk_A^\alpha) = d(\sigma_\alpha k_A^\alpha) , \quad (8.30)$$

where in the last step I used that $\sigma_B f_{\alpha C}^B = 0$ by (8.28). The operators proportional to σ_α are strictly G -invariant. Mere quasi-invariance of the Lagrangian thus requires nonzero σ_α , that is nonzero VEV of an unbroken charge in the ground state.

Example 8.2

In ferromagnets, $G/H \simeq \text{SU}(2)/\text{U}(1) \simeq S^2$. Up to an overall factor, there is a unique closed and G -invariant 2-form on S^2 , provided by the volume (area) form. In terms of the unit-vector parameterization of the sphere, this is proportional to $\varepsilon_{ijk} n^i dn^j \wedge dn^k$, cf. Example A.7. The solution (8.28) descends from the nonzero magnetization in the ferromagnetic ground state, corresponding to a nonzero VEV of the generator of the unbroken $\text{U}(1)$ subgroup. I will give an explicit form of the Lagrangian $\mathcal{L}_{\text{eff}}^{(0,1)}$ for ferromagnets in Sect. 9.2.

In case the symmetry group G is semisimple, its Lie algebra \mathfrak{g} cannot have any nontrivial central charges (Sect. 2.7 of [1]). Equation (8.29) is then the only possible solution for the 1-form $c(\pi)$ and the corresponding Lagrangian $\mathcal{L}_{\text{eff}}^{(0,1)}$. For G that are not semisimple, solutions with nonzero z_{ab} may exist.

Example 8.3

Consider the coset space $G/H \simeq U(1) \times U(1)/\{e\} \simeq T^2$. In this case, the symmetry group is Abelian so the conditions (8.24) are trivially satisfied. Up to an overall factor, there is a unique rank-2 antisymmetric tensor, $\sigma_{ab} = \varepsilon_{ab}$. According to (8.27), this necessarily arises from a central extension of the Lie algebra \mathfrak{g} of G . In the exponential parameterization (8.17), $\omega^a(\pi) = d\pi^a$, which leads to $c(\pi) = (1/2)\varepsilon_{ab}\pi^a d\pi^b$. This is merely quasi-invariant under the action of G as expected. Note that the NG fields π^a are angular variables on the torus T^2 that are only locally defined, and so is therefore the 1-form $c(\pi)$. What is well-defined globally on the whole coset space is only the G -invariant 2-form $dc(\pi)$.

To conclude the analysis of the Lagrangian $\mathcal{L}_{\text{eff}}^{(0,1)}$, let us see under what conditions we can expect it to actually shift by a surface term upon a G -transformation. Should $\mathcal{L}_{\text{eff}}^{(0,1)}$ be strictly G -invariant, it would necessarily assume the form

$$\mathcal{L}_{\text{eff}}^{(0,1)} = -\sigma_b \omega_a^b(\pi) \dot{\pi}^a, \quad (8.31)$$

with a constant H -invariant σ_a , that is $f_{\alpha a}^b \sigma_b = 0$. This follows by imposing directly the constraint $\mathcal{L}_{\xi_A}(\sigma_a \omega^a) = 0$. The corresponding closed 2-form is then

$$dc = -\sigma_a d\omega^a = -\frac{1}{2}\sigma_a f_{bc}^a \omega^b \wedge \omega^c, \quad (8.32)$$

that is, $\sigma_{ab} = -f_{ab}^c \sigma_c$. Quasi-invariant Lagrangians $\mathcal{L}_{\text{eff}}^{(0,1)}$ are therefore in a one-to-one correspondence with antisymmetric matrices σ_{ab} satisfying (8.24) that *cannot* be written as $\sigma_{ab} = -f_{ab}^c \sigma_c$ with some H -invariant tensor σ_a . Such antisymmetric matrices span the so-called second *Lie algebra cohomology* of \mathfrak{g} relative to the subalgebra \mathfrak{h} . See Sect. 3 of [4] for a mild introduction and further references.

Example 8.4

In case the symmetry under G is completely spontaneously broken, only the second of the conditions in (8.24) survives. In this case, nontrivial solutions for σ_{ab} , that is those that cannot be written as $\sigma_{ab} = -f_{ab}^c \sigma_c$ with some σ_c , are in a one-to-one correspondence with the central charges z_{ab} . Both of these are classified by the second Lie algebra cohomology of \mathfrak{g} .

8.1.4 Overview of the Lowest-Order Effective Lagrangian

Let me summarize what we have found so far. The basic assumptions I made are that an internal symmetry group G is spontaneously broken to its subgroup H , but the symmetry under continuous spacetime translations and spatial rotations remains

intact. Then in $d \geq 2$ spatial dimensions, the parts of the effective Lagrangian with up to two derivatives of NG fields take the generic form

$$\begin{aligned}\mathcal{L}_{\text{eff}}^{(0,1)} &= c_a(\pi) \dot{\pi}^a, \\ \mathcal{L}_{\text{eff}}^{(2,0)} &= -\frac{1}{2} \kappa_{cd} \omega_a^c(\pi) \omega_b^d(\pi) \nabla \pi^a \cdot \nabla \pi^b - \frac{1}{2} \lambda_{cd} \omega_a^c(\pi) \omega_b^d(\pi) \varepsilon^{rs} \partial_r \pi^a \partial_s \pi^b, \\ \mathcal{L}_{\text{eff}}^{(0,2)} &= \frac{1}{2} \bar{\kappa}_{cd} \omega_a^c(\pi) \omega_b^d(\pi) \dot{\pi}^a \dot{\pi}^b.\end{aligned}\quad (8.33)$$

Here κ_{ab} and $\bar{\kappa}_{ab}$ are constant symmetric H -invariant matrices,

$$f_{\alpha a}^c \kappa_{cb} + f_{\alpha b}^c \kappa_{ac} = 0, \quad f_{\alpha a}^c \bar{\kappa}_{cb} + f_{\alpha b}^c \bar{\kappa}_{ac} = 0. \quad (8.34)$$

Likewise, λ_{ab} is a constant antisymmetric matrix invariant under H ,

$$f_{\alpha a}^c \lambda_{cb} + f_{\alpha b}^c \lambda_{ac} = 0. \quad (8.35)$$

The λ_{ab} term can only exist in $d = 2$ spatial dimensions. In the special case of a Lorentz-invariant system, both $\mathcal{L}_{\text{eff}}^{(0,1)}$ and the λ_{ab} term in $\mathcal{L}_{\text{eff}}^{(2,0)}$ are forbidden. Moreover, $\kappa_{ab} = \bar{\kappa}_{ab}$ under the convention that the speed of light is set to unity. The entire effective Lagrangian with up to two derivatives then shrinks to a single term,

$$\mathcal{L}_{\text{eff}}^{(2)} = \frac{1}{2} \kappa_{cd} \omega_a^c(\pi) \omega_b^d(\pi) \partial_\mu \pi^a \partial^\mu \pi^b. \quad (8.36)$$

Both $\mathcal{L}_{\text{eff}}^{(2,0)}$ and $\mathcal{L}_{\text{eff}}^{(0,2)}$ are strictly G -invariant. On the other hand, $\mathcal{L}_{\text{eff}}^{(0,1)}$ may be only quasi-invariant. The functions $c_a(\pi)$ on G/H are constrained by the requirement that the 2-form $dc \equiv d(c_a d\pi^a) = (\partial c_b / \partial \pi^a) d\pi^a \wedge d\pi^b$ is closed and G -invariant. This implies

$$dc(\pi) = \frac{1}{2} \sigma_{cd} \omega_a^c(\pi) \omega_b^d(\pi) d\pi^a \wedge d\pi^b, \quad (8.37)$$

where σ_{ab} is a constant antisymmetric matrix subject to the conditions

$$f_{\alpha a}^c \sigma_{cb} + f_{\alpha b}^c \sigma_{ac} = 0, \quad f_{ab}^d \sigma_{cd} + f_{bc}^d \sigma_{ad} + f_{ca}^d \sigma_{bd} = 0. \quad (8.38)$$

Those Lagrangians $\mathcal{L}_{\text{eff}}^{(0,1)}$ that are strictly G -invariant assume the form

$$\mathcal{L}_{\text{eff}}^{(0,1)} = -\sigma_b \omega_a^b(\pi) \dot{\pi}^a, \quad \text{where} \quad f_{\alpha a}^b \sigma_b = 0. \quad (8.39)$$

In this case, $c_a(\pi) = -\sigma_b \omega_a^b(\pi)$ and $\sigma_{ab} = -f_{ab}^c \sigma_c$. In general, the 1-form $c(\pi)$ is determined by (8.37), hence by σ_{ab} , up to addition of a closed 1-form, $c(\pi) \rightarrow c(\pi) + d\tilde{c}(\pi)$. Such a shift only changes the Lagrangian by a total time derivative,

$$\mathcal{L}_{\text{eff}}^{(0,1)} \rightarrow \mathcal{L}_{\text{eff}}^{(0,1)} + \frac{\partial \tilde{c}(\pi)}{\partial \pi^a} \dot{\pi}^a = \mathcal{L}_{\text{eff}}^{(0,1)} + \frac{d\tilde{c}(\pi)}{dt}, \quad (8.40)$$

and so does not affect the dynamics of the low-energy EFT.

Let me conclude the overview with some remarks on the topological aspects of the effective Lagrangian. First of all, the broken part of the MC form ω_\perp is globally well-defined on the coset space G/H . This is because it furnishes a coframe, dual to the frame built out of the vector fields that realize the action of G on G/H . All the parts of the effective Lagrangian (8.33) that are strictly invariant under G are constructed out of products of $\omega^a(\pi)$ with constant tensor coefficients. As a consequence, all these parts are themselves globally well-defined on G/H , even though their explicit form may depend on the local coordinates π^a . In particular the $\kappa_{ab}, \bar{\kappa}_{ab}$ terms correspond to G -invariant Riemannian metrics on G/H .

The global existence of $\mathcal{L}_{\text{eff}}^{(0,1)}$, on the other hand, is not guaranteed by our construction. Problems may arise only if $\mathcal{L}_{\text{eff}}^{(0,1)}$ is merely quasi-invariant, otherwise the same argument as that for $\mathcal{L}_{\text{eff}}^{(2,0)}$ and $\mathcal{L}_{\text{eff}}^{(0,2)}$ applies. Quasi-invariant Lagrangians $\mathcal{L}_{\text{eff}}^{(0,1)}$ are classified by the second Lie algebra cohomology of \mathfrak{g} relative to the subalgebra \mathfrak{h} . Should the group G be compact and connected and the subgroup H closed and connected, this relative Lie algebra cohomology is isomorphic to the second de Rham cohomology of G/H (Theorem 7.4 in [5]). Cohomologically nontrivial σ_{ab} then ensures that the 2-form dc is closed but not exact. As a consequence, the 1-form $c(\pi)$ cannot be extended from the local coordinate patch to the entire coset space.

Example 8.5

Let us contrast the coset spaces $G/H \simeq G/\{e\}$ with respectively $G \simeq \text{U}(1) \times \text{U}(1)$ and $G \simeq \mathbb{R} \times \mathbb{R}$. These have identical Lie algebras $\mathfrak{g}, \mathfrak{h}$ that possess a single generator of the second (relative) Lie algebra cohomology, $\sigma_{ab} = \varepsilon_{ab}$. The single-derivative Lagrangian $\mathcal{L}_{\text{eff}}^{(0,1)} = (1/2)\varepsilon_{ab}\pi^a\partial_0\pi^b$ is quasi-invariant under G in both cases.

Now in the first case, $G/H \simeq \text{U}(1) \times \text{U}(1) \simeq T^2$. The 2-form $dc(\pi)$ is proportional to the volume form on the torus, and constitutes the single generator of the second de Rham cohomology of T^2 . This agrees with the fact that G is compact and connected and thus the two cohomology groups are necessarily isomorphic. The Lagrangian $\mathcal{L}_{\text{eff}}^{(0,1)}$ is obviously not globally well-defined on T^2 because the fields π^a are not.

In the second case, on the other hand, $G/H \simeq \mathbb{R}^2$. In the Euclidean plane, the Poincaré lemma holds and the second de Rham cohomology is trivial. Accordingly, $\mathcal{L}_{\text{eff}}^{(0,1)}$ is globally well-defined, since π^a are now nothing but the two Cartesian coordinates in the plane.

A pedant might object that a Lagrangian that is not well-defined everywhere on the coset space renders the whole EFT ill-defined. However, in classical field theory, one is largely interested just in the EoM. The latter only involves the functions $c_a(\pi)$ through the combinations $\partial c_b/\partial\pi^a - \partial c_a/\partial\pi^b$, that is, only depends on $dc(\pi)$. Unlike the Lagrangian, the classical EoM is therefore globally well-defined on G/H .

The situation is quite different in quantum theory where we need to be able to perform a functional integral over all field configurations on the coset space. Luckily, we do not really need the Lagrangian $\mathcal{L}_{\text{eff}}[\pi]$, or even the action S_{eff} , to be well-defined. What should be unambiguous is just the phase factor $\exp(iS_{\text{eff}})$. To ensure this, one needs to cover the coset space with a set of coordinate patches and define the action by piecewise integration. A detailed analysis is well-beyond the scope of this book. A curious reader will find more details and further references in [6], which puts forward a modern classification of topological terms in the action using a homology-based approach. Here I will just state without proof that consistency and G -invariance require that the so-called *Manton condition* be satisfied: the 1-forms $\iota_{\xi_A}(dc)$ must be exact for all the vector fields ξ_A realizing the action of G on G/H . This is a stronger condition than G -invariance of dc that I used above, which implies merely that $\iota_{\xi_A}(dc)$ is closed. The Manton condition is automatically satisfied by our $\mathcal{L}_{\text{eff}}^{(0,1)}$ whenever the first de Rham cohomology of G/H is trivial. This is the case for instance when G is compact and simply connected and H is connected; cf. Example A.28.

8.2 Effective Lagrangians from Background Gauge Invariance

In this section I will switch gears and outline another approach to the construction of effective Lagrangians for SSB. This is based on a technical assumption that rules out some of the more exotic theories covered by the method of Sect. 8.1, notably systems with a central extension of the symmetry algebra. The reward for making this sacrifice is a drastic simplification of the classification of possible contributions to the effective action. All we shall need will be elementary field theory with no recourse to differential geometry. The method presented here was pioneered by Leutwyler [7, 8], and I will largely follow the pedagogical account of [9].

The generating functional formalism constitutes an arsenal of important tools in both classical and quantum field theory. In this framework, one couples a given theory to a set of classical external (or background) fields and subsequently integrates, if only formally, over the dynamical degrees of freedom. The physical properties of the theory are encoded in the ensuing generating functional of the background fields. See Chap. 16 of [10] for an introduction to the formalism.

In principle, one has the freedom to choose the background at will. However, in presence of a continuous internal symmetry group G , it is convenient to introduce a set of background gauge fields A_μ^A , one for each generator Q_A of G . These can be clustered into a matrix-valued gauge field $A_\mu \equiv A_\mu^A Q_A$. The important technical assumption I am making here is that the symmetry under G is *gaugeable*. That is, it is possible to add the gauge fields so that the generating functional $W\{A\}$ of the

system³ is invariant under the gauge transformation

$$A_\mu \xrightarrow{g} T_g A_\mu \equiv g A_\mu g^{-1} + i g \partial_\mu g^{-1} . \quad (8.41)$$

Example 8.6

Consider a theory of a set of (not necessarily scalar) fields ϕ^i , whose Lagrangian density $\mathcal{L}[\phi]$ is strictly invariant under a linear representation of the symmetry group G , $\phi^i \xrightarrow{g} \mathcal{R}(g)^i_j \phi^j$. Such a Lagrangian can always be made invariant under the simultaneous local transformation of ϕ^i and (8.41). All one has to do is to replace derivatives of ϕ^i with gauge-covariant derivatives,

$$\partial_\mu \phi^i \rightarrow [\delta^i_j \partial_\mu - i \mathcal{R}(A_\mu)^i_j] \phi^j . \quad (8.42)$$

In order to assert gauge invariance of the generating functional $W\{A\}$, one still needs to check that the functional integral measure does not change upon the local transformation of ϕ^i . That would indicate an anomaly. In general, the assumption of gauge invariance of $W\{A\}$ essentially amounts to the absence of anomalies or other obstructions to gauging.

A linear realization of symmetry is typical for microscopic field theory models. There, one can usually check explicitly that the assumption of gauge invariance of the generating functional is satisfied. The low-energy EFT for the NG bosons should then, when coupled to the same external fields, reproduce the same generating functional. In other words, we expect that the effective action $S_{\text{eff}}\{\pi\} = \int d^D x \mathcal{L}_{\text{eff}}[\pi]$ can be replaced with a “gauged” action $S_{\text{eff}}\{\pi, A\} = \int d^D x \mathcal{L}_{\text{eff}}[\pi, A]$ such that

$$e^{iW\{A\}} = \int \mathcal{D}\pi \exp[iS_{\text{eff}}\{\pi, A\}] . \quad (8.43)$$

Leutwyler [7] showed that gauge invariance of $W\{A\}$ is sufficient to ensure that the effective action $S_{\text{eff}}\{\pi, A\}$ is invariant under the simultaneous local transformations (8.2) and (8.41) of the NG and gauge fields. The locality of (8.2) is implemented by allowing g therein to depend arbitrarily on spacetime coordinates as in (8.41).

It is common to gloss over this detail and simply assume right away that the effective action is G -invariant. I did the same in Sect. 8.1. It is therefore worth stressing that this is a consequence of the properties of the generating functional, established at the microscopic level. The proof is technical and I refer the reader to [7] for details.

³ I follow [7, 8] and use braces to indicate the arguments of a functional. Square brackets are reserved for local functions of fields and their derivatives.

We have replaced the problem of classifying possible G -invariant actions $S_{\text{eff}}\{\pi\}$ with that of constructing locally G -invariant actions $S_{\text{eff}}\{\pi, A\}$. How does that make our task easier? In short, a lot. I will now show in a series of simple steps how the problem of finding all possible gauged actions $S_{\text{eff}}\{\pi, A\}$ can be reduced to an exercise in group theory. If desired, one can then always discard the background gauge fields and thus recover (most of) the results of Sect. 8.1. But the gauged action in fact carries more information, as it tells us how the EFT responds to certain external perturbations.

8.2.1 Methodology of Construction of Effective Actions

In the first step, we exploit the gauge invariance to eliminate explicit dependence of the action on the NG fields. Namely, by setting $g = U(\pi)^{-1}$, we get

$$S_{\text{eff}}\{\pi, A\} = S_{\text{eff}}\{0, T_{U(\pi)^{-1}}A\} . \quad (8.44)$$

By combining (8.2) and (8.41), we find that under the local action of G , the composite field $T_{U(\pi)^{-1}}A_\mu$ transforms as

$$T_{U(\pi)^{-1}}A_\mu \xrightarrow{g} T_{U(\pi'(\pi, g))^{-1}}T_g A_\mu = T_{h(\pi, g)}T_{U(\pi)^{-1}}A_\mu . \quad (8.45)$$

This is a special case of a local transformation of $T_{U(\pi)^{-1}}A_\mu$ from H . Thus, gauged actions $S_{\text{eff}}\{\pi, A\}$ locally invariant under G are in a one-to-one correspondence with locally H -invariant functionals of $T_{U(\pi)^{-1}}A_\mu$.

Let us have a closer look at this composite field. By (8.41),

$$\begin{aligned} T_{U(\pi)^{-1}}A_\mu &= U(\pi)^{-1}A_\mu U(\pi) + iU(\pi)^{-1}\partial_\mu U(\pi) \\ &= iU(\pi)^{-1}(\partial_\mu - iA_\mu)U(\pi) . \end{aligned} \quad (8.46)$$

Up to an overall sign, this is just a gauged version of the MC form (8.5). We can make the analogy explicit by turning the gauge field into a 1-form, $A \equiv A_\mu dx^\mu$, and defining the gauged MC form via⁴

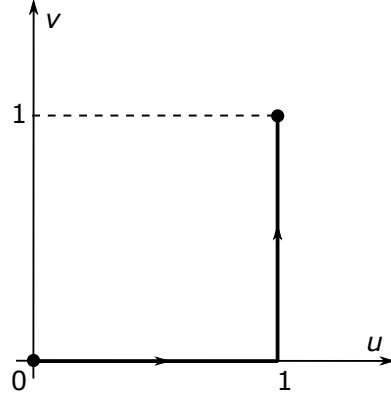
$$\Omega(\pi, A) \equiv \Omega_\mu^A(\pi, A)Q_A dx^\mu \equiv -T_{U(\pi)^{-1}}A = -iU(\pi)^{-1}(d - iA)U(\pi) . \quad (8.47)$$

This \mathfrak{g} -valued 1-form can again be split into unbroken and broken parts, $\Omega = \Omega_\parallel + \Omega_\perp$. These inherit the transformation rules (8.9) under the local action of G ,

$$\begin{aligned} \Omega_\parallel(\pi, A) &\xrightarrow{g} \Omega_\parallel(\pi', A') = h(\pi, g)\Omega_\parallel(\pi, A)h(\pi, g)^{-1} - ih(\pi, g)dh(\pi, g)^{-1} , \\ \Omega_\perp(\pi, A) &\xrightarrow{g} \Omega_\perp(\pi', A') = h(\pi, g)\Omega_\perp(\pi, A)h(\pi, g)^{-1} . \end{aligned} \quad (8.48)$$

⁴ This is a slight abuse of notation. The gauge field A is a 1-form on the spacetime. In order for $\Omega(\pi, A)$ to be a well-defined 1-form on the spacetime as well, the MC form $\omega(\pi)$ has to be pulled back by the map $\pi^a : x \rightarrow \pi^a(x)$.

Fig. 8.1 Integration path in the space of parameters u, v (shown by the oriented *solid line*), used to reconstruct the effective action from its partial derivatives (8.50)



The second step is to realize that we can temporarily forget about the origin of $\Omega_{\parallel}, \Omega_{\perp}$ in terms of the NG fields π^a and gauge fields A_{μ}^A . For the purposes of constructing the effective action, all we need is that $-\Omega_{\parallel}$ transforms as a gauge field of H , whereas Ω_{\perp} transforms linearly under the adjoint action of H . Once a locally H -invariant action $S_{\text{eff}}\{\Omega_{\parallel}, \Omega_{\perp}\}$ has been found, we can reconstruct the dependence on π^a, A_{μ}^A using (8.47). Now define two new vector fields by taking a variation of the action with respect to Ω_{\parallel} and Ω_{\perp} ,

$$J_{\alpha}^{\mu}[\Omega_{\parallel}, \Omega_{\perp}] \equiv \frac{\delta S_{\text{eff}}\{\Omega_{\parallel}, \Omega_{\perp}\}}{\delta \Omega_{\mu}^{\alpha}}, \quad \Sigma_a^{\mu}[\Omega_{\parallel}, \Omega_{\perp}] \equiv \frac{\delta S_{\text{eff}}\{\Omega_{\parallel}, \Omega_{\perp}\}}{\delta \Omega_{\mu}^a}. \quad (8.49)$$

In spite of the different transformation properties of Ω_{\parallel} and Ω_{\perp} , both J_{α}^{μ} and Σ_a^{μ} transform linearly under the adjoint action of H . This is just an auxiliary statement, and I therefore refer the reader to Appendix C of [9] for a detailed proof. The main message is that to construct such covariantly transforming objects is straightforward, and once we have done so, we can reconstruct the action.

To that end, introduce real scaling parameters u, v . It follows from (8.49) that

$$\begin{aligned} \frac{\partial S_{\text{eff}}\{u\Omega_{\parallel}, v\Omega_{\perp}\}}{\partial u} &= \int d^Dx \, \Omega_{\mu}^{\alpha}(x) J_{\alpha}^{\mu}[u\Omega_{\parallel}, v\Omega_{\perp}](x), \\ \frac{\partial S_{\text{eff}}\{u\Omega_{\parallel}, v\Omega_{\perp}\}}{\partial v} &= \int d^Dx \, \Omega_{\mu}^a(x) \Sigma_a^{\mu}[u\Omega_{\parallel}, v\Omega_{\perp}](x). \end{aligned} \quad (8.50)$$

Setting without loss of generality $S_{\text{eff}}\{0, 0\} = 0$, the full action is then obtained by integration along the path in the u, v space, shown in Fig. 8.1,

$$\begin{aligned} S_{\text{eff}}\{\Omega_{\parallel}, \Omega_{\perp}\} &= \int d^Dx \int_0^1 du \, \Omega_{\mu}^{\alpha}(x) J_{\alpha}^{\mu}[u\Omega_{\parallel}, 0](x) \\ &\quad + \int d^Dx \int_0^1 dv \, \Omega_{\mu}^a(x) \Sigma_a^{\mu}[\Omega_{\parallel}, v\Omega_{\perp}](x). \end{aligned} \quad (8.51)$$

Equivalently, the effective Lagrangian $\mathcal{L}_{\text{eff}}[\Omega_{\parallel}, \Omega_{\perp}]$ can be split into two pieces, $\mathcal{L}_{\text{eff}}[\Omega_{\parallel}, \Omega_{\perp}] = \mathcal{L}_{\text{CS}}[\Omega_{\parallel}] + \mathcal{L}_{\text{inv}}[\Omega_{\parallel}, \Omega_{\perp}]$, such that

$$\begin{aligned}\mathcal{L}_{\text{CS}}[\Omega_{\parallel}] &= \int_0^1 du \, \Omega_{\mu}^{\alpha} J_{\alpha}^{\mu}[u\Omega_{\parallel}, 0] , \\ \mathcal{L}_{\text{inv}}[\Omega_{\parallel}, \Omega_{\perp}] &= \int_0^1 dv \, \Omega_{\mu}^{\alpha} \Sigma_a^{\mu}[\Omega_{\parallel}, v\Omega_{\perp}] .\end{aligned}\tag{8.52}$$

As a consequence of the linear transformation properties of Ω_{μ}^{α} and Σ_a^{μ} , the part $\mathcal{L}_{\text{inv}}[\Omega_{\parallel}, \Omega_{\perp}]$ is strictly invariant under the local action of G . Possibly quasi-invariant contributions to the effective Lagrangian reside in the part $\mathcal{L}_{\text{CS}}[\Omega_{\parallel}]$, which only depends on the \mathfrak{h} -valued gauge field $-\Omega_{\parallel}$. I will refer to this part as *Chern–Simons* (CS) due to its resemblance of the CS theory with gauge group H .

The strategy to reconstruct the effective Lagrangian is now as follows. The gauge-invariant part of the Lagrangian, $\mathcal{L}_{\text{inv}}[\Omega_{\parallel}, \Omega_{\perp}]$, can be built directly out of Ω_{μ}^{α} , the field-strength of $-\Omega_{\mu}^{\alpha}$, $G_{\mu\nu}^{\alpha} \equiv -\partial_{\mu}\Omega_{\nu}^{\alpha} + \partial_{\nu}\Omega_{\mu}^{\alpha} + f_{\beta\gamma}^{\alpha}\Omega_{\mu}^{\beta}\Omega_{\nu}^{\gamma}$, and their covariant derivatives. This is common lore but if desired, a detailed proof can be found in Appendix D of [9]. In the same way, one can construct possible gauge-covariant “currents” J_{α}^{μ} , except that now the available building blocks are only $G_{\mu\nu}^{\alpha}$ and its covariant derivatives. In both cases, products of the covariant building blocks must be contracted with tensor coefficients that ensure invariance under the residual linearly realized symmetries (spatial rotations and the unbroken subgroup H). Finally, the CS part of the Lagrangian is obtained from the current J_{α}^{μ} by integration over u as indicated in (8.52). Here the quasi-invariance under H must be checked explicitly and may impose further constraints on the tensor couplings.

Up to any desired order in the derivative expansion, only a finite number of covariant operators can contribute to $\mathcal{L}_{\text{inv}}[\Omega_{\parallel}, \Omega_{\perp}]$ or $J_{\alpha}^{\mu}[\Omega_{\parallel}]$, and they can be enumerated by inspection. See [9] for a full list up to order four in derivatives. The problem of constructing all gauge-invariant actions then boils down to solving the group-theoretic linear constraints on the tensor couplings of the individual operators. The full dependence on the fields π^a, A_{μ}^A is already completely fixed by the structure of the operators. The practical use of this algorithmic procedure is best illustrated on the sample analysis of effective Lagrangians with up to two derivatives, worked out below.

8.2.2 Lagrangians up to Order Two in Derivative Expansion

For the derivative expansion to be consistent, the Lorentz index on A_{μ}^A has to count equally to the derivative $\partial/\partial x^{\mu}$. Each of $\Omega_{\mu}^{\alpha}, \Omega_{\mu}^a$ is then of order one in (spatial or temporal) derivatives. Up to order two, the invariant Lagrangian $\mathcal{L}_{\text{inv}}[\Omega_{\parallel}, \Omega_{\perp}]$ may contain the following contributions,

$$\Omega_{\mu}^a \quad (\text{order } 1) , \quad \Omega_{\mu}^a \Omega_{\nu}^b, D_{\mu}\Omega_{\nu}^a, G_{\mu\nu}^{\alpha} \quad (\text{order } 2) , \tag{8.53}$$

where $D_\mu \Omega_\nu^\alpha \equiv \partial_\mu \Omega_\nu^\alpha + i[\Omega_{\parallel\mu}, \Omega_{\perp\nu}]^\alpha$. However, a G -invariant Lagrangian density built out of $D_\mu \Omega_\nu^\alpha$ alone would be a pure surface term. Similarly, a term of the type $c_\alpha G_{\mu\nu}^\alpha$ would boil down to $c_\alpha f_{\beta\gamma}^\alpha \Omega_\mu^\beta \Omega_\nu^\gamma$, the rest of $G_{\mu\nu}^\alpha$ being a surface term. But the combination $c_\alpha f_{\beta\gamma}^\alpha$ would necessarily vanish due to the H -invariance condition imposed on the coefficient c_α . Altogether, upon adding the appropriate tensor couplings, the most general gauge-invariant and rotationally invariant Lagrangian up to second order in derivatives is

$$\mathcal{L}_{\text{inv}}[\Omega_{\parallel}, \Omega_{\perp}] = -\sigma_\alpha \Omega_0^\alpha + \frac{1}{2} \bar{\kappa}_{ab} \Omega_0^a \Omega_0^b - \frac{1}{2} \kappa_{ab} \delta^{rs} \Omega_r^a \Omega_s^b - \frac{1}{2} \lambda_{ab} \varepsilon^{rs} \Omega_r^a \Omega_s^b. \quad (8.54)$$

From (8.48), the coefficients $\sigma_\alpha, \kappa_{ab}, \bar{\kappa}_{ab}, \lambda_{ab}$ must be invariant under the adjoint action of H , that is satisfy the constraints

$$\begin{aligned} f_{\alpha a}^b \sigma_b &= 0, & f_{\alpha a}^c \lambda_{cb} + f_{\alpha b}^c \lambda_{ac} &= 0, \\ f_{\alpha a}^c \kappa_{cb} + f_{\alpha b}^c \kappa_{ac} &= 0, & f_{\alpha a}^c \bar{\kappa}_{cb} + f_{\alpha b}^c \bar{\kappa}_{ac} &= 0. \end{aligned} \quad (8.55)$$

In order to pin down $\mathcal{L}_{\text{CS}}[\Omega_{\parallel}]$ up to second order in derivatives, we need $J_\alpha^\mu[\Omega_{\parallel}]$ up to order one. But the basic building block available, $G_{\mu\nu}^\alpha$, already has order two. The only possibility left therefore is a constant current, restricted by rotation invariance to $J_\alpha^\mu[\Omega_{\parallel}] = -\delta_0^\mu \sigma_\alpha$. By (8.52), this leads in turn to

$$\mathcal{L}_{\text{CS}}[\Omega_{\parallel}] = -\sigma_\alpha \Omega_0^\alpha \quad (8.56)$$

as the sole possibility. Quasi-invariance under H imposes a constraint on σ_α ,

$$f_{\alpha\beta}^\gamma \sigma_\gamma = 0. \quad (8.57)$$

Note how effortlessly we recovered the effective Lagrangian, obtained in Sect. 8.1 and summarized in (8.33). The only difference is that out of all the possible solutions for the $c_a(\pi)$ functions, those found here correspond to vanishing VEV of the central charges z_{ab} in (8.27). This class of solutions is selected by the requirement that the symmetry under G is gaugeable.

However, our newly deduced effective Lagrangian consisting of (8.54) and (8.56) does not just reproduce the previously found results. It also tells us how the NG fields couple to the background gauge fields A_μ^A . To see how the Lagrangian depends on the latter, we combine the definition (8.47) with (8.7) and (8.8) to write

$$\Omega_\perp(\pi, A) = \omega_\perp(\pi) - A^A \nu_A^a(\pi) Q_a = \omega_b^a(\pi) [d\pi^b - A^A \xi_A^b(\pi)] Q_a. \quad (8.58)$$

This makes perfect sense: the vector fields ξ_A define infinitesimal group motions on the coset space. As a consequence,

$$D_\mu \pi^a \equiv \partial_\mu \pi^a - A_\mu^A \xi_A^a(\pi) \quad (8.59)$$

is the correct definition of a derivative of the NG field, covariant under local G -transformations. Altogether, the effective Lagrangian up to second order in deriva-

tives can then be written as

$$\begin{aligned}\mathcal{L}_{\text{eff}}^{(0,1)} &= -\sigma_A \omega_a^A(\pi) \dot{\pi}^a + \sigma_A v_B^A(\pi) A_0^B, \\ \mathcal{L}_{\text{eff}}^{(2,0)} &= -\frac{1}{2} \kappa_{cd} \omega_a^c(\pi) \omega_b^d(\pi) \mathbf{D}\pi^a \cdot \mathbf{D}\pi^b - \frac{1}{2} \lambda_{cd} \omega_a^c(\pi) \omega_b^d(\pi) \varepsilon^{rs} D_r \pi^a D_s \pi^b, \\ \mathcal{L}_{\text{eff}}^{(0,2)} &= \frac{1}{2} \bar{\kappa}_{cd} \omega_a^c(\pi) \omega_b^d(\pi) D_0 \pi^a D_0 \pi^b.\end{aligned}\tag{8.60}$$

Example 8.7

A chemical potential describing a statistical many-body state of a system can be introduced into its Lagrangian as a constant temporal gauge field (see Chap. 2 of [11]). Replacing $A_\mu^A \rightarrow \delta_{\mu 0} \mu^A$ in (8.60) therefore allows us to analyze the effect on the EFT of a set of chemical potentials μ^A . In particular, one can thus determine the ground state triggered by the chemical potentials as well as the spectrum of NG modes above it. A detailed discussion can be found in [12], where this setup was used to pin down the spectrum of massive NG bosons, introduced in Sect. 6.4.3.

In spin systems, the chemical potentials μ^A for the generators of $G \simeq \text{SU}(2)$ can be interpreted as the components of an external magnetic field \mathbf{B} . This is because the magnetic field couples to the conserved charge of $\text{SU}(2)$: spin. With the help of Example 7.14, one can easily check that the second term in $\mathcal{L}_{\text{eff}}^{(0,1)}$ in (8.60) is proportional to $\mathbf{B} \cdot \mathbf{n}$. This is the Zeeman interaction of spins with the magnetic field. Its normalization is fixed by σ_A , hence by the magnetization of the ground state.

8.2.3 Effects of Explicit Symmetry Breaking

So far in this chapter, I have assumed the symmetry of the EFT to be perfect. There are however good reasons to consider the effects of (presumably small) perturbations breaking the symmetry *explicitly*. First, I already used this idea in Sect. 5.2 to isolate a unique ground state in presence of SSB. Second, symmetries of real physical systems are almost always just approximate, even if to a high precision.

I will not attempt a general analysis of explicit symmetry breaking in the EFT, but rather focus on a special case that is commonplace and admits a very simple treatment. Suppose that the microscopic Lagrangian of the system contains a contribution $m_\varrho O^\varrho$ where m_ϱ are real constant parameters and O^ϱ a set of local operators. Suppose these operators transform under some (real) linear representation \mathcal{R} of G ,

$$O^\varrho \xrightarrow{g} O'^\varrho = \mathcal{R}(g)^\varrho_\sigma O^\sigma, \quad g \in G. \tag{8.61}$$

Such a perturbation violates the invariance of the action under G unless $m_\varrho = 0$ for all the components O^ϱ that belong to a nontrivial irreducible representation of G .

In order to understand how the effects of the perturbation propagate to the low-energy EFT, it is convenient to promote the parameters m_ϱ to background fields. The generating functional $W\{A, m\}$ then remains exactly invariant under the simultaneous local transformation

$$A_\mu \xrightarrow{g} T_g A_\mu = g A_\mu g^{-1} + i g \partial_\mu g^{-1}, \quad m_\varrho \xrightarrow{g} m'_\varrho = \mathcal{R}(g^{-1})^\sigma{}_\varrho m_\sigma. \quad (8.62)$$

This translates to the invariance of the gauged effective action $S_{\text{eff}}\{\pi, A, m\}$ under a simultaneous transformation of $\pi^a, A_\mu^A, m_\varrho$, generalizing (8.44) to

$$S_{\text{eff}}\{\pi, A, m\} = S_{\text{eff}}\{0, T_{U(\pi)^{-1}} A, m \mathcal{R}(U(\pi))\}. \quad (8.63)$$

The action therefore depends on π^a, A_μ^A and m_ϱ only through two composite fields, $T_{U(\pi)^{-1}} A_\mu$ and $\Xi_\varrho(\pi, m) \equiv \mathcal{R}(U(\pi))^\sigma{}_\varrho m_\sigma$. The field $\Xi_\varrho(\pi, m)$ transforms under G through a linear representation of H ,

$$\begin{aligned} \Xi_\varrho(\pi, m) &\xrightarrow{g} \Xi_\varrho(\pi', m') = \mathcal{R}(U(\pi'))^\sigma{}_\varrho m'_\sigma = \mathcal{R}(U(\pi))^\tau{}_\sigma \mathcal{R}(h(\pi, g)^{-1})^\sigma{}_\varrho m_\tau \\ &= \mathcal{R}(h(\pi, g)^{-1})^\sigma{}_\varrho \Xi_\sigma(\pi, m). \end{aligned} \quad (8.64)$$

By extension of the argument in Sect. 8.2.1, we can localize Ξ_ϱ in the strictly invariant part of the Lagrangian. The complete Lagrangian then takes the form

$$\mathcal{L}_{\text{eff}}[\pi, A, m] = \mathcal{L}_{\text{CS}}[\Omega_\parallel] + \mathcal{L}_{\text{inv}}[\Omega_\parallel, \Omega_\perp, \Xi]. \quad (8.65)$$

Exactly which operators containing Ξ_ϱ should be included at a given order of the derivative expansion depends on how we decide to count Ξ_ϱ . It does not contain any derivatives of NG fields but, being proportional to the perturbations m_ϱ , it makes sense to treat it as small. Depending on the concrete system, only a finite number of operators including Ξ_ϱ is then needed at any fixed order of the derivative expansion. The leading perturbation of the effective Lagrangian is linear in Ξ_ϱ ,

$$\mathcal{L}_{\text{pert}}[\pi, m] = \eta^\varrho \Xi_\varrho(\pi, m). \quad (8.66)$$

The effective coupling η^ϱ must be H -invariant, which amounts to $\mathcal{R}(Q_\alpha)^\varrho{}_\sigma \eta^\sigma = 0$ for any generator Q_α of H . A more complete list of operators containing Ξ_ϱ and contributing to the effective Lagrangian can be found in [9].

Example 8.8

The choice of background fields and their transformation properties is up to us. To illustrate this freedom, note that it is possible to treat the gauge fields A_μ^A themselves as linear perturbations. In this case, the operators O^ϱ are the Noether currents J_A^μ of the symmetry group G . These transform according to the dual of the adjoint representation of G ,

$$J_A^\mu \xrightarrow{g} J'^\mu_A = \mathcal{R}(g^{-1})^B{}_A J_B^\mu, \quad (8.67)$$

where \mathcal{R} is the adjoint representation itself. The gauge fields A_μ^A are now introduced to the microscopic Lagrangian through the linear coupling $A_\mu^A J_A^\mu$. Equation (8.62) assigns them the transformation rule $A_\mu^A \xrightarrow{g} \mathcal{R}(g)^A_B A_\mu^B$, which amounts to discarding the derivative piece of the background gauge transformation (8.41). Following (8.63) then leads to the composite field

$$\Xi_\mu^A(\pi, A) \equiv \mathcal{R}(U(\pi)^{-1})^A_B A_\mu^B = v_B^A(\pi) A_\mu^B. \quad (8.68)$$

The leading contribution of such a perturbation to the effective Lagrangian is then $\eta_A v_B^A(\pi) A_0^B$. The coupling η_A must be H -invariant, that is $f_{\alpha B}^A \eta_A = 0$.

We have successfully recovered the structure of the second term in $\mathcal{L}_{\text{eff}}^{(0,1)}$ in (8.60). Note that treating A_μ^A as a linear perturbation leaves the effective coupling η_A unfixed. One could in principle apply the same reasoning case by case to other operators in the effective Lagrangian containing A_μ^A . However, imposing full background gauge invariance is obviously much more efficient, as it fixes the dependence of the EFT on the gauge fields entirely without any free parameters.

8.2.4 Coupling to Matter Fields

Throughout the whole book, I mostly assume that the NG bosons are the *only* low-energy degrees of freedom present in the given system. This is justified in case any other, non-NG modes in the spectrum possess a gap. The validity of the low-energy EFT for NG bosons is then limited to energies well below this gap. However, there are physical systems where strictly gapless non-NG degrees of freedom naturally occur. One generic possibility is that the spectrum includes a Fermi sea of particles such as electrons, protons, neutrons, or quarks. Any local low-energy EFT must then necessarily include such additional gapless degrees of freedom. It is therefore worthwhile to digress and see how such modes fit into the EFT framework developed in this chapter.

The question how non-NG fields (also called *matter fields*) transform under the nonlinearly realized symmetry was already resolved in Chap. 7. Namely, any set of (not necessarily scalar) matter fields χ^ϱ can without loss of generality be assumed to transform under some linear representation D of the unbroken subgroup H . The action of the whole group G is then given by

$$\chi^\varrho \xrightarrow{g} \chi'^\varrho(\chi, \pi, g) = D(h(\pi, g))^\varrho_\sigma \chi^\sigma, \quad (8.69)$$

alongside (8.2) which defines the matrix $h(\pi, g) \in H$.

We are now looking for the most general effective action $S_{\text{eff}}\{\pi, A, m, \chi\}$ consistent with the background gauge invariance under G . Our basic trick has been to eliminate explicit dependence on the NG fields by performing a gauge transformation with $g = U(\pi)^{-1}$. For this transformation, $h(\pi, U(\pi)^{-1}) = e$, hence the matter

fields χ^ϱ remain unaffected. The effective action can therefore be built out of the composite fields $T_{U(\pi)^{-1}}A_\mu$ and $\Xi_\varrho(\pi, m)$, and χ^ϱ . The rest is just group theory. Since the matter fields may also enter the Lagrangian with derivatives, it is useful to have at hand their covariant derivative,

$$D_\mu \chi^\varrho \equiv \partial_\mu \chi^\varrho + i\Omega_\mu^\alpha D(Q_\alpha)^\varrho{}_\sigma \chi^\sigma . \quad (8.70)$$

The transformation rule (8.69) is not necessarily the only possible choice for the action of G on matter fields. Suppose that the (presumably reducible) representation D of H can be extended to a linear representation of the whole group G on the same set of fields χ^ϱ . The redefinition $\chi^\varrho \rightarrow \Psi^\varrho \equiv D(U(\pi))^\varrho{}_\sigma \chi^\sigma$ then gives variables that transform linearly under the whole G , $\Psi^\varrho \xrightarrow{g} D(g)^\varrho{}_\sigma \Psi^\sigma$. Such fields, while superficially natural, however conceal the physical structure of the spectrum. The degenerate energy levels are still organized in multiplets of the unbroken subgroup H . States from different multiplets of H , even if formally belonging to the same multiplet of G , will have different dispersion relations, and different interactions. In the extreme case, it may not even be possible to form complete multiplets of G . The arguably more complicated nonlinear transformation rule (8.69) is then the only option. An explicit example will illustrate this best.

Example 8.9

Most natural ferromagnets are metals, which betrays the presence of gapless, conducting electrons. While their appearance in spin doublets would seem natural, this is no longer mandatory once the $G \simeq \text{SU}(2)$ spin symmetry is spontaneously broken down to $H \simeq \text{U}(1)$. Let us denote the two possible polarizations of the electron as up (\uparrow) and down (\downarrow). Each of these carries a one-dimensional complex representation of H . The action of the full spin group G is then defined in accord with (8.69),

$$\chi_\uparrow \xrightarrow{g} D_\uparrow(h(\pi, g))\chi_\uparrow , \quad \chi_\downarrow \xrightarrow{g} D_\downarrow(h(\pi, g))\chi_\downarrow . \quad (8.71)$$

In the effective Lagrangian, Schrödinger-type terms for the up and down electrons can be added independently from each other,

$$\mathcal{L}_{\text{eff}} \ni \sum_{\sigma=\uparrow, \downarrow} c_\sigma \chi_\sigma^\dagger \left(iD_0 + \frac{\mathbf{D}^2}{2m_\sigma} + \mu_\sigma \right) \chi_\sigma + \cdots . \quad (8.72)$$

Here m_σ and μ_σ are the effective mass and chemical potential of χ_σ , and the ellipsis denotes other possible operators, compatible with all the assumed symmetries. The covariant derivatives induce coupling between the electrons and the magnon (spin wave) degrees of freedom. There is nothing that prevents us from setting one of c_\uparrow , c_\downarrow to zero: we do not need both spin polarizations to preserve the invariance under

SU(2). Physically, this amounts to the possibility of magnons interacting with a fully polarized Fermi sea of electrons or other spin-1/2 fermions such as neutrons.

Another important application of the formalism for coupling matter fields to NG bosons is the description of interactions between baryons and pseudoscalar mesons. The latter are pseudo-NG bosons of a spontaneously broken symmetry of quantum chromodynamics. Baryons play the role of matter fields that are much heavier than the NG modes, but may nevertheless be added to the low-energy EFT if desired. An interested reader will find more details in Chap. 4 of [13].

8.3 Equation of Motion

For many applications, it is useful to have at hand the EoM for the NG bosons. This is of obvious value in classical physics, it is however also very useful in quantum field theory. First, the linearized EoM can be used to identify the spectrum of excitations above the ground state and their perturbative propagators. Second, within a derivative expansion of the EFT, knowing the leading-order EoM helps to eliminate redundancies in the effective Lagrangian at higher orders.

Let us return to the general effective Lagrangian (8.33) without background gauge fields, this time augmented with the perturbation (8.66). Deriving the EoM from the effective action amounts in principle to a mere variation with respect to the NG field π^a . However, the complicated dependence of the Lagrangian on π^a makes this a rather odious task. I will therefore make a rare exception and leave out all details, simply displaying the final result. A reader wishing to verify it should be prepared to a repeated use of the MC equation (8.11) and of the algebraic constraints on all the effective couplings.

I will use the shorthand notation $\omega_\mu^A(\pi) \equiv \omega_a^A(\pi)\partial_\mu\pi^a$. Also, I will need the covariant derivative of the broken part of the MC form,

$$D_\mu\omega_\nu^a(\pi) \equiv \partial_\mu\omega_\nu^a(\pi) - f_{ab}^c\omega_\mu^b(\pi)\omega_\nu^c(\pi) . \quad (8.73)$$

With these preparations, the EoM descending from the lowest-order effective Lagrangian (8.33) can be written as

$$\begin{aligned} & \sigma_{ab}\omega_0^b - (\bar{\kappa}_{ab}D_0\omega_0^b - f_{ab}^d\bar{\kappa}_{cd}\omega_0^b\omega_0^c) + (\kappa_{ab}\delta^{rs}D_r\omega_s^b - f_{ab}^d\kappa_{cd}\delta^{rs}\omega_r^b\omega_s^c) \\ & + \frac{1}{2}(f_{ab}^d\lambda_{cd} + f_{bc}^d\lambda_{ad} + f_{ca}^d\lambda_{bd})\varepsilon^{rs}\omega_r^b\omega_s^c + m_\varrho\mathcal{R}(U(\pi))^\varrho{}_\sigma\mathcal{R}(\mathbf{i}Q_a)^\sigma{}_\tau\eta^\tau = 0 . \end{aligned} \quad (8.74)$$

This does not appear particularly elegant, but it is not that bad. In concrete applications, one or more of the terms in (8.74) are often missing. The last term obviously vanishes in the limit of exact symmetry. The term proportional to ε^{rs} is absent in $d \neq 2$ spatial dimensions. Moreover, all the terms containing a structure constant with three broken indices, f_{ab}^c or similar, drop for symmetric coset spaces (see

Sect. 7.3.2). One last special case, in which the form of (8.74) drastically simplifies, deserves spelling out explicitly.

Example 8.10

Consider the class of EFTs for Lorentz-invariant systems. This amounts to setting $\bar{\kappa}_{ab} = \kappa_{ab}$ as well as to dropping the single-time-derivative term and the purely spatial two-dimensional term proportional to ε^{rs} ,

$$\kappa_{ab} g^{\mu\nu} D_\mu \omega_\nu^b - f_{ab}^d \kappa_{cd} g^{\mu\nu} \omega_\mu^b \omega_\nu^c - m_\varrho \mathcal{R}(U(\pi))^\varrho {}_\sigma \mathcal{R}(\mathrm{i}Q_a)^\sigma {}_\tau \eta^\tau = 0. \quad (8.75)$$

Under the additional assumption that the coset space G/H is symmetric, which is the case for many relevant physical systems, the EoM takes the extremely compact form $\kappa_{ab} g^{\mu\nu} D_\mu \omega_\nu^b - m_\varrho \mathcal{R}(U(\pi))^\varrho {}_\sigma \mathcal{R}(\mathrm{i}Q_a)^\sigma {}_\tau \eta^\tau = 0$. This is essentially a nonlinear generalization of the Klein–Gordon equation to symmetric coset spaces.

A couple of remarks on the main result (8.74) are due. First, the last term proportional to m_ϱ looks like it might contribute a nonzero constant in the limit $\pi^a \rightarrow 0$. That would indicate an instability of the origin of the coset space under the symmetry-breaking perturbation. This is hardly surprising; in presence of the perturbation, we can no longer expect all the points of the coset space to correspond to physically equivalent vacua. In order to ensure that $\pi^a = 0$ is at least a stationary point of the potential induced by the perturbation, one has to demand that such a constant term in (8.74) is absent. That amounts to the condition

$$m_\varrho \mathcal{R}(Q_a)^\varrho {}_\sigma \eta^\sigma = 0. \quad (8.76)$$

Second, the contribution of the two-dimensional term proportional to ε^{rs} vanishes if $f_{ab}^d \lambda_{cd} + f_{bc}^d \lambda_{ad} + f_{ca}^d \lambda_{bd} = 0$. This implies by means of (8.24) that the 2-form $(1/2) \lambda_{ab} \omega^a(\pi) \wedge \omega^b(\pi)$ is closed, hence contributes a mere surface term to $\mathcal{L}_{\text{eff}}^{(2,0)}$. This makes sense, since surface terms do not affect the EoM.

What if we want to know how the EoM depends on the external gauge fields A_μ^A ? It might appear we could obtain a gauge-invariant EoM by replacing everywhere in (8.74) ω_μ^a and ω_μ^α with their gauged counterparts $\Omega_\mu^a, \Omega_\mu^\alpha$. Yet, there are subtleties hidden in this naive guess. First, we know from Sects. 8.1 and 8.2 that there are admissible values of σ_{ab} for which the symmetry under G cannot be gauged. The catch is that not every modification of (8.74) renders it a well-defined variational equation. That is, it may not be possible to construct a Lagrangian bottom-up from the gauged EoM. It is not difficult to show that consistency of the gauged EoM and background gauge invariance of the corresponding action require that σ_{ab} satisfies the constraints (8.28).

Second, it is not obvious that even when the symmetry under G can be gauged, the minimal replacement of ω_μ^a and ω_μ^α with Ω_μ^a and Ω_μ^α gives the correct gauged EoM. Repeating the derivation of the EoM but this time starting

with the gauged Lagrangian (8.60), one finds instead of (8.74) the following,

$$\begin{aligned} & -f_{ab}^C \sigma_C \Omega_0^b - (\bar{\kappa}_{ab} D_0 \Omega_0^b - f_{ab}^d \bar{\kappa}_{cd} \Omega_0^b \Omega_0^c) + (\kappa_{ab} \delta^{rs} D_r \Omega_s^b \\ & - f_{ab}^d \kappa_{cd} \delta^{rs} \Omega_r^b \Omega_s^c) + \frac{1}{2} (f_{ab}^d \lambda_{cd} + f_{bc}^d \lambda_{ad} + f_{ca}^d \lambda_{bd}) \varepsilon^{rs} \Omega_r^b \Omega_s^c \quad (8.77) \\ & - \frac{1}{2} \lambda_{ab} \varepsilon^{rs} [U(\pi)^{-1} F_{rs} U(\pi)]^b + m_\varrho \mathcal{R}(U(\pi))^\varrho{}_\sigma \mathcal{R}(iQ_a)^\sigma{}_\tau \eta^\tau = 0 . \end{aligned}$$

Here $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]$ is the field-strength tensor of A_μ . The slight modification of the first term agrees with the fact that the symmetry under G can only be gauged if $\sigma_{ab} = -f_{ab}^C \sigma_C$. Most importantly, however, there is an extra term that could not have been guessed by the naive gauging of (8.74). This has interesting consequences. Even when the 2-form $(1/2)\lambda_{ab}\omega^a(\pi) \wedge \omega^b(\pi)$ is closed, thus contributing a mere surface term, gauging it leads to a nontrivial modification of the Lagrangian as well as the EoM.

8.3.1 Spectrum of Nambu–Goldstone Bosons Revisited

It is common to analyze the excitation spectrum by expanding the Lagrangian to second order in fluctuations around the ground state. Here I will take an alternative route to illustrate the utility of the EoM. Namely, I will linearize (8.74), that is expand it to first order in the NG fields π^a . For the sake of simplicity, I will discard all external fields including the perturbations m_ϱ . By invoking the exponential parameterization (8.17) in which $\omega_b^a(0) = \delta_b^a$, we get at once

$$\sigma_{ab} \hat{\pi}^b - \bar{\kappa}_{ab} \ddot{\pi}^b + \kappa_{ab} \nabla^2 \pi^b \approx 0 . \quad (8.78)$$

The \approx symbol reminds us of the linearization done. This has plane-wave solutions

$$\pi^a(\mathbf{x}, t) = \hat{\pi}^a e^{-iEt} e^{i\mathbf{p} \cdot \mathbf{x}} , \quad (8.79)$$

where the amplitude $\hat{\pi}^a$, energy E and momentum \mathbf{p} satisfy

$$-iE \sigma_{ab} \hat{\pi}^b + E^2 \bar{\kappa}_{ab} \hat{\pi}^b - \mathbf{p}^2 \kappa_{ab} \hat{\pi}^b = 0 . \quad (8.80)$$

We would now like to understand how the dispersion relations of the various NG modes are related to the matrices κ_{ab} , $\bar{\kappa}_{ab}$ and σ_{ab} . From Sect. 6.3.2, we expect to find type-A and type-B NG bosons. I even declared therein that the dispersion of type-A NG modes is linear in momentum, whereas that of type-B NG modes is quadratic. We are now in the position to justify this claim. First, one can always change the basis of the variables $\hat{\pi}^a$ so that $\kappa_{ab} = \delta_{ab}$. Moreover, σ_{ab} can be brought by an additional orthogonal transformation to a block-diagonal form,

$$\sigma_{ab} = \left(\begin{array}{ccc|ccc} 0 & \sigma_{12} & & & & \\ \sigma_{21} & 0 & & & & \\ & & 0 & \sigma_{34} & & \\ & & \sigma_{43} & 0 & & \\ & & & & \ddots & \\ & & & & & \ddots \\ \hline & & & & & 0 \\ & & & & & \ddots \\ & & & & & 0 \end{array} \right) \equiv \left(\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & 0 \end{array} \right). \quad (8.81)$$

The upper-left block Σ corresponds to the type-B sector; these are the fields that enter the bilinear part of $\mathcal{L}_{\text{eff}}^{(0,1)}$. The lower-right block will analogously correspond to the type-A sector. The numbers of the different types of modes agree with our previous counting rule (6.24).

One last simplification we can make is to diagonalize the lower-right part of $\bar{\kappa}_{ab}$ by yet another orthogonal transformation without spoiling the already reduced forms of κ_{ab} and σ_{ab} . Thus, $\bar{\kappa}_{ab}$ can be assumed to take the generic form

$$\bar{\kappa}_{ab} = \left(\begin{array}{c|c} A & B \\ \hline B^T & \Delta \end{array} \right), \quad (8.82)$$

where A is symmetric and Δ is positive-definite and diagonal. Altogether, the spectrum of NG bosons is obtained by imposing the condition that the determinant of the matrix of coefficients in (8.80) vanishes,

$$\det \left(\begin{array}{c|c} -iE\Sigma + E^2A - \mathbf{p}^2 & E^2B \\ \hline E^2B^T & E^2\Delta - \mathbf{p}^2 \end{array} \right) = 0. \quad (8.83)$$

The exact solution $E(\mathbf{p})$ for the dispersion relation of the various modes will be very complicated. However, the asymptotic behavior of the dispersion in the limit of low momentum is easy to extract from (8.83). Namely, in this limit, the contribution of the off-diagonal blocks can be neglected, as can be the E^2A term in the upper-left block. The asymptotic dispersions in the type-A and type-B sectors therefore descend from the already diagonalized matrix equations

$$E^2\Delta - \mathbf{p}^2 \approx 0 \quad (\text{type-A}), \quad iE\Sigma + \mathbf{p}^2 \approx 0 \quad (\text{type-B}). \quad (8.84)$$

The presence of both first and second time derivatives in the type-B sector indicates that (8.83) has solutions $E(\mathbf{p})$ with nonzero limit as $\mathbf{p} \rightarrow \mathbf{0}$. Taken at face value, the linearized EoM (8.78) thus predicts the existence of *gapped* modes, accompanying type-B NG states. The required balance of operators with one and two time derivatives may however violate the derivative expansion of the EFT. In general, the presence of such gapped partners of type-B NG bosons cannot be asserted from the symmetry-breaking pattern alone.

8.3.2 More on the Geometry of the Coset Space

We had a first look at homogeneous spaces from the point of view of differential geometry in Sect. 7.4. I showed that any coset space G/H possesses a collection of G -invariant (pseudo-)Riemannian metrics. In this chapter, we found a use for them: two such metrics, $g(\pi)$ and $\bar{g}(\pi)$, enter the part of the effective Lagrangian for NG bosons with two derivatives. In Sect. 7.4, I also introduced a class of affine connections on coset spaces. I have not made use of the ensuing curvature and torsion so far. However, we will see in Chap. 10 that these appear naturally in scattering amplitudes of NG bosons.

What I want to briefly discuss now is yet another geometric structure on homogeneous spaces that is intimately connected to the spectrum of NG bosons. The reader might, if needed, want to recall the contents of Sect. 4.3.1 before proceeding. The exposition below follows [2], to which the reader is referred for a much more thorough discussion of geometric and topological aspects of type-B NG bosons.

Suppose that the spectrum of NG bosons were purely type-B as in ferromagnets. The matrix σ_{ab} must then be nonsingular. The same applies to the 2-form $dc(\pi)$ that defines the part of the effective Lagrangian with one time derivative, $\mathcal{L}_{\text{eff}}^{(0,1)}$. Being simultaneously closed, this 2-form therefore establishes a symplectic structure on the coset space. The 1-form $c(\pi)$ is the corresponding symplectic potential. The NG fields that block-diagonalize the bilinear part of $\mathcal{L}_{\text{eff}}^{(0,1)}$ as in (6.23) constitute a set of (local) Darboux coordinates. The existence of a symplectic structure on G/H underlines that it should be treated as the phase space of the EFT. The EoM is of first order in time derivatives. The number of independent modes in the spectrum equals $(1/2) \dim G/H$.

What if the spectrum includes both type-A and type-B NG bosons? Then the 2-form $dc(\pi)$ becomes singular, yet it still defines a new structure on G/H , called *presymplectic*. Intuitively, the coset space becomes partially a phase space and partially a configuration space. The paired NG fields giving rise to type-B NG bosons are “phase space coordinates.” These are augmented with additional “configuration space coordinates,” corresponding to the type-A NG fields. The dynamics of the former and the latter is respectively of first and second order in time.

Let us try to be a bit more precise while remaining physically intuitive. Suppose that the commutator matrix (8.26) were the sole order parameter our system

possesses. This would break the symmetry group G to some subgroup K . Suppose also that G is compact and that its Lie algebra \mathfrak{g} does not have any nontrivial central charges (for instance because G is semisimple). According to the discussion in Sect. 6.3.2, our order parameter then corresponds to the VEVs of a set of conserved charges that belong to a Cartan subalgebra of \mathfrak{g} . These charges together generate an Abelian subgroup $T \subset G$, usually called a *torus*. In group theory terminology, the subgroup $K = \{g \in G \mid gh = hg \ \forall h \in T\}$ is the *centralizer* of the torus T in G .

The NG modes owing their existence to the order parameter σ_{ab} should be described by an EFT that lives on the coset space G/K . A coset space G/K where K is the centralizer of a torus in G is called a *flag manifold*; see Chap. 7 of [14] for an introduction. Flag manifolds are known to carry a natural symplectic structure, corresponding to (8.23) with our order parameter σ_{ab} . In physics terms, the spectrum of the EFT on G/K contains by construction only type-B NG modes. Their number is $(1/2) \dim G/K$.

Example 8.11

Consider $G \simeq U(n)$ and the torus T generated by the diagonal matrices

$$\left(\begin{array}{c|c} \lambda_1 \mathbb{1}_{m \times m} & 0 \\ \hline 0 & \lambda_2 \mathbb{1}_{(n-m) \times (n-m)} \end{array} \right), \quad (8.85)$$

where $\lambda_{1,2} \in \mathbb{R}$. The centralizer of this torus in G consists of all block-diagonal unitary matrices, $K \simeq U(m) \times U(n-m)$. The flag manifold $U(n)/[U(m) \times U(n-m)]$ is called the (complex) *Grassmannian*. The special case of $m = 1$ is the *complex projective space*, $\mathbb{C}P^n \simeq U(n+1)/[U(n) \times U(1)]$.

How do we add the type-A NG bosons then? In presence of additional order parameters, the symmetry will be broken further down from K to H . This will lead to additional $\dim K/H$ modes, which we can identify with the type-A NG bosons. All this of course does not mean that we will have two different EFTs, one living on K/H for the type-A NG bosons, and the other living on G/K for the type-B NG bosons. The two coset spaces are geometrically integrated into G/H through a *fiber bundle* structure,⁵ $K/H \rightarrow G/H \xrightarrow{\pi} G/K$. Ignoring the type-A NG degrees of freedom amounts to the projection $\pi : G/H \rightarrow G/K$ from the *total space* G/H to the *base space* G/K . The type-A NG fields span the *fiber* K/H above each point of the base space. The presymplectic structure on G/H is obtained by pulling back the symplectic 2-form on the base space G/K via the projection map π .

⁵ Since this is the only place in the book where the concept of a fiber bundle appears, I have decided not to include a detailed explanation in Appendix A. I hope that even a reader without the necessary mathematical background can extract some useful information from the discussion.

Example 8.12

Consider a theory with a $G \simeq U(n)$ symmetry. Suppose that the symmetry is broken down to $H \simeq U(n-1)$ by the expectation value of a complex scalar field Φ that transforms in the fundamental representation of G . Since the action of G preserves the norm of Φ , the coset space $U(n)/U(n-1)$ is obviously equivalent to S^{2n-1} . For convenience, we can choose the order parameter as $\langle \Phi \rangle = (1, 0, \dots, 0)^T$. The Lie algebra \mathfrak{g} of G includes two linearly independent singlets of H , $Q_1 \equiv \mathbb{1}$ and $Q_2 \equiv \text{diag}(-n+1, 1, \dots, 1)$. Their linear combination, $Q_{\parallel} \equiv [Q_2 + (n-1)Q_1]/n = \text{diag}(0, 1, \dots, 1)$, generates the sole $U(1)$ factor of H . The generator Q_1 is a singlet of the whole group G and its VEV therefore does not break the symmetry. On the other hand, the VEV of Q_2 constitutes a candidate order parameter responsible for type-B NG bosons in the spectrum. See the closely related Example 6.6.

If $\langle Q_2 \rangle = 0$, or $\langle Q_{\parallel} \rangle = (n-1)/n \langle Q_1 \rangle$, we have no order parameter, hence $K \simeq G$. In this case, the coset space G/H is pure type-A; there are $2n-1$ type-A NG modes. If, however, $\langle Q_2 \rangle \neq 0$, there are $n-1$ type-B NG modes that parameterize the coset space $G/K \simeq U(n)/[U(1) \times U(n-1)] \simeq \mathbb{C}P^{n-1}$. Above each point of this base space, there is a fiber $K/H \simeq U(1)$, carrying one type-A NG degree of freedom.

References

1. S. Weinberg, *The Quantum Theory of Fields*, vol. I (Cambridge University Press, Cambridge, UK, 1995)
2. H. Watanabe, H. Murayama, Phys. Rev. **X4**, 031057 (2014). DOI [10.1103/PhysRevX.4.031057](https://doi.org/10.1103/PhysRevX.4.031057)
3. H. Goldstein, C.P. Poole, J.L. Safko, *Classical Mechanics* (Pearson Education Limited, Harlow, UK, 2013)
4. G. Goon, K. Hinterbichler, A. Joyce, M. Trodden, JHEP **06**, 004 (2012). DOI [10.1007/JHEP06\(2012\)004](https://doi.org/10.1007/JHEP06(2012)004)
5. J.A. de Azcárraga, J.M. Izquierdo, J.C. Pérez Bueno, Rev. R. Acad. Cien. Exactas Fis. Nat. Ser. A Mat. **95**, 225 (2001)
6. J. Davighi, B. Gripaios, JHEP **09**, 155 (2018). DOI [10.1007/JHEP09\(2018\)155](https://doi.org/10.1007/JHEP09(2018)155)
7. H. Leutwyler, Ann. Phys. **235**, 165 (1994). DOI [10.1006/aphy.1994.1094](https://doi.org/10.1006/aphy.1994.1094)
8. H. Leutwyler, Phys. Rev. **D49**, 3033 (1994). DOI [10.1103/PhysRevD.49.3033](https://doi.org/10.1103/PhysRevD.49.3033)
9. J.O. Andersen, T. Brauner, C.P. Hofmann, A. Vuorinen, JHEP **08**, 088 (2014). DOI [10.1007/JHEP08\(2014\)088](https://doi.org/10.1007/JHEP08(2014)088)
10. S. Weinberg, *The Quantum Theory of Fields*, vol. II (Cambridge University Press, Cambridge, UK, 1996)
11. J.I. Kapusta, C. Gale, *Finite-temperature field theory: Principles and applications* (Cambridge University Press, Cambridge, UK, 2006)
12. H. Watanabe, T. Brauner, H. Murayama, Phys. Rev. Lett. **111**, 021601 (2013). DOI [10.1103/PhysRevLett.111.021601](https://doi.org/10.1103/PhysRevLett.111.021601)
13. S. Scherer, M.R. Schindler, *A Primer for Chiral Perturbation Theory, Lecture Notes in Physics*, vol. 830 (Springer, 2012). DOI [10.1007/978-3-642-19254-8](https://doi.org/10.1007/978-3-642-19254-8)
14. A. Arvanitoyeorgos, *An Introduction to Lie Groups and the Geometry of Homogeneous Spaces, Student Mathematical Library*, vol. 22 (American Mathematical Society, 2003)

Chapter 9

Applications to Particle and Condensed-Matter Physics

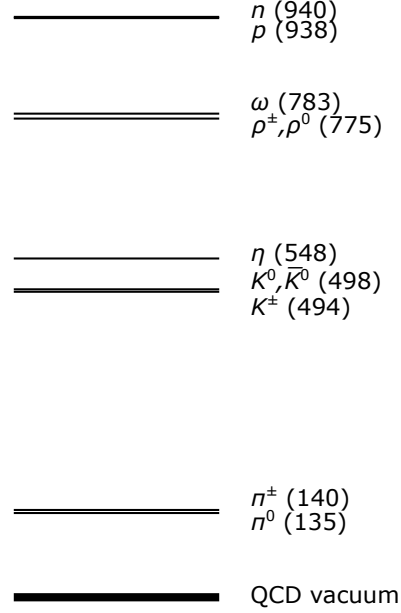
The material of Chaps. 7 and 8 is in principle sufficient to construct the low-energy *effective field theory* (EFT) for any system with a spontaneously broken internal symmetry. However, the implementation of the general methodology for a concrete physical system often still requires a nontrivial amount of effort. Moreover, the actual phenomenology of the EFT may depend sensitively on the internal and spacetime symmetries present. In this chapter, I will therefore work out in detail two applications of the general formalism, one in particle physics and one in condensed-matter physics. Apart from serving as an extensive illustration, this will allow us to dive deeper into some technical details of EFTs for *Nambu–Goldstone* (NG) *bosons*: the consistency of the derivative expansion of the effective Lagrangian, and the topological aspects of quasi-invariant Lagrangians.

9.1 Chiral Perturbation Theory of Mesons

Historically, the development of EFT for NG bosons was largely motivated by the need for a phenomenological description of low-energy hadron physics.¹ On the one hand, computation of hadron properties from first principles is challenging because the strong nuclear interaction is not amenable to a perturbative treatment at energies below the intrinsic scale of *quantum chromodynamics* (QCD). On the other hand, the spectrum of lightest hadrons exhibits scale separation; see Fig. 9.1. The eight lightest particles in the spectrum—the three pions, four kaons and the η -meson—are all pseudoscalars with varying quark flavor composition. Their masses are much lower than what one would expect given the constituent mass of about 300 MeV for the u, d quarks and nearly 500 MeV for the s quark. This suggests that the low-energy physics of the light pseudoscalar mesons should be captured by an EFT.

¹ This is also reflected by the widespread terminology. In the high-energy physics literature, it is still quite common to refer to NG bosons collectively as “pions.”

Fig. 9.1 Mass spectrum of light hadrons. All the masses in parentheses are shown in MeV. Current numerical values as of writing this book can be found in the Review of Particle Physics [1]



The key to understanding the observed scale separation in the hadron spectrum is provided by *spontaneous symmetry breaking* (SSB). The gauge interaction of QCD is vector-like, that is, gluons couple equally to left- and right-handed quarks. Moreover, the interaction is insensitive to the quark flavor. As a consequence, in the limit of vanishing quark masses, QCD with n_f quark flavors possesses a $G \simeq \text{SU}(n_f)_L \times \text{SU}(n_f)_R \times \text{U}(1)_B$ symmetry. Under an element $(g_L, g_R, e^{i\epsilon}) \in G$, the left- and right-handed quark spinors $\Psi_{L,R}$ transform as

$$\Psi_L \rightarrow e^{i\epsilon/3} g_L \Psi_L, \quad \Psi_R \rightarrow e^{i\epsilon/3} g_R \Psi_R. \quad (9.1)$$

The unitary matrices $g_{L,R}$ act on the flavor index of the quark field. The vector-like $\text{U}(1)_B$ group corresponds to conservation of baryon number, and the factors $1/3$ in (9.1) indicate the baryon number of a single quark. In addition to this internal *chiral symmetry*, QCD is invariant under the spacetime Poincaré group, and under the discrete symmetries of charge conjugation, parity and time reversal.

In the ground state of (still massless) QCD, the chiral symmetry is spontaneously broken down to the “vector” subgroup, $H \simeq \text{SU}(n_f)_V \times \text{U}(1)_B$, consisting of elements of the type $(g, g, e^{i\epsilon})$, that is, $g_L = g_R$. This implies the existence of $n_f^2 - 1$ pseudoscalar NG bosons. Restricting to $n_f = 2$ accounts for the lightest mesons: the pion triplet. The additional five modes that appear with $n_f = 3$ are the strange pseudoscalar mesons (kaons) and the η -meson. Of course, in reality, none of these is exactly massless. They are all pseudo-NG bosons owing to the fact that the chiral symmetry is only approximate, being explicitly broken by current quark masses. The fact that the strange mesons are heavier than the pions is a consequence of the s

quark being considerably heavier than the u and d quarks. The masses of the c , t and b quarks are so high (above the intrinsic scale of QCD) that with $n_f \geq 4$, QCD does not possess even an approximate chiral symmetry. In practice, it is therefore sufficient to focus on the $n_f = 2, 3$ cases.

The classical Lagrangian of massless QCD is also invariant under the *axial symmetry* $U(1)_A$, which amounts to $\Psi_{L,R} \rightarrow e^{\pm i\epsilon} \Psi_{L,R}$. Being also spontaneously broken in the QCD vacuum, this would suggest the existence of another, flavor-singlet, pseudoscalar meson. Yet the lightest available candidate is the η' -meson with the mass of 958 MeV. This is too high to be accounted for by explicit symmetry breaking due to quark masses. The resolution of this so-called $U(1)_A$ problem is that the axial symmetry is broken at the quantum level by the axial anomaly, arising from nonperturbative gluon dynamics. See Sect. 13.6 of [2] for further details.

The EFT for the light pseudoscalar mesons, constructed below, is called *chiral perturbation theory* (ChPT).² The reader will find a modern graduate-level exposition of ChPT in the dedicated monograph [3]. However, the pioneering works of Gasser and Leutwyler [4, 5] remain a valuable source of insight. The construction of ChPT will follow the general machinery developed in Sect. 8.2 with the appropriate coset space, $G/H \simeq [SU(n_f)_L \times SU(n_f)_R]/SU(n_f)_V$. Here I have tacitly dropped the $U(1)_B$ factor of the symmetry group of QCD. This remains unbroken and moreover leaves the meson fields intact, and so has no effect on the invariant part of the ChPT Lagrangian. I will reinstate the $U(1)_B$ symmetry in Sect. 9.1.4, where it will help us build the quasi-invariant (anomalous) part of the Lagrangian.

9.1.1 Power Counting

In a Lorentz-invariant EFT, the effective Lagrangian is organized by the total number, $n = s + t$, of spatial and temporal derivatives,

$$\mathcal{L}_{\text{eff}}[\pi] = \sum_{n \geq 2} \mathcal{L}_{\text{eff}}^{(n)}[\pi] . \quad (9.2)$$

Discounting possible tadpole operators, linear in π^a , a minimum of two derivatives is enforced by the nonlinearly realized internal symmetry and Lorentz invariance. The latter is ensured by contracting Lorentz indices of spacetime derivatives with the Minkowski metric $g_{\mu\nu}$ or the *Levi-Civita* (LC) *tensor*, $\varepsilon_{\mu\nu\lambda\dots}$. Hence, in an even

² I will adhere to this established name but stress that it is a misnomer. The derivative expansion of ChPT, worked out below, has nothing to do with ordinary perturbation theory known from quantum field theory.

number of spacetime dimensions D (which is the case of QCD, where $D = 4$), only operators with even n can contribute to (9.2).

In order to assess the relative importance of the individual parts $\mathcal{L}_{\text{eff}}^{(n)}$, consider a generic Feynman diagram Γ , contributing to a given observable. Suppose the diagram contains altogether I internal propagators, L loops and V_n interaction vertices from each of $\mathcal{L}_{\text{eff}}^{(n)}$. Upon Fourier transform, the diagram will evaluate to a homogeneous function of energy–momenta on the external legs. Adding up all the powers of energy–momentum from the vertices, propagators and loop integrals, the degree of the diagram as a function of the external energy–momenta then reads

$$\deg \Gamma = DL - 2I + \sum_{n \geq 2} nV_n . \quad (9.3)$$

The number of propagators I can be eliminated using the identity $I = L + \sum_n V_n - 1$, known from graph theory to hold for any connected graph (see, for instance, Sect. 13.4 of [6]). This leads to

$$\deg \Gamma = 2 + (D - 2)L + \sum_{n \geq 2} (n - 2)V_n . \quad (9.4)$$

Diagrammatic contributions to a given observable can now be organized by increasing powers of energy–momentum; higher powers imply stronger suppression at low energies. Any Feynman diagram will have $\deg \Gamma \geq 2$. The *leading-order* (LO) contribution to the observable corresponds to $\deg \Gamma = 2$ and consists of all tree-level ($L = 0$) diagrams with all vertices from $\mathcal{L}_{\text{eff}}^{(2)}$. This justifies a posteriori the focus of this book on classical Lagrangians with the lowest possible number of derivatives.

The degree $\deg \Gamma$ can be increased by adding vertices from higher-order ($n \geq 3$) parts of the Lagrangian or by adding loops. In $D = 4$ dimensions, the *next-to-leading order* (NLO) corresponds to $\deg \Gamma = 4$. This can be reached in two different ways. Either we restrict ourselves to tree-level diagrams but allow for one vertex from $\mathcal{L}_{\text{eff}}^{(4)}$, or allow one loop but keep all vertices from $\mathcal{L}_{\text{eff}}^{(2)}$. It is easy to extend this reasoning to classify possible contributions to any observable at even higher orders of the derivative expansion.

Let me conclude the discussion of power counting with several remarks. First, for pseudo-NG bosons, the propagator $i/(p^2 - m^2)$ is actually not a homogeneous function of the energy–momentum p^μ because of the mass m . That can be fixed by assigning the mass a formal counting degree 1 so that m^2 counts equally to p^2 . This affects the classification of operators in the effective Lagrangian of ChPT that incorporate the effects of explicit breaking of chiral symmetry by the quark mass, m_q . We expect based on Sect. 6.4.1 that m will scale as $\sqrt{m_q}$. Hence, we can continue using (9.4), as long as we count the quark mass as a small quantity of degree $\deg m_q = 2$. Similarly, we will want to couple ChPT to a set of background gauge fields, A_μ . In order for the power counting to be consistent with background gauge invariance, we must assign these fields the degree $\deg A_\mu = 1$.

Second, the counting rule (9.4) also tells us how to renormalize the EFT. In order to remove an overall divergence in a Feynman graph Γ , we need a counterterm among

the operators in $\mathcal{L}_{\text{eff}}^{(\deg \Gamma)}$. Since all diagrams at $\deg \Gamma = 2$ are tree-level, the effective couplings in $\mathcal{L}_{\text{eff}}^{(2)}$ are finite constants that do not run with the renormalization scale. The effective couplings in $\mathcal{L}_{\text{eff}}^{(n \geq 3)}$ must include divergences that cancel against loop diagrams with degree n . Their running will be determined by the corresponding renormalization group equation. Important is that to any finite order in the derivative expansion, only a finite number of operators, and a finite number of counterterms, is needed.

Third, note that for $D = 2$, all loop diagrams with the same $V_{n \geq 4}$ and arbitrary V_2 contribute at the same order of the derivative expansion. The lack of suppression of loop effects hints that the EFT is no longer weakly coupled. Eventually, it turns out that the infrared fluctuations of NG bosons are so wild in $D = 2$ dimensions that they destroy the order parameter leading to SSB. I will return to this in Sect. 15.2.

Finally, the basic counting rule (9.4) and much of the above comments applies equally to any EFT with only type-A NG bosons, relativistic or not. The LO Lagrangian will still be $\mathcal{L}_{\text{eff}}^{(2)}$. The LO contribution to any observable will come from tree-level diagrams with all vertices from $\mathcal{L}_{\text{eff}}^{(2)}$. The classification of contributions at higher orders may however be modified by the presence of operators with odd $n \geq 3$ regardless of the spacetime dimension D . The details for any particular choice of effective Lagrangian and D are easy to work out using (9.4).

9.1.2 Effective Lagrangian

The coset space of QCD, $G/H \simeq [\text{SU}(n_f)_L \times \text{SU}(n_f)_R]/\text{SU}(n_f)_V$, is symmetric. Let me therefore start by recalling some general properties of symmetric coset spaces; see Sect. 7.3.2 for full detail.

By definition, a coset space is symmetric if the Lie algebra \mathfrak{g} of G possesses an automorphism \mathcal{R} that acts as identity on the unbroken subalgebra \mathfrak{h} and as minus identity on the complementary subspace $\mathfrak{g}/\mathfrak{h}$ of \mathfrak{g} . Loosely speaking, \mathcal{R} changes the sign of all broken generators of G while leaving all unbroken generators intact. The automorphism \mathcal{R} can be, at least locally, lifted from the Lie algebra \mathfrak{g} to the Lie group G . One can then choose the coset representative $U(\pi)$ so that $\mathcal{R}(U(\pi)) = U(\pi)^{-1}$. With this choice, one can build a matrix-valued NG field that transforms linearly under the entire group G ,

$$\Sigma(\pi) \equiv U(\pi)^2, \quad \Sigma(\pi) \xrightarrow{g} \Sigma(\pi'(\pi, g)) = g \Sigma(\pi) \mathcal{R}(g)^{-1}. \quad (9.5)$$

The automorphism \mathcal{R} makes it easy to project out the broken component of the gauged *Maurer–Cartan* (MC) form (8.47),

$$\Omega_{\perp}(\pi, A) = \frac{1}{2} [\Omega(\pi, A) - \mathcal{R}(\Omega(\pi, A))] , \quad (9.6)$$

where $A \equiv A_{\mu} dx^{\mu}$ is the \mathfrak{g} -valued 1-form gauge field of G . Upon a brief manipulation using the definition of \mathcal{R} , we find that

$$\Omega_{\perp}(\pi, A) = -\frac{i}{2}U(\pi)^{-1}D\Sigma(\pi)U(\pi)^{-1} = \frac{i}{2}U(\pi)D\Sigma(\pi)^{-1}U(\pi), \quad (9.7)$$

where

$$D\Sigma(\pi) = d\Sigma(\pi) - iA\Sigma(\pi) + i\Sigma(\pi)\mathcal{R}(A) \quad (9.8)$$

is the G -covariant derivative of $\Sigma(\pi)$, and $D\Sigma(\pi)^{-1}$ is defined analogously.

What does all this translate to in case of ChPT? Here \mathcal{R} acts by swapping the transformations of left- and right-handed quarks, $\mathcal{R}(g_L, g_R) = (g_R, g_L)$. It is then natural to choose the coset representative as

$$U(\pi) = (u(\pi), u(\pi)^{-1}) \quad \text{where } u(\pi) \in \text{SU}(n_f). \quad (9.9)$$

The matrix variable $\Sigma(\pi) = (u(\pi)^2, u(\pi)^{-2})$ is subject to the linear transformation $\Sigma(\pi) \xrightarrow{(g_L, g_R)} (g_L, g_R)\Sigma(\pi)(g_R, g_L)^{-1}$. All information about the NG fields can then be encoded in a single $\text{SU}(n_f)$ -valued matrix variable,

$$\mathcal{U}(\pi) \equiv u(\pi)^2, \quad \mathcal{U}(\pi) \xrightarrow{(g_L, g_R)} \mathcal{U}(\pi'(\pi, g_L, g_R)) = g_L \mathcal{U}(\pi) g_R^{-1}. \quad (9.10)$$

The Lie algebra of $\text{SU}(n_f)_L \times \text{SU}(n_f)_R$ is $\mathfrak{su}(n_f)_L \oplus \mathfrak{su}(n_f)_R$. In our pair notation, the matrix-valued gauge field A_{μ} can thus be decomposed as $A_{\mu} = (A_{\mu}^L, \mathbb{1}) + (\mathbb{1}, A_{\mu}^R)$, where $A_{\mu}^{L,R}$ are independent $\mathfrak{su}(n_f)$ -valued gauge fields acting respectively on the left- and right-handed quarks. A straightforward manipulation then leads to $D\Sigma(\pi) = (D\mathcal{U}(\pi), \mathcal{U}(\pi)^{-1}) + (\mathcal{U}(\pi), D\mathcal{U}(\pi)^{-1})$, where

$$\begin{aligned} D_{\mu}\mathcal{U}(\pi) &= \partial_{\mu}\mathcal{U}(\pi) - iA_{\mu}^L\mathcal{U}(\pi) + i\mathcal{U}(\pi)A_{\mu}^R, \\ D_{\mu}\mathcal{U}(\pi)^{-1} &= \partial_{\mu}\mathcal{U}(\pi)^{-1} - iA_{\mu}^R\mathcal{U}(\pi)^{-1} + i\mathcal{U}(\pi)^{-1}A_{\mu}^L. \end{aligned} \quad (9.11)$$

This eventually gives

$$\Omega_{\perp} = -\frac{i}{2}[(u^{-1}(D\mathcal{U})u^{-1}, \mathbb{1}) + (\mathbb{1}, u(D\mathcal{U}^{-1})u)] . \quad (9.12)$$

Before we can write down even the LO effective Lagrangian of ChPT, we still have to discuss the explicit breaking of chiral symmetry by the quark masses. The mass term in the microscopic Lagrangian of QCD takes the form $\bar{\Psi}_L \mathcal{M} \Psi_R + \bar{\Psi}_R \mathcal{M}^{\dagger} \Psi_L$, where $\mathcal{M} = \text{diag}(m_u, m_d, \dots)$ is a real diagonal matrix, collecting the quark masses. In the spirit of Sect. 8.2.3, this is promoted to a complex matrix field that transforms under the chiral symmetry as $\mathcal{M} \xrightarrow{(g_L, g_R)} g_L \mathcal{M} g_R^{-1}$, thus restoring exact chiral invariance of the QCD Lagrangian. The basic building block for incorporating the effects of explicit symmetry breaking in ChPT is then the matrix field

$$\Xi(\pi, \mathcal{M}) = u(\pi)^{-1} \mathcal{M} u(\pi)^{-1}. \quad (9.13)$$

Being linear in quark masses, this is assigned the order $\text{deg } \Xi = 2$.

The operators we now have for building the LO effective Lagrangian of ChPT are, schematically, $\Omega_{\perp\mu}\Omega_{\perp}^{\mu}$ and Ξ . The only way to ensure invariance of these operators

under the linearly realized unbroken subgroup, $SU(n_f)_V$, is to take a trace. Moreover, Ξ has to enter the Lagrangian through $\text{tr}(\Xi + \Xi^\dagger)$. This is required by parity invariance of QCD and the fact that parity interchanges left- and right-handed quarks, thereby acting on both $u(\pi)$ and \mathcal{M} by Hermitian conjugation. At the end of the day, there are only two independent operators one can put into the LO Lagrangian,

$$\mathcal{L}_{\text{eff}}^{(2)} = \frac{f_\pi^2}{4} \text{tr}[D_\mu \mathcal{U}(\pi)^\dagger D^\mu \mathcal{U}(\pi)] + \frac{f_\pi^2 B}{2} \text{tr}[\mathcal{M} \mathcal{U}^\dagger(\pi) + \mathcal{M}^\dagger \mathcal{U}(\pi)] . \quad (9.14)$$

Accordingly, there are two independent parameters, conventionally denoted as f_π and B , both with mass dimension 1.

Example 9.1

Let us work out some immediate consequences of the LO Lagrangian (9.14). To that end, I will drop the background gauge fields and parameterize $\mathcal{U}(\pi)$ in terms of a Hermitian traceless matrix $\Pi(\pi)$ as $\mathcal{U}(\pi) = \exp[i\Pi(\pi)/f_\pi]$. This is a matrix version of the familiar exponential parameterization of the coset space; the factor f_π is inserted to give Π mass dimension 1. Upon expansion in powers of Π , the Lagrangian becomes

$$\begin{aligned} \mathcal{L}_{\text{eff}}^{(2)} = & \frac{f_\pi^2 B}{2} \text{tr}(\mathcal{M} + \mathcal{M}^\dagger) \\ & + \frac{1}{4} \text{tr}[\partial_\mu \Pi(\pi) \partial^\mu \Pi(\pi)] - \frac{B}{4} \text{tr}[(\mathcal{M} + \mathcal{M}^\dagger) \Pi(\pi)^2] + \mathcal{O}(\Pi^4) , \end{aligned} \quad (9.15)$$

where I used that \mathcal{M} is ultimately Hermitian to drop terms linear and cubic in $\Pi(\pi)$.

The first, constant term contributes to the energy density of the chiral-symmetry-breaking vacuum. Taking the derivative of the vacuum energy density with respect to any of the quark masses in turn gives the *vacuum expectation value* (VEV) of the corresponding mass operator; cf. (5.16). The latter serves as the order parameter for spontaneous breaking of chiral symmetry. The LO prediction of ChPT, based on (9.15), is that this so-called *chiral condensate* is independent of the quark flavor. The total condensate, summed over all quark flavors, equals

$$\langle \bar{\Psi} \Psi \rangle = -n_f f_\pi^2 B . \quad (9.16)$$

Let us now turn attention to the bilinear part of (9.15). It is convenient to further parameterize the matrix $\Pi(\pi)$ in terms of $n_f^2 - 1$ meson fields π^a as $\Pi(\pi) = \pi^a \lambda_a$. Here λ_a is a basis of traceless Hermitian $n_f \times n_f$ matrices, normalized as $\text{tr}(\lambda_a \lambda_b) = 2\delta_{ab}$. This ensures canonical normalization of the kinetic term for the meson fields. In case of $n_f = 2$, the λ_a s are just the usual Pauli matrices τ_a , whereas for $n_f = 3$, the so-called *Gell-Mann matrices* can be used instead. The next step is to find the eigenvalues of the mass matrix for π^a . Setting $n_f = 3$ and using the known flavor composition of the light pseudoscalar mesons, these eigenvalues can be identified with the masses of the individual states in Fig. 9.1 as

$$\begin{aligned}
m_{\pi^0}^2 &= \frac{2B}{3} \left(m_u + m_d + m_s - \sqrt{m_u^2 + m_d^2 + m_s^2 - m_u m_d - m_u m_s - m_d m_s} \right), \\
m_{\pi^\pm}^2 &= B(m_u + m_d), \\
m_{K^\pm}^2 &= B(m_u + m_s), \\
m_{K^0}^2 &= B(m_d + m_s), \\
m_\eta^2 &= \frac{2B}{3} \left(m_u + m_d + m_s + \sqrt{m_u^2 + m_d^2 + m_s^2 - m_u m_d - m_u m_s - m_d m_s} \right).
\end{aligned} \tag{9.17}$$

This determines the five different meson masses in terms of the three current quark masses, yet the latter are not uniquely fixed by (9.17). Indeed, any overall rescaling of the quark masses can be absorbed into a redefinition of the B parameter. It is however possible to use (9.17) to eliminate B and the quark masses altogether and thus obtain a constraint on the meson spectrum,

$$2(m_{\pi^\pm}^2 + m_{K^\pm}^2 + m_{K^0}^2) = 3(m_{\pi^0}^2 + m_\eta^2). \tag{9.18}$$

This so-called *Gell-Mann–Okubo formula* can be interpreted, for instance, as giving an estimate for the mass of the η -meson in terms of those of the pions and kaons. With the data shown in Fig. 9.1 as input, one gets $m_\eta \approx 568$ MeV, which is less than four per cent off the correct value.

The machinery developed in Sect. 8.2 can in principle be applied to an arbitrarily high order of the derivative expansion of ChPT. However, already at NLO ($n = 4$), constructing the effective Lagrangian is a nontrivial exercise. I will therefore content myself with spelling out the final result and refer the reader to Sect. 3.2 of [7], where all details are worked out rather pedantically. For both $n_f = 2$ and $n_f = 3$, the invariant part of the ChPT Lagrangian at NLO can be written as

$$\begin{aligned}
\mathcal{L}_{\text{eff}}^{(4)} &= c_1 \text{tr}(D_\mu \mathcal{U} D^\mu \mathcal{U}^\dagger D_\nu \mathcal{U} D^\nu \mathcal{U}^\dagger) + c_2 \text{tr}(D_\mu \mathcal{U} D^\mu \mathcal{U}^\dagger) \text{tr}(D_\nu \mathcal{U} D^\nu \mathcal{U}^\dagger) \\
&\quad + c_3 \text{tr}(D_\mu \mathcal{U} D_\nu \mathcal{U}^\dagger) \text{tr}(D^\mu \mathcal{U} D^\nu \mathcal{U}^\dagger) \\
&\quad + c_4 \text{tr}(F_{\mu\nu}^L D^\mu \mathcal{U} D^\nu \mathcal{U}^\dagger + F_{\mu\nu}^R D^\mu \mathcal{U}^\dagger D^\nu \mathcal{U}) \\
&\quad + c_5 \text{tr}(F_{\mu\nu}^L \mathcal{U} F^{\mu\nu} \mathcal{U}^\dagger) + c_6 (F_{\mu\nu}^L F^{\mu\nu} + F_{\mu\nu}^R F^{\mu\nu}) \\
&\quad + d_1 \text{tr}(\mathcal{M} \mathcal{U}^\dagger) \text{tr}(\mathcal{M}^\dagger \mathcal{U}) + d_2 [(\text{tr} \mathcal{M} \mathcal{U}^\dagger)^2 + (\text{tr} \mathcal{M}^\dagger \mathcal{U})^2] \\
&\quad + d_3 \text{tr}(\mathcal{M} \mathcal{U}^\dagger \mathcal{M} \mathcal{U}^\dagger + \mathcal{M}^\dagger \mathcal{U} \mathcal{M}^\dagger \mathcal{U}) + d_4 \text{tr}(\mathcal{M} \mathcal{M}^\dagger) \\
&\quad + d_5 \text{tr}(\mathcal{M} \mathcal{U}^\dagger + \mathcal{M}^\dagger \mathcal{U}) \text{tr}(D_\mu \mathcal{U} D^\mu \mathcal{U}^\dagger) \\
&\quad + d_6 \text{tr}[(\mathcal{M} \mathcal{U}^\dagger + \mathcal{U} \mathcal{M}^\dagger) D_\mu \mathcal{U} D^\mu \mathcal{U}^\dagger].
\end{aligned} \tag{9.19}$$

Here $F_{\mu\nu}^L$ and $F_{\mu\nu}^R$ are the field-strength tensors of the background gauge fields. The 12 effective couplings c_{1-6} and d_{1-6} are mutually independent in the $n_f = 3$ case. In

case of $n_f = 2$, special properties of 2×2 matrices make the c_1 operator redundant with c_2 , and the d_6 operator redundant with d_5 .

The form of the NLO Lagrangian shown in (9.19) matches the detailed derivation offered in [7]. However, in the literature, a somewhat different basis of operators is often used. For the reader's convenience, I list here detailed relations between the parameters c_{1-6}, d_{1-6} introduced above and the more common NLO couplings of ChPT L_{1-10}, H_{1-2} , cf. Sect. 3.5.1 of [3],

$$\begin{aligned} c_1 &= L_3, & c_2 &= L_1, & c_3 &= L_2, & c_4 &= -iL_9, \\ c_5 &= L_{10}, & c_6 &= H_1, & d_1 &= 2(L_6 - L_7), & d_2 &= L_6 + L_7, \\ d_3 &= L_8, & d_4 &= H_2, & d_5 &= L_4, & d_6 &= L_5. \end{aligned} \quad (9.20)$$

9.1.3 Interaction with External Fields

For further illustrations of the use of ChPT, I will utilize its simplest version: the $n_f = 2$ ChPT in the “isospin-symmetric” limit. In this limit, one sets $m_u = m_d = m$ so that all three pions have the same squared mass, $m_\pi^2 = 2Bm$. Accordingly, the LO Lagrangian (9.14) reduces to

$$\mathcal{L}_{\text{eff}}^{(2)} = \frac{f_\pi^2}{4} [\text{tr}(D_\mu \mathcal{U}^\dagger D^\mu \mathcal{U}) + m_\pi^2 \text{tr}(\mathcal{U} + \mathcal{U}^\dagger)] . \quad (9.21)$$

The kinetic term can be further expanded and organized by powers of the background gauge fields,

$$\begin{aligned} \text{tr}(D_\mu \mathcal{U}^\dagger D^\mu \mathcal{U}) &= \text{tr}[\partial_\mu \mathcal{U}^\dagger \partial^\mu \mathcal{U} - iA_\mu^L (\mathcal{U} \partial^\mu \mathcal{U}^\dagger - \partial^\mu \mathcal{U} \mathcal{U}^\dagger) \\ &\quad - iA_\mu^R (\mathcal{U}^\dagger \partial^\mu \mathcal{U} - \partial^\mu \mathcal{U}^\dagger \mathcal{U}) - 2\mathcal{U} A_\mu^R \mathcal{U}^\dagger A^{L\mu} + A_\mu^L A^{L\mu} + A_\mu^R A^{R\mu}] . \end{aligned} \quad (9.22)$$

Knowing the explicit dependence on the background fields is, among others, a useful starting point for deriving the Noether currents of chiral symmetry. I will however focus on illustrating the implications of some particular choices of actual, physical background fields.

Example 9.2

As explained in Example 8.7, a chemical potential parameterizing the statistical equilibrium of a many-body system can be introduced in the Lagrangian as a constant temporal gauge field. Thus, the effects of nonzero density of isospin, that is the diagonal generator of $\text{SU}(2)_V$, can be captured by setting

$$A_\mu^L = A_\mu^R = \delta_{\mu 0} \mu_1 \frac{\tau_3}{2}, \quad (9.23)$$

where μ_1 is the isospin chemical potential. The actual statistical ground state is found by minimizing the Hamiltonian of ChPT with respect to \mathcal{U} . A detailed analysis shows that the ground state can be represented by a real orthogonal matrix,

$$\langle \mathcal{U} \rangle = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}. \quad (9.24)$$

For $|\mu_1| \leq m_\pi$, the value of the angle α minimizing energy is $\alpha = 0$, implying $\langle \mathcal{U} \rangle = \mathbb{1}$. This is the usual QCD vacuum. On the other hand, for $|\mu_1| \geq m_\pi$, the ground state corresponds to $\cos \alpha = m_\pi^2 / \mu_1^2$. This state describes *Bose–Einstein condensation* of charged pions. The condensate carries nonzero isospin density obtained as minus the derivative of the Hamiltonian density with respect to μ_1 ,

$$n_1 = f_\pi^2 \mu_1 \sin^2 \alpha = f_\pi^2 \mu_1 \left(1 - \frac{m_\pi^4}{\mu_1^4} \right). \quad (9.25)$$

Next, let us look at the spectrum of excitations above the ground state (9.24). This can be extracted by expanding the Lagrangian (9.21) to second order in fluctuations around (9.24). In the vacuum phase ($|\mu_1| < m_\pi$), the neutral pion maintains its relativistic dispersion relation, $E_{\pi^0}(\mathbf{p}) = \sqrt{\mathbf{p}^2 + m_\pi^2}$. The energy of the charged pions is, on the other hand, trivially shifted by the chemical potential, $E_{\pi^\pm}(\mathbf{p}) = \sqrt{\mathbf{p}^2 + m_\pi^2} \mp \mu_1$. In the pion condensation phase ($|\mu_1| > m_\pi$), the dispersion of the neutral pion changes to

$$E_{\pi^0}(\mathbf{p}) = \sqrt{\mathbf{p}^2 + \mu_1^2}. \quad (9.26)$$

This is a relativistic-looking dispersion, except that the “mass” equals $|\mu_1|$. That is not a coincidence. In the pion condensation phase, the neutral pion mode can be interpreted as a massive NG boson of the isospin $SU(2)_V$ symmetry; see Sect. 6.4.3 and [8] for details. In the charged pion sector, it is no longer possible to distinguish isospin (or electric charge) eigenstates as a result of SSB. There are two excitation branches that are mixtures of π^+ and π^- , and their squared energies are

$$E_\pm(\mathbf{p})^2 = \mathbf{p}^2 + \frac{\mu_1^2}{2} (1 + 3 \cos^2 \alpha) \pm \frac{\mu_1}{2} \sqrt{(1 + 3 \cos^2 \alpha)^2 \mu_1^2 + 16 \mathbf{p}^2 \cos^2 \alpha}. \quad (9.27)$$

The lower of the two branches is gapless, $E_-(\mathbf{0}) = 0$. This is the NG boson of the spontaneously broken isospin symmetry. Further discussion of meson condensates in QCD can be found for instance in the pedagogical review [9].

QCD alone does not encompass all of particle physics. Hadrons can also interact via the weak and electromagnetic forces. ChPT makes it easy to couple pseudoscalar

mesons to the electroweak sector of the Standard Model (see, for instance, Sect. 20.2 of [10] for an overview). Indeed, we can imitate the coupling of quarks to the electroweak gauge bosons by setting

$$A_\mu^L = \frac{g}{2} \boldsymbol{\tau} \cdot \mathbf{A}_\mu + \frac{g'}{6} B_\mu, \quad A_\mu^R = g' Q B_\mu. \quad (9.28)$$

Here \mathbf{A}_μ is a triplet of potentials of the weak isospin gauge group, $SU(2)_I$. Similarly, B_μ is the potential of the hypercharge gauge group, $U(1)_Y$. The corresponding gauge couplings are g, g' . Finally, Q is the matrix of electric charges of the u and d quarks,

$$Q = \begin{pmatrix} 2/3 & 0 \\ 0 & -1/3 \end{pmatrix} = \frac{1}{6} \mathbb{1} + \frac{1}{2} \tau_3. \quad (9.29)$$

A complete set of LO electroweak interactions of pions is then obtained by inserting the above definitions into (9.22).

Example 9.3

A glance at (9.22) shows that adding the electroweak gauge sector leads to nontrivial effects even in the ground state of ChPT, $\langle \mathcal{U} \rangle = \mathbb{1}$. Namely, the last three terms in (9.22) generate a mass term for the electroweak gauge bosons,

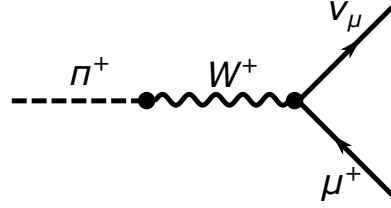
$$\begin{aligned} \frac{f_\pi^2}{4} \text{tr}(A_\mu^L A^{L\mu} + A_\mu^R A^{R\mu} - 2A_\mu^L A^{R\mu}) &= \frac{f_\pi^2}{4} \text{tr}[(A_\mu^L - A_\mu^R)(A^{L\mu} - A^{R\mu})] \\ &= \frac{f_\pi^2}{8} [(gA_\mu^1)^2 + (gA_\mu^2)^2 + (gA_\mu^3 - g'B_\mu)^2] \\ &= \frac{f_\pi^2}{8} [2g^2 W_\mu^+ W^{-\mu} + (g^2 + g'^2) Z_\mu Z^\mu]. \end{aligned} \quad (9.30)$$

In the last step, I introduced the charged gauge boson fields by $W_\mu^\pm \equiv (A_\mu^1 \pm iA_\mu^2)/\sqrt{2}$, and the neutral weak gauge boson via $Z_\mu \equiv A_\mu^3 \cos \theta_W - B_\mu \sin \theta_W$. Finally, the Weinberg angle θ_W is related to the gauge couplings by $\cos \theta_W = g/\sqrt{g^2 + g'^2}$. Thus, spontaneous breaking of chiral symmetry in QCD leads to the following contributions to the masses of the electroweak gauge bosons,

$$m_W^2 = \frac{1}{4} f_\pi^2 g^2, \quad m_Z^2 = \frac{1}{4} f_\pi^2 (g^2 + g'^2). \quad (9.31)$$

Given the characteristic scale of QCD, encoded in the value of f_π (fixed precisely below), these contributions are tiny. However, the idea that the gauge boson masses might be generated by a strong dynamics that spontaneously breaks chiral symmetry is intriguing. It lies behind the “technicolor” scenario of dynamical electroweak symmetry breaking. In this scenario, the Higgs boson is not elementary, but rather a composite bound state of constituent “techniquarks.” The value of f_π is expected to be of the order of the electroweak scale, that is a few hundreds of GeV. See [11] for a pedagogical introduction to technicolor models.

Fig. 9.2 Feynman diagram for the leading (tree-level) contribution to charged pion decay. The π^+-W^+ coupling is provided by ChPT whereas the interaction vertex between W^+ and the charged lepton current follows from the Standard Model of electroweak interactions



The physical value of f_π can be fixed by likewise utilizing the coupling of pions to the electroweak sector of the Standard Model. All we need is a single observable that does not depend on any other as yet unknown parameter. A suitable candidate is the leptonic decay of the charged pion.

Example 9.4

The charged pion π^+ decays with the probability of 99.988% into an antimuon μ^+ and a muon neutrino ν_μ . The leading perturbative contribution to the amplitude for this decay is shown in Fig. 9.2. The conversion of an on-shell pion into a virtual W -boson is described by our Lagrangian (9.21). Indeed, by expanding it to the first order in both W_μ^\pm and $\pi^\pm \equiv (\pi^1 \pm i\pi^2)/\sqrt{2}$, we find the bilinear term $-(f_\pi g/2)(W^{+\mu}\partial_\mu\pi^- + W^{-\mu}\partial_\mu\pi^+)$. The subsequent decay of the virtual W -boson into a lepton pair is governed by the “charged-current” interaction of the Standard Model, specifically the operator $(g/\sqrt{2})(W_\mu^+\bar{\mu}_L\gamma^\mu\nu_{\mu L} + W_\mu^-\bar{\nu}_{\mu L}\gamma^\mu\mu_L)$. I used an obvious notation for the spinor fields representing the leptons. Putting all the pieces together, the invariant amplitude for the $\pi^+ \rightarrow \mu^+ + \nu_\mu$ process becomes

$$\mathcal{A}_{\pi^+ \rightarrow \mu^+ + \nu_\mu} = i f_\pi G_F \bar{u}(p) \not{k} (1 - \gamma_5) v(q). \quad (9.32)$$

Here k^μ , p^μ , q^μ are respectively the four-momenta of the pion, muon neutrino and antimuon. Also, $u(p)$ and $v(q)$ are the corresponding Dirac spinors; polarization indices are suppressed for clarity. Finally, G_F is the Fermi coupling constant,

$$G_F = \frac{g^2}{4\sqrt{2}m_W^2} \approx 1.166 \times 10^{-5} \text{ GeV}^{-2}. \quad (9.33)$$

Upon squaring the amplitude and summing over polarizations of the particles in the final state, the integrated decay rate in the rest frame of the pion is found to be

$$\Gamma_{\pi^+ \rightarrow \mu^+ + \nu_\mu} = \frac{f_\pi^2 G_F^2}{4\pi} \frac{m_\mu^2 (m_\pi^2 - m_\mu^2)^2}{m_\pi^3}. \quad (9.34)$$

I have treated the neutrino as a massless particle. The masses of the pion and antimuon are, respectively, $m_\pi \approx 139.6 \text{ MeV}$ and $m_\mu \approx 105.7 \text{ MeV}$. Finally, we need an input on the lifetime of the charged pion, $\tau \approx 2.60 \times 10^{-8} \text{ s}$. This converts to the total decay rate of $\Gamma \approx 2.53 \times 10^{-8} \text{ eV}$. At the end of the day, we get an estimate for the *pion decay constant*,

$$f_\pi \approx 91 \text{ MeV} . \quad (9.35)$$

The result (9.34) tells us more than merely a good estimate for the total decay rate of the charged pion. Namely, leptons from different families have identical weak interactions. Upon replacing m_μ with the electron mass m_e , (9.34) therefore also gives us a decay rate for the process $\pi^+ \rightarrow e^+ + \nu_e$. This is more conveniently expressed in terms of the branching ratio,

$$R_{\pi^+ \rightarrow e^+ + \nu_e} \approx \frac{\Gamma_{\pi^+ \rightarrow e^+ + \nu_e}}{\Gamma_{\pi^+ \rightarrow \mu^+ + \nu_\mu}} = \frac{m_e^2 (m_\pi^2 - m_e^2)^2}{m_\mu^2 (m_\pi^2 - m_\mu^2)^2} \approx 1.28 \times 10^{-4} . \quad (9.36)$$

This is very close to the experimental value, which is about 1.23×10^{-4} [1]. The suppression of the electron decay channel compared to the muon one is purely kinematical. By angular momentum conservation, one of the leptons in the final state must be left-handed and one right-handed. Yet, the W -boson only couples to left-handed fermion fields. The combination of these two effects requires a helicity flip and is responsible for the proportionality of (9.34) to the lepton mass squared.

9.1.4 Effects of the Chiral Anomaly

So far I have tacitly assumed, following Sect. 8.2, that the effective action of ChPT is gauge-invariant in presence of the background fields. This allowed us to construct strictly gauge-invariant Lagrangians at LO (9.14) and NLO (9.19) of the derivative expansion of ChPT. Are there any contributions to the ChPT Lagrangian that are merely quasi-invariant? A detailed derivation of such contributions and their coupling to background gauge fields would require a differential-geometric approach akin to that of Sect. 8.1. The problem of finding all such *Wess–Zumino* (WZ), or *Wess–Zumino–Witten, terms* was studied thoroughly in the 1980s and 1990s. For a discussion close in spirit to this book, I refer the reader to [12, 13]. A pedagogical account of the method including explicit expressions for quasi-invariant Lagrangians for a broad class of coset spaces can be found in [14].

Here I will resort to a trick, which gives the right answer in case of $n_f = 2$ quark flavors. Suppose we were able to construct a current J^μ , conserved off-shell, that is without imposing the *equation of motion* (EoM) for the NG fields in our EFT. If in addition the EFT includes an Abelian gauge field A_μ , then the operator $A_\mu J^\mu$ is quasi-invariant and can be added to the Lagrangian density. It remains to guess what J^μ and A_μ might be within ChPT. The current is the tricky bit. For the moment, I will simply write it down, a partial a posteriori justification will be offered below;

$$J_{\text{GW}}^\mu = \lambda \varepsilon^{\mu\nu\alpha\beta} \text{tr} \left\{ (\mathcal{U} D_\nu \mathcal{U}^\dagger) (\mathcal{U} D_\alpha \mathcal{U}^\dagger) (\mathcal{U} D_\beta \mathcal{U}^\dagger) - \frac{3i}{2} [(D_\nu \mathcal{U} \mathcal{U}^\dagger) F_{\alpha\beta}^L - (D_\nu \mathcal{U}^\dagger \mathcal{U}) F_{\alpha\beta}^R] \right\}. \quad (9.37)$$

This is the *Goldstone–Wilczek* (GW) *current*; the overall factor λ is in principle arbitrary. In the absence of the background fields, the GW current would be manifestly conserved thanks to the antisymmetry of the LC tensor and cyclicity of trace. With the background in place, it is a nontrivial but straightforward exercise to show that³

$$\partial_\mu J_{\text{GW}}^\mu = \frac{3\lambda}{4} \varepsilon^{\mu\nu\alpha\beta} \text{tr} (-F_{\mu\nu}^L F_{\alpha\beta}^L + F_{\mu\nu}^R F_{\alpha\beta}^R). \quad (9.38)$$

The current is not conserved as promised, except for backgrounds that are purely vector-like, $A_\mu^L = A_\mu^R$. Luckily, this is not a problem. In fact, the nonconservation of the GW current turns out to be exactly what is needed to implement correctly the microscopic physics of QCD within ChPT.

To that end, recall that the flavor symmetry of QCD has a single U(1) factor, namely the baryon number U(1)_B. This can also be coupled to a background gauge field, A_μ^B , even if such a field may not have an obvious experimental realization. The presence of a coupling $A_\mu^B J_{\text{GW}}^\mu$ in the ChPT action implies that it is possible to create baryon number solely out of meson fields. This intriguing possibility was first proposed by Skyrme in the 1960s. In presence of the chiral background fields $A_\mu^{L,R}$, baryon number is not conserved due to the *chiral anomaly*. An explicit calculation (see Sect. 22.3 of [15]) shows that this anomaly is reproduced at the ChPT level by (9.38) if we set $\lambda = -1/(24\pi^2)$. This brings us to the final expression for the WZ term in the ChPT Lagrangian for $n_f = 2$ quark flavors,

$$\mathcal{L}_{\text{WZ}}^{(4)} = -\frac{1}{24\pi^2} \varepsilon^{\mu\nu\alpha\beta} A_\mu^B \text{tr} \left\{ (\mathcal{U} D_\nu \mathcal{U}^\dagger) (\mathcal{U} D_\alpha \mathcal{U}^\dagger) (\mathcal{U} D_\beta \mathcal{U}^\dagger) - \frac{3i}{2} [(D_\nu \mathcal{U} \mathcal{U}^\dagger) F_{\alpha\beta}^L - (D_\nu \mathcal{U}^\dagger \mathcal{U}) F_{\alpha\beta}^R] \right\}. \quad (9.39)$$

The superscript ⁽⁴⁾ indicates that the WZ term enters at the NLO of the derivative expansion of ChPT. Note also that it does not come with an arbitrary coupling. The normalization of the WZ term is fixed by anomaly matching and does not receive any radiative corrections.

The WZ term (9.39) is manifestly invariant under background gauge transformations from the SU(2)_L × SU(2)_R chiral group. On the contrary, under a background U(1)_B transformation, $A_\mu^B \rightarrow A_\mu^B + \partial_\mu \epsilon^B$, the corresponding WZ action changes by

³ A reader willing to check this should note that the GW current is manifestly gauge-invariant. It is thus sufficient, and advantageous, to compute its gauge-covariant derivative. This maintains gauge invariance at every step and simplifies the calculation thanks to the Bianchi identity for $F_{\mu\nu}^{L,R}$.

$$\delta S_{\text{WZ}} = - \int d^4x \epsilon^B(x) \partial_\mu J_{\text{GW}}^\mu(x) , \quad (9.40)$$

with the divergence of J_{GW}^μ given by (9.38). Watch the interplay of two symmetries: the variation of the action under $U(1)_B$ is proportional to the background fields for $SU(2)_L \times SU(2)_R$. What we have here is an example of a *mixed 't Hooft anomaly*. This should be contrasted to the naive $U(1)_A$ axial symmetry of QCD. The divergence of the axial current receives a contribution from the dynamical gluon fields, giving an example of an *Adler–Bell–Jackiw anomaly*. This kind of anomaly fundamentally invalidates a would-be classical symmetry of a quantum system. On the other hand, a symmetry exhibiting a 't Hooft anomaly still implies exact relations (Ward identities) for the generating functional of the theory. This makes 't Hooft anomalies a powerful tool for constraining low-energy EFTs, as I have illustrated here.

The construction of the WZ term (9.39) is not a mere academic exercise, as one might suspect from the presence of the “baryon number gauge field.” The term has measurable consequences for the electromagnetic interactions of pions. To see why, recall (9.29), which shows that electric charge does not belong to the chiral Lie algebra $\mathfrak{su}(2)_L \times \mathfrak{su}(2)_R$ due to not being traceless. Interactions of pions with an external electromagnetic field (with all other background fields switched off) can then be generated by setting

$$A_\mu^L = A_\mu^R = \frac{e}{2} \tau_3 A_\mu^Q , \quad A_\mu^B = \frac{e}{2} A_\mu^Q . \quad (9.41)$$

Here A_μ^Q is the electromagnetic gauge potential and e the electromagnetic coupling. The effects of interaction with the electromagnetic field via the WZ term are most striking in case of the neutral pion.

Example 9.5

Let us keep only the electromagnetic background field and the neutral pion π^0 . The charged pions are discarded by using the replacement $\mathcal{U} \rightarrow \exp(i\pi^0 \tau_3 / f_\pi)$. Upon simple integration by parts, the whole WZ term (9.39) then boils down to

$$\mathcal{L}_{\text{WZ}}^{(4)} \rightarrow - \frac{e^2}{32\pi^2 f_\pi} \pi^0 \varepsilon^{\mu\nu\alpha\beta} F_{\mu\nu}^Q F_{\alpha\beta}^Q . \quad (9.42)$$

This operator governs the electromagnetic decay of π^0 into a pair of photons. Denoting the four-momentum of the pion as k^μ and those of the photons as p^μ, q^μ , the invariant amplitude for the decay turns out to be

$$\mathcal{A}_{\pi^0 \rightarrow \gamma + \gamma} = - \frac{e^2}{4\pi^2 f_\pi} \varepsilon^{\mu\nu\alpha\beta} p_\mu \epsilon_\nu^*(p) q_\alpha \epsilon_\beta^*(q) . \quad (9.43)$$

Here $\epsilon_\mu^*(p)$ and $\epsilon_\mu^*(q)$ are the polarization vectors of the photons in the final state. It remains to take the square and sum over polarizations of the photons. The final result for the decay rate in the rest frame of the pion is

$$\Gamma_{\pi^0 \rightarrow \gamma\gamma} = \frac{\alpha^2 m_\pi^3}{64\pi^3 f_\pi^2}, \quad (9.44)$$

where $\alpha \equiv e^2/(4\pi)$ is the fine structure constant. Using the numerical input $m_\pi \approx 135.0 \text{ MeV}$ and $\alpha \approx 7.297 \times 10^{-3}$ along with the value for f_π found in Example 9.4, our final result is $\Gamma_{\pi^0 \rightarrow \gamma\gamma} \approx 8.0 \text{ eV}$. This is less than 3% off the experimental value of 7.81 eV [1].

The two-flavor WZ term (9.39) has remarkable physical consequences, yet vanishes by construction in the absence of external fields. A new twist in the story comes for $n_f = 3$ quark flavors. Here another WZ term appears, which remains nonzero even in the absence of background fields. This governs scattering processes with an odd number of mesons, such as $K^+ + K^- \rightarrow \pi^+ + \pi^- + \pi^0$, which would otherwise be forbidden in ChPT. The mathematical origin of this WZ term parallels that of the quasi-invariant Lagrangians with one time derivative, analyzed in Sect. 8.1. The Lagrangian density of the WZ term can be mapped to a 4-form ω_{WZ} such that the 5-form $d\omega_{\text{WZ}}$ is chirally invariant and closed but not exact. Such 5-forms are classified by the fifth de Rham cohomology group of the coset space. The coset space $[\text{SU}(3)_L \times \text{SU}(3)_R]/\text{SU}(3)_V$ has a unique generator of degree-5 cohomology,

$$d\omega_{\text{WZ}} \propto \text{tr}[(\mathcal{U}^{-1}d\mathcal{U}) \wedge (\mathcal{U}^{-1}d\mathcal{U}) \wedge (\mathcal{U}^{-1}d\mathcal{U}) \wedge (\mathcal{U}^{-1}d\mathcal{U}) \wedge (\mathcal{U}^{-1}d\mathcal{U})]. \quad (9.45)$$

The overall normalization is again fixed by matching to the flavor anomalies of QCD. More details about the geometric nature of this WZ term and its coupling to background gauge fields can be found in the original work of Witten [16].

The two-flavor coset space $[\text{SU}(2)_L \times \text{SU}(2)_R]/\text{SU}(2)_V$, being three-dimensional, obviously has vanishing fifth cohomology group. However, it does have a nontrivial third de Rham cohomology with a single generator,

$$\omega_{\text{GW}} \propto \text{tr}[(\mathcal{U}^{-1}d\mathcal{U}) \wedge (\mathcal{U}^{-1}d\mathcal{U}) \wedge (\mathcal{U}^{-1}d\mathcal{U})]. \quad (9.46)$$

When pulled back to the four-dimensional Minkowski spacetime, ω_{GW} is just the Hodge dual of the GW current (9.37) in absence of external fields. This explains why such an identically conserved current exists in the first place. Moreover, the integral of J_{GW}^0 over \mathbb{R}^3 defines a conserved charge that is a topological invariant. Upon a suitable normalization, it coincides with the Brouwer degree of the pion fields viewed as a map $S^3 \rightarrow S^3$; cf. Example A.25. A *skyrmion* is a configuration of pion fields for which the topological charge is nonvanishing. Thanks to the coupling to A_μ^B in (9.39), the topological charge has the interpretation as baryon number. This provides a mathematical foundation for the Skyrme model of baryons.

9.2 Spin Waves in Ferro- and Antiferromagnets

I have already used ferromagnets repeatedly to illustrate various features of SSB, including the peculiarities of the spectrum of NG bosons in nonrelativistic systems. In order to make the present section self-contained, I will however start with a concise summary of the basic facts.

Ferro- and antiferromagnets are phases of matter that exhibit spin order. Although such order may also be induced in relativistic matter, I will have implicitly in mind its realization in ordinary crystalline solids. The advantage of this restriction is that in the nonrelativistic limit, spin can be treated as an internal degree of freedom. One can then base the construction of EFT for (anti)ferromagnets on spontaneous breakdown of the internal $G \simeq \text{SU}(2)$ spin symmetry. In both types of systems, the unbroken subgroup is $H \simeq \text{U}(1)$, corresponding to spin rotations about the axis of spin alignment. The coset space is therefore $G/H \simeq \text{SU}(2)/\text{U}(1) \simeq S^2$. From the point of view of low-energy EFT, the only difference between ferro- and antiferromagnets is a nonzero VEV of spin in the former. This is directly reflected by the spectrum of NG bosons: spin waves, or *magnons*. In ferromagnets, there is a single type-B magnon, whereas antiferromagnets feature two type-A magnons.

The generators of $G \simeq \text{SU}(2)$ can be taken as $\tau_A/2$. Without loss of generality, we may choose the spin axes so that $H \simeq \text{U}(1)$ is generated by $\tau_3/2$. The coset space $\text{SU}(2)/\text{U}(1)$ is then symmetric thanks to the inner automorphism $\mathcal{R}(g) = R^{-1}gR$ with $g \in \text{SU}(2)$ and $R = i\tau_3$. This makes it possible to map the coset representative $U(\pi)$ on a unit-vector variable $\mathbf{n}(\pi) \in S^2$ via

$$\boldsymbol{\tau} \cdot \mathbf{n}(\pi) \equiv N(\pi) = U(\pi)^2 \tau_3 = U(\pi) \tau_3 U(\pi)^{-1} . \quad (9.47)$$

The matrix field $N(\pi)$ transforms linearly in the adjoint representation of G . As a consequence, the G -covariant derivative (9.8) boils down to

$$D\mathbf{n}(\pi) = d\mathbf{n}(\pi) + \mathbf{A} \times \mathbf{n}(\pi) , \quad (9.48)$$

where \mathbf{A}_μ is a triplet of background gauge fields of $\text{SU}(2)$. The covariant derivative $D_\mu \mathbf{n}(\pi)$ is the basic building block for construction of EFT for (anti)ferromagnets.

In addition to the internal spin symmetry, I will assume invariance under continuous spacetime translations and continuous spatial rotations. This is of course just a crude idealization of real materials, where such ideal symmetry may be explicitly broken by a variety of perturbations. These include especially the anisotropy induced by the underlying crystal lattice, and the effects of spin–orbit coupling. I will nevertheless initially assume the ideal, unperturbed limit. An outline of some phenomenological consequences of explicit symmetry breaking is deferred to Sect. 9.2.3.

9.2.1 Power Counting and Effective Lagrangian

The general philosophy of the derivative expansion of the EFT for (anti)ferromagnets copies closely that for ChPT, detailed in Sect. 9.1.1. However, the two cases differ substantially due to the qualitatively different spectra of (anti)ferromagnetic magnons. I will start with the more nontrivial, genuinely nonrelativistic case of ferromagnets.

9.2.1.1 Ferromagnets

The energy of ferromagnetic magnons is quadratic in momentum, at least in the long-wavelength limit. In order to assign to a given Feynman diagram a well-defined degree, we therefore count momentum as order one and energy as order two. The Schrödinger-like propagator of the magnon then has overall degree -2 . In close parallel with (9.3), the degree of a Feynman diagram Γ becomes

$$\deg \Gamma = (D + 1)L - 2I + \sum_{s,t} (s + 2t)V_{s,t} . \quad (9.49)$$

Here $V_{s,t}$ denotes the number of vertices from $\mathcal{L}_{\text{eff}}^{(s,t)}$, the part of effective Lagrangian with s spatial and t temporal derivatives. As before, the number of propagators I can be eliminated via $I = L + \sum_{s,t} V_{s,t} - 1$, which leads to the final result

$$\deg \Gamma = 2 + (D - 1)L + \sum_{s,t} (s + 2t - 2)V_{s,t} . \quad (9.50)$$

The LO of the derivative expansion, $\deg \Gamma = 2$, corresponds to tree-level diagrams ($L = 0$) with all vertices satisfying $s + 2t = 2$. Thus, the LO effective Lagrangian consists of $\mathcal{L}_{\text{eff}}^{(0,1)}$ and $\mathcal{L}_{\text{eff}}^{(2,0)}$. Before we construct these, let us briefly consider higher orders of the derivative expansion. First, note that unlike in ChPT, here we have a well-defined derivative expansion even in $D = 2$ spacetime dimensions. Indeed, ferromagnetic order that can be described by a derivatively coupled low-energy EFT exists even in one-dimensional spin chains. This is special to type-B NG bosons, as I will show in Sect. 15.2.

What exactly constitutes the NLO of the derivative expansion depends on the number of dimensions. For $D = 2$ (ferromagnetic chains), the NLO corresponds to $\deg \Gamma = 3$. It collects contributions from one-loop diagrams with all vertices from the LO Lagrangian, and from tree-level diagrams with one vertex from $\mathcal{L}_{\text{eff}}^{(3,0)}$, if any such operators exist. (They may be forbidden by parity.) In $D = 4$ dimensions (bulk ferromagnets), on the other hand, there are two options. In case $\mathcal{L}_{\text{eff}}^{(3,0)}$ is allowed, then tree-level diagrams with one such vertex constitute the sole contribution with $\deg \Gamma = 3$. Otherwise, the NLO corresponds to $\deg \Gamma = 4$, and consists of tree-level diagrams with one vertex from $\mathcal{L}_{\text{eff}}^{(4,0)}$, $\mathcal{L}_{\text{eff}}^{(2,1)}$, or $\mathcal{L}_{\text{eff}}^{(0,2)}$. Up to and including NLO, there are no quantum corrections; loops only start contributing at $\deg \Gamma = 5$. Perhaps the most interesting is the case of $D = 3$ (thin ferromagnetic films or layers). Barring

the possible existence of $\mathcal{L}_{\text{eff}}^{(3,0)}$, the NLO here is $\deg \Gamma = 4$. It includes both one-loop diagrams with all vertices from the LO Lagrangian and tree-level diagrams with one vertex from the NLO Lagrangian.

Clearly, the setup of the derivative expansion depends very sensitively on the specific choice of material (which affects discrete symmetries such as parity) and sample (which controls the dimension D). I will therefore limit the discussion to the effective Lagrangian at LO. A detailed classification of possible operators up to order four in derivatives, including the effects of the discrete crystal and time-reversal symmetries, can be found in [17].

The $\mathcal{L}_{\text{eff}}^{(2,0)}$ part of the LO Lagrangian is trivial. According to the general analysis in Sect. 8.2, we expect it to be of the type $\kappa_{ab} \delta^{rs} \Omega_r^a \Omega_s^b$, where the coupling κ_{ab} must be invariant under the adjoint action of $H \simeq \text{U}(1)$. The broken part of the gauged MC form (9.7) now takes the specific form

$$\Omega_{\perp}(\pi, A) = -\frac{i}{2} U(\pi)^{-1} D N(\pi) U(\pi) \tau_3 = \frac{i}{2} \tau_3 U(\pi)^{-1} D N(\pi) U(\pi) . \quad (9.51)$$

The H -invariant part of the symmetric tensor product of Ω_{\perp} with itself is projected out by taking the trace. This leads immediately to the Lagrangian

$$\mathcal{L}_{\text{eff}}^{(2,0)} = -\frac{\varrho_s}{4} \text{tr}[D N(\pi) \cdot D N(\pi)] = -\frac{\varrho_s}{2} \delta^{rs} D_r \mathbf{n}(\pi) \cdot D_s \mathbf{n}(\pi) . \quad (9.52)$$

The parameter ϱ_s is called *spin stiffness* and controls the gradient energy arising from “bending” the uniform ground state magnetization.

In $d = 2$ spatial dimensions, it is also possible to construct an invariant operator using antisymmetric tensor product, $\lambda_{ab} \varepsilon^{rs} \Omega_r^a \Omega_s^b$. This leads to the operator $\varepsilon^{rs} \mathbf{n} \cdot (D_r \mathbf{n} \times D_s \mathbf{n})$. In the absence of background fields, the latter is a pure surface term; its integral over \mathbb{R}^2 gives, up to normalization, the Brouwer degree of the map $\mathbf{n} : \mathbb{R}^2 \rightarrow S^2$. That is however no longer the case when the EFT is coupled to background gauge fields. A quick calculation shows that up to surface terms, $\varepsilon^{rs} \mathbf{n} \cdot (D_r \mathbf{n} \times D_s \mathbf{n}) \simeq \varepsilon^{rs} \mathbf{n} \cdot \mathbf{F}_{rs}$, where $\mathbf{F}_{\mu\nu} = \partial_{\mu} \mathbf{A}_{\nu} - \partial_{\nu} \mathbf{A}_{\mu} + \mathbf{A}_{\mu} \times \mathbf{A}_{\nu}$ is the field strength of the background. When added to the Lagrangian, this operator will modify the EoM for spin waves; cf. the discussion of EoM in Sect. 8.3. In the following, I will nevertheless disregard this contribution to the EFT. First, it only exists in $d = 2$ dimensions and moreover violates time reversal, under which $\mathbf{n}(\mathbf{x}, t) \rightarrow -\mathbf{n}(\mathbf{x}, -t)$. Second, it requires a specific, nontrivial background to be nonzero, and thus does not affect the propagation of free magnons.

Let us now focus on the $\mathcal{L}_{\text{eff}}^{(0,1)}$ part of the LO Lagrangian. According to the general discussion in Sect. 8.2, this reads

$$\mathcal{L}_{\text{eff}}^{(0,1)} = -M \omega_a^3(\pi) \pi^a + M v_A^3(\pi) A_0^A , \quad (9.53)$$

where M is the density of spin (magnetization) in the ferromagnetic ground state. The second term in (9.53) is seen to equal $MA_0 \cdot \mathbf{n}(\pi)$. To evaluate the first term, we project out the third component of the MC form by taking trace with τ_3 . Then we apply the exponential parameterization, $U(\pi) = \exp(i\pi^a \tau_a/2)$, and use (7.31),

$$\begin{aligned} -M\omega_a^3(\pi)\dot{\pi}^a &= iM \operatorname{tr}[\tau_3 U(\pi)^{-1} \partial_0 U(\pi)] = -\frac{M}{2} \dot{\pi}^a \int_0^1 d\tau \operatorname{tr}[\tau_3 U(\tau\pi)^{-1} \tau_a U(\tau\pi)] \\ &= -M\dot{\pi}^a \int_0^1 d\tau n_a(\tau\pi) \simeq M\pi^a \int_0^1 d\tau \dot{n}_a(\tau\pi), \end{aligned} \quad (9.54)$$

where \simeq indicates equality up to a total derivative. The index a runs over 1, 2, we can however formally extend π^a to a three-component vector $\boldsymbol{\pi}$ and write $\pi^a \dot{n}_a = \boldsymbol{\pi} \cdot \dot{\mathbf{n}} = (\mathbf{n} \times \boldsymbol{\pi}) \cdot (\mathbf{n} \times \dot{\mathbf{n}})$. Using the definition (9.47) to take the derivative $\partial_\tau N(\tau\pi)$, we find that $\mathbf{n}(\tau\pi) \times \boldsymbol{\pi} = \partial_\tau \mathbf{n}(\tau\pi)$. Putting all the pieces together, we then arrive at the final expression for the LO effective Lagrangian for ferromagnets,

$$\begin{aligned} \mathcal{L}_{\text{eff}}^{\text{LO}} &= M \int_0^1 d\tau \partial_\tau \mathbf{n}(\tau\pi) \cdot [\mathbf{n}(\tau\pi) \times \dot{\mathbf{n}}(\tau\pi)] + MA_0 \cdot \mathbf{n}(\pi) \\ &\quad - \frac{\mathcal{Q}_s}{2} \delta^{rs} D_r \mathbf{n}(\pi) \cdot D_s \mathbf{n}(\pi). \end{aligned} \quad (9.55)$$

The second term in (9.55) is the usual Zeeman coupling of spin to an external magnetic field, whereas the third term represents the gradient energy of the spin configuration. The first term, however, deserves a comment, since the presence of integration over the parameter τ makes it look nonlocal. To get rid of the integral, let us take a step back. Consider a one-parameter family of fields, $\mathbf{n}(\tau, \pi)$, $\tau \in [0, 1]$, such that $\mathbf{n}(0, \pi) = (0, 0, 1) \equiv \mathbf{n}_0$ (ground state) and $\mathbf{n}(1, \pi) = \mathbf{n}(\pi)$. The explicit choice of interpolation used in (9.55) corresponds to $\mathbf{n}(\tau, \pi) = \mathbf{n}(\tau\pi)$. It is easy to check that upon a smooth deformation of the field, $\delta \mathbf{n}(\tau, \pi)$, the variation of the action only depends on $\delta \mathbf{n}(1, \pi) = \delta \mathbf{n}(\pi)$. The concrete choice of interpolation between the $\tau = 0$ and $\tau = 1$ limits therefore does not matter. Assuming for simplicity that $n^3(\mathbf{x}, t)$ as a function on the spacetime is non-negative everywhere, we can change the interpolation to

$$\mathbf{n}(\tau, \pi) = (\tau n^1(\pi), \tau n^2(\pi), \sqrt{1 - \tau^2[(n^1(\pi))^2 + (n^2(\pi))^2]}), \quad (9.56)$$

which makes it possible to carry out the integral over τ ,

$$\mathcal{L}_{\text{eff}}^{\text{LO}} = -M \frac{\varepsilon_{ab} n^a(\pi) \dot{n}^b(\pi)}{1 + n^3(\pi)} + MA_0 \cdot \mathbf{n}(\pi) - \frac{\mathcal{Q}_s}{2} \delta^{rs} D_r \mathbf{n}(\pi) \cdot D_s \mathbf{n}(\pi). \quad (9.57)$$

This is as far as we can get. The Lagrangian is local and manifestly invariant under $H \simeq \text{U}(1)$. Moreover, it only depends on the NG fields π^a through $\mathbf{n}(\pi)$, and thus does not rely on the exponential parameterization of $U(\pi)$, originally used to derive (9.55).

9.2.1.2 Antiferromagnets

As far as the construction of the effective Lagrangian is concerned, antiferromagnets are much simpler than ferromagnets. The spectrum consists of two type-A NG bosons whose energy is, in the long-wavelength limit, linear in momentum. For the sake of power counting, we therefore have to treat spatial and temporal derivatives on equal footing. The resulting expression for the degree of a given Feynman diagram is a trivial generalization of (9.4) we found in ChPT,

$$\deg \Gamma = 2 + (D - 2)L + \sum_{s,t} (s + t - 2)V_{s,t} . \quad (9.58)$$

The only difference to ChPT is that spatial and temporal derivatives may enter the effective Lagrangian independently.

Owing to the same symmetry-breaking pattern, the building blocks for constructing the EFT are the same for ferro- and antiferromagnets. The main difference is that the part with a single time derivative, $\mathcal{L}_{\text{eff}}^{(0,1)}$, is missing in the latter; antiferromagnets have zero net magnetization. The LO Lagrangian then consists of $\mathcal{L}_{\text{eff}}^{(2,0)}$ and $\mathcal{L}_{\text{eff}}^{(0,2)}$, and we can write it down at once,

$$\mathcal{L}_{\text{eff}}^{\text{LO}} = \frac{\varrho_s}{2v^2} [D_0 \mathbf{n}(\pi) \cdot D_0 \mathbf{n}(\pi) - v^2 \delta^{rs} D_r \mathbf{n}(\pi) \cdot D_s \mathbf{n}(\pi)] . \quad (9.59)$$

There are two independent parameters which are easy to relate to physical observables. The spin stiffness ϱ_s measures the gradient energy of the order parameter, whereas v turns out to be the phase velocity of antiferromagnetic magnons.

Similarly to ferromagnets, the organization of the derivative expansion beyond LO depends sensitively on D and the presence of discrete symmetries such as parity or time reversal. I will therefore stop the discussion of power counting here, and turn to the consequences of the EFT at LO.

9.2.2 Equation of Motion and Magnon Spectrum

We already know the number and type of magnons in both ferro- and antiferromagnets. However, the EFT tells us more, in particular what the corresponding fluctuations of the order parameter look like. To that end, I will drop the background gauge fields and derive the EoM corresponding to the LO effective Lagrangian.

Let us start with ferromagnets. As hinted above, taking a variation of (9.55) gives a surface term in τ , which allows one to write the variation of the action solely in terms of $\delta \mathbf{n}(\pi)$,

$$\delta S_{\text{eff}}^{\text{LO}} = \int d^D x \, \delta \mathbf{n} \cdot (M \mathbf{n} \times \dot{\mathbf{n}} + \varrho_s \nabla^2 \mathbf{n}) . \quad (9.60)$$

The variation $\delta \mathbf{n}$ is not arbitrary but rather should keep \mathbf{n} on the coset space, that is the unit sphere S^2 . In other words, $\delta \mathbf{n}$ should be a tangent vector to the sphere. The

vanishing of $\delta S_{\text{eff}}^{\text{LO}}$ therefore requires that

$$M\mathbf{n} \times \dot{\mathbf{n}} + \varrho_s \nabla^2 \mathbf{n} = \lambda \mathbf{n} ; \quad (9.61)$$

λ can be interpreted as a Lagrange multiplier for the constraint $\mathbf{n} \cdot \mathbf{n} = 1$. We can get rid of it by taking a cross product with \mathbf{n} , which gives the *Landau–Lifshitz equation*,

$$\dot{\mathbf{n}} = \frac{\varrho_s}{M} \mathbf{n} \times \nabla^2 \mathbf{n} . \quad (9.62)$$

Previously, in Sect. 4.3.1, I derived this equation using the Hamiltonian (symplectic) formulation of field theory from the postulated Poisson bracket for the spin variable $\mathbf{n}(\mathbf{x})$; cf. (4.29). The Lagrangian and Hamiltonian descriptions of ferromagnets are of course equivalent. For instance, the symplectic 2-form (4.33) can be recovered by noting that according to (9.53), the symplectic potential is $-M\omega^3$. The exterior derivative thereof is easily evaluated using the MC equation (8.11). This shows that the fundamental Poisson bracket for $\mathbf{n}(\mathbf{x})$ is already automatically built in the low-energy Lagrangian EFT for ferromagnets.

To solve the Landau–Lifshitz equation (9.62) is a hard problem due to its non-linearity. What one can do easily is to linearize the equation in small fluctuations around the ground state. Inserting $\mathbf{n} = \mathbf{n}_0 + \delta\mathbf{n}$ and keeping only terms linear in $\delta\mathbf{n}$, we get

$$\delta\dot{\mathbf{n}} = \frac{\varrho_s}{M} \mathbf{n}_0 \times \nabla^2 \delta\mathbf{n} . \quad (9.63)$$

We can now look for plane-wave solutions by using the ansatz

$$\delta\mathbf{n}(\mathbf{x}, t) = A e^{-iEt} e^{i\mathbf{p} \cdot \mathbf{x}} , \quad (9.64)$$

where A is a complex amplitude orthogonal to $\mathbf{n}_0 = (0, 0, 1)$. Inserting the ansatz in (9.63) shows that the energy and momentum satisfy the dispersion relation

$$E(\mathbf{p}) = \frac{\varrho_s}{M} \mathbf{p}^2 , \quad (9.65)$$

typical for type-B NG bosons. The amplitude must satisfy the constraint $i\mathbf{A} = \mathbf{n}_0 \times \mathbf{A}$. This is solved by any $\mathbf{A} \propto (1, -i, 0)$. Ferromagnetic spin waves are circularly polarized in the plane transverse to the direction of the ground state magnetization, \mathbf{n}_0 , regardless of the direction of momentum \mathbf{p} . This can be understood as Larmor precession of individual spins around the effective magnetic field generated by the spin-polarized background.

Antiferromagnets can be treated in the same way, without the complications due to the single-time-derivative operator in the Lagrangian. The EoM obtained from (9.59) can be written as

$$\frac{\varrho_s}{v^2} (\partial_0^2 \mathbf{n} - v^2 \nabla^2 \mathbf{n}) = \lambda \mathbf{n} , \quad (9.66)$$

where λ is a Lagrange multiplier. Upon linearization around the ground state, \mathbf{n}_0 , we find plane-wave solutions of the same general form as in (9.64). However, the dispersion relation is now $E(\mathbf{p}) = v |\mathbf{p}|$, typical for type-A NG bosons. The complex

amplitude \mathbf{A} must be orthogonal to \mathbf{n}_0 . We conclude that antiferromagnetic spin waves are also polarized in the plane transverse to the direction of \mathbf{n}_0 , regardless of the direction of momentum. However, the polarization can be both linear and circular, or in general elliptic. There are therefore two different, independent types of antiferromagnetic magnons, which can be chosen to be linearly polarized.

9.2.3 Effects of Symmetry-Breaking Perturbations

So far, we have assumed exact symmetry under $SU(2)$ spin rotations, continuous spacetime translations and spatial rotations. Let us now briefly consider the effects of some phenomenologically important perturbations. These can be classified into two broad groups: perturbations controlled by external fields, and those intrinsic to the given material.

The most natural tunable perturbation that the spins in (anti)ferromagnets can be exposed to is an external magnetic field, \mathbf{B} . Insofar as its effect on orbital degrees of freedom can be neglected, the magnetic field couples directly to the conserved charge of $G \simeq SU(2)$: the total spin. It can therefore be treated as a vector-valued chemical potential. In the low-energy EFT, this is implemented by setting $\mathbf{A}_\mu(x) = \delta_{\mu 0} \mathbf{B}(x)$; the magnetic moment of the spins is absorbed into the definition of \mathbf{B} . I will now show that treating \mathbf{B} as a background gauge field allows us to make some exact statements about magnon spectrum. Importantly, we do not have to introduce any new arbitrary parameters into the Lagrangian.

Example 9.6

According to (9.55), the effect of an external magnetic field on ferromagnets is taken into account by adding the Zeeman term, $M\mathbf{B} \cdot \mathbf{n}(\pi)$, to the Lagrangian. As long as the magnetic field is uniform (which I will from now on assume), the ground state \mathbf{n}_0 will remain uniform as well. However, its orientation is no longer arbitrary, but rather has to be aligned parallel to \mathbf{B} . The magnetic field selects a unique stable equilibrium state. The effect on the magnon spectrum is also easy to work out. The left-hand side of the Landau–Lifshitz equation (9.62) has to be modified by replacing $\dot{\mathbf{n}} \rightarrow D_0 \mathbf{n} = \dot{\mathbf{n}} + \mathbf{B} \times \mathbf{n}$. Upon linearization, the plane-wave magnon solutions are still found to be circularly polarized in the plane transverse to $\mathbf{n}_0 \parallel \mathbf{B}$. The only effect of the magnetic field is a constant shift of the dispersion relation (9.65),

$$E(\mathbf{p}) = |\mathbf{B}| + \frac{\rho_s}{M} \mathbf{p}^2. \quad (9.67)$$

The reason why the response of ferromagnets to a uniform magnetic field is so simple is that the conserved charge that \mathbf{B} couples to remains unbroken. Energy levels can therefore be labeled by the projection of spin into the direction of \mathbf{B} . The excitation energy of a state with spin S (relative to the ground state) will be shifted by the magnetic field by $-S|\mathbf{B}|$. The ferromagnetic ground state is maximally polarized,

and a single magnon carries a unit of spin less than the ground state. This explains why the magnon receives a gap equal to $|\mathbf{B}|$. This is an exact result valid to all orders of the derivative expansion; the ferromagnetic magnon is an example of a massive NG boson in the sense of Sect. 6.4.3.

Example 9.7

The effect of magnetic fields on antiferromagnets is somewhat less trivial. According to (9.59), their LO Lagrangian (in absence of other perturbations than \mathbf{B}) becomes

$$\mathcal{L}_{\text{eff}}^{\text{LO}} = \frac{\mathcal{Q}_s}{2v^2} \{ [\partial_0 \mathbf{n}(\pi) + \mathbf{B} \times \mathbf{n}(\pi)]^2 - v^2 \delta^{rs} \partial_r \mathbf{n}(\pi) \cdot \partial_s \mathbf{n}(\pi) \} . \quad (9.68)$$

The corresponding Hamiltonian density is

$$\mathcal{H}_{\text{eff}}^{\text{LO}} = \frac{\mathcal{Q}_s}{2v^2} \{ [\partial_0 \mathbf{n}(\pi)]^2 - [\mathbf{B} \times \mathbf{n}(\pi)]^2 + v^2 \delta^{rs} \partial_r \mathbf{n}(\pi) \cdot \partial_s \mathbf{n}(\pi) \} . \quad (9.69)$$

The response of the ground state is quite different from ferromagnets: the energy is minimized by any $\mathbf{n}_0 \perp \mathbf{B}$. Hence the U(1) group of spin rotations around the direction of \mathbf{B} is spontaneously broken. We expect the spectrum to contain one true NG boson, whereas the other magnon should receive a gap from the magnetic field.

To see this explicitly, let us choose $\mathbf{B} = (0, 0, |\mathbf{B}|)$ and $\mathbf{n}_0 = (1, 0, 0)$. We can use n^2 and n^3 as two independent fluctuations of the order parameter, and parameterize the latter as

$$\mathbf{n}(x) = (\sqrt{1 - [(n^2(x))^2 + (n^3(x))^2]}, n^2(x), n^3(x)) . \quad (9.70)$$

To the second order in the fluctuations and up to a total time derivative, the Lagrangian then reads

$$\mathcal{L}_{\text{eff}}^{\text{LO}} \simeq \frac{\mathcal{Q}_s}{2v^2} \left\{ \sum_{i=2,3} [(\partial_0 n^i)^2 - v^2 \nabla n^i \cdot \nabla n^i] + \mathbf{B}^2 [1 - (n^3)^2] \right\} + \dots . \quad (9.71)$$

This makes it clear that $n^2(x)$ remains gapless, as expected. On the other hand, the $n^3(x)$ mode receives a gap, its full dispersion relation being $E(\mathbf{p}) = \sqrt{v^2 \mathbf{p}^2 + \mathbf{B}^2}$. The conclusion that $E(\mathbf{0}) = |\mathbf{B}|$ is exact. The spectrum of an ideal antiferromagnet in a uniform magnetic field contains one true NG boson and one massive NG boson.

As opposed to the effect of external fields, perturbations induced by the underlying crystal lattice are intrinsic to the given material and therefore cannot be “switched off.” A prominent position among these is occupied by anisotropy in either spatial or spin structure of the microscopic interactions. I will content myself with the simplest illustrative example of such a crystal anisotropy, whereby one spin axis is distinguished from the other two,

$$\mathcal{L}_{\text{pert}} = \epsilon(n^3)^2 . \quad (9.72)$$

For $\epsilon > 0$, the perturbation favors spin alignment along the third axis; this kind of anisotropy is called *easy-axis*. In the opposite case of $\epsilon < 0$, spin alignment along the first or second axis is preferred. This is an *easy-plane* anisotropy. The effects of easy-axis and easy-plane anisotropy on ferro- and antiferromagnets are very different. It is therefore best to discuss them separately. In each individual case, I will proceed by first identifying the ground state and then expanding the Lagrangian to second order in fluctuations.

Example 9.8

Let us start with ferromagnets. In the absence of background gauge fields but presence of the anisotropy, the Lagrangian (9.57) becomes

$$\mathcal{L}_{\text{eff}}^{\text{LO}} = -M \frac{\varepsilon_{ab} n^a \dot{n}^b}{1 + n^3} - \frac{\varrho_s}{2} \delta^{rs} \partial_r \mathbf{n} \cdot \partial_s \mathbf{n} + \epsilon(n^3)^2 . \quad (9.73)$$

In easy-axis ferromagnets, the ground state is unique up to overall sign, $\mathbf{n}_0 = (0, 0, 1)$. The two independent fluctuations can be taken as n^1 and n^2 . Both the anisotropy and the ground state preserve the U(1) group of spin rotations around the third axis. We can thus look for normal modes as eigenstates of this symmetry. This motivates the introduction of a complex field, $\psi \equiv (n^1 + in^2)/\sqrt{2}$. Dropping the energy density of the ground state and expanding the Lagrangian to second order in ψ , we get

$$\mathcal{L}_{\text{eff}}^{\text{LO}} \simeq iM\psi^\dagger \partial_0 \psi - \varrho_s \nabla \psi^\dagger \cdot \nabla \psi - 2\epsilon \psi^\dagger \psi + \dots . \quad (9.74)$$

This leads to a Schrödinger-like equation, describing circularly polarized spin waves with dispersion relation

$$E(\mathbf{p}) = \frac{2\epsilon}{M} + \frac{\varrho_s}{M} \mathbf{p}^2 . \quad (9.75)$$

The anisotropy gives the magnon a gap since both generators of SU(2), spontaneously broken in the preferred ground state, are also broken explicitly. In contrast to (9.67), the gap predicted by (9.75), $E(\mathbf{0}) = 2\epsilon/M$, is only a LO result and will receive corrections at higher orders of the derivative expansion. The same is true for all the other magnon dispersion relations, derived below. The magnon has become a pseudo-NG boson.

In the easy-plane case, any uniform state with $\langle n^3 \rangle = 0$ minimizes the energy. I will choose the ground state as $\mathbf{n}_0 = (1, 0, 0)$ and the independent fluctuations as n^2 and n^3 . Upon series expansion in the latter using the parameterization (9.70), the Lagrangian becomes

$$\mathcal{L}_{\text{eff}}^{\text{LO}} \simeq Mn^3 \partial_0 n^2 - \frac{\varrho_s}{2} \sum_{i=2,3} \nabla n^i \cdot \nabla n^i - |\epsilon| (n^3)^2 + \dots . \quad (9.76)$$

We already found the spectrum of this type of Lagrangian back in Sect. 3.2.1, I can therefore just write down the final result,

$$E(\mathbf{p}) = \frac{\varrho_s}{M} |\mathbf{p}| \sqrt{\mathbf{p}^2 + \frac{2|\epsilon|}{\varrho_s}}. \quad (9.77)$$

The dispersion relation remains gapless but has become linear. This is because the $U(1)$ group of spin rotations, left intact by the anisotropy, is now spontaneously broken. The anisotropy has turned the magnon into a type-A NG boson.

Example 9.9

Next we turn to antiferromagnets, which are now for a change much easier to analyze. In the absence of background gauge fields but upon adding the anisotropy term, the Lagrangian (9.59) turns into

$$\mathcal{L}_{\text{eff}}^{\text{LO}} = \frac{\varrho_s}{2v^2} [(\partial_0 \mathbf{n})^2 - v^2 \delta^{rs} \partial_r \mathbf{n} \cdot \partial_s \mathbf{n}] + \epsilon (n^3)^2. \quad (9.78)$$

This Lagrangian is diagonal in the spin index of n^i . We can therefore read off the spectrum immediately upon identification of the ground state and its independent fluctuations. In easy-axis antiferromagnets, the ground state is $\mathbf{n}_0 = (0, 0, 1)$ up to a sign, and its fluctuations are n^1 and n^2 . These excite two gapped magnons,

$$E_{1,2}(\mathbf{p}) = \sqrt{v^2 \mathbf{p}^2 + \frac{2v^2 \epsilon}{\varrho_s}}. \quad (9.79)$$

We can choose the basis of independent spin waves freely, either as linear, circular, or generally elliptic. In the easy-plane case, on the other hand, we can choose the ground state as $\mathbf{n}_0 = (1, 0, 0)$ and its fluctuations as n^2 and n^3 . Here we find two different excitation branches with different dispersion relations, corresponding to linearly polarized spin waves,

$$E_2(\mathbf{p}) = v |\mathbf{p}|, \quad E_3(\mathbf{p}) = \sqrt{v^2 \mathbf{p}^2 + \frac{2v^2 |\epsilon|}{\varrho_s}}. \quad (9.80)$$

The reason why one of the excitations remains gapless is that there is an exact $U(1)$ symmetry, left intact by the perturbation, that is spontaneously broken.

I will conclude the discussion of perturbations in spin systems with a very peculiar example that leads to fascinating phenomenology. Until now, I have ruled out the existence of $\mathcal{L}_{\text{eff}}^{(1,0)}$ based on invariance under spatial rotations. However, one can, in fact, construct an operator with a single spatial derivative that does not break rotations, as long as two conditions are satisfied. First, parity must be broken, typically by the structure of the underlying crystal lattice. Second, spin–orbit coupling must be taken into account. This breaks the separate symmetries under spatial (orbital)

and spin rotations to a single SU(2), under which spatial coordinates \mathbf{x} and the spin vector \mathbf{n} transform simultaneously. One can then add to the Lagrangian of (anti)ferromagnets the *Dzyaloshinskii–Moriya* (DM) term,

$$\mathcal{L}_{\text{eff}}^{(1,0)} = -\frac{2\pi\varrho_s}{\lambda_{\text{DM}}} \mathbf{n} \cdot (\nabla \times \mathbf{n}) , \quad (9.81)$$

where λ_{DM} is a new parameter with the dimension of length. This length scale is fixed by the choice of concrete material, and is usually much larger than the scale of the underlying crystal lattice. For instance, in MnSi one finds $\lambda_{\text{DM}} \approx 18$ nm and in FeGe $\lambda_{\text{DM}} \approx 70$ nm [18]. This justifies treating the perturbation coupling $2\pi\varrho_s/\lambda_{\text{DM}}$ as a small parameter of degree one in the power counting. The Lagrangian $\mathcal{L}_{\text{eff}}^{(1,0)}$ can then be consistently included in the LO of the derivative expansion.

Example 9.10

The ground state induced by the DM interaction can be discussed jointly for ferro- and antiferromagnets if one restricts to time-independent spin configurations. In the absence of external gauge fields and anisotropy, the LO Hamiltonian then reduces to

$$\mathcal{H}_{\text{eff}}^{\text{LO}} = \frac{\varrho_s}{2} \delta^{rs} \partial_r \mathbf{n} \cdot \partial_s \mathbf{n} + \frac{2\pi\varrho_s}{\lambda_{\text{DM}}} \mathbf{n} \cdot (\nabla \times \mathbf{n}) . \quad (9.82)$$

The condition for minimum energy follows by “completing the square,”

$$\begin{aligned} \mathcal{H}_{\text{eff}}^{\text{LO}} &\simeq \frac{\varrho_s}{2} [(\nabla \cdot \mathbf{n})^2 + (\nabla \times \mathbf{n})^2] + \frac{2\pi\varrho_s}{\lambda_{\text{DM}}} \mathbf{n} \cdot (\nabla \times \mathbf{n}) \\ &= \frac{\varrho_s}{2} (\nabla \cdot \mathbf{n})^2 + \frac{\varrho_s}{2} \left(\nabla \times \mathbf{n} + \frac{2\pi}{\lambda_{\text{DM}}} \mathbf{n} \right)^2 - \frac{2\pi^2\varrho_s}{\lambda_{\text{DM}}^2} . \end{aligned} \quad (9.83)$$

The second term will be minimized by any spin configuration satisfying the first-order differential equation $\nabla \times \mathbf{n} = -(2\pi/\lambda_{\text{DM}})\mathbf{n}$. This guarantees without further conditions minimization of the first term, $(\nabla \cdot \mathbf{n})^2$, and thus of the whole Hamiltonian.

Let us now temporarily forget that $\mathbf{n}(\mathbf{x})$ should be a unit vector at any \mathbf{x} , and Fourier-transform it to momentum space. The complex amplitude $\mathbf{n}_{\mathbf{p}}$ of the plane wave with momentum \mathbf{p} is then subject to the condition $\mathbf{i}\mathbf{p} \times \mathbf{n}_{\mathbf{p}} = -(2\pi/\lambda_{\text{DM}})\mathbf{n}_{\mathbf{p}}$. This requires that $\mathbf{p} \perp \mathbf{n}_{\mathbf{p}}$ and $|\mathbf{p}| = 2\pi/\lambda_{\text{DM}}$ for any \mathbf{p} such that $\mathbf{n}_{\mathbf{p}}$ is nonzero. This gives the parameter λ_{DM} the interpretation as the wavelength of a spatially inhomogeneous order, induced by the DM interaction. An arbitrary linear combination of plane waves with fixed $|\mathbf{p}|$ is, however, not allowed by the constraint $\mathbf{n} \cdot \mathbf{n} = 1$. Without going into further technical details, let me spell out the final result. The ground state corresponds to a real helix that is right-handed for $\lambda_{\text{DM}} > 0$ and left-handed for $\lambda_{\text{DM}} < 0$. The direction of the axis of the helix is arbitrary and spontaneously breaks the symmetry under spatial and spin rotations. Choosing it for illustration along the z -axis, and \mathbf{n} to point along the x -axis at $z = 0$, the ground state becomes

$$\langle \mathbf{n}(\mathbf{x}) \rangle = (\cos 2\pi z/\lambda_{\text{DM}}, \sin 2\pi z/\lambda_{\text{DM}}, 0) . \quad (9.84)$$

In presence of the DM interaction, the spin system becomes a *helimagnet*.

The magnon spectrum of helimagnets is no less fascinating than their ground state. In case the underlying spin order is ferromagnetic, there is a single magnon branch (cf. Example 6.5). It has an unusual, strongly anisotropic dispersion relation. Along the axis of the helix, the energy is linear in momentum for wavelengths much longer than λ_{DM} , whereas in the two transverse directions it is quadratic [19]. One can show in addition that in the former case, the spin wave is linearly polarized, while in the latter case it is polarized circularly as usual in ferromagnets.

9.2.4 Some Topological Aspects of Ferromagnets

The list of interesting features of spin systems does not end with the multitude of textures that can be induced by various perturbations. In particular ferromagnets exhibit a number of intriguing topological properties that can be traced to the part $\mathcal{L}_{\text{eff}}^{(0,1)}$ of the effective Lagrangian. A brief sample is presented below. A reader wishing to learn more about the topology of spin systems is encouraged to consult [20].

Let us start by rewriting the single-time-derivative part of the ferromagnet Lagrangian (9.55) in a way that exposes its geometric nature. Suppose the field $\mathbf{n}(\mathbf{x}, t)$ converges to a constant for $t \rightarrow \pm\infty$; this is a usual assumption when setting up the variational principle for fields. Viewed as a function of τ and t , \mathbf{n} then maps $[0, 1] \times \mathbb{R}$ to a “disk” D on S^2 whose boundary Γ carries the physical field ($\tau = 1$). The action defined by (9.55) can be cast as

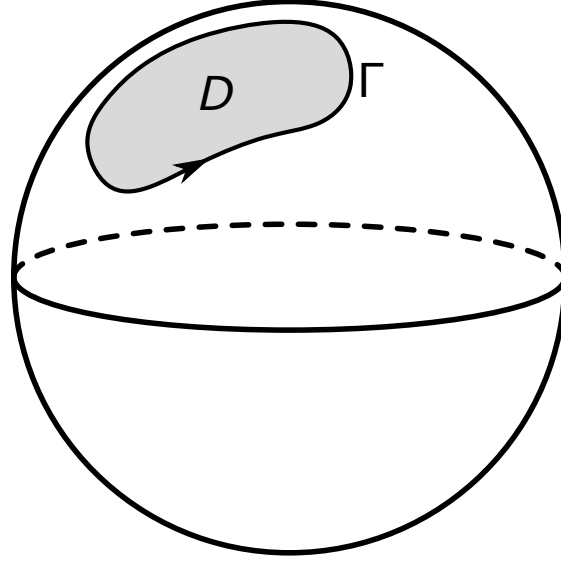
$$S_{\text{eff}}^{\text{LO}}\{\mathbf{n}, A\} = M \int d^d\mathbf{x} \int_D \Omega[\mathbf{n}](\mathbf{x}, t, \tau) + \cdots, \quad (9.85)$$

where $\Omega[\mathbf{n}] \equiv \mathbf{n} \cdot (\partial_t \mathbf{n} \times \partial_\tau \mathbf{n}) dt \wedge d\tau$; the ellipsis stands for the spacetime integral of the second and third term in (9.55). Note that $\mathbf{n} \cdot (d\mathbf{n} \times d\mathbf{n})$ is the area form on S^2 . The action is therefore determined geometrically by the area of the domain on S^2 , bounded by the curve Γ ; see Fig. 9.3 for a visualization.

But how do we know which domain? I showed already that smoothly varying the interpolation $\mathbf{n}(\tau, \pi)$ between the fixed limits at $\tau = 0$ and $\tau = 1$ does not change the action. However, there are distinct classes of maps that cannot be smoothly deformed into each other. We could, for instance, think of the interpolation as filling the domain D shown by shading in Fig. 9.3, or its complement on the sphere. There is no a priori way to distinguish the “inside” and “outside” of the curve Γ . The interpolation $\mathbf{n}(\tau, \pi)$ could even cover the whole sphere multiple times, before converging to the curve Γ in the limit $\tau \rightarrow 1$. The only way around this intrinsic ambiguity is to ensure that it is not physically observable.

To that end, consider two domains, D_1 and D_2 , swept by two different interpolations $\mathbf{n}_1(\tau, \pi)$ and $\mathbf{n}_2(\tau, \pi)$ of the same physical field, $\mathbf{n}_1(1, \pi) = \mathbf{n}_2(1, \pi) = \mathbf{n}(\pi)$. We can glue the two maps into a single one, $\tilde{\mathbf{n}}(\tau, \pi)$, $\tau \in [0, 2]$, by setting

Fig. 9.3 Geometric interpretation of the single-time-derivative term in the effective Lagrangian (9.55) for ferromagnets. The spin configuration $\mathbf{n}(\mathbf{x}, t)$ maps the time axis to a closed curve Γ on the coset space S^2 . Manifest invariance of the Lagrangian under $SU(2)$ spin rotations can be saved at the cost of extending the integration from Γ to the disk D (shaded area) whose boundary is Γ



$$\begin{aligned}\tilde{\mathbf{n}}(\tau, \pi) &\equiv \mathbf{n}_1(\tau, \pi) && \text{for } \tau \in [0, 1] , \\ \tilde{\mathbf{n}}(\tau, \pi) &\equiv \mathbf{n}_2(2 - \tau, \pi) && \text{for } \tau \in [1, 2] .\end{aligned}\tag{9.86}$$

This new map satisfies $\tilde{\mathbf{n}}(0, \pi) = \tilde{\mathbf{n}}(2, \pi) = \mathbf{n}_0$. Thanks to the compactification of the time axis to the circle S^1 , we can think of it as a map $S^2 \rightarrow S^2$ with the domain spanned by the variables τ, t . The points $\tau = 0$ and $\tau = 2$ are the poles of the domain and $\tau = 1$ is the equator. As a consequence,

$$\int_{D_1} \Omega[\mathbf{n}_1] - \int_{D_2} \Omega[\mathbf{n}_2] = 4\pi w[\tilde{\mathbf{n}}] ,\tag{9.87}$$

where $w[\tilde{\mathbf{n}}]$ is the integer-valued Brouwer degree (A.145) of $\tilde{\mathbf{n}}$. The conclusion is that the classical action of a ferromagnet suffers from a topological ambiguity, shifting it by $4\pi MV w[\tilde{\mathbf{n}}]$ where V is spatial volume. The functional integral of the EFT as a quantum theory will still be well-defined provided the action is only ambiguous up to an integer multiple of 2π . This requires that the total spin MV be quantized in half-integers. The low-energy EFT for ferromagnets, constructed solely based on the geometry of the coset space, “knows” about quantization of spin!

Now that we have established the topological nature of the $\mathcal{L}_{\text{eff}}^{(0,1)}$ Lagrangian, let us look at some of its consequences. The most immediate one is the presence of a *Berry phase*; see Sect. 1.5 of [21] for a general introduction and [22] for the application to EFTs for NG bosons. Let us denote the quantum-mechanical ground state of a ferromagnet, corresponding to the order parameter $\langle \mathbf{n} \rangle = \mathbf{n}_0$, as $|\mathbf{n}_0\rangle$. Suppose we expose the ferromagnet to a weak, uniform magnetic field \mathbf{B} . This forces it to align its magnetization with the field so that $\langle \mathbf{n} \rangle \equiv \mathbf{n}(\mathbf{B}) = \mathbf{B}/|\mathbf{B}|$, with the corresponding vacuum state $|\mathbf{n}(\mathbf{B})\rangle$.

Consider now a time-dependent magnetic field $\mathbf{B}(t)$ and arrange the ferromagnet state $|\psi(0)\rangle$ at $t = 0$ to be the vacuum state $|\mathbf{n}(\mathbf{B}(0))\rangle$. If the variation of $\mathbf{B}(t)$ with time is sufficiently slow, the evolution of the state of the system $|\psi(t)\rangle$ will be adiabatic. We can set the ground state energy E to zero to eliminate the trivial phase factor of e^{-iEt} . Even then, we cannot expect $|\psi(t)\rangle$ to equal $|\mathbf{n}(\mathbf{B}(t))\rangle$. Rather,

$$|\psi(t)\rangle = \exp[i\gamma_B(t)] |\mathbf{n}(\mathbf{B}(t))\rangle, \quad (9.88)$$

where $\gamma_B(t)$ is the Berry phase. Taking the time derivative of (9.88) and projecting the result to $|\psi(t)\rangle$ leads to $\dot{\gamma}_B(t) = i \langle \mathbf{n}(\mathbf{B}(t)) | \partial_0 | \mathbf{n}(\mathbf{B}(t)) \rangle$. The total phase accumulated over an interval $[t_1, t_2]$, $\int_{t_1}^{t_2} \dot{\gamma}_B(t) dt$, only depends on the path followed by \mathbf{B} , not on its precise time dependence. To underline the geometric nature of the Berry phase, we introduce the *Berry connection*,

$$\omega_B(\mathbf{B}) \equiv i \langle \mathbf{n}(\mathbf{B}) | d | \mathbf{n}(\mathbf{B}) \rangle, \quad (9.89)$$

which is a 1-form on the space of all quantum ground states of the ferromagnet. The gauge freedom associated with the Berry connection amounts to the arbitrary choice of phase of $|\mathbf{n}(\mathbf{B})\rangle$ for every value of \mathbf{B} . Since the state $|\mathbf{n}(\mathbf{B})\rangle$ only depends on the direction of \mathbf{B} , the Berry connection naturally induces a 1-form on the coset space S^2 of the ferromagnet. When the magnetic field varies so that it traces a closed loop in the parameter space, the total Berry phase is

$$\gamma_B(\Gamma) = \int_{\Gamma} \omega_B = \int_D d\omega_B. \quad (9.90)$$

Here Γ is the loop on S^2 traced by the order parameter in the process and D the disk bounded by it, as in Fig. 9.3. The 2-form $d\omega_B$ is called the *Berry curvature*.

To see the connection between the Berry phase and the effective Lagrangian for ferromagnets, recall the parameterization of the coset space by matrices $U(\pi)$. With a slight abuse of notation, one can write $|\mathbf{n}(\mathbf{B})\rangle = U(\pi(\mathbf{B})) |\mathbf{n}_0\rangle$. The Berry connection then reads

$$\omega_B = i \langle \mathbf{n}_0 | U(\pi)^{-1} dU(\pi) | \mathbf{n}_0 \rangle = - \langle \mathbf{n}_0 | \omega(\pi) | \mathbf{n}_0 \rangle = -MV\omega^3(\pi). \quad (9.91)$$

Comparing this to (9.53), we see that up to the factor of volume, the Berry connection is precisely the 1-form that defines the Lagrangian density $\mathcal{L}_{\text{eff}}^{(0,1)}$. On the other hand, it follows from (9.85) and the Stokes theorem that $-d\omega^3$ is the area form on S^2 . The Berry phase (9.90) therefore equals the total spin of the system times the area of D .

Mathematically, the area 2-form is the single generator of the second de Rham cohomology group of S^2 . The topological features of $\mathcal{L}_{\text{eff}}^{(0,1)}$ stem from pulling this 2-form back to the space of variables τ, t . However, interesting physics also arises from pulling it back to the whole spacetime. For simplicity, I will now restrict to $d = 2$ spatial dimensions. The resulting closed spacetime 2-form is then Hodge-dual to a current, conventionally normalized as

$$J^\mu[\mathbf{n}] = \frac{1}{8\pi} \varepsilon^{\mu\nu\lambda} \varepsilon_{ijk} n^i \partial_\nu n^j \partial_\lambda n^k . \quad (9.92)$$

This current is conserved off-shell, similarly to the GW current introduced in Sect. 9.1.4. Following the analogy, we expect the density $J^0[\mathbf{n}]$ to give rise to a topological charge, that is a topological invariant of the field $\mathbf{n}(\mathbf{x}, t)$. Indeed, $\int d^2\mathbf{x} J^0[\mathbf{n}](\mathbf{x}, t)$ is just the Brouwer degree $w[\mathbf{n}]$ of the map $\mathbf{n} : \mathbb{R}^2 \rightarrow S^2$. In $d = 2$ dimensions, there are smooth spin configurations carrying nonzero value of $w[\mathbf{n}]$, called skyrmions, or sometimes *baby skyrmions*, to distinguish them from their counterparts in QCD. See the dedicated monograph [20] for further details about this fascinating subject.

Recall now the symplectic approach to ferromagnets, outlined in Sect. 4.3.1. It is easy to check explicitly that as a functional on the phase space of a ferromagnet, $w[\mathbf{n}]$ has a vanishing Poisson bracket with any other functional. This is an example of a general feature that topological charges do not generate any flow on the phase space. Let us see what happens if we deform $w[\mathbf{n}]$ by inserting an arbitrary function of spatial coordinates. Changing for convenience the overall normalization, we define a class of functionals,

$$Q_f[\mathbf{n}] \equiv \frac{M}{2} \int d^2\mathbf{x} f(\mathbf{x}) \varepsilon^{rs} \varepsilon_{ijk} n^i(\mathbf{x}) \partial_r n^j(\mathbf{x}) \partial_s n^k(\mathbf{x}) . \quad (9.93)$$

It is a simple exercise to verify that the Poisson algebra of these functionals reproduces the algebra of functions on \mathbb{R}^2 ,

$$\{Q_f, Q_g\} = Q_{\varepsilon^{rs} \partial_r f \partial_s g} \equiv Q_{\{f, g\}} . \quad (9.94)$$

In addition, the functional Q_f generates a flow on the phase space through

$$\{\mathbf{n}(\mathbf{x}), Q_f[\mathbf{n}]\} = \varepsilon^{rs} \partial_r f(\mathbf{x}) \partial_s \mathbf{n}(\mathbf{x}) . \quad (9.95)$$

Curiously, this makes it possible to identify $P^r \equiv Q_{\varepsilon^{rs} x^s}$ with the generator of spatial translations, that is momentum. The real surprise comes now: according to (9.94),

$$\{P^r[\mathbf{n}], P^s[\mathbf{n}]\} = 4\pi M \varepsilon^{rs} w[\mathbf{n}] . \quad (9.96)$$

For skyrmions, which carry nonzero $w[\mathbf{n}]$, the two components of momentum do not commute with each other. This has a simple physical interpretation. Namely, for nonzero $w[\mathbf{n}]$,

$$x_w^r[\mathbf{n}] \equiv \frac{1}{8\pi w[\mathbf{n}]} \int d^2\mathbf{x} x^r \varepsilon^{\mu\nu} \varepsilon_{ijk} n^i(\mathbf{x}) \partial_\mu n^j(\mathbf{x}) \partial_\nu n^k(\mathbf{x}) = -\frac{\varepsilon^{rs} P_s[\mathbf{n}]}{4\pi M w[\mathbf{n}]} \quad (9.97)$$

represents the center of the topological charge distribution. Conservation of momentum in the absence of external forces then implies conservation of the center of charge. Skyrmions are pinned to a fixed position unless an external field is applied to them. The fact that $\{x_w^r[\mathbf{n}], x_w^s[\mathbf{n}]\} = \varepsilon^{rs} / (4\pi M w[\mathbf{n}])$ hints that the dynamics of skyrmions closely resembles that of a charged particle in a magnetic field.

References

1. R.L. Workman, et al., PTEP **2022**, 083C01 (2022). DOI [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097)
2. V.P. Nair, *Quantum Field Theory: A Modern Perspective* (Springer Science+Business Media, New York, 2005)
3. S. Scherer, M.R. Schindler, *A Primer for Chiral Perturbation Theory, Lecture Notes in Physics*, vol. 830 (Springer, 2012). DOI [10.1007/978-3-642-19254-8](https://doi.org/10.1007/978-3-642-19254-8)
4. J. Gasser, H. Leutwyler, Ann. Phys. **158**, 142 (1984). DOI [10.1016/0003-4916\(84\)90242-2](https://doi.org/10.1016/0003-4916(84)90242-2)
5. J. Gasser, H. Leutwyler, Nucl. Phys. **B250**, 465 (1985). DOI [10.1016/0550-3213\(85\)90492-4](https://doi.org/10.1016/0550-3213(85)90492-4)
6. J. Matoušek, J. Nešetřil, *Invitation to Discrete Mathematics* (Oxford University Press, Oxford, UK, 2009)
7. J.O. Andersen, T. Brauner, C.P. Hofmann, A. Vuorinen, JHEP **08**, 088 (2014). DOI [10.1007/JHEP08\(2014\)088](https://doi.org/10.1007/JHEP08(2014)088)
8. H. Watanabe, T. Brauner, H. Murayama, Phys. Rev. Lett. **111**, 021601 (2013). DOI [10.1103/PhysRevLett.111.021601](https://doi.org/10.1103/PhysRevLett.111.021601)
9. M. Mannarelli, Particles **2**, 411 (2019). DOI [10.3390/particles2030025](https://doi.org/10.3390/particles2030025)
10. M.E. Peskin, D.V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, Reading, USA, 1995)
11. J.R. Andersen, et al., Eur. Phys. J. Plus **126**, 81 (2011). DOI [10.1140/epjp/i2011-11081-1](https://doi.org/10.1140/epjp/i2011-11081-1)
12. E. D'Hoker, S. Weinberg, Phys. Rev. **D50**, 6050 (1994). DOI [10.1103/PhysRevD.50.R6050](https://doi.org/10.1103/PhysRevD.50.R6050)
13. E. D'Hoker, Nucl. Phys. **B451**, 725 (1995). DOI [10.1016/0550-3213\(95\)00265-T](https://doi.org/10.1016/0550-3213(95)00265-T)
14. T. Brauner, H. Kolečová, Nucl. Phys. **B945**, 114676 (2019). DOI [10.1016/j.nuclphysb.2019.114676](https://doi.org/10.1016/j.nuclphysb.2019.114676)
15. S. Weinberg, *The Quantum Theory of Fields*, vol. II (Cambridge University Press, Cambridge, UK, 1996)
16. E. Witten, Nucl. Phys. **B223**, 422 (1983). DOI [10.1016/0550-3213\(83\)90063-9](https://doi.org/10.1016/0550-3213(83)90063-9)
17. J.M. Román, J. Soto, Int. J. Mod. Phys. **B13**, 755 (1999). DOI [10.1142/S0217979299000655](https://doi.org/10.1142/S0217979299000655)
18. Y. Togawa, Y. Kousaka, K. Inoue, J.i. Kishine, J. Phys. Soc. Jpn. **85**, 112001 (2016). DOI [10.7566/JPSJ.85.112001](https://doi.org/10.7566/JPSJ.85.112001)
19. T.R. Kirkpatrick, D. Belitz, Phys. Rev. **B72**, 180402 (2005). DOI [10.1103/PhysRevB.72.180402](https://doi.org/10.1103/PhysRevB.72.180402)
20. J.H. Han, *Skyrmions in Condensed Matter, Springer Tracts in Modern Physics*, vol. 278 (Springer, 2017). DOI [10.1007/978-3-319-69246-3](https://doi.org/10.1007/978-3-319-69246-3)
21. D. Tong, [arXiv:1606.06687](https://arxiv.org/abs/1606.06687)
22. H. Watanabe, H. Murayama, Phys. Rev. **X4**, 031057 (2014). DOI [10.1103/PhysRevX.4.031057](https://doi.org/10.1103/PhysRevX.4.031057)

Chapter 10

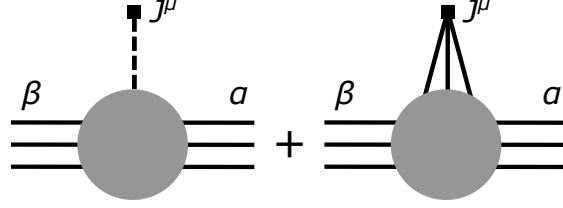
Scattering of Nambu–Goldstone Bosons

In our survey of *spontaneous symmetry breaking* (SSB), we have so far been mostly concerned with the ground state and excitation spectrum. Yet, the *effective field theory* (EFT) framework developed in Chap. 8 captures full nonlinear dependence of the action on the *Nambu–Goldstone* (NG) fields. It thus also includes all interactions among NG bosons and their interactions with other excitations, if present. Microscopic interactions in quantum systems may have many different macroscopic manifestations. However, there is an important class of observables that are amenable to both systematic computation and experiment: scattering amplitudes of free asymptotic states. I will therefore conclude the part of the book devoted to spontaneous breaking of internal symmetry with a short primer on scattering of NG bosons.

I will start in Sect. 10.1 by revisiting the *Adler zero* principle that scattering amplitudes of NG bosons tend to vanish in the long-wavelength limit. I will offer a general justification of this phenomenon, and underline the nature of exceptions to the rule. A modern approach to scattering of NG bosons that utilizes the geometry of the coset space of broken symmetry is introduced in Sect. 10.2. In Sect. 10.3, I demonstrate that there are exceptional EFTs whose long-wavelength scattering amplitudes are even softer than naively expected from the Adler zero principle. These theories set a benchmark for many modern developments in the rapidly expanding field of scattering amplitudes. One important line of development concerns the recursive reconstruction of all (tree-level) scattering amplitudes from a finite number of “seed” amplitudes. This is the subject of Sect. 10.4.

Scattering amplitudes as observables are arguably more important in fundamental (subatomic) physics than in other branches such as condensed-matter physics. I will therefore restrict the discussion to Lorentz-invariant systems. Here NG bosons are massless particles and the kinematics of their scattering is well-understood. Much of the material of this chapter can however be generalized to nonrelativistic EFTs with continuous spatial rotation symmetry; see [1, 2] for details. Let me stress that even amplitudes in relativistic systems have grown into a vast subject, of which this chapter only exposes a small corner. A reader interested in a broader introduction to the physics of scattering amplitudes is recommended to start with the lecture notes [3] or the book [4].

Fig. 10.1 Sketch of contributions to the matrix element $\langle\beta|J^\mu(x)|\alpha\rangle$. In the former, the current creates an intermediate one-particle NG state $|\pi(\mathbf{p})\rangle$ (dashed line) that couples to $|\alpha\rangle$ and $|\beta\rangle$ through the amplitude $\mathcal{A}_{\alpha\rightarrow\beta+\pi(\mathbf{p})}$. In the latter, the current connects to $|\alpha\rangle$, $|\beta\rangle$ via a multiparticle intermediate state



10.1 Adler Zero Revisited

One of the hallmarks of SSB is the existence of a local conserved current J^μ that couples to the NG boson state; cf. the proof of Goldstone’s theorem in Sect. 6.2.1. This can be used to extract nonperturbative information about scattering processes involving one or more NG bosons. We start with a generic process whose initial and final states $|\alpha\rangle$ and $|\beta\rangle$ may include any set of particles of arbitrary mass and spin, and inspect the matrix element $\langle\beta|J^\mu(x)|\alpha\rangle$. This receives two qualitatively different types of contributions, schematically shown in Fig. 10.1. We shall focus on the first contribution, which amounts to the creation or annihilation of a one-particle NG boson by the current.

The NG boson may be an elementary excitation of a field in the action. The contributions with an intermediate one-particle NG state then correspond to one-particle-reducible diagrams where cutting a single propagator disconnects the current from the $\alpha \rightarrow \beta$ process. However, the NG boson may also be composite. The NG state then reveals itself through a single-particle pole in the matrix element $\langle\beta|J^\mu(x)|\alpha\rangle$ at energy–momentum $p^\mu \equiv p_\alpha^\mu - p_\beta^\mu$. Contributions to the matrix element without such a single-particle pole are represented by the second diagram in Fig. 10.1.

We use the translation property of the current operator, $J^\mu(x) = e^{iP \cdot x} J^\mu(0) e^{-iP \cdot x}$, where P^μ is the operator of energy–momentum. This holds whenever the current does not depend explicitly on spacetime coordinates, which is the case for internal symmetry. According to the rules of polology (Sect. 19.2 of [5]), the residue of $\langle\beta|J^\mu(x)|\alpha\rangle$ at the one-particle NG pole is then $e^{-iP \cdot x} \langle 0|J^\mu(0)|\pi(\mathbf{p})\rangle \mathcal{A}_{\alpha\rightarrow\beta+\pi(\mathbf{p})}$. Here $\mathcal{A}_{\alpha\rightarrow\beta+\pi(\mathbf{p})}$ is the invariant amplitude for the process where a NG boson of momentum \mathbf{p} is added to the final state.¹ With a Lorentz-invariant vacuum $|0\rangle$, the matrix element $\langle 0|J^\mu(0)|\pi(\mathbf{p})\rangle$ transforms as a Lorentz vector, as one readily verifies using the properties of single-particle states (Sect. 2.5 of [6]). Lorentz invariance therefore dictates that $\langle 0|J^\mu(0)|\pi(\mathbf{p})\rangle = i p_{\text{on}}^\mu F$, where F is a constant and $p_{\text{on}}^\mu \equiv (|\mathbf{p}|, \mathbf{p})$ is an on-shell energy–momentum with spatial part \mathbf{p} . The matrix element $\langle\beta|J^\mu(0)|\alpha\rangle$ can now be split into a pole part and a regular part as

¹ In contrast to Chap. 6, I use here the normalization $\langle\pi(\mathbf{p})|\pi(\mathbf{q})\rangle = (2\pi)^d 2|\mathbf{p}| \delta^d(\mathbf{p} - \mathbf{q})$ of the one-particle NG states. The amplitude $\mathcal{A}_{\alpha\rightarrow\beta}$ for a generic process $\alpha \rightarrow \beta$ is normalized by its relation to the S -matrix, $S_{\alpha\rightarrow\beta} = -(2\pi)^D i \mathcal{A}_{\alpha\rightarrow\beta} \delta^D(p_\alpha - p_\beta)$. Finally, adding the NG boson to $|\beta\rangle$ rather than $|\alpha\rangle$ is a convention, amounting to the choice of orientation of p^μ .

$$\langle \beta | J^\mu(0) | \alpha \rangle = \langle 0 | J^\mu(0) | \pi(\mathbf{p}) \rangle_{\text{off}} \frac{1}{p^2} \mathcal{A}_{\alpha \rightarrow \beta + \pi}(\mathbf{p}) + R_{\beta\alpha}^\mu(p), \quad (10.1)$$

where $\langle 0 | J^\mu(0) | \pi(\mathbf{p}) \rangle_{\text{off}} \equiv i p^\mu F$. The as yet unspecified function $R_{\beta\alpha}^\mu(p)$ collects all the nonpole contributions, and thus remains regular in the on-shell limit $p^2 \rightarrow 0$.

Polology suggests that the numerator of the pole in (10.1) should be proportional to $\langle 0 | J^\mu(0) | \pi(\mathbf{p}) \rangle$. Replacing this with $\langle 0 | J^\mu(0) | \pi(\mathbf{p}) \rangle_{\text{off}}$ amounts to trading p_{on}^μ for p^μ , and in turn to a redefinition of $R_{\beta\alpha}^\mu(p)$ without changing the residue at the pole. The reason for doing so will become clear below when we work out an explicit example. Here let me just stress that $\langle 0 | J^\mu(0) | \pi(\mathbf{p}) \rangle_{\text{off}}$ is a mere symbolic notation for $i p^\mu F$. This object is not a well-defined matrix element but rather depends on the energy-momenta p_α^μ and p_β^μ .

We now combine (10.1) with current conservation expressed as $p_\mu \langle \beta | J^\mu(0) | \alpha \rangle = 0$. This leads to a master identity which allows one to reconstruct the on-shell scattering amplitude from the remainder function $R_{\beta\alpha}^\mu(p)$,

$$\mathcal{A}_{\alpha \rightarrow \beta + \pi}(\mathbf{p}) = \frac{i}{F} p_\mu R_{\beta\alpha}^\mu(p). \quad (10.2)$$

Example 10.1

Recall the toy model (2.1), introduced in Chap. 2. To simplify the notation, I will only consider the limit of exact symmetry, $\epsilon \rightarrow 0$, and disregard the fermionic degrees of freedom. In the linear parameterization of the complex field, $\phi(x) = e^{i\theta} [v + \chi(x) + i\pi(x)]/\sqrt{2}$, the axial current in (2.3) becomes

$$J_A^\mu = -2i(\phi^* \partial^\mu \phi - \partial^\mu \phi^* \phi) = 2v \partial^\mu \pi + 2(\chi \partial^\mu \pi - \pi \partial^\mu \chi), \quad (10.3)$$

whence we read off $F = -2v$. Let us have a look at the process $\chi\pi \rightarrow \chi\pi$ in which a NG boson scatters off a massive (Higgs) particle. We set $|\alpha\rangle = |\chi(\mathbf{p}_1)\pi(\mathbf{p}_2)\rangle$ and $|\beta\rangle = |\chi(\mathbf{p}_3)\rangle$. The “missing” energy-momentum p_4^μ is supplied by the current. Using the Feynman rules (2.15), the remainder function $R_{\beta\alpha}^\mu(p_4)$ collecting nonpole contributions to $\langle \beta | J_A^\mu(0) | \alpha \rangle$ turns out to be

$$\begin{aligned} R_{\beta\alpha}^\mu(p_4) &= \text{diagram 1} + \text{diagram 2} + \text{diagram 3} \\ &= -4i\lambda v \frac{(2p_3 + p_4)^\mu}{s} + 12i\lambda v \frac{(p_4 - 2p_2)^\mu}{t - m_\chi^2} + 4i\lambda v \frac{(2p_1 - p_4)^\mu}{u}, \end{aligned} \quad (10.4)$$

where the solid squares indicate an insertion of the axial current operator. Contracting this with p_4^μ gives

$$ip_{4\mu} R_{\beta\alpha}^\mu(p_4) = -4\lambda v \left[1 + m_\chi^2 \left(\frac{1}{s} + \frac{3}{t - m_\chi^2} + \frac{1}{u} \right) \right]. \quad (10.5)$$

Using finally (10.2), this reproduces the previously calculated amplitude (2.20).

Note that when evaluating the remainder function $R_{\beta\alpha}^\mu(p_4)$, I ignored the contributions of Feynman diagrams where an external π -type leg terminates in the current operator. These arise from the part of the current linear in π , and are included in the pole part of (10.1). Had we used therein the on-shell matrix element $\langle 0|J^\mu(0)|\pi(\mathbf{p})\rangle$, we would have had to include in (10.4) the contributions of the mentioned diagrams with the pole canceled by the numerator factor $p^\mu - p_{\text{on}}^\mu$. At the same time, the right-hand side of (10.2) would receive an additional factor of 2, owing to the fact that the on-shell limit of $p \cdot p_{\text{on}}/p^2$ is $1/2$. The conclusion that the amplitude can be reconstructed from $R_{\beta\alpha}^\mu(p_4)$ would however remain.

It is instructive to reproduce the same result also with the other parameterization considered in Chap. 2, $\phi(x) = e^{i\theta} \exp[i\pi(x)/v][v + \chi(x)]/\sqrt{2}$. In this case, the axial current reads

$$J_A^\mu = 2v\partial^\mu\pi + 4\chi\partial^\mu\pi + \frac{2}{v}\chi^2\partial^\mu\pi, \quad (10.6)$$

whereas the Feynman rules change to (2.28). The presence of the cubic term in the current adds another Feynman diagram to the expression for the remainder function,

$$\begin{aligned} R_{\beta\alpha}^\mu(p_4) &= \text{diagram 1} + \text{diagram 2} + \text{diagram 3} + \text{diagram 4} \\ &= \frac{4i}{v} \frac{s - m_\chi^2}{s} (p_3 + p_4)^\mu - \frac{24i\lambda v p_2^\mu}{t - m_\chi^2} - \frac{4i}{v} \frac{u - m_\chi^2}{u} (p_1 - p_4)^\mu - \frac{4ip_2^\mu}{v}. \end{aligned} \quad (10.7)$$

This reproduces the same scattering amplitude, in the form shown in (2.30).

What is most interesting about (10.2) is the extra factor of p_μ on the right-hand side, arising from current conservation. This shows that

$$\lim_{\mathbf{p} \rightarrow \mathbf{0}} \mathcal{A}_{\alpha \rightarrow \beta + \pi(\mathbf{p})} = 0, \quad (10.8)$$

provided the remainder function is nonsingular in this limit. This is the *Adler zero*. The same conclusion applies to processes with a NG boson inserted in the initial state. This distinction is trivial in Lorentz-invariant systems; a pedagogical proof of (10.8) for systems with mere spatial rotation invariance can be found in Sect. 3.1 of [1]. I will not reproduce the details here since they have no bearing on the result.

The function $R_{\beta\alpha}^\mu(p)$ is by construction nonsingular in the on-shell limit $p^2 \rightarrow 0$. The absence of singularities for $\mathbf{p} \rightarrow \mathbf{0}$ (or equivalently $p^\mu \rightarrow 0$) is an additional assumption. Finding a mechanism that leads to a violation of this assumption is required for understanding the origin of exceptions to the Adler zero principle.

10.1.1 Generalized Soft Theorem

Before we embark on a more detailed analysis of the conditions under which the Adler zero (10.8) is realized, let me introduce some terminology. Taking the momentum of a massless particle (a NG boson or, for instance, a gauge boson) to zero is generally referred to as the (single) *soft limit*. The statement of Adler zero is equivalent to the vanishing of the soft limit of scattering amplitudes of NG bosons. A universal statement about the soft limit of scattering amplitudes such as (10.8) is called a *soft theorem*. We will see that even in theories where the soft limit of $\mathcal{A}_{\alpha \rightarrow \beta + \pi(p)}$ is nonvanishing, one can still establish a generalized soft theorem. This relates the soft limit of $\mathcal{A}_{\alpha \rightarrow \beta + \pi(p)}$ to the amplitude $\mathcal{A}_{\alpha \rightarrow \beta}$ with the NG boson removed.

We noticed in Chap. 2 that scattering the NG boson off the fermion within our toy model violates the Adler zero principle. Operationally, this could be understood as a consequence of the singularity of the fermion propagator in (2.21) in the soft limit for the NG boson. This singularity in turn arises from the cubic coupling between the NG boson and the fermion. The origin of the singularity is however purely kinematical and not specific to coupling of NG bosons to fermions; the same obstruction to the Adler zero may occur in purely bosonic theories. It is, in fact, commonplace in EFTs where all degrees of freedom are NG bosons, as constructed in Chap. 8. The Adler zero is therefore a much less robust consequence of SSB than, say, the very existence of gapless NG bosons.

Let us now examine the hypothesis that the violation of the Adler zero principle arises from the presence of cubic interaction vertices in the theory. To keep things simple, I will from now on restrict to relativistic EFTs of NG bosons, disregarding any non-NG degrees of freedom. The argument below closely follows [7].

Consider a set of NG fields π^a in a theory defined by the generic Lagrangian

$$\mathcal{L}_{\text{eff}} = \frac{1}{2} \delta_{ab} \partial_\mu \pi^a \partial^\mu \pi^b + \frac{1}{2} \lambda_{abc} \partial_\mu \pi^a \partial^\mu \pi^b \pi^c + \mathcal{O}(\pi^4, \partial^4) . \quad (10.9)$$

The coupling λ_{abc} can be assumed to be symmetric in its a, b indices. There are no terms without derivatives as guaranteed by the broken symmetry. Also, we can restrict the discussion to operators with two derivatives in line with the power counting laid out in Sect. 9.1.1. Finally, I assume that the kinetic term is diagonal in the NG flavor indices and properly normalized. This can always be ensured by a linear transformation of the NG fields. Other than that, however, I do not implicitly assume any particular parameterization of the fields.

The first thing to notice is that the cubic vertex in (10.9) alone cannot be the sole culprit in violating the Adler zero principle. Observable properties of scattering amplitudes must be unchanged by a field redefinition of the type

$$\pi^a \rightarrow \pi'^a(\pi) = \pi^a + \frac{1}{2} c_{bc}^a \pi^b \pi^c + \mathcal{O}(\pi^3) , \quad (10.10)$$

where c_{bc}^a is symmetric in b, c . In particular, choosing $c_{bc}^a = (1/2)\delta^{ad}(\lambda_{dbc} + \lambda_{dcb} - \lambda_{bcd})$ eliminates the cubic interaction vertex altogether. In fact, even cubic operators with more than two derivatives can be removed analogously. This redundancy of cubic couplings of NG bosons is inherent to relativistic EFTs for SSB.

For the Lagrangian (10.9) to actually represent an EFT for NG bosons, it must preserve a symmetry that is spontaneously broken. We do not need to invoke the full machinery of nonlinear realizations at this stage. It is sufficient to assume the existence of a set of spontaneously broken symmetries, one for each NG field,

$$\delta\pi^a \equiv \epsilon^b [F_b^a + G_{bc}^a \pi^c + O(\pi^2)] . \quad (10.11)$$

The independence of these transformations and the fact that they are all spontaneously broken amount to the condition that F_b^a is an invertible square matrix. The assumed invariance of (10.9) then leads to the following set of Noether currents,

$$J_a^\mu = \delta_{bc} F_a^c \partial^\mu \pi^b + K_{abc} \partial^\mu \pi^b \pi^c + O(\pi^3) , \quad (10.12)$$

where $K_{abc} \equiv \delta_{bd} G_{ac}^d + F_a^d \lambda_{dbc}$.

Let us now evaluate, schematically, the amplitude $\mathcal{A}_{\alpha \rightarrow \beta + \pi^a(p)}$ for a process where $|\alpha\rangle, |\beta\rangle$ are states with an arbitrary number of NG bosons. The soft NG boson state $|\pi^a(p)\rangle$ is added to the final state $|\beta\rangle$. Then (10.1) is replaced with

$$\langle \beta | J_a^\mu(0) | \alpha \rangle = \sum_b \langle 0 | J_a^\mu(0) | \pi^b(p) \rangle_{\text{off}} \frac{1}{p^2} \mathcal{A}_{\alpha \rightarrow \beta + \pi^b(p)} + R_{\beta\alpha a}^\mu(p) . \quad (10.13)$$

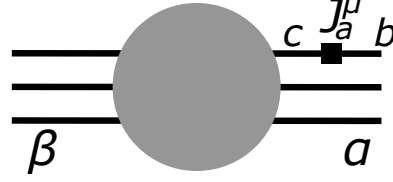
From (10.12) we find that $\langle 0 | J_a^\mu(0) | \pi^b(p) \rangle = -ip_{\text{on}}^\mu F_a^b$. Applying momentum conservation leads to $\sum_b F_a^b \mathcal{A}_{\alpha \rightarrow \beta + \pi^b(p)} = -ip_\mu R_{\beta\alpha a}^\mu(p)$. We expect the Adler zero property of $\mathcal{A}_{\alpha \rightarrow \beta + \pi^b(p)}$ to be violated if the limit $p \rightarrow 0$ makes some of the propagators in the process on-shell. Barring fine tuning of the kinematics, this happens when the NG state $|\pi^b(p)\rangle$ is “attached” to an external leg of the $\alpha \rightarrow \beta$ process via a cubic interaction vertex. In terms of $R_{\beta\alpha a}^\mu(p)$, such singular contributions map to insertions of the bilinear part of J_a^μ into one of the external legs of $\alpha \rightarrow \beta$. See Fig. 10.2 for a visualization and the index notation.

Near the limit of vanishing p^μ , inserting the current into an external leg with flavor b , carrying energy–momentum p_b^μ , gives the singular contributions

$$\begin{aligned} R_{\beta\alpha a}^\mu(p) &\supset \sum_c (-i\mathcal{A}_{\alpha \rightarrow \beta}^{b \rightarrow c}) \frac{i}{(p_b - p)^2} [-iK_{abc} p_b^\mu + iK_{acb} (p_b - p)^\mu] \\ &\approx \frac{i}{2} \sum_c \mathcal{A}_{\alpha \rightarrow \beta}^{b \rightarrow c} \frac{p_b^\mu}{p \cdot p_b} (K_{abc} - K_{acb}) . \end{aligned} \quad (10.14)$$

The superscript $b \rightarrow c$ indicates a replacement of the flavor index on the leg where the insertion has been made. Upon summing over all possible insertions, we arrive

Fig. 10.2 Sketch of non-pole contributions to $\langle \beta | J_a^\mu(0) | \alpha \rangle$ that are singular in the limit $p^\mu \equiv p_\alpha^\mu - p_\beta^\mu \rightarrow 0$. These contributions arise from inserting the bilinear part of the Noether current (10.12) into an external leg of the $\alpha \rightarrow \beta$ process. The insertion is indicated by the *solid square*



at the final form of our generalized soft theorem,

$$\lim_{p \rightarrow 0} \sum_b F_a^b \mathcal{A}_{\alpha \rightarrow \beta + \pi^b(p)} = \frac{1}{2} \sum_{b \in \alpha \cup \beta} \sum_c (K_{abc} - K_{acb}) \mathcal{A}_{\alpha \rightarrow \beta}^{b \rightarrow c}. \quad (10.15)$$

Let me spell out some properties of this result explicitly. First, the soft theorem (10.15) relates the soft limit of an n -particle amplitude to a set of $(n-1)$ -particle amplitudes, in which the soft NG leg is removed and the label on one other leg is changed. The only other ingredient needed is the set of algebraic coefficients K_{abc} , which is independent of the concrete scattering process considered. In this sense, the soft theorem (10.15) is universal.

Second, as is evident from (10.12), the coefficients K_{abc} receive two types of contributions: one from the cubic coupling λ_{abc} , the other from the part G_{bc}^a of the symmetry transformation (10.11), linear in NG fields. Importantly, the combination $K_{abc} - K_{acb}$, antisymmetric under the exchange $b \leftrightarrow c$, is manifestly invariant under the field redefinition (10.10). This ensures that the soft theorem (10.15) has a physical meaning independent of the choice of field parameterization.

10.1.2 Application to Coset Effective Theories

The above argument is fairly general and in principle also applies to scattering amplitudes in theories where some of the particles in $|\alpha\rangle$ and $|\beta\rangle$ are not NG bosons. The only essential constraint is that there are no cubic operators without derivatives in the Lagrangian. The sum in (10.15) is then to be restricted to c such that π^c has the same mass as π^b . However, once the Lagrangian (10.9) does represent an EFT with only NG boson degrees of freedom, we can use the wealth of information accumulated in Chaps. 7 and 8.

Let us recall some of the basic relations that we are going to use here. First, the two-derivative part of the effective Lagrangian for the coset space G/H reads

$$\mathcal{L}_{\text{eff}}^{(2)} = \frac{1}{2} \kappa_{cd} \omega_a^c(\pi) \omega_b^d(\pi) \partial_\mu \pi^a \partial^\mu \pi^b, \quad (10.16)$$

where the constant symmetric matrix κ_{ab} satisfies the constraint $f_{\alpha a}^c \kappa_{cb} + f_{\alpha b}^c \kappa_{ac} = 0$. The index α labels generators of the unbroken subgroup H . With the exponential parameterization of the coset space, $U(\pi) = \exp(i\pi^a Q_a)$, the *Maurer–Cartan* (MC) form can be computed explicitly as a power series in the NG fields,

$$\omega_a^A(\pi) = \delta_a^A - \frac{1}{2} f_{ab}^A \pi^b + \mathcal{O}(\pi^2), \quad (10.17)$$

where the index A runs over all generators of G . Likewise, infinitesimal transformations induced on the coset space by G , $\delta\pi = \epsilon^A \xi_A^a(\pi)$, are determined by a set of Killing vectors,

$$\xi_A^a(\pi) = \delta_A^a - \left(f_{Ab}^a - \frac{1}{2} \delta_A^e f_{eb}^a \right) \pi^b + \mathcal{O}(\pi^2). \quad (10.18)$$

To make the bilinear (kinetic) part of the Lagrangian (10.16) canonically normalized, one can set $\kappa_{ab} = \delta_{ab}$ by a suitable linear transformation of the NG fields. To preserve the form of the exponential parameterization and hence of (10.17) and (10.18), this has to be compensated by a linear transformation of the basis of broken generators Q_a . That in turn affects the values of the structure constants. Below, I will assume that such a choice of basis of broken generators has already been made.

With all these pieces at hand, it is easy to construct the Noether currents for the spontaneously broken generators of G ,

$$\begin{aligned} J_a^\mu &= \frac{\partial \mathcal{L}_{\text{eff}}^{(2)}}{\partial (\partial_\mu \pi^b)} \xi_a^b(\pi) \\ &= \delta_{ab} \partial^\mu \pi^b - \left(\delta_{bd} f_{ac}^d + \frac{1}{2} \delta_{ad} f_{bc}^d \right) \partial^\mu \pi^b \pi^c + \mathcal{O}(\pi^3). \end{aligned} \quad (10.19)$$

By matching this to (10.12), we can read off the algebraic coefficients needed in the soft theorem (10.15), namely $F_a^b = \delta_a^b$ and

$$K_{abc} - K_{acb} = \delta_{cd} f_{ab}^d - \delta_{bd} f_{ac}^d - \delta_{ad} f_{bc}^d. \quad (10.20)$$

Example 10.2

Let G be compact and semisimple. Using a faithful matrix representation of G , one can define the Cartan–Killing form on the Lie algebra \mathfrak{g} as $\Delta_{AB} \equiv \text{tr}(Q_A Q_B)$. The cyclicity of trace, $\text{tr}([Q_A, Q_B] Q_C) = \text{tr}(Q_A [Q_B, Q_C])$, implies the identity

$$f_{CA}^D \Delta_{DB} + f_{CB}^D \Delta_{AD} = 0. \quad (10.21)$$

This encodes the invariance of the Cartan–Killing form under the adjoint action of G on \mathfrak{g} . At the same time, (10.21) ensures that the covariant structure constant $f_{ABC} \equiv \Delta_{AD} f_{BC}^D$ is fully antisymmetric in its three indices.

It is convenient to choose the basis of broken generators Q_a so that the space $\mathfrak{g}/\mathfrak{h}$ is “orthogonal” to \mathfrak{h} , that is, $\Delta_{a\beta} = 0$. The unbroken and broken indices can then be raised and lowered independently using $\Delta_{\alpha\beta}$ and Δ_{ab} , respectively. The constraint $f_{\alpha a}^c \kappa_{cb} + f_{\alpha b}^c \kappa_{ac} = 0$ can be rewritten as $\kappa^a_c f_{\alpha b}^c - f_{\alpha c}^a \kappa^c_b = 0$. Thus, the matrix κ^a_b commutes with the representation of H on the space $\mathfrak{g}/\mathfrak{h}$. Suppose in particular that this representation is irreducible, that is, all the NG modes of G/H span a single irreducible multiplet of H . This is the case for a number of physically important coset spaces such as $\text{SO}(n+1)/\text{SO}(n) \simeq S^n$, $\text{ISO}(n)/\text{SO}(n) \simeq \mathbb{R}^n$, or $G_L \times G_R/G_V$ with simple G . Then by Schur’s lemma, κ^a_b must be proportional to δ^a_b , and thus κ_{ab} to Δ_{ab} . The choice $\kappa_{ab} = \delta_{ab}$ here corresponds to the common normalization of symmetry generators such that $\text{tr}(Q_A Q_B) \propto \delta_{AB}$. Finally, (10.20) reduces to

$$K_{abc} - K_{acb} \propto f_{cab} - f_{bac} - f_{abc} = f_{abc} . \quad (10.22)$$

In this special case, the right-hand side of (10.15) becomes $(1/2) \sum_{b \in \alpha \cup \beta} \sum_c f_{abc} \mathcal{A}_{\alpha \rightarrow \beta}^{b \rightarrow c}$ up to an overall factor.

It is obvious from (10.15) and (10.20) that a sufficient condition for the presence of Adler zero is that the coset space G/H is symmetric so that $f_{bc}^a = 0$. In fact, for symmetric coset spaces, the reason for vanishing of the right-hand side of (10.15) is twofold. Namely, in the exponential parameterization, the automorphism $\mathcal{R}(U(\pi)) = U(\pi)^{-1}$ is equivalent to the inversion $\pi^a \rightarrow -\pi^a$. As a consequence, the components of the broken part of the MC form, $\omega_b^a(\pi)$, are manifestly even functions of the NG fields. Hence, for symmetric coset spaces, the effective Lagrangian (10.16) can only give scattering amplitudes with an even number of particles. For theories with only even interaction vertices, the soft theorem (10.15) applied to an amplitude with an even number of particles yields Adler zero without any further requirements on the coefficients K_{abc} .

10.2 Geometric Framework for Scattering Amplitudes

In the analysis in Sect. 10.1.1, the freedom to redefine the field variables π^a was something of a nuisance. Indeed, we had to demonstrate explicitly that our soft theorem (10.15) is invariant under field reparameterization and thus physically meaningful. Here we will take control and make the freedom to redefine fields work for us. The discussion in this section loosely follows [8].

Suppose we are given a set of fields π^a taking values from some manifold \mathcal{M} . For the moment, we do not even have to assume that these are NG fields and \mathcal{M} is a coset space, $\mathcal{M} \simeq G/H$. All we need is that the low-energy physics of the system is captured by the Lagrangian

$$\mathcal{L}_{\text{eff}} = \frac{1}{2} g_{ab}(\pi) \partial_\mu \pi^a \partial^\mu \pi^b . \quad (10.23)$$

In particular, this requires that there are no interactions without derivatives, and any higher-derivative operators can be treated as subleading. Under a point transformation, $\pi^a \rightarrow \pi'^a(\pi)$, the symmetric matrix function $g_{ab}(\pi)$ on \mathcal{M} behaves as a rank-2 covariant tensor. Moreover, it must be positive-definite in some neighborhood of the origin, $\pi^a = 0$, to give a kinetic term with the correct signature. It can therefore be interpreted as a Riemannian metric on the target manifold \mathcal{M} .

Perturbative interactions among the fields π^a are generated by expanding the metric in powers of the fields. Individual interaction vertices are thus determined by derivatives of the metric at the origin. However, physical observables must at the same time be invariant under an arbitrary redefinition of the fields. We can use this fact and switch to the Riemannian (geodesic) normal coordinates on \mathcal{M} , in which (see Appendix A.6.4)

$$g_{ab}(\pi) = g_{ab}(0) - \frac{1}{3}\pi^c\pi^d R_{acbd}(0) - \frac{1}{6}\pi^c\pi^d\pi^e (\hat{\nabla}_e R)_{acbd}(0) + O(\pi^4) . \quad (10.24)$$

Moreover, we can always assume (and I will henceforth do so) that $g_{ab}(0) = \delta_{ab}$ so that the kinetic term in (10.23) is canonically normalized. This shows that any observable such as an on-shell scattering amplitude depends on $g_{ab}(\pi)$ solely through the Riemann curvature tensor and its covariant derivatives at the origin.

To see how the curvature tensor enters the amplitudes, let me spell out explicitly the simplest cases of the 4-particle and 5-particle amplitude,

$$\begin{aligned} \mathcal{A}_{a_1 a_2 a_3 a_4}(p_1, p_2, p_3, p_4) &= i \quad \text{[diagram: circle with 4 incoming lines]} = -s_{24}R_{a_1 a_2 a_3 a_4} - s_{34}R_{a_1 a_3 a_2 a_4} , \\ \mathcal{A}_{a_1 a_2 a_3 a_4 a_5}(p_1, p_2, p_3, p_4, p_5) &= i \quad \text{[diagram: circle with 5 incoming lines]} \\ &= -s_{45}\hat{\nabla}_{a_3}R_{a_1 a_4 a_2 a_5} - s_{25}\hat{\nabla}_{a_4}R_{a_1 a_2 a_3 a_5} \\ &\quad - s_{35}\hat{\nabla}_{a_4}R_{a_1 a_3 a_2 a_5} - s_{34}\hat{\nabla}_{a_5}R_{a_1 a_3 a_2 a_4} \\ &\quad - (s_{24} + s_{45})\hat{\nabla}_{a_5}R_{a_1 a_2 a_3 a_4} . \end{aligned} \quad (10.25)$$

Here I have introduced a compact notation for a scattering amplitude with particles of energy–momentum p_i^μ and flavor a_i . All the energy–momenta are (now and henceforth) oriented inwards. The variables $s_{ij} \equiv (p_i + p_j)^2$ generalize the standard Mandelstam variables to processes with an arbitrary number of particles. Finally, I have dropped the argument (0) of the curvature tensor for notation simplicity.

The presentation of the amplitudes is highly ambiguous due to the redundancies among the Mandelstam variables and the components of the curvature tensor. To arrive at the expressions in (10.25), I have adopted the following conventions. First, the variables s_{1i} with any $i \geq 2$ can be eliminated through

$s_{1i} = -\sum_{j \geq 2} s_{ji}$, which follows from overall energy–momentum conservation for massless particles. Second, one of the remaining Mandelstam variables can be eliminated by means of $0 = p_1^2 = (p_2 + p_3 + \dots)^2$. I use this to get rid of s_{23} . Similarly, all components $R_{a_i a_j a_k a_l}$ can be expressed solely in terms of those for which $i \leq \min(j, k, l)$ simultaneously with $k < l$ and $j \leq \max(k, l)$. This is guaranteed by the symmetries of the curvature tensor and the algebraic Bianchi identity (see Appendix A.6.3). Finally, the differential Bianchi identity (A.99) can be utilized to remove $\hat{\nabla}_{a_m} R_{a_i a_j a_k a_l}$ where $m < \min(k, l)$.

With the above remarks out of the way, let us have a look at the result. First, the 4-particle amplitude vanishes in the limit where any of the energy–momenta p_i^μ is taken to zero. This boils down to the very special kinematics of relativistic 4-particle scattering processes. The 5-particle amplitude is more interesting. Our expression is designed to make the soft limit for the last particle transparent,²

$$\begin{aligned} \lim_{p_5 \rightarrow 0} \mathcal{A}_{a_1 a_2 a_3 a_4 a_5}(p_1, p_2, p_3, p_4, p_5) &= -s_{24} \hat{\nabla}_{a_5} R_{a_1 a_2 a_3 a_4} - s_{34} \hat{\nabla}_{a_5} R_{a_1 a_3 a_2 a_4} \\ &= \hat{\nabla}_{a_5} \mathcal{A}_{a_1 a_2 a_3 a_4}(p_1, p_2, p_3, p_4) . \end{aligned} \quad (10.26)$$

This is a special case of a geometric soft theorem,

$$\lim_{p_n \rightarrow 0} \mathcal{A}_{a_1 \dots a_n}(p_1, \dots, p_n) = \hat{\nabla}_{a_n} \mathcal{A}_{a_1 \dots a_{n-1}}(p_1, \dots, p_{n-1}) , \quad (10.27)$$

valid for any scalar theory that does not contain nonderivative interactions. There may be other, higher-order derivative couplings additional to (10.23). The covariant derivative of the amplitude in (10.27) is taken in the target space \mathcal{M} with respect to the *Levi-Civita* (LC) connection defined by the metric $g_{ab}(\pi)$.

The reader is referred to [8] for a general nonperturbative proof of (10.27). Here we are mostly interested in EFTs for NG bosons, defined on a coset space of spontaneously broken symmetry. For this class of theories, I will derive below a dedicated geometric soft theorem using the more elementary approach of Sect. 10.1.

10.2.1 Geometric Soft Theorem for Nambu–Goldstone Bosons

Inspired by the above discussion, we shall now reinterpret the results of Sect. 10.1.2 in terms of the geometric properties of the coset space G/H . The Noether current for the spontaneously broken generator Q_a can be written as

$$J_a^\mu = \frac{\partial \mathcal{L}_{\text{eff}}^{(2)}}{\partial (\partial_\mu \pi^b)} \xi_a^b(\pi) = g_{bc}(\pi) \xi_a^b(\pi) \partial^\mu \pi^c . \quad (10.28)$$

² The amplitude is of course invariant under any simultaneous permutation of the energy–momenta and the flavor labels. This permutation invariance is just hidden in the economic expression (10.25).

Expanding this in powers of the NG fields π^a gives

$$J_a^\mu = g_{bc}(0)\xi_a^b(0)\partial^\mu\pi^c + \partial_d(g_{bc}\xi_a^b)(0)\partial^\mu\pi^c\pi^d + \mathcal{O}(\pi^3) . \quad (10.29)$$

I have already made the assumption that the kinetic term is canonically normalized, that is $g_{ab}(0) = \delta_{ab}$. Accordingly, the coupling of the current $J_a^\mu(x)$ to the one-particle state $|\pi^b(\mathbf{p})\rangle$ is defined by the matrix element $\langle 0|J_a^\mu(0)|\pi^b(\mathbf{p})\rangle = -ip_{0n}\xi_a^b(0)$. The bilinear part of the current (10.29) yields in turn the algebraic coefficient $K_{abc} = \partial_c(g_{ab}\xi_a^d)(0) = \partial_c(\xi_a^b)_b(0)$. We only need the part of this, antisymmetric under the exchange $b \leftrightarrow c$, which can be simplified to

$$K_{abc} - K_{acb} = (\hat{\nabla}_c \xi_a^b)_b(0) - (\hat{\nabla}_b \xi_a^b)_c(0) = 2(\hat{\nabla}_c \xi_a^b)_b(0) \quad (10.30)$$

using the Killing equation (A.92). Putting all the pieces together, we infer from (10.15) a geometric soft theorem for the amplitude $\mathcal{A}_{a_1 \dots a_n}(p_1, \dots, p_n)$ for scattering of n NG bosons with flavors a_i and energy–momenta p_i^μ ,

$$\begin{aligned} & \lim_{p_n \rightarrow 0} \xi_a^b(0) \mathcal{A}_{a_1 \dots a_{n-1} b}(p_1, \dots, p_n) \\ &= \sum_{i=1}^{n-1} (\hat{\nabla}^b \xi_a^b)_{a_i}(0) \mathcal{A}_{a_1 \dots a_{i-1} b a_{i+1} \dots a_{n-1}}(p_1, \dots, p_{n-1}) . \end{aligned} \quad (10.31)$$

In [8], the same result was obtained from (10.27) by invoking the G -invariance of the scattering amplitude in the schematic form

$$\mathcal{L}_{\xi_a} \mathcal{A}_{a_1 \dots a_{n-1}} = \xi_a^b \hat{\nabla}_b \mathcal{A}_{a_1 \dots a_{n-1}} + \sum_{i=1}^{n-1} (\hat{\nabla}_{a_i} \xi_a)^b \mathcal{A}_{a_1 \dots a_{i-1} b a_{i+1} \dots a_{n-1}} = 0 . \quad (10.32)$$

The formulation (10.31) of the soft theorem is manifestly covariant under field redefinitions that preserve the normalization of the kinetic term. One can take advantage of this and compute the algebraic coefficient $(\hat{\nabla}^b \xi_a^b)_{a_i}(0)$ in suitably chosen local coordinates on G/H . Specifically, in the exponential parameterization, we already have an expression for the Killing vector in (10.18). Likewise, the expression (10.17) for the MC form can be used to recover the metric $g_{ab}(\pi)$ and in turn the Christoffel symbols, needed to evaluate the covariant derivative of the Killing vector. Putting everything together, one finds

$$(\hat{\nabla}_c \xi_a^b)_b(0) = \frac{1}{2}(\delta_{cd} f_{ab}^d - \delta_{bd} f_{ac}^d - \delta_{ad} f_{bc}^d) , \quad (10.33)$$

in accord with (10.20). This confirms that the soft theorems (10.15) and (10.31), formulated respectively in terms of the algebraic and geometric properties of the coset space, are equivalent.

10.2.2 Adler Zero or Not?

We have already observed that if the coset space G/H is symmetric, the right-hand side of the soft theorem (10.15) or (10.31) automatically vanishes. Is the symmetry of G/H also a necessary condition for the Adler zero? We cannot exclude the possibility that the soft limit of a specific amplitude in a given theory vanishes due to fine tuning of the effective couplings. However, in order that *all* amplitudes of the EFT vanish in the soft limit, the coefficients (10.33) must vanish for any choice of a, b, c . Using the shorthand notation $f_{abc} \equiv \delta_{ad} f_{bc}^d$, the vanishing of (10.33) is in turn equivalent to $f_{abc} = f_{bca} + f_{cab}$. Applying this twice gives

$$f_{abc} = f_{bca} + f_{cab} = (f_{cab} + f_{abc}) + (f_{abc} + f_{bca}) = 2f_{abc} + f_{bca} + f_{cab} . \quad (10.34)$$

This is only possible if $f_{abc} = 0$. Thus, the set of all scattering amplitudes of the EFT satisfies the Adler zero principle if and only if the coset space G/H is symmetric.

Incidentally, there is a neat geometric way to understand the one-way implication, ensuring Adler zero for symmetric coset spaces. Namely, the latter have the property that all covariant derivatives of the Riemann curvature tensor identically vanish (Theorem 10.19 in [9]). The Adler zero can then be seen as a direct consequence of (10.27). On coset spaces that are not symmetric, the nontrivial soft limit of scattering amplitudes is governed by the structure constants f_{bc}^a . According to Sect. 7.4.1, these are in a one-to-one correspondence with the (frame components of the) torsion 2-form of the canonical connection on G/H . Thus, the nonvanishing soft limit can be said to arise geometrically from the torsion of the coset space.

Example 10.3

One of the simplest symmetry groups leading to a nonvanishing soft limit of some scattering amplitudes is the Heisenberg group H_3 . This is a three-dimensional nilpotent Lie group. One can choose a basis of its Lie algebra \mathfrak{h}_3 , Q_a with $a = 1, 2$ and Q , so that the only nontrivial commutation relation among the generators is

$$[Q_a, Q_b] = i\varepsilon_{ab}Q . \quad (10.35)$$

Suppose the symmetry of a system under the Heisenberg group is completely broken. Denoting the NG fields associated with Q_a and Q respectively as π^a and θ , the coset space $H_3/\{e\} \simeq H_3$ can be parameterized by

$$U(\pi, \theta) = \exp\left(\frac{i}{v}\pi^a Q_a\right) \exp\left(\frac{i}{v}\theta Q\right) , \quad (10.36)$$

where v is a positive dimensionful constant. The corresponding components of the MC form are $\omega^a = (1/v)d\pi^a$ and $\omega^\theta = (1/v)d\theta + 1/(2v^2)\varepsilon_{ab}\pi^a d\pi^b$. The most general two-derivative effective Lagrangian would now be given by a generic rank-2 symmetric tensor built out of the MC form. For illustration, it is however sufficient to consider a particularly simple special case,

$$\mathcal{L}_{\text{eff}}^{(2)} = \frac{1}{2} \delta_{ab} \partial_\mu \pi^a \partial^\mu \pi^b + \frac{1}{2} \left(\partial_\mu \theta + \frac{1}{2v} \varepsilon_{ab} \pi^a \partial_\mu \pi^b \right)^2. \quad (10.37)$$

This makes the calculation of various scattering amplitudes easy. Namely, the sole cubic interaction operator equals, up to a surface term, $-1/(2v) \theta \varepsilon_{ab} \pi^a \square \pi^b$. As a consequence, any Feynman diagram where a θ -propagator is attached to two external π -type legs will vanish on-shell.

Any on-shell 3-particle amplitude in a derivatively coupled relativistic theory of massless scalars automatically vanishes. Hence we need to consider a process with at least five particles to have a hope for a nonzero soft limit. Leaving out straightforward details, a simple example is

$$\mathcal{A}_{\theta 1222}(p_1, p_2, p_3, p_4, p_5) = \frac{s_{12}}{2v^3}. \quad (10.38)$$

In order to compare this to the prediction of our soft theorem (10.31), we have to keep in mind the factors of v in (10.36). These make the only nonzero structure constant effectively $f_{ab}^\theta = \varepsilon_{ab}/v$. Together with $\xi_a^b(0) = \delta_a^b$, (10.31) and (10.33) then predict

$$\lim_{p_5 \rightarrow 0} \mathcal{A}_{\theta 1222}(p_1, \dots, p_5) = \frac{1}{2v} [\mathcal{A}_{1122}(p_1, \dots, p_4) - \mathcal{A}_{\theta \theta 22}(p_1, \dots, p_4)]. \quad (10.39)$$

This is easily verified by calculating the two 4-particle amplitudes and finding that

$$\mathcal{A}_{1122}(p_1, p_2, p_3, p_4) = \frac{3s_{12}}{4v^2}, \quad \mathcal{A}_{\theta \theta 22}(p_1, p_2, p_3, p_4) = -\frac{s_{12}}{4v^2}. \quad (10.40)$$

By permutation invariance, the same kind of result applies to the soft limit in which either p_3^μ or p_4^μ is taken to zero. On the other hand, the soft limit of $\mathcal{A}_{\theta 1222}(p_1, \dots, p_5)$ for p_1^μ or p_2^μ manifestly vanishes, as is clear from (10.38).

The cubic interaction operator in (10.37) can be removed by the field redefinition $\pi^a = \phi^a + c \theta \varepsilon^a_b \phi^b$ with $c \equiv 1/(2v)$. This leads to the Lagrangian

$$\begin{aligned} \mathcal{L}_{\text{eff}}^{(2)} = & \frac{1}{2} (1 + c^2 \theta^2) \partial_\mu \boldsymbol{\phi} \cdot \partial^\mu \boldsymbol{\phi} + \frac{1}{2} (1 - c^2 \boldsymbol{\phi}^2) (\partial_\mu \theta)^2 + \frac{c^2}{4} \partial_\mu (\theta^2) \partial^\mu (\boldsymbol{\phi}^2) \\ & + \frac{c^3}{3} \partial_\mu (\theta^3) \boldsymbol{\phi} \times \partial^\mu \boldsymbol{\phi} + \frac{c^2}{2} [(1 + c^2 \theta^2) \boldsymbol{\phi} \times \partial_\mu \boldsymbol{\phi} - c \boldsymbol{\phi}^2 \partial_\mu \theta]^2, \end{aligned} \quad (10.41)$$

where I used the notation $\boldsymbol{\phi} \equiv (\phi^1, \phi^2)$ and $\boldsymbol{\phi} \times \partial_\mu \boldsymbol{\phi} \equiv \varepsilon_{ab} \phi^a \partial_\mu \phi^b$ to avoid proliferation of indices. The first term on the second line of (10.41) is equivalent to $-(c^3/3) \theta^3 \boldsymbol{\phi} \times \square \boldsymbol{\phi}$ by partial integration, and thus does not contribute to tree-level 5-particle amplitudes of the theory. All such amplitudes are therefore encoded in a single operator, $-c^3 \partial^\mu \theta \boldsymbol{\phi}^2 (\boldsymbol{\phi} \times \partial_\mu \boldsymbol{\phi})$. This shows that $\mathcal{A}_{\theta 1222}$ and $\mathcal{A}_{\theta 2111}$ are the only nonzero 5-particle amplitudes, and makes it easy to reproduce (10.38).

10.2.3 Symmetric Coset Spaces

Expressing the scattering amplitudes in terms of the curvature tensor and its covariant derivatives is elegant and gives us useful insight. However, calculating the covariant derivatives of the curvature tensor from the metric $g_{ab}(\pi)$ in practice may be a tiresome task. The situation is much better for symmetric coset spaces where all the covariant derivatives of the curvature tensor vanish. Here we have a closed expression for the Riemannian metric in the normal coordinates (see Appendix A.6.4),

$$g_{ab}(\pi) = g_{ac}(0) \left[\frac{\sin^2 \sqrt{\hat{R}(\pi)}}{\hat{R}(\pi)} \right]_b^c = g_{ac}(0) \sum_{k=0}^{\infty} \frac{(-1)^k 2^{2k+1}}{(2k+2)!} [\hat{R}(\pi)^k]_b^c, \quad (10.42)$$

where $\hat{R}^a_b(\pi) \equiv R^a_{cbd}(0) \pi^c \pi^d$. All we need to know are the constants $R^a_{cbd}(0)$. These can be extracted from our discussion of the geometry of coset spaces in Sect. 7.4. First, however, a word of caution is in place.

Throughout Appendix A, I dutifully distinguish components of tensors in a local frame from those in a local coordinate basis by using different types of indices. This has not been necessary in the main text of the book so far. Here, to avoid confusion, I will use lowercase indices a, b, \dots to indicate a local coordinate system, and underlined indices $\underline{a}, \underline{b}, \dots$ for components in the frame defined by the MC form. Thus, for instance, a comparison of (10.16) and (10.23) shows that the G -invariant Riemannian metric on the coset space is $g(\pi) = \kappa_{ab} \omega^{\underline{a}}(\pi) \otimes \omega^{\underline{b}}(\pi)$. The components of the metric in the MC form basis are constant, as shown explicitly in Sect. 8.1.2.

According to Sect. 7.4.1, the same is true for the curvature 2-form of the LC connection on symmetric coset spaces,

$$R^a_{\underline{b}}(\pi) = -\frac{1}{2} f^a_{\alpha\underline{b}} f^{\alpha}_{\underline{c}\underline{d}} \omega^{\underline{c}}(\pi) \wedge \omega^{\underline{d}}(\pi). \quad (10.43)$$

Combining this with (10.42) gives upon a brief manipulation a practically more useful expression for the metric,

$$g(\pi) = \kappa_{ac} \left[\frac{\sin^2 \sqrt{\hat{R}(\pi)}}{\hat{R}(\pi)} \right]_{\underline{b}}^c d\pi^{\underline{a}} \otimes d\pi^{\underline{b}}, \quad \hat{R}^a_{\underline{b}}(\pi) \equiv -f^a_{\alpha\underline{c}} f^{\alpha}_{\underline{b}\underline{d}} \pi^{\underline{c}} \pi^{\underline{d}}, \quad (10.44)$$

where $\pi^{\underline{a}} \equiv \omega^{\underline{a}}_a(0) \pi^a$. Here $\omega^{\underline{a}}_a(0)$ is a constant matrix that can easily be evaluated from the definition of the MC form, $\omega_{\perp}(\pi) = \omega^{\underline{a}}(\pi) Q_{\underline{a}} = \omega^{\underline{a}}_a(\pi) Q_a d\pi^a$. Combining (10.23) with (10.44) determines the complete two-derivative effective Lagrangian on symmetric coset spaces in a closed form.

The structure constant f^a_{ac} in (10.44) defines the adjoint action of the unbroken subgroup H on the broken generators, and thus also the NG fields. For compact and semisimple groups G , the other type of structure constant in (10.44), f^{α}_{bd} , can be related to the former using the Cartan–Killing form on \mathfrak{g} ; cf. Example 10.2. This leads to the intriguing conclusion that the effective Lagrangian (10.16) is completely fixed by the set of low-energy constants κ_{ab} , spanning a symmetric invariant tensor

of H , and certain linear representation of H . No specific information about the full symmetry group G is needed, as was first emphasized in [10].

Example 10.4

The requirement that G be compact and semisimple is essential. Contrast the coset spaces $\text{SO}(n+1)/\text{SO}(n) \simeq S^n$ and $\text{ISO}(n)/\text{SO}(n) \simeq \mathbb{R}^n$. Both of these are symmetric, share the same H , and in both the set of broken generators transforms as a vector of $\text{SO}(n)$. Yet, the corresponding EFTs are different; $\text{SO}(n+1)$ is compact but $\text{ISO}(n)$ is not. In case of $\text{SO}(n+1)/\text{SO}(n)$, the low-energy EFT is most easily presented as a nonlinear sigma model for a linearly-transforming unit vector field, $\mathbf{n} \in S^n$. It implies nontrivial interactions among the n NG bosons. On the other hand, in case of $\text{ISO}(n)/\text{SO}(n)$, the leading-order EFT (10.16) reduces to a noninteracting theory, $\mathcal{L}_{\text{eff}} = (1/2)\delta_{ab}\partial_\mu\pi^a\partial^\mu\pi^b$. This follows from the fact that the broken generators of $\text{ISO}(n)$ commute with each other so that $f_{bc}^\alpha = 0$. The curvature tensor is trivial. After all, $\text{ISO}(n)/\text{SO}(n)$ is just the flat Euclidean space.

10.3 Beyond Adler Zero

Until now, we have been solely preoccupied with the question whether or not a given scattering amplitude vanishes in the soft limit for a chosen particle. This amounts to restricting the amplitude to a special hypersurface in the space of all allowed (on-shell and energy–momentum-conserving) kinematical configurations. There are however other interesting kinematical regimes that also probe the algebraic and geometric structure of the underlying EFT. One simple possibility is to take a consecutive soft limit for two particles. Here the geometric soft theorem (10.27) proves invaluable, giving immediately

$$\lim_{p_{n-1} \rightarrow 0} \lim_{p_n \rightarrow 0} \mathcal{A}_{a_1 \dots a_n}(p_1, \dots, p_n) = \hat{\nabla}_{a_n} \hat{\nabla}_{a_{n-1}} \mathcal{A}_{a_1 \dots a_{n-2}}(p_1, \dots, p_{n-2}) . \quad (10.45)$$

Of particular interest is the commutator of the two limits, probing the extent to which the limits interfere with each other,

$$\begin{aligned} & \left[\lim_{p_{n-1} \rightarrow 0}, \lim_{p_n \rightarrow 0} \right] \mathcal{A}_{a_1 \dots a_n}(p_1, \dots, p_n) \\ &= \sum_{i=1}^{n-2} R_{a_i a_{n-1} a_n}^b \mathcal{A}_{a_1 \dots a_{i-1} b a_{i+1} \dots a_{n-2}}(p_1, \dots, p_{n-2}) , \end{aligned} \quad (10.46)$$

where I used (A.79). Another possibility is to take a simultaneous soft limit for two or even more particles. This may lead to a kinematic singularity even in the absence of cubic interaction vertices. See [8, 11] for a precise formulation of such a simultaneous double-soft theorem and further details.

Here I will remain within the confines of the single soft limit, but look more closely at the asymptotic behavior of the scattering amplitude near the limit. I will promote the energy-momenta p_i^μ in an n -particle process to functions $\hat{p}_i^\mu(z)$ of a scaling parameter $z \in \mathbb{C}$ such that $\hat{p}_i^\mu(1) = p_i^\mu$ for all i and moreover $\hat{p}_n^\mu(z) = zp_n^\mu$. The other functions $\hat{p}_i^\mu(z)$ with $i = 1, \dots, n-1$ must be chosen so as to preserve overall energy-momentum conservation, $\sum_{i=1}^n \hat{p}_i^\mu(z) = 0$, and remain on-shell, $[\hat{p}_i(z)]^2 = 0$, for any $z \in \mathbb{C}$. Also, it is required that $\hat{p}_i^\mu(z)$ with $i = 1, \dots, n-1$ have a nonzero limit for $z \rightarrow 0$, that is, only the n -th particle becomes soft for small z .

For a generic kinematical configuration, the complexified tree-level amplitude $\hat{\mathcal{A}}_{a_1 \dots a_n}(\hat{p}_1(z), \dots, \hat{p}_n(z))$ will presumably be analytic in z in some neighborhood of the origin. It can thus be expanded in a series with a nonzero radius of convergence,

$$\hat{\mathcal{A}}_{a_1 \dots a_n}(\hat{p}_1(z), \dots, \hat{p}_{n-1}(z), zp_n) = \sum_{k=0}^{\infty} z^k c_{a_1 \dots a_n}^{(k)} . \quad (10.47)$$

The soft theorems derived in Sects. 10.1 and 10.2 determine the leading term, $c_{a_1 \dots a_n}^{(0)}$, in terms of $(n-1)$ -particle amplitudes with energy-momenta $\hat{p}_1^\mu(0), \dots, \hat{p}_{n-1}^\mu(0)$. The asymptotic behavior of the amplitude at small z is dominated by the first term in (10.47) for which $c_{a_1 \dots a_n}^{(k)}$ is nonzero. In the following, I will refer to the corresponding value of k as the *soft scaling parameter* and use the symbol σ ;

$$\hat{\mathcal{A}}_{a_1 \dots a_n}(\hat{p}_1(z), \dots, \hat{p}_{n-1}(z), zp_n) = z^\sigma c_{a_1 \dots a_n}^{(\sigma)} + \mathcal{O}(z^{\sigma+1}) . \quad (10.48)$$

The notation used in (10.47) may suggest that the coefficients $c_{a_1 \dots a_n}^{(k)}$ are independent of the detailed choice of functions $\hat{p}_i(z)$ beyond the basic requirements laid out above (10.47). I am not aware of any proof of this implicit assumption, or even a study addressing the issue. Still, it is plausible to assume that at least the leading term in (10.48) is independent of such arbitrary choices.

Clearly, $\sigma = 0$ if the soft limit of the amplitude is nontrivial (nonvanishing). The Adler zero amounts to $\sigma \geq 1$. Barring accidental cancellations, we expect a generic EFT satisfying the Adler zero principle to have $\sigma = 1$.

Example 10.5

Recall that for $n = 4$, all the Mandelstam variables s, t, u vanish in the soft limit for any of the four particles. This automatically guarantees Adler zero for the 4-particle amplitude in any derivatively coupled EFT of massless scalars. Moreover, in EFTs of a single species of massless scalar, permutation invariance allows only one linear function of Mandelstam variables, $s + t + u$, which is identically zero. Hence, in EFTs of a single NG boson, the 4-particle amplitude automatically satisfies $\sigma \geq 2$.

In case $\sigma \geq 2$ for *all* (tree-level) scattering amplitudes in a given EFT, we say that the soft limit of the amplitudes is *enhanced*. The rest of the present section is devoted to two immediate questions. Are there, in fact, any theories where scattering amplitudes feature an enhanced soft limit? If yes, is there any generic mechanism that makes the soft limit enhanced beyond the simple Adler zero?

10.3.1 Dirac–Born–Infeld Theory

The cleanest way to answer the first question is to find an explicit example. We will make an educated guess, loosely following [12]. Consider the class of EFTs of a single real NG field, π . We already know that $\sigma \geq 2$ for the 4-particle amplitude. In fact, the constraint $s + t + u = 0$ leaves us with a single candidate, algebraically independent permutation-invariant 4-particle amplitude with $\sigma = 2$, namely $\mathcal{A}(p_1, p_2, p_3, p_4) = (s^2 + t^2 + u^2)/(4\Lambda_D)$. Here Λ_D is a nonzero constant of mass dimension D , the dimension of spacetime; the factor of 4 is a mere convention. This amplitude can be reproduced by a local quartic interaction operator, $-(\partial_\mu \pi)^2/(8\Lambda_D)$. This interaction in turn gives nontrivial amplitudes also for any even number of particles higher than four.

The problem is that already the next, $n = 6$ amplitude fails to satisfy the desired scaling with $\sigma = 2$. By dimensional analysis, the $n = 6$ amplitude is a homogeneous function of degree 6 in energy–momenta. A detailed computation shows that the leading, $c^{(1)}$ term in its expansion (10.47) can be canceled by adding a new interaction operator carrying six derivatives, $[(\partial_\mu \pi)^2]^3/(16\Lambda_D)$. This pattern extends inductively to all higher orders. For any n , the $\sigma = 2$ scaling of the (tree-level) $(2n)$ -particle amplitude can be rescued by adding an interaction operator $[(\partial_\mu \pi)^2]^n$ and tuning its coefficient to a unique value. This ensures cancellation of the $c^{(1)}$ contribution to the $(2n)$ -particle amplitude, generated by exchange diagrams with lower-point interaction vertices. Altogether, we end up with a unique theory modulo the choice of Λ_D ; all the iteratively constructed interaction operators fold neatly into a closed-form Lagrangian,

$$\mathcal{L}_{\text{DBI}} = \Lambda_D \sqrt{1 + (\partial_\mu \pi)^2/\Lambda_D} . \quad (10.49)$$

This is the so-called *Dirac–Born–Infeld* (DBI) *theory*.

A combination of (10.49) with the volume element, $d^D x \sqrt{1 + (\partial_\mu \pi)^2/\Lambda_D}$, admits a remarkable geometric interpretation. This is the induced volume measure on a D -dimensional hypersurface (“brane”), embedded in a flat $(D + 1)$ -dimensional spacetime and parameterized by the Minkowski coordinates x^μ . The NG field $\pi(x)$ arises from spontaneous breaking of the symmetry under translations in the extra dimension by the brane. Properly normalized, the displacement of the brane along the extra dimension is $\pi(x)/\sqrt{|\Lambda_D|}$. The sign of Λ_D determines the signature of the metric in the extra dimension. For positive Λ_D , the spacetime Lorentz group $\text{SO}(d, 1)$ is thus extended to $\text{SO}(d, 2)$, whereas for negative Λ_D it is extended to

$\text{SO}(d+1, 1)$. Up to an arbitrary choice of scale, we end up with two distinct DBI theories, differing by the geometry in the extra dimension.

The fact that the DBI Lagrangian (10.49) does not contain any fields without derivatives automatically guarantees the Adler zero. The realization of the $\sigma = 2$ enhancement is however highly nontrivial. At first sight, it is not even obvious why the order-by-order cancellations required to make the soft limit enhanced should be possible at all. In Chap. 11, I will show that this is a consequence of the extended symmetry of the DBI theory. Let us therefore have a closer look at this symmetry. Under the translation in the extra dimension, the NG field shifts simply as $\pi(x) \rightarrow \pi'(x) = \pi(x) + \epsilon$. What is more interesting are the rotations connecting the D physical dimensions to the extra dimension. These can be parameterized by a Lorentz vector ϵ^μ and their infinitesimal form reads

$$\delta\pi(x) = \sqrt{|\Lambda_D|} \epsilon_\mu x^\mu, \quad \delta x^\mu = -\frac{\text{sgn } \Lambda_D}{\sqrt{|\Lambda_D|}} \epsilon^\mu \pi(x). \quad (10.50)$$

Note that this is not a spacetime symmetry by the definition given in Sect. 4.1, since δx^μ depends on the field $\pi(x)$. Rather, (10.50) is an example of a generalized local transformation. Its evolutionary form is

$$\pi'(x) - \pi(x) = \sqrt{|\Lambda_D|} \epsilon_\mu x^\mu + \frac{\text{sgn } \Lambda_D}{\sqrt{|\Lambda_D|}} \epsilon^\mu \pi(x) \partial_\mu \pi(x), \quad (10.51)$$

which depends on both $\pi(x)$ and its derivative. It is easy to verify that (10.49) is quasi-invariant under (10.51).

10.3.2 Galileon and Special Galileon Theory

Operationally, the form of the DBI theory is so strongly constrained because its Lagrangian only depends on the first derivatives of the NG field. One might hope to find more examples of theories with scattering amplitudes enhanced beyond the plain Adler zero by allowing for higher derivatives. Indeed, one can realize $\sigma = 2$ trivially by setting

$$\mathcal{L}_{\text{eff}} = \frac{1}{2} (\partial_\mu \pi)^2 + \mathcal{L}_{\text{int}}(\partial\partial\pi), \quad (10.52)$$

where \mathcal{L}_{int} is an arbitrary function of second derivatives of $\pi(x)$ whose Taylor expansion in π starts at the fourth or higher order. This, while not very interesting per se, points in the right direction. Namely, the class of theories (10.52) is manifestly invariant under the so-called *Galileon* symmetry,

$$\delta\pi(x) = \epsilon + \epsilon_\mu x^\mu. \quad (10.53)$$

The interaction Lagrangian is strictly invariant, whereas the kinetic term is quasi-invariant. Mere invariance under the constant shift with scalar parameter ϵ , together

with the absence of cubic interaction vertices, already guarantees Adler zero. However, it is ultimately the full Galileon symmetry (10.53) that makes the soft limit of the scattering amplitudes enhanced; I will again return to this in Chap. 11. It is now natural to ask whether there might be any interactions that, similarly to the kinetic term, are merely quasi-invariant under (10.53). If present, such interactions would necessarily contain less than two derivatives per field, and thus realize a soft limit with $\sigma = 2$ in a nontrivial manner. Moreover, by power counting, they would dominate over the interactions of type $\mathcal{L}_{\text{int}}(\partial\partial\pi)$ in (10.52).

Finding all quasi-invariant Lagrangians requires an inspection of the relative Lie algebra cohomology of the spontaneously broken symmetry, as observed in a special case in Sect. 8.1. In case of the Galileon symmetry, a detailed analysis shows [13] that in D spacetime dimensions, there are $D+1$ algebraically independent quasi-invariant Galileon Lagrangians,

$$\mathcal{L}_{\text{Gal}}^{(n)} \equiv \varepsilon^{\mu_1 \dots \mu_n \lambda_{n+1} \dots \lambda_D} \varepsilon^{\nu_1 \dots \nu_n} \pi(\partial_{\mu_1} \partial_{\nu_1} \pi) \cdots (\partial_{\mu_n} \partial_{\nu_n} \pi), \quad (10.54)$$

where $n = 0, \dots, D$. The quasi-invariance of all $\mathcal{L}_{\text{Gal}}^{(n)}$ under the Galileon symmetry (10.53) is manifest. The first of these, $\mathcal{L}_{\text{Gal}}^{(0)} = (-1)^{D-1} D! \pi$, is a mere tadpole. This should be discarded in order for the EFT to be perturbatively well-defined. The next Galileon Lagrangian, $\mathcal{L}_{\text{Gal}}^{(1)} = (-1)^{D-1} (D-1)! \pi \square \pi$, is just the kinetic term up to a rescaling and integration by parts. Each of the interaction Lagrangians $\mathcal{L}_{\text{Gal}}^{(n)}$ with $n \geq 2$, or any linear combination thereof, realizes the $\sigma = 2$ enhanced soft limit of scattering amplitudes. One might be concerned about the presence of the cubic operator, $\mathcal{L}_{\text{Gal}}^{(2)}$. This is however harmless. Namely, within the class of Lagrangians

$$\mathcal{L}_{\text{Gal}} = \sum_{n=0}^D c_n \mathcal{L}_{\text{Gal}}^{(n)}, \quad (10.55)$$

the couplings c_n can be recalibrated by a one-parameter group of transformations, known as the *Galileon duality* [14, 15]. This amounts to a field redefinition, and thus does not affect the S -matrix of the theory. It can always be used to set c_2 to zero.

The 4-particle amplitude in the Galileon theory (10.55) is a homogeneous function of degree 6 in the particle energy–momenta. Permutation invariance together with the constraint $s + t + u = 0$ implies that the amplitude equals $s^3 + t^3 + u^3$ up to overall normalization. Hence, in Galileon theories, the 4-particle amplitude is doubly enhanced, $\sigma = 3$. Is it possible to adjust the values of c_n in (10.55) so as to maintain $\sigma = 3$ for all multiparticle amplitudes? The answer to this question is positive, and the resulting one-parameter family of theories is dubbed *special Galileon*.

The further enhancement of the soft limit of scattering amplitudes is associated with yet another, hidden symmetry [16]. Consider the infinitesimal transformation

$$\delta_\epsilon \pi(x) \equiv \epsilon^{\mu\nu} \left[x_\mu x_\nu + \frac{1}{\Lambda_{D+2}} \partial_\mu \pi(x) \partial_\nu \pi(x) \right], \quad (10.56)$$

where Λ_{D+2} is a (nonzero) constant with mass dimension $D+2$ and $\epsilon^{\mu\nu}$ is a traceless symmetric tensor of parameters. This is another nontrivial example of a generalized local transformation. To check whether (10.56) leaves the Lagrangian (10.55) quasi-invariant, it is sufficient to inspect the variation of the shift of the corresponding action, $\delta_\epsilon S_{\text{Gal}}$, with respect to π . A direct calculation gives

$$\frac{\delta(\delta_\epsilon S_{\text{Gal}})}{\delta\pi} = 2 \sum_{n=0}^D c_n \left[n(n+1)X_n + \frac{1}{\Lambda_{D+2}}(D-n)(D-n-1)X_{n+2} \right], \quad (10.57)$$

$$X_n \equiv \epsilon^{\mu_1 \dots \mu_n \lambda_{n+1} \dots \lambda_D} \epsilon^{\nu_1 \dots \nu_n}_{\lambda_{n+1} \dots \lambda_D} \epsilon_{\mu_n \nu_n} \prod_{i=1}^{n-1} (\partial_{\mu_i} \partial_{\nu_i} \pi);$$

a reader wishing to verify this might want to take advantage of the auxiliary identity

$$\begin{aligned} & \epsilon^{\mu_1 \dots \mu_{n+2} \lambda_{n+3} \dots \lambda_D} \epsilon^{\nu_1 \dots \nu_{n+2}}_{\lambda_{n+3} \dots \lambda_D} \epsilon_{\mu_{n+2} \nu_{n+2}} \prod_{i=1}^{n+1} (\partial_{\mu_i} \partial_{\nu_i} \pi) \\ &= \frac{n+1}{(D-n)(D-n-1)} \epsilon^{\mu_1 \dots \mu_n \lambda_{n+1} \dots \lambda_D} \epsilon^{\nu_1 \dots \nu_n}_{\lambda_{n+1} \dots \lambda_D} \\ & \quad \times \epsilon^{\alpha\beta} \left[n(\partial_\alpha \partial_{\mu_n} \pi)(\partial_\beta \partial_{\nu_n} \pi) \prod_{i=1}^{n-1} (\partial_{\mu_i} \partial_{\nu_i} \pi) - (\partial_\alpha \partial_\beta \pi) \prod_{i=1}^n (\partial_{\mu_i} \partial_{\nu_i} \pi) \right]. \end{aligned} \quad (10.58)$$

The quasi-invariance of the Lagrangian (10.55) is now equivalent to the vanishing of (10.57), which leads to a recurrence relation for the couplings c_n , solved by

$$c_{2n} = \frac{(-1)^n}{\Lambda_{D+2}^n} \binom{D}{2n} \frac{c_0}{2n+1}, \quad c_{2n+1} = \frac{(-1)^n}{\Lambda_{D+2}^n} \frac{1}{D} \binom{D}{2n+1} \frac{c_1}{n+1}. \quad (10.59)$$

What we have here are in principle two independent types of theories. However, the even set c_{2n} inevitably includes the tadpole operator $\mathcal{L}_{\text{Gal}}^{(0)}$ that we have discarded on physical grounds. We are thus left only with the odd terms c_{2n+1} , that is those Galileon operators (10.54) that are even in the NG field π . The parameter c_1 is fixed by demanding proper normalization of the kinetic term. The only genuine parameter of the special Galileon theory is therefore the scale Λ_{D+2} that enters the symmetry transformation (10.56). Specifically in $D=4$ spacetime dimensions, we have $c_3 = -c_1/(2\Lambda_6)$ as the sole effective coupling. Upon some integration by parts, the Lagrangian of the special Galileon theory can then be cast as

$$\mathcal{L}_{\text{Gal}}^{D=4} \simeq \frac{1}{2}(\partial_\mu \pi)^2 - \frac{1}{12\Lambda_6}(\partial_\mu \pi)^2 [(\Box \pi)^2 - (\partial_\nu \partial_\lambda \pi)^2]. \quad (10.60)$$

See [17] for a recent overview of the many intriguing properties of the special Galileon theory.

So far we have discovered three theories that realize in a nontrivial manner an enhanced soft limit of scattering amplitudes: DBI, Galileon and special Galileon. Remarkably, this is it as far as relativistic EFTs of a single NG boson are concerned [18].

Mapping the landscape of theories with $\sigma \geq 2$ and multiple NG bosons remains, as of writing this book, an open problem. Below, I outline a symmetry-based approach that makes the identification of candidate multiflavor EFTs in principle straightforward. In Sect. 10.4, I will then introduce an entirely different method, which utilizes solely known infrared properties of on-shell amplitudes and general quantum-field-theoretic principles. This approach is very efficient in proving that other nontrivial soft behavior than that observed in the above three theories is not possible.

10.3.3 Effective Theories with Enhanced Soft Limit from Symmetry

I have already hinted, so far without proof, that the enhanced soft limit of scattering amplitudes in the DBI and Galileon theories is a consequence of symmetry. The additional symmetry cannot be internal, for that would lead to mere linear constraints on scattering amplitudes or, even worse, additional NG bosons if spontaneously broken. Indeed, a glance at (10.50) and (10.53) shows that we are dealing with transformations with nontrivial dependence on spacetime coordinates, parameterized by a Lorentz vector. Guided by this observation, I will now show that the presence of such symmetries is severely constrained by group-theoretical consistency requirements. The discussion in this subsection is a special case of a framework developed in [19].

Consider an EFT for a single species of NG bosons. The symmetry of this EFT is generated by the operators of angular momentum $J_{\mu\nu}$ and energy–momentum P_μ , and a spontaneously broken scalar generator Q . The commutation relations between $J_{\mu\nu}$ and P_μ are fixed by the Poincaré algebra. Furthermore, $[J_{\mu\nu}, Q] = 0$ expresses the fact that Q is a scalar, whereas $[P_\mu, Q] = 0$ is needed for Q to generate an internal symmetry. Let us now add a Lorentz-vector operator K_μ , generating the tentative hidden symmetry responsible for enhancement of the soft limit of scattering amplitudes. Lorentz invariance restricts the commutation relations between K_μ and the other generators to³

$$\begin{aligned} [J_{\mu\nu}, K_\lambda] &= i(g_{\nu\lambda}K_\mu - g_{\mu\lambda}K_\nu), & [P_\mu, K_\nu] &= i(ag_{\mu\nu}Q + bJ_{\mu\nu}), \\ [K_\mu, K_\nu] &= icJ_{\mu\nu}, & [K_\mu, Q] &= i(dP_\mu + eK_\mu). \end{aligned} \quad (10.61)$$

The parameters a, b, c, d, e cannot take arbitrary values, since to correspond to a well-defined Lie algebra, the commutators must satisfy a set of Jacobi identities. A straightforward calculation leads to the algebraically independent constraints $ae = 0$, $b = 0$ and $c = -ad$. Should $[P_\mu, K_\nu]$ be nonzero, which is necessary for the symmetry generated by K_μ to affect the momentum dependence of scattering amplitudes, a must be nonzero as well. Upon absorbing it into a redefinition of Q via $Q \rightarrow Q/a$, the three commutators in (10.61) not fixed by Lorentz invariance reduce to

³ Additional terms containing the LC tensor can be added to some of the commutators in $D = 3$ or $D = 4$ spacetime dimensions. These however turn out to be irrelevant in the sense that they are either forbidden by Jacobi identities or can be absorbed into a redefinition of the generators.

$$[P_\mu, K_\nu] = i g_{\mu\nu} Q, \quad [K_\mu, K_\nu] = -i \nu J_{\mu\nu}, \quad [K_\mu, Q] = i \nu P_\mu, \quad (10.62)$$

with the shorthand notation $\nu \equiv ad$.

The case of $\nu = 0$ describes the Lie algebra of Galileon transformations (10.53). Here Q generates the constant shift with parameter ϵ , whereas K_μ generates the linear shift with parameter ϵ_μ . It remains to address the case $\nu \neq 0$. Here the parameter can be nearly eliminated by rescaling both Q and K_μ by $\sqrt{|\nu|}$. What is left of ν is just its sign, entering through $[K_\mu, K_\nu] = \mp i J_{\mu\nu}$ and $[K_\mu, Q] = \pm i P_\mu$. Finally, merge K_μ with $J_{\mu\nu}$ and Q with P_μ by identifying $J_{\mu, D+1} \equiv K_\mu$ and $P_{D+1} \equiv Q$ together with $g_{D+1, D+1} \equiv \pm 1$ and voilà, we find the $(D+1)$ -dimensional Poincaré algebra. Its $\nu = +1$ version is based on the $SO(d, 2)$ group of spacetime rotations, whereas the $\nu = -1$ version is based on $SO(d+1, 1)$. These correspond precisely to the two mutations of the DBI theory.

We have successfully recovered the symmetries of the DBI and Galileon theories by a simple Lie-algebraic argument. The utility of the symmetry-based approach of course does not end here. It can be pursued further towards an explicit construction of effective actions, respecting the symmetry. This however requires a generalization of the techniques developed in Chaps. 7 and 8 to coordinate-dependent symmetries.

The classification of candidate symmetries leading to enhancement of the soft limit of scattering amplitudes, outlined here, can be extended in various directions. First, we may want to see how the special Galileon theory fits in. As suggested by (10.56), this requires adding another set of generators that span a traceless symmetric tensor representation of the Lorentz group. While the derivation of all the Lie-algebraic constraints is now more laborious, one eventually recovers (10.56) along with (10.60) as the only physically relevant solution [19]. A similar reasoning was used in [20] to show that in $D = 4$ dimensions, there are no EFTs of a single NG boson that would realize $\sigma \geq 4$ enhancement of the soft limit in a nontrivial manner. Finally, the Lie algebra (10.61) can be extended to allow for multiple flavors of NG bosons. This turns out to be an efficient tool for carving out the landscape of potentially interesting EFTs [21]. The DBI and Galileon theories have a natural multiflavor generalization. However, the precise relation between the symmetry and the asymptotic behavior of scattering amplitudes in the soft limit that would allow us to unambiguously predict the value of σ , remains unknown.

10.4 Soft Recursion

No primer on scattering of NG bosons can be complete without at least briefly mentioning on-shell methods for scattering amplitudes, developed in the last decades. Given the context of this book, I will restrict the discussion to on-shell recursion relations for EFTs of NG bosons. These emerged as an adaptation of the framework, originally designed for the Yang–Mills theory by Britto, Cachazo, Feng and Witten. An interested reader will find an introduction to the latter in Chap. 3 of [4].

We return to the broad class of EFTs for multiple NG boson flavors. For notation simplicity, I will however suppress flavor and momentum labels and denote the on-shell tree-level amplitude for an n -particle scattering process simply as \mathcal{A}_n . Suppose now that we are able to complexify the energy–momenta in the process, $p_i^\mu \rightarrow \hat{p}_i^\mu(z)$, so that all the $\hat{p}_i^\mu(z)$ are linear functions of $z \in \mathbb{C}$, satisfying $\hat{p}_i^\mu(0) = p_i^\mu$. We of course still require that the $\hat{p}_i^\mu(z)$ add up to zero and remain on-shell, $[\hat{p}_i(z)]^2 = 0$, for any $z \in \mathbb{C}$. Then the scattering amplitude \mathcal{A}_n is complexified to a meromorphic function $\hat{\mathcal{A}}_n(z)$ in the complex plane. Furthermore, for any holomorphic function $F_n(z)$ such that $F_n(0) = 1$, the function $\hat{\mathcal{A}}_n(z)/[zF_n(z)]$ is also meromorphic and has a simple pole at $z = 0$ with residue $\hat{\mathcal{A}}_n(0) = \mathcal{A}_n$. The residue theorem then implies that upon integration along an infinitesimal circle enclosing the origin,

$$\mathcal{A}_n = \frac{1}{2\pi i} \oint dz \frac{\hat{\mathcal{A}}_n(z)}{zF_n(z)}. \quad (10.63)$$

By the same residue theorem, the integral can also be expressed in terms of residues at the other poles in the complex plane, possibly including the residue at infinity. This is the central idea behind on-shell recursion techniques. To progress further, we need to be more specific about the complexification of the energy–momentum variables. We also need input on the asymptotic behavior of the amplitude at $z \rightarrow \infty$. Together, these will allow us to tailor the function $F_n(z)$ to the EFT at hand.

10.4.1 Complexified Kinematics

Our requirements that $\hat{p}_i^\mu(z)$ be linear in z and reduce to p_i^μ at $z = 0$ are solved by

$$\hat{p}_i^\mu(z) = p_i^\mu + zq_i^\mu \quad (10.64)$$

with arbitrary q_i^μ . However, the on-shell condition $[\hat{p}_i(z)]^2 = 0$ is most easily satisfied if q_i^μ is parallel to p_i^μ . I will parameterize such q_i^μ by a real constant c_i ,⁴

$$\hat{p}_i^\mu(z) \equiv p_i^\mu(1 - c_i z). \quad (10.65)$$

This is called the *all-line soft shift* of the energy–momentum variables. The last requirement we have to take care of is overall energy–momentum conservation, which leads to

$$\sum_{i=1}^n c_i p_i^\mu = 0. \quad (10.66)$$

I will always assume a generic kinematical configuration, disregarding accidental linear dependencies among the energy–momenta p_i^μ . This is equivalent to the assumption that p_i^μ as a $D \times n$ matrix has the maximum possible rank consistent

⁴ Within this section, I always indicate summation over particles in the scattering process explicitly. Hence, no summation is implied by the repeated index “ i ” in (10.65).

with energy–momentum conservation, that is $\min(D, n - 1)$. By the rank–nullity theorem, the set of solutions to (10.66) then spans a vector space of dimension $n - \min(D, n - 1) \geq 1$. The one-dimensional subspace that is always guaranteed to exist corresponds to $c_1 = \dots = c_n$. However, it is desirable to have solutions for which all the c_i s are different. This will allow us to probe the single soft limit for the individual particles by tuning $z \rightarrow 1/c_i$. The existence of such solutions requires that $D < n - 1$, or better $n \geq D + 2$.

The all-line soft shift (10.65) cannot access the single soft limit of all scattering amplitudes, and the problem gets worse with increasing the spacetime dimension D . This motivated [18] to introduce alternative prescriptions, combining (10.64) and (10.65) for disjoint subsets of the n particles. To illustrate the idea, I will briefly outline the minimal modification of (10.65), dubbed the *all-but-one-line soft shift*. This applies (10.65) to the first $n - 1$ particles but uses (10.64) for the last one,

$$\hat{p}_i^\mu(z) \equiv p_i^\mu(1 - c_i z) \quad \text{for } i = 1, \dots, n - 1, \quad \hat{p}_n^\mu(z) \equiv p_n^\mu + z q_n^\mu. \quad (10.67)$$

Energy–momentum conservation and the on-shell condition now dictate that

$$q_n^\mu = \sum_{i=1}^{n-1} c_i p_i^\mu, \quad q_n^2 = p_n \cdot q_n = 0. \quad (10.68)$$

The first relation in (10.68) can be viewed as a definition of q_n^μ . The second relation then constitutes two homogeneous constraints on c_1, \dots, c_{n-1} , one linear and one quadratic. This defines an $(n - 3)$ -dimensional hypersurface in the \mathbb{R}^{n-1} space of all c_i . The hypersurface always includes a one-dimensional linear subspace where $c_1 = \dots = c_{n-1}$, as guaranteed by energy–momentum conservation. The existence of other, nontrivial solutions for c_i requires that $n - 3 > 1$, or $n \geq 5$, independently of the spacetime dimension D . We find that the all-but-one-line shift is applicable to a broader set of amplitudes than the all-line shift for any $D \geq 4$. The price to pay is that it only allows us to access the soft limit for $n - 1$ particles out of n .

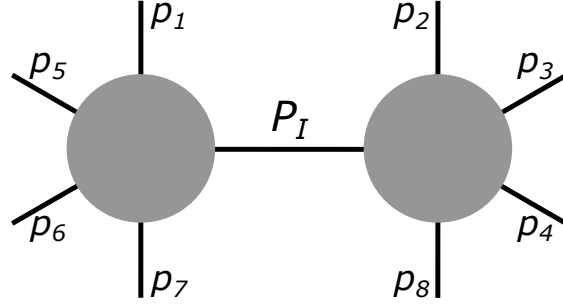
10.4.2 Recursion Relation for On-Shell Amplitudes

We are now ready to deal with the complexified amplitudes. Consider an EFT with soft scaling parameter σ . By definition, any complexified n -particle amplitude $\hat{\mathcal{A}}_n(z)$ in this theory has a zero of order σ at $z = 1/c_i$ for any of the particles to which (10.65) has been applied. It is then advantageous to set

$$F_{n,\sigma}(z) \equiv \prod_{i=1}^{n_s} (1 - c_i z)^\sigma, \quad (10.69)$$

where $n_s = n$ for the all-line shift and $n_s = n - 1$ for the all-but-one-line shift. This choice of $F_n(z)$ is optimized for maximum suppression of the integrand in (10.63)

Fig. 10.3 Example of a factorization channel in an 8-particle scattering process. The two sides of the factorization channel correspond respectively to $I = \{1, 5, 6, 7\}$ and $\tilde{I} = \{2, 3, 4, 8\}$. The gray disks represent the sum of all possible Feynman diagrams with the given external legs



at large z without introducing any new poles in it. This suppression is sufficient to eliminate the pole at infinity provided

$$m < n_s \sigma, \quad (10.70)$$

where m is the degree of \mathcal{A}_n as a function of energy–momenta.

Example 10.6

The constraint (10.70) is satisfied for a large class of EFTs. First, any EFT of the type (10.16), defined on a symmetric coset space, has $(m, \sigma) = (2, 1)$ for any n . This follows from the relation $V = I + 1$ between the numbers of interaction vertices V and internal propagators I , valid for any connected tree-level Feynman diagram. Second, the Lagrangian of the DBI theory (10.49) contains one derivative on each field. As a result, the negative powers of energy–momentum due to propagators are exactly canceled by the adjacent interaction vertices. What is left is in effect one derivative per each external leg of the Feynman diagram, hence $(m, \sigma) = (n, 2)$ for the DBI theory. Finally, the special Galileon theory in $D = 4$ dimensions contains a single quartic interaction vertex with six derivatives. This gives $m = 6V - 2I$ and $n = 4V - 2I$, hence $(m, \sigma) = (2n - 2, 3)$.

With no pole at infinity and no new poles from $F_{n,\sigma}(z)$, all the $z \neq 0$ poles of the integrand in (10.63) come from the propagators inside $\hat{\mathcal{A}}_n(z)$. Consider a partition of all the particles in the scattering process into two disjoint subsets, I and \tilde{I} . Among all diagrams contributing to the tree-level amplitude $\hat{\mathcal{A}}_n(z)$, there is a subset of graphs where the particles in I and \tilde{I} can be separated by cutting a single propagator. This subset of diagrams defines a *factorization channel*, which I will label simply with the letter I ; see Fig. 10.3 for illustration. One can always assign the labels I, \tilde{I} so that the n -th particle belongs to \tilde{I} . Then c_i is well-defined for all particles in I even when the all-but-one-line shift is used. With the shorthand notation

$$P_I^\mu \equiv \sum_{i \in I} p_i^\mu, \quad Q_I^\mu \equiv \sum_{i \in I} c_i p_i^\mu, \quad (10.71)$$

the complex energy–momentum in the propagator separating I and \tilde{I} is $P_I^\mu - zQ_I^\mu$. This leads to two poles at $z = z_I^\pm$ where

$$z_I^\pm = \frac{1}{Q_I^2} \left[P_I \cdot Q_I \pm \sqrt{(P_I \cdot Q_I)^2 - P_I^2 Q_I^2} \right]. \quad (10.72)$$

Near these poles, the complex amplitude $\hat{\mathcal{A}}_n(z)$ factorizes as

$$\hat{\mathcal{A}}_n(z) \xrightarrow{z \rightarrow z_I^\pm} \frac{\hat{\mathcal{A}}_I(z_I^\pm) \hat{\mathcal{A}}_{\tilde{I}}(z_I^\pm)}{(P_I - zQ_I)^2} = \frac{z_I^+ z_I^-}{(z - z_I^+)(z - z_I^-)} \frac{\hat{\mathcal{A}}_I(z_I^+) \hat{\mathcal{A}}_{\tilde{I}}(z_I^+)}{P_I^2}. \quad (10.73)$$

Here $\hat{\mathcal{A}}_I(z_I^\pm)$ is the complexified *on-shell* amplitude for scattering of particles in I , and likewise for $\hat{\mathcal{A}}_{\tilde{I}}(z_I^\pm)$. We arrive at the key result that the n -particle amplitude \mathcal{A}_n can be reconstructed from on-shell amplitudes with lower numbers of particles, as long as it can be complexified using one of the prescriptions introduced in Sect. 10.4.1. Importantly, this *soft recursion* applies to the entire amplitudes. This observation lies at the heart of an approach to symbolic computation of scattering amplitudes that avoids the combinatorial explosion of ordinary perturbation theory. In practice, one only needs to calculate explicitly a small number of “seed” amplitudes. All higher-point amplitudes can then be reconstructed recursively.

Example 10.7

In $D = 4$ spacetime dimensions, the all-line shift (10.65) can be used for any amplitude with $n \geq 6$. This leaves us with \mathcal{A}_4 and \mathcal{A}_5 as the seeds that need to be known a priori. For theories with just even interaction vertices, such as those from Example 10.6, one only needs to calculate the 4-particle amplitude explicitly. This is consistent with what we learned previously about the structure of EFTs on symmetric coset spaces. Namely, it follows from (10.42) that the quartic interaction vertex already fixes uniquely the entire effective Lagrangian. The same applies to the DBI theory.

Let us now see the recursion program through to the end. Upon summation over all factorization channels, (10.63) becomes [22]

$$\mathcal{A}_n = \sum_I \frac{1}{P_I^2} \left[\frac{\hat{\mathcal{A}}_I(z_I^+) \hat{\mathcal{A}}_{\tilde{I}}(z_I^+)}{(1 - z_I^+/z_I^-) F_{n,\sigma}(z_I^+)} + \frac{\hat{\mathcal{A}}_I(z_I^-) \hat{\mathcal{A}}_{\tilde{I}}(z_I^-)}{(1 - z_I^-/z_I^+) F_{n,\sigma}(z_I^-)} \right]. \quad (10.74)$$

The sum runs over all partitions of the particles, that is each pair I, \tilde{I} is only counted once. This is our master equation, which allows recursive construction of tree-level scattering amplitudes by symbolic computation. However, the first step of the recursion can often be performed manually. Namely, if the functions $\hat{\mathcal{A}}_I(z), \hat{\mathcal{A}}_{\tilde{I}}(z)$ themselves do not contain any factorization poles, the summand in (10.74) can be interpreted in terms of the residues of the meromorphic func-

tion $\hat{\mathcal{A}}_I(z)\hat{\mathcal{A}}_{\tilde{I}}(z)/[zF_{n,\sigma}(z)(P_I - zQ_I)^2]$. This function does have also poles at $z = 1/c_i$, since unlike $\hat{\mathcal{A}}_n(z)$ itself, $\hat{\mathcal{A}}_I(z)$ and $\hat{\mathcal{A}}_{\tilde{I}}(z)$ are not on-shell and thus do not necessarily vanish at these points. We can then use the residue theorem to rewrite the sum over poles at z_I^\pm in terms of a sum over poles at $z = 0$ and $z = 1/c_i$,

$$\mathcal{A}_n = \sum_I \left[\frac{\hat{\mathcal{A}}_I(0)\hat{\mathcal{A}}_{\tilde{I}}(0)}{P_I^2} + \sum_{i=1}^{n_s} \text{Res}_{z=1/c_i} \frac{\hat{\mathcal{A}}_I(z)\hat{\mathcal{A}}_{\tilde{I}}(z)}{zF_{n,\sigma}(z)(P_I - zQ_I)^2} \right]. \quad (10.75)$$

The first term takes into account diagrammatic contributions to \mathcal{A}_n with an internal propagator. The second term must therefore match the contributions to \mathcal{A}_n from contact n -point operators in the effective Lagrangian.

Example 10.8

The special Galileon theory (10.60) in $D = 4$ dimensions has a single, quartic interaction vertex. When applied to the 6-particle amplitude, we therefore expect the second term in (10.75) to vanish. To check this, rewrite the Lagrangian as

$$\mathcal{L}_{\text{sGal}}^{D=4} \simeq \frac{1}{2}(\partial_\mu \pi)^2 - \frac{1}{24\Lambda_6} \varepsilon^{\mu_1 \mu_2 \mu_3 \lambda} \varepsilon^{\nu_1 \nu_2 \nu_3} \lambda \pi \prod_{i=1}^3 (\partial_{\mu_i} \partial_{\nu_i} \pi). \quad (10.76)$$

The Feynman rule for the interaction vertex is $[-i/(6\Lambda_6)]G(p_1, p_2, p_3)$, where $p_1^\mu, p_2^\mu, p_3^\mu$ are any three of the four four-momenta in the vertex and G denotes the corresponding Gram determinant. Since the latter is a homogeneous function of all its arguments of degree two, we have

$$\hat{\mathcal{A}}_I(z)\hat{\mathcal{A}}_{\tilde{I}}(z) = \frac{1}{(6\Lambda_6)^2} G(\{p_i\}_{i \in I}) G(\{p_j\}_{j \in \tilde{I}}) \prod_{k=1}^6 (1 - c_k z)^2 \quad (10.77)$$

for any partition of the six particles into two triples I, \tilde{I} . I have implicitly used the all-line shift. Note that the criterion (10.70) is satisfied even if we take $F_{n,\sigma}(z)$ with $\sigma = 2$ instead of $\sigma = 3$. The advantage of this choice is that the denominator term $F_{n,\sigma}(z)$ is then completely canceled by the last factor in (10.77). This indeed makes the second term in (10.75) disappear since the residues at $z = 1/c_i$ trivially vanish.

The derivation of (10.74) does not apply to theories with a nonvanishing soft limit, where $\sigma = 0$ and (10.70) cannot hold. Here we have two alternatives: either use $F_{n,0}(z) = 1$ and deal with the residue at infinity, or use $F_{n,\sigma}(z)$ with $\sigma \geq 1$ and deal with the ensuing poles at $z = 1/c_i$. The latter option is feasible, at least for two-derivative EFTs of the type (10.16), where sufficient suppression of the integrand in (10.63) is ensured by $F_{n,1}(z)$. The residues at the simple poles at $z = 1/c_i$ can then be extracted from the geometric soft theorem (10.27) or its more explicit version (10.31). A minor modification of the steps leading to (10.74) now gives [8]

$$\mathcal{A}_n = \sum_I \frac{1}{p_I^2} \sum_{\pm} \frac{\hat{\mathcal{A}}_I(z_I^{\pm}) \hat{\mathcal{A}}_I(z_I^{\pm})}{(1 - z_I^{\pm}/z_I^{\mp}) F_{n,1}(z_I^{\pm})} + \sum_{i=1}^n \frac{\hat{\mathbf{v}}_{a_i} \hat{\mathcal{A}}_{n-1}(1/c_i)}{\prod_{j \neq i} (1 - c_j/c_i)}. \quad (10.78)$$

Here I used the shorthand notation $\hat{\mathcal{A}}_{n-1}$ for the amplitude where the i -th particle with flavor index a_i has been removed. It is of course still possible to rewrite the first term in (10.78) as (10.75).

10.4.3 Soft Bootstrap

The modern scattering amplitude program strives to construct amplitudes, or even define what a quantum field theory is, without recourse to a Lagrangian or the symmetry thereof. It is therefore of great interest to understand what kind of soft behavior of scattering amplitudes is allowed on general grounds. Recursion relations such as (10.74) or its generalization (10.78) provide a versatile tool for constraining the landscape of possible EFTs.

In [18], soft recursion was used to derive a set of bounds on the soft scaling parameter σ . The most striking result is that in $D \geq 4$ spacetime dimensions, $\sigma = 3$ is the maximum value that can be realized in a nontrivial manner. (Any σ can be realized trivially by interactions with at least σ derivatives per field.) Apart from this universal bound, the maximum achievable value of σ in an EFT with a fixed set of interaction operators is also constrained by the number of derivatives per field. Consider a contact operator of the schematic type $(\partial\pi)^2 \partial^{m-2} \pi^{n-2}$, and characterize the number of derivatives per field by the parameter $\varrho \equiv (m-2)/(n-2)$. We thus have $\varrho = 1$ for the DBI theory and $\varrho = 2$ for all the Galileon interactions (10.54). Then, the soft scaling parameter is bounded by $\sigma \leq \varrho + 1$. This shows that the DBI and special Galileon theories are “exceptional” in the sense that they maximize σ in their respective classes of theories with fixed ϱ . Similar bounds can be derived for translationally and rotationally invariant nonrelativistic EFTs [1].

The values of σ and ϱ alone do not necessarily specify a unique EFT. Here one can gain further insight by a procedure known as the *soft bootstrap*, see for instance [23, 24]. In $D \leq 4$ dimensions, the only seed amplitudes needed to apply soft recursion in combination with the all-line shift are \mathcal{A}_4 and \mathcal{A}_5 . These are contact amplitudes without any exchange contributions. As such, they are given by polynomials of degree m in the particle energy–momenta, or $m/2$ in the Mandelstam variables. These polynomials are restricted by energy–momentum conservation and permutation invariance, and their complete classification is usually straightforward. With a list of candidate seed amplitudes at hand, one then proceeds to construct higher-point amplitudes via (10.74) or (10.78). The result of the recursion must be independent of the choice of parameters c_i satisfying (10.66) or (10.68). Failure to pass this c_i -independence test indicates that the seed amplitudes do not correspond to any well-defined local field theory.

Soft bootstrap cannot, strictly speaking, be used to prove that a theory based on a fixed set of seed amplitudes exists. That would require checking the c_i -independence of recursively constructed amplitudes to all orders of the recursion. In this aspect, soft bootstrap is nicely complemented by more conventional EFT approaches based on symmetry. The latter are efficient in isolating a set of candidate EFTs with desired particle composition and soft behavior, as illustrated in Sect. 10.3.3. What soft bootstrap can do is show that other theories with given soft behavior than those known a priori do not exist.

References

1. M.A. Mojahed, T. Brauner, JHEP **03**, 086 (2022). DOI [10.1007/JHEP03\(2022\)086](https://doi.org/10.1007/JHEP03(2022)086)
2. C. Cheung, M. Derda, A. Helset, J. Parra-Martinez, JHEP **08**, 103 (2023). DOI [10.1007/JHEP08\(2023\)103](https://doi.org/10.1007/JHEP08(2023)103)
3. C. Cheung, [arXiv:1708.03872](https://arxiv.org/abs/1708.03872)
4. H. Elvang, Y.t. Huang, *Scattering Amplitudes in Gauge Theory and Gravity* (Cambridge University Press, Cambridge, UK, 2015)
5. S. Weinberg, *The Quantum Theory of Fields*, vol. II (Cambridge University Press, Cambridge, UK, 1996)
6. S. Weinberg, *The Quantum Theory of Fields*, vol. I (Cambridge University Press, Cambridge, UK, 1995)
7. K. Kampf, J. Novotný, M. Shifman, J. Trnka, Phys. Rev. Lett. **124**, 111601 (2020). DOI [10.1103/PhysRevLett.124.111601](https://doi.org/10.1103/PhysRevLett.124.111601)
8. C. Cheung, A. Helset, J. Parra-Martinez, JHEP **04**, 011 (2022). DOI [10.1007/JHEP04\(2022\)011](https://doi.org/10.1007/JHEP04(2022)011)
9. J.M. Lee, *Introduction to Riemannian Manifolds* (Springer International Publishing, Cham, Switzerland, 2018)
10. I. Low, Phys. Rev. **D91**, 105017 (2015). DOI [10.1103/PhysRevD.91.105017](https://doi.org/10.1103/PhysRevD.91.105017)
11. I. Low, Phys. Rev. **D93**, 045032 (2016). DOI [10.1103/PhysRevD.93.045032](https://doi.org/10.1103/PhysRevD.93.045032)
12. C. Cheung, K. Kampf, J. Novotný, J. Trnka, Phys. Rev. Lett. **114**, 221602 (2015). DOI [10.1103/PhysRevLett.114.221602](https://doi.org/10.1103/PhysRevLett.114.221602)
13. G. Goon, K. Hinterbichler, A. Joyce, M. Trodden, JHEP **06**, 004 (2012). DOI [10.1007/JHEP06\(2012\)004](https://doi.org/10.1007/JHEP06(2012)004)
14. C. de Rham, M. Fasiello, A.J. Tolley, Phys. Lett. **B733**, 46 (2014). DOI [10.1016/j.physletb.2014.03.061](https://doi.org/10.1016/j.physletb.2014.03.061)
15. K. Kampf, J. Novotný, JHEP **10**, 006 (2014). DOI [10.1007/JHEP10\(2014\)006](https://doi.org/10.1007/JHEP10(2014)006)
16. K. Hinterbichler, A. Joyce, Phys. Rev. **D92**, 023503 (2015). DOI [10.1103/PhysRevD.92.023503](https://doi.org/10.1103/PhysRevD.92.023503)
17. F. Přeucil, J. Novotný, JHEP **11**, 166 (2019). DOI [10.1007/JHEP11\(2019\)166](https://doi.org/10.1007/JHEP11(2019)166)
18. C. Cheung, K. Kampf, J. Novotný, C.H. Shen, J. Trnka, JHEP **02**, 020 (2017). DOI [10.1007/JHEP02\(2017\)020](https://doi.org/10.1007/JHEP02(2017)020)
19. M.P. Bogers, T. Brauner, JHEP **05**, 076 (2018). DOI [10.1007/JHEP05\(2018\)076](https://doi.org/10.1007/JHEP05(2018)076)
20. D. Roest, D. Stefanyzyn, P. Werkman, JHEP **08**, 081 (2019). DOI [10.1007/JHEP08\(2019\)081](https://doi.org/10.1007/JHEP08(2019)081)
21. M.P. Bogers, T. Brauner, Phys. Rev. Lett. **121**, 171602 (2018). DOI [10.1103/PhysRevLett.121.171602](https://doi.org/10.1103/PhysRevLett.121.171602)
22. C. Cheung, K. Kampf, J. Novotný, C.H. Shen, J. Trnka, Phys. Rev. Lett. **116**, 041601 (2016). DOI [10.1103/PhysRevLett.116.041601](https://doi.org/10.1103/PhysRevLett.116.041601)
23. H. Elvang, M. Hadjiantonis, C.R.T. Jones, S. Paranjape, JHEP **01**, 195 (2019). DOI [10.1007/JHEP01\(2019\)195](https://doi.org/10.1007/JHEP01(2019)195)
24. I. Low, Z. Yin, JHEP **11**, 078 (2019). DOI [10.1007/JHEP11\(2019\)078](https://doi.org/10.1007/JHEP11(2019)078)

Part IV
Spontaneously Broken
Spacetime Symmetry

Chapter 11

Locally Equivalent Symmetries

Our discussion of *effective field theory* (EFT) for *Nambu–Goldstone* (NG) *bosons* has so far been limited to internal symmetries. The application of the EFT program to symmetries that depend on or affect spacetime coordinates is associated with numerous subtleties. The first of these lies in the identification of the degrees of freedom of the low-energy EFT. We saw already in Chap. 6 that the fluctuations of the order parameter induced by independent broken symmetry generators may be related to each other. Very often, this is a consequence of a redundancy between the localized actions of different symmetry transformations. This chapter explores such local equivalence of coordinate-dependent symmetries as a preparation for the more detailed discussion of broken spacetime symmetries in the following chapters.

The first two sections follow the pedagogical exposition in [1]. In Sect. 11.1, I show that local equivalence of symmetries leads to a linear relation between the corresponding Noether currents. This lifts the intuitive classical picture of redundancy among NG modes, outlined in Chap. 6, to the language of quantum field theory. In Sect. 11.2, I then reinterpret the local equivalence of symmetries in terms of their simultaneous gauging. The whole Sect. 11.3 is reserved for several physically relevant examples, illustrating the general theory of locally equivalent symmetries. Finally, in Sect. 11.4, I return to scattering of NG bosons of internal symmetries. I show how the relations between Noether currents derived in this chapter help explain the enhancement of the soft limit of scattering amplitudes in the *Dirac–Born–Infeld* (DBI) and Galileon theories.

11.1 Relations Between Noether Currents

Consider a theory of a set of (not necessarily scalar) fields ψ^i , and suppose that its action is invariant under two different classes of continuous symmetries. As we know from Chap. 4, the associated Noether currents can be extracted by evaluating the variation of the action under the corresponding localized transformations. The evolutionary form of these transformations can be written generally as

$$\begin{aligned}\delta_1\psi^i(x) &= \epsilon_1^A(x)F_A^i[\psi, x](x) + \partial_\mu\epsilon_1^A(x)K_A^{i\mu}[\psi, x](x) + \cdots, \\ \delta_2\psi^i(x) &= \epsilon_2^{\mathfrak{A}}(x)F_{\mathfrak{A}}^i[\psi, x](x) + \partial_\mu\epsilon_2^{\mathfrak{A}}(x)K_{\mathfrak{A}}^{i\mu}[\psi, x](x) + \cdots.\end{aligned}\quad (11.1)$$

As before, the square bracket notation indicates local functions of the fields and their derivatives, possibly also depending explicitly on the spacetime coordinates. Furthermore, $\epsilon_1^A(x)$ and $\epsilon_2^{\mathfrak{A}}(x)$ are the localized parameters of the transformations. The presence of derivatives of the parameters in (11.1) hints that I allow for a generic localization of the transformations; the ellipses represent terms with higher derivatives of $\epsilon_1^A(x)$ and $\epsilon_2^{\mathfrak{A}}(x)$. The original symmetries are recovered by reducing the coordinate-dependent parameters to constants, $\epsilon_1^A(x) \rightarrow \epsilon_1^A$ and $\epsilon_2^{\mathfrak{A}}(x) \rightarrow \epsilon_2^{\mathfrak{A}}$. I will call the two symmetries *locally equivalent* if there is a set of smooth functions $f_{\mathfrak{A}}^A(x)$ such that setting

$$\epsilon_1^A(x) = f_{\mathfrak{A}}^A(x)\epsilon_2^{\mathfrak{A}}(x) \quad (11.2)$$

makes the two transformations in (11.1) identical, $\delta_1\psi^i(x) = \delta_2\psi^i(x)$.

Locally equivalent symmetry transformations should give the same variation of the action as expressed in terms of the corresponding Noether currents,

$$\begin{aligned}\delta S &= \int d^Dx J_{2\mathfrak{A}}^\mu[\psi, x](x)\partial_\mu\epsilon_2^{\mathfrak{A}}(x) = \int d^Dx J_{1A}^\mu[\psi, x](x)\partial_\mu\epsilon_1^A(x) \\ &= \int d^Dx J_{1A}^\mu[\psi, x](x) \left[f_{\mathfrak{A}}^A(x)\partial_\mu\epsilon_2^{\mathfrak{A}}(x) + \epsilon_2^{\mathfrak{A}}(x)\partial_\mu f_{\mathfrak{A}}^A(x) \right].\end{aligned}\quad (11.3)$$

However, the latter expression does not automatically vanish for constant $\epsilon_2^{\mathfrak{A}}$ as it should. In order to ensure this, we have to impose the *integrability condition*

$$J_{1A}^\mu[\psi, x]\partial_\mu f_{\mathfrak{A}}^A = \partial_\mu N_{\mathfrak{A}}^\mu[\psi, x], \quad (11.4)$$

where $N_{\mathfrak{A}}^\mu[\psi, x]$ is some local vector function of the fields and their derivatives. Combining (11.4) with (11.3) and integrating by parts in the latter then leads to the Noether current *equivalence relation*

$$J_{2\mathfrak{A}}^\mu[\psi, x] = f_{\mathfrak{A}}^A J_{1A}^\mu[\psi, x] - N_{\mathfrak{A}}^\mu[\psi, x]. \quad (11.5)$$

This is the main result of the chapter that deserves a few comments. First, an alternative approach to local equivalence is to start with constant $\epsilon_2^{\mathfrak{A}}$, that is the actual symmetry with generators $Q_{2\mathfrak{A}}$. One then demands the existence of smooth functions $f_{\mathfrak{A}}^A(x)$ such that this symmetry is reproduced by a localized transformation of the type $\delta_1\psi^i$ with parameter $\epsilon_1^A(x) = f_{\mathfrak{A}}^A(x)\epsilon_2^{\mathfrak{A}}$. This is the origin of the integrability condition (11.4). If needed, the localized transformation $\delta_2\psi^i$ can then be *defined* as $\delta_1\psi^i$ with the parameter $\epsilon_1^A(x)$ as given in (11.2). It is therefore reasonable to think of $\delta_1\psi^i$ as a *parent* symmetry and of $\delta_2\psi^i$ as its *descendant*.

Second, I have not imposed the equation of motion. Thus, both the equivalence relation (11.5) and the integrability condition (11.4) must hold off-shell. By taking a divergence of (11.5) and using (11.4), we obtain another off-shell relation,

$$\partial_\mu J_{2\mathfrak{A}}^\mu[\psi, x] = f_{\mathfrak{A}}^A \partial_\mu J_{1A}^\mu[\psi, x]. \quad (11.6)$$

This makes the Ward identities for the correlation functions of $J_{2\mathfrak{A}}^\mu$ a consequence of those for J_{1A}^μ . By the same token, the two sets of currents couple to the same one-particle NG states. This explains in a quantum-field-theoretic language the redundancy of NG modes generated by locally equivalent symmetries.

Third, should $f_{\mathfrak{A}}^A$ be constant, the integrability condition (11.4) is fulfilled for $N_{\mathfrak{A}}^\mu = 0$. The equivalence relation (11.5) is then strictly linear in the currents. The corresponding generators Q_{1A} and $Q_{2\mathfrak{A}}$ are also linearly related, $Q_{2\mathfrak{A}} = f_{\mathfrak{A}}^A Q_{1A}$. This special case therefore amounts to a mere change of basis of symmetry generators, or a choice of a subset thereof, and is of little interest. A nontrivial current equivalence relation only arises from a set of non-constant functions $f_{\mathfrak{A}}^A(x)$. Without loss of generality, one can drop the constant part of $f_{\mathfrak{A}}^A(x)$ and assume that $f_{\mathfrak{A}}^A(0) = 0$. This assumption is satisfied for all the examples worked out below. It also guarantees a posteriori that the relation (11.2) cannot be inverted. In other words, the distinction between a parent symmetry and its descendant is well-defined and cannot be reversed.

Finally, recall that the Noether current is determined by the symmetry transformation only up to addition of a vector field whose divergence vanishes off-shell. Suppose that we modify the parent Noether current by $J_{1A}^\mu \rightarrow J_{1A}^\mu + \delta J_{1A}^\mu$ such that $\partial_\mu \delta J_{1A}^\mu = 0$. Then the integrability condition (11.4) can still be satisfied if we shift $N_{\mathfrak{A}}^\mu$ appropriately, $N_{\mathfrak{A}}^\mu \rightarrow N_{\mathfrak{A}}^\mu + f_{\mathfrak{A}}^A \delta J_{1A}^\mu$. It follows that the ambiguity of the parent current does not affect the descendant current $J_{2\mathfrak{A}}^\mu$ as given by (11.5). The descendant current of course suffers from its own ambiguity under $J_{2\mathfrak{A}}^\mu \rightarrow J_{2\mathfrak{A}}^\mu + \delta J_{2\mathfrak{A}}^\mu$ such that $\partial_\mu \delta J_{2\mathfrak{A}}^\mu = 0$ off-shell. The ambiguities of the two currents are therefore independent of each other and do not constrain the validity of the equivalence relation (11.5).

11.2 Noether Currents from Background Gauging

The derivation of the Noether current of a given symmetry, based on (4.7), treats the localized symmetry transformation as a mere technical trick. There is however an alternative approach to Noether's theorem that takes the local transformation seriously and promotes it to an actual symmetry of the action. This offers additional insight into the structure of Noether currents, the price to pay being a technical assumption that has to be verified case by case.

This key assumption is that it is possible to make the action of the theory exactly invariant under the localized symmetry transformation by adding a set of background gauge fields A_μ^A . More concretely, we want to promote the original action $S\{\psi\}$ to a functional $\tilde{S}\{\psi, A\}$ such that $\tilde{S}\{\psi, A\}|_{\text{vac}} = S\{\psi\}$, where the subscript “vac” indicates removing the background. I allow for a generic local transformation of ψ^i with parameters $\epsilon^A(x)$, following (11.1),

$$\delta\psi^i(x) = \epsilon^A(x) F_A^i[\psi, x](x) + \partial_\mu \epsilon^A(x) K_A^{i\mu}[\psi, x](x) + \cdots \quad (11.7)$$

As for the local transformation of the gauge field, we can assume a similar generic expansion in derivatives of the parameters $\epsilon^A(x)$,

$$\delta A_\mu^A(x) = \epsilon^B(x) \tilde{F}_{B\mu}^A[A, x](x) + \partial_\nu \epsilon^B(x) \tilde{K}_{B\mu}^{A\nu}[A, x](x) + \cdots ; \quad (11.8)$$

the ellipsis again stands for terms with higher derivatives of $\epsilon^A(x)$. It is important that δA_μ^A does not depend on the dynamical fields ψ^i so that A_μ^A can be treated as a fixed external background.

The variation of action under a simultaneous transformation of ψ^i and A_μ^A consists of contributions from varying the former and the latter, $\delta \tilde{S} \equiv \delta_\psi \tilde{S} + \delta_A \tilde{S}$, where

$$\delta_A \tilde{S} = \int d^D x \frac{\delta \tilde{S}}{\delta A_\mu^A} (\epsilon^B \tilde{F}_{B\mu}^A + \partial_\nu \epsilon^B \tilde{K}_{B\mu}^{A\nu} + \cdots) . \quad (11.9)$$

Next we remove the background. In this limit, $\delta_\psi \tilde{S}$ turns by construction into $\int d^D x J_A^\mu \partial_\mu \epsilon^A$. In order that $\delta \tilde{S}$ vanishes at least when ϵ^A is constant, there must be a local function $R_A^\mu[\psi, x]$ such that

$$\left. \frac{\delta \tilde{S}\{\psi, A\}}{\delta A_\mu^B(x)} \tilde{F}_{A\mu}^B[A, x](x) \right|_{\text{vac}} = \partial_\mu R_A^\mu[\psi, x](x) . \quad (11.10)$$

With this consistency condition satisfied, the vanishing of $\delta \tilde{S}$ for any $\epsilon^A(x)$ leads to the identification

$$J_A^\mu[\psi, x](x) = R_A^\mu[\psi, x](x) - \left. \frac{\delta \tilde{S}\{\psi, A\}}{\delta A_\nu^B(x)} \tilde{K}_{A\nu}^{B\mu}[A, x](x) \right|_{\text{vac}} + \cdots . \quad (11.11)$$

Equations (11.10) and (11.11) already constitute a fairly general tool for derivation of Noether currents. Yet, the approach based on background gauge invariance can be further extended to higher-rank tensor background fields. Instead of cluttering the formalism with additional indices, I will content myself with a couple of illustrative examples, worked out in Sects. 11.3.2 and 11.3.3.

The notation used in (11.10) and (11.11) is tailored to what I called the parent symmetry. What if the system possesses an additional, descendant set of symmetries? By (11.2), the localized transformations of the descendant type form a subset of all localized transformations of the parent type. All we therefore have to do is to gauge the parent symmetry by adding its background gauge fields A_μ^A . That we can in fact do so is the sole technical assumption we have to make. The descendant symmetry is then automatically gauged as well without the need for any additional gauge fields. This once again underlines the redundancy of the descendant symmetry. In particular, the background gauging makes it possible to recover the equivalence relation (11.5) for Noether currents; see [1] for a detailed proof.

11.3 Examples

I will now illustrate the general theory developed above with several examples. These are chosen mostly for their relevance to EFTs discussed elsewhere in the book. I will therefore not shy away from showing certain amount of details.

11.3.1 Galileon Symmetry

Consider a class of theories of a single real relativistic scalar field ϕ , invariant under the Galileon symmetry

$$\delta\phi(x) = \epsilon_1 + \epsilon_2^\mu x_\mu . \quad (11.12)$$

This can be localized by replacing the constant parameters with arbitrary functions, $\epsilon_1 \rightarrow \epsilon_1(x)$ and $\epsilon_2^\mu \rightarrow \epsilon_2^\mu(x)$. The shift by ϵ_1 is the parent symmetry. The descendant symmetry with parameter ϵ_2^μ can be recovered locally by setting $\epsilon_1(x) = \epsilon_2^\mu(x)x_\mu$. This is the local equivalence condition (11.2) where $f_\mu(x) = x_\mu$. The integrability condition (11.4) and the current equivalence relation (11.5) then translate to

$$J_{1\nu}[\phi, x] = \partial_\mu N_\nu^\mu[\phi, x] , \quad J_{2\nu}^\mu[\phi, x](x) = x_\nu J_1^\mu[\phi, x](x) - N_\nu^\mu[\phi, x](x) . \quad (11.13)$$

Interestingly, the set of functions $N_\nu^\mu[\phi, x]$ is in this case the primary object, which by means of (11.13) determines both currents.

Let us see how this works explicitly. Take the class of Lagrangians of the type

$$\mathcal{L} = \sum_{n=0}^D c_n \mathcal{L}_{\text{Gal}}^{(n)} + \mathcal{L}_{\text{int}}(\partial\partial\phi) . \quad (11.14)$$

Here \mathcal{L}_{int} is an arbitrary function of second derivatives of ϕ , which is manifestly invariant under (11.12). Furthermore, $\mathcal{L}_{\text{Gal}}^{(n)}$ with $n = 0, \dots, D$ is the set of quasi-invariant Galileon Lagrangians, cf. Sect. 10.3.2,

$$\mathcal{L}_{\text{Gal}}^{(n)} = \varepsilon^{\mu_1 \dots \mu_n \lambda_{n+1} \dots \lambda_D} \varepsilon^{\nu_1 \dots \nu_n}{}_{\lambda_{n+1} \dots \lambda_D} \phi (\partial_{\mu_1} \partial_{\nu_1} \phi) \dots (\partial_{\mu_n} \partial_{\nu_n} \phi) . \quad (11.15)$$

It is easy to extract the current J_1^μ by evaluating the variation of the action under the shift $\phi(x) \rightarrow \phi(x) + \epsilon_1(x)$. With the shorthand notation for the symmetric tensor

$$X^{\mu\nu} \equiv - \sum_{n=0}^D (n+1) c_n \varepsilon^{\mu\mu_2 \dots \mu_n \lambda_{n+1} \dots \lambda_D} \varepsilon^{\nu\nu_2 \dots \nu_n}{}_{\lambda_{n+1} \dots \lambda_D} \prod_{i=2}^n (\partial_{\mu_i} \partial_{\nu_i} \phi) , \quad (11.16)$$

the current takes the compact form

$$J_1^\mu = X^{\mu\nu} \partial_\nu \phi - \partial_\nu \frac{\partial \mathcal{L}_{\text{int}}}{\partial (\partial_\mu \partial_\nu \phi)} . \quad (11.17)$$

This matches the integrability condition in (11.13) since $X^{\mu\nu}$ obviously has vanishing divergence and so $X^{\mu\nu}\partial_\nu\phi = \partial_\nu(X^{\mu\nu}\phi)$. It follows that

$$N_\nu^\mu = X_\nu^\mu\phi - \frac{\partial\mathcal{L}_{\text{int}}}{\partial(\partial_\mu\partial^\nu\phi)} , \quad (11.18)$$

and the current $J_{2\nu}^\mu$ is then given by the second relation in (11.13). This result can be easily verified by evaluating the variation of the action under $\phi(x) \rightarrow \phi(x) + \epsilon_2^\mu(x)x_\mu$.

Example 11.1

The simplest example of a theory from the class (11.14) is the theory of a free massless scalar, $\mathcal{L} = (1/2)(\partial_\mu\phi)^2$. This amounts to setting $c_1 = (-1)^D/[2(D-1)!]$ and discarding all the other operators contributing to (11.14). By (11.16), this leads to $X^{\mu\nu} = 2c_1 g^{\mu\nu}(-1)^D(D-1)! = g^{\mu\nu}$. We then find in turn

$$J_1^\mu = \partial^\mu\phi , \quad N_\nu^\mu = \delta_\nu^\mu\phi , \quad J_{2\nu}^\mu = x_\nu\partial^\mu\phi - \delta_\nu^\mu\phi . \quad (11.19)$$

To illustrate the background gauging formalism, we need to gauge the parent symmetry. This is most easily done by first integrating the quasi-invariant Galileon terms by parts so that each factor of ϕ carries at least one derivative. All one then has to do is to replace every $\partial_\mu\phi$ with $D_\mu\phi \equiv \partial_\mu\phi - A_\mu$. This gives the gauged action

$$\begin{aligned} \tilde{S}\{\phi, A\} = & - \sum_{n=0}^D c_n \epsilon^{\mu_1 \dots \mu_n \lambda_{n+1} \dots \lambda_D} \epsilon^{\nu_1 \dots \nu_n}{}_{\lambda_{n+1} \dots \lambda_D} \\ & \times \int d^Dx D_{\mu_1}\phi(x) D_{\nu_1}\phi(x) \prod_{i=2}^n [\partial_{\mu_i} D_{\nu_i}\phi(x)] + \int d^Dx \mathcal{L}_{\text{int}}(\partial D\phi)(x) . \end{aligned} \quad (11.20)$$

This is manifestly invariant under the simultaneous gauge transformation $\delta\phi(x) = \epsilon_1(x)$ and $\delta A_\mu(x) = \partial_\mu\epsilon_1(x)$. The latter corresponds to (11.8) with $\tilde{F}_{1\mu} = 0$ and $\tilde{K}_{1\mu}^\nu = \delta_\mu^\nu$; no terms with higher derivatives of $\epsilon_1(x)$ are present. The consistency condition (11.10) is trivially satisfied with $R^\mu = 0$, and (11.11) then gives

$$J_1^\mu[\phi, x](x) = - \left. \frac{\delta\tilde{S}\{\phi, A\}}{\delta A_\mu(x)} \right|_{\text{vac}} . \quad (11.21)$$

It takes little effort to convince oneself that this exactly reproduces (11.17).

To extract the descendant current $J_{2\nu}^\mu$ from the same gauged action, we set $\epsilon_1(x) = \epsilon_2^\mu(x)x_\mu$, which changes the transformation rule for A_μ to

$$\delta A_\mu(x) = \epsilon_{2\mu}(x) + x_\nu\partial_\mu\epsilon_2^\nu(x) . \quad (11.22)$$

This matches (11.8) with $\tilde{F}_{2\mu\nu} = g_{\mu\nu}$ and $\tilde{K}_{2\alpha\mu}^\nu = \delta_\mu^\nu x_\alpha$, and (11.10) then becomes

$$\partial_\mu R_\nu^\mu[\phi, x](x) = g_{\mu\nu} \left. \frac{\delta \tilde{S}\{\phi, A\}}{\delta A_\mu(x)} \right|_{\text{vac}} = -J_{1\nu}[\phi, x](x) . \quad (11.23)$$

With the help of the first relation in (11.13) this is seen to be solved by $R_\nu^\mu = -N_\nu^\mu$. The master equation (11.11) then gives $J_{2\nu}^\mu$ in agreement with the current equivalence relation, that is the second identity in (11.13).

11.3.2 Spacetime Translations and Rotations

In relativistic theories with only scalar fields, spacetime translations and rotations can be implemented through their action on Minkowski coordinates,

$$x^\mu \rightarrow x^\mu + \epsilon_1^\mu + \epsilon_2^{\mu\nu} x_\nu . \quad (11.24)$$

The vector ϵ_1^μ parameterizes infinitesimal translations and the antisymmetric tensor $\epsilon_2^{\mu\nu}$ infinitesimal spacetime rotations. The evolutionary form (11.1) of the transformations induced on the fields ψ^i is obtained by setting, respectively,¹

$$\delta_1 \psi^i = \epsilon_1^\mu F_\mu^i = -\epsilon_1^\mu \partial_\mu \psi^i , \quad \delta_2 \psi^i = \frac{1}{2} \epsilon_2^{\mu\nu} F_{\mu\nu}^i = \epsilon_2^{\mu\nu} x_\mu \partial_\nu \psi^i . \quad (11.25)$$

Translations and rotations are locally equivalent via (11.2); the latter can be recovered from the former by setting

$$\epsilon_1^\mu(x) = \frac{1}{2} f_{\alpha\beta}^\mu(x) \epsilon_2^{\alpha\beta}(x) , \quad f_{\alpha\beta}^\mu(x) \equiv \delta_\alpha^\mu x_\beta - \delta_\beta^\mu x_\alpha . \quad (11.26)$$

The translations are the parent symmetry and the rotations its descendant. The Noether current corresponding to the parent symmetry is the *canonical energy-momentum* (EM) *tensor* $T^\mu{}_\nu$, defined by

$$\delta_1 S = \int d^D x T^\mu{}_\nu[\psi, x](x) \partial_\mu \epsilon_1^\nu(x) . \quad (11.27)$$

The left-hand side of the integrability condition (11.4) then becomes

$$T^\mu{}_\nu \partial_\mu f_{\alpha\beta}^\nu = T_{\beta\alpha} - T_{\alpha\beta} . \quad (11.28)$$

It is known that in purely scalar theories, the canonical EM tensor is symmetric; I will justify this below using the background gauging approach. The integrability condition is then satisfied by $N_{\alpha\beta}^\mu = 0$. The equivalence relation (11.5) in turn gives the Noether current for spacetime rotations, that is the angular momentum tensor,

¹ The factor 1/2 in $\delta_2 \psi^i$ is conventional and compensates for the fact that the components $\epsilon_2^{\alpha\beta}$ and $\epsilon_2^{\beta\alpha}$ with fixed $\alpha \neq \beta$ parameterize the same transformation. The derivation of the angular momentum tensor $M^\mu{}_{\alpha\beta}$ below assumes that the same factor 1/2 is used in its definition via (4.7).

$$M^\mu_{\alpha\beta} = f^\nu_{\alpha\beta} T^\mu_\nu = x_\beta T^\mu_\alpha - x_\alpha T^\mu_\beta . \quad (11.29)$$

This is just the familiar relation between momentum and angular momentum.

The equivalence relation (11.29) does not apply to theories of fields with nonzero spin without further qualification. Namely, in such theories, spacetime translations and rotations are not automatically locally equivalent. The reason for this is that the components of fields with spin undergo a transformation under spacetime rotations, which cannot be reproduced by the naive local translation, $\delta_1 \psi^i(x) = -\epsilon_1^\mu(x) \partial_\mu \psi^i(x)$, for any choice of $\epsilon_1^\mu(x)$. The problem can be circumvented by a judicious choice of the localized transformation under translations. Instead of developing a general theory, let me illustrate the idea on a simple example.

Example 11.2

Recall Example 4.3 where I introduced the following theory of a real scalar field ϕ and a vector field A_μ ,

$$\mathcal{L} = A^\mu \partial_\mu \phi - \frac{1}{2} A^\mu A_\mu . \quad (11.30)$$

We found that the canonical EM tensor of this theory is not symmetric. However, we could solve the problem by considering a modified local translation,

$$\delta_1 \phi(x) = -\epsilon_1^\nu(x) \partial_\nu \phi(x) , \quad \delta_1 A_\mu(x) = -\epsilon_1^\nu(x) \partial_\nu A_\mu(x) - A_\nu(x) \partial_\mu \epsilon_1^\nu(x) . \quad (11.31)$$

This is motivated by the fact that under a general local coordinate transformation, A_μ should transform as a covariant vector. Noether's theorem then gives the improved, off-shell-symmetric EM tensor

$$\tilde{T}^{\mu\nu} = g^{\mu\nu} \mathcal{L} + A^\mu A^\nu - (A^\mu \partial^\nu \phi + A^\nu \partial^\mu \phi) . \quad (11.32)$$

The prescription (11.31) restores local equivalence of spacetime translations and rotations in spite of the presence of the vector field. Namely, inserting $\epsilon_1^\mu(x) = (1/2) f^\mu_{\alpha\beta}(x) \epsilon_2^{\alpha\beta}$ with constant $\epsilon_2^{\alpha\beta}$ in (11.31) gives

$$\delta_1 A_\mu(x) \xrightarrow{\epsilon_1^\mu(x) = \epsilon_2^{\mu\nu} x_\nu} \epsilon_2^{\alpha\gamma} x_\alpha \partial_\gamma A_\mu(x) + \epsilon_{2\mu\nu} A^\nu(x) . \quad (11.33)$$

This is exactly how A_μ should transform under a spacetime rotation. We can then localize both translations and rotations by using (11.31) augmented with (11.26). Since $\tilde{T}^{\mu\nu}$ is symmetric, the integrability condition is now satisfied and we reproduce the angular momentum tensor by replacing (11.29) with

$$M^\mu_{\alpha\beta} = x_\beta \tilde{T}^\mu_\alpha - x_\alpha \tilde{T}^\mu_\beta . \quad (11.34)$$

Let us now address the relation between spacetime translations and rotations within the background gauging formalism. We start by gauging the parent symmetry: spacetime translations. This amounts to promoting the flat Minkowski spacetime to a possibly curved spacetime manifold. Here it proves useful to invoke the differential-geometric language developed in Appendix A, to which I refer the reader for the basic terminology. Thus, the geometry of the spacetime manifold can be characterized by a local coframe (basis of differential 1-forms), $e^{*\alpha} \equiv e_{\mu}^{*\alpha} dx^{\mu}$.² Under an infinitesimal coordinate transformation, $x^{\mu} \rightarrow x^{\mu} + \epsilon_1^{\mu}(x)$, the coordinate components of the coframe change as

$$\delta e_{\mu}^{*\alpha}(x) = -\epsilon_1^{\nu}(x) \partial_{\nu} e_{\mu}^{*\alpha}(x) - e_{\nu}^{*\alpha}(x) \partial_{\mu} \epsilon_1^{\nu}(x) . \quad (11.35)$$

This matches (11.8) provided we identify $\tilde{F}_{1\nu\mu}^{\alpha} = -\partial_{\nu} e_{\mu}^{*\alpha}$ and $\tilde{K}_{1\lambda\mu}^{\alpha\nu} = -\delta_{\mu}^{\nu} e_{\lambda}^{*\alpha}$. A gauge-invariant action $\tilde{S}\{\psi, e^{*}\}$ is obtained from $S\{\psi\}$ by replacing derivatives of ψ^i therein with covariant derivatives, and the flat Minkowski metric $g_{\mu\nu}$ with the spacetime metric, $g_{\mu\nu}(x) = g_{\alpha\beta} e_{\mu}^{*\alpha}(x) e_{\nu}^{*\beta}(x)$. An appropriate volume element is built using the determinant of the coframe $e_{\mu}^{*\alpha}(x)$ as a matrix.

Eventually, we want to remove the background and go back to the flat Minkowski spacetime. I assume that the latter is described by global coordinates x^{μ} in which the components $e_{\mu}^{*\alpha}$ are constant. Then the consistency condition (11.10) is solved by $R_{\nu}^{\mu} = 0$, and (11.11) reduces to

$$T^{\mu}_{\nu}[\psi, x](x) = e_{\nu}^{*\alpha}(x) \left. \frac{\delta \tilde{S}\{\psi, e^{*}\}}{\delta e_{\mu}^{*\alpha}(x)} \right|_{\text{vac}} . \quad (11.36)$$

Suppose now that the theory at hand can be coupled to the background geometry in a way that the action $\tilde{S}\{\psi, e^{*}\}$ depends on $e_{\mu}^{*\alpha}$ only through the spacetime metric, $g_{\mu\nu}$. This is an example of a higher-rank tensor background hinted at above. Then the variation of the action (11.9) is replaced with

$$\delta_g \tilde{S} = \int d^D x \frac{\delta \tilde{S}}{\delta g_{\mu\nu}} (-\epsilon_1^{\lambda} \partial_{\lambda} g_{\mu\nu} - g_{\lambda\nu} \partial_{\mu} \epsilon_1^{\lambda} - g_{\mu\lambda} \partial_{\nu} \epsilon_1^{\lambda}) . \quad (11.37)$$

On a flat Minkowski background, this automatically vanishes for constant ϵ_1^{μ} thanks to $\partial_{\lambda} g_{\mu\nu}|_{\text{vac}} = 0$. The EM tensor is then identified as

$$T^{\mu\nu}[\psi, x](x) = 2 \left. \frac{\delta \tilde{S}\{\psi, g\}}{\delta g_{\mu\nu}(x)} \right|_{\text{vac}} . \quad (11.38)$$

This is the *Hilbert EM tensor*. Should our theory be Lorentz-invariant and only contain scalar fields, it can always be coupled to the spacetime background using the metric $g_{\mu\nu}$. At the same time, the scalar fields retain the transformation rule

² Following the notation introduced in Sect. 10.2.3, I use underlined indices to indicate components within a local (co)frame. Ordinary Lorentz indices refer to a set of local spacetime coordinates, x^{μ} .

$\delta_1 \psi^i(x) = -\epsilon_1^\mu(x) \partial_\mu \psi^i(x)$. This guarantees that the canonical and Hilbert EM tensors coincide, and shows that the canonical EM tensor is necessarily symmetric.

Theories of fields with spin can be coupled to background geometry in different manners, depending on how the spin index of the fields is treated. Let me illustrate this in the case of a vector field A_μ . One already mentioned possibility is to treat this as a covariant vector that transforms under a general coordinate transformation via the second relation in (11.31). It is then possible to construct a generally covariant action $\tilde{S}\{A, g\}$ that only depends on the background through the metric. The resulting EM tensor (11.38) is symmetric.

On the other hand, we may project the field on the local frame $e_{\underline{\alpha}}^\mu$ and treat $A_{\underline{\alpha}} \equiv e_{\underline{\alpha}}^\mu A_\mu$ as a set of spacetime scalars. This corresponds to keeping the naive transformation under local translations, $\delta A_{\underline{\alpha}}(x) = -\epsilon_1^\mu(x) \partial_\mu A_{\underline{\alpha}}(x)$. In this case, the gauged action $\tilde{S}\{A, e^*\}$ may depend explicitly on the local frame. We then have to use (11.36) instead of (11.38). This gives the canonical EM tensor which is not necessarily symmetric.

The general moral is that different choices of gauging the symmetry may lead to different expressions for the Noether current. This is tantamount to different choices of the localized symmetry transformation (11.7). The resulting alternative Noether currents however only differ by terms that vanish on-shell, modulo the inevitable ambiguity with respect to contributions whose divergence vanishes off-shell.

With the EM tensor at hand, we can extract the Noether current for any descendant symmetry that can be locally reproduced by spacetime translations (see e.g. Sect. 14.3 of [2]). Consider a one-parameter group of symmetries, generated by a vector field $\xi(x)$. The corresponding local transformation with parameter $\epsilon_2(x)$ can be written as $x^\mu \rightarrow x^\mu + \epsilon_1^\mu(x)$ with $\epsilon_1^\mu(x) = \epsilon_2(x) \xi^\mu(x)$. This turns (11.35) to

$$\delta e_{\underline{\mu}}^{*\alpha} = -\epsilon_2(\xi^\nu \partial_\nu e_{\underline{\mu}}^{*\alpha} + e_{\underline{\nu}}^{*\alpha} \partial_{\underline{\mu}} \xi^\nu) - \xi^\nu e_{\underline{\nu}}^{*\alpha} \partial_{\underline{\mu}} \epsilon_2, \quad (11.39)$$

whence we extract $\tilde{F}_{2\underline{\mu}}^\alpha = -(\xi^\nu \partial_\nu e_{\underline{\mu}}^{*\alpha} + e_{\underline{\nu}}^{*\alpha} \partial_{\underline{\mu}} \xi^\nu)$ and $\tilde{K}_{2\underline{\mu}}^{\alpha\nu} = -\delta_{\underline{\mu}}^\nu \xi^\lambda e_{\underline{\lambda}}^{*\alpha}$. The consistency condition (11.10) then boils down to

$$\partial_{\underline{\mu}} R^{\underline{\mu}} = -T^{\underline{\mu}}{}_{\underline{\nu}} \partial_{\underline{\mu}} \xi^{\underline{\nu}} = -\frac{1}{2} T^{\underline{\mu}\underline{\nu}} (\partial_{\underline{\mu}} \xi_{\underline{\nu}} + \partial_{\underline{\nu}} \xi_{\underline{\mu}}). \quad (11.40)$$

In the last step, I assumed that $T^{\underline{\mu}\underline{\nu}}$ is symmetric, which we now know to be guaranteed in case the action $\tilde{S}\{\psi, e^*\}$ only depends on the coframe through the metric. But should $\xi(x)$ actually generate a symmetry of the flat Minkowski spacetime, $\partial_{\underline{\mu}} \xi_{\underline{\nu}} + \partial_{\underline{\nu}} \xi_{\underline{\mu}}$ must vanish by the Killing equation (A.92). Then (11.10) is satisfied with $R^{\underline{\mu}} = 0$ and (11.11) gives immediately

$$J^\mu[\psi, x](x) = T^{\mu\nu}[\psi, x](x) \xi_\nu(x). \quad (11.41)$$

This confirms, among others, a general relation between the Hilbert EM tensor and the angular momentum tensor, anticipated in (11.34).

11.3.3 Galilei Invariance

The previous two examples were relativistic in spirit. Let us therefore have a look at one more example that is intrinsically nonrelativistic. Consider a theory of a complex Schrödinger field $\psi(\mathbf{x}, t)$ whose excitations are nonrelativistic particles with mass m . Suppose that the theory is invariant under spatial translations and under the internal U(1) group of phase transformations of the field. These symmetries act respectively on the spatial coordinates, $\mathbf{x}' = \mathbf{x} + \boldsymbol{\epsilon}$, and the field, $\psi' = e^{i\alpha}\psi$, and so correspond to

$$\delta_1 \psi(\mathbf{x}, t) = i\alpha \psi(\mathbf{x}, t) - \boldsymbol{\epsilon} \cdot \boldsymbol{\nabla} \psi(\mathbf{x}, t) . \quad (11.42)$$

Finally, it is well-known from elementary quantum mechanics that the Schrödinger equation for a free particle is invariant under Galilei boosts,

$$\psi'(\mathbf{x}', t) = \psi'(\mathbf{x} + \mathbf{v}t, t) = \exp \left[im \left(\mathbf{v} \cdot \mathbf{x} + \frac{1}{2} \mathbf{v}^2 t \right) \right] \psi(\mathbf{x}, t) . \quad (11.43)$$

Here \mathbf{v} is the boost velocity which plays the role of the transformation parameter. Let us assume that the invariance under (11.43) is inherited by our possibly interacting nonrelativistic field theory.

The evolutionary form of an infinitesimal Galilei boost is

$$\delta_2 \psi(\mathbf{x}, t) = im \mathbf{v} \cdot \mathbf{x} \psi(\mathbf{x}, t) - t \mathbf{v} \cdot \boldsymbol{\nabla} \psi(\mathbf{x}, t) . \quad (11.44)$$

This can be locally recovered as a combination of a spatial translation and a phase transformation if we set

$$\alpha(\mathbf{x}, t) = m \mathbf{x} \cdot \mathbf{v}(\mathbf{x}, t) , \quad \boldsymbol{\epsilon}(\mathbf{x}, t) = t \mathbf{v}(\mathbf{x}, t) . \quad (11.45)$$

In this case, therefore, the combination of spatial translations and internal U(1) is the parent symmetry, whereas the Galilei boosts are its descendants. The rest follows a familiar pattern. The integrability condition (11.4) requires that there is a local function $N^{\mu r}[\psi, \mathbf{x}, t]$ of the field and its derivatives such that

$$m J^r + T^{0r} = \partial_\mu N^{\mu r} , \quad (11.46)$$

where J^μ is the current of the internal U(1) symmetry. The general equivalence relation (11.5) then gives the current $B^{\mu r}$ corresponding to Galilei boosts,

$$B^{\mu r} = m x^r J^\mu + t T^{\mu r} - N^{\mu r} . \quad (11.47)$$

Should $N^{\mu r}$ happen to be zero, the integrability condition (11.46) would boil down to $T^{0r} = -mJ^r$. Except for the sign convention in the definition of the two currents, this is the usual nonrelativistic relation between momentum and velocity. Moreover, (11.47) would imply $B^{0r} = mx^r J^0 + tT^{0r} = m(x^r J^0 - tJ^r)$. This generalizes to field theory the well-known fact that the conserved charge for a Galilei boost of a particle of mass m is $m\mathbf{x} - \mathbf{p}t$, where \mathbf{p} is the particle momentum. Whether or not, or under what conditions, $N^{\mu r}$ actually is zero needs to be inspected case by case.

The story of Galilei invariance becomes very interesting, and nontrivial, once we consider background gauging. What we need is to gauge simultaneously spatial translations, Galilei boosts and the internal U(1) phase transformations. Since the Galilei boosts are descendant, we want to primarily gauge spatial translations and the internal U(1). The actions of these cannot be trivially separated as one could do in a Lorentz-invariant field theory. Namely, the gauged translations must affect the phase of the Schrödinger field in a way that reproduces (11.43).

It turns out that the minimal setup that does the job includes a background spatial metric $g_{rs}(x)$ and a gauge field $A_\mu(x)$ for the U(1) symmetry [3]. Under a combination of an infinitesimal local translation, $\mathbf{x}' = \mathbf{x} + \boldsymbol{\epsilon}(\mathbf{x}, t)$, and a U(1) transformation with parameter $\alpha(\mathbf{x}, t)$, the background fields vary by

$$\begin{aligned}\delta g_{rs}(x) &= -\epsilon''(x)\partial_u g_{rs}(x) - g_{us}(x)\partial_r \epsilon''(x) - g_{ru}(x)\partial_s \epsilon''(x) , \\ \delta A_0(x) &= \partial_0 \alpha(x) - \epsilon''(x)\partial_u A_0(x) - A_u(x)\partial_0 \epsilon''(x) , \\ \delta A_r(x) &= \partial_r \alpha(x) - \epsilon''(x)\partial_u A_r(x) - A_u(x)\partial_r \epsilon''(x) - mg_{rs}(x)\partial_0 \epsilon^s(x) .\end{aligned}\tag{11.48}$$

The first two lines take the expected form. The only nontrivial ingredient really is the last term on the third line of (11.48), which depends on the mass parameter m . This reflects the fact that the realization of spatial translations and Galilei boosts on a Schrödinger field is centrally extended, with m playing the role of the central charge. The associated nonrelativistic spacetime geometry is known as *Newton–Cartan*. See [4, 5] for more details on nonrelativistic field theory in Newton–Cartan spacetimes, and [6] for broader background on nonrelativistic geometry.

The localized transformation of the Schrödinger field itself is given by (11.42). By combining this with (11.48), we then generate at once the Noether currents of all the symmetries in terms of a gauged action $\tilde{S}\{\psi, g, A\}$. First, the current of the U(1) symmetry is given by a trivial modification of (11.21),

$$J^\mu[\psi, x](x) = - \frac{\delta \tilde{S}\{\psi, g, A\}}{\delta A_\mu(x)} \Big|_{\text{vac}} ,\tag{11.49}$$

where “vac” now indicates setting g_{rs} to the flat Euclidean metric δ_{rs} and A_μ to zero. Next comes the EM tensor. By (11.27), the only components that the invariance under (time-dependent) spatial coordinate transformations gives us access to are

$$T^{0r}[\psi, x](x) = m \frac{\delta \tilde{S}\{\psi, g, A\}}{\delta A_r(x)} \Big|_{\text{vac}} , \quad T^{rs}[\psi, x](x) = 2 \frac{\delta \tilde{S}\{\psi, g, A\}}{\delta g_{rs}(x)} \Big|_{\text{vac}} .\tag{11.50}$$

Note that (11.49) and (11.50) together imply a relation between momentum density and the particle number current, $T^{0r} = -mJ^r$. This corresponds to (11.46) with vanishing $N^{\mu r}$. It might come as a surprise that the background gauge invariance yields a stronger constraint on the Noether currents than the original, physical symmetry. After all, background gauge invariance per se is merely a handy tool to encode the consequences of the symmetry. However, recall that the realization of Galilei transformations on the Schrödinger field is centrally extended. This is an obstruction that makes gauging of the symmetry nontrivial. It is ultimately the assumption that the gauging is possible that is responsible for the vanishing of $N^{\mu r}$. This does not add any new physics, but rather constrains the class of theories to which the background gauging procedure can be applied.

Next, the current $B^{\mu r}$ for Galilei boosts can be identified by using (11.9) together with $\delta_\psi \tilde{S} = \int d^D x B^{\mu r} \partial_\mu v^r$. Upon inserting (11.45) in (11.48), we find that the consistency condition (11.10) is satisfied with $R_r^\mu = 0$. This leads via (11.11) to

$$\begin{aligned} B^{0r}[\psi, x](x) &= -mx^r \left. \frac{\delta \tilde{S}\{\psi, g, A\}}{\delta A_0(x)} \right|_{\text{vac}} + mt \left. \frac{\delta \tilde{S}\{\psi, g, A\}}{\delta A_r(x)} \right|_{\text{vac}}, \\ B^{rs}[\psi, x](x) &= -mx^s \left. \frac{\delta \tilde{S}\{\psi, g, A\}}{\delta A_r(x)} \right|_{\text{vac}} + 2t \left. \frac{\delta \tilde{S}\{\psi, g, A\}}{\delta g_{rs}(x)} \right|_{\text{vac}}. \end{aligned} \quad (11.51)$$

These can be combined using (11.49) and (11.50) into $B^{\mu r} = mx^r J^\mu + tT^{\mu r}$, which is just (11.47) without the $N^{\mu r}$ term. This is consistent with our conclusion above that $N^{\mu r} = 0$ as a consequence of the assumption that the gauged action $\tilde{S}\{\psi, g, A\}$ is invariant under (11.48).

11.3.4 Changing the Background: Magnetic Translations

Our discussion in this chapter started with declaring the existence of certain symmetry and its subsequent localization. However, it is also possible to adopt the opposite approach. Suppose we know a priori the type of background gauge fields and the local transformations that make the action $\tilde{S}\{\psi, A\}$ invariant. It is then possible to identify the original symmetry as the subset of the gauge transformations that preserve a trivial background.

Example 11.3

Recall the Galileon symmetry of Sect. 11.3.1. Out of all the gauge transformations $A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \epsilon(x)$, only those with constant ϵ preserve the background, in this case regardless of the specific choice of $A_\mu(x)$. Constant ϵ corresponds exactly to the shift symmetry, acting on the Galileon scalar field ϕ .

Example 11.4

In case of relativistic theories of fields with integer spin, we can consider as the background a generic spacetime metric $g_{\mu\nu}(x)$. The set of local general coordinate transformations that preserve such a background corresponds to (the continuous part of) its isometry group (see Appendix A.6.2 for the necessary mathematical background). In case of the flat Minkowski spacetime, this is just the Poincaré group, as follows from a trivial modification of Example A.15. Note that choosing the background geometry is not equivalent to fixing the coframe $e_\mu^{*\alpha}$ as (11.36) might naively suggest. The reason is that the coframe is not uniquely determined for a given spacetime manifold. This underlines the necessity to carefully identify the geometric data, tailored to the given spacetime symmetry and uniquely representing the spacetime background.

Once we have the gauged action $\tilde{S}\{\psi, A\}$, we may use it to study physics on nontrivial backgrounds. Indeed, the vanishing of the variation of $\tilde{S}\{\psi, A\}$ under a localized symmetry transformation is an exact statement valid for any choice of A_μ^A . Imposing the equation of motion on the dynamical fields ψ^i amounts to setting $\delta_\psi \tilde{S} = 0$, which turns (11.9) into a generalized “conservation law,”

$$\int d^D x \left. \frac{\delta \tilde{S}}{\delta A_\mu^A} (\epsilon^B \tilde{F}_{B\mu}^A + \partial_\nu \epsilon^B \tilde{K}_{B\mu}^{A\nu} + \dots) \right|_{\text{on-shell}} = 0. \quad (11.52)$$

For this to represent an actual divergence-type conservation law, the chosen background must possess *some* symmetry. This may however be different from the symmetry that we originally gauged by introducing the fields A_μ^A in the first place. I will illustrate this on an interesting example starting from Galilei invariance in $d = 3$ spatial dimensions.

The flat nonrelativistic spacetime corresponds to $g_{rs} = \delta_{rs}$ and $A_\mu = 0$. The only infinitesimal transformations of type (11.48) that preserve this background are a linear combination of Euclidean translations and rotations, Galilei boosts, and the internal U(1) transformations. This is not obvious but is straightforward to check.

Let us now change the background by introducing a uniform magnetic field \mathbf{B} . Such a uniform background field should preserve some notion of translation invariance. This however cannot be the naive Euclidean translations due to the coordinate-dependence of the vector potential $\mathbf{A}(\mathbf{x})$ of the magnetic field. The details turn out to depend on the choice of gauge; I will adopt the symmetric gauge, $\mathbf{A}(\mathbf{x}) = (\mathbf{B} \times \mathbf{x})/2$. The translation invariance can now be rescued if it is accompanied by a local U(1) transformation. In terms of the Schrödinger field $\psi(\mathbf{x}, t)$, this so-called *magnetic translation symmetry* takes the form

$$\psi'(\mathbf{x}', t) = \psi'(\mathbf{x} + \boldsymbol{\epsilon}, t) = \exp \left[\frac{i}{2} \boldsymbol{\epsilon} \cdot (\mathbf{x} \times \mathbf{B}) \right] \psi(\mathbf{x}, t) . \quad (11.53)$$

One remarkable consequence of this twisted translation symmetry is that the components of its generator P_r no longer commute with each other. It is easy to check that they satisfy the commutation relation $[P_r, P_s] = -i\epsilon_{rsu}B^uQ$, where Q is the generator of the $U(1)$ symmetry.

The magnetic translation symmetry (11.53) can be localized by promoting $\boldsymbol{\epsilon}$ to a function of spacetime coordinates and setting

$$\psi'(\mathbf{x} + \boldsymbol{\epsilon}(x), t) = \exp[-i\boldsymbol{\epsilon}(x) \cdot \mathbf{A}(x)]\psi(x, t) , \quad (11.54)$$

where $\mathbf{A}(x)$ is now the spatial part of the variable $A_\mu(x)$. At the infinitesimal level, this is equivalent to (11.42) with $\alpha(x) = -\boldsymbol{\epsilon}(x) \cdot \mathbf{A}(x)$. Following the same reasoning as in Sect. 11.3.3, we then get modified relations for the EM tensor,

$$\begin{aligned} T^{0r}[\psi, x](x) &= -mJ^r[\psi, x](x) - 2A^r(x)J^0[\psi, x](x)|_{\text{vac}} , \\ T^{rs}[\psi, x](x) &= 2\frac{\delta\tilde{S}\{\psi, g, A\}}{\delta g_{rs}(x)}\Big|_{\text{vac}} - 2A^s(x)J^r[\psi, x](x)|_{\text{vac}} , \end{aligned} \quad (11.55)$$

which generalize (11.50) to a uniform magnetic background. Here “vac” refers to the background vector potential in the symmetric gauge, $A_\mu(x)|_{\text{vac}} = (1/2)\delta_\mu^r\epsilon_{rsu}B^sx^u$. The components of $J^\mu[\psi, x]$ are still given by (11.49).

The magnetic translations illustrate in a very nontrivial manner many of the concepts introduced in Chap. 4 and this chapter. On the magnetic background, the Lagrangian of the Schrödinger field depends explicitly on the spacetime coordinates, yet possesses a nontrivial notion of translation symmetry. The Lie algebra of infinitesimal translations is centrally extended. The translation symmetry can be gauged simultaneously with the internal $U(1)$ symmetry by introducing a spatial metric and allowing the background fields to transform according to (11.48). This allows one to identify a nontrivial relation between the EM tensor and the Noether current of the $U(1)$ symmetry, shown in (11.55).

11.4 Application to Scattering of Nambu–Goldstone Bosons

I will now return to the promise I made in Sect. 10.3.³ Therein, I suggested that the enhanced soft limit of scattering amplitudes of NG bosons in the DBI and Galileon theories arises from the presence of extended symmetry. We now have all we need to be able to understand why. The material of this section loosely follows [7].

³ A reader interested primarily in the development of the general theory for spontaneously broken spacetime symmetry may proceed directly to the next chapter.

11.4.1 Galileon Theory

Let us start with the simpler Galileon theory and recall the discussion of the Galileon symmetry in Sect. 11.3.1. In any translationally-invariant theory endowed with the Galileon symmetry, the parent current $J_1^\mu[\phi]$ will not depend explicitly on the coordinates. By the integrability condition in (11.13), the same holds for the local function $N_\nu^\mu[\phi]$. Following the philosophy of Sect. 10.1, we now consider an arbitrary scattering process $\alpha \rightarrow \beta$. Using the translation property $J_1^\mu[\phi](x) = e^{iP \cdot x} J_1^\mu[\phi](0) e^{-iP \cdot x}$ and analogously $N_\nu^\mu[\phi](x) = e^{iP \cdot x} N_\nu^\mu[\phi](0) e^{-iP \cdot x}$, we deduce

$$\langle \beta | J_{1\nu}[\phi](0) | \alpha \rangle = -i p_\mu \langle \beta | N_\nu^\mu[\phi](0) | \alpha \rangle, \quad (11.56)$$

where $p^\mu \equiv p_\alpha^\mu - p_\beta^\mu$.

The general rules of polology dictate that when p^μ approaches the mass shell of a massless particle, the matrix element $\langle \beta | J_{1\nu}[\phi](0) | \alpha \rangle$ develops a pole. This corresponds to the emission of a NG boson with momentum \mathbf{p} and allows one to extract the on-shell scattering amplitude $\mathcal{A}_{\alpha \rightarrow \beta + \pi(\mathbf{p})}$; cf. (10.1). The same reasoning can however also be applied to the operator $N_\nu^\mu[\phi]$, leading to

$$\langle \beta | N_\nu^\mu[\phi](0) | \alpha \rangle = \langle 0 | N_\nu^\mu[\phi](0) | \pi(\mathbf{p}) \rangle_{\text{off}} \frac{1}{p^2} \mathcal{A}_{\alpha \rightarrow \beta + \pi(\mathbf{p})} + R_{\beta\alpha\nu}^\mu(p), \quad (11.57)$$

where $R_{\beta\alpha\nu}^\mu(p)$ is by definition free of poles. By combining this with (11.56), we find a relation between $R_{\beta\alpha\nu}^\mu(p)$ and the function $R_{\beta\alpha}^\mu(p)$ in (10.1),

$$R_{\beta\alpha\nu}(p) = -i p_\mu R_{\beta\alpha\nu}^\mu(p). \quad (11.58)$$

From (10.2) we then finally get an explicit expression for the on-shell amplitude,

$$\mathcal{A}_{\alpha \rightarrow \beta + \pi(\mathbf{p})} = \frac{1}{F} p_\mu p^\nu R_{\beta\alpha\nu}^\mu(p), \quad (11.59)$$

where F is determined by $\langle 0 | J_1^\mu[\phi](0) | \pi(\mathbf{p}) \rangle = i p_{\text{on}}^\mu F$, and $p_{\text{on}}^\mu \equiv (|\mathbf{p}|, \mathbf{p})$.

The general relation (10.2) allows one to assert the existence of Adler zero, provided the function $R_{\beta\alpha}^\mu(p)$ remains regular in the soft limit $p^\mu \rightarrow 0$. A sufficient condition to ensure this is, as we know from Chap. 10, the absence of bilinear operators in the current $J_1^\mu[\phi]$. This automatically translates into the absence of bilinear operators in $N_\nu^\mu[\phi]$. We then conclude from (11.59) without further assumptions that in the soft limit, the scattering amplitudes in the Galileon theory vanish at least with the second power of momentum. In other words, the soft limit of scattering amplitudes in the Galileon theory is enhanced with the soft scaling parameter $\sigma = 2$.

11.4.2 Theories with Generalized Shift Symmetry

Having warmed up on the Galileon theory, we can now generalize the above argument to a much broader class of theories. For simplicity of notation, I will restrict the discussion to Lorentz-invariant theories of a single NG field ϕ that are invariant under the constant shift $\phi \rightarrow \phi + \epsilon$ [7]. Further generalization to theories invariant under spacetime translations but mere spatial rotations can be found in [8]. A generalization to theories of multiple flavors of NG bosons presumably exists but, as far as I know, has not been worked out in the literature.

Suppose that in addition to the constant shift symmetry, our theory is also invariant under a generalized shift of the type

$$\delta\phi(x) = \epsilon \{f(x) + \Phi[\phi, x](x)\} . \quad (11.60)$$

Here $f(x)$ is assumed to be a polynomial function of the spacetime coordinates of degree $\deg f \equiv n \geq 1$. The local operator $\Phi[\phi, x]$ may depend explicitly on the coordinates as well, but is required to be polynomial in x^μ of degree lower than n . Examples of symmetries of this type include both the symmetry of the DBI theory (10.51) and the special Galileon symmetry (10.56).

Let us now localize (11.60) and treat it as a generalized coordinate-dependent shift of the field, that is $\delta\phi(x) = \tilde{\epsilon}[\phi, x](x)$ with $\tilde{\epsilon}[\phi, x](x) \equiv \epsilon(x) \{f(x) + \Phi[\phi, x](x)\}$. The corresponding variation of the action is

$$\begin{aligned} \delta S = \int d^D x J_1^\mu[\phi](x) \{ \epsilon(x) \partial_\mu [f(x) + \Phi[\phi, x](x)] \\ + [f(x) + \Phi[\phi, x](x)] \partial_\mu \epsilon(x) \} , \end{aligned} \quad (11.61)$$

where $J_1^\mu[\phi]$ is the parent current due to the constant shift symmetry. An integrability condition required so that (11.60) really is a symmetry of the action reads

$$J_1^\mu[\phi](x) \partial_\mu [f(x) + \Phi[\phi, x](x)] = \partial_\mu N^\mu[\phi, x](x) . \quad (11.62)$$

Unlike the parent current $J_1^\mu[\phi]$, the local function $N^\mu[\phi, x]$ may depend explicitly on the coordinates. For (11.62) to hold, however, the polynomial degree of $N^\mu[\phi, x]$ in x^μ can be at most that of $\partial_\mu f(x)$, that is $n - 1$.

The next step is to rewrite (11.62) in the form

$$\begin{aligned} J_1^\mu[\phi] \partial_\mu f &= \Phi[\phi, x] \partial_\mu J_1^\mu[\phi] - \partial_\mu (\Phi[\phi, x] J_1^\mu[\phi] - N^\mu[\phi, x]) \\ &\equiv \Phi[\phi, x] \partial_\mu J_1^\mu[\phi] + \partial_\mu M^\mu[\phi, x] . \end{aligned} \quad (11.63)$$

The new local function $M^\mu[\phi, x]$ of the field and its derivatives is again of polynomial degree at most $n - 1$ in the spacetime coordinates. It satisfies the translation property $M^\mu[\phi, x](x) = e^{iP \cdot x} M^\mu[\phi, x](0) e^{-iP \cdot x}$, with the fields on the right-hand side evaluated at the origin but the explicit coordinate dependence left intact. Applying the polology rules to this operator then gives, in analogy to (11.57),

$$\begin{aligned} \langle \beta | M^\mu[\phi, x](0) | \alpha \rangle &= \langle 0 | M^\mu[\phi, x](0) | \pi(\mathbf{p}) \rangle_{\text{off}} \frac{1}{p^2} \mathcal{A}_{\alpha \rightarrow \beta + \pi(\mathbf{p})} \\ &+ R_{M\beta\alpha}^\mu(x, p) . \end{aligned} \quad (11.64)$$

The coordinate dependence of the remainder function $R_{M\beta\alpha}^\mu(x, p)$ comes entirely from that of $M^\mu[\phi, x](0)$. Therefore, $R_{M\beta\alpha}^\mu(x, p)$ is a polynomial function of the coordinates with degree at most $n - 1$.

The crucial observation is that due to current conservation, the on-shell matrix element of $\Phi[\phi, x] \partial_\mu J_1^\mu[\phi]$ between any initial and final states, $|\alpha\rangle$ and $|\beta\rangle$, vanishes. By combining (11.63) and (11.64), we then get a constraint on the remainder functions, generalizing (11.58),

$$R_{\beta\alpha}^\mu(p) \partial_\mu f(x) = -i p_\mu R_{M\beta\alpha}^\mu(x, p) + \partial_\mu R_{M\beta\alpha}^\mu(x, p) . \quad (11.65)$$

To understand the implications of this identity, we expand $R_{\beta\alpha}^\mu(p)$ and $R_{M\beta\alpha}^\mu(x, p)$ in a formal power series in p^μ . Let us denote their coefficients of order k in p^μ respectively as a_k and $b_k(x)$, and write (11.65) symbolically as

$$a_k \cdot \partial f(x) \simeq b_{k-1}(x) + \partial \cdot b_k(x) . \quad (11.66)$$

This must hold as an equality between polynomial functions of the coordinates. We now use repeatedly the fact that $\deg b_k(x) \leq n - 1$ for all k . Setting $k = 0$, we infer that $a_0 = 0$ and simultaneously $\partial \cdot b_0(x) = 0$, which implies that $b_0(x)$ is constant, $\deg b_0(x) = 0$. In case $n \geq 2$, we can next take (11.66) with $k = 1$. This can only hold if $a_1 = 0$ and simultaneously $\partial \cdot b_1(x) \simeq -b_0(x)$, that is, $\deg b_1(x) \leq 1$. By induction, we find that for any $k \leq n - 1$, $a_k = 0$ and simultaneously $\deg b_k(x) \leq k$.

The main conclusion is that for a generalized shift symmetry of the type (11.60) where $\deg f(x) = n$, the series expansion of $R_{\beta\alpha}^\mu(p)$ starts at order n in energy-momentum. By (10.2), the soft limit of scattering amplitudes of the theory is then enhanced with the soft degree $\sigma = n + 1$. This confirms in particular that $\sigma = 2$ for the DBI theory and $\sigma = 3$ for the special Galileon theory.

References

1. T. Brauner, Phys. Scr. **95**, 035004 (2020). DOI [10.1088/1402-4896/ab50a5](https://doi.org/10.1088/1402-4896/ab50a5)
2. C.P. Burgess, *Introduction to Effective Field Theory* (Cambridge University Press, Cambridge, UK, 2021). DOI [10.1017/9781139048040](https://doi.org/10.1017/9781139048040)
3. D.T. Son, M. Wingate, Ann. Phys. **321**, 197 (2006). DOI [10.1016/j.aop.2005.11.001](https://doi.org/10.1016/j.aop.2005.11.001)
4. M. Geracie, D.T. Son, C. Wu, S.F. Wu, Phys. Rev. **D91**, 045030 (2015). DOI [10.1103/PhysRevD.91.045030](https://doi.org/10.1103/PhysRevD.91.045030)
5. K. Jensen, SciPost Phys. **5**, 011 (2018). DOI [10.21468/SciPostPhys.5.1.011](https://doi.org/10.21468/SciPostPhys.5.1.011)
6. E. Bergshoeff, J. Figueroa-O'Farrill, J. Gomis, SciPost Phys. Lect. Notes **69** (2023). DOI [10.21468/SciPostPhysLectNotes.69](https://doi.org/10.21468/SciPostPhysLectNotes.69)
7. C. Cheung, K. Kampf, J. Novotný, C.H. Shen, J. Trnka, JHEP **02**, 020 (2017). DOI [10.1007/JHEP02\(2017\)020](https://doi.org/10.1007/JHEP02(2017)020)
8. M.A. Mojahed, T. Brauner, JHEP **03**, 086 (2022). DOI [10.1007/JHEP03\(2022\)086](https://doi.org/10.1007/JHEP03(2022)086)

Chapter 12

Nonlinear Realization of Spacetime Symmetry

In Part III of the book, I assumed that whatever the full symmetry of the system, only internal symmetry is spontaneously broken. Moreover, I only considered two possibilities for the symmetry of the underlying spacetime. For nonrelativistic systems, I assumed invariance under spacetime translations and spatial rotations; this symmetry is sometimes called *Aristotelian*. In case of relativistic systems, the Aristotelian symmetry is turned into Poincaré symmetry by adding invariance under Lorentz boosts. Both of these spacetime symmetries are easy to implement provided they are not spontaneously broken. Invariance under spacetime translations is ensured by demanding that the Lagrangian density of the theory does not depend explicitly on spacetime coordinates. Invariance under spatial or spacetime rotations can be guaranteed by contracting indices appropriately using ordinary tensor calculus.

The situation becomes subtle when the spacetime symmetry is spontaneously broken, or when a different spacetime symmetry than Poincaré or Aristotelian is present. It is known that there are in fact multiple mathematically consistent kinematical groups including spacetime translations, spatial rotations and boosts [1]. Here I will consider in particular the Galilei symmetry, relevant for *effective field theories* (EFTs) of many nonrelativistic systems. Due to the mathematical structure of the Galilei group, the theory of its representations, and thus construction of Galilei-invariant actions, is highly nontrivial. For an overview of other possible non-Lorentzian kinematics and the associated spacetime geometry, see [2].

The EFT for broken spacetime symmetries finds natural application in cosmology. To make the methodology developed below useful also there, I will make an exception and initially allow for theories defined on a generic spacetime manifold. Some simplifying assumptions on the spacetime manifold will be necessary, but will only be introduced when needed. Otherwise, the general philosophy of construction of EFTs for spontaneously broken spacetime symmetry is the same as in Part III. I will start in this chapter by classifying possible nonlinear realizations of spacetime symmetry. In Chaps. 13 and 14, I will then utilize the results to build effective actions. Unfortunately, there are no general explicit expressions for effective Lagrangians akin to those for internal symmetries, worked out in Chap. 8. I will therefore have to resort to outlining the basic algorithm and working out some illustrative examples.

12.1 Reminder of Nonlinear Realization of Internal Symmetry

The material of this chapter closely parallels Chap. 7, which the reader is advised to remind themselves of before proceeding further. In order to underline the basic logic, let me however give at least a brief summary of the steps we took in Chap. 7.

We used the fact that internal symmetries in the sense of Sect. 4.1.1 are point transformations that only act on fields and leave spacetime coordinates intact. This allows for a geometric reformulation of the problem of classifying all nonlinear realizations of internal symmetry. Namely, if the fields take values from a manifold \mathcal{M} , then realizing the internal symmetry group G on them amounts to defining an *action* of G on \mathcal{M} . While the action of the symmetry group is assumed to exist globally on the whole manifold \mathcal{M} , its explicit expression requires a set of local coordinates. Thus, all statements depending on a specific choice of coordinates are without further qualification valid only in a local coordinate patch.

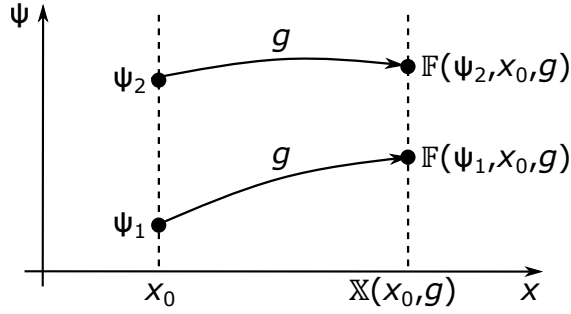
We start by picking a point $\psi_0 \in \mathcal{M}$.¹ The subgroup H_{ψ_0} of elements of G that map ψ_0 to itself is called its *isotropy group*. Whenever H_{ψ_0} is compact, there are local coordinates (π^a, χ^ϱ) on \mathcal{M} such that ψ_0 corresponds to $\pi^a = \chi^\varrho = 0$, and H_{ψ_0} acts separately and linearly on π^a and χ^ϱ . Moreover, the set of points $(\pi^a, 0)$ spans a submanifold of \mathcal{M} , equivalent to the coset space G/H_{ψ_0} . This makes it possible to encode π^a in a unique representative element $U(\pi)$ of the corresponding left coset of H_{ψ_0} in G . The representative element can be chosen so that the point ψ_0 is mapped to $U(0) = e$. In addition, the linear representation of the isotropy group H_{ψ_0} on π^a is realized by the adjoint action, $U(\pi) \rightarrow hU(\pi)h^{-1}$, $h \in H_{\psi_0}$. Altogether, the action of G on \mathcal{M} in the standard coordinates (π^a, χ^ϱ) is fixed by the choice of representative $U(\pi)$ and of the linear representation of H_{ψ_0} on χ^ϱ .

In physics terms, every point $\psi_0 \in \mathcal{M}$ defines a specific value of the order parameter for *spontaneous symmetry breaking* (SSB). The isotropy group H_{ψ_0} corresponds to the unbroken subgroup of G . Once treated as maps from the spacetime to \mathcal{M} , π^a become the *Nambu–Goldstone* (NG) fields implied by SSB, whereas χ^ϱ are usually called *matter fields*. The same symmetry-breaking pattern $G \rightarrow H$ may be realized by different order parameters, and thus on different target manifolds \mathcal{M} . This however only affects the number and type of matter fields present. The NG fields are identified unambiguously as coordinates on the coset space G/H , and are in a one-to-one correspondence with the generators of broken symmetry.

Moving now to spacetime symmetries, we will be able to follow the mathematical construction of nonlinear realization of internal symmetry very closely. We will only have to introduce one important modification, taking into account the distinction between fields and coordinates. However, the mapping to the physical concepts of SSB and NG bosons will turn out to be much more intricate.

¹ In Chap. 7, I used the symbol x_0 instead of ψ_0 , which I will now reserve for a reference point on the spacetime manifold.

Fig. 12.1 Visualization of the action of a spacetime symmetry on the product manifold $\mathcal{M} \times M$. The transformation of the spacetime coordinates x^μ is independent of the fields ψ^i . For fixed $x_0 \in M$ and $g \in G$, the symmetry thus acts as a map between the slices $\mathcal{M} \times \{x_0\}$ and $\mathcal{M} \times \{X(x_0, g)\}$, indicated by the *dashed* lines



12.2 Spacetime Symmetry as a Point Transformation

Consider a theory of a set of bosonic fields ψ^i living on a D -dimensional spacetime manifold M . It is temporarily convenient to treat ψ^i and the spacetime coordinates x^μ on equal footing as independent variables spanning the product manifold $\mathcal{M} \times M$. This allows us to view the action of a spacetime symmetry (in the sense of Sect. 4.1.1) as a special case of a point transformation on $\mathcal{M} \times M$. See Fig. 12.1 for a visualization. Formally, the action of a spacetime symmetry group G on $\mathcal{M} \times M$ constitutes a set of maps T_g , or equivalently functions \mathbb{F}^i and \mathbb{X}^μ , such that

$$T_g : (\psi^i, x^\mu) \rightarrow (\psi'^i, x'^\mu) \equiv (\mathbb{F}^i(\psi, x, g), \mathbb{X}^\mu(x, g)) , \quad g \in G . \quad (12.1)$$

The functions \mathbb{F}^i and \mathbb{X}^μ cannot be chosen arbitrarily, but have to respect the structure of G . First, the unit element $e \in G$ must be realized by an identity transformation,

$$T_e = \text{id} \quad \Leftrightarrow \quad \mathbb{F}^i(\psi, x, e) = \psi^i , \quad \mathbb{X}^\mu(x, e) = x^\mu . \quad (12.2)$$

Second, the group multiplication law dictates that for any $g_1, g_2 \in G$,

$$T_{g_1 g_2} = T_{g_1} \circ T_{g_2} \quad \Leftrightarrow \quad \begin{cases} \mathbb{F}^i(\psi, x, g_1 g_2) = \mathbb{F}^i(\mathbb{F}(\psi, x, g_2), \mathbb{X}(x, g_2), g_1) , \\ \mathbb{X}^\mu(x, g_1 g_2) = \mathbb{X}^\mu(\mathbb{X}(x, g_2), g_1) . \end{cases} \quad (12.3)$$

Finally, for any $g \in G$ the map $T_{g^{-1}}$ must be the inverse of T_g ,

$$T_{g^{-1}} = (T_g)^{-1} \quad \Leftrightarrow \quad \begin{cases} \mathbb{F}^i(\mathbb{F}(\psi, x, g), \mathbb{X}(x, g), g^{-1}) = \psi^i , \\ \mathbb{X}^\mu(\mathbb{X}(x, g), g^{-1}) = x^\mu . \end{cases} \quad (12.4)$$

The special case of (12.1) where $\psi'^i = \mathbb{F}^i(\psi, g)$ and $\mathbb{X}^\mu(x, g) = x^\mu$ corresponds to an internal symmetry.² Another special case is a *purely* spacetime transformation,

² Strictly speaking, my previous definition of a spacetime symmetry included the assumption that $\mathbb{X}^\mu(x, g) \neq x^\mu$, at least for some $x \in M$. This assumption is however immaterial in the present context. It will actually be convenient to think of the formalism developed here as a generalization rather than a modification of that in Part III of the book.

where $\mathbb{F}^i(\psi, x, g) = \psi^i$ for all $\psi \in \mathcal{M}$ and $x \in M$. It follows from the conditions on the functions \mathbb{X}^μ in (12.2)–(12.4) that purely spacetime transformations form a subgroup of the symmetry group G . Here I will make a technical assumption that will prove essential for setting up standard coordinates on the product manifold $\mathcal{M} \times M$. Namely, I will assume that the subgroup of G of purely spacetime transformations acts transitively on every slice $\{\psi\} \times M$ with fixed $\psi \in \mathcal{M}$. Loosely speaking, this requires that the spacetime manifold M has sufficient symmetry that turns it into a homogeneous space. Now choose an arbitrary $x_0 \in M$ and keep it fixed. By our assumption, for any $x \in M$ there is a purely spacetime transformation $\mathcal{T}_{x_0 \rightarrow x} \in G$ that maps x_0 to x . Typically, there will be multiple such transformations; we then have to choose one $\mathcal{T}_{x_0 \rightarrow x}$ by convention. For flat spacetimes, it is natural to use a translation connecting the two points. However, we do not have to be that specific at this stage, keeping in mind that M may be curved.

Eventually, we would like to treat ψ^i as functions on the spacetime. Mathematically, this amounts to replacing the action of the symmetry group G on $\mathcal{M} \times M$ with one on maps $M \rightarrow \mathcal{M}$. Such an induced action is easy to write down thanks to the assumption that the transformation of the coordinates is independent of the fields. Thus, the fields $\psi^i(x)$ are mapped to $\psi'^i(x, g)$ such that

$$\begin{aligned} \psi'^i(\mathbb{X}(x, g), g) &= \mathbb{F}^i(\psi(x), x, g) , \quad \text{or equivalently} \\ \psi'^i(x, g) &= \mathbb{F}^i(\psi(\mathbb{X}(x, g^{-1})), \mathbb{X}(x, g^{-1}), g) , \end{aligned} \quad (12.5)$$

for any $g \in G$. We will need (12.5) once we want to construct a G -invariant effective action for the fields. For the time being, however, I will pursue the analogy with Chap. 7 and focus on the action of G on the product manifold $\mathcal{M} \times M$.

12.3 Standard Nonlinear Realization

For every point $(\psi, x) \in \mathcal{M} \times M$ there is an isotropy group consisting of elements of G that map the point to itself,

$$H_{(\psi, x)} \equiv \{h \in G \mid T_h(\psi, x) = (\psi, x)\} . \quad (12.6)$$

The setup of nonlinear realization of internal symmetry as reviewed in Sect. 12.1 would naively suggest that we now choose a point $(\psi_0, x_0) \in \mathcal{M} \times M$. In its neighborhood, we could then establish the standard nonlinear realization of G in the usual manner. There are however two problems with this naive approach. First, we eventually want to treat fields as functions, defined globally on the whole spacetime M , as long as global coordinates on M exist. Second, the fields should be functions of the *original*, physical coordinates x^μ , not of some new variables whose dependence on ψ^i and x^μ is beyond our control.

In order to reach these goals, we take an intermediate step, introducing an isotropy group of a chosen spacetime point,

$$H_x \equiv \{g \in G \mid \mathbb{X}^\mu(x, g) = x^\mu\} . \quad (12.7)$$

Obviously, $H_{(\psi, x)}$ is a subgroup of H_x for any $\psi \in \mathcal{M}$. Moreover, having fixed a reference point $x_0 \in M$, we can use the maps $\mathcal{T}_{x_0 \rightarrow x}$ to show that, in analogy with (7.7), isotropy groups at different spacetime points are related by conjugation,

$$H_x = \mathcal{T}_{x_0 \rightarrow x} H_{x_0} \mathcal{T}_{x_0 \rightarrow x}^{-1} , \quad H_{(\psi, x)} = \mathcal{T}_{x_0 \rightarrow x} H_{(\psi, x_0)} \mathcal{T}_{x_0 \rightarrow x}^{-1} . \quad (12.8)$$

Example 12.1

Let ψ be a complex Schrödinger field so that $\mathcal{M} \simeq \mathbb{C}$, and let M be the flat Galilei spacetime. The action (11.43) of a Galilei boost with velocity \mathbf{v} corresponds to

$$(\psi, \mathbf{x}, t) \xrightarrow{\mathbf{v}} (\exp[i\mathbf{m}(\mathbf{v} \cdot \mathbf{x} + \mathbf{v}^2 t/2)]\psi, \mathbf{x} + \mathbf{v}t, t) . \quad (12.9)$$

We assume that in addition to boosts, the symmetry group G also includes spacetime translations, spatial rotations, and the internal $U(1)$ symmetry of phase transformations. For $x_0^\mu = (x_0, t_0) = (\mathbf{0}, 0)$, we find $H_{x_0} \simeq [\text{SO}(d) \ltimes \mathbb{R}^d] \times U(1)$, consisting of spatial rotations, Galilei boosts, and the phase transformations. In the special case of $\psi_0 = 0$, $H_{(\psi_0, x_0)} \simeq H_{x_0}$. Otherwise, for any $\psi_0 \neq 0$, $H_{(\psi_0, x_0)} \simeq \text{SO}(d) \ltimes \mathbb{R}^d \subsetneq H_{x_0}$. The isotropy groups for other spacetime points are obtained by conjugation (12.8) using the spacetime translation $\mathcal{T}_{x_0 \rightarrow x}(\psi, \mathbf{x}', t') \equiv (\psi, \mathbf{x}' + \mathbf{x} - \mathbf{x}_0, t' + t - t_0)$.³

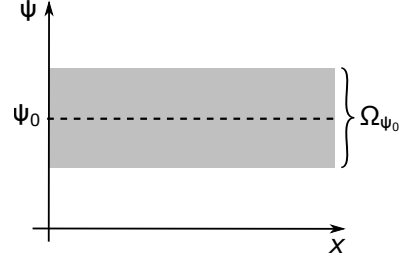
For any fixed $x \in M$, the transformation of (ψ^i, x^μ) under $g \in H_x$ is determined by the functions $\mathbb{F}^i(\psi, x, g)$. In other words, these functions define an action of H_x on the slice $\mathcal{M} \times \{x\} \simeq \mathcal{M}$. This makes it possible, following now Chap. 7 verbatim, to introduce a standard set of field coordinates on \mathcal{M} . Together with the spacetime coordinates x^μ , this defines a parameterization of the whole manifold $\mathcal{M} \times M$, in which the action of G takes a standard form. In order to keep the main results of this chapter together in one place, I will first give a concise overview of the ensuing standard realization of spacetime symmetry. Afterwards, I will stress some of its subtleties. A number of examples is worked out in detail in Sect. 12.4; the reader may want to consult these alongside the formal construction developed below.

12.3.1 Summary of the Construction

Consider the action of a continuous group G on the product manifold $\mathcal{M} \times M$ via a set of point transformations of the type (12.1). It is assumed that G includes a set of purely spacetime transformations that act transitively on every slice $\{\psi\} \times M$ with

³ In order to avoid cluttered notation, I use here and in the following the same symbol $\mathcal{T}_{x_0 \rightarrow x}$ to denote both the element of G and the corresponding map on $\mathcal{M} \times M$.

Fig. 12.2 Schematic visualization of the domain (*shaded area*) on the manifold $\mathcal{M} \times \mathcal{M}$ where the standard coordinates $(\pi^a, \chi^\varrho, x^\mu)$ are well-defined. The domain $\Omega_{\psi_0} \subset \mathcal{M}$ carries a local coordinate patch with the field variables (π^a, χ^ϱ)



fixed $\psi \in \mathcal{M}$. I will for simplicity choose the reference point $x_0 \in M$ with $x_0^\mu = 0$ and denote as $\mathcal{T}_x \equiv \mathcal{T}_{0 \rightarrow x}$ the purely spacetime transformation that maps x_0 to $x \in M$.

The slice $\mathcal{M} \times \{0\} \simeq \mathcal{M}$ carries an action of the isotropy group H_0 . Choose a fixed point $\psi_0 \in \mathcal{M}$. Assuming that the isotropy subgroup $H_{(\psi_0, 0)}$ is compact, it is always possible to find coordinates (π^a, χ^ϱ) on $\mathcal{M} \times \{0\}$ with the following properties:

- The point ψ_0 corresponds to $(\pi^a, \chi^\varrho) = (0, 0)$. Also, the set of points $\{(\pi^a, 0)\}$ spans a submanifold of $\mathcal{M} \times \{0\}$, equivalent to the coset space $H_0/H_{(\psi_0, 0)}$.
- The coordinates π^a parameterize uniquely a representative $U(\pi) \in H_0$ of the corresponding left coset of $H_{(\psi_0, 0)}$ in H_0 such that $U(0) = e$.
- The subgroup $H_{(\psi_0, 0)}$ acts on $H_0/H_{(\psi_0, 0)}$ by adjoint action, $U(\pi) \rightarrow hU(\pi)h^{-1}$ with $h \in H_{(\psi_0, 0)}$, which induces a linear transformation of the coordinates π^a .
- The whole group H_0 acts on $\mathcal{M} \times \{0\}$ by left multiplication as

$$\begin{aligned} U(\pi) &\xrightarrow{g_0} U(\pi'(\pi, g_0)) = g_0 U(\pi) h(\pi, g_0)^{-1}, \\ \chi^\varrho &\xrightarrow{g_0} \chi'^\varrho(\chi, \pi, g_0) = D(h(\pi, g_0))^\varrho{}_\sigma \chi^\sigma, \end{aligned} \quad (12.10)$$

where $g_0 \in H_0$ and $h(\pi, g_0) \in H_{(\psi_0, 0)}$. Moreover, $D(h)$ is a matrix representation of $H_{(\psi_0, 0)}$. Altogether, the action of H_0 on $\mathcal{M} \times \{0\}$ is fixed by the choice of representation D of $H_{(\psi_0, 0)}$ and the choice of parameterization $U(\pi)$ of the coset space $H_0/H_{(\psi_0, 0)}$.

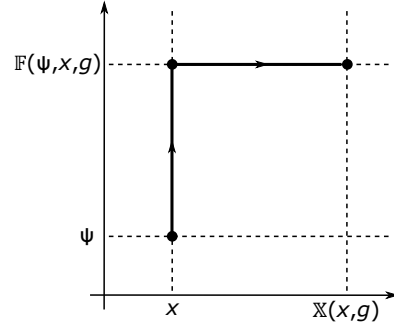
This finishes the setup of coordinates and the action of H_0 on the domain $\Omega_{\psi_0} \times \{0\}$; Ω_{ψ_0} is a neighborhood of ψ_0 in \mathcal{M} where the coordinates (π^a, χ^ϱ) are well-defined. As the next step, we transport the coordinates (π^a, χ^ϱ) to the whole spacetime using the maps \mathcal{T}_x . This leads to coordinates $(\pi^a, \chi^\varrho, x^\mu)$ on the domain $\Omega_{\psi_0} \times M$ (see Fig. 12.2). Intuitively, we require that all points of the slice $\{\psi\} \times M$ with fixed $\psi \in \Omega_{\psi_0}$ have the same coordinates π^a, χ^ϱ . Technically, we set

$$(\pi^a, \chi^\varrho, x^\mu) \equiv \mathcal{T}_x(\pi^a, \chi^\varrho, 0). \quad (12.11)$$

By (12.8), this fixes the action of the isotropy group H_x on $\mathcal{M} \times \{x\}$. Indeed, for any $g_x \in H_x$ there is a unique $g_0 \in H_0$ such that $g_x = \mathcal{T}_x g_0 \mathcal{T}_x^{-1}$. This leads to

$$\begin{aligned} T_{g_x}(\pi^a, \chi^\varrho, x^\mu) &= \mathcal{T}_x \circ T_{g_0}(\pi^a, \chi^\varrho, 0) = \mathcal{T}_x(\pi'^a(\pi, g_0), \chi'^\varrho(\chi, \pi, g_0), 0) \\ &= (\pi'^a(\pi, g_0), \chi'^\varrho(\chi, \pi, g_0), x^\mu). \end{aligned} \quad (12.12)$$

Fig. 12.3 For any fixed point $(\psi, x) \in \mathcal{M} \times M$, the action of a chosen element $g \in G$ can be composed from actions of an element of H_x and a purely spacetime transformation



Moreover, (12.8) guarantees that for any $g_x \in H_{(\psi_0, x)}$, $g_0 \in H_{(\psi_0, 0)}$. Therefore, the isotropy subgroup $H_{(\psi_0, x)}$ is realized on the slice $\mathcal{M} \times \{x\}$ by linear transformations of π^a and χ^ϱ for any $x \in M$.

The structure we already have extends uniquely to an action of the entire group G on the entire manifold $\mathcal{M} \times M$, or at least on $\Omega_{\psi_0} \times M$. Namely, note that the action of any $g \in G$ on a chosen point (ψ, x) can be composed of the action of an element of H_x and a purely spacetime transformation, see Fig. 12.3. Indeed, decompose

$$g = \mathcal{T}_{x \rightarrow \mathbb{X}(x, g)} \mathcal{T}_{x \rightarrow \mathbb{X}(x, g)}^{-1} g \equiv \mathcal{T}_{x \rightarrow \mathbb{X}(x, g)} g_x(x, g), \quad (12.13)$$

with the shorthand notation $\mathcal{T}_{x \rightarrow x'} \equiv \mathcal{T}_{x_0 \rightarrow x'} \mathcal{T}_{x_0 \rightarrow x}^{-1}$. The group element $g_x(x, g)$ maps (ψ^i, x^μ) to $(\mathbb{F}^i(\psi, x, g), x^\mu)$ for any $\psi \in \mathcal{M}$ and thus belongs to H_x . It corresponds to an element $g_0(x, g) \in H_0$ by $g_0(x, g) = \mathcal{T}_x^{-1} g_x(x, g) \mathcal{T}_x = \mathcal{T}_{\mathbb{X}(x, g)}^{-1} g \mathcal{T}_x$. Together with (12.12), this leads to the final expression for the action of G within $\Omega_{\psi_0} \times M$,

$$T_g(\pi^a, \chi^\varrho, x^\mu) = (\pi'^a(\pi, g_0(x, g)), \chi'^\varrho(\chi, \pi, g_0(x, g)), \mathbb{X}^\mu(x, g)). \quad (12.14)$$

This completes the setup of the nonlinear realization of G . I assumed an a priori knowledge of the transformation properties of the spacetime coordinates, that is the functions $\mathbb{X}^\mu(x, g)$. With this provision, the action of G is uniquely fixed by the structure of the isotropy groups H_0 and $H_{(\psi_0, 0)}$, and by the representation $D(h)$ of $H_{(\psi_0, 0)}$. In the process, we made some choices, including the choice of the representative $U(\pi)$ parameterized by π^a and of the element $\mathcal{T}_x \in G$ representing a spacetime point $x \in M$. These are just two sides of the same coin, pertinent respectively to the manifolds \mathcal{M} and M .

Complicated as it might seem, (12.14) actually encodes in a rather simple manner the action of G by left matrix multiplication. Indeed, let us represent the point $(\pi^a, \chi^\varrho, x^\mu)$ formally by $\mathcal{T}_x(U(\pi), \chi^\varrho)$. Then left multiplication by any $g \in G$ gives

$$\begin{aligned} T_g \circ \mathcal{T}_x(U(\pi), \chi^\varrho) &= \mathcal{T}_{\mathbb{X}(x, g)} \circ T_{g_0(x, g)}(U(\pi), \chi^\varrho) \\ &= \mathcal{T}_{\mathbb{X}(x, g)}(U(\pi'(\pi, g_0(x, g)))h(\pi, g_0(x, g))), \\ &\quad D(h(\pi, g_0(x, g)))^\varrho{}_\sigma \chi^\sigma, \end{aligned} \quad (12.15)$$

which exactly copies (12.14) in the matrix notation of (12.10). This shows that it is not actually necessary to decompose every $g \in G$ into a product of an element of H_x and a purely spacetime transformation as in (12.13). It is sufficient to observe that for any $x \in M$ and $g \in G$, $\mathcal{T}_{\mathbb{X}(x,g)}^{-1} g \mathcal{T}_x$ fixes the origin of M and thus belongs to H_0 .

12.3.2 Relation to Physics of Broken Spacetime Symmetry

Mathematically, the generalization of the standard nonlinear realization from internal to spacetime symmetry was relatively straightforward, perhaps even deceptively so. A number of its features thus deserve pointing out before we proceed to examples.

Universality of the Construction In Chap. 7 we found out that provided the isotropy group H is compact, the standard nonlinear realization captures all possible actions of the symmetry group G . Furthermore, I remarked that even if H is noncompact, the construction still goes through, but is no longer guaranteed to be exhaustive, provided the coset space G/H is reductive. The same applies here with the necessary modifications. The universality of the standard nonlinear realization is ensured if $H_{(\psi_0,0)}$ is compact. Even if it is not, the construction is still consistent, yet not necessarily exhaustive, if the coset space $H_0/H_{(\psi_0,0)}$ is reductive. Reference [3] gives an explicit example of a group action that cannot be obtained from the standard nonlinear realization. There $G \simeq \text{SO}(2,1)$ acts on $\mathcal{M} \simeq \mathbb{R}$, the isotropy subgroup being the noncompact affine group $\text{Aff}(1) \simeq \mathbb{R}^+ \ltimes \mathbb{R}$.

A possible point of concern might be that I assumed an a priori knowledge of the action of G on spacetime coordinates. This is however not a source of any hidden ambiguity. For one thing, the action of symmetry on spacetime coordinates is usually known. It is the classification of NG fields and their transformation properties that we are after. Moreover, even if unknown a priori, the functions $\mathbb{X}^\mu(x, g)$ can be constrained following the same philosophy. Namely, restricting (12.1) to the coordinates x^μ gives a group of well-defined transformations on M . If desired, possible forms of these transformations can be classified using the setup of Chap. 7.

Fate of Descendant Symmetries Recall the discussion of locally equivalent symmetries in Sect. 11.1. As explained therein, one can assume without loss of generality that the functions $f_{\mathfrak{A}}^A(x)$ vanish at the origin, $x^\mu = 0$. Provided the localized parent transformation in (11.1) does not involve any derivatives of $\epsilon_1^A(x)$, the action of the descendant symmetry at the origin vanishes. In the standard nonlinear realization, such a descendant symmetry therefore automatically belongs to $H_{(\psi_0,0)}$ and does not give rise to any NG variables π^a .

Physical Unbroken Symmetry In contrast to the internal symmetry case, a single point $(\psi, x) \in \mathcal{M} \times M$ cannot unambiguously represent the order parameter for SSB. The order parameter rather corresponds to a function $\varphi^i : M \rightarrow \mathcal{M}$. A particular value of the order parameter then defines a submanifold, $\mathcal{M}_\varphi \equiv \{(\varphi^i(x), x^\mu) \mid x \in M\} \subset \mathcal{M} \times M$. Accordingly, neither H_x nor $H_{(\varphi(x), x)}$ can be identified with the subgroup of unbroken symmetries, H_φ . The latter rather consists of all elements of

G that preserve the manifold \mathcal{M}_φ , that is map points on \mathcal{M}_φ to other points on \mathcal{M}_φ . With the help of (12.5), this translates to the condition $\varphi^i(\mathbb{X}(x, g)) = \mathbb{F}^i(\varphi(x), x, g)$ for all $x \in M$. This constraint is highly implicit; in practice it is usually much easier to check its infinitesimal form,

$$F_A^i(\varphi(x), x) - X_A^\mu(x) \partial_\mu \varphi^i(x) = 0 \quad \text{for all } x \in M. \quad (12.16)$$

The functions F_A^i and X_A^μ define the action of the generator Q_A of G . The reconstruction of the group H_φ from the solutions to (12.16) is based on the correspondence between a Lie group and its Lie algebra. An example of a possible pitfall this approach may fall into, in the context of isometries of a Riemannian manifold, is given in Appendix A.6.2.

Example 12.2

Suppose that the order parameter is constant, $\varphi^i(x) \equiv \varphi_0^i$. It follows that the “vacuum submanifold” is $\mathcal{M}_\varphi = \{\varphi_0\} \times M$. According to (12.1), the unbroken subgroup is

$$H_{\varphi_0} = \{g \in G \mid \mathbb{F}^i(\varphi_0, x, g) = \varphi_0^i \text{ for all } x \in M\}. \quad (12.17)$$

Any purely spacetime symmetry automatically belongs to H_{φ_0} . In other words, purely spacetime symmetries can only be broken by a coordinate-dependent order parameter. On the other hand, for internal symmetries, the condition in (12.17) reduces to $\mathbb{F}^i(\varphi_0, g) = \varphi_0^i$. Since internal symmetries by construction do not affect the coordinates x^μ , the definitions of $H_{(\varphi_0, x)}$ and H_{φ_0} in this special case coincide. At the same time, for internal symmetries we have trivially $H_x \simeq G$ for any $x \in M$, hence $H_x/H_{(\varphi_0, x)} \simeq G/H_{\varphi_0}$. This verifies that for internal symmetries, the more general formalism developed here boils down to that of Chap. 7.

Classification of Order Parameter Fluctuations For internal symmetries, independent fluctuations of the order parameter are in a one-to-one correspondence with the NG fields π^a . It would therefore be tempting to conclude that in the more general case of spacetime symmetries, they are classified by the coset space $H_0/H_{(\psi_0, 0)}$. The latter however only counts the NG variables π^a , realizing nonlinearly the action of H_0 . Our parameterization (12.11) of the manifold $\mathcal{M} \times M$ also involves the action of the “translations” \mathcal{T}_x . Being purely spacetime, these may be spontaneously broken if $\partial_\mu \varphi^i(x) \neq 0$ for some $x \in M$. We will only be able to introduce the corresponding NG modes in Chap. 13 once we treat ψ^i as maps $M \rightarrow \mathcal{M}$. The resulting EFT can be expected to feature a nontrivial realization of spacetime translations.

It is possible to be more explicit in the special case of a constant order parameter, $\varphi^i(x) = \varphi_0^i$, where purely spacetime symmetries remain unbroken. Here we expect a low-energy EFT in terms of NG fields π^a on which purely spacetime symmetries act trivially, $\pi'^a(\mathbb{X}(x, g), g) = \pi^a(x)$. The number of these NG fields is identified by the dimension of $H_0/H_{(\varphi_0, 0)}$. This generalizes our previous counting rule for NG fields of broken internal symmetry.

Example 12.3

Consider a theory of a single real relativistic scalar field ϕ that is invariant under the Galileon transformations,

$$(\phi, x^\mu) \xrightarrow{\epsilon, \epsilon_\mu} (\phi'(\phi, x, \epsilon), x'^\mu(x, \epsilon)) = (\phi + \epsilon + \epsilon_\mu x^\mu, x^\mu). \quad (12.18)$$

The symmetry group is $G \simeq \text{ISO}(d, 1) \ltimes \mathbb{R}^{D+1}$, where $\text{ISO}(d, 1)$ is the D -dimensional Poincaré group and the \mathbb{R}^{D+1} factor collects the constant and linear shifts of ϕ in (12.18). Suppose the dynamics of the theory is such that the ground state has a constant *vacuum expectation value* (VEV), $\langle \phi(x) \rangle \equiv \varphi$. This order parameter breaks by (12.16) the entire Galileon symmetry, that is $H_\varphi \simeq \text{ISO}(d, 1)$. The coset space $G/H_\varphi \simeq \mathbb{R}^{D+1}$ obviously does not correctly identify the NG modes. After all, our theory has a Lorentz-invariant vacuum and a single scalar field. Indeed, we find that $H_0 \simeq \text{SO}(d, 1) \ltimes \mathbb{R}^{D+1}$ where $\text{SO}(d, 1)$ is the Lorentz group. Likewise, $H_{(\varphi, 0)} \simeq \text{SO}(d, 1) \ltimes \mathbb{R}^D$. This only includes the linear shifts of ϕ in (12.18), which act trivially at $x^\mu = 0$. In the end, we thus find $H_0/H_{(\varphi, 0)} \simeq \mathbb{R}$, corresponding to the symmetry under constant shifts of ϕ . This gives the correct number of NG modes: one. The symmetry under linear shifts of ϕ is descendant. While it constrains the effective Lagrangian, cf. Sect. 11.3.1, it cannot affect the physical spectrum.

Domain of the Standard Nonlinear Realization As emphasized rather stubbornly, the standard field coordinates (π^a, χ^ϱ) are in principle only well-defined in some domain $\Omega_{\psi_0} \subset \mathcal{M}$. In order for our formalism to be able to capture all fluctuations of the order parameter, the manifold \mathcal{M}_φ had better lie entirely in $\Omega_{\psi_0} \times M$. This is obviously the case when the order parameter is constant as in Example 12.2. It is however not guaranteed in general and has to be checked case by case. In fact, it is easy to imagine a situation where this condition is very nontrivial. Consider a system with an order parameter $\varphi^i(x)$ where the isotropy group $H_{(\varphi(x), x)}$ varies from place to place. This is in contrast to our setup where $H_{(\psi_0, x)}$ for all $x \in M$ are mutually isomorphic. It may happen e.g. in systems where the order parameter for spontaneous breaking of internal symmetry vanishes at some spacetime points. We might then be lucky and still get away with our standard nonlinear realization. The price to pay is the presence of π^a -type variables that locally excite gapped modes in the spectrum, or χ^ϱ -type variables that locally excite NG modes. I am not aware of any work in the literature that would systematically address this kind of situation.

Example 12.4

A two-dimensional superfluid vortex is a map $\psi : \mathbb{R}^2 \rightarrow \mathbb{C}$, expressed in the polar coordinates ϱ, θ on \mathbb{R}^2 as $\psi(\mathbf{x}) = f(\varrho)e^{in\theta}$. Here $f(\varrho)$ is a strictly increasing profile function such that $f(0) = 0$ and the limit of $f(\varrho)$ for $\varrho \rightarrow \infty$ is finite. The nonzero parameter $n \in \mathbb{Z}$ is the so-called winding number of the vortex. The target manifold $\mathcal{M} \simeq \mathbb{C}$ carries an action of $G \simeq \text{U}(1)$ under which $\psi \xrightarrow{\epsilon} e^{i\epsilon}\psi$. This is an internal symmetry and thus $H_0 \simeq G \simeq \text{U}(1)$. For $\psi_0 = 0$, corresponding to the core of the

vortex, we find $H_{(\psi_0,0)} \simeq G \simeq \text{U}(1)$ as well. There are no NG variables, only two matter variables χ^ρ , equivalent to the real and imaginary parts of ψ . These are well-defined in the entire complex plane, hence $\Omega_{\psi_0} = \mathbb{C}$. On the other hand, choosing $\psi_0 \neq 0$ as appropriate for the vortex away from the core leads to $H_{(\psi_0,0)} \simeq \{e\}$. The coset space $H_0/H_{(\psi_0,0)} \simeq \text{U}(1)$ is now parameterized by a single NG variable π . The local parameterization of $\mathcal{M} \simeq \mathbb{C}$ is completed by one matter variable χ . One can view χ, π as polar coordinates on the target space. As such, their domain of validity Ω_{ψ_0} can be extended at most to \mathbb{C} with a half-line starting at the origin removed.

Some Literature on the Subject Nonlinear realizations of internal symmetry were classified in the pioneering work [4]. First attempts to generalize their results to spacetime symmetries appeared soon afterwards [5, 6]. These early works addressed a narrow class of relativistic systems where an extended symmetry such as the conformal symmetry is spontaneously broken to the Poincaré group. They followed an approach, based on an abstract coset space wherein all spontaneously broken symmetries induce relevant, independent degrees of freedom. Formally, this can be viewed as working with a set of order parameters that is sufficiently large to prevent any symmetries from being locally equivalent. Alternatively, it can be interpreted as being “agnostic” about the choice of order parameter. It is then necessary to include in the EFT the maximum possible set of order parameter fluctuations that can be enforced by a given symmetry-breaking pattern.

The price of this *agnostic nonlinear realization* is that it introduces extra degrees of freedom that at best correspond to gapped modes in the spectrum, and at worst are outright unphysical. Writing down an EFT solely in terms of the physical NG modes requires eliminating the superfluous degrees of freedom by an operational procedure known as *inverse Higgs constraint* (IHC) [7]. In spite of its obvious shortcomings, this framework has influenced the narrative surrounding spontaneous breaking of spacetime symmetries for decades. Much of the subsequent work focused on the machinery of IHCs and its physical interpretation. Namely, it has been known for some time that fluctuations of the order parameter, corresponding to locally equivalent symmetries, are redundant [8, 9]. Eliminating the would-be NG degrees of freedom corresponding to descendant symmetries by imposing a set of IHCs can then be interpreted as gauge-fixing [10]. The alternative possibility that the modes eliminated by the IHC may be physical but are necessarily gapped was recognized in [11, 12]. I will elaborate on the two scenarios later when we have at hand explicit examples. Finally, much work has been done on the question to what extent imposing an IHC might interfere with the universality of the standard nonlinear realization [13–15]. The moral is that a point transformation relating different parameterizations of an EFT may become nonlocal upon imposing the IHC.

As far as I know, the fact that the would-be NG fields of descendant symmetries need not be included in the nonlinear realization at all was first pointed out in [16]. The approach developed here makes this explicit by identifying the relevant NG variables π^a with coordinates on the coset space $H_0/H_{(\psi_0,0)}$. This has a simple classical interpretation, whereby π^a represent the deviation of the order parameter

from its local value. They can be further augmented with additional NG fields, arising from the spacetime variation of the order parameter. To incorporate such additional NG fields will be the most challenging problem we will have to face in Chap. 13.

12.4 Examples

To illustrate the general construction of Sect. 12.3, I will now work out in detail several examples, covering a range of different symmetries. In all the examples, I assume flat (relativistic or not) spacetime where the map $\mathcal{T}_{x \rightarrow x'}$ is realized by the unique translation that moves the point $x \in M$ to $x' \in M$. The isotropy group H_0 is fixed by the transformation properties of spacetime coordinates, known a priori. The subgroup $H_{(\psi_0, 0)}$, on the other hand, may depend sensitively on the choice of the reference point ψ_0 . This is demonstrated by the examples in Sects. 12.4.2 to 12.4.4.

12.4.1 Lorentz Scalars with Internal Symmetry

A good starting point is to check that our new algorithm can reproduce what we know from before about internal symmetries. Consider for simplicity a theory of a set of relativistic scalars ψ^i with a symmetry group $G \simeq \text{ISO}(d, 1) \times G_{\text{int}}$. Here $\text{ISO}(d, 1)$ is the Poincaré group, acting as a purely spacetime symmetry, and G_{int} is an internal symmetry. The actions of the purely spacetime and internal parts of G are completely independent of each other. Mathematically speaking, $\text{ISO}(d, 1)$ and G_{int} possess a well-defined action on, respectively, $M \simeq \mathbb{R}^D$ and \mathcal{M} . Both actions are trivially extended to $\mathcal{M} \times M$ by embedding.

It follows at once that $H_0 \simeq \text{SO}(d, 1) \times G_{\text{int}}$ and $H_{(\psi_0, 0)} \simeq \text{SO}(d, 1) \times H_{\text{int}}$, where $H_{\text{int}} \subset G_{\text{int}}$ is the isotropy group of the point $\psi_0 \in \mathcal{M}$. Unsurprisingly, we end up with a set of NG variables π^a , parameterizing the coset space $H_0/H_{(\psi_0, 0)} \simeq G_{\text{int}}/H_{\text{int}}$. By (12.10) and (12.14), a group element $(g_{\text{s.t.}}, g_{\text{int}}) \in \text{ISO}(d, 1) \times G_{\text{int}}$ acts on the standard coordinates $(\pi^a, \chi^\varrho, x^\mu)$ via

$$\begin{aligned} U(\pi) &\xrightarrow{(g_{\text{s.t.}}, g_{\text{int}})} U(\pi'(\pi, g_{\text{int}})) = g_{\text{int}} U(\pi) h_{\text{int}}(\pi, g_{\text{int}})^{-1}, \\ \chi^\varrho &\xrightarrow{(g_{\text{s.t.}}, g_{\text{int}})} \chi'^\varrho(\chi, \pi, g_{\text{int}}) = D(h_{\text{int}}(\pi, g_{\text{int}}))^\varrho_{\sigma} \chi^\sigma, \\ x^\mu &\xrightarrow{(g_{\text{s.t.}}, g_{\text{int}})} x'^\mu(x, g_{\text{s.t.}}) = (T_{g_{\text{s.t.}}} x)^\mu. \end{aligned} \quad (12.19)$$

Here $h_{\text{int}} \in H_{\text{int}}$ and $D(h_{\text{int}})$ is a matrix representation of H_{int} . Finally, $T_{g_{\text{s.t.}}}$ defines the action of the Poincaré group on the Minkowski spacetime. This is of course well-known, so the abstract notation just takes explicitly into account the possibility of using other coordinates x^μ than Minkowski.

This might look like a mere idiosyncratic reformulation of the standard non-linear realization of internal symmetry as laid out in Chap. 7. Indeed, should the order parameter be constant, $\varphi^i(x) \equiv \varphi_0^i$, the unbroken subgroup is $H_{\varphi_0} \simeq \text{ISO}(d, 1) \times H_{\text{int}}$ so that $G/H_{\varphi_0} \simeq G_{\text{int}}/H_{\text{int}}$, in accord with Example 12.2. However, the parameterization of $\mathcal{M} \times M$ by (π^a, χ^e, x^μ) and the corresponding group action (12.19) are also valid for order parameters $\varphi^i(x)$ with arbitrary coordinate dependence. The true unbroken subgroup H_φ can then be smaller than $\text{ISO}(d, 1) \times H_{\text{int}}$ and in extreme cases even trivial. This is not mere pedantry. For instance, dense relativistic matter can often be described by a time-dependent condensate of scalar fields. Although such a background spontaneously breaks boosts, our construction guarantees that one can parameterize its fluctuations by the same degrees of freedom as in a Lorentz-invariant vacuum. A similar remark applies to all the other examples discussed below.

The same setup can be used even if we replace $\text{ISO}(d, 1)$ with any other purely spacetime symmetry group that contains spacetime translations. The message is fairly simple: it does not matter what the spacetime symmetry is, as long as it does not act directly on the fields.

12.4.2 Lorentz Scalars with Scale Invariance

Next, we look at a simple example of a symmetry whose actions on coordinates and fields cannot be trivially separated. Consider a set of relativistic scalar fields ψ^i , carrying the action of $G \simeq \mathbb{R}^+ \ltimes \text{ISO}(d, 1)$. Here $\text{ISO}(d, 1)$ is again the Poincaré group. What is new is the factor \mathbb{R}^+ , representing scale transformations of the spacetime coordinates. It would be possible to add an internal symmetry factor G_{int} in the same way as in Sect. 12.4.1, but I will not do so to keep the notation simple.

It is now convenient to fix the spacetime coordinates x^μ as the standard Minkowski ones. The maps \mathcal{T}_x are then realized explicitly as $\mathcal{T}_x = e^{ix \cdot P}$ where P_μ is the generator of spacetime translations. The action of translations on the Minkowski coordinates is $x'^\mu(x, \epsilon) \equiv e^{i\epsilon \cdot P} x^\mu = x^\mu + \epsilon^\mu$. Likewise, an element $e^\alpha \in \mathbb{R}^+$ with real parameter α is realized on the coordinates as $x'^\mu(x, \alpha) \equiv e^{i\alpha D} x^\mu = e^\alpha x^\mu$, where D is the dilatation operator. Accordingly, the action of the dilatation group \mathbb{R}^+ on the Poincaré group is fixed by the commutation relations $[D, P_\mu] = -iP_\mu$ and $[D, J_{\mu\nu}] = 0$.

It is now clear that $H_0 \simeq \mathbb{R}^+ \times \text{SO}(d, 1)$. Since Lorentz transformations act trivially on scalar fields, we have only two options for $H_{(\psi_0, 0)}$, depending on whether or not ψ_0 also preserves scale invariance. Let us start with $H_{(\psi_0, 0)} \simeq \mathbb{R}^+ \times \text{SO}(d, 1)$, which corresponds to a scale-invariant order parameter $\psi_0 \in \mathcal{M}$. This implies $H_0/H_{(\psi_0, 0)} \simeq \{e\}$, hence there are no NG variables π^a , only matter fields. These should carry a linear representation of the dilatation symmetry. Since \mathbb{R}^+ only has one-dimensional irreducible representations, we can always choose a basis χ^e of

coordinates on $\mathcal{M} \times \{0\}$ such that $\chi'^{\varrho}(\chi, \alpha) \equiv e^{i\alpha D} \chi^{\varrho} = \exp(-\alpha \Delta_{\varrho}) \chi^{\varrho}$.⁴ The parameter Δ_{ϱ} is the scaling dimension of χ^{ϱ} . The action of dilatations on the entire manifold $\mathcal{M} \times M \simeq \mathcal{M} \times \mathbb{R}^D$ is then given by

$$(\chi^{\varrho}, x^{\mu}) \xrightarrow{\alpha} (\chi'^{\varrho}(\chi, \alpha), x'^{\mu}(x, \alpha)) = (\exp(-\alpha \Delta_{\varrho}) \chi^{\varrho}, e^{\alpha} x^{\mu}) . \quad (12.20)$$

It is instructive to see how this transformation rule is reproduced by (12.15). Namely, the commutator $[D, P_{\mu}] = -iP_{\mu}$ together with the Hadamard lemma (7.30) gives

$$e^{i\alpha D} P_{\mu} e^{-i\alpha D} = e^{\alpha} P_{\mu} \quad \text{or} \quad e^{i\alpha D} e^{i\epsilon \cdot P} e^{-i\alpha D} = \exp(i\epsilon^{\alpha} \epsilon \cdot P) . \quad (12.21)$$

This ensures that for $g = e^{i\alpha D}$, $g_0(x, g) = e^{i\alpha D}$ for any $x \in M$. The scale transformation of χ^{ϱ} is independent of the spacetime coordinate as it should. Mathematically, the conjugation relation (12.21) boils down to the fact that the translation generators P_{μ} carry a representation of the dilatation group \mathbb{R}^+ .

The other option for the isotropy group is $H_{(\psi_0, 0)} \simeq \text{SO}(d, 1)$, in which case $H_0/H_{(\psi_0, 0)} \simeq \mathbb{R}$. This can be thought of as arising from a Lorentz-invariant but dimensionful order parameter. We now have one NG variable, the *dilaton* π , plus possibly a set of $\dim \mathcal{M} - 1$ matter fields χ^{ϱ} . With the exponential parameterization, $U(\pi) = e^{i\pi D}$, (12.10) tells us that $\pi'(\pi, \alpha) = \pi + \alpha$ and $\chi'^{\varrho}(\chi, \pi, \alpha) = \chi^{\varrho}$. The extension of the action to other spacetime points works exactly the same as in the previous case. The final result for the action of dilatations therefore is

$$(\pi, \chi^{\varrho}, x^{\mu}) \xrightarrow{\alpha} (\pi'(\pi, \alpha), \chi'^{\varrho}(\chi, \pi, \alpha), x'^{\mu}(x, \alpha)) = (\pi + \alpha, \chi^{\varrho}, e^{\alpha} x^{\mu}) . \quad (12.22)$$

Note that the scaling dimension of all the matter fields is now zero. This is just a matter of a choice of variables. Namely, in presence of the dilaton, any field Ψ^{ϱ} with scaling dimension Δ_{ϱ} can be redefined to $\chi^{\varrho} = \exp(\Delta_{\varrho} \pi) \Psi^{\varrho}$.

I have not spelled out explicitly the action of spacetime translations and rotations. However, these are purely spacetime transformations and only affect the spacetime coordinates, similarly to the last line of (12.19).

12.4.3 Lorentz Vector with(out) Lorentz Scalar

Another possibility how to make the action of symmetry on coordinates and fields entangled is to keep the Poincaré group $G \simeq \text{ISO}(d, 1)$, but take a nonscalar field. Consider for simplicity a single Lorentz-vector field, A^{μ} , defining a D -dimensional target manifold $\mathcal{M} \simeq \mathbb{R}^D$. We then have $H_0 \simeq \text{SO}(d, 1)$, but the isotropy group $H_{(A_0, 0)}$ depends on the choice of A_0^{μ} . There are four qualitatively different options: $A_0^{\mu} = 0$, or nonvanishing A_0^{μ} that is respectively timelike, lightlike, or spacelike.

In the simplest case of $A_0^{\mu} = 0$, we find $H_{(A_0, 0)} \simeq H_0 \simeq \text{SO}(d, 1)$ and consequently $H_0/H_{(A_0, 0)} \simeq \{e\}$. There are no NG variables and the sole degree of freedom, A^{μ}

⁴ From now until the end of Chap. 12, a repeated index ϱ does not imply any summation.

itself, is of the matter type. It carries the vector representation of the Lorentz group, $A^\mu \rightarrow A'^\mu(A, g) = \Lambda(g)^\mu{}_\nu A^\nu$ for any $g \in \text{SO}(d, 1)$. Since the translation generators P_μ also carry the vector representation of the Lorentz group, we have a conjugation relation similar to (12.21) for dilatations. This guarantees that $g_0(x, g) = g$ for any $g \in \text{SO}(d, 1)$ and $x \in M \simeq \mathbb{R}^D$. From (12.15), we then extract the expected result for the action of Lorentz transformations,

$$T_g(A^\mu, x^\mu) = (\Lambda(g)^\mu{}_\nu A^\nu, \Lambda(g)^\mu{}_\nu x^\nu), \quad g \in \text{SO}(d, 1). \quad (12.23)$$

Spacetime translations only affect the coordinates, $x'^\mu(x, \epsilon) = e^{i\epsilon \cdot P} x^\mu = x^\mu + \epsilon^\mu$.

For illustration, I will work out in detail one more special case. Suppose that A_0^μ is nonzero and timelike. Without loss of generality, we can assume that $A_0^\mu = a\delta^{\mu 0}$ with $a \neq 0$. Then $H_{(A_0, 0)} \simeq \text{SO}(d)$, the group of spatial rotations. The coset space $H_0/H_{(A_0, 0)} \simeq \text{SO}(d, 1)/\text{SO}(d)$ is now a noncompact d -dimensional manifold, best viewed as the mass shell of a massive relativistic particle. A convenient implicit way to parameterize it is by treating it as the d -dimensional hyperboloid in $\mathcal{M} \simeq \mathbb{R}^D$ satisfying the constraint $A_\mu A^\mu = a^2$. That has the advantage of maintaining the linear transformation property (12.23) under all Lorentz transformations. This is just the noncompact version of the $\text{SO}(d+1)/\text{SO}(d) \simeq S^d$ coset space, conventionally parameterized by a unit vector $\mathbf{n} \in \mathbb{R}^{d+1}$.

If needed, it is possible to introduce explicit coordinates π^r on $H_0/H_{(A_0, 0)}$, transforming linearly as a vector under $H_{(A_0, 0)} \simeq \text{SO}(d)$. One suitable coordinatization arises from thinking of A^μ as the energy-momentum of a particle of mass $|a|$. The d independent coordinates π^r are then the components of its spatial momentum. A pure Lorentz boost can be represented by $e^{i\eta \cdot K}$, where K is the boost generator and η the rapidity. The latter is parallel to the velocity \mathbf{v} of the boost; its magnitude is determined implicitly by $|\mathbf{v}| = \tanh |\eta|$. The action of the boost on π^r in terms of the velocity \mathbf{v} reads

$$\pi'(\pi, \mathbf{v}) = \pi + \frac{\gamma_{\mathbf{v}} - 1}{\mathbf{v}^2} (\mathbf{v} \cdot \pi) \mathbf{v} + \gamma_{\mathbf{v}} \mathbf{v} \sqrt{a^2 + \pi^2} \text{sgn } a, \quad \gamma_{\mathbf{v}} \equiv \frac{1}{\sqrt{1 - \mathbf{v}^2}}. \quad (12.24)$$

The parameterization of the target manifold $\mathcal{M} \simeq \mathbb{R}^D$ is completed by adding to π^r a matter field χ , transforming in the singlet representation of $\text{SO}(d)$. By (12.15), it then automatically transforms trivially under the whole Poincaré group. The transformation of coordinates x^μ remains of course the same as in (12.23).

From internal symmetry, we are already used to the fact that the number and type of matter fields $\chi^{\mathcal{Q}}$ present depend on the choice of order parameter. It is only the NG fields π^a that are fixed by the symmetry-breaking pattern. Here comes the surprise: for spacetime symmetries, even the number of NG variables π^a may depend on the order parameter. For an example, take a theory of a relativistic complex scalar field ϕ where an internal $\text{U}(1)$ symmetry is broken by the VEV $\langle \phi(x) \rangle \equiv \varphi(x) = \varphi_0 e^{-i\mu x}$ with constant nonzero φ_0 and μ . This state describes matter with nonzero density of the $\text{U}(1)$ charge; μ is the

corresponding chemical potential. The state breaks the $G \simeq \text{ISO}(d, 1) \times \text{U}(1)$ symmetry spontaneously down to $H_\varphi \simeq \text{ISO}(d) \times \mathbb{R}$, where $\text{ISO}(d)$ now includes spatial translations and rotations and \mathbb{R} stands for a combination of time translations and internal $\text{U}(1)$ transformations. In line with our discussion in Sect. 12.4.1, there is a single NG variable, parameterizing the coset space $H_0/H_{(\varphi_0, 0)} \simeq [\text{SO}(d, 1) \times \text{U}(1)]/\text{SO}(d, 1)$. Now add a vector A^μ as a secondary order parameter. A constant background, $\langle A^\mu(x) \rangle \equiv A_0^\mu = a\delta^{\mu 0}$, can be interpreted for instance as the VEV of the current of the $\text{U}(1)$ symmetry. This does not affect the unbroken subgroup H_φ , yet it does add a vector π^r of NG variables. We will deal with this puzzle in Chap. 13. It will turn out that the number of physical gapless NG modes in the spectrum is still uniquely fixed by the symmetry-breaking pattern $G \rightarrow H_\varphi$. The new variables π^r represent either spurious, nondynamical degrees of freedom or dynamical but gapped excitations of the order parameter.

12.4.4 Schrödinger Scalars with Galilei Symmetry

To work out at least one nonrelativistic example, we now return to Galilei invariance. The action of the Galilei group on spacetime coordinates is well-known. A spatial translation by ϵ shifts Cartesian coordinates by $\mathbf{x} \rightarrow \mathbf{x} + \epsilon$, and likewise $t \rightarrow t + \epsilon$ represents a temporal translation by ϵ . Accordingly, the translation operator \mathcal{T}_x can be parameterized as $\mathcal{T}_{x,t} = e^{itH} e^{i\mathbf{x}\cdot\mathbf{P}}$, where H and \mathbf{P} are the respective generators. A Galilei boost with velocity \mathbf{v} acts on the coordinates as $(\mathbf{x}, t) \rightarrow (\mathbf{x} + \mathbf{v}t, t)$, and will be represented by $e^{i\mathbf{v}\cdot\mathbf{K}}$ with \mathbf{K} being the generator. One should also add the generator of spatial rotations, J_{rs} . However, since we will be mostly concerned with the boosts, I will not spell out the action of rotations explicitly.

From the transformation of the coordinates, we extract the conjugation property

$$e^{i\epsilon H} \mathbf{K} e^{-i\epsilon H} = \mathbf{K} - \epsilon \mathbf{P} \quad \text{or} \quad [H, \mathbf{K}] = i\mathbf{P} . \quad (12.25)$$

The actions of spatial translations and boosts on the coordinates commute with each other. However, it is known that representations of the Galilei symmetry admit a central extension of the commutator $[P_r, K_s]$. Let us therefore introduce a tentative central charge Q so that⁵

$$[P_r, K_s] = i\delta_{rs}Q \quad \text{or} \quad e^{i\epsilon\cdot\mathbf{P}} \mathbf{K} e^{-i\epsilon\cdot\mathbf{P}} = \mathbf{K} - \epsilon Q . \quad (12.26)$$

The resulting central extension of the Galilei group is known as the *Bargmann group*. This has the structure $G \simeq \text{SO}(d) \ltimes \{\mathbb{R}_K^d \ltimes [\mathbb{R}^D \times \text{U}(1)_Q]\}$. The first factor stands for spatial rotations, \mathbb{R}^D for spacetime translations, and \mathbb{R}_K^d for boosts. Finally, I use

⁵ In $d = 2$ spatial dimensions, the Galilei group admits another, exotic central extension, $[K_r, K_s] = i\kappa \epsilon_{rs}$. The parameter κ is however related to two-dimensional spin [17], and I will thus drop it.

the notation $U(1)_Q$ for the transformations generated by Q although at this stage it is not clear, or even important, that the symmetry is compact. What matters is that the action of Q on the coordinates is trivial. This ensures $H_0 \simeq [SO(d) \ltimes \mathbb{R}_K^d] \times U(1)_Q$.

The isotropy group $H_{(\psi_0,0)}$ depends sensitively on the type of fields included. I will initially restrict to rotation scalars, ψ^i , which makes the rotation $SO(d)$ into a purely spacetime symmetry. Moreover, the \mathbb{R}_K^d group of Galilei boosts is now descendant, as we saw in Sect. 11.3.3. This is confirmed by an explicit classification of low-spin indecomposable representations of the Galilei group [18]. In the end, the only part of the symmetry whose action on ψ^i at the origin, $(\mathbf{x}, t) = (\mathbf{0}, 0)$, may be nontrivial is the $U(1)_Q$. We have a freedom to decide whether or not $U(1)_Q$ belongs to $H_{(\psi_0,0)}$ by a suitable choice of the reference point ψ_0^i .

Let us first assume that $H_{(\psi_0,0)} \simeq H_0$ so that $H_0/H_{(\psi_0,0)} \simeq \{e\}$. It is then possible to find a set of *complex* field coordinates χ^ϱ that form one-dimensional linear representations of $U(1)_Q$, $\chi'^\varrho(\chi, \alpha) \equiv e^{i\alpha Q} \chi^\varrho = \exp(i\alpha m_\varrho) \chi^\varrho$, where m_ϱ are the corresponding charges of Q . This simple transformation property survives at all (\mathbf{x}, t) since Q is a central charge and so commutes with spacetime translations. Likewise, all of the spacetime translations and spatial rotations will act solely on the coordinates. The only nontrivial piece is the action of Galilei boosts. This is determined with the help of (12.15) and the relation

$$e^{i\mathbf{v}\cdot\mathbf{K}} e^{itH} e^{i\mathbf{x}\cdot\mathbf{P}} = e^{itH} e^{i(\mathbf{x}+\mathbf{v}t)\cdot\mathbf{P}} \exp\left[i\left(\mathbf{v}\cdot\mathbf{x} + \frac{1}{2}\mathbf{v}^2 t\right)Q\right] e^{i\mathbf{v}\cdot\mathbf{K}}, \quad (12.27)$$

which follows from (12.25) and (12.26). The final result is

$$e^{i\mathbf{v}\cdot\mathbf{K}}(\chi^\varrho, \mathbf{x}, t) = (\exp[i m_\varrho(\mathbf{v}\cdot\mathbf{x} + \mathbf{v}^2 t/2)] \chi^\varrho, \mathbf{x} + \mathbf{v}t, t), \quad (12.28)$$

which reproduces using solely the Lie algebra of the Bargmann group the transformation rule (12.9) I simply postulated before. We can see that the central charge Q , introduced above ad hoc, measures the nonrelativistic (rest) mass. In systems of identical particles of a fixed mass, this is proportional to the number of particles.

The other option for the isotropy group is $H_{(\psi_0,0)} \simeq SO(d) \ltimes \mathbb{R}_K^d$. This implies that $H_0/H_{(\psi_0,0)} \simeq U(1)_Q$. The particle number symmetry $U(1)_Q$ is going to be realized nonlinearly as for instance in superfluids. We need one NG variable, π . With the exponential parameterization of the coset space, $U(\pi) = e^{i\pi Q}$, this transforms under $U(1)_Q$ as $\pi'(\pi, \alpha) = \pi + \alpha$. In addition, there will be a set of *real* matter fields χ^ϱ , invariant under $U(1)_Q$. Using again (12.27), we then find quite a different result than above,

$$e^{i\mathbf{v}\cdot\mathbf{K}}(\pi, \chi^\varrho, \mathbf{x}, t) = (\pi + \mathbf{v}\cdot\mathbf{x} + \mathbf{v}^2 t/2, \chi^\varrho, \mathbf{x} + \mathbf{v}t, t). \quad (12.29)$$

This completes the range of options accessible with scalar fields. Let us see at least briefly what may happen when higher-spin fields are present. For simplicity, consider a single multiplet $A^\mu \equiv (A^0, \mathbf{A})$ that transforms under boosts as a Galilean vector [18], $A'^\mu(A, \mathbf{v}) \equiv e^{i\mathbf{v}\cdot\mathbf{K}}(A^0, \mathbf{A}) = (A^0, \mathbf{A} + \mathbf{v}A^0)$. Similarly to Sect. 12.4.3, one can think of $\langle A^\mu(x) \rangle \equiv A_0^\mu(x)$ as the VEV of the density and current of the

$U(1)_Q$ symmetry, respectively. In this interpretation, a state with $A_0^\mu(x) = (a, \mathbf{0})$ and constant $a \neq 0$ corresponds to uniform dense matter in its rest frame. With this secondary order parameter, the isotropy group $H_{((\psi_0, A_0), 0)}$ is reduced to $SO(d) \times U(1)_Q$ or $SO(d)$, depending on whether or not the primary order parameter ψ_0^i preserves $U(1)_Q$. The coset space $H_0/H_{((\psi_0, A_0), 0)}$ then necessarily includes a noncompact d -dimensional submanifold \mathbb{R}_K^d , carrying a vector of NG variables, ξ^r . The natural parameterization is $U(\xi) = e^{i\xi \cdot K}$. Following our algorithm for the standard nonlinear realization, we find that (12.28) and (12.29) remain valid under the respective assumptions on $U(1)_Q$. The only change is a new transformation rule for ξ^r ,

$$U(\xi'(\xi, \mathbf{v})) = e^{i\mathbf{v} \cdot \mathbf{K}} U(\xi) = U(\xi + \mathbf{v}) . \quad (12.30)$$

The modification of the realization of the Bargmann group by adding ξ^r is nearly trivial, so why exactly have we done this? The reason will become clear in the next chapter where we address the problem of construction of invariant actions. Namely, invariance under the linearly realized isotropy subgroup $H_{(\psi_0, 0)}$ has to be ensured by brute force using representation theory. The mathematical structure of the Bargmann group makes this a difficult task. With Galilei boosts realized nonlinearly by the NG field ξ^r , all that remains is to impose invariance under spatial rotations and possibly $U(1)_Q$, which is straightforward.

References

1. H. Bacry, J.M. Lévy-Leblond, J. Math. Phys. **9**, 1605 (1968). DOI [10.1063/1.1664490](https://doi.org/10.1063/1.1664490)
2. E. Bergshoeff, J. Figueroa-O'Farrill, J. Gomis, SciPost Phys. Lect. Notes **69** (2023). DOI [10.21468/SciPostPhysLectNotes.69](https://doi.org/10.21468/SciPostPhysLectNotes.69)
3. A. Joseph, A.I. Solomon, J. Math. Phys. **11**, 748 (1970). DOI [10.1063/1.1665205](https://doi.org/10.1063/1.1665205)
4. S.R. Coleman, J. Wess, B. Zumino, Phys. Rev. **177**, 2239 (1969). DOI [10.1103/PhysRev.177.2239](https://doi.org/10.1103/PhysRev.177.2239)
5. D.V. Volkov, Sov. J. Part. Nucl. **4**, 1 (1973)
6. V.I. Ogievetsky, Acta Univ. Wratislav. **207**, 117 (1974)
7. E. Ivanov, V.I. Ogievetsky, Theor. Math. Phys. **25**, 1050 (1975)
8. I. Low, A.V. Manohar, Phys. Rev. Lett. **88**, 101602 (2002). DOI [10.1103/PhysRevLett.88.101602](https://doi.org/10.1103/PhysRevLett.88.101602)
9. H. Watanabe, H. Murayama, Phys. Rev. Lett. **110**, 181601 (2013). DOI [10.1103/PhysRevLett.110.181601](https://doi.org/10.1103/PhysRevLett.110.181601)
10. A. Nicolis, R. Penco, F. Piazza, R.A. Rosen, JHEP **11**, 055 (2013). DOI [10.1007/JHEP11\(2013\)055](https://doi.org/10.1007/JHEP11(2013)055)
11. S. Endlich, A. Nicolis, R. Penco, Phys. Rev. **D89**, 065006 (2014). DOI [10.1103/PhysRevD.89.065006](https://doi.org/10.1103/PhysRevD.89.065006)
12. T. Brauner, H. Watanabe, Phys. Rev. **D89**, 085004 (2014). DOI [10.1103/PhysRevD.89.085004](https://doi.org/10.1103/PhysRevD.89.085004)
13. P. Creminelli, M. Serone, G. Trevisan, E. Trincherini, JHEP **02**, 037 (2015). DOI [10.1007/JHEP02\(2015\)037](https://doi.org/10.1007/JHEP02(2015)037)
14. R. Klein, D. Roest, D. Stefanyshyn, JHEP **10**, 051 (2017). DOI [10.1007/JHEP10\(2017\)051](https://doi.org/10.1007/JHEP10(2017)051)
15. B. Finelli, JHEP **03**, 075 (2020). DOI [10.1007/JHEP03\(2020\)075](https://doi.org/10.1007/JHEP03(2020)075)
16. I. Kharuk, A. Shkerin, Phys. Rev. **D98**, 125016 (2018). DOI [10.1103/PhysRevD.98.125016](https://doi.org/10.1103/PhysRevD.98.125016)
17. R. Jackiw, V.P. Nair, Phys. Lett. **B480**, 237 (2000). DOI [10.1016/S0370-2693\(00\)00379-8](https://doi.org/10.1016/S0370-2693(00)00379-8)
18. M. de Montigny, J. Niederle, A.G. Nikitin, J. Phys. **A39**, 9365 (2006). DOI [10.1088/0305-4470/39/29/026](https://doi.org/10.1088/0305-4470/39/29/026)

Chapter 13

Broken Spacetime Symmetry in Quantum Matter

With the nonlinear realization of spacetime symmetry put forward in Chap. 12, we are now in the position to construct the low-energy *effective field theory* (EFT). As I already occasionally stressed, the EFT machinery for spontaneously broken spacetime symmetry is less developed than that for internal symmetry. As a consequence, I will not be able to offer an explicit general expression for the effective Lagrangian akin to what I did in Chap. 8. Instead, I will describe the basic framework in an algorithmic fashion, and then work out some illustrative examples.

In this chapter, I only consider systems where the symmetry group G is known and dictated by the microscopic dynamics. This has been the underlying assumption of the whole book so far. Yet, there are physical systems whose low-energy physics may feature additional symmetry which itself may but need not be spontaneously broken. Such symmetry is usually called *emergent*, and appears among others in classical systems such as fluids or solids. I will defer a detailed discussion of such emergent symmetries and their spontaneous breaking to the next chapter.

Following some preliminary work in Sect. 13.1, the survey of EFTs for spontaneously broken spacetime symmetry starts in Sect. 13.2. Here I deal with systems where the values of the order parameter at all points in space belong to the same orbit of the symmetry group. This requires a minimal modification of the setup for internal symmetries, but already involves some of the subtleties associated with the spacetime ones. A major issue is that the number of *Nambu–Goldstone* (NG) variables π^a depends on the choice of order parameter realizing the given symmetry-breaking pattern. This means that some would-be NG variables are not required by the symmetry-breaking pattern per se. In Sect. 12.3.2, I suggested that such fields may, but need not, be physical. In case they are, they necessarily excite gapped modes in the spectrum, not true NG modes. I illustrate this dichotomy by several examples in Sect. 13.3. Then I offer some comments on the *inverse Higgs constraints* (IHCs) as an operational prescription for eliminating such spurious degrees of freedom.

The generic case of systems where the order parameter may depend on coordinates in an a priori arbitrary manner is considered in Sect. 13.4. Here the methods available for the construction of EFT appear to be least developed. I will therefore have to

content myself with a careful discussion of the special case of one-dimensional order parameter modulation, augmented with a couple of illustrative examples.

13.1 Building Blocks for Construction of Effective Actions

In Chap. 8, I showed that the effective action for NG bosons of spontaneously broken internal symmetry can always be expressed in terms of the associated *Maurer–Cartan* (MC) form. I am not aware of any proof of the equivalent statement for spacetime symmetries. I will however make the common assumption that this is still the case.

13.1.1 Maurer–Cartan Form

Recall that we deal generally with a set of bosonic fields ψ^i , taking values from a target manifold \mathcal{M} . The spacetime is treated as a possibly curved manifold M of dimension D , parameterized by local coordinates x^μ . Together, ψ^i and x^μ span the product manifold $\mathcal{M} \times M$. In order to avoid excessive repetition, I will not review all the details of the standard nonlinear realization of spacetime symmetry. The reader is invited to consult Sect. 12.3 to refresh their memory. Let me just stress the important roles of the isotropy groups H_x of a spacetime point $x \in M$, and $H_{(\psi, x)}$ of the point $(\psi, x) \in \mathcal{M} \times M$. For chosen fixed $\psi_0 \in \mathcal{M}$, the coset space $H_0/H_{(\psi_0, 0)}$ spans a submanifold of $\mathcal{M} \times \{0\} \simeq \mathcal{M}$, parameterized by NG coordinates π^a . These are generally accompanied by a set of matter variables χ^ϱ . Together, $(\pi^a, \chi^\varrho, x^\mu)$ constitute of a complete set of local coordinates on $\mathcal{M} \times M$.

Let us first focus on the variables (π^a, x^μ) , *required* for a successful nonlinear realization of spacetime symmetry. I will represent these jointly as

$$\mathfrak{U}(\pi, x) \equiv \mathcal{T}_x U(\pi) , \quad (13.1)$$

where \mathcal{T}_x is a fixed purely spacetime transformation that transports the spacetime origin to the point x (cf. Sect. 12.2). Moreover, $U(\pi)$ is a matrix representative of an element of the coset space $H_0/H_{(\psi_0, 0)}$. By (12.15), the action of any $g \in G$ on (π^a, x^μ) can then be expressed compactly as

$$\mathfrak{U}(\pi, x) \xrightarrow{g} \mathfrak{U}(\pi'(\pi, g_0(x, g)), \mathbb{X}(x, g)) = g \mathfrak{U}(\pi, x) h(\pi, g_0(x, g))^{-1} , \quad (13.2)$$

where $g_0(x, g) = \mathcal{T}_{\mathbb{X}(x, g)}^{-1} g \mathcal{T}_x$. Moreover, $\pi'^a(\pi, g_0(x, g))$ and $h(\pi, g_0(x, g))$ are defined by (12.10). The MC form can now be introduced through

$$\begin{aligned} \omega(\pi, x) &\equiv -i \mathfrak{U}(\pi, x)^{-1} d \mathfrak{U}(\pi, x) \\ &= -i U(\pi)^{-1} d U(\pi) + U(\pi)^{-1} (-i \mathcal{T}_x^{-1} d \mathcal{T}_x) U(\pi) . \end{aligned} \quad (13.3)$$

The first piece that depends solely on π^a is familiar from our analysis of internal symmetries. The second piece is new and in some, as yet unclear, way reflects the action of the symmetry on the spacetime. Equation (13.2) induces the following transformation of the MC form,

$$\begin{aligned}\omega(\pi, x) &\xrightarrow{g} \omega(\pi'(\pi, g_0), \mathbb{X}(x, g)) \\ &= h(\pi, g_0)\omega(\pi, x)h(\pi, g_0)^{-1} - ih(\pi, g_0)dh(\pi, g_0)^{-1},\end{aligned}\quad (13.4)$$

where I for brevity dropped the arguments of $g_0(x, g)$.

In order to understand better the structure of the MC form, we need to make a digression. By definition of spacetime symmetry, the transformation of coordinates x^μ is independent of the field variables ψ^i (or π^a, χ^ϱ). One can thus restrict the action of G to the spacetime manifold, where it is defined by the maps $x^\mu \xrightarrow{g} x'^\mu \equiv \mathbb{X}^\mu(x, g)$. Under this restricted action, the spacetime behaves as a homogeneous space, $M \simeq G/H_0$. This implies the following *local* homeomorphism of manifolds,

$$G \simeq H_{(\psi_0, 0)} \times H_0/H_{(\psi_0, 0)} \times G/H_0 \simeq H_{(\psi_0, 0)} \times H_0/H_{(\psi_0, 0)} \times M. \quad (13.5)$$

With the obvious notation for the Lie algebras of the three groups G , H_0 and $H_{(\psi_0, 0)}$, the corresponding isomorphism of tangent spaces reads

$$\mathfrak{g} \simeq \mathfrak{h}_{(\psi_0, 0)} \oplus \mathfrak{h}_0/\mathfrak{h}_{(\psi_0, 0)} \oplus \mathfrak{g}/\mathfrak{h}_0. \quad (13.6)$$

Here $\mathfrak{h}_0/\mathfrak{h}_{(\psi_0, 0)}$ and $\mathfrak{g}/\mathfrak{h}_0$ should be viewed as shorthand notation for the respective tangent spaces; these are themselves not Lie algebras without further qualifications.

I previously assumed that at the very least, the coset space $H_0/H_{(\psi_0, 0)}$ is reductive. This ensures that both $\mathfrak{h}_{(\psi_0, 0)}$ and $\mathfrak{h}_0/\mathfrak{h}_{(\psi_0, 0)}$ are invariant subspaces under the adjoint action of $H_{(\psi_0, 0)}$. In the following, I will need a stronger assumption, namely that all three components of the direct sum (13.6) are invariant under the adjoint action of $H_{(\psi_0, 0)}$. This is guaranteed if $H_{(\psi_0, 0)}$ is compact, but in general has to be viewed as an additional assumption.

Example 13.1

Let us check the validity of this assumption for the examples of nonlinear realization worked out in Sect. 12.4. In case of a theory of Lorentz scalars with an internal symmetry group G_{int} (Sect. 12.4.1), we have $H_{(\psi_0, 0)} \simeq \text{SO}(d, 1) \times H_{\text{int}}$. As long as the coset space $G_{\text{int}}/H_{\text{int}}$ itself is reductive, the tangent space $\mathfrak{h}_0/\mathfrak{h}_{(\psi_0, 0)} \simeq \mathfrak{g}_{\text{int}}/\mathfrak{h}_{\text{int}}$ is naturally invariant under the adjoint action of $H_{(\psi_0, 0)}$. Likewise, the tangent space $\mathfrak{g}/\mathfrak{h}_0$ of $G/H_0 \simeq M \simeq \mathbb{R}^D$ carries the vector representation of $\text{SO}(d, 1)$. It is easy to check that also for the examples in Sects. 12.4.2 and 12.4.3, our assumption on the decomposition (13.6) is satisfied.

The case of Galilei symmetry (Sect. 12.4.4) is most nontrivial. Here the commutation relation $[P_r, K_s] = i\delta_{rs}Q$ implies that the d -dimensional space spanned on the components of the momentum operator does not carry a representation of Galilei boosts. Hence the boosts should not be included in the isotropy subgroup $H_{(\psi_0, 0)}$.

The way out, as shown in Sect. 12.4.4, is to introduce a vector order parameter A^μ that gives a vector of NG variables ξ^r , carrying a nonlinear realization of the boosts. Then, $H_{((\psi_0, A_0), 0)}$ is either $\text{SO}(d)$ or $\text{SO}(d) \times \text{U}(1)_Q$. In both cases, the Lie algebra \mathfrak{g} of G naturally splits into invariant subspaces of $H_{((\psi_0, A_0), 0)}$ as in (13.6). This observation lends further support to the mathematical realization of Galilei boosts in terms of the auxiliary vector of NG fields ξ^r . I will return to this in Sect. 13.3.3.

In order to reformulate the above basic assumption in a language more common in physics, we need a notation for the symmetry generators. In analogy with the treatment of internal symmetries in Chap. 7, I will denote generators that form a basis of \mathfrak{g} as $Q_{A,B,\dots}$. A subset of these generators that spans a basis of $\mathfrak{h}_{(\psi_0, 0)}$ will be $Q_{\alpha, \beta, \dots}$. Analogously, the basis of $\mathfrak{h}_0/\mathfrak{h}_{(\psi_0, 0)}$ will be $Q_{a, b, \dots}$. Finally, for the basis of the complementary space $\mathfrak{g}/\mathfrak{h}_0$, I will use the notation $P_{\underline{\mu}, \underline{\nu}, \dots}$. In the fixed basis Q_A , the structure of the Lie algebra \mathfrak{g} is defined by the structure constants f_{AB}^C via $[Q_A, Q_B] = i f_{AB}^C Q_C$. Then, our assumption that all the components of the direct sum (13.6) carry a representation of $H_{(\psi_0, 0)}$ restricts the commutators of Q_α to

$$[Q_\alpha, Q_\beta] = i f_{\alpha\beta}^\gamma Q_\gamma, \quad [Q_\alpha, Q_b] = i f_{\alpha b}^c Q_c, \quad [Q_\alpha, P_{\underline{\mu}}] = i f_{\alpha \underline{\mu}}^\nu P_{\underline{\nu}}. \quad (13.7)$$

The remaining commutators of the Lie algebra \mathfrak{g} take the generic form

$$\begin{aligned} [Q_a, Q_b] &= i(f_{ab}^\gamma Q_\gamma + f_{ab}^c Q_c), & [Q_a, P_{\underline{\mu}}] &= i(f_{a \underline{\mu}}^\beta Q_\beta + f_{a \underline{\mu}}^b Q_b + f_{a \underline{\mu}}^\nu P_{\underline{\nu}}), \\ [P_{\underline{\mu}}, P_{\underline{\nu}}] &= i(f_{\underline{\mu}\underline{\nu}}^\alpha Q_\alpha + f_{\underline{\mu}\underline{\nu}}^a Q_a + f_{\underline{\mu}\underline{\nu}}^\lambda P_{\underline{\lambda}}). \end{aligned} \quad (13.8)$$

There is no $f_{ab}^\mu P_{\underline{\mu}}$ term in $[Q_a, Q_b]$ since all the Q_α and Q_a together span a basis of \mathfrak{h}_0 . As to $P_{\underline{\mu}}$, at this stage we do not need to assume that they are mutually commuting translation generators.

We are now ready to return to the MC form. By its definition (13.3), $\omega \equiv \omega^A Q_A$ is a (locally defined) 1-form on $H_0/H_{(\psi_0, 0)} \times M$ that takes values in the Lie algebra \mathfrak{g} of G . It consists of three parts, corresponding to the three spaces on the right-hand side of (13.6),

$$\begin{aligned} \omega &\equiv \omega_{\parallel} + \omega_{\perp} + \omega_P, \\ \omega_{\parallel} &\equiv \omega^\alpha Q_\alpha, & \omega_{\perp} &\equiv \omega^a Q_a, & \omega_P &\equiv e^{*\underline{\mu}} P_{\underline{\mu}}. \end{aligned} \quad (13.9)$$

The notation for ω_{\parallel} and ω_{\perp} follows the conventions introduced in Chap. 7 for internal symmetries. The ω_P component, proportional to the generators $P_{\underline{\mu}}$, is new here. Thanks to the assumption that both spaces $\mathfrak{h}_0/\mathfrak{h}_{(\psi_0, 0)}$ and $\mathfrak{g}/\mathfrak{h}_0$ are invariant under the adjoint action of $H_{(\psi_0, 0)}$, the transformation rule (13.4) splits as

$$\begin{aligned} \omega_{\parallel}(\pi, x) &\xrightarrow{g} h(\pi, g_0) \omega_{\parallel}(\pi, x) h(\pi, g_0)^{-1} - i h(\pi, g_0) d h(\pi, g_0)^{-1}, \\ \omega_{\perp}(\pi, x) &\xrightarrow{g} h(\pi, g_0) \omega_{\perp}(\pi, x) h(\pi, g_0)^{-1}, \\ \omega_P(\pi, x) &\xrightarrow{g} h(\pi, g_0) \omega_P(\pi, x) h(\pi, g_0)^{-1}. \end{aligned} \quad (13.10)$$

Following Chap. 7, it would now be possible to work out further technical details. These include for instance an explicit expression for the action of G on the coordinates (π^a, x^μ) , or the MC equations for the exterior derivative of $\omega(\pi, x)$. I will however not do so even though it would be straightforward, for we will not need these details. Let us instead turn to the physical interpretation of the various parts of the MC form.

13.1.2 Covariant Derivatives of Fields

The first part of the MC form in (13.3), $-iU(\pi)^{-1}dU(\pi)$, obviously takes values from the Lie algebra \mathfrak{h}_0 . It follows that ω_P does not contain any contributions proportional to $d\pi^a$. In other words, ω_P is a well-defined 1-form on the slice $\{\pi\} \times M \simeq M$ for any fixed point π^a in $H_0/H_{(\psi_0,0)}$,

$$e^{*\mu}(\pi, x) \equiv e_{\nu}^{*\mu}(\pi, x)dx^{\nu} . \quad (13.11)$$

By assumption, the action of G on the spacetime M is transitive. This implies that at $\pi^a = 0$, the components of $\omega_P(0, x) = (-i\mathcal{T}_x^{-1}d\mathcal{T}_x)^{\mu}P_{\mu}$ span the whole cotangent space to M . By continuity, the same must be true in some neighborhood of the origin of $H_0/H_{(\psi_0,0)}$, $\pi^a = 0$. We conclude that the 1-forms $e^{*\mu}(\pi, x)$ with fixed π^a define a (local) coframe on the spacetime manifold M . We are of course free to use any (local) frame or coframe on M we wish. Nevertheless, it will be convenient to stick to ω_P thanks to its covariant transformation properties under G , as shown in (13.10). This justifies a posteriori the notation $e^{*\mu}$ for the components of ω_P .

Next we have a look at the ω_{\perp} part of the MC form,

$$\omega^a(\pi, x) \equiv \omega_b^a(\pi)d\pi^b + \omega_{\mu}^a(\pi, x)dx^{\mu} . \quad (13.12)$$

The ω_b^a piece comes from the $-iU(\pi)^{-1}dU(\pi)$ term in (13.3) and is thus independent of the spacetime coordinates. The full 1-form ω_{\perp} however contains components from cotangent spaces to both $H_0/H_{(\psi_0,0)}$ and M . Unlike in the case of internal symmetries, it therefore cannot be interpreted as giving rise to a coframe on $H_0/H_{(\psi_0,0)}$.

At this point, recall that we eventually want to treat π^a as fields. These naturally define maps from M to $H_0/H_{(\psi_0,0)}$ that take any $x \in M$ to $\pi^a(x) \in H_0/H_{(\psi_0,0)}$. Slightly more formally, one can view the fields as maps from M to $H_0/H_{(\psi_0,0)} \times M$ that assign to $x \in M$ the pair $(\pi^a(x), x^{\mu})$. This makes it possible to pull differential forms on $H_0/H_{(\psi_0,0)} \times M$ back to M . Thus, the coframe (13.11) becomes $e^{*\mu}(\pi(x), x) = e_{\nu}^{*\mu}(\pi(x), x)dx^{\nu}$.¹ More interesting is the pull-back of (13.12) to M ,

$$\omega^a(\pi(x), x) = [\omega_b^a(\pi(x))\partial_{\mu}\pi^b(x) + \omega_{\mu}^a(\pi(x), x)]dx^{\mu} . \quad (13.13)$$

¹ I take the liberty to denote 1-forms on $H_0/H_{(\psi_0,0)} \times M$ and their pull-backs to M using the same symbol. It should be clear from the context which of the two is meant.

This obviously carries information about the derivatives of the NG fields $\pi^a(x)$. To extract the *covariant derivative* $\nabla_\mu \pi^a$ of the NG field, we decompose the spacetime 1-form $\omega^a(\pi(x), x)$ in the coframe $e^{*\mu}(\pi(x), x)$,

$$\omega^a \equiv (\nabla_\mu \pi^a) e^{*\mu} = (\nabla_\mu \pi^a) e_\nu^{*\mu} dx^\nu . \quad (13.14)$$

Comparison with (13.13) then gives

$$\nabla_\mu \pi^a(x) = [\omega_b^a(\pi(x)) \partial_\nu \pi^b(x) + \omega_\nu^a(\pi(x), x)] e_\mu^\nu(\pi(x), x) , \quad (13.15)$$

where $e_\mu^\nu(\pi, x) \equiv e_\mu^\nu(\pi, x) \partial_\nu$ is the local frame on M , dual to $e^{*\mu}(\pi, x)$. The covariant derivative can also be expressed compactly as $\nabla_\mu \pi^a = \omega^a(e_\mu)$.

I am using here the same symbol ∇_μ for the covariant derivative as in Appendix A.5, yet the two objects do not seem to be the same. Let me clarify the difference. First, the first term on the right-hand side of (13.15) is not just the gradient of π^a projected to the frame e_μ . The extra factor ω_b^a , as well as the ω_ν^a term, is needed to ensure covariance under the action of G . In particular the ω_b^a factor is nontrivial even if G is an internal symmetry group. Second, we are not introducing here an all-purpose connection on M that would make $\nabla_\mu \pi^a$ covariant under the maximal structure group $\text{GL}(\dim M)$. As follows from (13.10), we only allow changes of the local frame induced by a representation of the structure group $H_{(\psi_0, 0)}$. This procedure is designed to give a G -invariant EFT for the NG fields. In spite of using covariant derivatives, the EFTs constructed below are not covariant under general coordinate transformations on M .

The remaining part of the MC form (13.9) that we have not discussed yet is ω_\parallel . This, as the transformation rule (13.10) suggests, plays the role of an $\mathfrak{h}_{(\psi_0, 0)}$ -valued connection. Let us define the spacetime components of the connection after pull-back to M through

$$\omega_\parallel(\pi(x), x) \equiv \omega_\mu^\alpha(\pi(x), x) Q_\alpha dx^\mu . \quad (13.16)$$

Then for a matter field χ^ϱ that transforms linearly in a representation D of $H_{(\psi_0, 0)}$, cf. (12.10), the covariant derivative can be defined as

$$\nabla_\mu \chi^\varrho(x) \equiv [\partial_\nu \chi^\varrho(x) + i \omega_\nu^\alpha(\pi(x), x) D(Q_\alpha)^\varrho{}_\sigma \chi^\sigma(x)] e_\mu^\nu(\pi(x), x) . \quad (13.17)$$

This generalizes the covariant derivative we constructed in the context of internal symmetries in Sect. 8.2.4.

We conclude by checking in what precise sense the derivatives (13.15) and (13.17) of the NG and matter fields are covariant. To that end, consider a transformation by an element of $g \in G$ infinitesimally close to unity, $g \approx e + i\epsilon^A Q_A$, where ϵ^A is a set of small constant parameters. The element $h(\pi, g_0(x, g)) \in H_{(\psi_0, 0)}$, defined by (12.10) with $g_0(x, g) = \mathcal{T}_{\mathbb{X}(x, g)}^{-1} g \mathcal{T}_x$, then reads, to linear order in ϵ^A ,

$$h(\pi, g_0(x, g)) \approx e + i\epsilon^A k_A^\alpha(\pi, x) Q_\alpha . \quad (13.18)$$

Here $k_A^\alpha(\pi, x)$ is a set of functions that are calculable following the guidance of Sect. 7.3, but we need not do so explicitly. It is easy to check that the last two lines of (13.10) now translate into the infinitesimal transformations

$$\begin{aligned} \delta\omega^a(\pi, x) &= -\epsilon^A k_A^\alpha(\pi, x) f_{\alpha b}^a \omega^b(\pi, x) , \\ \delta e^{*\mu}(\pi, x) &= -\epsilon^A k_A^\alpha(\pi, x) f_{\alpha \nu}^\mu e^{*\nu}(\pi, x) . \end{aligned} \quad (13.19)$$

By the duality between the frame e_μ and the coframe $e^{*\mu}$, the transformation of the latter gives $\delta e_\mu = \epsilon^A k_A^\alpha f_{\alpha \mu}^\nu e_\nu$. Equipped with these auxiliary identities, we arrive at the infinitesimal transformations of the covariant derivatives (13.15) and (13.17),

$$\begin{aligned} \delta(\nabla_\mu \pi^a) &= \epsilon^A k_A^\alpha (-f_{\alpha b}^a \nabla_\mu \pi^b + f_{\alpha \mu}^\nu \nabla_\nu \pi^a) , \\ \delta(\nabla_\mu \chi^\varrho) &= \epsilon^A k_A^\alpha [iD(Q_\alpha)^\varrho{}_\sigma \nabla_\mu \chi^\sigma + f_{\alpha \mu}^\nu \nabla_\nu \chi^\varrho] . \end{aligned} \quad (13.20)$$

The covariant derivatives transform linearly. This makes it possible to construct G -invariant Lagrangian densities out of the (possibly higher) covariant derivatives using standard methods of tensor algebra and representation theory.

Example 13.2

Recall the example worked out in Sect. 12.4.2. Here $G \simeq \mathbb{R}^+ \ltimes \text{ISO}(d, 1)$ acts on a set of relativistic scalars, where $\text{ISO}(d, 1)$ is the Poincaré group and \mathbb{R}^+ represents scale transformations of spacetime coordinates. Suppose that the ground state of the system carries a uniform condensate, spontaneously breaking the scale invariance. We choose accordingly $\psi_0 \in \mathcal{M}$ that breaks the dilatation group \mathbb{R}^+ . Thus, $H_{(\psi_0, 0)} \simeq \text{SO}(d, 1)$ whereas $H_0 \simeq \mathbb{R}^+ \times \text{SO}(d, 1)$.

Let us disregard possible matter fields and focus on the dilaton field π , parameterizing the coset space $H_0/H_{(\psi_0, 0)} \simeq \mathbb{R}$. We fix x^μ to be the standard Minkowski coordinates, and P_μ the energy–momentum operator. The manifold $H_0/H_{(\psi_0, 0)} \times \mathcal{M} \simeq \mathbb{R}^{d+2}$ is naturally parameterized by $\mathfrak{U}(\pi, x) = e^{ix \cdot P} e^{i\pi D}$.² With the help of the commutation relation $[D, P_\mu] = -iP_\mu$, we compute the MC form

$$\omega(\pi, x) = -i\mathfrak{U}(\pi, x)^{-1} d\mathfrak{U}(\pi, x) = Dd\pi + e^{-\pi} P \cdot dx . \quad (13.21)$$

From the second term, we extract the spacetime coframe $e^{*\mu} = \delta_\nu^\mu e^{-\pi} dx^\nu$ and the corresponding dual frame, $e_\mu = \delta_\mu^\nu e^\pi \partial_\nu$. The dual frame can be used as a covariant derivative operator, ∇_μ , that is scale-invariant.

To produce an action, we still need a volume element. This is most easily obtained by constructing a volume form on the Minkowski spacetime using the coframe,

$$\text{vol} = e^{*0} \wedge \cdots \wedge e^{*d} = e^{-(d+1)\pi} dx^0 \wedge \cdots \wedge dx^d . \quad (13.22)$$

² In order to avoid notation clash with the dilatation operator D , I temporarily denote the dimension of spacetime as $d + 1$.

A generic Poincaré- and scale-invariant effective action for the dilaton then reads

$$S_{\text{eff}}\{\pi\} = \int d^{d+1}x e^{-(d+1)\pi} \mathcal{L}_{\text{eff}}(\nabla\pi, \nabla\nabla\pi, \dots) . \quad (13.23)$$

The Lagrangian \mathcal{L}_{eff} can be built by writing down the most general Lorentz-invariant operator in terms of the derivatives of π , and then replacing ∂_μ with ∇_μ everywhere.

13.2 Twisting Order Parameter for Internal Symmetry

We have reached the point where we can start dealing with concrete systems featuring spontaneous breakdown of spacetime symmetry. As mentioned above, I will always assume that the effective action can be constructed solely out of the MC form. Any claim of having found the most general invariant action therefore has to be interpreted accordingly. It is not always straightforward to construct the most general effective action even in this restricted sense. To make further progress, it is convenient to narrow down the landscape of systems that we consider.

In this section, I will elaborate on the example of nonlinear realization worked out in Sect. 12.4.1. Here is a brief reminder. Suppose that the symmetry group factorizes as $G \simeq G_{\text{s.t.}} \times G_{\text{int}}$, where “s.t.” and “int” stand respectively for “spacetime” and “internal.” Moreover, suppose that the target manifold \mathcal{M} is parameterized by a set of fields ψ^i that are scalar in the sense that $\mathbb{F}^i(\psi, x, g_{\text{s.t.}}) = \psi^i$ for all $g_{\text{s.t.}} \in G_{\text{s.t.}}$. This is equivalent to requiring that $G_{\text{s.t.}}$ is a purely spacetime symmetry. Likewise, G_{int} is internal in the sense that $\mathbb{F}^i(\psi, x, g_{\text{int}})$ is independent of x^μ and $\mathbb{X}^\mu(x, g_{\text{int}}) = x^\mu$ for all $g_{\text{int}} \in G_{\text{int}}$. The action of the symmetry group G on the standard coordinates (π^a, χ^a, x^μ) is then given by (12.19). The isotropy group of the spacetime origin is $H_0 \simeq H_{\text{s.t.}} \times G_{\text{int}}$, and the spacetime manifold \mathcal{M} is equivalent to the homogeneous space $G_{\text{s.t.}}/H_{\text{s.t.}}$. Similarly, the isotropy group of $(\psi_0, 0) \in \mathcal{M} \times \mathcal{M}$ is $H_{\text{s.t.}} \times H_{\text{int}}$, hence $H_0/H_{(\psi_0, 0)} \simeq G_{\text{int}}/H_{\text{int}}$.

I will now make one additional, important assumption. Let us choose the reference point $\psi_0 \in \mathcal{M}$ so that it represents the actual value of the order parameter at the spacetime origin, $\langle \psi^i(0) \rangle = \psi_0^i$. Suppose that in the resulting standard coordinates on \mathcal{M} , the *vacuum expectation value* (VEV) of all the matter fields, $\langle \chi^a(x) \rangle$, is zero. This is equivalent to requiring that the order parameter is fully specified by the VEVs $\langle \pi^a(x) \rangle$. The values of the order parameter at different spacetime points lie on the same orbit of G_{int} in \mathcal{M} . It is then mathematically consistent to eliminate the matter fields from the EFT altogether. After all, they are expected to excite gapped modes in the spectrum, which can be ignored at sufficiently low energies.

Example 13.3

Consider a theory of a single real relativistic scalar field ϕ , equipped with an internal shift symmetry, $(\phi, x^\mu) \xrightarrow{\epsilon} (\phi + \epsilon, x^\mu)$. This corresponds to $G_{\text{s.t.}} \simeq \text{ISO}(d, 1)$,

$H_{s.t.} \simeq \text{SO}(d, 1)$, and $G_{\text{int}} \simeq \mathbb{R}$. For any choice of the reference point ϕ_0 , the internal isotropy group is trivial, $H_{\text{int}} \simeq \{e\}$. In this case, the only degree of freedom, that is ϕ itself, is of the NG type. There are no matter-type variables. Therefore, our assumption that the VEV of all χ^ϱ be vanishing is trivially satisfied.

A slightly less trivial example is that of a complex relativistic scalar ϕ , subject to the action of $G_{\text{int}} \simeq \text{U}(1)$ via $(\phi, x^\mu) \xrightarrow{\epsilon} (e^{i\epsilon}\phi, x^\mu)$. For any nonzero $\phi_0 \in \mathbb{C}$, we have again $H_{\text{int}} \simeq \{e\}$. The field is then naturally parameterized by its modulus and phase, $\phi = \varrho e^{i\theta}$, where θ is the sole NG variable. Provided $\langle \varrho(x) \rangle \equiv \varrho_0$ is constant, ϱ can be traded for a matter field with vanishing VEV as desired, $\chi \equiv \varrho - \varrho_0$. On the other hand, the VEV of the phase, $\langle \theta(x) \rangle$, can have arbitrary coordinate dependence. This kind of order parameter describes relativistic superfluids.

Discarding the matter fields amounts to restricting our basic setup to the manifold $H_0/H_{(\psi_0,0)} \times M \simeq G_{\text{int}}/H_{\text{int}} \times G_{s.t.}/H_{s.t.}$. Henceforth, I will only consider flat spacetimes such that $G_{s.t.}$ possesses a set of D mutually commuting translation generators P_μ . With the parameterization $\mathcal{U}(\pi, x) = e^{ix \cdot P} U(\pi)$, the MC form becomes

$$\omega(\pi, x) = -iU(\pi)^{-1} dU(\pi) + P \cdot dx. \quad (13.24)$$

The spacetime coframe is trivial, $e^{*\mu} = \delta^\mu_\nu dx^\nu$, and it is not necessary to distinguish frame and coordinate-basis indices. The ω_\parallel and ω_\perp parts of the MC form (13.24) are fixed by the internal symmetry. We can therefore reuse the wealth of information accumulated in Chap. 7 to evaluate them.

So far, I have not made any assumptions about the coordinate dependence of the order parameter beyond the requirement that all $\langle \chi^\varrho(x) \rangle$ vanish. Whether or not the order parameter breaks any spacetime symmetries has no bearing on the set of degrees of freedom π^a of the EFT, their transformation under G , or the MC form. Yet, the devil is in the details, as we shall now see.

13.2.1 New Features of the Old Setup

A consistent low-energy EFT setup requires a number of ingredients to be in place. The first is of course having a set of well-defined degrees of freedom, well separated from whatever high-energy modes have been ignored. This is generally guaranteed by the broken symmetry. These low-energy degrees of freedom should map to positive-energy excitations of a stable ground state. Failure to satisfy this condition indicates that we have not identified the correct ground state. Finally, in the absence of further physical input, it is necessary to include in the effective Lagrangian all operators consistent with the given symmetry. Typically, infinitely many such operators exist. Hence, we need an organizing principle to decide which operators are relevant and which can be neglected: power counting.

Stability of the Ground State Our choice of the reference point, $\psi_0^i \equiv \langle \psi^i(0) \rangle$, corresponds to $\langle \pi^a(0) \rangle = 0$, or $\langle U(\pi(0)) \rangle = e$. The actual order parameter defines a map from M to the coset space $G_{\text{int}}/H_{\text{int}}$, $\langle U(\pi(x)) \rangle \equiv U_0(x)$. Nontrivial coordinate dependence indicates that the order parameter is *twisted*. Importantly, not every choice of $U_0(x)$ is physically consistent. A necessary condition is that $U_0(x)$ is a local minimum of the energy functional of the EFT. This is straightforward to check in two steps. First, $U_0(x)$ should be a stationary state, that is a solution of the *equation of motion* (EoM); see Sect. 8.3 for details on the latter. Second, the spectrum of small fluctuations about this state should only contain positive-energy excitations.

Example 13.4

Recall the helimagnetic state, introduced in Example 6.5,

$$\langle \mathbf{n}(\mathbf{x}, t) \rangle = (\cos kz, \sin kz, 0) , \quad (13.25)$$

where k is a positive constant. In this case, the order parameter is $\mathbf{n} \in \text{SO}(3)/\text{SO}(2) \simeq S^2$. It is spatially twisted so that it forms a helix oriented along the z -axis, with pitch $2\pi/k$. In Example 9.10, I showed that this order parameter minimizes the energy in the EFT for ferromagnets augmented with the *Dzyaloshinskii–Moriya* (DM) interaction. The precise value of k is fixed by the coupling of the DM term.

Ensuring the stability of the ground state is an integral part of the construction of the low-energy EFT. The requirement was already implicitly present in our analysis of internal symmetries in Part III of the book. The fact that the vacuum, $\pi^a = 0$, is a solution of the EoM was guaranteed by the absence of terms linear in π^a in the effective Lagrangian. Such terms were typically forbidden by the nonlinearly realized symmetry. Only in one exceptional example, namely the class of Galileon theories (Sect. 10.3.2), I enforced the absence of linear terms ad hoc by discarding the tadpole operator. Furthermore, stability under small fluctuations was ensured by fixing the signs of the parts of the effective Lagrangian bilinear in π^a , carrying respectively two temporal or two spatial derivatives.

Power Counting Eventually, we would like to parameterize the EFT by NG fields $\tilde{\pi}^a(x)$ such that $\langle \tilde{\pi}^a(x) \rangle = 0$. This makes it easy to distinguish the ground state from its fluctuations. It is often convenient to factorize the matrix variable $U(\pi)$ as

$$U(\pi(x)) \equiv U_0(x)U(\tilde{\pi}(x)) . \quad (13.26)$$

This maintains a simple transformation rule for $\tilde{\pi}^a$ under the action of G_{int} . Indeed, a comparison with (12.19) shows that

$$U(\tilde{\pi}) \xrightarrow{g_{\text{int}}} U(\tilde{\pi}'(\tilde{\pi}, g_{\text{int}})) = (U_0^{-1}g_{\text{int}}U_0)U(\tilde{\pi})h_{\text{int}}(\tilde{\pi}, g_{\text{int}})^{-1} . \quad (13.27)$$

In other words, the action of G_{int} by left multiplication is simply conjugated by U_0^{-1} .

The challenge to face is that following blindly the machinery of Chap. 8, we arrive at an effective Lagrangian where derivatives act on $U(\pi)$ rather than on $U(\tilde{\pi})$. This is not a problem if the characteristic length scale of the symmetry breaking $G_{\text{int}} \rightarrow H_{\text{int}}$ is much shorter than the scale of variations of the ground state $U_0(x)$. Such a *hierarchical* symmetry breaking occurs for instance in helimagnets. Indeed, the helix pitch in FeGe is about 70 nm [1], which is much longer than the scale of the crystal lattice at which the magnetic order is formed. In such systems, derivatives of both $U(\tilde{\pi}(x))$ and $U_0(x)$ can be considered small and counted together. Should, however, the order parameter feature short-distance variations, a different power counting is needed. The precise setup of the power counting that renders the EFT predictive is then best considered case by case.

13.2.2 Case Study: Relativistic Superfluids

From the point of view of symmetry, a superfluid is an ordered phase of matter where an internal $U(1)$ symmetry is spontaneously broken. We therefore need an order parameter charged under this symmetry. The simplest choice is to take a complex scalar field ϕ carrying the action of $G_{\text{int}} \simeq U(1)$ via $(\phi, x^\mu) \xrightarrow{\epsilon} (e^{i\epsilon} \phi, x^\mu)$. We then assume that due to microscopic dynamics, this field develops a nonzero VEV. That is however not enough. We want the vacuum to describe *matter*, that is a state with nonzero density of the $U(1)$ charge. In a Lorentz-invariant scalar theory, a state of uniform nonzero charge density corresponds to a time-dependent VEV,

$$\langle \phi(x) \rangle \equiv \varphi(x) = \varphi_0 e^{-i\mu t} . \quad (13.28)$$

The constants φ_0 and μ are without loss of generality assumed to be positive. See Chap. 3 of [2] for an introduction to superfluidity from the field theory point of view.

The same superfluid state can also be represented by a constant order parameter at the cost of redefining the Lagrangian. Indeed, introduce a new scalar field, $\Phi(x) \equiv e^{i\mu t} \phi(x)$, so that our superfluid state amounts to $\langle \Phi(x) \rangle = \varphi_0$. As a consequence of the field redefinition, $\partial_0 \phi = (\partial_0 - i\mu)\Phi$. This agrees with my previous observation (see Example 8.7) that an equilibrium state parameterized by the chemical potential μ can be described in the Lagrangian formalism by a constant background gauge field, $A_\mu = (\mu, \mathbf{0})$. It is now clear that the order parameter (13.28) corresponds to a superfluid at rest. Using Lorentz invariance, a superfluid state in uniform motion can likewise be parameterized by a constant timelike A_μ , or equivalently by the VEV $\varphi(x) = \varphi_0 e^{-iA \cdot x}$.

As we already saw in Example 13.3, the condensate (13.28) satisfies our technical assumption underlying the construction of EFT for systems with a twisted order parameter. Here we have $U(\pi(x)) = e^{i\pi(x)Q}$, where $\pi(x)$ is the phase of $\phi(x)$ and

Q is the generator of $G_{\text{int}} \simeq \text{U}(1)$. The NG field transforms under the internal symmetry as $\pi \xrightarrow{\epsilon} \pi + \epsilon$. The MC form (13.24) reduces to $\omega(\pi, x) = Qd\pi + P \cdot dx$. Upon dropping the tadpole operator, π , which is itself quasi-invariant, the most general Poincaré- and $\text{U}(1)$ -invariant effective action can be written as

$$S_{\text{eff}}\{\pi\} = \int d^Dx \mathcal{L}_{\text{eff}}(\partial\pi, \partial\partial\pi, \dots) . \quad (13.29)$$

The effective Lagrangian \mathcal{L}_{eff} is an arbitrary function of the derivatives of π in which all Lorentz indices have been contracted in a Lorentz-invariant manner.

Let us check that the tentative ground state, $\langle\pi(x)\rangle = -\mu t$, satisfies the EoM. This follows from the equivalence of the EoM and the conservation of the Noether current of the internal $\text{U}(1)$ symmetry; see Example 4.2 for a proof of the equivalence. The current carried by the condensate (13.28) is necessarily constant, and thus automatically conserved for any values of φ_0 and μ .

Next, we have to deal with power counting. I will follow closely the logic of Sect. 9.1.1, where a power-counting scheme for relativistic EFTs was worked out. A new complication is that we do not know a priori how large the chemical potential μ is, and thus whether $\partial_\mu\pi$ can be considered small. We therefore cannot naively expand the effective Lagrangian in the derivatives of π . However, we can use the fact that every π in the Lagrangian always carries at least one “persistent” derivative and only count *additional* derivatives. This makes sense: introducing the fluctuation field $\tilde{\pi}$ via (13.26), that is $\pi(x) \equiv \tilde{\pi}(x) - \mu t$, any higher derivative kills the chemical potential and thus captures directly the variation of $\tilde{\pi}$.

To formalize this observation, expand the effective Lagrangian as

$$\mathcal{L}_{\text{eff}}[\pi] = \sum_{n \geq 0} \mathcal{L}_{\text{eff}}^{(n)}[\pi] , \quad (13.30)$$

where $\mathcal{L}_{\text{eff}}^{(n)}$ collects all Lorentz-invariant operators with n additional derivatives. Now consider a generic Feynman diagram Γ , contributing to a given observable. Denote as I the number of internal propagators in Γ , as E the number of external legs, as L the number of loops, and as V_n the number of interaction vertices from each of $\mathcal{L}_{\text{eff}}^{(n)}$. Think of the Fourier representation, in which the diagram amounts to a homogeneous function of energy–momenta on the external legs. Loop integration contributes altogether DL powers of energy–momenta, the internal propagators contribute $-2I$ powers. The persistent derivatives on each field give $2I + E$ powers of energy–momenta. Finally, the additional derivatives add up to $\sum_{n \geq 0} nV_n$. Altogether, the naive degree of the diagram becomes

$$\deg \Gamma = DL + E + \sum_{n \geq 0} nV_n . \quad (13.31)$$

For any given observable, E is fixed and $\deg \Gamma \geq E$. The *leading-order* (LO) contribution comes from diagrams with $L = 0$ and $V_n = 0$ for any $n \geq 1$. Accordingly, the LO effective Lagrangian for a relativistic superfluid is $\mathcal{L}_{\text{eff}}^{(0)}$: an arbitrary Lorentz-

invariant function of $\partial_\mu \pi$. The *next-to-leading-order* (NLO) Lagrangian requires two additional derivatives, $n = 2$. For any $D \geq 3$, the NLO contribution to any observable corresponds to $\deg \Gamma = E + 2$ and amounts to tree-level ($L = 0$) diagrams with one interaction vertex from $\mathcal{L}_{\text{eff}}^{(2)}$ and all others from $\mathcal{L}_{\text{eff}}^{(0)}$. Interestingly, the contributions from loop diagrams only start to matter beyond NLO of the derivative expansion. Most important is that at any finite degree, only a finite number of Lagrangians $\mathcal{L}_{\text{eff}}^{(n)}$ and a finite number of Feynman diagrams is required. This renders the power-counting scheme consistent and predictive.

It might still look troublesome that the LO Lagrangian can be an arbitrary Lorentz-invariant function of $\partial_\mu \pi$. This surely contains an infinite number of unknown parameters, so how could such an EFT be useful? To address this question, note there is only one way to make such Lagrangians Lorentz-invariant, namely by contracting the Lorentz indices on $\partial_\mu \pi$ pairwise. Then, upon separating the ground state from the fluctuations, the LO Lagrangian becomes

$$\mathcal{L}_{\text{eff}}^{(0)}[\tilde{\pi}] = P\left(\sqrt{(\partial_\mu \pi)^2}\right) = P\left(\sqrt{(\partial_0 \tilde{\pi} - \mu)^2 - (\nabla \tilde{\pi})^2}\right), \quad (13.32)$$

where P is some as yet unknown function. In the superfluid ground state, the Lagrangian reduces to a function of the chemical potential, $P(\mu)$. This can be related to the energy density of the equilibrium state by Legendre transformation,

$$\mathcal{H}_{\text{eff}}^{(0)}\Big|_{\tilde{\pi}=0} \equiv U(\mu) = \mu P'(\mu) - P(\mu). \quad (13.33)$$

The prime indicates a derivative of P with respect to its argument. Also, I used that $\partial \mathcal{L}_{\text{eff}}^{(0)} / \partial(\partial_0 \pi) = -\partial \mathcal{L}_{\text{eff}}^{(0)} / \partial \mu$ and that $\langle \pi(x) \rangle = -\mu t$. Equation (13.33) looks familiar: $P(\mu)$ can be interpreted as the thermodynamic pressure of the superfluid at zero temperature, and $P'(\mu) \equiv n(\mu)$ as the density of the U(1) charge. Our main result therefore is that the LO EFT for superfluids at zero temperature (13.32) is completely fixed by the thermodynamic equation of state [3].

To get a flavor of the physical content of the EFT, let us expand (13.32) to second order in the NG field,

$$\mathcal{L}_{\text{eff}}^{(0)}[\tilde{\pi}] \simeq P(\mu) + \frac{P''(\mu)}{2}(\partial_0 \tilde{\pi})^2 - \frac{P'(\mu)}{2\mu}(\nabla \tilde{\pi})^2 + \dots \quad (13.34)$$

This shows that the NG bosons propagate with the phase velocity

$$v = \sqrt{\frac{P'(\mu)}{\mu P''(\mu)}} = \sqrt{\frac{n}{\mu} \frac{d\mu}{dn}} = \sqrt{\frac{dP}{dU}}, \quad (13.35)$$

where I used the thermodynamic relations $dU = \mu dn$ and $dP = n d\mu$. It is remarkable that the phase velocity of the superfluid NG boson at zero temperature is given by the same expression (13.35) as that of hydrodynamic sound.

Example 13.5

For an illustration of these general results, let us have a look at dense matter consisting of relativistic (Dirac) fermions in $d = 3$ dimensions [3]. Provided the interactions between the particles are sufficiently weak and their mass sufficiently small, the equation of state can be well approximated by that of a free gas of massless fermions,³

$$P(\mu) = \frac{\mu^4}{12\pi^2}, \quad n(\mu) = \frac{\mu^3}{3\pi^2}, \quad U(\mu) = \frac{\mu^4}{4\pi^2}. \quad (13.36)$$

Suppose that the interaction between the fermions is attractive so that they form Cooper pairs in the spin-singlet, s -wave state. The fermionic quasiparticles near the Fermi surface then become gapped and decouple at sufficiently low energies. The low-energy physics of the system reduces to the dynamics of the condensate of scalar Cooper pairs. This is an example of a fermionic superfluid. According to (13.32), the low-energy EFT of such a system is dominated by a LO Lagrangian that is polynomial in the NG field,

$$\mathcal{L}_{\text{eff}}^{(0)}[\tilde{\pi}] = \frac{1}{12\pi^2} [(\partial_0 \tilde{\pi} - \mu)^2 - (\nabla \tilde{\pi})^2]^2. \quad (13.37)$$

The NG bosons propagate at the speed $v = 1/\sqrt{3}$, characteristic of a gas of weakly interacting ultrarelativistic particles. Note that in this case, the microscopic energy scale of the system is set by the chemical potential itself. The scale associated with the spontaneous breakdown of the $U(1)$ symmetry is assumed to be much smaller, which is ensured by the weak binding of the Cooper pairs. Thus, μ could not have been treated as a small parameter in any reasonable sense.

13.3 Vector Modes: the Relevant, the Irrelevant and the Unphysical

As demonstrated in the previous chapter, the same symmetry-breaking pattern can be realized with various choices of order parameter. This ambiguity is innocuous for internal symmetries. On the other hand, in case of spacetime symmetries, it can lead to nonlinear realizations with apparently different numbers of NG variables. For instance, the superfluid order parameter (13.28) breaks the combined spacetime and internal symmetry in a rather nontrivial manner. First, it obviously breaks Lorentz boosts. Second, it breaks time translations and the internal $U(1)$ symmetry down to a “diagonal” subgroup, where the phase generated by a time translation is compensated by a $U(1)$ rotation. In the setup of Sect. 13.2.2, all this is implicitly taken into account. However, as pointed out in Sect. 12.4.3, it might be meaningful to add a secondary,

³ This restriction is of course just a matter of convenience. Our general result for the LO EFT (13.32) applies equally well to a strongly-interacting gas of massive fermions.

vector order parameter, representing the VEV of the Noether current of the $U(1)$ symmetry. This requires introducing an additional vector of NG fields. Similarly, we saw in Sect. 12.4.4 that nonlinear realization of Galilei symmetry is easier to implement if one adds a vector of NG variables corresponding to Galilei boosts.

These observations raise obvious questions. Are the additional vector NG variables merely a useful mathematical tool, or do they correspond to actual modes in the spectrum? In case they are physical, what is their exact nature? To answer these questions, I will now work out several physically distinct examples, from nonlinear realization to concrete effective actions. With the additional insight, I will then in Sect. 13.3.4 attempt to draw general conclusions.

13.3.1 The Relevant: Helimagnets

Let us start with a familiar system where a vector field turns out to excite a physical, gapless NG mode. This will set a benchmark for our discussion of the role of vector NG modes. At the same time, it will provide the first example of an EFT going beyond the class of systems covered by Sect. 13.2. I already addressed spin systems in detail in Chap. 9. Moreover, I mentioned the exotic case of helimagnets on several occasions. I will therefore not spend time on reviewing the basic phenomenology, and instead start right away with a symmetry-based analysis.

We are talking about a condensed-matter system whose symmetry is captured by the Aristotelian group in three spatial dimensions, $G \simeq SO(3) \ltimes \mathbb{R}^4$. Unlike in Chap. 9, I will not include a separate internal symmetry under spin rotations. This is broken by the spin-orbit coupling, which plays a key role in the formation of the helimagnetic order. The order parameter is the local magnetization density, which is a vector under spatial rotations, hence $\mathcal{M} \simeq \mathbb{R}^3$. We are of course interested in the physics of ordered states where the magnetization is nonzero. This fixes the two isotropy groups $H_0 \simeq SO(3)$ and $H_{(\psi_0,0)} \simeq SO(2)$. The local manifold decomposition (13.5) in this case becomes

$$G \simeq SO(2) \times S^2 \times \mathbb{R}^4. \quad (13.38)$$

The submanifold relevant for the nonlinear realization of the symmetry on the NG variables is $H_0/H_{(\psi_0,0)} \times M \simeq S^2 \times \mathbb{R}^4$. I will parameterize the spacetime using the translation operator $\mathcal{T}_{x,t} = e^{itH} e^{ix \cdot \mathbf{P}}$. The coset space S^2 is parameterized by two local coordinates π^a , encoded in an $SO(3)$ -valued matrix $U(\pi)$.

Having fixed the basic setup, we can read off the MC form from (13.3),

$$\begin{aligned} \omega(\pi, x) &= -iU(\pi)^{-1} dU(\pi) + Hdt + U(\pi)^{-1} (\mathbf{P} \cdot d\mathbf{x}) U(\pi) \\ &= -iU(\pi)^{-1} dU(\pi) + Hdt + P_{\underline{r}} [U(\pi)^{-1}]^r_s dx^s. \end{aligned} \quad (13.39)$$

The second equality follows from the fact that the momentum operator carries a vector representation of $\text{SO}(3)$ under the adjoint action thereof.⁴ Using (13.9), we subsequently identify the elements of the spacetime coframe,

$$\mathbf{e}^{*0}(\pi, x) = dt, \quad \mathbf{e}^{*L}(\pi, x) = [U(\pi)^{-1}]^L_s dx^s. \quad (13.40)$$

The ω_\perp part of the MC form comes entirely from $-iU(\pi)^{-1}dU(\pi)$, i.e. is given by $\omega^a(\pi) = \omega^a_b(\pi)d\pi^b$. The related covariant derivatives of the NG variables π^a are extracted using (13.15),

$$\nabla_0 \pi^a = \omega^a_b(\pi) \partial_0 \pi^b, \quad \nabla_r \pi^a = \omega^a_b(\pi) U(\pi)^s_r \partial_s \pi^b. \quad (13.41)$$

Under the action of $H_{(\psi_0,0)} \simeq \text{SO}(2)$, these split into the irreducible multiplets

$$\begin{aligned} & \delta^a_a \nabla_a \pi^a, \quad \varepsilon^{ab} \delta^a_a \nabla_a \pi_b, \\ & \nabla_0 \pi^a, \quad \nabla_a \pi^a, \quad \delta^a_a \nabla_a \pi_b + \delta^a_b \nabla_a \pi^a - \delta_{ab} \delta^c_c \nabla_c \pi^c. \end{aligned} \quad (13.42)$$

Those on the first line of (13.42) are singlets, while the first two items on the second line transform as vectors under $\text{SO}(2)$. Finally, the last operator in (13.42) transforms as a traceless symmetric tensor of $\text{SO}(2)$. That is also a two-dimensional real representation, but carries a double charge of $\text{SO}(2)$ compared to the vector representation. This is important when combining the irreducible multiplets into an effective Lagrangian in a way that preserves invariance under $H_{(\psi_0,0)} \simeq \text{SO}(2)$.

Before we can construct an effective action, we still need to consider power counting. Here I will, unlike for superfluids in Sect. 13.2.2, assume that the spatial variation of the order parameter is slow. This allows us to apply the naive scheme where each (covariant) derivative counts. The EFT will then be dominated by operators with the lowest number of derivatives. It is convenient to express the effective Lagrangian in terms of a unit-vector parameterization of the coset space, $\mathbf{n}(\pi) \in S^2$. This is related to $U(\pi)$ via $\mathbf{n}(\pi) = U(\pi)\mathbf{n}_0$, where $\mathbf{n}_0 \in S^2$ is an arbitrary but fixed reference vector. The advantage of using the \mathbf{n} variable is that it makes invariance under the whole $H_0 \simeq \text{SO}(3)$ manifest. Dropping surface terms, the most general invariant Lagrangian with up to two derivatives consists of

$$\begin{aligned} \mathcal{L}_{\text{eff}}^{(1,0)} &= c_1^{(1,0)} \mathbf{n} \cdot (\nabla \times \mathbf{n}), \\ \mathcal{L}_{\text{eff}}^{(2,0)} &= c_1^{(2,0)} \delta^{rs} \partial_r \mathbf{n} \cdot \partial_s \mathbf{n} + c_2^{(2,0)} (\nabla \cdot \mathbf{n})^2 + c_3^{(2,0)} [\mathbf{n} \cdot (\nabla \times \mathbf{n})]^2 \\ &\quad + c_4^{(2,0)} (\nabla \cdot \mathbf{n}) [\mathbf{n} \cdot (\nabla \times \mathbf{n})], \\ \mathcal{L}_{\text{eff}}^{(1,1)} &= c_1^{(1,1)} \partial_0 \mathbf{n} \cdot [(\mathbf{n} \cdot \nabla) \mathbf{n}], \\ \mathcal{L}_{\text{eff}}^{(0,2)} &= c_1^{(0,2)} (\partial_0 \mathbf{n})^2. \end{aligned} \quad (13.43)$$

⁴ In Chap. 9, I worked out the EFT for spin systems using the fundamental (spinor) representation of $\text{SO}(3)$. It goes without saying that this is just a matter of convenience. The nonlinear realization of symmetry on coset spaces as detailed in Sect. 7.3 only depends on the structure of the symmetry group. In the present case, the use of the vector representation of $H_0 \simeq \text{SO}(3)$ is suggested by the transformation properties of the momentum operator \mathbf{P} .

All the operators in (13.43) are clearly invariant under spatial rotations. What is less obvious is that there are no other algebraically independent contributions to the invariant Lagrangian with up to two derivatives. This follows from the formalism based on the MC form, but would not be straightforward to check in the language using the \mathbf{n} field. Let me at least clarify why some apparent candidate operators do not give anything new. Starting with $\mathcal{L}_{\text{eff}}^{(2,0)}$, there is no operator of the type $(\nabla \times \mathbf{n})^2$, since this equals $\delta^{rs} \partial_r \mathbf{n} \cdot \partial_s \mathbf{n} - (\nabla \cdot \mathbf{n})^2$ up to a surface term. Likewise, the operator $[\mathbf{n} \times (\nabla \times \mathbf{n})]^2$ is absent, being equal to $(\nabla \times \mathbf{n})^2 - [\mathbf{n} \cdot (\nabla \times \mathbf{n})]^2$. Other candidate operators can be eliminated using the identity $(\mathbf{n} \cdot \nabla) \mathbf{n} = -\mathbf{n} \times (\nabla \times \mathbf{n})$, valid for a unit vector field \mathbf{n} . Finally, there is no $\partial_0 \mathbf{n} \cdot (\nabla \times \mathbf{n})$ in $\mathcal{L}_{\text{eff}}^{(1,1)}$, since this is itself a surface term. Similar reasoning shows that no new operators can be produced using second covariant derivatives such as $\delta_a^{\underline{a}} \nabla_{\underline{0}} \nabla_{\underline{a}} \pi^a$ or $\delta_a^{\underline{a}} \nabla_{\underline{\alpha}} \nabla_{\underline{a}} \pi^a$.

The presence of $\mathcal{L}_{\text{eff}}^{(1,0)}$ with a single spatial derivative is a new feature, arising from the vector order parameter. There is no invariant operator with a single time derivative. However, we know from Chap. 9 that an internal SU(2) spin symmetry together with spatial rotational invariance admit a quasi-invariant Lagrangian,

$$\mathcal{L}_{\text{eff}}^{(0,1)} = c_1^{(0,1)} \frac{\varepsilon_{ab} n^a \partial_0 n^b}{1 + n^3}, \quad (13.44)$$

where $a, b \in \{1, 2\}$. Being free of spatial derivatives, this is also quasi-invariant under our $H_0 \simeq \text{SO}(3)$ that acts on both \mathbf{n} and spatial coordinates. Finally, note that the coframe (13.40) yields the standard volume measure $d^3\mathbf{x} \, dt$ thanks to the fact that $U(\pi)$ is an SO(3)-valued matrix. This justifies a posteriori dropping surface terms from (13.43), and makes it actually possible to include a quasi-invariant Lagrangian such as (13.44).

In ferromagnets where the coupling $c_1^{(0,1)}$ is nonzero, every time derivative counts as two spatial derivatives (see Sect. 9.2.1 for a detailed justification). We can then discard the $\mathcal{L}_{\text{eff}}^{(1,1)}$ and $\mathcal{L}_{\text{eff}}^{(0,2)}$ Lagrangians as subleading. Even with this simplification, there are still six independent operators in the $\mathcal{L}_{\text{eff}}^{(1,0)}$, $\mathcal{L}_{\text{eff}}^{(0,1)}$ and $\mathcal{L}_{\text{eff}}^{(2,0)}$ Lagrangians. In order to make the analysis of the physical consequences of the EFT feasible, we need to take one more step to reduce the Lagrangian. Namely, the spin-orbit interaction in real materials is typically much weaker than other interactions. It therefore makes sense to treat the couplings of all the operators that lock the spin and orbital rotations together, namely $c_1^{(1,0)}$, $c_2^{(2,0)}$, $c_3^{(2,0)}$ and $c_4^{(2,0)}$, as small. Counting all of these formally as degree one in the derivative expansion, the entire LO (degree-two) effective Lagrangian will be just

$$\mathcal{L}_{\text{eff}}^{\text{LO}} = -M \frac{\varepsilon_{ab} n^a \partial_0 n^b}{1 + n^3} - \frac{\varrho_s}{2} \delta^{rs} \partial_r \mathbf{n} \cdot \partial_s \mathbf{n} - \frac{2\pi \varrho_s}{\lambda_{\text{DM}}} \mathbf{n} \cdot (\nabla \times \mathbf{n}). \quad (13.45)$$

Here I have already switched to the physical notation for the effective couplings: M for the local magnetization density, ϱ_s for the spin stiffness, and λ_{DM} for the length scale of the DM interaction. The latter is all that is left of the effects of the spin-orbit coupling at the LO of the derivative expansion.

I showed already in Example 9.10 that the DM interaction inevitably leads to the formation of a helical structure in the ground state. One can without loss of generality choose a Cartesian system of coordinates such that

$$\langle \mathbf{n}(\mathbf{x}, t) \rangle \equiv \mathbf{n}_0(\mathbf{x}) = (\cos kz, \sin kz, 0) , \quad (13.46)$$

where $k \equiv 2\pi/\lambda_{\text{DM}}$. Let us next focus on the excitation spectrum. The first step in that regard is to reparameterize the field following the logic of (13.26),

$$\mathbf{n}(\mathbf{x}, t) \equiv \begin{pmatrix} \cos kz - \sin kz & 0 \\ \sin kz & \cos kz & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{N}(\mathbf{x}, t) . \quad (13.47)$$

The variable $\mathbf{N}(\mathbf{x}, t)$ is still a unit vector. Moreover, in the ground state it is constant, $\langle \mathbf{N}(\mathbf{x}, t) \rangle \equiv \mathbf{N}_0 = (1, 0, 0)$. This allows us to identify independent fluctuations around the ground state with the N^2 and N^3 components of \mathbf{N} .

The task to express the Lagrangian (13.45) in terms of \mathbf{N} is a bit tedious, and the final result is not elegant. I will therefore skip the straightforward details and only spell out the part of the Lagrangian bilinear in N^2, N^3 . Up to a surface term,

$$\mathcal{L}_{\text{eff}}^{\text{LO}} \simeq \frac{\varrho_s}{2} (N^2 \ N^3) \begin{pmatrix} \nabla^2 & \Delta(z) \\ -\Delta(z) & \nabla^2 - k^2 \end{pmatrix} \begin{pmatrix} N^2 \\ N^3 \end{pmatrix} + \dots , \quad (13.48)$$

where

$$\Delta(z) \equiv -\frac{M}{\varrho_s} \partial_0 + \frac{4\pi}{\lambda_{\text{DM}}} [(\cos kz) \partial_1 + (\sin kz) \partial_2] . \quad (13.49)$$

The excitation spectrum is determined by the zero modes of the matrix differential operator in (13.48). Due to the explicit coordinate dependence, this is a complicated problem. I will therefore content myself with some simple observations that illustrate the peculiarities of helimagnets. To start with, the presence of operators with a single time derivative makes the two degrees of freedom N^2, N^3 canonically conjugate to each other. The spectrum should therefore consist of a single type of magnon, just like in uniform ferromagnets in the absence of the DM interaction. The magnon dispersion relation can be calculated explicitly in the special case of propagation along the helix axis. The single-particle wave function is then independent of the x, y coordinates so that $\Delta \rightarrow -(M/\varrho_s) \partial_0$. A Fourier transform in time and the z -coordinate then gives the energy as a function of momentum,

$$E(p_z) = \frac{\varrho_s}{M} |p_z| \sqrt{p_z^2 + k^2} . \quad (13.50)$$

The physical momentum p_z should be much smaller than the inverse of the length scale associated with the formation of the ferromagnetic order, which can be thought of as $\sqrt{\varrho_s}$. This defines the range of validity of the EFT. Within this range, however, there are two qualitatively different regimes. At long wavelengths, $p_z \ll k$, (13.50) is well approximated by $E(p_z) \approx (k\varrho_s/M) |p_z|$. The dispersion relation is linear. The corresponding one-particle state is a linearly polarized spin wave with $N^3 \approx 0$. This amounts to spin oscillations in the xy plane along the direction perpendicular locally to $\mathbf{n}_0(\mathbf{x})$. On the other hand, at intermediate wavelengths, $k \ll p_z \ll \sqrt{\varrho_s}$, the dispersion relation (13.50) turns into $E(p_z) \approx (\varrho_s/M) p_z^2$. This reproduces the standard ferromagnetic magnon spectrum we found in Sect. 9.2.2. Physically, this means that such short-distance oscillations of spin are insensitive to the helical structure of the ground state. The magnon is circularly polarized in the plane perpendicular to the local magnetization $\mathbf{n}_0(\mathbf{x})$.

This concludes our excursion to helimagnetism. We have discovered very interesting physics, but no surprises as to the number and type of NG modes in the spectrum so far. To that end, we will capitalize on the experience we have collected here to explore another fascinating physical system: liquid crystals.

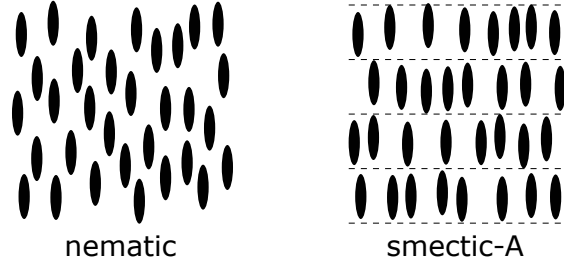
13.3.2 The Irrelevant: Smectic Liquid Crystals

Liquid crystals are an intriguing exotic state of matter; see [4] for a thorough account of the subject. The basic ingredient underlying the liquid crystal order is a strongly anisotropic microscopic constituent. This is typically an elongated organic molecule or a polymer, with a rigid, rod-like structure. An ordered ground state arises from a geometric arrangement of the molecules that breaks invariance under spatial translations or rotations, or both.

I will be mostly concerned with the interplay of two simple liquid crystal phases: the *nematic* phase and the *smectic* (or more precisely *smectic-A*) phase. In the nematic phase, the microscopic constituents are aligned along a common direction but their spatial positions remain disordered as in an ordinary liquid. See the left panel of Fig. 13.1 for a visualization. In the smectic phase, the alignment persists but the constituents are in addition organized in parallel layers. Within each layer, however, their positions remain disordered; see the right panel of Fig. 13.1. Both of these phases feature an order parameter that defines the axis of alignment, and can be viewed as an “unoriented vector.” To describe smectics, one needs in addition a secondary order parameter that accounts for the layered structure of the medium.

The microscopic symmetry governing the physics of liquid crystals is the same as that of spin systems with spin-orbit coupling, $G \simeq \text{SO}(3) \ltimes \mathbb{R}^4$. In the simpler nematic phase, which I will start with, the target space is $\mathcal{M} \simeq \mathbb{R}^3/\mathbb{Z}_2$, where any vector in \mathbb{R}^3 is identified with its opposite. This setup is almost identical to that of Sect. 13.3.1, and we can therefore largely reuse the EFT developed therein. The change in the global structure of \mathcal{M} from \mathbb{R}^3 to $\mathbb{R}^3/\mathbb{Z}_2$ affects the nature of topological defects in the medium. However, as far as the dynamics of small fluctuations of the order

Fig. 13.1 Schematic visualization of nematic and smectic-A orders. A nematic consists of a set of mutually aligned undirected rods with random positions. In a smectic-A, the rods are in addition organized in layers. Their positions within each layer remain disordered



parameter is concerned, all we have to do is impose invariance under the reflection $\mathbf{n} \rightarrow -\mathbf{n}$. This kills, first of all, the quasi-invariant Lagrangian (13.44), for there is no equivalent of “magnetization” in liquid crystals that the coupling $c_1^{(0,1)}$ could measure. Moreover, the $c_4^{(2,0)}$ and $c_1^{(1,1)}$ terms in (13.43) drop out, being odd in \mathbf{n} . In materials that respect spatial parity, we can finally discard the $c_1^{(1,0)}$ term.⁵ All in all, the LO Lagrangian for the fluctuations of the order parameter in nematics consists of the $c_1^{(2,0)}$, $c_2^{(2,0)}$, $c_3^{(2,0)}$ and $c_1^{(0,2)}$ operators. However, a somewhat different basis of operators with spatial derivatives is more common in the literature, namely

$$\mathcal{L}_{\text{eff}}^{\text{LO}} \supset -\frac{1}{2} \{K_1(\nabla \cdot \mathbf{n})^2 + K_2[\mathbf{n} \cdot (\nabla \times \mathbf{n})]^2 + K_3[\mathbf{n} \times (\nabla \times \mathbf{n})]^2\} . \quad (13.51)$$

The advantage of this parameterization is that there are field configurations for which all but one of the operators in (13.51) vanish. Global stability of the uniform state with constant $\langle \mathbf{n}(x) \rangle \equiv \mathbf{n}_0$ thus requires that all the couplings $K_{1,2,3}$ are positive. For completeness, let me add that they are related to the parameters in (13.43) by $K_1 = -2(c_1^{(2,0)} + c_2^{(2,0)})$, $K_2 = -2(c_1^{(2,0)} + c_3^{(2,0)})$ and $K_3 = -2c_1^{(2,0)}$ up to integration by parts.

Unlike in the case of helimagnets, I will not use the Lagrangian (13.51) to analyze the spectrum of excitations of a nematic. Namely, to get a physically correct picture of the dynamics, one would need to consider the interplay of the vector variable \mathbf{n} with hydrodynamic degrees of freedom (see Sect. 8.5 of [6]). After all, we are talking about *liquids*. Our restricted setting will however be sufficient to contrast the NG degrees of freedom in the nematic and smectic-A phases.

To describe the smectic state, we need an additional order parameter accounting for the layered structure as shown in the right panel of Fig. 13.1. This can be done with a single real scalar field $\phi(x)$. The layers of the medium are then defined as surfaces of constant ϕ . The translation across one layer of molecules changes $\phi(x)$ by a fixed amount; I will use a normalization of $\phi(x)$ where the increment equals 2π . Interestingly, this picture implies the existence of a new, emergent symmetry. Namely, for the sake of identification of the individual strata of the smectic, it does not matter what the attached value of ϕ is. None of the surfaces of constant ϕ changes if we shift $\phi(x)$ by a constant. The EFT for smectic liquid crystals should therefore be

⁵ The $c_1^{(1,0)}$ term is relevant for *cholesteric* liquid crystals whose building blocks possess intrinsic chirality. The mathematics of cholesterics is largely identical to that of helimagnets [5].

invariant under the internal symmetry $(\mathbf{n}, \phi, x^\mu) \xrightarrow{\epsilon} (\mathbf{n}, \phi + \epsilon, x^\mu)$. This is our first example of an emergent symmetry arising from the spatial distribution of a medium; I will give a more systematic account of such symmetries in Chap. 14. The complete symmetry group of a smectic is then $G \simeq [\text{SO}(3) \ltimes \mathbb{R}^4] \times \mathbb{R}$. The corresponding isotropy groups are $H_0 \simeq \text{SO}(3) \times \mathbb{R}$ and $H_{(\psi_0, 0)} \simeq \text{SO}(2)$. The degrees of freedom of the EFT are determined by the coset space $H_0/H_{(\psi_0, 0)} \simeq S^2 \times \mathbb{R}$.

The values of the two order parameters in the smectic ground state are correlated. As indicated in Fig. 13.1, the axis of alignment \mathbf{n}_0 of the molecules is perpendicular to the entire stack of parallel layers. Moreover, the individual molecular layers are equidistantly spaced. This implies that the VEV of the gradient $\langle \nabla \phi(x) \rangle$ is a constant nonzero vector, parallel to \mathbf{n}_0 . I will use the notation

$$\langle \phi(\mathbf{x}, t) \rangle = k \mathbf{n}_0 \cdot \mathbf{x} , \quad (13.52)$$

where the parameter k defines the distance between neighboring layers, $2\pi/k$.

The shift symmetry \mathbb{R} guarantees that $\phi(x)$ can only enter the effective Lagrangian with derivatives. Being a scalar under rotations, its covariant derivatives are

$$\nabla_0 \phi = \partial_0 \phi , \quad \nabla_r \phi = U(\pi)^s_r \partial_s \phi , \quad (13.53)$$

where $U(\pi)$ is still the matrix related to \mathbf{n} by $\mathbf{n}(\pi) = U(\pi)\mathbf{n}_0$. Also, I used the previously derived coframe (13.40), which is not affected by adding the scalar field $\phi(x)$. The appearance of these covariant derivatives in the effective Lagrangian is constrained by the isotropy group $H_{(\psi_0, 0)} \simeq \text{SO}(2)$ and the reflection $\mathbf{n} \rightarrow -\mathbf{n}$. To these we should add the discrete symmetry under $\phi \rightarrow -\phi$ which, just like the internal shifts of ϕ , preserves the identification of the layers of the smectic. With these constraints, the contributions of $\phi(x)$ to the LO effective Lagrangian are reduced to

$$\mathcal{L}_{\text{eff}}^{\text{LO}} \supset c_0 (\partial_0 \phi)^2 + c_{\parallel} (\mathbf{n} \cdot \nabla \phi)^2 + c_{\perp} (\mathbf{n} \times \nabla \phi)^2 . \quad (13.54)$$

In order for the alignment of $\nabla \phi$ and \mathbf{n} to be energetically preferred, the spatial gradient couplings c_{\parallel} and c_{\perp} should satisfy $c_{\perp} < c_{\parallel}$.

Let us now turn the argument around and see how the background (13.52) affects the vector field \mathbf{n} . I will set without loss of generality $\mathbf{n}_0 = (0, 0, 1)$ so that the two independent fluctuations of \mathbf{n} correspond to n^1 and n^2 . This reduces (13.54) to

$$\mathcal{L}_{\text{eff}}^{\text{LO}} \supset -(c_{\parallel} - c_{\perp}) k^2 [(n^1)^2 + (n^2)^2] . \quad (13.55)$$

Thanks to the stability constraint on the couplings, $c_{\parallel} - c_{\perp} > 0$. What we have here is therefore a well-defined, positive-definite mass term for the fluctuations of \mathbf{n} .

Let us pause to appreciate the implications of this result. In the nematic phase, we dealt with two NG degrees of freedom, parameterizing via the unit vector \mathbf{n} the coset space S^2 . In the smectic phase, we extended the coset space to $S^2 \times \mathbb{R}$. The corresponding new NG variable parameterizes the fluctuations of ϕ around its VEV (13.52). Given that ϕ only enters the effective Lagrangian

with derivatives, it necessarily couples to a NG mode in the spectrum. However, at the same time, the fluctuations of \mathbf{n} received a gap. Adding an extra NG field has *reduced* the number of NG modes from two to one!

To understand what is going on, think of the order parameter (13.52) alone. Being a spatial vector, the gradient $\nabla\phi$ spontaneously breaks spatial rotations in addition to the internal shift symmetry of ϕ and translations in the direction of \mathbf{n}_0 . We may then treat $\langle\mathbf{n}(x)\rangle \equiv \mathbf{n}_0$ as a secondary order parameter, since it does not break any symmetries that would be left intact by $\langle\phi(x)\rangle$. This makes the \mathbf{n} mode redundant: its presence is not required to reproduce the symmetry-breaking pattern of a smectic. Consequently, the fluctuations of \mathbf{n} are not protected by symmetry from acquiring a gap.

This is an explicit realization of an issue I alerted the reader to already in Chap. 12. The number of true, gapless NG degrees of freedom (one) is fixed by the symmetry-breaking pattern. However, the additional, would-be NG modes contained in $\mathbf{n}(x)$ are physical and may manifest themselves as low-lying excitations of the smectic ground state. The nematic and smectic phases of liquid crystals are separated by a continuous (second-order) phase transition. At the transition point, the scale k vanishes. Accordingly, near the phase transition, the gap of the \mathbf{n} -type fluctuations is very small and they need to be included in the low-energy EFT.

Further away from the phase transition, we expect the gapped modes to have little effect on the low-energy physics. In this regime, it should be possible to construct an EFT for smectics in terms of $\phi(x)$ alone. This can be arrived at in two different ways. One possibility is to start from a combination of (13.51) and (13.54) and integrate out the vector \mathbf{n} . The other possibility is to construct the EFT for $\phi(x)$ alone from scratch. This is arguably more straightforward but less predictive, since the effective couplings of the EFT will be unrelated to those in (13.51) and (13.54). For illustration, I will nevertheless choose the latter option as it highlights the physics of the NG mode in the smectic phase.

The task to find an EFT for a scalar field ϕ , subject to an internal shift symmetry \mathbb{R} , copies closely the construction of Sect. 13.2.2, with two differences. The first difference is minor. Namely, unlike in Sect. 13.2.2, we do not impose full Poincaré invariance, but merely the Aristotelian symmetry. The resulting effective Lagrangian is, in analogy with (13.29), a function of derivatives of $\phi(x)$ where spatial indices are contracted in a way preserving invariance under spatial rotations. The second difference is absolutely essential: we want our EFT to stabilize the anisotropic smectic state (13.52). To that end, let us parameterize the fluctuations of $\phi(x)$ by a NG field $\pi(x)$ such that $\phi(\mathbf{x}, t) = k\mathbf{n}_0 \cdot \mathbf{x} + \pi(\mathbf{x}, t)$. The basic rotationally invariant building block for the construction of the Lagrangian is $(\nabla\phi)^2 = k^2 + 2k\mathbf{n}_0 \cdot \nabla\pi + (\nabla\pi)^2$. Dropping the constant piece, the part of the effective Lagrangian with one derivative per field will be a generic function of $\partial_0\pi$ and $k\mathbf{n}_0 \cdot \nabla\pi + (\nabla\pi)^2/2$. The static part of the effective Lagrangian, bilinear in the NG field $\pi(x)$, is thus contained in

$$\mathcal{L}_{\text{eff}} \supset c_1 \left[k \mathbf{n}_0 \cdot \nabla \pi + \frac{(\nabla \pi)^2}{2} \right] + c_2 \left[k \mathbf{n}_0 \cdot \nabla \pi + \frac{(\nabla \pi)^2}{2} \right]^2 + \dots \quad (13.56)$$

The ellipsis includes operators that are bilinear in $\pi(x)$ but contain more than one derivative per field. The leading operator of this kind is $c_3(\nabla^2 \phi)^2 = c_3(\nabla^2 \pi)^2$.

Remarkably, the c_1 operator is forbidden. The easiest way to see this is to think of the gradient $\nabla \phi$ as a vector field, $\mathbf{A} \equiv \nabla \phi$. The part of the static Lagrangian with one derivative per ϕ is then a mere function of \mathbf{A} . Should the corresponding Hamiltonian density have a minimum for nonzero $\langle \mathbf{A} \rangle = k \mathbf{n}_0$ as implied by (13.52), its Hessian matrix at the minimum must have two zero modes. This is just the statement of the Goldstone theorem in terms of the eigenvalues of the mass matrix, which we proved back in Sect. 3.1. Hence, for the smectic state (13.52) to be stable within the EFT, the Lagrangian (13.56) must not contain any terms quadratic in the part of the gradient of π , perpendicular to \mathbf{n}_0 . It is the stability criterion together with the spontaneously broken rotational invariance that forbids the c_1 operator.

We conclude that the leading contributions to the bilinear Lagrangian consist of the operators $(\partial_0 \pi)^2$, $(\nabla_{\parallel} \pi)^2$ and $(\nabla_{\perp}^2 \pi)^2$, where \parallel and \perp denote projections to directions parallel and transverse to \mathbf{n}_0 . This suggests an unusual, anisotropic power counting. For all these operators to contribute at the same order, the time derivative should have the same counting degree as ∇_{\parallel} and ∇_{\perp}^2 . The presence of a higher power of the transverse gradient has a profound impact on the stability of the smectic phase at nonzero temperature. I will relegate a more detailed discussion and further generalization of this observation to Chap. 15.

13.3.3 The Unphysical: Nonrelativistic Superfluids

The final example in our exploration of the role of vector modes will be nonrelativistic superfluids. Here we do know a priori that it is possible to construct an EFT solely in terms of a scalar degree of freedom. All one has to do is to take the EFT for a relativistic superfluid, developed in Sect. 13.2.2, and perform a nonrelativistic limit. It is however instructive to carry out the construction of the nonrelativistic version of the EFT from scratch.

To get started, recall the basic properties of Galilei symmetry from Sect. 12.4.4. The Lie algebra of the Galilei group in d spatial dimensions includes the generators $P_{\underline{r}}$ and H of space and time translations and J_{rs} of spatial rotations. In addition, there is a vector generator $K_{\underline{r}}$ of Galilei boosts. I will add from the outset a central charge Q whose eigenvalues measure nonrelativistic mass. This turns the Galilei group into the Bargmann group. Altogether, the nontrivial part of the local structure of the symmetry group is determined by the commutation relations

$$[H, K_{\underline{r}}] = i P_{\underline{r}} \ , \quad [P_{\underline{r}}, K_{\underline{s}}] = i \delta_{rs} Q \ . \quad (13.57)$$

The commutators of J_{rs} are fixed by rotation invariance and will not be needed. All the other commutators not listed here are vanishing. Mathematically, the Bargmann group has the structure $G \simeq \text{SO}(d) \ltimes \{\mathbb{R}_K^d \ltimes [\mathbb{R}^D \times \text{U}(1)_Q]\}$. All the symmetry transformations but translations leave the spacetime origin fixed, hence its isotropy group is $H_0 \simeq [\text{SO}(d) \ltimes \mathbb{R}_K^d] \times \text{U}(1)_Q$. The translations themselves will be represented by the operator $\mathcal{T}_{\mathbf{x},t} \equiv e^{it\hat{H}} e^{i\mathbf{x}\cdot\hat{\mathbf{P}}}$.

In order to describe the superfluid order, we need an order parameter that carries a nontrivial action of the internal $\text{U}(1)_Q$ subgroup. Similarly to relativistic superfluids, the simplest choice is a complex scalar field ψ . For any $\psi_0 \neq 0$, the corresponding isotropy group is then $H_{(\psi_0,0)} \simeq \text{SO}(d) \ltimes \mathbb{R}_K^d$. However, as pointed out before, this is not satisfactory, since it violates our requirements on the decomposition (13.6) of the Lie algebra \mathfrak{g} of G . Namely, the operators P_r and H span a basis of $\mathfrak{g}/\mathfrak{h}_0$ but do not carry a representation of $H_{(\psi_0,0)}$. The way out is to add a secondary order parameter $A^\mu \equiv (A^0, \mathbf{A})$ that transforms as a vector under Galilei boosts. Choosing $A_0^\mu = (a, \mathbf{0})$ with $a \neq 0$ can be interpreted as specifying the density of the $\text{U}(1)_Q$ charge in the rest frame of the superfluid. It reduces the isotropy subgroup further to $H_{((\psi_0, A_0), 0)} \simeq \text{SO}(d)$. The coset space $H_0/H_{((\psi_0, A_0), 0)} \simeq \mathbb{R}_K^d \times \text{U}(1)_Q$ can be parameterized by NG variables π and ξ^r through $U(\pi, \boldsymbol{\xi}) \equiv e^{i\pi Q} e^{i\boldsymbol{\xi} \cdot \mathbf{K}}$. In this parameterization, a Galilei boost with velocity \mathbf{v} acts on the NG variables and the spacetime coordinates as

$$e^{i\mathbf{v}\cdot\mathbf{K}}(\pi, \boldsymbol{\xi}, \mathbf{x}, t) = (\pi + \mathbf{v} \cdot \mathbf{x} + \mathbf{v}^2 t/2, \boldsymbol{\xi} + \mathbf{v}, \mathbf{x} + \mathbf{v}t, t) . \quad (13.58)$$

With all the pieces at hand, we can now compute the MC form (13.3),

$$\omega(\pi, \boldsymbol{\xi}, \mathbf{x}, t) = Q[d\pi - \boldsymbol{\xi} \cdot d\mathbf{x} + (1/2)\boldsymbol{\xi}^2 dt] + \mathbf{K} \cdot d\boldsymbol{\xi} + Hdt + \mathbf{P} \cdot (d\mathbf{x} - \boldsymbol{\xi}dt) . \quad (13.59)$$

The ω_{\parallel} part of the MC form is trivial. From $\omega_P = Hdt + \mathbf{P} \cdot (d\mathbf{x} - \boldsymbol{\xi}dt)$, we extract the spacetime coframe, $e^0 = dt$ and $e^r = \delta_s^r(dx^s - \xi^s dt)$. The $\delta_s^r \xi^s dt$ term in e^r does not affect the spacetime volume form, we can thus use the volume measure $d^d \mathbf{x} dt$ to turn effective Lagrangians into effective actions. To construct covariant derivatives of the NG fields, we will also need the spacetime frame, dual to the above coframe,

$$e_0^0 = 1 , \quad e_0^r = \xi^r , \quad e_r^0 = 0 , \quad e_r^s = \delta_r^s . \quad (13.60)$$

The covariant derivatives of π and ξ^r are then extracted from the ω_{\perp} part of the MC form (13.59) with the help of (13.15),

$$\begin{aligned} \nabla_0 \pi &= \partial_0 \pi + \boldsymbol{\xi} \cdot \nabla \pi - \boldsymbol{\xi}^2/2 , & \nabla_r \pi &= \delta_r^s (\partial_s \pi - \xi_s) , \\ \nabla_0 \xi^r &= (\partial_0 + \boldsymbol{\xi} \cdot \nabla) \xi^r , & \nabla_s \xi^r &= \delta_s^u \partial_u \xi^r . \end{aligned} \quad (13.61)$$

It is an easy exercise to check explicitly that all these covariant derivatives are invariant under the Galilei boost (13.58). Thanks to their simple form, it is not necessary to distinguish frame and coordinate-basis indices. In the following, I will therefore happily use $\nabla_r \pi = \partial_r \pi - \xi_r$ and $\nabla_s \xi^r = \partial_s \xi^r$.

In accord with the general machinery of nonlinear realizations, we could also add matter fields, organized in linear multiplets of $H_{((\psi_0, A_0), 0)} \simeq \text{SO}(d)$. By (13.17), the covariant derivatives of a matter field χ^e would then be $\nabla_0 \chi^e = (\partial_0 + \xi \cdot \nabla) \chi^e$ and $\nabla_r \chi^e = \partial_r \chi^e$. This opens the possibility to promote (almost) any nonrelativistic theory with Aristotelian symmetry to a theory that is Galilei-invariant. Indeed, recall that in Part III of the book, we dealt largely with Aristotelian EFTs where we imposed invariance under linearly realized rotations by hand. What prevents us from introducing an auxiliary vector field ξ^r and replacing $\partial_0 \rightarrow \nabla_0 \equiv \partial_0 + \xi \cdot \nabla$ everywhere? This will certainly work for strictly invariant Lagrangians. (The naive replacement $\partial_0 \rightarrow \nabla_0$ might spoil the quasi-invariance of Lagrangians that shift upon a symmetry transformation by a total time derivative.) The only fly in the ointment is that we do not have a physical interpretation for the ξ^r field in terms of the NG degrees of freedom of the Aristotelian EFT. Rather, ξ^r should be viewed as the local velocity of a medium carrying the NG modes; the combination $\partial_0 + \xi \cdot \nabla$ is often referred to as the *material derivative*. The Aristotelian EFTs constructed in Chap. 8 describe the low-energy physics in the rest frame of the medium. The same remark applies to relativistic systems. There, adding a NG field for Lorentz boosts allows one to promote an Aristotelian theory to one invariant under the full Poincaré group.

With the basic building blocks at hand, we can proceed to the construction of an EFT for the superfluid. Following closely the relativistic counterpart in Sect. 13.2.2, we first fix a power-counting scheme. We again count both $\partial_0 \pi$ and $\nabla \pi$ as degree zero. For the covariant derivatives in (13.61) to have a consistent counting degree, we have to assign ξ^r degree zero as well. Any other derivative acting on π or ξ^r will then have degree one. It follows that at the LO of the derivative expansion, we can ignore $\nabla_\mu \xi^r$. The LO Lagrangian will be some function of the rotationally invariant operators $\nabla_0 \pi$ and $\delta^{rs} \nabla_r \pi \nabla_s \pi$. This underlines the fact that the field ξ^r is merely auxiliary and does not represent independent degrees of freedom. Indeed, it only enters the LO Lagrangian without any derivatives and can be eliminated algebraically by imposing its EoM,

$$(\partial_r \pi - \xi_r) \left[\frac{\partial \mathcal{L}_{\text{eff}}^{(0)}}{\partial (\nabla_0 \pi)} - 2 \frac{\partial \mathcal{L}_{\text{eff}}^{(0)}}{\partial (\delta^{su} \nabla_s \pi \nabla_u \pi)} \right] = 0. \quad (13.62)$$

The natural solution is $\xi = \nabla \pi$, which turns the two building blocks for the LO Lagrangian to $\nabla_0 \pi \rightarrow \partial_0 \pi + (\nabla \pi)^2/2$ and $\nabla_r \pi \rightarrow 0$. Of course, (13.62) can also be satisfied if the expression in square brackets vanishes. This would imply that the LO Lagrangian only depends on $\nabla_0 \pi + (1/2) \delta^{rs} \nabla_r \pi \nabla_s \pi = \partial_0 \pi + (\nabla \pi)^2/2$. Either way, we end up with the same LO Lagrangian in terms of π alone.

Finally, recall the superfluid ground state can be described by a time-dependent VEV of a complex scalar field. This amounts to $\langle \pi(x) \rangle = -\mu t$ where μ is the chemical

potential. The fluctuations around this background can be parameterized by a field $\tilde{\pi}$ such that $\pi(x) = \tilde{\pi}(x) - \mu t$. Following again the analogy with Sect. 13.2.2, we conclude that the LO effective Lagrangian for Galilei-invariant superfluids reads

$$\mathcal{L}_{\text{eff}}^{(0)}[\tilde{\pi}] = P\left(\mu - \partial_0 \tilde{\pi} - (\nabla \tilde{\pi})^2/2\right). \quad (13.63)$$

The function $P(\mu)$ represents the thermodynamic pressure of the superfluid in equilibrium at zero temperature.

Example 13.6

As a simple application of our new EFT, expand the Lagrangian (13.63) to second order in the NG field $\tilde{\pi}$. Up to a surface term, we find

$$\mathcal{L}_{\text{eff}}^{(0)}[\tilde{\pi}] \simeq P(\mu) + \frac{P''(\mu)}{2}(\partial_0 \tilde{\pi})^2 - \frac{P'(\mu)}{2}(\nabla \tilde{\pi})^2 + \dots \quad (13.64)$$

This gives immediately the phase velocity of the NG boson,

$$v = \sqrt{\frac{P'(\mu)}{P''(\mu)}} = \sqrt{\frac{dP}{dn}}, \quad (13.65)$$

where $n(\mu) \equiv P'(\mu)$ should be interpreted as the mass density of the superfluid. This matches the relativistic result $v = \sqrt{dP/dU}$ in (13.35), since in the nonrelativistic limit, the energy density $U(\mu)$ reduces to the density of the rest mass. Note that in spite of the appearance, the EFT (13.63) describes a type-A NG mode with a linear dispersion relation. This justifies a posteriori the use of the power-counting scheme where temporal and spatial derivatives are counted equally.

Before closing the discussion of nonrelativistic superfluids, I have a debt to pay off. Namely, when constructing the EFT, I tacitly assumed the effective Lagrangian to be strictly invariant. Could the Lagrangian also include some quasi-invariant contributions? The answer to this question is surprisingly rich, but luckily will not affect the LO Lagrangian (13.63).

As pointed out in Sect. 8.1, quasi-invariant Lagrangians for broken internal symmetries are classified by relative Lie algebra cohomology. It is not clear whether this elegant result survives all the complications of nonlinear realization of spacetime symmetries. Let us nevertheless take a leap of faith and see where it will lead us. We thus seek $(D+1)$ -forms that are closed and H_0 -invariant. I will first split the MC form (13.59) into components corresponding to the individual generators,

$$\begin{aligned} \omega_Q &= d\pi - \xi \cdot d\mathbf{x} + (1/2)\xi^2 dt, & \omega_K^r &= d\xi^r, \\ \omega_H &= dt, & \omega_P^r &= dx^r - \xi^r dt. \end{aligned} \quad (13.66)$$

We are now to take the exterior product of $D+1$ of these 1-forms. Invariance under the whole isotropy group H_0 is ensured by contracting vector indices in a rotationally invariant manner. It is the closedness requirement that poses a challenge. To that end, we take note of the MC equations, reflecting the commutation relations (13.57) between the generators,

$$d\omega_Q = \delta_{rs}\omega_P^r \wedge \omega_K^s, \quad d\omega_K^r = 0, \quad d\omega_H = 0, \quad d\omega_P^r = \omega_H \wedge \omega_K^r. \quad (13.67)$$

These suggest a set of $d+1$ naturally closed, rotationally invariant $(D+1)$ -forms,

$$\begin{aligned} \Omega_{D+1}^{(n)} &\equiv \frac{1}{d!} \varepsilon_{r_1 \dots r_d} \omega_H \wedge \omega_Q \wedge \omega_K^{r_1} \wedge \dots \wedge \omega_K^{r_n} \wedge \omega_P^{r_{n+1}} \wedge \dots \wedge \omega_P^{r_d} \\ &= \frac{1}{d!} \varepsilon_{r_1 \dots r_d} dt \wedge (d\pi - \xi \cdot d\mathbf{x}) \wedge d\xi^{r_1} \wedge \dots \wedge d\xi^{r_n} \wedge dx^{r_{n+1}} \wedge \dots \wedge dx^{r_d}, \end{aligned} \quad (13.68)$$

where $n = 0, \dots, d$. The corresponding effective Lagrangians are equivalent to the D -form potentials of these $(D+1)$ -forms, that is such $\Omega_D^{(n)}$ that $d\Omega_D^{(n)} = \Omega_{D+1}^{(n)}$. These D -form potentials are obviously not unique, since they can be shifted without consequences by the exterior derivative of any $(D-1)$ -form. A suitable choice is

$$\begin{aligned} \Omega_D^{(n)} &= -\frac{1}{d!} \varepsilon_{r_1 \dots r_d} dt \wedge \left[\pi d\xi^{r_1} \wedge \dots \wedge d\xi^{r_n} \wedge dx^{r_{n+1}} \wedge \dots \wedge dx^{r_d} \right. \\ &\quad \left. + \frac{n}{2(d-n+1)} \xi^2 d\xi^{r_1} \wedge \dots \wedge d\xi^{r_{n-1}} \wedge dx^{r_n} \wedge \dots \wedge dx^{r_d} \right]. \end{aligned} \quad (13.69)$$

A reader wishing to check this may find the following identity useful,

$$\begin{aligned} \varepsilon_{r_1 \dots r_d} d\xi^s \wedge d\xi^{r_1} \wedge \dots \wedge d\xi^{r_{n-1}} \wedge dx^{r_n} \wedge \dots \wedge dx^{r_d} \\ = -\frac{d-n+1}{n} \varepsilon_{r_1 \dots r_d} dx^s \wedge d\xi^{r_1} \wedge \dots \wedge d\xi^{r_n} \wedge dx^{r_{n+1}} \wedge \dots \wedge dx^{r_d}. \end{aligned} \quad (13.70)$$

Next, we eliminate the auxiliary field ξ^r by imposing the same constraint as above, $\xi = \nabla\pi$. After some manipulation, the D -form potentials (13.69) then give a set of $d+1$ quasi-invariant Lagrangians (with adjusted overall normalization),

$$\mathcal{L}_{\text{quasi}}^{(n)} = \varepsilon^{r_1 \dots r_n u_{n+1} \dots u_d} \varepsilon^{s_1 \dots s_n} \pi \prod_{i=1}^n (\partial_{r_i} \partial_{s_i} \pi). \quad (13.71)$$

The form of the Lagrangians (13.71) should ring a bell. Indeed, except for replacing spacetime indices with spatial ones, these are exactly the Galileon Lagrangians (10.54). The similarity is not accidental. The generators Q, K_μ of the constant and linear parts of the Galileon symmetry (10.53) and the energy-momentum operator P_μ satisfy the commutation relation $[P_\mu, K_\nu] = i g_{\mu\nu} Q$. This is identical to the centrally extended commutator in the Bargmann algebra. We have thus killed two birds with one stone, explaining the origin of the Galileon Lagrangians as a byproduct of our discussion of Galilei-invariant superfluids. Now back to the physi-

cal content of (13.71). The $n = 0$ Lagrangian, $\mathcal{L}_{\text{quasi}}^{(0)}$, is a tadpole operator. This must be dropped to keep the EFT stable. The $n = 1$ Lagrangian is equivalent to $(\nabla\pi)^2$ by partial integration. This is, in spite of appearance, already included in the LO effective Lagrangian (13.63). Namely, it is a part of $\partial_0\pi + (\nabla\pi)^2/2$ where $\partial_0\pi$ itself is a surface term. Finally, all the quasi-invariant Lagrangians (13.71) with $n \geq 2$ contain more than one derivative per field. Consequently, they will only contribute at higher, subleading orders of the derivative expansion of the superfluid EFT.

13.3.4 Inverse Higgs Constraints

The analysis of concrete examples has given us enough experience to appreciate the broader spectrum of subtleties associated with spontaneous breaking of spacetime symmetries. Let us reevaluate what we have learned in particular about the role of vector fields. I should warn the reader that this matter is not yet quite settled. Hence, the content and organization of the discussion below is necessarily more subjective than elsewhere in the book.

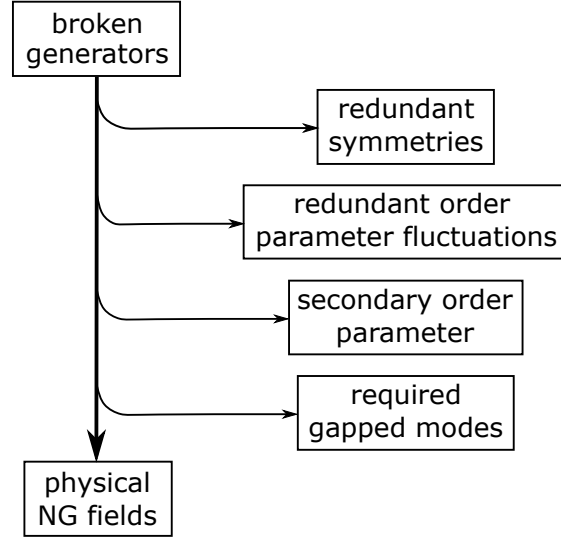
In case of internal symmetries, the situation is clear. There is a one-to-one mapping between broken symmetry generators and NG fields in the low-energy EFT. The construction of the EFT itself is fairly streamlined. All one needs to know is the symmetry-breaking pattern. As shown in Chap. 8, finding the EFT then boils down to solving certain linear-algebraic constraints for the tensor couplings in the effective Lagrangian. The functional dependence of the Lagrangian on the NG fields is completely fixed by symmetry.

In case of spontaneously broken spacetime symmetry, we would hope to be able to do the same. This desire lies behind the *agnostic nonlinear realization* that I already mentioned in Sect. 12.3.2. In this framework, one starts with a separate NG field for each spontaneously broken symmetry generator. The actions of all the generators in an abstract group space are by construction independent of each other. This however leads to the surprising observation that the number of fields that ultimately couple to gapless excitations may be lower than that of the broken generators.⁶ I will now outline the various mechanisms how the naively expected number of NG fields based on mere counting broken generators may be reduced. This will add nuance to the intuitive picture painted in Sect. 6.1 and summarized in Fig. 6.1 therein. The different mechanisms are listed in Fig. 13.2.

Redundancy of Symmetries Two symmetries that are locally equivalent necessarily lead to identical fluctuations of the order parameter. This is the simplest and most important of all the mechanisms, to which I already devoted the whole Chap. 11. Its advantage is that such a redundancy can be checked at the level of classical symme-

⁶ Remember that these independent NG fields may still not be in a one-to-one correspondence with NG modes in the spectrum. Some of the fields may be pairwise canonically conjugated, leading to type-B NG bosons. The canonical conjugation mechanism is the same for internal and spacetime symmetries, and thus requires no further discussion.

Fig. 13.2 Various mechanisms whereby the number of physical NG fields coupling to gapless, true NG modes can be lower than the number of generators parameterizing the coset space $H_0/H_{(\psi_0,0)}$. The detailed meaning of the keywords involved is explained in the main text. The physical NG fields are in turn in a one-to-one (type-A) or two-to-one (type-B) correspondence with the NG modes in the spectrum



try transformations. No specific choice of an order parameter or knowledge of the corresponding symmetry-breaking pattern is needed. See Sect. 11.3 for examples.

Redundancy of Order Parameter Fluctuations Moving down the decision tree, the next possibility is that some symmetries are locally distinct, yet lead to equivalent fluctuations of the order parameter. An example of this type is the helimagnet; see Example 6.5 for details. Here spatial translations and rotations are not trivially redundant due to the vector nature of the order parameter. However, thanks to the particular spatial modulation of the order parameter, three independent broken symmetry generators lead to mere two distinct types of fluctuations. This type of redundancy requires detailed knowledge of the order parameter, but can still be detected straightforwardly following the approach of Sect. 6.3.1.

Both of the above types of redundancy are automatically accounted for by the standard nonlinear realization of Chap. 12. This ensures a one-to-one parameterization of whatever fields carrying the action of the symmetry one starts with. However, it is sometimes advantageous to add auxiliary would-be NG fields, even if these are known beforehand not to correspond to separate physical degrees of freedom. We saw an example in Sect. 13.3.3. The unphysical nature of the vector field ξ^r therein manifested itself in the fact that we could eliminate it algebraically using its EoM. Such a procedure, while physically certainly correct, may however become exceedingly cumbersome beyond LO of the derivative expansion. There is an alternative that allows one to eliminate ξ^r algebraically *before* one starts constructing the effective Lagrangian [7]. Indeed, note that the covariant derivative $\nabla_r \pi$ in (13.61) is linear in ξ^r without any derivatives on it. Setting it to zero gives $\xi = \nabla \pi$, which is an algebraic condition that is covariant under all the symmetries of the EFT. This is an example of an IHC. Imposing the IHC brings us back to the nonlinear realization of Galilei symmetry solely in terms of the scalar π . Yet, invariance under Galilei boosts

is still automatically implemented. Galilei-invariant Lagrangians can be constructed using the same building blocks (13.61) while imposing by hand just invariance under spatial rotations. The latter can be ensured using standard tensor methods.

The auxiliary field ξ^r in nonrelativistic superfluids is associated with Galilei boosts. In this concrete case, the corresponding would-be NG modes are clearly unphysical. However, it seems to be true generally that there are no observed single-particle NG states for which boosts would be the sole associated broken symmetry. Curiously, the Goldstone theorem for boosts can be saturated by states from a continuum spectrum rather than by stable bosonic quasiparticles, as in Fermi liquids [8, 9]. A class of mathematically consistent EFTs that feature NG bosons originating in boost symmetry has been proposed in [10].

Eliminating the auxiliary vector field using its EoM or using an IHC must lead to EFTs with the same functional form. The only difference is possibly in the identification of the effective couplings, but these are usually determined subsequently by matching the predictions of the EFT to selected observables. This makes the two approaches entirely equivalent. The advantage of using the classical EoM (or generally, at loop level, integrating the auxiliary field out) is that it is physically transparent. The IHC, on the other hand, is usually technically simpler. The drawback is that there is no universal algorithm how to find a suitable IHC. As a rule, one has to follow intuition combined with insight in the concrete problem at hand. See [11] for a discussion of some of the traps involved.

Gapped Modes from Secondary Order Parameter Eliminating the redundancy of either type, we are back to the standard nonlinear realization of spacetime symmetry, developed in Sect. 12.3. Here all the fields, whether NG fields π^a or matter fields χ^ϱ , describe independent, physical degrees of freedom. Barring possible NG modes of spontaneously broken translation symmetry, the NG variables are classified as coordinates on the coset space $H_0/H_{(\psi_0,0)}$. However, there may be different choices of the order parameter(s) leading to the same symmetry-breaking pattern but different coset spaces $H_0/H_{(\psi_0,0)}$, hence different numbers of NG variables. This possibility owes its existence to the distinction between the isotropy groups H_0 , $H_{(\psi_0,0)}$ and the unbroken subgroup H_φ in case of spacetime symmetries.

A phenomenologically relevant example is the smectic liquid crystal, discussed in Sect. 13.3.2. (Another, relativistic example of this type can be found in [12].) Let us inductively generalize what we observed therein. Suppose there is a secondary order parameter that can be removed from the EFT without affecting the symmetry-breaking pattern. Then the associated NG variables are not required to be present by the broken symmetry. By the same token, the corresponding modes in the spectrum are not protected by the symmetry from acquiring a gap.⁷

⁷ A similar phenomenon exists even for internal symmetries. As stressed in Sect. 8.3.2, the primary order parameter responsible for type-B NG modes in the spectrum is the commutator matrix (8.26). The presence of an additional, secondary order parameter may lead to gapped “partners” of the

Here one has two choices how to go about the construction of the EFT. If it is possible to identify the secondary order parameter beforehand, one can drop it from the outset. The resulting *minimal* nonlinear realization of the given symmetry-breaking pattern then only includes genuine NG fields. Accordingly, all modes described by the EFT are gapless, that is true NG bosons. Alternatively, one may proceed with the nonlinear realization that includes the secondary order parameter. The extra would-be NG fields are then dynamical and couple to gapped modes in the spectrum. If desired, these fields can be eliminated either by using their EoM or by imposing a suitable IHC.

Example 13.7

Let us have a closer look at smectic liquid crystals (Sect. 13.3.2). Here the primary order parameter is $\langle \phi(\mathbf{x}, t) \rangle = k \mathbf{n}_0 \cdot \mathbf{x}$. This breaks rotations that do not preserve the axis of alignment defined by \mathbf{n}_0 . Moreover, it breaks translations along \mathbf{n}_0 together with the emergent shift symmetry of $\phi(x)$ down to the “diagonal” subgroup. Altogether, there are three spontaneously broken symmetry generators. One of these corresponds to the genuine NG mode described by the EFT (13.56). In addition, there are two gapped modes that can be viewed as fluctuations of the secondary order parameter, $\langle \mathbf{n}(x) \rangle = \mathbf{n}_0$.

Suppose we start from an EFT including all three degrees of freedom. This might even be desirable near the nematic–smectic phase transition at which the modes excited by the vector field $\mathbf{n}(x)$ become gapless. However, should we decide to eliminate the gapped modes, we need a covariant IHC. To that end, note that the covariant derivative $\nabla_{\underline{r}} \phi$ in (13.53) splits under the action of the isotropy group $H_{(\psi_0, 0)} \simeq \text{SO}(2)$ into a singlet, $\nabla_{\underline{a}} \phi$, and a doublet, $\nabla_{\underline{a}} \phi$. The condition $\nabla_{\underline{a}} \phi = 0$ is covariant under all the symmetries of the smectic and moreover is algebraic in π^a . In a somewhat more human notation, it is equivalent to $\mathbf{n} \times \nabla \phi = 0$. The solution to this IHC is, up to overall sign, $\mathbf{n} = \nabla \phi / |\nabla \phi|$.

In fact, the IHC can do more than to eliminate unwanted degrees of freedom. According to (13.20), covariant derivatives of NG fields π^a extracted from the ω_{\perp} part of the MC form transform exactly as matter fields with respect to some linear representation of $H_{(\psi_0, 0)}$. The very fact that the constraint $\nabla_{\underline{a}} \phi = 0$ can be solved for π^a means that the Jacobian matrix $\partial(\nabla_{\underline{a}} \phi) / \partial \pi^b$ is nonsingular, at least around the reference point $\pi^a = 0$. Instead of setting it to zero, we can therefore treat $\nabla_{\underline{a}} \phi \equiv \chi_{\underline{a}}$ as a new field variable that replaces π^a in the EFT. We conclude that the gapped degrees of freedom are indistinguishable from ordinary matter fields. This is the underlying reason why they can be removed from the EFT without spoiling the nonlinear realization of the broken symmetry. And for the same reason, it is no wonder that the broken symmetry does not force them to remain gapless.

type-B NG modes. See [13] for a detailed discussion within Hamiltonian formalism, or Sect. 8.3.1 for an outline of how these gapped partner states manifest themselves in the effective Lagrangian.

The same reasoning applies to Galilei-invariant superfluids (Sect. 13.3.3). Even if one does not wish to eliminate the auxiliary field ξ^r from the EFT, one can still trade it for $\chi \equiv \xi - \nabla\pi$. This is invariant under Galilei boosts and transforms as a matter field in the vector representation of the linearly realized isotropy group $SO(d)$. The observation that whenever a set of would-be NG fields can be eliminated from the EFT by imposing some IHCs, they can instead be kept and traded for a set of matter fields, is general [14]. The change from the would-be NG fields to the matter fields is a field redefinition that is nonsingular. This is guaranteed by the fact that the IHCs are algebraically solvable.

The distinction between the three mechanisms discussed so far crucially relies on the type of order parameter breaking the symmetry of the system. Within a naive EFT based on the agnostic nonlinear realization, there is therefore no way to distinguish these scenarios. One has to accept the possibility that some would-be NG degrees of freedom of the EFT couple to gapped states. Within the EFT itself, it is not possible to tell whether such states are physical or mere artifacts of the EFT setup. This suggests that when it comes to spacetime symmetries, the symmetry-breaking pattern itself is not restrictive enough. To get more detailed insight, one needs additional low-energy data such as the order parameter. See [15, 16] for further discussion.

Gapped Modes Required by Order Parameter So far, generators not leading to independent NG degrees of freedom could always be identified by inspecting the classical symmetry transformations and the order parameter(s). Interestingly, that is not the end of the story. There are theories featuring would-be NG fields that couple to gapped modes which cannot be discarded by dropping a secondary order parameter. See [17] for an example and more details. It is then still possible to eliminate such fields using either the EoM or a suitably chosen IHC. One should however expect the resulting minimal nonlinear realization of symmetry to involve generalized local transformations that depend on derivatives of the NG fields. As long as one insists on using solely point symmetries, it is not possible to recover the minimal nonlinear realization by reducing the order parameter. In this sense, the presence of the gapped modes is required for a description of the symmetry-breaking pattern based on a local order parameter.

In all the scenarios outlined above and summarized in Fig. 13.2, the IHCs play a prominent role. It is therefore important to stress that they are just a technical tool. We always have the option to work with the full EFT based either on the agnostic nonlinear realization or the standard nonlinear realization of Chap. 12. We just have to live with the possibility that the EFT may include some gapped modes. These of course become irrelevant at sufficiently low energies. The set of genuine NG degrees of freedom is fixed unambiguously by the symmetry-breaking pattern.

How to choose the IHC in practice is a kind of art in its own right. The examples I presented above were all mercifully simple. Nonetheless, there is a useful rule of thumb rooted in the expression (13.3) for the MC form. Namely, if the commutator $[P_\mu, Q_a]$ contains Q_b , then the NG field π^a appears linearly and without derivatives in the ω^b component of the MC form. It therefore seems possible to eliminate π^a by setting ω^b (or the irreducible multiplet of components of the MC form containing

it) to zero. This naive algorithm however requires further conditions to work [18], and moreover depends sensitively on the exact choice of field parameterization [19]. It is therefore comforting to know that whether or not one chooses to impose some IHCs has no effect on the low-energy physics of the theory.

13.4 Genuine Breaking of Translation Invariance

Until now I assumed that whatever the pattern of symmetry breaking, the low-energy physics is captured by an EFT that lives on the coset space $H_0/H_{(\psi_0,0)}$. As we saw, this is appropriate for systems where the value of the order parameter at all space-time points lies on the same orbit of the symmetry group. Any matter fields χ^{ϱ} that might be present can then be discounted. In this last section of the chapter, I will briefly outline the changes that are necessary to account for fluctuations of generic coordinate-dependent order parameters. The nonlinear realization of spacetime symmetry developed in Chap. 12 remains valid even in this general case. This is because the construction does not depend on the order parameter as a function on spacetime, merely on the choice of the representative point $\psi_0 \in \mathcal{M}$. By the same token, everything said in Sect. 13.1 about the MC form and covariant derivatives still applies.

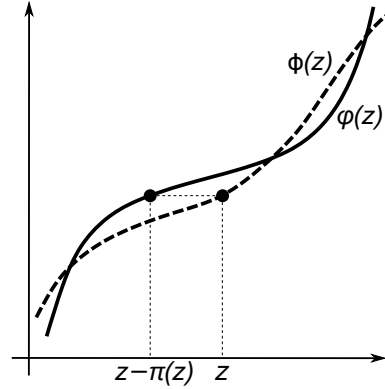
The new complication is that we now have to explicitly include the matter fields $\chi^{\varrho}(x)$ in the game. These may contain NG degrees of freedom that are not detected by the $H_0/H_{(\psi_0,0)}$ coset space, namely those of spontaneously broken translations. The challenge we face is to find a parameterization of the matter fields that makes the existence of the translation NG modes manifest. As of writing this book, there does not seem to be a general parameterization that would do the job in presence of other, possibly spontaneously broken, symmetries. I will thus merely sketch the basic idea in the simplest case of spacetime symmetry breaking by an otherwise featureless order parameter.

13.4.1 One-Dimensional Modulation of the Order Parameter

To further simplify the problem, consider a single real scalar field $\phi(x)$ in a theory invariant under a purely spacetime symmetry group G . This may be the Aristotelian group or one of its extensions including boosts such as the Galilei or Poincaré group. The isotropy group H_0 contains the transformations that act linearly on spacetime coordinates, that is rotations and boosts. For any choice of $\phi_0 \in \mathbb{R}$, the isotropy group $H_{(\phi_0,0)} \simeq H_0$ makes the coset space $H_0/H_{(\phi_0,0)}$ trivial. There are no NG variables and the sole degree of freedom, namely ϕ itself, is of the matter type.

Suppose now that the VEV $\langle \phi(x) \rangle$ varies along one particular direction in space. In order to maintain Poincaré (or Galilei) covariance if desired, I will characterize this preferred direction by a fixed spacelike vector n^{μ} . Our order parameter for

Fig. 13.3 Two parameterizations of the fluctuations of a strictly monotonic function $\varphi(z)$. The first amounts to a local shift $\pi(z)$ of the graph of $\varphi(z)$ in the horizontal direction. The second is given by the function $\phi(z)$ corresponding to the shifted graph. Mathematically, the two parameterizations are related by $\phi(z) = \varphi(z - \pi(z))$



spacetime symmetry breaking is then defined by a function φ of a single variable such that

$$\langle \phi(x) \rangle \equiv \varphi(n \cdot x) . \quad (13.72)$$

We would eventually like to trade $\phi(x)$ for a NG field $\pi(x)$ that manifests spontaneous breaking of translations in the direction of n^μ . Before we proceed, let me stress that the order parameter (13.72) will as a rule also break rotations and (if present) Lorentz or Galilei boosts. However, these never enter our setup explicitly. For scalar fields, we already know that both spatial rotations and boosts are descendant symmetries in the sense of Chap. 11. Their nonlinear realization will therefore be automatically taken care of by the same NG field $\pi(x)$.

Let us now start with a purely mathematical problem on functions of a single variable. I will call the latter z to distinguish it from the spacetime or spatial coordinates, x^μ and x^r . Suppose we are given a function $\varphi(z)$ that “breaks translation invariance,” that is, it is strictly monotonic, $\varphi'(z) \neq 0$. We would like to parameterize the “fluctuations” of $\varphi(z)$ in a way that makes the breaking of translation invariance manifest. We do so by locally shifting the graph of the function by $\pi(z)$ along the z -axis. This defines uniquely a new function $\phi(z)$ via $\phi(z) \equiv \varphi(z - \pi(z))$, see Fig. 13.3. Thanks to the assumed strict monotonicity of $\varphi(z)$, the mapping between $\pi(z)$ and $\phi(z)$ for fixed $\varphi(z)$ is one-to-one.

This little trick can be promoted to an arbitrary number D of spacetime dimensions, provided φ as assumed only varies in one particular direction. The coordinates corresponding to all the other directions in spacetime play a role of fixed parameters in the mapping between ϕ and π . Thus, I assume that $\varphi(n \cdot x)$ in (13.72) is a strictly monotonic function of its argument. This is necessary for translations along n^μ to be spontaneously broken at each spacetime point. The scalar field $\phi(x)$ can then be traded for $\pi(x)$, defined implicitly by

$$\phi(x) \equiv \varphi(n \cdot x - \pi(x)) . \quad (13.73)$$

The function φ as defined by (13.72) obviously depends on the choice of coordinates. For instance, $\varphi(n \cdot x)$ and $\varphi(n \cdot x + c)$ with any $c \in \mathbb{R}$ describe the same order

parameter, differing only by a coordinate shift along n^μ . This is a hallmark of translation symmetry breaking. The introduction of the NG field $\pi(x)$ makes it possible to map a *fixed* function φ to a *scalar* field ϕ . This requires that under a spacetime transformation, be it a translation, rotation or boost, the NG field transforms as

$$\pi'(x') = \pi(x) + n' \cdot x' - n \cdot x . \quad (13.74)$$

In the special case of spatial translations, $x \rightarrow x + \epsilon$, this boils down to $\pi'(x + \epsilon, t) = \pi(x, t) - n \cdot \epsilon$, where n is the spatial part of n^μ . Shifting the NG field by a constant is typical for spontaneously broken Abelian symmetries. We are on the right track.

Now that we have settled the parameterization, we would like to construct the most general effective action consistent with all the symmetries present. Since there were no NG variables to start with, the MC form (13.3) reduces to $P \cdot dx$. This gives a trivial spacetime coframe, $e^{*\mu} = \delta^\mu_\nu dx^\nu$. Also, by (13.17), the covariant derivative of the matter field ϕ is trivial, $\nabla_\mu \phi = \delta^\nu_\mu \partial_\nu \phi$. Therefore, an invariant action can be built by taking any Lagrangian density that does not depend explicitly on spacetime coordinates. It can contain $\phi(x)$ with an arbitrary number of derivatives (including zero). The indices on the derivatives are to be contracted in a way that preserves the linearly realized group H_0 . In case of a relativistic, that is Poincaré-invariant theory, any Lorentz-invariant Lagrangian density $\mathcal{L}_{\text{eff}}[\phi]$ will do.

It looks like we have not made much progress. In particular, we have not learned anything so far about the possible dependence of the effective Lagrangian on the NG field $\pi(x)$. However, there are still two requirements we have not addressed yet. The first of these is that the order parameter (13.72) is at least a local minimum of the energy functional of the EFT. This will constrain what the effective Lagrangian may look like. The second requirement is to have a consistent power-counting scheme that would allow us to decide which operators in the Lagrangian are most important. In order to understand what is at stake, it is best to work out an illustrative example.

13.4.2 Case Study: Fluctuations of a Domain Wall

Following Chap. 5 of [20], consider the class of relativistic scalar field theory models

$$\mathcal{L}_{\text{eff}} = \frac{1}{2}(\partial_\mu \phi)^2 - V(\phi) , \quad (13.75)$$

where $V(\phi)$ is a potential function, required for stability to be bounded from below. We can assume without loss of generality that $\min V(\phi) = 0$. The classical EoM of the model reads

$$\square \phi + \frac{dV(\phi)}{d\phi} = 0 . \quad (13.76)$$

Suppose the potential has (at least) two degenerate minima. Then the EoM has *domain wall* (or *kink*) solutions φ that interpolate between the minima. Such solutions depend on a single spatial coordinate, which I will denote as z , and converge to the

respective minima of the potential for $z \rightarrow \pm\infty$. Upon reduction to the single variable z , the EoM (13.76) has the first integral $(1/2)(d\phi/dz)^2 - V(\phi)$. For the domain wall solutions, this first integral evaluates to zero since for $z \rightarrow \pm\infty$, $\varphi(z)$ converges to a constant such that $\lim_{z \rightarrow \pm\infty} V(\varphi(z)) = 0$. This reduces the EoM to the following first-order differential equation for $\varphi(z)$,

$$\varphi'(z) = \pm\sqrt{2V(\varphi(z))} . \quad (13.77)$$

The prime on φ indicates a derivative with respect to its argument.

Example 13.8

One of the most common scalar potentials is the double-well potential,

$$V(\phi) \equiv \frac{\lambda}{2}(\phi^2 - v^2)^2 , \quad (13.78)$$

where λ and v are positive parameters. A simple calculation gives the corresponding domain wall solution, interpolating between the minima at $\phi = \pm v$,

$$\varphi(z) = \pm v \tanh[\sqrt{\lambda}v(z - z_0)] . \quad (13.79)$$

Here z_0 is an integration constant that determines the center of the domain wall where $\varphi(z_0) = 0$. The width of the domain wall is given by $1/(\sqrt{\lambda}v)$.

Another example of a phenomenologically important potential is

$$V(\phi) \equiv m^2 v^2 \left(1 - \cos \frac{\phi}{v}\right) , \quad (13.80)$$

where m and v are positive parameters. This appears for instance in the low-energy effective theory (9.21) of quantum chromodynamics, when restricted to neutral pions. In this realization, m corresponds to the pion mass m_π and v to the pion decay constant, f_π . The potential $V(\phi)$ is now periodic with minima at $\phi = 2n\pi v$ for any $n \in \mathbb{Z}$. The domain wall solution, connecting two neighboring minima, is

$$\varphi(z) = 4v \arctan \exp[\pm m(z - z_0)] . \quad (13.81)$$

In this case, the width of the domain wall is fixed by the mass parameter to be $1/m$.

Both concrete domain wall solutions in Example 13.8 satisfy our criteria for the translation-breaking order parameter modulated in one dimension. We would eventually like to derive a low-energy EFT for the fluctuations of the domain wall. The first step towards this goal is to use the general parameterization (13.73) with $n \cdot x = z$. The potential term in (13.75) can then be eliminated by means of (13.77), upon which the Lagrangian becomes

$$\mathcal{L}_{\text{eff}} = [\varphi'(z - \pi)]^2 \left[\frac{1}{2} (\partial_\mu \pi)^2 - (1 - \partial_z \pi) \right]. \quad (13.82)$$

The contribution of the $1 - \partial_z \pi$ piece to action is a mere boundary term independent of local variations of $\pi(x)$. Namely, $\int dz [\varphi'(z - \pi)]^2 (1 - \partial_z \pi) = \int_{\tilde{z}_-}^{\tilde{z}_+} d\tilde{z} [\varphi'(\tilde{z})]^2$ where $\tilde{z} \equiv z - \pi(x)$ and $\tilde{z}_\pm \equiv \lim_{z \rightarrow \pm\infty} \tilde{z}$. Dropping this contribution, we are left with the following effective action [16],

$$S_{\text{eff}}\{\pi\} = \frac{1}{2} \int d^D x [\varphi'(z - \pi(x))]^2 [\partial_\mu \pi(x)]^2. \quad (13.83)$$

The EFT (13.83) apparently describes a derivatively coupled NG field $\pi(x)$ as we wanted. However, it is still not quite satisfactory due to the explicit appearance of the domain wall background $\varphi(z)$. It is not even clear what the quasiparticle spectrum of the EFT is. To that end, we utilize the EoM, linearized in the NG field,

$$\square \pi(x) - \frac{2\varphi''(z)}{\varphi'(z)} \partial_z \pi(x) \approx 0, \quad (13.84)$$

where the \approx symbol indicates the linear approximation. It is convenient to switch to the linear fluctuation of the order parameter, $\chi(x) \equiv \phi(x) - \varphi(z)$.⁸ The linearized EoM (13.84) is thus transformed to

$$\square \chi(x) + \frac{\varphi'''(z)}{\varphi'(z)} \chi(x) \approx 0, \quad \chi(x) \approx -\varphi'(z) \pi(x). \quad (13.85)$$

As a simple check, note that the domain wall solution satisfies $\varphi'''/\varphi' = d^2 V(\varphi)/d\varphi^2$. Equation (13.85) then descends directly from the Lagrangian (13.75) by expanding the latter to second order in $\chi(x)$. The form of (13.85) allows us to Fourier-transform in time and the transverse spatial coordinates (denoted collectively as \mathbf{x}_\perp). Parameterizing plane-wave solutions by energy E and transverse momentum \mathbf{p}_\perp ,

$$\chi(\mathbf{x}_\perp, z, t) = \hat{\chi}(z) \exp(-iEt + i\mathbf{p}_\perp \cdot \mathbf{x}_\perp), \quad (13.86)$$

the profile function $\hat{\chi}(z)$ solves the one-dimensional eigenvalue problem

$$\left[-\partial_z^2 + \frac{\varphi'''(z)}{\varphi'(z)} \right] \hat{\chi}(z) = (E^2 - \mathbf{p}_\perp^2) \hat{\chi}(z). \quad (13.87)$$

In order to appreciate the content of this equation for single-particle states, let us return to the two concrete choices of potential introduced in Example 13.8. In both cases, I will for simplicity consider only the solution $\varphi(z)$ centered at $z_0 = 0$.

⁸ While $\pi(z)$ is defined as the horizontal distance of the graphs of $\phi(z)$ and $\varphi(z)$ in Fig. 13.3 (the other spacetime coordinates being fixed), $\chi(z)$ can be viewed as their vertical distance.

Example 13.9

For the double-well potential with the corresponding domain wall solution (13.79), $\varphi'''(z)/\varphi'(z) = 2\lambda v^2[2 - 3 \operatorname{sech}^2(\sqrt{\lambda} v z)]$. Introducing a dimensionless coordinate $Z \equiv \sqrt{\lambda} v z$ turns (13.87) into

$$\left(-\partial_Z^2 - \frac{6}{\cosh^2 Z}\right) \hat{\chi}(Z) = \left(\frac{E^2 - \mathbf{p}_\perp^2}{\lambda v^2} - 4\right) \hat{\chi}(Z). \quad (13.88)$$

The operator on the left-hand side is a special case of the *Pöschl–Teller Hamiltonian*, $H_n \equiv -\partial_Z^2 - n(n+1)/\cosh^2 Z$ with positive $n \in \mathbb{Z}$. The spectrum of this Hamiltonian is well-known; see e.g. Chap. 11 of [21]. There are n bound states with eigenvalues $-n^2, -(n-1)^2, \dots, -1$. The continuous part of the spectrum covers the open interval $(0, \infty)$. In the $n = 2$ case of interest to us, the ground state with eigenvalue -4 gives $E = |\mathbf{p}_\perp|$. The corresponding profile function is $\hat{\chi}(Z) \propto 1/\cosh^2 Z$. This solution is localized in the z -direction to the vicinity of the domain wall. It propagates as a plane wave in the transverse directions with a massless relativistic dispersion relation. The upper bound state of H_2 with eigenvalue -1 gives $E^2 = \mathbf{p}_\perp^2 + 3\lambda v^2$. Its profile function is $\hat{\chi}(Z) \propto \sinh Z/\cosh^2 Z$. This solution is also localized to the domain wall and propagates only in the transverse directions, but it has a nonzero mass of $\sqrt{3}\lambda v$. Finally, the continuum part of the spectrum of the Pöschl–Teller Hamiltonian can be parameterized by the momentum variable p_z so that for $z \rightarrow \pm\infty$, $\hat{\chi}(z) \propto \exp(ip_z z)$. The corresponding energy is $E^2 = \mathbf{p}_\perp^2 + p_z^2 + 4\lambda v^2$. This describes delocalized fluctuations of the domain wall that propagate in the entire space. They have a relativistic dispersion relation with mass $2\sqrt{\lambda} v$.

In case of the cosine potential with its domain wall solution (13.81), $\varphi'''(z)/\varphi'(z) = m^2[1 - 2 \operatorname{sech}^2(mz)]$. With the dimensionless coordinate $Z \equiv mz$, the eigenvalue problem (13.87) becomes

$$\left(-\partial_Z^2 - \frac{2}{\cosh^2 Z}\right) \hat{\chi}(Z) = \left(\frac{E^2 - \mathbf{p}_\perp^2}{m^2} - 1\right) \hat{\chi}(Z). \quad (13.89)$$

The left-hand side is again one of the Pöschl–Teller Hamiltonians, H_1 . Its sole bound state has the profile function $\hat{\chi}(Z) \propto 1/\cosh Z$. This is a state localized to the domain wall that propagates in the transverse directions with the massless relativistic dispersion, $E = |\mathbf{p}_\perp|$. The continuum part of the spectrum of this Pöschl–Teller Hamiltonian again amounts to delocalized fluctuations of the domain wall. Their dispersion relation is $E^2 = \mathbf{p}_\perp^2 + p_z^2 + m^2$. Far away from the domain wall, they describe particles of mass m propagating in the bulk.

We are now ready to draw some general conclusions. Starting with any potential $V(\phi)$ with degenerate minima, the domain wall solution will be localized in the z -direction. The linearized EoM (13.84) always admits plane-wave solutions, propagating in the transverse directions with the massless dispersion relation, $E = |\mathbf{p}_\perp|$, and with a constant longitudinal profile, $\hat{\pi}(z) \propto 1$. By (13.85), this corresponds to

$\hat{\chi}(z) \propto \varphi'(z)$. The existence of this solution is guaranteed by the spontaneously broken translation invariance. Indeed, the fact that $\hat{\pi}(z)$ is constant indicates that this is the NG boson we are after. It does not propagate in the z -direction since translations in the z -direction are completely broken, hence there is no physical momentum p_z to label one-particle states with. The NG mode propagates solely along the directions in which translations remain unbroken.

The second observation is that there is a continuum of plane-wave-like solutions that propagate in the bulk. Far away from the domain wall where $\varphi(z)$ goes to one of the minima of $V(\phi)$, φ_0 , these are simple plane waves, $\hat{\chi}(z) \propto \exp(ip_z z)$. They have a relativistic dispersion relation with squared mass $V''(\varphi_0)$. In fact, we could have guessed this much by expanding the Lagrangian (13.75) around the uniform solution $\langle \phi(x) \rangle = \varphi_0$. In addition to the NG mode and the bulk plane-wave continuum, there may be further solutions to (13.85) that are localized on the domain wall and propagate only in the transverse directions. These correspond to excited bound states of the one-dimensional Hamiltonian on the left-hand side of (13.87). The presence of such massive excitations propagating along the domain wall depends on the concrete choice of the potential $V(\phi)$.

It is not a priori obvious why the NG mode with $\hat{\pi}(z) \propto 1$ is necessarily the lowest-lying excitation, that is, why the domain wall solution is stable. To understand this, let us assume without loss of generality that $\phi(x)$ is monotonic in z and estimate its energy from below by completing the square,

$$\begin{aligned} & \int d^{d-1}\mathbf{x}_\perp dz \left[\frac{1}{2}(\partial_0\phi)^2 + \frac{1}{2}(\nabla_\perp\phi)^2 + \frac{1}{2}(\partial_z\phi)^2 + V(\phi) \right] \\ &= \int d^{d-1}\mathbf{x}_\perp dz \left\{ \frac{1}{2}(\partial_0\phi)^2 + \frac{1}{2}(\nabla_\perp\phi)^2 + \frac{1}{2}[\phi' \mp \sqrt{2V(\phi)}]^2 \pm \phi' \sqrt{2V(\phi)} \right\} \\ &\geq \int d^{d-1}\mathbf{x}_\perp \left| \int_{\phi_-}^{\phi_+} d\phi \sqrt{2V(\phi)} \right|, \end{aligned} \quad (13.90)$$

where $\phi_\pm \equiv \lim_{z \rightarrow \pm\infty} \phi(x)$. This estimate, known as the *Bogomolny bound*, shows that static solutions satisfying (13.77) minimize the energy per unit transverse volume on the class of field configurations with a fixed boundary condition at $z \rightarrow \pm\infty$.

We have discovered that our EFT (13.83) inevitably describes both the translation NG boson and specific gapped excitations. This is hardly surprising given that (13.83) is equivalent via a field redefinition to the original model (13.75). One might naively expect that an EFT for the NG mode alone could be extracted by integrating over z . After all, we found that the NG solution for $\pi(x)$ does not depend on z at all. However, inserting a z -independent $\pi(\mathbf{x}_\perp, t)$ makes it possible to shift the z -coordinate in $\varphi'(z - \pi)$, thus reducing (13.83) to

$$S_{\text{eff}}\{\pi\} = \frac{1}{2} \int dz [\varphi'(z)]^2 \int d^{d-1}\mathbf{x}_\perp dt [(\partial_0\pi)^2 - (\nabla_\perp\pi)^2]. \quad (13.91)$$

This is a noninteracting theory, describing free surface waves on the domain wall. We conclude that *all* interactions among the NG modes in the EFT (13.83) are mediated by the gapped modes.

The problem of disentangling the NG mode from the gapped excitations is related to our ability to organize contributions to the EFT by a derivative expansion. Also, we have so far been forced to work with specific domain wall solutions $\varphi(z)$, descending from the chosen model potential $V(\phi)$. We would rather have a model-independent EFT where the stability of the ground state and the existence of a power counting serve merely as constraints. I will now address these issues jointly in a framework inspired by the background gauge invariance approach of Sect. 8.2.

13.4.3 Effective Action from Background Gauge Invariance

The idea to use background gauging to derive an EFT for spontaneously broken translations has been utilized in multiple contexts, including inflationary cosmology [22] and superconductivity [23]. Here I will loosely follow the exposition in [16]. We are looking for an effective action $S_{\text{eff}}\{\pi\}$ for the NG mode of spontaneously broken translations in the z -direction. In the spirit of Sect. 8.2, we first promote the physical symmetry to a local invariance by adding a set of background gauge fields. Localizing spacetime translations leads to the group of spacetime diffeomorphisms. The precise choice of the corresponding background field is constrained by the physical symmetry and field content. In case of relativistic theories of scalar fields, one can use a spacetime metric, $g_{\mu\nu}(x)$, to make the gauged action $S_{\text{eff}}\{\pi, g\}$ diffeomorphism-invariant.

The transformation of the metric under a diffeomorphism of the spacetime manifold, $x^\mu \rightarrow x'^\mu(x) \equiv x^\mu + \epsilon^\mu(x)$, is given by the push-forward by the inverse of the diffeomorphism; see Appendix A.3.1. This must be augmented with a transformation rule for the NG field $\pi(x)$. In order to maintain the relation $\phi(x) = \varphi(z - \pi(x))$ between a scalar field $\phi(x)$ and a fixed function $\varphi(z)$, $\pi(x)$ should transform as

$$\pi'(x') = \pi(x) + \epsilon^z(x) . \quad (13.92)$$

The NG field can now be eliminated by choosing the diffeomorphism as $\epsilon^\mu(x) = -\delta_z^\mu \pi(x)$. With a slight abuse of notation, I will denote this operation as \mathcal{T}_π . This reduces the action to $S_{\text{eff}}\{\pi, g\} = S_{\text{eff}}\{0, \mathcal{T}_\pi g\}$. The dependence of the action on the composite gauge field $(\mathcal{T}_\pi g)_{\mu\nu}$ is constrained by invariance under the residual group of diffeomorphisms that fix the point $\pi = 0$. These are (possibly z -dependent) $(D-1)$ -dimensional diffeomorphisms with $\epsilon^z(x) = 0$. Once the most general admissible action $S_{\text{eff}}\{0, \mathcal{T}_\pi g\}$ has been found, the EFT in flat Minkowski spacetime can be recovered by setting $g_{\mu\nu}(x)$ to the Minkowski metric.

To constrain the form of $S_{\text{eff}}\{0, \mathcal{T}_\pi g\}$, it is convenient to separate the z -coordinate from the other spacetime coordinates. I will indicate those using lowercase Fraktur

indices $\mathbf{m}, \mathbf{n}, \dots$. It will also be more practical to work with the inverse of the metric, $g^{\mu\nu}(x)$, rather than $g_{\mu\nu}(x)$ itself. The transformation of the metric under diffeomorphisms then splits as

$$\begin{aligned} g'^{\mathbf{mn}} &= g^{\mathbf{ab}}(\delta_{\mathbf{a}}^{\mathbf{m}} + \partial_{\mathbf{a}}\epsilon^{\mathbf{m}})(\delta_{\mathbf{b}}^{\mathbf{n}} + \partial_{\mathbf{b}}\epsilon^{\mathbf{n}}) + g^{\mathbf{az}}(\delta_{\mathbf{a}}^{\mathbf{m}} + \partial_{\mathbf{a}}\epsilon^{\mathbf{m}})\partial_z\epsilon^{\mathbf{n}} + g^{\mathbf{zb}}\partial_z\epsilon^{\mathbf{m}}(\delta_{\mathbf{b}}^{\mathbf{n}} + \partial_{\mathbf{b}}\epsilon^{\mathbf{n}}) \\ &\quad + g^{\mathbf{zz}}\partial_z\epsilon^{\mathbf{m}}\partial_z\epsilon^{\mathbf{n}}, \\ g'^{\mathbf{mz}} &= g^{\mathbf{ab}}(\delta_{\mathbf{a}}^{\mathbf{m}} + \partial_{\mathbf{a}}\epsilon^{\mathbf{m}})\partial_{\mathbf{b}}\epsilon^{\mathbf{z}} + g^{\mathbf{az}}(\delta_{\mathbf{a}}^{\mathbf{m}} + \partial_{\mathbf{a}}\epsilon^{\mathbf{m}})(1 + \partial_z\epsilon^{\mathbf{z}}) + g^{\mathbf{zb}}\partial_z\epsilon^{\mathbf{m}}\partial_{\mathbf{b}}\epsilon^{\mathbf{z}} \\ &\quad + g^{\mathbf{zz}}\partial_z\epsilon^{\mathbf{m}}(1 + \partial_z\epsilon^{\mathbf{z}}), \\ g'^{\mathbf{zz}} &= g^{\mathbf{ab}}\partial_{\mathbf{a}}\epsilon^{\mathbf{z}}\partial_{\mathbf{b}}\epsilon^{\mathbf{z}} + g^{\mathbf{az}}\partial_{\mathbf{a}}\epsilon^{\mathbf{z}}(1 + \partial_z\epsilon^{\mathbf{z}}) + g^{\mathbf{zb}}(1 + \partial_z\epsilon^{\mathbf{z}})\partial_{\mathbf{b}}\epsilon^{\mathbf{z}} + g^{\mathbf{zz}}(1 + \partial_z\epsilon^{\mathbf{z}})^2. \end{aligned} \quad (13.93)$$

For the sake of brevity, I suppressed the arguments x of $g^{\mu\nu}(x)$ and x' of $g'^{\mu\nu}(x')$. The transformation rule for $g^{\mathbf{mz}}$ is analogous to that for $g^{\mathbf{mz}}$.

Under the restricted $(D-1)$ -dimensional diffeomorphisms with $\epsilon^z(x) = 0$, $g^{\mathbf{zz}}(x)$ behaves as a scalar, $g'^{\mathbf{zz}}(x') = g^{\mathbf{zz}}(x)$. On the other hand, $g'^{\mathbf{mz}} = g^{\mathbf{az}}(\delta_{\mathbf{a}}^{\mathbf{m}} + \partial_{\mathbf{a}}\epsilon^{\mathbf{m}})$ depends on both $g^{\mathbf{mz}}$ and $g^{\mathbf{zz}}$. Finally, $g'^{\mathbf{mn}} = g^{\mathbf{ab}}(\delta_{\mathbf{a}}^{\mathbf{m}} + \partial_{\mathbf{a}}\epsilon^{\mathbf{m}})(\delta_{\mathbf{b}}^{\mathbf{n}} + \partial_{\mathbf{b}}\epsilon^{\mathbf{n}})$ depends on all $g^{\mathbf{mn}}, g^{\mathbf{mz}}, g^{\mathbf{zn}}$ and $g^{\mathbf{zz}}$. This suggests that our main ingredient for constructing diffeomorphism-invariant actions will be

$$\begin{aligned} (\mathcal{T}_{-\pi}g)^{\mathbf{zz}} &= g^{\mathbf{ab}}\partial_{\mathbf{a}}\pi\partial_{\mathbf{b}}\pi - 2g^{\mathbf{az}}\partial_{\mathbf{a}}\pi(1 - \partial_z\pi) + g^{\mathbf{zz}}(1 - \partial_z\pi)^2 \\ &= g^{\mathbf{zz}} - 2g^{\mathbf{z\alpha}}\partial_{\alpha}\pi + g^{\mathbf{ab}}\partial_{\mathbf{a}}\pi\partial_{\mathbf{b}}\pi, \end{aligned} \quad (13.94)$$

where the argument of $(\mathcal{T}_{-\pi}g)^{\mathbf{zz}}$ is $\tilde{x}^{\mu} \equiv x^{\mu} - \delta_z^{\mu}\pi(x)$. The z -component of \tilde{x}^{μ} , that is $\tilde{z} \equiv z - \pi(x)$, is itself invariant under spacetime diffeomorphisms and can thus appear in the effective action without restrictions. Finally, the effective action may also contain $(\mathcal{T}_{-\pi}g)^{\mathbf{mz}}$ and $(\mathcal{T}_{-\pi}g)^{\mathbf{mn}}$. These, not being scalars, can however only enter through higher-derivative tensors such as the Riemann curvature tensor. See Appendix A of [22] for a discussion of geometric structures allowed by the restricted $(D-1)$ -dimensional diffeomorphism invariance. For our purposes, the main conclusion is that the part of the effective action dominant at low energies will be included in

$$\begin{aligned} S_{\text{eff}}\{\pi, g\} &= \int d^D\tilde{x} \text{vol}(\mathcal{T}_{-\pi}g)(\tilde{x}) f(\tilde{z}, (\mathcal{T}_{-\pi}g)^{\mathbf{zz}}(\tilde{x})) \\ &= \int d^Dx \text{vol}(g)(x) f(z - \pi(x), (\mathcal{T}_{-\pi}g)^{\mathbf{zz}}(\tilde{x})). \end{aligned} \quad (13.95)$$

Here f is a generic smooth function of two variables. The volume measure $\text{vol}(g)$ for a metric of (timelike) Lorentzian signature equals $\sqrt{(-1)^d \det g}$; cf. Appendix A.7.2.

Example 13.10

Using the specific choice $f(u, v) \equiv (1/2)[\varphi'(u)]^2(v-1)$ and subsequently going back to the flat Minkowski spacetime gives

$$S_{\text{eff}}\{\pi\} = \frac{1}{2} \int d^D x [\varphi'(z - \pi)]^2 [(\partial_\mu \pi)^2 - 2(1 - \partial_z \pi)] . \quad (13.96)$$

This reproduces our previous model result (13.82).

A systematic expansion in derivatives of the NG field is easier if one uses the combination $(\mathcal{T}_{\pi g})^{zz} + 1$, which goes to $2\partial_z \pi + (\partial_\mu \pi)^2$ in the flat-spacetime limit. This suggests expanding the function $f(u, v)$ in (13.95) in powers of $v + 1$. Up to second order in $v + 1$ and thus in derivatives, we have $f(u, v) = c_0(u) + c_1(u)(v + 1) + c_2(u)(v + 1)^2 + \dots$, which translates to the Minkowski-spacetime action

$$S_{\text{eff}}\{\pi\} = \int d^D x \{ c_0(z - \pi) + c_1(z - \pi)[2\partial_z \pi + (\partial_\mu \pi)^2] \\ + c_2(z - \pi)[2\partial_z \pi + (\partial_\mu \pi)^2]^2 \} + \dots . \quad (13.97)$$

This is the low-energy EFT we have been looking for. It remains to elucidate the consistency constraints on the functions $c_i(z - \pi)$. To that end, we expand the Lagrangian in (13.97) to second order in the NG field,

$$\mathcal{L}_{\text{eff}}[\pi] = c_0(z) + [-c'_0(z) + 2c_1(z)\partial_z]\pi \\ + \frac{1}{2}[c''_0(z) - 2c'_1(z)\partial_z]\pi^2 + c_1(z)(\partial_\mu \pi)^2 + 4c_2(z)(\partial_z \pi)^2 + \dots . \quad (13.98)$$

Bulk stability requires absence of any terms linear in π . This leads to the constraint

$$c'_0(z) + 2c'_1(z) = 0 . \quad (13.99)$$

In fact, one can demand even $c_0(z) + 2c_1(z) = 0$, since any constant offset of $c_0(z)$ can be dropped from (13.97). This sharper relation is indeed satisfied by the choice of $f(u, v)$ in Example 13.10, where $c_0(u) = -[\varphi'(u)]^2$ and $c_1(u) = (1/2)[\varphi'(u)]^2$. With the constraint (13.99), the first term on the second line of (13.98) automatically drops, and the Lagrangian boils down to $c_1(z)(\partial_\mu \pi)^2 + 4c_2(z)(\partial_z \pi)^2$ up to a surface term. A necessary (though not inevitably sufficient) condition for the corresponding energy functional to be bounded from below is therefore

$$c_1(z) \geq 0 \quad \text{and} \quad c_1(z) \geq 4c_2(z) \quad \text{for any } z \in \mathbb{R} . \quad (13.100)$$

So far, I have not assumed any specific choice or properties of the profile functions $c_i(z)$. The effective action (13.97) defines a generic EFT for the NG mode of spontaneously broken translations in the z -direction in a relativistic system. In particular, the functions $c_i(z)$ in (13.97) do not have to arise from a topologically nontrivial background such as a domain wall. However, if that happens to be the case, we expect them to be localized to a finite range in z . The NG mode will then correspond to a surface wave and propagate solely in the transverse directions. I would now like to use the general EFT (13.95) to shed some light on the physics of these surface waves beyond the noninteracting approximation (13.91). To that end, it is convenient to

switch to the \tilde{x}^μ coordinates. Namely, $\pi(x)$ measures the local fluctuation-induced displacement of the domain wall in the z -direction. Geometrically, the dynamical state of the domain wall can thus be viewed as a hypersurface in \mathbb{R}^D defined by a constant value of \tilde{z} .

With the shorthand notation $\tilde{g} \equiv \mathcal{T}_{-\pi} g$, the determinant of this D -dimensional metric can be decomposed as $\det \tilde{g} = (\tilde{g}^{zz})^{-1} \det \tilde{\mathbf{g}}$. Here $\tilde{\mathbf{g}}_{mn}$ is a projection of $\tilde{g}_{\mu\nu}$ to the $(D-1)$ -dimensional space of x^m , which is nothing but the metric on the domain wall induced by the bulk Lorentzian metric in \mathbb{R}^D . The factorization of $\det \tilde{g}$ follows from the mathematical properties of block matrices [24]. Augmenting this observation with the constraint $c_0(z) + 2c_1(z) = 0$ turns the first line of (13.95) to

$$S_{\text{eff}}\{\pi, g\} = \int d^D \tilde{x} \, \text{vol}(\tilde{\mathbf{g}})(\tilde{x}) \times \left\{ c_0(\tilde{z}) + \left[\frac{c_0(\tilde{z})}{8} + c_2(\tilde{z}) \right] [\tilde{g}^{zz}(\tilde{x}) + 1]^2 + \dots \right\}. \quad (13.101)$$

So far this is an exact rewriting of (13.95) except for the truncation of the power expansion in $\tilde{g}^{zz} + 1$. Now we take the limit of flat spacetime. Simultaneously, we restrict to fields independent of the \tilde{z} coordinate, as appropriate for the NG mode. This makes it possible to do the integral over \tilde{z} independently of the NG field. We thus arrive at an effective $(D-1)$ -dimensional theory for the surface waves on the domain wall which, up to higher-order corrections, takes the form

$$S_{\text{eff}} = \int dz \, c_0(z) \int d^{d-1} \mathbf{x}_\perp dt \, \text{vol}(\tilde{\mathbf{g}})(\mathbf{x}_\perp, t) + \dots. \quad (13.102)$$

In this approximation, the EFT for the domain wall fluctuations is completely geometric and fixed by the induced metric,

$$\tilde{\mathbf{g}}_{mn} = g_{mn} - \partial_m \pi \partial_n \pi, \quad \text{vol}(\tilde{\mathbf{g}}) = \sqrt{1 - g^{mn} \partial_m \pi \partial_n \pi}. \quad (13.103)$$

The expression for $\text{vol}(\tilde{\mathbf{g}})$ follows from the so-called Weinstein–Aronszajn identity for matrix determinants. The reader may recognize this as the *Dirac–Born–Infeld* (DBI) theory that we met previously in Sect. 10.3.1. All that is left of the profile of the domain wall is the overall prefactor in (13.102).

Example 13.11

The prefactor $\int dz \, c_0(z)$ measures, up to overall sign, the energy of the domain wall per unit transverse volume, that is its surface tension. For the class of models (13.75), it equals $-\int dz [\varphi'(z)]^2$; see also (13.90). In case of the double-well potential with the domain wall solution (13.79), this gives $-(4/3)\sqrt{\lambda}v^3$. For the cosine potential with the corresponding domain wall (13.81), one finds analogously $-8mv^2$.

The $(D - 1)$ -dimensional theory (13.102) obviously describes only the NG mode as we wanted; the gapped modes we found in Sect. 13.4.2 are gone. The price to pay is that the D -dimensional Lorentz symmetry is now realized in a nontrivial fashion beyond our spacetime symmetry paradigm; cf. (10.50).

The fact that all the modes propagating in the D -dimensional bulk are gapped is essential for the validity of the EFT (13.102). There are physical systems where the bulk modes are naturally gapless, for instance when two immiscible superfluids are separated by an interface. In this case, integrating the bulk modes out makes the EFT for the surface waves on the interface non-local. This leads to a fractional-power dispersion relation of the surface modes, $E(\mathbf{p}) \propto |\mathbf{p}|^{3/2}$. See [25] for a detailed discussion from an EFT perspective.

13.4.4 Further Possible Applications

In discussing spontaneous breaking of spatial translations, I have deliberately merely outlined the general approach and then focused on a specific type of system to illustrate it. There are however several natural, physically motivated modifications or generalizations of the setup. I will now at least briefly mention some of the possible avenues one might wish to follow.

Periodic Modulation of Order Parameter Mathematically the simplest modification is one with a single, otherwise featureless real order parameter where the profile $\varphi(z)$, or $c_0(z)$, is not spatially localized. The most interesting situation arises when $\varphi(z)$ is a periodic function of z . In this case the physics changes qualitatively. The eigenvalue problem (13.87) for linear fluctuations will have no bound states. Instead, its continuous spectrum will have a band structure. There should still be a NG mode of the spontaneously broken translations, but this will assume the form of a sound wave (phonon) on the crystalline background $\varphi(z)$. In the limit where $\varphi(z)$ consists of widely separated localized kinks, the propagation of sound in the z -direction arises from tunneling between states localized on the individual kinks.

The detailed structure of the dispersion relation of the low-lying phonon excitations will be no less interesting. Namely, it turns out that the spontaneously broken symmetry under spatial rotations forbids the c_1 term in (13.97). As a consequence, the derivative expansion of the effective Lagrangian starts at second order in longitudinal derivatives ∂_z but fourth order in the transverse gradient ∇_\perp [16]. We already saw the same behavior in smectic liquid crystals (Sect. 13.3.2), and I will return to its physical implications in Chap. 15.

Other Spacetime Symmetries Another natural possibility is to consider systems with a real order parameter modulated in one spatial dimension, but with a different spacetime symmetry than Poincaré. This is relevant in particular for any condensed-matter system. As long as one has a concrete microscopic model, the

analysis proceeds along the same steps as in Sect. 13.4.2. One can however also adopt the model-independent approach of Sect. 13.4.3 based on background gauge invariance. This requires coupling the theory to an appropriate spacetime geometry. A basic discussion can be found in Appendix A of [16]. For a more complete, if mathematically also more advanced, overview of non-Lorentzian geometry, see [26].

Additional Degrees of Freedom One might also want to study spontaneous breaking of translation invariance alongside other, possibly internal, broken symmetries. The field parameterization required then follows the general standard nonlinear realization of Sect. 12.3 combined with the identification of the translation NG field à la Sect. 13.4.1. The background gauge invariance approach may again prove helpful. Its application however relies on our ability to simultaneously gauge all the relevant symmetries by adding suitable background gauge fields. This is not guaranteed a priori without further qualifications. Finally, there are also physical systems where spacetime translations are spontaneously broken in more than one direction. In this case, the idea behind Sect. 13.4.1 might still be possible to apply provided one has at hand a sufficient number of fields to parameterize the order fluctuations uniquely.

References

1. Y. Togawa, Y. Kousaka, K. Inoue, J.i. Kishine, J. Phys. Soc. Jpn. **85**, 112001 (2016). DOI [10.7566/JPSJ.85.112001](https://doi.org/10.7566/JPSJ.85.112001)
2. A. Schmitt, *Introduction to Superfluidity: Field-theoretical Approach and Applications, Lecture Notes in Physics*, vol. 888 (Springer, 2015). DOI [10.1007/978-3-319-07947-9](https://doi.org/10.1007/978-3-319-07947-9)
3. D.T. Son, [arXiv:hep-ph/0204199](https://arxiv.org/abs/hep-ph/0204199)
4. P.G. de Gennes, J. Prost, *The Physics of Liquid Crystals* (Clarendon Press, Oxford, UK, 1993)
5. L. Radzihovsky, T.C. Lubensky, Phys. Rev. **E83**, 051701 (2011). DOI [10.1103/PhysRevE.83.051701](https://doi.org/10.1103/PhysRevE.83.051701)
6. P.M. Chaikin, T.C. Lubensky, *Principles of condensed matter physics* (Cambridge University Press, Cambridge, UK, 1995)
7. E. Ivanov, V.I. Ogievetsky, Theor. Math. Phys. **25**, 1050 (1975)
8. I.Z. Rothstein, P. Shrivastava, JHEP **05**, 014 (2018). DOI [10.1007/JHEP05\(2018\)014](https://doi.org/10.1007/JHEP05(2018)014)
9. L. Alberte, A. Nicolis, JHEP **07**, 076 (2020). DOI [10.1007/JHEP07\(2020\)076](https://doi.org/10.1007/JHEP07(2020)076)
10. A. Nicolis, R. Penco, F. Piazza, R. Rattazzi, JHEP **06**, 155 (2015). DOI [10.1007/JHEP06\(2015\)155](https://doi.org/10.1007/JHEP06(2015)155)
11. B. Finelli, JHEP **03**, 075 (2020). DOI [10.1007/JHEP03\(2020\)075](https://doi.org/10.1007/JHEP03(2020)075)
12. I. Kharuk, A. Shkerin, Phys. Rev. **D98**, 125016 (2018). DOI [10.1103/PhysRevD.98.125016](https://doi.org/10.1103/PhysRevD.98.125016)
13. T. Hayata, Y. Hidaka, Phys. Rev. **D91**, 056006 (2015). DOI [10.1103/PhysRevD.91.056006](https://doi.org/10.1103/PhysRevD.91.056006)
14. T. Brauner, H. Watanabe, Phys. Rev. **D89**, 085004 (2014). DOI [10.1103/PhysRevD.89.085004](https://doi.org/10.1103/PhysRevD.89.085004)
15. A. Nicolis, R. Penco, F. Piazza, R.A. Rosen, JHEP **11**, 055 (2013). DOI [10.1007/JHEP11\(2013\)055](https://doi.org/10.1007/JHEP11(2013)055)
16. Y. Hidaka, T. Noumi, G. Shiu, Phys. Rev. **D92**, 045020 (2015). DOI [10.1103/PhysRevD.92.045020](https://doi.org/10.1103/PhysRevD.92.045020)
17. S. Endlich, A. Nicolis, R. Penco, Phys. Rev. **D89**, 065006 (2014). DOI [10.1103/PhysRevD.89.065006](https://doi.org/10.1103/PhysRevD.89.065006)
18. R. Klein, D. Roest, D. Stefanyszyn, JHEP **10**, 051 (2017). DOI [10.1007/JHEP10\(2017\)051](https://doi.org/10.1007/JHEP10(2017)051)
19. I.N. McArthur, JHEP **11**, 140 (2010). DOI [10.1007/JHEP11\(2010\)140](https://doi.org/10.1007/JHEP11(2010)140)
20. N. Manton, P. Sutcliffe, *Topological Solitons* (Cambridge University Press, Cambridge, UK, 2004)

21. K.T. Hecht, *Quantum Mechanics* (Springer-Verlag, New York, US, 2000)
22. C. Cheung, P. Creminelli, A.L. Fitzpatrick, J. Kaplan, L. Senatore, JHEP **03**, 014 (2008). DOI [10.1088/1126-6708/2008/03/014](https://doi.org/10.1088/1126-6708/2008/03/014)
23. K. Samokhin, Phys. Rev. **B81**, 224507 (2010). DOI [10.1103/PhysRevB.81.224507](https://doi.org/10.1103/PhysRevB.81.224507)
24. P.D. Powell, [arXiv:1112.4379](https://arxiv.org/abs/1112.4379)
25. H. Watanabe, H. Murayama, Phys. Rev. **D89**, 101701 (2014). DOI [10.1103/PhysRevD.89.101701](https://doi.org/10.1103/PhysRevD.89.101701)
26. E. Bergshoeff, J. Figueroa-O'Farrill, J. Gomis, SciPost Phys. Lect. Notes **69** (2023). DOI [10.21468/SciPostPhysLectNotes.69](https://doi.org/10.21468/SciPostPhysLectNotes.69)

Chapter 14

Broken Spacetime Symmetry in Classical Matter

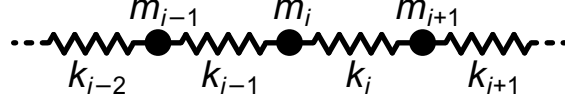
The first example of a *Nambu–Goldstone* (NG) *boson* I mentioned in the introduction (Chap. 1) was hydrodynamic sound. However, our subsequent exploration of *spontaneous symmetry breaking* (SSB) took us very far away from this initial target. Indeed, I focused almost exclusively on the quantum world. This had a good reason: many conceptual subtleties of SSB are rooted in the description of quantum symmetries in terms of operators on the Hilbert space of states. We then did much work to establish the methods of nonlinear realization of symmetry and *effective field theory* (EFT) for NG bosons. Eventually, we discovered that the *leading order* (LO) of the derivative expansion of the EFT for quantum systems with SSB is a classical field theory. In this chapter, we shall close the circle and see how the techniques we have developed can be applied to purely classical systems.

To build the necessary intuition, I will start in Sect. 14.1 with a primitive toy model for elastic solids. While physically inadequate in the details, this is good enough to show that a classical medium may possess an emergent symmetry, reflecting its internal structure. A proper identification of such emergent symmetries is key for distinguishing thermodynamic phases of matter with otherwise identical microscopic dynamics, such as solids and fluids. The subsequent analysis is straightforward and follows the standard workflow from nonlinear realization to effective actions. The nonlinear realization of a purely spacetime symmetry augmented with the emergent symmetry of classical matter is detailed in Sect. 14.2. In the following Sect. 14.3, I then work out explicit examples of EFTs for different phases of matter.

14.1 Emergent Symmetry of Classical Matter

Classical matter is characterized by the possibility to uniquely label its elements and individually track their evolution. While doing so, one has to carefully distinguish two different types of coordinates. The first type are the genuine spacetime coordinates, independent of whatever matter is present in the spacetime. The second type are the coordinates that label the individual elements of the medium. These are

Fig. 14.1 One-dimensional chain of point masses, connected by linear springs. The springs are only allowed to vibrate in the horizontal direction. The masses and spring constants may vary along the chain as indicated



known in classical mechanics as *body coordinates* (see e.g. Sect. 8.1 of [1]) or *material coordinates* (Chap. 4 of [2]). They capture internal variations of the material properties of the medium. The difference between the two types of coordinates is best elucidated by an example.

14.1.1 Introduction: Spring Model of Elasticity

One of the most basic mechanical models, illustrating a transition between mechanics of particles and continuous field theory, is the linear spring chain, shown in Fig. 14.1. For the sake of discussion, I will assume that the particles connected by the springs have generally unequal masses. Likewise, the springs themselves may have different spring constants, although I will for simplicity assume equal rest (unloaded) length a . Denoting the position of the i -th particle as x_i , the Lagrangian of the system is

$$L = \sum_i \left[\frac{1}{2} m_i \dot{x}_i^2 - \frac{1}{2} k_i (x_{i+1} - x_i - a)^2 \right]. \quad (14.1)$$

The ground state of this system corresponds to a chain of point masses at rest, placed equidistantly at distance a between the nearest neighbors. One can choose the origin of coordinates so that in the ground state, $\langle x_i \rangle = ia$. We would now like to perform the continuum limit, assuming that a is very small. This requires that the displacement $x_i - ia$ from the equilibrium position only varies appreciably over distances much longer than a . I will replace the discrete index i with a continuous body coordinate X through $i \rightarrow X/a$. The individual masses m_i are replaced with the linear mass density $\varrho(X)$ via $m_i \rightarrow a\varrho(X)$, and the spring constants k_i with the local elastic (Young) modulus $E(X)$ via $k_i \rightarrow E(X)/a$. Treating the Lagrangian (14.1) as a Riemann sum then leads to the continuous approximation

$$L \rightarrow \int dX \left\{ \frac{1}{2} \varrho(X) [\partial_0 x(X, t)]^2 - \frac{1}{2} E(X) [\partial_X x(X, t) - 1]^2 \right\}. \quad (14.2)$$

This is a field theory in one spatial dimension with the dynamical degree of freedom $x(X, t)$. Provided both $\varrho(X)$ and $E(X)$ are positive for any $X \in \mathbb{R}$, which follows from their origin in the parameters m_i and k_i , the ground state is $\langle x(X, t) \rangle = X$ up to an additive constant. This residual freedom stems from spontaneous breakdown

of spatial translations. Parameterizing the displacement of the medium from equilibrium by the field $\phi(X, t) \equiv x(X, t) - X$, the fluctuations around the ground state are governed by the *equation of motion* (EoM),

$$\varrho(X) \partial_0^2 \phi(X, t) = \partial_X [E(X) \partial_X \phi(X, t)] . \quad (14.3)$$

In the limit of equal masses m_i and spring constants k_i , the functions $\varrho(X)$ and $E(X)$ become constant. The EoM (14.3) then describes compression waves, propagating along the chain with phase velocity $v = \sqrt{E/\varrho}$. In this limit, the action for $\phi(X, t)$ features a set of emergent continuous symmetries that were not present in the original discrete model (14.1). The most obvious of these is the invariance under the “translation” $X \xrightarrow{\epsilon} X + \epsilon$. A straightforward application of Noether’s theorem shows that the corresponding integral charge is, up to overall normalization,

$$P_{\text{pseudo}} = - \int dX \varrho \partial_0 \phi(X, t) \partial_X \phi(X, t) . \quad (14.4)$$

Despite the suggestive analogy with translation invariance, this is *not* the momentum carried by the oscillating masses of the original discrete chain. The latter rather equals

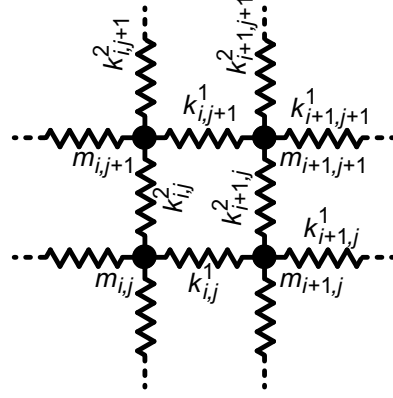
$$P = \sum_i m_i \dot{x}_i \rightarrow \int dX \varrho(X) \partial_0 \phi(X, t) . \quad (14.5)$$

The resolution of this puzzle is that in presence of a uniform classical medium, we have two different coordinates and consequently two different translation symmetries. The position of the mass m_i with respect to the laboratory inertial reference frame is given by the coordinate x_i . Accordingly, a genuine spatial translation amounts to the shift $x \xrightarrow{\epsilon} x + \epsilon$. In the formulation (14.2) of our field theory, this acts like an “internal symmetry.” That is, it transforms the dependent variable $x(X, t)$ while leaving the independent variable X intact. Applying Noether’s theorem recovers the integral momentum (14.5). This remains conserved for any choice of functions $\varrho(X)$ and $E(X)$. That is because momentum conservation reflects the uniformity of the underlying space itself; it does not care about the properties of the medium. The corresponding local conservation law is equivalent to (14.3), which is the continuous limit of the Newtonian EoM for the point masses m_i .

The other integral charge, (14.4), is usually called *pseudomomentum*. This arises, as already stressed, from invariance under $X \xrightarrow{\epsilon} X + \epsilon$. Given that X was introduced as a material coordinate, conservation of P_{pseudo} should reflect the uniformity of the medium. Sure enough, it was essential for derivation of (14.4) to assume that $\varrho(X)$ and $E(X)$ are constant. The distinction between momentum and pseudomomentum has historically led to much confusion in the research on the continuum mechanics of fluids. See [3] for a pedagogical account of some of the related subtleties.

The comparison of momentum and pseudomomentum illustrates best the striking contrast between the fundamental spacetime symmetries and the emergent symmetries of matter. However, the field theory (14.2) also features other

Fig. 14.2 Rectangular network of springs as a toy model of a two-dimensional elastic solid. All the point masses $m_{i,j}$ are allowed to differ, as are the spring constants $k_{i,j}^1$ and $k_{i,j}^2$. The rest (unloaded) length of all the springs is assumed to be the same and equal to a



emergent symmetries than translations. Namely, for constant $\varrho(X)$ and $E(X)$ its action is invariant under the whole Poincaré group $\text{ISO}(1, 1)$ of transformations acting on X, t . These include the “translations” of X , giving rise to conservation of pseudomomentum, and the usual time translations, leading to conservation of energy. In addition, the theory has an emergent $\text{SO}(1, 1)$ Lorentz invariance under “boosts” that mix X and t . It is straightforward to work out the corresponding conservation law but I will not do so since we will not need it. Finally, the dynamics of our nonrelativistic spring chain should be invariant under Galilei boosts. Under the corresponding transformations, $x \xrightarrow{v} x + vt$ with v being the velocity of the boost, the Lagrangian (14.2) shifts by a total time derivative. Similarly to spatial translations, this is a genuine spacetime symmetry that is present for any choice of $\varrho(X)$ and $E(X)$.

Before closing the discussion of our toy model, let us briefly look at its generalization to a higher number of spatial dimensions d . For illustration, it is sufficient to take $d = 2$ and consider a rectangular network of springs as in Fig. 14.2. The position of the particle $m_{i,j}$ is now given by a two-component vector $\mathbf{x}_{i,j}$. Assuming otherwise the same dynamics as before, whereby the potential energy of each spring is a quadratic function of its extension, the Lagrangian becomes

$$L = \sum_{i,j} \left[\frac{1}{2} m_{i,j} \dot{\mathbf{x}}_{i,j}^2 - \frac{1}{2} k_{i,j}^1 (|\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}| - a)^2 - \frac{1}{2} k_{i,j}^2 (|\mathbf{x}_{i,j+1} - \mathbf{x}_{i,j}| - a)^2 \right]. \quad (14.6)$$

The ground state of this model corresponds to a rectangular network of equidistantly placed particles at rest. One can choose Cartesian coordinates in the plane so that $\langle \mathbf{x}_{i,j} \rangle = (ia, ja)$. To perform the continuum limit, we make the replacement $(i, j) \rightarrow X/a$ and $m_{i,j} \rightarrow a^2 \varrho(X)$, where X is a two-component vector of body coordinates and $\varrho(X)$ the local mass density. We also need two elastic moduli, $k_{i,j}^1 \rightarrow E_1(X)$

and $k_{i,j}^2 \rightarrow E_2(\mathbf{X})$. In terms of the displacement vector $\phi^r(\mathbf{X}, t) \equiv x^r(\mathbf{X}, t) - X^r$,¹ the Lagrangian of the resulting continuous two-dimensional theory reads

$$L \rightarrow \int d^2\mathbf{X} \left\{ \frac{1}{2} \varrho(\mathbf{X}) [\partial_0 \phi(\mathbf{X}, t)]^2 - \frac{1}{2} E_1(\mathbf{X}) [\partial_1 \phi^1(\mathbf{X}, t)]^2 - \frac{1}{2} E_2(\mathbf{X}) [\partial_2 \phi^2(\mathbf{X}, t)]^2 \right\} + \dots \quad (14.7)$$

The ellipsis stands for terms of higher order in ϕ^r ; in $d \geq 2$ dimensions, the dynamics of the spring network does not resolve into a superposition of one-dimensional harmonic motions. Moreover, the model (14.7) does not correctly capture the physics of elastic solids even in the harmonic approximation. Namely, for purely transverse oscillations such that $\partial \phi^r / \partial X^r = 0$ for each fixed r , there is no linear restoring force and the motion is strongly anharmonic. In spite of these flaws, the model (14.7) is sufficient to shed light on the fate of the emergent symmetries.

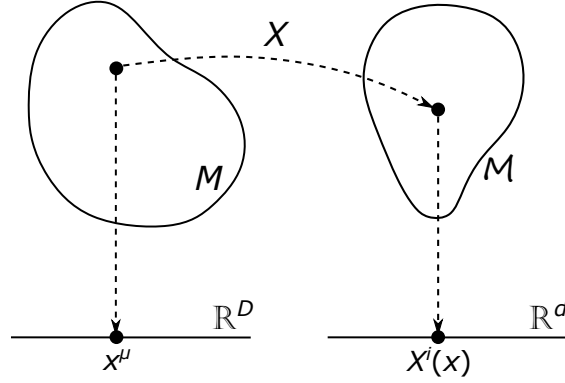
We always find exact invariance under genuine spacetime translations, spatial rotations and Galilei boosts regardless of the choice of the functions $\varrho(\mathbf{X})$ and $E_{1,2}(\mathbf{X})$. This much is obvious from the discrete version (14.6) of the model. These symmetries reflect the properties of spacetime itself and are insensitive to the material structure of the medium. Should the medium be uniform, with constant $\varrho(\mathbf{X})$ and $E_{1,2}(\mathbf{X})$, we will in addition have invariance under continuous “internal translations” of \mathbf{X} . This symmetry ensures conservation of pseudomomentum, and we can expect it to arise in the long-distance description of real crystalline materials. However, we cannot in general expect invariance under continuous “internal rotations” of \mathbf{X} , unless we demand $E_1 = E_2$ and restrict the Lagrangian (14.7) to the lowest order in the power expansion in ϕ^r . This suggests that the macroscopic properties of real crystals tend to be homogeneous but anisotropic even in the continuum limit $a \rightarrow 0$, which agrees with empirical observations. The dependence of the Lagrangian on X^r is then constrained by the continuous internal translation invariance and the discrete group of point symmetries of the crystal lattice.

14.1.2 Emergent Symmetries of Solids and Fluids

We shall now synthesize the above observations into a general setup for describing classical matter. Much of the discussion in this subsection is inspired by Chap. 4 of [2]. The basic assumption is that we are dealing with a thermodynamic state where quantum correlations are limited to short distances, typically due to thermal fluctuations. It is then possible to identify elements of the medium that are mu-

¹ I previously introduced the symbol $X^\mu[\psi, x]$ to indicate the shift of the spacetime coordinate x^μ under an infinitesimal symmetry transformation. In this chapter, I will not use this notation in order to avoid confusion with the conventional symbol X^r for body coordinates. The same applies to the symbol $F^i[\psi, x]$, previously introduced to denote an infinitesimal transformation of a field ψ^i . In this chapter, I will reserve the letter F for the function defining the LO effective Lagrangian.

Fig. 14.3 The time evolution of the classical medium as a map from the spacetime M to the target space \mathcal{M} of body coordinates. The spacetime coordinates x^μ in M and the body coordinates X^i in \mathcal{M} can be chosen arbitrarily and independently of each other



tually distinguishable and can be assigned unique labels. This requirement limits the validity of the EFTs developed below to distances much longer than the scales characterizing quantum correlations and the discrete structure of matter.

The labels X^i on the medium elements take values from a target space \mathcal{M} , which is typically some domain in the real space, \mathbb{R}^d . The domain would be finite for an isolated material object of a finite size. However, here we will mostly be concerned with properties of matter filling the entire space, where it is natural to take $\mathcal{M} \simeq \mathbb{R}^d$. In both cases, the labels X^i represent comoving body coordinates, attached to a fixed medium element. The trajectory of the element is defined by giving its position as a function of time, $\mathbf{x}(X, t)$. This is the standard Lagrangian picture of continuum mechanics, which I used to formulate the toy models (14.2) and (14.7). In order to highlight the genuine spacetime symmetries, it is however more convenient to invert the relation between x^r and X^i . The time evolution of the medium is then specified by a map from the spacetime M to \mathcal{M} that assigns to every point $x^\mu \in M$ the material coordinates $X^i(\mathbf{x}, t) \equiv X^i(x)$ of the element occupying this point. The uniqueness of the labeling of the medium elements is ensured by requiring that for any fixed time t , the map $x^r \rightarrow X^i(\mathbf{x}, t)$ is a diffeomorphism between space and \mathcal{M} . The advantage of this picture is that the labels X^i take the same values in any reference frame. In other words, the functions $X^i(x)$ are scalar fields with respect to any spacetime symmetry. On the other hand, emergent symmetries acting solely on X^i can then be treated as internal symmetries, that is point transformations on \mathcal{M} .

The explicit realization of the fields $X^i(x)$ depends on a choice of coordinates. Importantly, the coordinates x^μ in the spacetime and the material coordinates X^i can be fixed independently of each other. See Fig. 14.3 for a visualization of this freedom. In the following, I will deal exclusively with matter in flat spacetimes. Accordingly, I will use standard Minkowski coordinates x^μ for relativistic systems, and Cartesian coordinates x^r augmented with Newtonian time t for nonrelativistic systems. This still leaves us with the freedom to choose the body coordinates X^i at will, which should be distinguished from any internal symmetry acting on \mathcal{M} . In Sect. 14.1.1, I fixed this freedom tacitly

by the definition of X^r in terms of the discrete labels i, j . In order to avoid misunderstanding and to highlight the scalar nature of the body coordinates, I have now changed the notation from X^r to X^i .

The scalar fields $X^i(\mathbf{x}, t)$ are the generalized coordinates of our continuum field-theoretic description of classical matter. The corresponding generalized velocity is $\partial_0 X^i(\mathbf{x}, t)$. How is this related to the actual local (Eulerian) velocity of the medium with respect to the laboratory frame? Recall that X^i are comoving labels that cannot change along the trajectory of the medium element. Parameterizing the trajectory as $\mathbf{x}(t)$, the condition $dX^i(\mathbf{x}(t), t)/dt = 0$ leads to $\partial_0 X^i = -\dot{\mathbf{x}}^r \partial_r X^i$, where $\dot{\mathbf{x}} \equiv d\mathbf{x}/dt$ is the desired local velocity. The matrix $M^i_r(\mathbf{x}, t) \equiv \partial_r X^i(\mathbf{x}, t)$ is invertible thanks to the assumption that the map $x^r \rightarrow X^i(\mathbf{x}, t)$ is a diffeomorphism. It follows that

$$\dot{x}^r = -(M^{-1})^r_i \partial_0 X^i. \quad (14.8)$$

There is also another, more elegant albeit slightly more mathematically advanced, way to express the local medium velocity in terms of derivatives of $X^i(\mathbf{x})$. First pick any function $f(X)$ on \mathcal{M} and define the current

$$J_f^\mu \equiv \frac{f(X)}{d!} \varepsilon^{\mu\nu_1 \dots \nu_d} \varepsilon_{i_1 \dots i_d} \partial_{\nu_1} X^{i_1} \dots \partial_{\nu_d} X^{i_d}. \quad (14.9)$$

By the antisymmetry of the *Levi-Civita* (LC) symbol, $J_f^\mu \partial_\mu X^i = 0$ for any i . Expanding this as $J_f^0 \partial_0 X^i + J_f^r M^i_r = 0$ gives $J_f^r = -J_f^0 (M^{-1})^r_i \partial_0 X^i$. Using (14.8), we can thus rewrite the current as

$$J_f^\mu = (J_f^0, J_f^0 \dot{\mathbf{x}}). \quad (14.10)$$

It also follows from the antisymmetry of the LC symbol that the current (14.9) is conserved, $\partial_\mu J_f^\mu = 0$, for any choice of the function $f(X)$. It is therefore natural to identify $J_f^0 = f(X) \det M$ with the density of a conserved charge as observed in the laboratory frame. The spatial part of the current, $J_f^r = J_f^0 \dot{x}^r$, then represents the flow of this charge. Finally, $f(X)$ itself is the density of the same charge in the comoving material coordinates, since $\det M$ is just the Jacobian of the transformation between the spacetime and material coordinate systems.

The origin of the currents J_f^μ can be traced back to the fact that the individual elements of our classical medium are distinguishable. Indeed, we could in principle attach to each element a dedicated “charge.” All such charges would be conserved by construction. The freedom to choose the function $f(X)$ in (14.9) at will is just a continuous version of this observation. Operationally, the existence of infinitely many conservation laws, one for each $f(X)$, is not a problem. Namely, J_f^μ are not Noether currents in that their conservation does not require the EoM. The presence of such identically conserved currents does not constrain the local classical dynamics of the medium.

The above said does not mean that the currents J_f^μ are inevitably artifacts of the Lagrangian picture of continuum mechanics that are not macroscopically observable. Suppose our medium carries a conserved charge. This could count for instance the number of particles or, in the nonrelativistic limit, their mass. By choosing $f(X)$ as the density of the charge in the material coordinates, we get a current J_f^μ that describes macroscopic flow of this charge.

Mathematically, the currents (14.9) descend from the d -forms $f(X)dX^1 \wedge \cdots \wedge dX^d$ on \mathcal{M} . Being top-dimensional, any such d -form is automatically closed. The closedness is preserved when the d -form is pulled back to the spacetime by the map $x^\mu \rightarrow X^i(x)$. Taking the Hodge dual of the ensuing closed spacetime d -form then gives the current J_f^μ with vanishing divergence.

Having established the basic setup, we are now finally in a position to discuss the emergent symmetries of various types of classical matter. I will focus on the two most common classical phases of matter: crystalline solids and fluids. Since this will to a certain extent merely collect and organize some observations I have already made previously, I can afford to be brief.

Solids We will be interested in the quasiequilibrium dynamics of crystalline solids at distances much longer than the lattice spacing of the microscopic crystal structure. In this regime, one can expect the solid to be internally uniform, and thus be symmetric under the group \mathbb{R}^d of continuous internal translations. These translations take a particularly simple form in “Cartesian coordinates” on \mathcal{M} , namely $X^i \xrightarrow{\epsilon} X^i + \epsilon^i$. Barring possible quasi-invariant contributions, the effective Lagrangian can then only depend on $X^i(x)$ through their derivatives. In addition, the dependence on $X^i(x)$ is constrained by the point symmetry of the crystal lattice. This can be imposed order by order in powers of $X^i(x)$ using tensor methods. I will not go into detail here, and rather refer the reader to Sect. 10 of [4] for further discussion. For the sake of illustration, I will only deal with the special case of *isotropic* solids, which feature a full $\text{SO}(d)$ symmetry under continuous rotations of $X^i(x)$. Exact, full isotropy cannot really be achieved for any crystalline material. However, some polycrystalline materials are effectively isotropic at distances much longer than the typical size of a grain, thanks to the random orientation of the grains.

Fluids The key difference between solids and fluids is that in the latter, there are no restoring elastic forces that would counteract shear strain. This enables macroscopic flow, whose physics is largely affected by dissipation. Unfortunately, dissipative effects are notoriously difficult to include in Lagrangian field theory. In order that the EFT for fluids we construct be meaningful, we thus need to make some simplifying assumptions. Namely, we shall restrict ourselves to physical processes that do not involve entropy production. Also, we will only consider states of the fluid that are perturbations of a uniform equilibrium where the entropy density in the material coordinates is constant. Such fluids are called *barotropic*; their thermodynamic state can be described by a single variable such as pressure or density. Thus, our EFT

will be able to capture processes such as sound propagation, but not, for instance, convective heat transfer due to an initial temperature gradient.

By definition, the local density of energy and entropy, and thus also of any other extensive property, of a barotropic fluid corresponds to a constant function f on \mathcal{M} . At the same time, the macrostate of the fluid is completely specified by its local Eulerian velocity and density J_f^0 observed in the laboratory frame. The same should be true for the Lagrangian, which is directly connected to observables such as energy density and pressure. The field-theoretic description of fluids in terms of the scalar fields $X^i(x)$ should therefore be invariant under any point transformation on \mathcal{M} that preserves the currents (14.9). This is not mere freedom of choice of coordinates on \mathcal{M} . It corresponds to an actual physical reshuffling of the elements of the fluid that does not affect its macroscopic properties. Now $J_f^0 = f \det M$ requires that the point transformation on \mathcal{M} preserves $\det M$, and so must have a unit Jacobian. We conclude that the dynamics of a barotropic fluid possesses an emergent symmetry under the group of *volume-preserving diffeomorphisms* (VPDs) on \mathcal{M} , $\text{SDiff}(\mathcal{M})$. This is an infinite-dimensional symmetry group that includes the internal translations and rotations of solids as a subgroup. See [5] for a detailed account of the role of the diffeomorphism group in hydrodynamics, and [6, 7] for two somewhat complementary overviews of the variational approach to fluids.

14.2 Nonlinear Realization of Emergent Symmetry

We have done a serious amount of work to carefully identify the emergent symmetries of classical matter. Our effort will now pay dividends in that the next steps will be fairly straightforward. In the absence of degrees of freedom sensitive to other symmetries, the symmetry group of a classical medium is $G \simeq G_{\text{s.t.}} \times G_{\text{int}}$. Here $G_{\text{s.t.}}$ is whatever spacetime symmetry is appropriate for the system at hand, such as Aristotelian, Galilei or Poincaré. The group G_{int} collects all the emergent symmetries that act on the scalar fields $X^i(x)$ as internal. For the reader's convenience, the choices of G_{int} in case of uniform solids and barotropic fluids are summarized in the first two columns of Table 14.1. The product structure, $G \simeq G_{\text{s.t.}} \times G_{\text{int}}$, of the symmetry group already appeared in Sect. 13.2, and we will be able to largely follow the path paved therein.

14.2.1 Field Variables and Unbroken Symmetry

To start with, the isotropy group of the spacetime origin is $H_0 \simeq H_{\text{s.t.}} \times G_{\text{int}}$, where $H_{\text{s.t.}}$ collects all transformations from $G_{\text{s.t.}}$ that fix the origin (typically rotations and boosts). I will always assume that G_{int} includes a set of mutually commuting translations that act transitively on $\mathcal{M} \simeq \mathbb{R}^d$. The isotropy subgroup $H_{(X_0,0)}$ is then the same, up to conjugation by an element of G_{int} , for any choice of the reference

Table 14.1 Overview of the emergent symmetries and related subgroups in different phases of classical matter. The G_{cryst} symbol denotes the crystallographic point group of a crystal lattice. Also, $\text{SDiff}_0(\mathbb{R}^d)$ is the group of VPDs in \mathbb{R}^d that fix the origin. The coset space $G_{\text{int}}/H_{\text{int}}$ is equivalent to \mathbb{R}^d in all cases. The unbroken subgroup H_φ indicates the symmetry of a uniform, static equilibrium state. The subscript “diag” denotes diagonal symmetries whose actions on the spatial and body coordinates are locked to each other. The values shown in the last column assume that the subgroup of time-independent transformations in $G_{\text{s.t.}}$, acting only on the spatial coordinates, is the Euclidean group $\text{ISO}(d)$. The factor of \mathbb{R} in H_φ corresponds to time translations

Material	Emergent symmetry G_{int}	Isotropy group H_{int}	Unbroken subgroup H_φ
Solid	$G_{\text{cryst}} \ltimes \mathbb{R}^d$	G_{cryst}	$(G_{\text{cryst}} \ltimes \mathbb{R}^d)_{\text{diag}} \times \mathbb{R}$
Isotropic solid	$\text{ISO}(d)$	$\text{SO}(d)$	$[\text{ISO}(d)]_{\text{diag}} \times \mathbb{R}$
Fluid	$\text{SDiff}(\mathbb{R}^d)$	$\text{SDiff}_0(\mathbb{R}^d)$	$[\text{ISO}(d)]_{\text{diag}} \times \mathbb{R}$

point X_0^i . It is convenient to set $X_0^i = 0$ so that $H_{(X_0,0)} \simeq H_{\text{s.t.}} \times H_{\text{int}}$ where H_{int} consists of all emergent symmetries that fix the origin in \mathcal{M} . The concrete isotropy subgroups for solids and fluids are listed in the third column of Table 14.1. Note that in case of fluids, both $G_{\text{int}} \simeq \text{SDiff}(\mathbb{R}^d)$ and $H_{\text{int}} \simeq \text{SDiff}_0(\mathbb{R}^d)$ are infinite-dimensional. However, the coset space $G_{\text{int}}/H_{\text{int}}$ is equivalent to \mathbb{R}^d for any choice of G_{int} thanks to the assumed transitive action of G_{int} on \mathcal{M} . This reconfirms that the dynamical degrees of freedom of the EFT for classical matter will always be the d NG fields $X^i(x)$. That is in contrast to the *agnostic nonlinear realization* (cf. Sects. 12.3.2 and 13.3.4), whose application to fluids would suggest an EFT with an infinite number of would-be NG fields; see [8] for details.

Before we proceed to the construction of effective actions, let us check which of the symmetries in G actually are spontaneously broken. Suppose we are interested in the physics of fluctuations around a static equilibrium state, characterized by the time-independent *vacuum expectation values* (VEVs)²

$$\langle X^i(x, t) \rangle \equiv \varphi^i(x) . \quad (14.11)$$

Due to the product structure of G , an element $(g_{\text{s.t.}}, g_{\text{int}}) \in G$ transforms the space-time and body coordinates as $T_g : (X^i, x^\mu) \rightarrow (X'^i, x'^\mu) \equiv (\mathbb{F}^i(X, g_{\text{int}}), \mathbb{X}^\mu(x, g_{\text{s.t.}}))$. Here the x^μ -independent functions \mathbb{F}^i realize the action of G_{int} on \mathcal{M} , whereas the X^i -independent functions \mathbb{X}^μ realize the action of $G_{\text{s.t.}}$ on the spacetime. The unbroken subgroup H_φ consists of transformations preserving (14.11),

$$\mathbb{F}^i(\varphi(x), h_{\text{int}}) = \varphi^i(\mathbb{X}(x, h_{\text{s.t.}})) , \quad (h_{\text{s.t.}}, h_{\text{int}}) \in H_\varphi . \quad (14.12)$$

This condition has the geometric meaning of *equivariance* of φ^i as a map $M \rightarrow \mathcal{M}$ under the action of H_φ , that is $\mathbb{F} \circ \varphi = \varphi \circ \mathbb{X}$. Since $\varphi^i(x)$ is a diffeomorphism between the coordinate space and \mathcal{M} , (14.12) defines a one-to-one correspondence between

² A different choice of the reference state of our classical medium might require a modification of the discussion below. The change would however not affect the general nonlinear realization setup, which only depends on the groups $G_{\text{s.t.}}$, $H_{\text{s.t.}}$, G_{int} and H_{int} .

the maps \mathbb{F}^i and \mathbb{X}^r . In this sense, H_φ is the “diagonal subgroup” of $G_{\text{s.t.}} \times G_{\text{int.}}$. It also includes those transformations from $G_{\text{s.t.}}$ that act on x^r trivially, that is time translations. An overview of the unbroken subgroups in solids and fluids is given in the last column of Table 14.1.

14.2.2 Building Blocks for Construction of Effective Actions

In the absence of other degrees of freedom, we are dealing with the d scalar fields $X^i(x)$.³ All of these are of the NG type, realizing nonlinearly the emergent internal translations on $\mathcal{M} \simeq \mathbb{R}^d$. Let us denote the generators of the translations as Π_i . I will now finally fix the freedom to choose coordinates on $\mathcal{M} \simeq G_{\text{int.}}/H_{\text{int.}} \simeq \mathbb{R}^d$ at will by parameterizing it as $U(X) \equiv \exp(iX^i \Pi_i)$. This gives a precise definition of the previously mentioned “Cartesian coordinates” on \mathcal{M} , in which the translations from $G_{\text{int.}}$ act on X^i by trivial shifts, $X^i \xrightarrow{\epsilon} X^i + \epsilon^i$. According to (13.24), the *Maurer–Cartan* (MC) form is then given by

$$\omega(X, x) = \Pi_i dX^i + P \cdot dx, \quad (14.13)$$

where P_μ generates spacetime translations. The spacetime coframe is trivial, that is, $e^{*\mu} = \delta^\mu_\nu dx^\nu$. Accordingly, it is not necessary to use different notations for frame and coordinate-basis indices. The covariant derivatives of the NG fields are simply $\nabla_\mu X^i = \partial_\mu X^i$. This makes it possible to use the same power counting as for superfluids (Sect. 13.2.2), whereby an n -th partial derivative of X^i is assigned the counting degree $n - 1$. The LO of the EFT is thus defined by a Lagrangian density where every field X^i carries exactly one derivative.

Invariance of the effective action is ensured as follows. First, invariance under the whole internal symmetry $G_{\text{int.}}$ is guaranteed by using the MC form as a building block and imposing solely the linearly realized isotropy group $H_{\text{int.}}$. Similarly, one has to impose explicitly invariance under the linearly realized spacetime isotropy group, $H_{\text{s.t.}}$. Finally, invariance under spacetime translations requires that the Lagrangian density does not depend explicitly on the spacetime coordinates.

In case of fluids, the coset space $G_{\text{int.}}/H_{\text{int.}} \simeq \text{SDiff}(\mathbb{R}^d)/\text{SDiff}_0(\mathbb{R}^d)$ is not reductive, hence the line of reasoning using the transformation properties (13.10) of the MC form does not necessarily apply. In such a situation, we may have to impose by hand invariance under the entire emergent symmetry group $G_{\text{int.}}$. Luckily, we know a priori how VPDs act on the coordinates X^i and thus also on the 1-forms dX^i on \mathcal{M} . This will make it possible for us to construct an EFT for fluids in Sect. 14.3.4.

³ For relativistic systems, there is an alternative formulation of the EFT where the X^i are completed to a Lorentz vector of dynamical variables. See Sect. 6 of [9] for an introduction and references.

14.3 Effective Field Theory of Classical Matter

I will now carry out the above-outlined program for several types of physical systems of interest. I will start with what in many ways is the simplest case: an isotropic relativistic solid. Having warmed up, we shall then have a look at the phenomenologically more relevant nonrelativistic solids. At the very end, we will return to the most nontrivial case of fluids with their infinite-dimensional symmetry group.

14.3.1 Relativistic Solids

Consider a system with relativistic microscopic dynamics that settles to a thermodynamic equilibrium state with the symmetries of an isotropic solid. The full symmetry is $G_{\text{s.t.}} \times G_{\text{int}} \simeq \text{ISO}(d, 1) \times \text{ISO}(d)$, where the first factor is the spacetime Poincaré group and the second one is the emergent internal symmetry. With one derivative per field, the only way to obey the linearly realized Lorentz symmetry, $H_{\text{s.t.}} \simeq \text{SO}(d, 1)$, is to pairwise contract indices on $\partial_\mu X^i$. It follows that the LO effective Lagrangian is some function of the Lorentz-invariant matrix operator

$$\Xi^{ij} \equiv \partial_\mu X^i \partial^\mu X^j . \quad (14.14)$$

It remains to impose invariance under $H_{\text{int}} \simeq \text{SO}(d)$. Since Ξ^{ij} is a symmetric tensor, any H_{int} -invariant function of Ξ^{ij} can only depend on its d real eigenvalues. These can be encoded in any set of d algebraically independent invariants constructed out of Ξ^{ij} , for instance the traces of the first d powers of Ξ^{ij} . The LO effective action for relativistic isotropic solids then reads

$$S_{\text{eff}}\{X\} = \int d^D x F(\text{tr } \Xi(x), \text{tr } \Xi(x)^2, \dots, \text{tr } \Xi(x)^d) + \dots , \quad (14.15)$$

where F is a function of the displayed arguments and the ellipsis indicates corrections of higher order in the derivative expansion. These include operators with on average more than one derivative per X^i , and Lorentz and internal indices contracted in a way respecting the $H_{\text{s.t.}} \times H_{\text{int}} \simeq \text{SO}(d, 1) \times \text{SO}(d)$ symmetry.

With all the background we had built up, the path to (14.15) was very short. There are however a couple of potential loopholes to close. First, I tacitly assumed the effective Lagrangian to be strictly invariant under all the symmetries. Can there be any quasi-invariant contributions to the Lagrangian? Here I refer the reader to [10], which showed that the symmetries of a relativistic isotropic solid do not allow any genuinely quasi-invariant Lagrangians in $d = 2$ or 3 spatial dimensions. It is plausible to assume that the same conclusion holds for any $d \geq 2$.

Second, we still need to check whether the tentative equilibrium (14.11) is a stable state of the EFT (14.15). To start with, note that the LO EoM of the EFT is

$$\partial_\mu \left(\frac{\partial F}{\partial \Xi^{ij}} \partial^\mu X^j \right) = 0 . \quad (14.16)$$

In a uniform equilibrium, the VEV of the invariants (14.14), $\langle \Xi^{ij} \rangle = -\nabla \varphi^i \cdot \nabla \varphi^j$, should be coordinate-independent. The EoM then implies that the functions $\varphi^i(\mathbf{x})$ must be harmonic, $\nabla^2 \varphi^i = 0$. Acting with the Laplace operator on $\langle \Xi^{ij} \rangle$, we get in turn that $(\partial_r \nabla \varphi^i) \cdot (\partial_r \nabla \varphi^j) = 0$. Setting $i = j$ herein shows that all partial derivatives of $\nabla \varphi^i$ vanish. Hence, $\varphi^i(\mathbf{x})$ is a linear function of coordinates, $\varphi^i(\mathbf{x}) = M^i_r x^r + c^i$ with a constant invertible matrix M^i_r and constant c^i . The set of constants c^i can be removed by shifting the origin of coordinates. Finally, suppose that the generators Π_i of internal translations were chosen as a basis of the standard vector representation of $H_{\text{int}} \simeq \text{SO}(d)$. By using the singular value decomposition augmented with appropriate orthogonal rotations of the fields X^i and coordinates x^r , the matrix M^i_r can be made diagonal and positive-semidefinite. Invariance under the unbroken diagonal $\text{SO}(d)$ rotations of spatial and body coordinates then requires that $M^i_r \propto \delta^i_r$. Finally, the proportionality factor can be absorbed into a simultaneous rescaling of the body coordinates X^i and the generators Π_i . We therefore conclude that, without loss of generality, the uniform equilibrium of the solid can be described by

$$\langle X^i(\mathbf{x}, t) \rangle = \delta^i_r x^r . \quad (14.17)$$

Below, I will implicitly assume the same ground state for both nonrelativistic solids and (relativistic or nonrelativistic) fluids.

Further constraints on the parameters of the effective Lagrangian arise from requiring that the state (14.17) is (meta)stable with respect to small fluctuations. To that end, we parameterize the fluctuations of X^i by a set of NG fields, $\pi^i(x) \equiv X^i(x) - \delta^i_r x^r$. In terms of these fields, we have

$$\Xi^{ij} = -\delta^{ij} - (\delta^{ir} \partial_r \pi^j + \delta^{jr} \partial_r \pi^i) + \partial_\mu \pi^i \partial^\mu \pi^j . \quad (14.18)$$

It is now a matter of straightforward algebra to expand the action (14.15) to second order in π^i and compute the excitation spectrum. The reader is most welcome to do this exercise, or check [11] where the special case of $d = 3$ spatial dimensions is analyzed. I will not work out the details here, since I will address the same problem in the arguably more realistic setting of nonrelativistic solids below.

14.3.2 Nonrelativistic Supersolids

Most naturally occurring solid materials lie safely within the nonrelativistic domain. It therefore appears more appropriate to demonstrate the utility of our EFT formalism in a setting that connects more directly to classical theory of elasticity [4]. However, in order that the discussion below is not a mere rehash of Sect. 14.3.1 using nonrelativistic notation, I will add a new physical ingredient. Namely, some materials are known to enter at low temperatures a quantum phase called suggestively *supersolid*.

In this phase, matter exhibits a combination of crystalline solid order and superflow. The latter stems from the presence of an internal, spontaneously broken $U(1)$ symmetry. Once we have found an effective action for supersolids, we will be able to recover an EFT for ordinary solids by decoupling the superfluid NG boson. The content of this subsection is heavily influenced by [12].

The mathematical setup closely follows the discussion of nonrelativistic superfluids in Sect. 13.3.3, to which I refer the reader for details. In order to avoid excessive cross-references, I will however spell out the main features of the setup explicitly. The full symmetry of the nonrelativistic supersolid is given by the Bargmann group augmented with the emergent $ISO(d)$ symmetry of a classical isotropic solid, that is

$$G \simeq SO(d) \ltimes \{\mathbb{R}_K^d \ltimes [\mathbb{R}^D \times U(1)_Q]\} \times ISO(d) . \quad (14.19)$$

Here the $SO(d)$ factor represents spatial rotations, \mathbb{R}_K^d Galilei boosts, \mathbb{R}^D spacetime translations, and $U(1)_Q$ the internal symmetry counting the number of particles. The symmetry group does not have the simple direct product structure $G_{s.t.} \times G_{int.}$. This is due to the dual nature of $U(1)_Q$, which is an internal symmetry but simultaneously centrally extends the spacetime Galilei group. We will however still be able to follow the philosophy of Sect. 14.2 with minor modifications. In particular the spacetime isotropy group H_0 is in this case $H_0 \simeq [SO(d) \ltimes \mathbb{R}_K^d] \times U(1)_Q \times ISO(d)$.

In order to be able to apply our basic EFT framework for spacetime symmetry, we need to realize the Galilei boosts nonlinearly. To that end, we need a Galilei vector order parameter $A^\mu = (A^0, \mathbf{A})$, choosing a timelike reference point, $A_0^\mu = (a, \mathbf{0})$ with $a \neq 0$. This is accompanied by two other order parameters, a complex scalar ψ charged under $U(1)_Q$, and the body coordinates X^i . Taking any $\psi_0 \neq 0$ and $X_0^i = 0$ leads to $H_{((\psi_0, A_0, X_0), 0)} \simeq SO(d) \times SO(d)$, where the two factors act respectively on the spatial and body coordinates. The coset space relevant for the nonlinear realization of the symmetry is $H_0/H_{((\psi_0, A_0, X_0), 0)} \simeq \mathbb{R}_K^d \times U(1)_Q \times \mathbb{R}^d$. The last factor of \mathbb{R}^d is new compared to Sect. 13.3.3 and carries the solid degrees of freedom. The whole coset space is parameterized by the NG variables π , ξ^r and X^i through

$$U(\pi, \xi, X) \equiv e^{i\pi Q} e^{i\xi \cdot K} \exp(iX^i \Pi_i) , \quad (14.20)$$

where Q is the generator of $U(1)_Q$ and K^r that of Galilei boosts. The MC form is calculated using the commutation relations of the Bargmann group,

$$\begin{aligned} \omega(\pi, \xi, X, \mathbf{x}, t) = & Q[d\pi - \xi \cdot d\mathbf{x} + (1/2)\xi^2 t] + \mathbf{K} \cdot d\xi + \Pi_i dX^i \\ & + Hdt + \mathbf{P} \cdot (d\mathbf{x} - \xi dt) . \end{aligned} \quad (14.21)$$

The second line herein defines the spacetime coframe, $e^{*0} = dt$ and $e^{*r} = \delta_S^r(dx^s - \xi^s dt)$. This allows one to extract the covariant derivatives of all the NG fields from the ω_\perp part of the MC form,

$$\begin{aligned}
\nabla_0 \pi &= \partial_0 \pi + \boldsymbol{\xi} \cdot \boldsymbol{\nabla} \pi - \boldsymbol{\xi}^2/2, & \nabla_r \pi &= \partial_r \pi - \xi_r, \\
\nabla_0 \xi^r &= (\partial_0 + \boldsymbol{\xi} \cdot \boldsymbol{\nabla}) \xi^r, & \nabla_s \xi^r &= \partial_s \xi^r, \\
\nabla_0 X^i &= (\partial_0 + \boldsymbol{\xi} \cdot \boldsymbol{\nabla}) X^i, & \nabla_r X^i &= \partial_r X^i.
\end{aligned} \tag{14.22}$$

For simplicity of notation, I have already dropped the underscores distinguishing spacetime and frame indices.

The would-be NG field $\xi^r(x)$ is unphysical and can be eliminated by imposing the *inverse Higgs constraint* (IHC) $\nabla_r \pi = 0$, or $\boldsymbol{\xi}(x) = \boldsymbol{\nabla} \pi(x)$. This assigns ξ^r the counting degree zero. As a consequence, $\nabla_\mu \xi^r$ has degree one, whereas both $\nabla_\mu \pi$ and $\nabla_\mu X^i$ are of degree zero. The building blocks we have to construct the LO effective Lagrangian are therefore

$$\nabla_0 \pi \rightarrow \partial_0 \pi + (\boldsymbol{\nabla} \pi)^2/2, \quad \nabla_0 X^i \rightarrow (\partial_0 + \boldsymbol{\nabla} \pi \cdot \boldsymbol{\nabla}) X^i, \quad \nabla_r X^i \rightarrow \partial_r X^i. \tag{14.23}$$

These take automatically care of the nonlinearly realized symmetries under Galilei boosts, $U(1)_Q$ transformations and internal translations of X^i . To ensure invariance under spatial rotations, note that both $\nabla_0 \pi$ and $\nabla_0 X^i$ are scalars, whereas $\nabla_r X^i$ is a vector under spatial $SO(d)$. All spatial indices thus have to be contracted into

$$\tilde{\Xi}^{ij} \equiv \delta^{rs} \nabla_r X^i \nabla_s X^j = \boldsymbol{\nabla} X^i \cdot \boldsymbol{\nabla} X^j. \tag{14.24}$$

This symmetric tensor is related to the previously defined matrix $M^i_r(x) \equiv \partial_r X^i(x)$ by $\tilde{\Xi} = M M^T$. Note that there is no additional, algebraically independent rotationally invariant operator where the spatial indices on $\nabla_r X^i$ would be contracted with the LC symbol, since $\varepsilon^{r_1 \dots r_d} \nabla_{r_1} X^{i_1} \dots \nabla_{r_d} X^{i_d} = \varepsilon^{i_1 \dots i_d} \det M = \varepsilon^{i_1 \dots i_d} \sqrt{\det \tilde{\Xi}}$. Altogether, the LO effective action for nonrelativistic supersolids is given by [12]

$$S_{\text{eff}}\{\pi, X\} = \int d^D x F(\nabla_0 \pi(x), \nabla_0 X(x), \tilde{\Xi}(x)) + \dots. \tag{14.25}$$

The indices on $\nabla_0 X^i$ and $\tilde{\Xi}^{ij}$ are to be contracted in a way that respects invariance under the internal $SO(d)$ rotations.

Again, I have not explicitly considered the possibility of quasi-invariant contributions to the Lagrangian. There are $d+1$ such terms, constructed solely out of the superfluid mode π ; cf. Sect. 13.3.3. However, these are only relevant beyond the LO of the derivative expansion. An additional quasi-invariant Lagrangian existing in $d=2$ dimensions and including the solid variable X^i was found in [10]. This likewise contributes only at higher orders of the derivative expansion.

14.3.3 Nonrelativistic Solids

Let us now see whether we can recover the physics of classical (isotropic) solids from the supersolid EFT (14.25). It appears we should be able to simply erase the NG field $\pi(x)$. This certainly does not interfere with the nonlinear realization

of the symmetries of solids on X^i , since that is entirely independent of π . The building blocks (14.23) then boil down to $\partial_0 X^i$ and $\partial_r X^i$. Invariance under spatial rotations again requires that $\partial_r X^i$ only enters the LO EFT through the combination $\tilde{\Xi}^{ij}$. However, the time derivative, $\partial_0 X^i$, is potentially problematic since unlike $\nabla_0 X^i$, it is not invariant under Galilei boosts. This actually makes sense: recall that the nonrelativistic kinetic term in (14.2) or (14.7) is only quasi-invariant under Galilei boosts. Guided by the analogy, the kinetic term for the solid should be $(1/2)\varrho(\mathbf{x}, t)[\dot{\mathbf{x}}(X, t)]^2$. Here $\mathbf{x}(X, t)$ is the Lagrangian variable we worked with in Sect. 14.1 and $\varrho(\mathbf{x}, t)$ is the local mass density of the solid in the laboratory frame. Now $\varrho(x) = \varrho_0 \det M(x) = \varrho_0 \sqrt{\det \tilde{\Xi}(x)}$, where ϱ_0 is the mass density of the solid in the body coordinates, which is constant thanks to the assumed uniformity of the solid. Using finally (14.8), the LO effective action for the classical nonrelativistic isotropic solid can be written as

$$S_{\text{eff}}\{X\} = \int d^D x \left\{ \frac{\varrho_0}{2} \sqrt{\det \tilde{\Xi}(x)} [\tilde{\Xi}(x)^{-1}]_{ij} \partial_0 X^i(x) \partial_0 X^j(x) - \mathcal{F}(\tilde{\Xi}(x)) \right\} + \dots \quad (14.26)$$

The function $\mathcal{F}(\tilde{\Xi})$ defines the free energy of static elastic deformations of the solid. Our derivation of the kinetic term in (14.26) was based on an educated guess. It can however also be recovered by starting from an exactly Galilei-invariant operator in the EFT (14.25) for supersolids and adding a surface term, proportional to the gradient of $\pi(x)$. This explains why setting $\pi \rightarrow 0$ does not destroy Galilei invariance altogether but renders (14.26) quasi-invariant. See [12] for details.

I will now review the basic elastic properties of isotropic solids, following closely [4]. As in Sect. 14.1.1, the displacement of the solid element carrying label X^i from its equilibrium position is given by $\phi^i(X, t) \equiv \delta_r^i x^r(X, t) - X^i$. This parameterization is however not suitable for an EFT where the scalar fields $X^i(x)$ are the dynamical variables. We need to invert the relation between x^r and ϕ^i , which is not possible in a closed form. Luckily, we will only need the deviation of X^i from its VEV to first order in ϕ^i ,

$$X^i(x) = \delta_r^i x^r - \phi^i(x) + \mathcal{O}(\phi^2) \quad (14.27)$$

To the same order, the matrix variable $\tilde{\Xi}^{ij}$ reads

$$\tilde{\Xi}^{ij}(x) = \delta^{ij} - [\delta^{ir} \partial_r \phi^j(x) + \delta^{jr} \partial_r \phi^i(x)] + \mathcal{O}(\phi^2) \quad (14.28)$$

With the shorthand notation $\phi_r(x) \equiv \delta_{ri} \phi^i(x)$, small deformations of the solid are therefore encoded in the *strain tensor*

$$e_{rs}(x) \equiv \frac{1}{2} [\partial_r \phi_s(x) + \partial_s \phi_r(x)] \quad (14.29)$$

By its definition, $\phi^i(x)$ transforms as a vector under the unbroken diagonal $\text{SO}(d)$ subgroup. The strain tensor therefore behaves as a rank-2 symmetric tensor.

The energy cost of a small static elastic deformation of the solid is governed by the expansion of $\mathcal{F}(\tilde{\Xi})$ in powers of $\tilde{\Xi}^{ij} - \delta^{ij}$, or directly the strain tensor. The expansion cannot contain a linear term, as that would immediately imply an instability of the equilibrium state where $\langle \phi^i(x) \rangle = 0$. The physics of small elastic deformations is therefore dominated by the part of $\mathcal{F}(\tilde{\Xi})$ quadratic in the strain tensor. Invariance under unbroken diagonal rotations admits two independent quadratic operators at the lowest order in derivatives,

$$\mathcal{F} = \mu \operatorname{tr} e^2 + \frac{\lambda}{2} (\operatorname{tr} e)^2 + \dots \quad (14.30)$$

Here λ, μ are known as the *Lamé coefficients*, and the ellipsis denotes contributions of higher order in ϕ^i or derivatives. It is convenient to rewrite (14.30) as

$$\mathcal{F} = \mu \operatorname{tr} \left(e - \frac{\operatorname{tr} e}{d} \mathbb{1} \right)^2 + \frac{K}{2} (\operatorname{tr} e)^2 + \dots, \quad (14.31)$$

where $K \equiv \lambda + (2\mu/d)$. Importantly, there are configurations of the solid for which either of the two terms in (14.31) vanishes whereas the other is nonzero. The μ -term is zero if and only if e_{rs} is proportional to δ_{rs} , that is for uniform dilatations or contractions of the solid. Such overall volume change is detected by $\operatorname{tr} e = \nabla \cdot \phi$, which measures the deviation of the Jacobian $\det M$ from unity. For this reason, the parameter K is usually called *bulk modulus*. On the other hand, for deformations of shape that do not affect the volume of the solid, the K -term in (14.31) vanishes. The energy of such deformations is measured by the *shear modulus* μ . Since the two terms in (14.31) can be set to zero independently of each other, bulk stability of the solid requires that

$$K > 0, \quad \mu > 0. \quad (14.32)$$

Example 14.1

The constraints (14.32) may look trivial, but they have very observable consequences for small oscillations of the solid. To see this, it is sufficient to consider the part of (14.26), bilinear in the fields $\phi^i(x)$,

$$\mathcal{L}_{\text{eff}} = \frac{\varrho_0}{2} \delta^{rs} \partial_0 \phi_r \partial_0 \phi_s - \left[\mu \operatorname{tr} e^2 + \frac{\lambda}{2} (\operatorname{tr} e)^2 \right] + \dots \quad (14.33)$$

The corresponding EoM reads

$$\varrho_0 \partial_0^2 \phi_r \approx \mu \nabla^2 \phi_r + (\lambda + \mu) \partial_r \nabla \cdot \phi, \quad (14.34)$$

where the symbol \approx indicates a linear approximation. This has plane-wave solutions of the type

$$\phi_r(\mathbf{x}, t) = \hat{\phi}_r e^{-iEt} e^{i\mathbf{p} \cdot \mathbf{x}}, \quad (14.35)$$

where $\hat{\phi}_r$ is an amplitude. Solutions for which $\hat{\phi} \parallel \mathbf{p}$ and $\hat{\phi} \perp \mathbf{p}$ represent respectively *longitudinal* and *transverse* sound. The corresponding phase velocities are

$$v_L = \sqrt{\frac{\lambda + 2\mu}{\varrho_0}} = \sqrt{\frac{K + 2\mu(1 - 1/d)}{\varrho_0}}, \quad v_T = \sqrt{\frac{\mu}{\varrho_0}}. \quad (14.36)$$

The stability condition $K > 0$ implies a specific hierarchy between the velocities of longitudinal and transverse sound,

$$\frac{v_L}{v_T} > \sqrt{2 \left(1 - \frac{1}{d}\right)}. \quad (14.37)$$

This relation would have been completely invisible, had we focused solely on the spectrum of sound waves itself. When looking for the constraints on parameters of an EFT imposed by the stability requirement, it is therefore all-important to consider the whole configuration space of the EFT.

14.3.4 Perfect Fluids

As the last stop of our exploration of classical matter, we now return to fluids. I will restore full Poincaré invariance, partially because relativistic fluids are common in astrophysics and particle physics, and partially because it makes the analysis simpler.

As argued in Sect. 14.3.1, the NG fields $X^i(x)$ necessarily enter the effective Lagrangian through the combination $\Xi^{ij} \equiv \partial_\mu X^i \partial^\mu X^j$. However, not all actions of the solid type (14.15) are consistent with the required invariance of the fluid action under $G_{\text{int}} \simeq \text{SDiff}(\mathbb{R}^d)$. The transition from solids to fluids can be viewed as a “fine-tuning” of the effective couplings of the former. There is only one algebraically independent function of Ξ^{ij} that fits the bill, namely $\det \Xi \equiv |\Xi|$. Thus, the LO effective Lagrangian of a relativistic perfect (barotropic) fluid reads

$$S_{\text{eff}}\{X\} = \int d^D x F(|\Xi(x)|) + \dots. \quad (14.38)$$

With the help of the identity $\partial|\Xi|/\partial\Xi^{ij} = |\Xi|(\Xi^{-1})_{ji}$, we readily extract the corresponding LO EoM,

$$\partial_\mu \left[F'(|\Xi|) |\Xi| (\Xi^{-1})_{ij} \partial^\mu X^j \right] = 0, \quad (14.39)$$

where the prime indicates a derivative of $F(|\Xi|)$ with respect to its argument $|\Xi|$.

We could in principle stop here, for we have accomplished our goal to construct an EFT for perfect fluids. However, to shed light on the physical content of (14.38) and (14.39) requires additional work. The first step is to derive the *energy–momentum*

(EM) *tensor* of the EFT. This is a standard problem (see for instance Example 4.3),⁴ the result being

$$T^{\mu\nu} = 2F'(|\Xi|) |\Xi| (\Xi^{-1})_{ij} \partial^\mu X^i \partial^\nu X^j - g^{\mu\nu} F(|\Xi|) . \quad (14.40)$$

This already tells us, upon some manipulation, that the EoM (14.39) is equivalent to the local conservation laws for energy and momentum. The next step is to trade the explicit dependence on the derivatives $\partial_\mu X^i$ for the local Eulerian velocity of the fluid. To that end, consider the current (14.9) with $f(X) = 1$,

$$J^\mu \equiv J_1^\mu = \frac{1}{d!} \varepsilon^{\mu\nu_1 \dots \nu_d} \varepsilon_{i_1 \dots i_d} \partial_{\nu_1} X^{i_1} \dots \partial_{\nu_d} X^{i_d} . \quad (14.41)$$

It is a matter of straightforward algebra to verify that $J^\mu J_\mu = (-1)^d |\Xi|$. The extra sign compensates for our set of conventions, in which $\langle \Xi^{ij} \rangle = -\delta^{ij}$ but $\langle J^\mu \rangle = \delta^{\mu 0}$. The spacetime vector

$$u^\mu \equiv \frac{J^\mu}{\sqrt{(-1)^d |\Xi|}} \quad (14.42)$$

is then normalized to unity, $u^2 = 1$. Moreover, according to (14.10), its spatial part is proportional to the local Eulerian velocity $\dot{\mathbf{x}}$ of the fluid. It follows that u^μ is the velocity spacetime vector as observed in the laboratory frame.

Consider now the symmetric tensor $G_{\mu\nu} \equiv (\Xi^{-1})_{ij} \partial_\mu X^i \partial_\nu X^j + u_\mu u_\nu$. Since $\partial_\mu X^i$ as a spacetime vector is orthogonal to J^μ and thus u^μ , it follows that $G_{\mu\nu} u^\nu = u_\mu$. At the same time, it is easy to see that $G_{\mu\nu} \partial^\nu X^i = \partial_\mu X^i$. Together, $\partial^\mu X^i$ for $i = 1, \dots, d$ and u^μ constitute a basis of D linearly independent vectors on the spacetime, hence $G_{\mu\nu} A^\nu = A_\mu = g_{\mu\nu} A^\nu$ for any spacetime vector A^μ . This implies the useful identity

$$(\Xi^{-1})_{ij} \partial_\mu X^i \partial_\nu X^j = g_{\mu\nu} - u_\mu u_\nu , \quad (14.43)$$

which allows us to rewrite the EM tensor (14.40) as

$$T^{\mu\nu} = 2F'(|\Xi|) |\Xi| (g^{\mu\nu} - u^\mu u^\nu) - g^{\mu\nu} F(|\Xi|) . \quad (14.44)$$

In an inertial reference frame locally comoving with the fluid (local rest frame), the EM tensor should be diagonal and isotropic, $T^{\mu\nu} = \text{diag}(U, P, \dots, P)$. Here U is the local energy density and P the pressure of the fluid. This can be written in a covariant form as $T^{\mu\nu} = (U + P)u^\mu u^\nu - P g^{\mu\nu}$. The *scalar* functions U and P can be projected out of the EM tensor using $T^{\mu\nu} u_\mu u_\nu = U$ and $T^\mu_\mu = U - dP$. It follows that the as yet unknown function $F(|\Xi|)$ is related to the energy and pressure by

$$U(|\Xi|) = -F(|\Xi|) , \quad P(|\Xi|) = F(|\Xi|) - 2F'(|\Xi|) |\Xi| . \quad (14.45)$$

⁴ I use an opposite overall sign of the EM tensor as compared to Example 4.3 to ensure that the T^{00} component coincides with the canonical Hamiltonian of the EFT.

Example 14.2

To check that we have got the basic physics right, let us use our EFT to calculate the speed of sound in the fluid. For that, we need to find the bilinear part of the effective Lagrangian in the fluctuations $\pi^i(x) \equiv X^i(x) - \delta_r^i x^r$. I will outline the main steps but skip straightforward details. First, the fluctuation of Ξ^{ij} itself is given by (14.18),

$$\Xi^{ij} = -\delta^{ij} + \delta\Xi^{ij}, \quad \delta\Xi^{ij} \equiv -(\delta^{ir}\partial_r\pi^j + \delta^{jr}\partial_r\pi^i) + \partial_\mu\pi^i\partial^\mu\pi^j. \quad (14.46)$$

Using the identity $\det \Xi = \exp \text{tr} \log \Xi$, we next obtain the expansion

$$(-1)^d |\Xi| = 1 - \text{tr} \delta\Xi + \frac{1}{2} [(\text{tr} \delta\Xi)^2 - \text{tr}(\delta\Xi^2)] + \mathcal{O}(\delta\Xi^3). \quad (14.47)$$

What remains to be done is just a Taylor expansion of the Lagrangian, $\mathcal{L}_{\text{eff}} = F(|\Xi|)$, in π^i . Dropping a constant and a surface term, we find

$$\mathcal{L}_{\text{eff}} \simeq -(-1)^d F'_0 (\partial_0 \pi)^2 + [(-1)^d F'_0 + 2F''_0] (\nabla \cdot \pi)^2 + \dots, \quad (14.48)$$

where F'_0 and F''_0 are derivatives of $F(|\Xi|)$ taken in the equilibrium, $\langle |\Xi| \rangle = (-1)^d$. Note that (14.48) does not contain any transverse gradients of $\pi^i(x)$. This is expected, saying merely that transverse fluctuations do not propagate via harmonic oscillations due to the absence of shear elastic forces in a fluid. Fluids support only longitudinal sound waves, whose phase velocity follows as

$$v_L = \sqrt{1 + 2(-1)^d \frac{F''_0}{F'_0}}. \quad (14.49)$$

It is easy to see with the help of (14.45) that this equals $\sqrt{P'(|\Xi|)/U'(|\Xi|)} = \sqrt{dP/dU}$, as expected for thermodynamic sound.

Example 14.3

Another informative check of our EFT is to take the nonrelativistic limit. To that end, we use the relation (14.45) between the function $F(|\Xi|)$ and the energy density $U(|\Xi|)$ to decompose the former as

$$F(|\Xi|) = -\varrho_0 \sqrt{(-1)^d |\Xi|} - \mathcal{F}((-1)^d |\Xi|). \quad (14.50)$$

The constant ϱ_0 is the mass density of the fluid in the body coordinates. In the local rest frame, $\varrho_0 \sqrt{(-1)^d |\Xi|} \rightarrow \varrho_0 \sqrt{|\tilde{\Xi}|}$ where $\tilde{\Xi}^{ij}$ is defined by (14.24), is the density of the rest mass of the fluid. This is expected to dominate the energy in the nonrelativistic limit. On the other hand, $\mathcal{F}((-1)^d |\Xi|) \rightarrow \mathcal{F}(|\tilde{\Xi}|)$ becomes the elastic free energy of the fluid. Its contribution to $F(|\Xi|)$ is assumed to be suppressed relatively to the leading term in (14.50) by two inverse powers of the speed of light.

In the laboratory frame, we write $\Xi^{ij} = \partial_0 X^i \partial_0 X^j - \tilde{\Xi}^{ij}$ and again use the identity $\det \Xi = \exp \operatorname{tr} \log \Xi$. This gives

$$(-1)^d |\Xi| = |\tilde{\Xi}| \left[1 - (\tilde{\Xi}^{-1})_{ij} \partial_0 X^i \partial_0 X^j + \cdots \right], \quad (14.51)$$

where the ellipsis stands for higher-order corrections, negligible in the nonrelativistic limit. In turn, (14.45) becomes

$$\begin{aligned} U &= \varrho_0 \sqrt{|\tilde{\Xi}|} - \frac{\varrho_0}{2} \sqrt{|\tilde{\Xi}|} (\tilde{\Xi}^{-1})_{ij} \partial_0 X^i \partial_0 X^j + \mathcal{F}(|\tilde{\Xi}|) + \cdots, \\ P &= -\mathcal{F}(|\tilde{\Xi}|) + 2\mathcal{F}'(|\tilde{\Xi}|) |\tilde{\Xi}| + \cdots. \end{aligned} \quad (14.52)$$

This has a simple interpretation. The three contributions to U correspond, up to a sign, to the relativistic rest energy, the nonrelativistic kinetic energy and the thermodynamic (internal) energy, while P is the nonrelativistic pressure. The same expansion converts the action (14.38) to

$$\begin{aligned} S_{\text{eff}}\{X\} &= \int d^D x \left\{ -\varrho_0 \sqrt{|\tilde{\Xi}(x)|} + \frac{\varrho_0}{2} \sqrt{|\tilde{\Xi}(x)|} [\tilde{\Xi}(x)^{-1}]_{ij} \partial_0 X^i(x) \partial_0 X^j(x) \right. \\ &\quad \left. - \mathcal{F}(|\tilde{\Xi}(x)|) \right\} + \cdots. \end{aligned} \quad (14.53)$$

This copies the EFT (14.26) for nonrelativistic solids that we previously obtained by an educated guess. The only difference is that the gradient free energy \mathcal{F} is now allowed to depend only on $|\tilde{\Xi}|$. (The $\varrho_0 \sqrt{|\tilde{\Xi}(x)|}$ term gives upon integration a constant and can be dropped.) The EFT (14.53) for nonrelativistic fluids inherits the symmetry under VPDs of the target space \mathcal{M} from its relativistic ancestor (14.38). The spacetime part of its symmetry is however different, and includes spacetime translations, spatial rotations, and Galilei boosts.

The symmetry of fluids under VPDs, $G_{\text{int}} \simeq \text{SDiff}(\mathbb{R}^d)$, implies the existence of infinitely many conserved currents. Indeed, the action of an infinitesimal VPD can be expressed as $\delta X^i = \epsilon V^i(X)$, where ϵ is a parameter and $V^i(X)$ any smooth vector field on $\mathcal{M} \simeq \mathbb{R}^d$ with vanishing divergence, $\partial V^i(X)/\partial X^i = 0$. A straightforward application of Noether's theorem gives the current

$$J_V^\mu = F'(|\Xi|) |\Xi| (\Xi^{-1})_{ij} \partial^\mu X^i V^j(X). \quad (14.54)$$

Its on-shell conservation is also seen as a direct consequence of the EoM (14.39). The conservation of the corresponding integral charges generalizes the so-called Kelvin circulation theorem to relativistic fluids; see [13] for further details.

14.4 Coupling Nambu–Goldstone Bosons to Classical Matter

Before we wrap up the discussion of EFTs for classical matter, let us step back to see how it connects to the rest of the book. We started the analysis of spontaneously broken spacetime symmetries with a fairly general construction of nonlinear realization thereof in Chap. 12. However, in its subsequent applications, we gradually increased the assumptions on the symmetry and the order parameter. In the present chapter, we ended up discarding altogether the possible presence of NG modes associated with long-range order in the quantum ground state. Here it is possible to tie up some loose ends with little effort. An attentive reader might have noticed two related observations I made in quite different contexts. In Sect. 13.3.3, I pointed out that it might be possible to make an EFT with Aristotelian symmetry invariant under boosts, whether Galilei or Lorentz. All one needs is an auxiliary variable representing the velocity of the medium in which the EFT lives. On the other hand, we learned in Sect. 14.1.2 that this velocity is naturally encoded in the identically conserved current (14.9). Following this link, I will now sketch concretely how an Aristotelian EFT for broken internal symmetry can be made boost-invariant by coupling the EFT to a classical medium. A general EFT framework for such hybrid systems does not seem to exist as yet. For the sake of illustration, I will restrict the discussion to systems where the classical medium is a fluid. The construction is loosely inspired by [14].

Instead of lengthy reminders, I refer the reader to Chap. 8 for the details of construction of EFTs for broken internal symmetries. (See Sect. 8.1.4 for an executive summary.) All these EFTs were built assuming Aristotelian symmetry, that is symmetry under spacetime translations and spatial rotations. There are two details we have to attend to if we want to augment this spacetime symmetry with boosts. First, the temporal and spatial derivatives, which appear independently in the general Aristotelian effective Lagrangian (8.33), have to be combined in a way that respects boost invariance. Second, any operator invariant under all the spacetime and internal symmetries can be multiplied with an arbitrary function of $|\Xi|$ (for relativistic fluids) or $|\tilde{\Xi}|$ (for nonrelativistic fluids). From this point on, it is convenient to split the discussion of the two cases.

Suppose we want to make the Aristotelian EFT invariant under Galilei boosts. We need not do anything about spatial derivatives, which are already Galilei-invariant on their own. Temporal derivatives can be fixed by contraction with the current (14.41). In terms of the body coordinates X^i and the velocity $\dot{\mathbf{x}}$ of the medium, this amounts to replacing the time derivative of the NG field π^a of the broken internal symmetry with $J^\mu \partial_\mu \pi^a = \det M (\partial_0 \pi^a + \dot{\mathbf{x}} \cdot \nabla \pi^a) \equiv \det M (\nabla_0 \pi^a)$. The minimal modification of the two-derivative part of the effective Lagrangian (8.33) then reads

$$\begin{aligned} \mathcal{L}_{\text{eff}}^{(2,0)} &= -\frac{1}{2} \kappa_{cd}(|\tilde{\Xi}|) \omega_a^c(\pi) \omega_b^d(\pi) \nabla \pi^a \cdot \nabla \pi^b \\ &\quad - \frac{1}{2} \lambda_{cd}(|\tilde{\Xi}|) \omega_a^c(\pi) \omega_b^d(\pi) \varepsilon^{rs} \partial_r \pi^a \partial_s \pi^b, \\ \mathcal{L}_{\text{eff}}^{(0,2)} &= \frac{1}{2} \bar{\kappa}_{cd}(|\tilde{\Xi}|) \omega_a^c(\pi) \omega_b^d(\pi) \nabla_0 \pi^a \nabla_0 \pi^b. \end{aligned} \quad (14.55)$$

The two factors of $\det M$ coming from the replacement $\partial_0 \pi^a \rightarrow J^\mu \partial_\mu \pi^a$ are without loss of generality absorbed into $\bar{\kappa}_{ab}(|\tilde{\Xi}|)$. To preserve the internal symmetry acting on π^a , the coupling functions $\kappa_{ab}(|\tilde{\Xi}|)$, $\bar{\kappa}_{ab}(|\tilde{\Xi}|)$ and $\lambda_{ab}(|\tilde{\Xi}|)$ must satisfy the constraints (8.34) and (8.35) for all values of $X^i(x)$. The precise dependence of the coupling functions on $|\tilde{\Xi}|$ can be fixed by measuring the effective couplings κ_{ab} , $\bar{\kappa}_{ab}$ and λ_{ab} as a function of the density of the underlying medium.

The only nontrivial bit of the “Galileanization” of the EFT (8.33) lies in the part of the Lagrangian with a single time derivative, $\mathcal{L}_{\text{eff}}^{(0,1)}$. In case this is also strictly invariant under the internal symmetry, it can be promoted to

$$\mathcal{L}_{\text{eff}}^{(0,1)} = -\sigma_b(|\tilde{\Xi}|)\omega_a^b(\pi)\nabla_0\pi^a, \quad (14.56)$$

where $\sigma_b(|\tilde{\Xi}|)$ is subject to the condition $f_{ab}^b\sigma_b(|\tilde{\Xi}|) = 0$; cf. (8.39). However, in case $\mathcal{L}_{\text{eff}}^{(0,1)}$ is merely quasi-invariant, we cannot multiply it by an arbitrary function of $|\tilde{\Xi}|$. The only way out then seems to be to set

$$\mathcal{L}_{\text{eff}}^{(0,1)} = c_a(\pi)J^\mu\partial_\mu\pi^a, \quad (14.57)$$

where the 1-form $c(\pi) \equiv c_a(\pi)d\pi^a$ is constrained by the required strict invariance of its exterior derivative, $dc(\pi)$. This makes invariance under Galilei boosts manifest, while quasi-invariance under the internal symmetry acting on π^a is preserved thanks to the identical conservation of J^μ . Note that in this special case, imposing Galilei invariance produces a unique coupling between the NG fields π^a and the matter variables X^i , without any new a priori unknown parameters. Mathematically, the Lagrangian (14.57) descends from the $(d+1)$ -form $c(\pi) \wedge dX^1 \wedge \cdots \wedge dX^d$ on the coset space of broken internal symmetry.

Example 14.4

Our main example of a system with a $\mathcal{L}_{\text{eff}}^{(0,1)}$ Lagrangian have been ferromagnets (see Sect. 9.2). While most natural ferromagnets are metals, there are also ferromagnetic materials that are crystalline insulators. The low-energy EFT of such materials brings together ferromagnetic magnons and the phonons of the crystal lattice. In this case, the dependence of the two-derivative Lagrangian on the tensor $\tilde{\Xi}^{ij}$ is more complicated than in (14.55). The reason for this is the reduced emergent symmetry of crystals as compared to fluids; see [14] for further details. However, the one-derivative part of the Lagrangian is still (14.57), generating a unique coupling between magnons and phonons.

Let us finally see how the above argument needs to be modified, should we replace the requirement of Galilei invariance with the relativistic Poincaré invariance. In this case, the two-derivative Lagrangian may only feature properly contracted Lorentz indices. Embedding the antisymmetric λ_{ab} -term into a Lorentz-invariant operator is possible by replacing $\varepsilon^{rs}\partial_r\pi^a\partial_s\pi^b \rightarrow \varepsilon^{\lambda\mu\nu}J_\lambda\partial_\mu\pi^a\partial_\nu\pi^a$. Analogously, the spatial gradient operator $\nabla\pi^a \cdot \nabla\pi^b$ should be embedded in $-g^{\mu\nu}\partial_\mu\pi^a\partial_\nu\pi^b$. In order to

maintain the form of the two-derivative Lagrangian in the local rest frame of the underlying medium, it is then convenient to write it as

$$\begin{aligned}\mathcal{L}_{\text{eff}}^{(2,0)} &= \frac{1}{2}\kappa_{cd}(|\Xi|)\omega_a^c(\pi)\omega_b^d(\pi)\partial_\mu\pi^a\partial^\mu\pi^b \\ &\quad - \frac{1}{2}\lambda_{cd}(|\Xi|)\omega_a^c(\pi)\omega_b^d(\pi)\varepsilon^{\lambda\mu\nu}J_\lambda\partial_\mu\pi^a\partial_\nu\pi^b, \\ \mathcal{L}_{\text{eff}}^{(0,2)} &= \frac{1}{2}[\bar{\kappa}_{cd}(|\Xi|) - \kappa_{cd}(|\Xi|)]\omega_a^c(\pi)\omega_b^d(\pi)(J^\mu\partial_\mu\pi^a)(J^\nu\partial_\nu\pi^b).\end{aligned}\quad (14.58)$$

This naturally incorporates the possibility for type-A NG bosons to propagate with a velocity unrelated to the speed of light while maintaining full Poincaré invariance. The coupling functions $\kappa_{ab}(|\Xi|)$, $\bar{\kappa}_{ab}(|\Xi|)$ and $\lambda_{ab}(|\Xi|)$ are still subject to the constraints (8.34) and (8.35) for all values of $X^I(x)$. The one-derivative Lagrangian $\mathcal{L}_{\text{eff}}^{(0,1)}$, should it be strictly invariant, can be coupled to the medium by a slight modification of (14.56),

$$\mathcal{L}_{\text{eff}}^{(0,1)} = -\sigma_b(|\Xi|)\omega_a^b(\pi)J^\mu\partial_\mu\pi^a. \quad (14.59)$$

When $\mathcal{L}_{\text{eff}}^{(0,1)}$ is merely quasi-invariant, it can still be coupled to the classical medium via (14.57). Namely, (14.57) in fact respects the infinite-dimensional group of transformations preserving the spacetime volume form thanks to its origin in the $(d+1)$ -form $c(\pi) \wedge dX^1 \wedge \cdots \wedge dX^d$. This group includes both Galilei and Lorentz boosts.

References

1. J.V. José, E.J. Saletan, *Classical Dynamics: A Contemporary Approach* (Cambridge University Press, Cambridge, UK, 1998)
2. D.E. Soper, *Classical Field Theory* (Dover Publications, 2008)
3. M. Stone, in *Artificial Black Holes*, ed. by M. Novello, M. Visser, G. Volovik (World Scientific, Singapore, 2002), pp. 335–364
4. L.D. Landau, E.M. Lifshitz, *Theory of Elasticity, Course of Theoretical Physics*, vol. 7, 3rd edn. (Pergamon Press, Oxford, UK, 1986)
5. V.I. Arnold, B.A. Khesin, *Topological Methods in Hydrodynamics* (Springer-Verlag, New York, 1998)
6. R. Jackiw, V.P. Nair, S.Y. Pi, A.P. Polychronakos, J. Phys. **A37**, R327 (2004). DOI [10.1088/0305-4470/37/42/R01](https://doi.org/10.1088/0305-4470/37/42/R01)
7. N. Andersson, G.L. Comer, Living Rev. Rel. **24**, 3 (2021). DOI [10.1007/s41114-021-00031-6](https://doi.org/10.1007/s41114-021-00031-6)
8. A. Nicolis, R. Penco, R.A. Rosen, Phys. Rev. **D89**, 045002 (2014). DOI [10.1103/PhysRevD.89.045002](https://doi.org/10.1103/PhysRevD.89.045002)
9. H. Liu, P. Glorioso, PoS **TASI2017**, 008 (2018). DOI [10.22323/1.305.0008](https://doi.org/10.22323/1.305.0008)
10. L.V. Delacrétaz, A. Nicolis, R. Penco, R.A. Rosen, Phys. Rev. Lett. **114**, 091601 (2015). DOI [10.1103/PhysRevLett.114.091601](https://doi.org/10.1103/PhysRevLett.114.091601)
11. S. Endlich, A. Nicolis, J. Wang, JCAP **10**, 011 (2013). DOI [10.1088/1475-7516/2013/10/011](https://doi.org/10.1088/1475-7516/2013/10/011)
12. D.T. Son, Phys. Rev. Lett. **94**, 175301 (2005). DOI [10.1103/PhysRevLett.94.175301](https://doi.org/10.1103/PhysRevLett.94.175301)
13. S. Dubovsky, T. Gregoire, A. Nicolis, R. Rattazzi, JHEP **03**, 025 (2006). DOI [10.1088/1126-6708/2006/03/025](https://doi.org/10.1088/1126-6708/2006/03/025)
14. S. Pavaskar, R. Penco, I.Z. Rothstein, SciPost Phys. **12**, 155 (2022). DOI [10.21468/SciPostPhys.12.5.155](https://doi.org/10.21468/SciPostPhys.12.5.155)

Part V
Epilogue

Chapter 15

Topics Not Covered in This Book

The story of *spontaneous symmetry breaking* (SSB) and its *effective field theory* (EFT) description does not end here. There are several further facets of SSB that would have deserved place in the table of contents of the book. However, giving them proper credit would either increase the volume of the book beyond reasonable limits, or require a substantial amount of additional background. In this chapter, I will give a brief primer on some of these exciting advanced aspects of SSB. I will be able to go to detail where it is feasible based on the material covered elsewhere in the book. For topics that depart substantially from the main text, I will resort to a few basic comments augmented with references for further reading.

15.1 Effects of Nonzero Temperature

The great majority of Parts [III](#) and [IV](#) the book is restricted to the leading order of the derivative expansion of the EFT for *Nambu–Goldstone* (NG) *bosons*. This is just the classical approximation to the EFT, based on the leading-order effective Lagrangian and using only tree-level Feynman diagrams. At a few exceptional spots, such as the proof of the Goldstone theorem in Sect. [6.2](#) or of the Adler zero property in Sect. [10.1](#), the presented argument is clearly valid beyond the classical approximation. However, I have not discussed explicitly perturbative loop corrections except for an outline of their role in the derivative expansion; cf. Sects. [9.1.1](#) and [9.2.1](#).

The effects of nonzero temperature have thus fallen through the cracks together with other loop corrections. Namely, in the so-called imaginary time formalism, the temperature of a system in thermodynamic equilibrium enters through discrete sums over Matsubara frequencies in loop diagrams. This is a standard part of field theory and I will therefore not dwell on details. A reader interested in the application of thermal perturbation theory to EFT for NG bosons will find more information for instance in [\[1\]](#) (chiral perturbation theory of mesons) or [\[2\]](#) (EFT for ferromagnetic magnons). One generic aspect of SSB worth mentioning explicitly is that the order parameter tends to be reduced by thermal fluctuations. As a rule, the phase where a

symmetry is spontaneously broken will persist only up to certain critical temperature, above which the symmetry is “restored.” This is of course mere jargon. The symmetry of a system constrains its dynamics at all temperatures. However, above the critical temperature, the order parameter for SSB vanishes and excitations in the spectrum are organized in multiplets of the full symmetry group. See Chap. 7 of [3] for a discussion of symmetry restoration within thermal field theory.

What are the thermal effects on the spectrum of NG bosons? First of all, the Goldstone theorem remains valid at nonzero temperature as long as the symmetry is spontaneously broken. See Chap. 26 of [4] for a proof of the existence of a stable gapless quasiparticle, the NG boson, at nonzero temperature. Likewise, the distinction between type-A and type-B NG bosons survives at nonzero temperature. A detailed investigation of the thermal spectrum of NG bosons was carried out in [5]. At nonzero temperature, quasiparticles manifest themselves by poles in the propagator in the lower half-plane of complex energy. The two types of NG modes differ in the way their thermal width scales with momentum in the long-wavelength limit. As a rule, the complex energy of a type-A NG boson is thus schematically $E(\mathbf{p}) \propto |\mathbf{p}| - i\mathbf{p}^2$, whereas that of a type-B NG boson is $E(\mathbf{p}) \propto \mathbf{p}^2 - i\mathbf{p}^4$. In both cases, the ratio of thermal width and the real part of the energy tends to zero in the limit $\mathbf{p} \rightarrow \mathbf{0}$. This ensures the stability of the NG modes.

The imaginary time formalism is only suitable for describing matter in thermodynamic equilibrium. In the past couple of decades, much progress has been made in the development of quantum-field-theoretic methods for thermodynamic systems out of equilibrium. The extension of these techniques to EFTs based on nonlinear realization of symmetry is fairly recent. The reader will find further details in [6–8].

15.2 No-Go Theorems for Spontaneous Symmetry Breaking

The entire book is based on the *assumption* that the symmetry of a given system is spontaneously broken. It is however equally interesting to try to understand under what circumstances SSB may in fact occur. The most striking results in this regard form a collection of “no-go” theorems, forbidding SSB in a specific class of theories. What follows below is, to the best of my knowledge, a representative list. The results included differ somewhat in the level of rigor at which they have been proven, and in the mathematical techniques required to establish them. Having just discussed the effect of loop corrections, it is natural to start with several related statements where the fluctuations of the order parameter play a key role.

Coleman Theorem In Lorentz-invariant systems, NG bosons necessarily have a relativistic dispersion relation, $E(\mathbf{p}) = |\mathbf{p}|$. It is easy to see that the corresponding free propagator in coordinate space,

$$\int \frac{d^D p}{(2\pi)^D} \frac{e^{-i\mathbf{p} \cdot \mathbf{x}}}{p^2}, \quad (15.1)$$

Table 15.1 Subdivision of NG bosons into classes based on the noninteracting part of the effective Lagrangian. The latter determines the form of the (inverse) propagator, which in turn gives the dispersion relation of the NG mode. The last two columns of the table indicate spacetime dimensions, allowed by the condition that the coordinate-space propagator does not diverge at long distances

Type	Inverse propagator	Dispersion relation	Dimension ($T = 0$)	Dimension ($T > 0$)
A_n	$E^2 - \mathbf{p}^{2n}$	$E(\mathbf{p}) \propto \mathbf{p} ^n$	$D \geq n + 2$	$D \geq 2n + 2$
B_{2n}	$E - \mathbf{p}^{2n}$	$E(\mathbf{p}) \propto \mathbf{p} ^{2n}$	any $D \geq 2$	$D \geq 2n + 2$

is infrared-divergent for $D = 2$ spacetime dimensions. This translates into a divergence of the two-point correlation function of the order parameter at long distances. In other words, the fluctuations arising from the NG boson destroy the assumed order parameter. This is the essence of Coleman's theorem [9], which forbids spontaneous breaking of a continuous symmetry in two-dimensional relativistic systems.

This argument lends itself to a broad generalization. Consider a generic effective Lagrangian whose bilinear part starts at order $2n$ in spatial derivatives. While the $n = 1$ option is most natural, we also saw in Sect. 13.3.2 an example of a system where $n = 2$, at least in some spatial directions. More generally, the lack of low-derivative contributions to the spatial part of the kinetic term can be naturally explained by the presence of a coordinate-dependent symmetry [10, 11]. As to the part of the effective Lagrangian carrying time derivatives, we expect either one or two derivatives respectively for type-A and type-B NG bosons. This leads to a refined classification of NG bosons as type- A_n or type- B_{2n} . See the first three columns of Table 15.1 for an overview of the schematic forms of the corresponding free-particle propagators and dispersion relations. A simple modification of the argument leading to Coleman's theorem now shows that at zero temperature, type- A_n NG bosons are forbidden in $D \leq n + 1$ spacetime dimensions. Intriguingly, no such a constraint exists for type- B_{2n} NG bosons. In order to have any spectrum of quasiparticles, at least one spatial dimension must of course be present. However, for any $D \geq 2$, the propagator of a type- B_{2n} boson at zero temperature is infrared-finite. See the fourth column of Table 15.1 for a summary.

Hohenberg–Mermin–Wagner Theorem At nonzero temperature, the situation changes dramatically. Here the Minkowski-spacetime integral (15.1) is replaced with an imaginary-time sum-integral of the schematic type

$$T \sum_{k=-\infty}^{+\infty} \int \frac{d^d \mathbf{p}}{(2\pi)^d} \frac{\exp(-i\omega_k t + i\mathbf{p} \cdot \mathbf{x})}{(i\omega_k \text{ or } \omega_k^2) + \mathbf{p}^{2n}}, \quad (15.2)$$

where $\omega_k \equiv 2k\pi T$ is a bosonic Matsubara frequency and T the temperature. The type- A_n and type- B_{2n} cases correspond respectively to $i\omega_k$ and ω_k^2 in the denominator. Due to the discrete nature of the Matsubara sum, the contributions of all terms with $k \neq 0$ are infrared-finite. An infrared divergence can only arise from the zero Matsubara mode, and will be absent provided $d \geq 2n + 1$ or $D \geq 2n + 2$ [12] (the

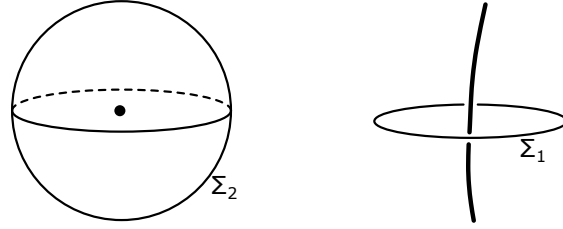
last column of Table 15.1). While the critical dimension is the same for type- A_n and type- B_{2n} NG bosons, note that n here counts the number of spatial derivatives in the Lagrangian. For fixed power of momentum in the dispersion relation, the critical dimension for type- A_{2n} NG bosons is higher than that for type- B_{2n} ones.

The original results on the absence of SSB at nonzero temperature are due to Hohenberg [13] (for superfluids) and Mermin and Wagner [14] (for isotropic lattice models of ferro- and antiferromagnets). These covered implicitly the cases of type- A_1 and type- B_2 NG bosons. However, their approach was different from the above intuitive argument and did not rely on the propagator of the would-be NG mode. Rather, they used the so-called Bogoliubov inequality to place an upper bound on the order parameter. This bound shows that the order parameter goes to zero in the limit of infinite volume and vanishing symmetry-breaking perturbation. See [15] for a review and further references.

Landau–Peierls Instability There is another related result that applies to systems where the order parameter is spatially modulated in one direction, thus spontaneously breaking translations. This is the case for instance for liquid crystals in the smectic-A phase, as we saw in Sect. 13.3.2. Let us briefly recall the line of reasoning therein to stress its generality. Suppose that the order parameter is given by a scalar field ϕ that develops a nonzero gradient in the ground state. We can treat $\langle \nabla \phi \rangle$ as a secondary, vector order parameter. The assumption that translations are spontaneously broken only in one dimension amounts to the condition that $\langle \nabla \phi \rangle$ points in the same direction everywhere in space. Next, we expand ϕ around its expectation value as $\phi \equiv \langle \phi \rangle + \pi$ and focus on the part of the effective Lagrangian bilinear in the fluctuation π . The assumed (and also spontaneously broken) *rotation* invariance dictates that there cannot be any term in the Lagrangian, bilinear in the part of $\nabla \pi$, perpendicular to $\langle \nabla \phi \rangle$. The gradient expansion of the Lagrangian starts with terms proportional to $(\nabla_{\parallel} \pi)^2$ and $(\nabla_{\perp}^2 \pi)^2$, where \parallel and \perp denote projections to subspaces parallel and perpendicular to $\langle \nabla \phi \rangle$. The denominator of the large fraction in (15.2) should then be replaced with $(i\omega_k \text{ or } \omega_k^2) + \mathbf{p}_{\parallel}^2 + \mathbf{p}_{\perp}^4$. Integrating the $k = 0$ term over \mathbf{p}_{\parallel} shows that the order parameter is washed out by thermal fluctuations at any nonzero temperature whenever $d \leq 3$. This is known as the *Landau–Peierls instability*; see Sect. 1.6 of [16] and [17] for a more detailed discussion within the condensed-matter and nuclear physics context, respectively.

Absence of Time Crystals Speaking of spontaneously breaking translations in a single direction, it is mandatory to consider the possibility that this direction corresponds to time. The idea that time translations could be spontaneously broken was proposed in [18, 19]. Such systems have since been known as *time crystals*. The reason why this possibility had not been noticed until the 21st century is that time crystals are not easy to realize. It was shown soon after the original proposal that their existence is forbidden in thermodynamic equilibrium of any Hamiltonian with

Fig. 15.1 Examples of topological defects in $d = 3$ spatial dimensions together with the surfaces defining their topological charge. The left panel shows a point defect (monopole), the right panel a line defect (vortex)



sufficiently short-range interactions [20, 21]. There is however a nonequilibrium route towards quantum time crystals; see [22] for a recent review of the subject.

Vafa–Witten Theorem Finally, the list of no-go theorems would not be complete without the Vafa–Witten theorem [23]. This is of quite a different nature than all the other results above, and applies to Lorentz-invariant gauge theories coupled to fermionic matter in a vector-like manner. An important example of such a theory is the *quantum chromodynamics* (QCD). Suppose the theory possesses a vector-like symmetry, that is one that acts in the same way on left- and right-handed fermions. In QCD, this could be for instance the $U(1)_B$ baryon number symmetry or, in the limit of equal quark masses, the $SU(2)_V$ isospin symmetry. The theorem states that such vector-like symmetries cannot be spontaneously broken in the ground state. The proof is technical. See also [24] for a review of the relevant mathematical methods and their applications to QCD and hadron physics.

15.3 Topological Aspects of Spontaneous Symmetry Breaking

We already encountered various topological aspects of EFTs for NG bosons on several occasions. Below, I will generalize some of the observations made previously, and add new interesting physics.

Topological Defects and Solitons The classification of *topological defects* was the first major application of topology to physics; see [25] for an introduction to the relevant mathematics. Recall that the coset space G/H can be viewed as the vacuum manifold of a theory with SSB, whose points indicate possible values of the order parameter in the ground state. A defect is a nonuniform configuration of the G/H -valued order parameter that is singular on some subdimensional domain in space. See Fig. 15.1 for the simple examples of a point and line defect. The presence of a defect can be established from the properties of the order parameter away from the singularity. Thus, a p -dimensional defect in \mathbb{R}^d can be enclosed by a $(d - p - 1)$ -dimensional hypersurface Σ_{d-p-1} that is topologically a sphere, S^{d-p-1} . Equivalence classes of defects can then be determined by studying maps $S^{d-p-1} \rightarrow G/H$, which are classified by the *homotopy group* $\pi_{d-p-1}(G/H)$. This assigns the defect a unique label: its *topological charge*. An exception are domain walls, which are of codimension one in space, dividing it into two halves. A domain

Table 15.2 Classification of basic topological defects by their codimension in space \mathbb{R}^d

Defect	Dimension p	Classifying group
Domain wall	$d - 1$	$\pi_0(G/H) \times \pi_0(G/H)$
Vortex	$d - 2$	$\pi_1(G/H)$
Monopole	$d - 3$	$\pi_2(G/H)$

wall is thus characterized by the values of the order parameter on both sides. See Table 15.2 for an overview of basic types of topological defects.

Example 15.1

We saw examples of domain walls in Sect. 13.4. The double-well potential (13.78) has a discrete $G \simeq \mathbb{Z}_2$ symmetry under $\phi \rightarrow -\phi$. This is broken in its minima to $H \simeq \{e\}$, hence $G/H \simeq \mathbb{Z}_2$ and likewise $\pi_0(G/H) \simeq \mathbb{Z}_2$. The cosine potential (13.80), on the other hand, has the symmetry group $G \simeq \mathbb{Z}_2 \ltimes \mathbb{Z}$. The \mathbb{Z}_2 factor is generated by the inversion $\phi \rightarrow -\phi$ whereas the \mathbb{Z} factor by the shift $\phi \rightarrow \phi + 2\pi v$. In any of the minima of the potential, the symmetry is broken to a subgroup isomorphic to $H \simeq \mathbb{Z}_2$. Therefore, in this case $G/H \simeq \mathbb{Z}$ and accordingly $\pi_0(G/H) \simeq \mathbb{Z}$. The elements of $\pi_0(G/H) \times \pi_0(G/H)$ shown in Table 15.2 correspond to domain wall solutions interpolating between different pairs of minima.

The focus on singular field configurations may be surprising; the physics should not change if we “smooth down” the order parameter. This is however only possible at the cost of embedding the coset space G/H in a larger order parameter manifold \mathcal{M} .

Example 15.2

Consider a superfluid in $d = 2$ dimensions. Here the order parameter $\langle\psi\rangle$ takes values from \mathbb{C} , while the vacuum manifold is $G/H \simeq \text{U}(1)/\{e\} \simeq S^1$ for any $\langle\psi\rangle \neq 0$. A point defect (vortex) in the superfluid is described by a smooth complex field $\psi(\mathbf{x})$. The corresponding topological charge is the *winding number*, defining an element of $\pi_1(G/H) \simeq \pi_1(S^1) \simeq \mathbb{Z}$; see also Example 12.4. Importantly, nonzero winding number implies by the argument principle (Chap. 7 of [26]) the existence of a point $\mathbf{x} \in \mathbb{R}^2$ where $\psi(\mathbf{x}) = 0$: the core of the vortex. This is our “singularity” where the order parameter cannot lie on the vacuum manifold $G/H \simeq S^1$.

For another example, consider a spin system in $d = 3$ dimensions with $G/H \simeq S^2$. Suppose that on the sphere Σ_2 as shown in Fig. 15.1, the order parameter represented by a unit vector $\mathbf{n} \in S^2$ points radially outwards everywhere. This spin configuration is sometimes called *hedgehog*. If we now try to continue the field to the inside of the sphere, we will inevitably encounter a point singularity: a monopole. Possible values of its topological charge are classified by $\pi_2(G/H) \simeq \pi_2(S^2) \simeq \mathbb{Z}$. The singularity can be avoided only at the cost of deforming the spin configuration out of the coset

space G/H . The natural way to do so is by embedding $G/H \simeq S^2$ in \mathbb{R}^3 and allowing \mathbf{n} to vary its magnitude. One can then extend the hedgehog field on Σ_2 to the inside by scaling down its magnitude so that it vanishes at the location of the monopole.

In contrast to defects are *topological solitons* (Sect. IX of [25]). These are fields in \mathbb{R}^d that are smooth and take values from G/H everywhere. Nontrivial topology arises from imposing a specific boundary condition at spatial infinity. Thus, the field configuration of a p -dimensional soliton is required to converge to a constant far from a fixed p -dimensional hypersurface C_p in \mathbb{R}^d . Now consider an open hypersurface Σ_{d-p} that is “transverse” to C_p . The boundary condition effectively compactifies this hypersurface to a sphere, S^{d-p} . Equivalence classes of p -dimensional solitons in \mathbb{R}^d are therefore classified by the homotopy group $\pi_{d-p}(G/H)$.

Example 15.3

The simplest type of a topological soliton is that with $p = 0$. Such solitons can be viewed as quasiparticles localized in all directions in space, and typically carry a finite amount of energy. The transverse hypersurface Σ_d is in this case the entire space \mathbb{R}^d . We already met two examples of such solitons in Chap. 9: *skyrmions* in QCD, corresponding to $d = 3$ and $\pi_3(G/H) \simeq \pi_3(S^3) \simeq \mathbb{Z}$, and *baby skyrmions* in two-dimensional ferromagnets, for which $\pi_2(G/H) \simeq \pi_2(S^2) \simeq \mathbb{Z}$. The latter are easily generalized to higher dimensions, simply by making the fields independent of whatever extra coordinates are present. Thus a skyrmion in a three-dimensional ferromagnet is localized to a line, C_1 . Its topological charge is determined by the properties of a unit vector field on a two-dimensional surface Σ_2 , transverse to C_1 .

Topological Terms in the Action We saw in Sect. 8.1 that the part of the effective Lagrangian for NG bosons with a single time derivative may be topologically nontrivial. Similarly, it turned out in Sect. 9.1 that the part of the low-energy EFT of QCD, taking into account the chiral anomaly, is topological in nature. We shall now generalize these observations, thus uncovering a broad class of contributions to EFTs for NG bosons, arising from the topology of the coset space G/H .

Suppose that G/H is compact and consider a closed p -form $\omega^{(p)}$ that belongs to a nontrivial cohomology class in the p -th de Rham cohomology group, $H^p(G/H)$.¹ Upon pulling $\omega^{(p)}$ back to a spacetime M of dimension D by the map $M \rightarrow G/H$ defining the NG fields, it becomes a likewise closed p -form on M . There are now three different possibilities depending on the relation between p and D . If $p = D$, the spacetime p -form can be directly integrated and added to the action of the EFT. This is called a θ -term. Due to its origin in a closed D -form, it is a mere surface term in the Lagrangian and thus does not affect the perturbative dynamics of the EFT. This is the reason why θ -terms have not featured in the book at all. If present, they

¹ I follow the convention common in high-energy physics and indicate the degree of a differential form with a superscript in parentheses.

Table 15.3 Physical interpretation of generators of the de Rham cohomology group $H^p(G/H)$, depending on the dimension D of spacetime they are pulled back to. For $p < D$, the pull-back leads to a closed p -form on the spacetime, Hodge-dual to an identically conserved tensor current of rank $D - p$. In the table, the rank of currents $J^{\mu\nu\cdots}$ is indicated by the number of indices

Dimension	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
$D = 1$	θ -term	WZ term	—	—	—
$D = 2$	J^μ	θ -term	WZ term	—	—
$D = 3$	$J^{\mu\nu}$	J^μ	θ -term	WZ term	—
$D = 4$	$J^{\mu\nu\lambda}$	$J^{\mu\nu}$	J^μ	θ -term	WZ term

contribute a phase factor to the generating functional though, which may affect the nonperturbative physics of the theory.

Another relevant possibility is $p = D+1$. In this case, there is a locally well-defined D -form $\omega^{(D)}$ on G/H such that $d\omega^{(D)} = \omega^{(p)}$. Upon pulling back to the spacetime, $\omega^{(D)}$ gives a contribution to the Lagrangian that is quasi-invariant under the action of G . This is a *Wess–Zumino (WZ) term*. The anomalous contribution to the low-energy effective Lagrangian of QCD is obviously of this kind; cf. Sect. 9.1.4. However, the single-time-derivative part of the effective Lagrangian, $\mathcal{L}_{\text{eff}}^{(0,1)}$, constructed in Sect. 8.1.3, belongs to the same category. Here only the variation of the NG fields in time matters and thus effectively $D = 1$, which makes $H^2(G/H)$ the relevant de Rham cohomology group. Following the line of reasoning of Sect. 9.2.4, we see moreover that such WZ terms generally give rise to a Berry phase when the ground state is driven by an external field [27]. See Table 15.3 for a summary. Further technical details on the construction of topological WZ and θ -terms in the action can be found in the recent study [28]. For a somewhat more condensed-matter oriented perspective on topological terms, see Chap. 9 of [29].

15.4 Generalized Symmetries

It remains to clarify the content of the lower left corner of Table 15.3. For $p < D$, the spacetime p -form arising from pulling back $\omega^{(p)}$ can be expressed as the Hodge dual of a $(D - p)$ -form, $\star J^{(D-p)}$. The form $J^{(D-p)}$ has a vanishing codifferential; cf. Appendix A.6.5. In a flat spacetime, this can be expressed component-wise as

$$\partial_\mu J^{\mu\nu\cdots} = 0, \quad (15.3)$$

where $J^{\mu\nu\cdots}$ is fully antisymmetric in all its indices. Equation (15.3) is nothing but a local conservation law with a tensor current of rank $D - p$. It encodes, on the classical level, a *generalized symmetry*. This (and other) kind of generalization of conservation laws has attracted much attention in the last decade. An interested reader will find a comprehensive introduction for instance in [30–32].

In the context of EFT for NG bosons, the identically conserved currents $J^{\mu\nu\cdots}$ give us a tool for calculation of topological charges of defects and solitons, classified in Sect. 15.3. Mathematically, this is based on the correspondence between the homotopy and de Rham cohomology groups of G/H via the Hurewicz theorem (see Appendix A.8.2). Physically, the conservation law (15.3) suggests the existence of a conserved charge, obtained by integrating the temporal component, $J^{0rs\cdots}$, over a spatial hypersurface. For a static defect or soliton, it is easiest to pull $\omega^{(p)}$ back directly to the space \mathbb{R}^d and express it therein as $\star \mathcal{J}^{(d-p)}$. The $(d-p)$ -form $\mathcal{J}^{(d-p)}$ is the density of our topological charge. The topological charge itself is obtained by integration over an effectively closed p -dimensional surface Σ_p in \mathbb{R}^d ,

$$Q(\Sigma_p) \equiv \int_{\Sigma_p} \star \mathcal{J}^{(d-p)} . \quad (15.4)$$

The fact that Σ_p is closed guarantees that the value of $Q(\Sigma_p)$ does not change under smooth deformations of the surface. Depending on circumstances, (15.4) represents the topological charge of a $(d-p-1)$ -dimensional defect enclosed by Σ_p , or that of a $(d-p)$ -dimensional soliton intersecting Σ_p .

Example 15.4

The coset space $G/H \simeq \text{U}(1) \simeq S^1$ of superfluids has nontrivial first de Rham cohomology group, $H^1(G/H) \simeq \mathbb{R}$. Pulling its generator back to space and integrating over a closed loop Σ_1 defines the winding number of a $(d-2)$ -dimensional defect: the vortex. The coset space $G/H \simeq \text{SU}(2)/\text{U}(1) \simeq S^2$ of ferro- and antiferromagnets has nontrivial second de Rham cohomology group, $H^2(G/H) \simeq \mathbb{R}$. Pulling its generator back to \mathbb{R}^2 and integrating over the entire plane, $\Sigma_2 = \mathbb{R}^2$, defines the charge of a 0-dimensional soliton: the baby skyrmion.

All the conservation laws of the type (15.3) discussed until now were *emergent*. Namely, their existence follows from the topology of the coset space G/H , which itself parameterizes the breaking of the symmetry we started with. There are however also other, important examples of generalized conservation laws. For instance, the Maxwell equations in absence of matter take the form of a rank-two conservation law, $\partial_\mu F^{\mu\nu} = 0$. In this case, the charge (15.4) obtained by integration over a closed spatial surface is nothing but the electric flux through that surface. The “defect” is a localized electric charge enclosed by the surface. Intriguingly, the spontaneous breakdown of the generalized symmetry of electrodynamics is the ultimate reason why the photon is massless. This brings us back to the beginning of the book: we are constantly surrounded by a sea of NG bosons, namely sound and light. A reader wondering whether the classification of NG bosons into type-A and type-B also applies to generalized symmetries will be glad to hear that this is indeed the case. See [33] for a discussion of the classification and counting of NG bosons of

generalized symmetries, and [34] for a nontrivial example of a system where the photon coupled to a particular type of matter behaves as a type-B NG mode.

References

1. P. Gerber, H. Leutwyler, Nucl. Phys. **B321**, 387 (1989). DOI [10.1016/0550-3213\(89\)90349-0](https://doi.org/10.1016/0550-3213(89)90349-0)
2. C.P. Hofmann, Phys. Rev. **B65**, 094430 (2002). DOI [10.1103/PhysRevB.65.094430](https://doi.org/10.1103/PhysRevB.65.094430)
3. J.I. Kapusta, C. Gale, *Finite-temperature field theory: Principles and applications* (Cambridge University Press, Cambridge, UK, 2006)
4. F. Strocchi, *Symmetry Breaking*. Theoretical and Mathematical Physics (Springer Nature Switzerland AG, 2021). DOI [10.1007/978-3-662-62166-0](https://doi.org/10.1007/978-3-662-62166-0)
5. T. Hayata, Y. Hidaka, Phys. Rev. **D91**, 056006 (2015). DOI [10.1103/PhysRevD.91.056006](https://doi.org/10.1103/PhysRevD.91.056006)
6. M.J. Landry, JHEP **07**, 200 (2020). DOI [10.1007/JHEP07\(2020\)200](https://doi.org/10.1007/JHEP07(2020)200)
7. M. Hongo, S. Kim, T. Noumi, A. Ota, Phys. Rev. **D103**, 056020 (2021). DOI [10.1103/PhysRevD.103.056020](https://doi.org/10.1103/PhysRevD.103.056020)
8. C.O. Akyuz, G. Goon, R. Penco, [arXiv:2306.17232](https://arxiv.org/abs/2306.17232)
9. S.R. Coleman, Commun. Math. Phys. **31**, 259 (1973). DOI [10.1007/BF01646487](https://doi.org/10.1007/BF01646487)
10. T. Griffin, K.T. Grosvenor, P. Hořava, Z. Yan, Commun. Math. Phys. **340**, 985 (2015). DOI [10.1007/s00220-015-2461-2](https://doi.org/10.1007/s00220-015-2461-2)
11. T. Griffin, K.T. Grosvenor, P. Hořava, Z. Yan, Phys. Rev. Lett. **115**, 241601 (2015). DOI [10.1103/PhysRevLett.115.241601](https://doi.org/10.1103/PhysRevLett.115.241601)
12. R. Argurio, D. Naegels, A. Pasternak, Phys. Rev. **D100**, 066002 (2019). DOI [10.1103/PhysRevD.100.066002](https://doi.org/10.1103/PhysRevD.100.066002)
13. P.C. Hohenberg, Phys. Rev. **158**, 383 (1967)
14. N.D. Mermin, H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966)
15. A. Gelfert, W. Nolting, J. Phys.: Condens. Matter **13**, R505 (2001). DOI [10.1088/0953-8984/13/27/201](https://doi.org/10.1088/0953-8984/13/27/201)
16. P.G. de Gennes, J. Prost, *The Physics of Liquid Crystals* (Clarendon Press, Oxford, UK, 1993)
17. G. Baym, B.L. Friman, G. Grinstein, Nucl. Phys. **B210**, 193 (1982). DOI [10.1016/0550-3213\(82\)90239-5](https://doi.org/10.1016/0550-3213(82)90239-5)
18. F. Wilczek, Phys. Rev. Lett. **109**, 160401 (2012). DOI [10.1103/PhysRevLett.109.160401](https://doi.org/10.1103/PhysRevLett.109.160401)
19. A. Shapere, F. Wilczek, Phys. Rev. Lett. **109**, 160402 (2012). DOI [10.1103/PhysRevLett.109.160402](https://doi.org/10.1103/PhysRevLett.109.160402)
20. P. Bruno, Phys. Rev. Lett. **111**, 070402 (2013). DOI [10.1103/PhysRevLett.111.070402](https://doi.org/10.1103/PhysRevLett.111.070402)
21. H. Watanabe, M. Oshikawa, Phys. Rev. Lett. **114**, 251603 (2015). DOI [10.1103/PhysRevLett.114.251603](https://doi.org/10.1103/PhysRevLett.114.251603)
22. M.P. Zaletel, M. Lukin, C. Monroe, C. Nayak, F. Wilczek, N.Y. Yao, Rev. Mod. Phys. **95**, 031001 (2023). DOI [10.1103/RevModPhys.95.031001](https://doi.org/10.1103/RevModPhys.95.031001)
23. C. Vafa, E. Witten, Nucl. Phys. **B234**, 173 (1984). DOI [10.1016/0550-3213\(84\)90230-X](https://doi.org/10.1016/0550-3213(84)90230-X)
24. S. Nussinov, M.A. Lampert, Phys. Rept. **362**, 193 (2002). DOI [10.1016/S0370-1573\(01\)00091-6](https://doi.org/10.1016/S0370-1573(01)00091-6)
25. N.D. Mermin, Rev. Mod. Phys. **51**, 591 (1979). DOI [10.1103/RevModPhys.51.591](https://doi.org/10.1103/RevModPhys.51.591)
26. T. Needham, *Visual Complex Analysis* (Oxford University Press, Oxford, UK, 1997)
27. H. Watanabe, H. Murayama, Phys. Rev. **X4**, 031057 (2014). DOI [10.1103/PhysRevX.4.031057](https://doi.org/10.1103/PhysRevX.4.031057)
28. J. Davighi, B. Gripaios, JHEP **09**, 155 (2018). DOI [10.1007/JHEP09\(2018\)155](https://doi.org/10.1007/JHEP09(2018)155)
29. A. Altland, B. Simons, *Condensed matter field theory* (Cambridge University Press, Cambridge, UK, 2010)
30. P.R.S. Gomes, SciPost Phys. Lect. Notes **74** (2023). DOI [10.21468/SciPostPhysLectNotes.74](https://doi.org/10.21468/SciPostPhysLectNotes.74)
31. T.D. Brennan, S. Hong, [arXiv:2306.00912](https://arxiv.org/abs/2306.00912)
32. L. Bhardwaj, L.E. Bottini, L. Fraser-Taliente, L. Gladden, D.S.W. Gould, A. Plattschorre, H. Tillim, Phys. Rept. **1051**, 1 (2024). DOI [10.1016/j.physrep.2023.11.002](https://doi.org/10.1016/j.physrep.2023.11.002)
33. Y. Hidaka, Y. Hirono, R. Yokokura, Phys. Rev. Lett. **126**, 071601 (2021). DOI [10.1103/PhysRevLett.126.071601](https://doi.org/10.1103/PhysRevLett.126.071601)
34. N. Sogabe, N. Yamamoto, Phys. Rev. **D99**, 125003 (2019). DOI [10.1103/PhysRevD.99.125003](https://doi.org/10.1103/PhysRevD.99.125003)

Chapter 16

Some Open Questions

I will close the book by highlighting some technical issues that I could not have addressed properly. I label them as “open questions,” at the risk of revealing my own ignorance rather than a gap in the existing knowledge. Since the book revolves around developing a *formalism* for spontaneously broken symmetry, some of the items are purely mathematical. The first commandment of *effective field theory* (EFT) dictates that the effective Lagrangian must contain all operators that respect the symmetry of the system. Accordingly, an important ingredient of the EFT formalism is the classification of all possible Lagrangians consistent with the symmetry. This requirement lies behind most of the issues listed below.

Universality of the Standard Nonlinear Realization The classification of group actions on a manifold in Chap. 7 relies heavily on the assumption that the isotropy subgroup H of the symmetry group G is compact. Any progress towards classification of group actions such that H is noncompact would be most welcome. This applies in particular to coset spaces G/H that are not reductive, for which many of the nice features of our standard nonlinear realization are lost.

Global Existence of the Group Action The standard nonlinear realization of symmetry developed in Chap. 7 is restricted to a single local coordinate patch. Accordingly, the explicitly constructed action of the group G is limited to its elements near unity. Yet, the global existence of the group action on the whole manifold \mathcal{M} , or coset space G/H , has been implicitly assumed throughout the book. A better understanding of possible topological obstructions to extending the group action from a local coordinate patch to the entire manifold would be desirable. This issue is particularly pressing for Lie groups consisting of several connected components. As far as I know, the first attempt to systematically deal with such cases was made in [1]. However, more work is needed to establish a general, practically useful formalism.

Group Action via Generalized Local Transformations Most of the book focuses on symmetries realized by point transformations on a target space \mathcal{M} or its Cartesian product with the spacetime M . This allows one to use the mathematical language of group actions on a finite-dimensional manifold. However, we have also seen some

physically relevant examples of generalized local symmetries, where the transformation of fields is allowed to depend on their derivatives. This suggests working directly in the infinite-dimensional space of fields as maps $M \rightarrow \mathcal{M}$. With the knowledge of the symmetry group and the symmetry-breaking pattern, one may then still apply the *agnostic nonlinear realization*, mentioned in Chaps. 12 and 13. To what extent this exhausts all possible actions of the symmetry remains unclear. Yet, a setup that makes the implementation of generalized local symmetries systematic has long been in use in the theory of differential equations [2]. First attempts to apply this setup to EFT appeared very recently [3, 4]. Hopefully, they will help to place the treatment of generalized local symmetries on the same footing as that of point symmetries.

Interplay of Nonlinear Realization and Inverse Higgs Constraints As stressed in Chap. 13, the operational way of eliminating would-be *Nambu–Goldstone* (NG) fields that excite gapped modes using an *inverse Higgs constraint* (IHC) is optional. The EFT for solely genuine NG degrees of freedom, obtained by applying an IHC, may carry a realization of the symmetry by generalized local transformations. On the other hand, the EFT before the IHC is imposed is ambiguous in that its field content may depend on the choice of order parameter. Either way, the universality of the nonlinear realization of the symmetry may be compromised. Here, too, more work is needed to establish that the existing formalism(s) for spontaneously broken spacetime symmetry give(s) the most general effective Lagrangian.

Nonlinear Realization of Translation Symmetry Possibly the largest gap in the narrative of the book is the treatment of spontaneously broken translation symmetry. The main challenge is to establish a unique local parameterization of given fields in terms of NG variables, one for each broken translation generator. Thus, the discussion in Chap. 13 is limited to order parameters spatially modulated just in one direction. A generalization of the formalism to ordered states of matter, modulated in several dimensions, is pending. The same obstacle hinders the application of the background gauge approach, which proved extremely efficient in case of internal symmetries (Chap. 8). Some concrete applications of this approach beyond the single example worked out in Chap. 13 can be found in [5]. However, a fully general background gauge formalism for construction of EFTs for spontaneously broken spacetime symmetry is not available as yet.

References

1. I. Kharuk, Phys. Rev. **D98**, 025006 (2018). DOI [10.1103/PhysRevD.98.025006](https://doi.org/10.1103/PhysRevD.98.025006)
2. P.J. Olver, *Applications of Lie Groups to Differential Equations* (Springer-Verlag, New York, 1986)
3. N. Craig, Y.T. Lee, Phys. Rev. Lett. **132**, 061602 (2024). DOI [10.1103/PhysRevLett.132.061602](https://doi.org/10.1103/PhysRevLett.132.061602)
4. M. Alminawi, I. Brivio, J. Davighi, [arXiv:2308.00017](https://arxiv.org/abs/2308.00017)
5. Y. Hidaka, T. Noumi, G. Shiu, Phys. Rev. **D92**, 045020 (2015). DOI [10.1103/PhysRevD.92.045020](https://doi.org/10.1103/PhysRevD.92.045020)

Appendix A

Elements of Differential Geometry

Most of the book relies on mathematical background familiar to any graduate student of theoretical physics, that is mostly advanced calculus and basic group theory. However, some topics are greatly illuminated by taking a more geometric viewpoint. For instance, differential calculus on manifolds and Riemannian geometry provide useful insight into the nature of coset spaces on which effective field theories for spontaneously broken symmetry are defined. This is discussed in Chap. 7. Likewise, the classification of effective theories for broken internal symmetries worked out in Chap. 8 is greatly simplified by using differential calculus on manifolds and elements of de Rham cohomology. Last but not least, Riemannian geometry provides a practically convenient language for the analysis of scattering of Nambu–Goldstone bosons in Chap. 10.

The purpose of this appendix is to give the reader the background needed to follow such more advanced parts of the book. I try to keep the discussion elementary, focusing on qualitative understanding rather than on mathematical rigor. Wherever possible, I justify the need for new concepts and motivate their formal definition. This necessarily makes the exposition biased. My approach might be natural for someone familiar with quantum field theory who seeks to develop advanced calculus methods to deal with fields that live on a manifold. The appendix can however be studied independently without any reference to the main text of the book. A more complete introduction to differential geometry at a similar level as here can be found in [1]. A reader looking for a more thorough treatment of the subject at a level still accessible to a graduate student of theoretical physics is advised to consult [2–4].

A.1 Smooth Manifolds

The basic concept of differential geometry is that of a *manifold*. This is an abstraction that arose historically from the study of non-Euclidean geometry and of the geometry of surfaces. Formally, a manifold \mathcal{M} of dimension n is a set equipped with the notion of closeness (topology) that can be locally mapped to \mathbb{R}^n in a one-to-one manner.

More precisely, one assumes the existence of a collection of pairs, $\{(U_i, \varphi_i)\}$, where U_i is an open subset of \mathcal{M} and $\varphi_i : U_i \rightarrow \mathbb{R}^n$ is a *homeomorphism*. (The latter is defined as a continuous map that is invertible and whose inverse is also continuous.) Borrowing terminology from cartography, each pair (U_i, φ_i) is called a *chart*, and the collection of all charts is an *atlas*. It is required that each point $x \in \mathcal{M}$ belongs to at least one of the sets U_i . The point $\varphi_i(x) \in \mathbb{R}^n$ corresponds to an n -tuple of real numbers x^a , $a = 1, \dots, n$, called the *coordinates* of x . It is common if a bit sloppy to identify the point $x \in \mathcal{M}$ with its image $\varphi_i(x)$ or even the coordinates x^a , and I will often follow this convention.

Mapping a point $x \in \mathcal{M}$ to its coordinates is both a blessing and a curse. On the one hand, the coordinates allow us to import to the manifold mathematical structures we are familiar with from ordinary calculus. On the other hand, it may not be possible to cover the whole manifold with a single set U_i . The coordinates are in general only defined locally in some neighborhood of x . We must therefore be able to switch from one set of coordinates to another. The first commandment of differential geometry is that whatever structure we build on the manifold, it must have an intrinsic meaning regardless of a particular choice of local coordinates.

Example A.1

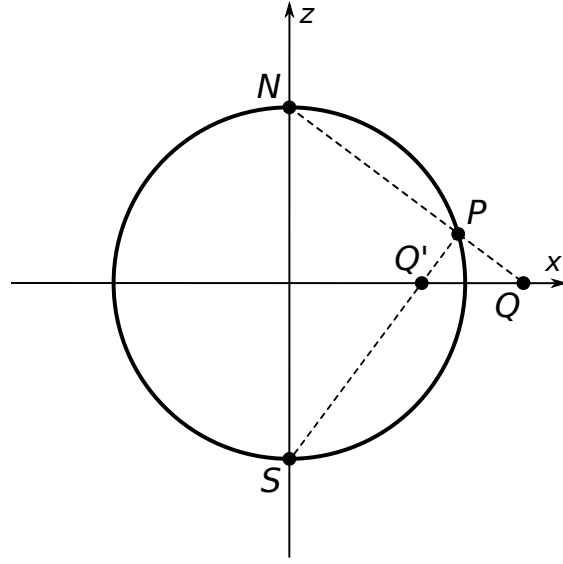
The $\mathcal{M} = \mathbb{R}^2$ is itself an n -dimensional manifold. It can be covered with a single coordinate chart, $(U, \varphi) = (\mathbb{R}^2, \text{id})$, corresponding to usual Cartesian coordinates. The coordinates can however also be chosen differently, in which case a single chart may not be sufficient to cover the whole \mathbb{R}^2 . For example, let us consider \mathbb{R}^2 and use the polar coordinates ϱ, θ instead of the Cartesian coordinates x, y . The map

$$(x, y) \rightarrow (\varrho, \theta) \equiv \left(\sqrt{x^2 + y^2}, \text{sgn } y \arccos \frac{x}{\sqrt{x^2 + y^2}} \right) \quad (\text{A.1})$$

is continuous and invertible on $U = \mathbb{R}^2 \setminus \{(x, 0) \mid x \leq 0\}$, making U homeomorphic to the domain $(0, +\infty) \times (-\pi, +\pi) \subset \mathbb{R}^2$. To cover the whole plane $\mathcal{M} = \mathbb{R}^2$, we would need to add at least one more chart.

The first thing we can do with the local coordinates is to introduce a differential structure on \mathcal{M} . Let us choose a chart (U_i, φ_i) . We then define a *smooth* real function on U_i as a map $f : U_i \rightarrow \mathbb{R}$ such that $f \circ \varphi_i^{-1}$ is an infinitely differentiable function on $\varphi_i(U_i) \subset \mathbb{R}^n$. We would of course like the concept of smoothness to be independent of the choice of chart. This is ensured by requiring that for any two overlapping open sets U_i, U_j , the map $\varphi_i \circ \varphi_j^{-1}$ that switches between the two coordinate systems is itself infinitely differentiable. A function defined on the whole manifold, $f : \mathcal{M} \rightarrow \mathbb{R}$, is now called smooth if it is smooth when restricted to any single coordinate chart.

Fig. A.1 Stereographic projection of the n -sphere S^n to \mathbb{R}^n , here for $n = 1$. First, S^n is naturally embedded in \mathbb{R}^{n+1} as a unit sphere centered at the origin. A point P on S^n is then projected to the point Q in \mathbb{R}^n using a straight line starting at the north pole N of the sphere. An alternative projection from the south pole S would map P to Q'



Example A.2

One way to define local coordinates on the n -sphere S^n is through stereographic projection, see Fig. A.1. We imagine that S^n is realized (embedded) in \mathbb{R}^{n+1} as a unit sphere centered at the origin. The embedding is defined implicitly by $(x^1)^2 + \dots + (x^n)^2 + z^2 = 1$, where (x^1, \dots, x^n, z) are Cartesian coordinates in \mathbb{R}^{n+1} . A given point $P \in S^n$ is then projected from the north pole N to the hyperplane \mathbb{R}^n defined by $z = 0$. The image of P is the point

$$Q \equiv (y^1, \dots, y^n) = \left(\frac{x^1}{1-z}, \dots, \frac{x^n}{1-z} \right). \quad (\text{A.2})$$

The set (y^1, \dots, y^n) constitutes the desired local coordinates on S^n . These coordinates are well-defined on the whole sphere except for the north pole N itself. To cover the entire sphere, we need another coordinate patch. This can be defined through a similar projection from the south pole S . The latter maps P to

$$Q' \equiv (y'^1, \dots, y'^n) = \left(\frac{x^1}{1+z}, \dots, \frac{x^n}{1+z} \right). \quad (\text{A.3})$$

Noting that

$$|y| = \frac{|x|}{1-z} = \sqrt{\frac{1+z}{1-z}}, \quad |y'| = \frac{|x|}{1+z} = \sqrt{\frac{1-z}{1+z}}, \quad (\text{A.4})$$

we find that the two sets of local coordinates are related by inversion, $y'^a = y^a/|y|^2$. This ensures smooth transition between the two coordinate systems wherever they are both well-defined, that is on S^n with both poles removed.

A.2 Linear Structures on Manifolds

We have now learned what a smooth function on a manifold is. Since our primary goal is to develop calculus on manifolds, we would eventually like to differentiate and integrate such functions. In order to be able to do so independently of the choice of local coordinates, we will however need additional structure on the manifold.

A.2.1 Tangent Vectors and Vector Fields

The first step towards setting up calculus on manifolds is to define a derivative of a function. In \mathbb{R}^n with its Cartesian coordinates, this naturally leads to the notion of a partial derivative. On a general manifold \mathcal{M} , there is however no such preferred set of coordinates. The way out is to generalize the concept of a directional derivative in \mathbb{R}^n . Thus, for a fixed point $x \in U_i \subset \mathcal{M}$ and a set of coefficients v^1, \dots, v^n , we define a linear operator \mathbf{v} acting on smooth functions f on U_i through

$$\mathbf{v} = v^a \frac{\partial}{\partial x^a} : f \rightarrow \mathbf{v}[f] \equiv v^a \left. \frac{\partial(f \circ \varphi_i^{-1})}{\partial x^a} \right|_x. \quad (\text{A.5})$$

Such a linear operator is usually called a *tangent vector* to \mathcal{M} at x . Accordingly, the set of all tangent vectors at x is called the *tangent space* of \mathcal{M} at x and denoted as $\mathcal{T}_x \mathcal{M}$. Let me stress that the adjective “tangent” here is just a convention. There is no way to give it a literal geometric meaning unless we imagine \mathcal{M} as a hypersurface embedded in a higher-dimensional Euclidean space.

The definition (A.5) relies on the components v^a in a fixed *coordinate basis* of the tangent space, $\partial_a \equiv \partial/\partial x^a$. However, the concept of a tangent vector has a meaning independent of a specific choice of local coordinates. We just have to remember to recalculate the components of the vector if the local coordinates change, as the following example illustrates.

Example A.3

Suppose that we trade the Cartesian coordinates x, y in \mathbb{R}^2 for the polar coordinates ϱ, θ through the usual prescription $(x, y) = (\varrho \cos \theta, \varrho \sin \theta)$. The coordinate basis of the tangent space is then converted by using the chain rule for partial derivatives,

$$\partial_\varrho = (\cos \theta) \partial_x + (\sin \theta) \partial_y, \quad \partial_\theta = -(\varrho \sin \theta) \partial_x + (\varrho \cos \theta) \partial_y. \quad (\text{A.6})$$

In order for a vector \mathbf{v} to be well-defined independently of the choice of local coordinates, its components (v^x, v^y) and (v^ϱ, v^θ) in the two coordinate systems must be related by

$$v^\varrho = v^x \cos \theta + v^y \sin \theta, \quad v^\theta = -\frac{v^x}{\varrho} \sin \theta + \frac{v^y}{\varrho} \cos \theta. \quad (\text{A.7})$$

These results agree with what elementary planar geometry would tell us if we treated ∂_ϱ and $(1/\varrho)\partial_\theta$ as unit vectors in the radial and azimuthal direction, respectively.

In general, if we change the local coordinates x^a smoothly to $\tilde{x}^a(x)$, the coordinate bases and vector components in the new and old coordinate systems are related by

$$\tilde{\partial}_a = \frac{\partial x^b}{\partial \tilde{x}^a} \partial_b, \quad \tilde{v}^a = \frac{\partial \tilde{x}^a}{\partial x^b} v^b. \quad (\text{A.8})$$

The first of these is just the chain rule, whereas the latter follows from the requirement that the vector $\mathbf{v} = v^a \partial_a = \tilde{v}^a \tilde{\partial}_a$ is independent of the choice of coordinates. Note that $\partial \tilde{x}^a / \partial x^b$ is just a shorthand notation. If x^a and \tilde{x}^a correspond to coordinate maps φ_i, φ_j , this notation represents the partial derivatives of $\varphi_j \circ \varphi_i^{-1}$, evaluated at $\varphi_i(x) \in \mathbb{R}^n$. In the following, I will occasionally use such intuitive notation without further warnings.

For most applications in physics, a *vector field* is of greater interest than a (tangent) vector defined at a single point $x \in \mathcal{M}$. Just like for real functions on \mathcal{M} , it is natural to require that a vector field varies smoothly as a function of x . This is most easily ensured by demanding that, within a local coordinate patch, the components v^a are themselves smooth functions of x . It is also possible to define a vector field without referring to specific coordinates. One thus treats it as a collection of vectors $\mathbf{v}(x) \in \mathcal{T}_x \mathcal{M}$ such that $\mathbf{v}(x)[f]$ is a smooth function of x for any smooth “test function” f on \mathcal{M} .

With the notion of a vector field at hand, let us briefly return to the choice of basis of the tangent space. With local coordinates x^a defined on an open set U_i , we get a set of coordinate basis fields ∂_a that are likewise well-defined only on U_i . It is however often advantageous to have a basis that is not limited to a particular coordinate patch, or not related to a specific set of coordinates at all. One thus defines a (*local*) *frame* as a set of vector fields \mathbf{e}_A , $A = 1, \dots, n$, defined on some open set $U \subset \mathcal{M}$, such that the set of vectors $\{\mathbf{e}_A(x)\}_{A=1}^n$ is a basis of $\mathcal{T}_x \mathcal{M}$ at any $x \in U$.¹ Any vector field can then be expressed in terms of its components with respect to the frame, $\mathbf{v}(x) = v^A(x) \mathbf{e}_A(x)$. I will sometimes use a local frame instead of coordinate basis fields; the latter can always be viewed as a special case of the former.

¹ Within this appendix, I will carefully distinguish coordinate indices a, b, \dots from frame indices A, B, \dots . In practice (including the main text of this book), the two notations are often blended.

Restricting the definition of a local frame to an open subset of \mathcal{M} is not as innocuous as it might appear. The existence of a frame well-defined on the whole of \mathcal{M} depends on the global topology of \mathcal{M} . Even on a manifold as simple as the 2-sphere S^2 , there is no smooth vector field that would be nonzero everywhere! This is the “hairy ball theorem” of algebraic topology; see Sect. 13.7 of [1] for an elementary discussion.

A second remark of caution is that a local frame is not merely a coordinate basis in disguise. By the definition of a basis, one can always relate a frame \mathbf{e}_A to any coordinate basis ∂_a by $\mathbf{e}_A = c_A^a \partial_a$, where c_A^a is a matrix of coefficients with nonzero determinant. In general, there is however no set of local coordinates y^A , $A = 1, \dots, n$ such that $\mathbf{e}_A = \partial/\partial y^A$. As I will show later, this is only possible if the condition $\mathbf{e}_A[c_B^a] = \mathbf{e}_B[c_A^a]$ is satisfied for all A, B and a .

A.2.2 Tensors and Tensor Fields

Once we know how to construct tangent vectors and vector fields on a manifold, we can quickly build tensors and tensor fields of an arbitrary rank. While the motivation for doing so may not be as immediate as for tangent vectors, several types of tensors are of central importance to differential geometry. To start with, the *cotangent space* of \mathcal{M} at x , $\mathcal{T}_x^* \mathcal{M}$, is defined as the dual of $\mathcal{T}_x \mathcal{M}$. Its elements, called *cotangent vectors* or *covectors*, are linear maps from $\mathcal{T}_x \mathcal{M}$ to \mathbb{R} . The counterpart of a coordinate basis ∂_a on $\mathcal{T}_x \mathcal{M}$ is the dual coordinate basis on $\mathcal{T}_x^* \mathcal{M}$, usually denoted as dx^a . Hence $dx^a(\partial_b) = \delta_b^a$. The action of any covector on any vector then follows from linearity and can be expressed component-wise,

$$\omega = \omega_a dx^a, \quad \mathbf{v} = v^a \partial_a \Rightarrow \omega(\mathbf{v}) = \omega_a v^a, \quad \omega \in \mathcal{T}_x^* \mathcal{M}, \quad \mathbf{v} \in \mathcal{T}_x \mathcal{M}. \quad (\text{A.9})$$

The choice of notation for the dual coordinate basis is not accidental. Any smooth function f defined in a neighborhood of x gives rise to a covector df through

$$df(\mathbf{v}) \equiv \mathbf{v}[f] = v^a \frac{\partial f}{\partial x^a}, \quad \mathbf{v} \in \mathcal{T}_x \mathcal{M}. \quad (\text{A.10})$$

I have simplified the notation by treating f as a function of the local coordinates x^a ; cf. (A.5). By the definition of the dual basis, we have $v^a = dx^a(\mathbf{v})$. Hence $df = (\partial f / \partial x^a) dx^a$, generalizing the differential of a function known from ordinary calculus. It follows that upon switching the local coordinates from x^a to $\tilde{x}^a(x)$, the dual coordinate basis and components of covectors transform to

$$d\tilde{x}^a = \frac{\partial \tilde{x}^a}{\partial x^b} dx^b, \quad \tilde{\omega}_a = \frac{\partial x^b}{\partial \tilde{x}^a} \omega_b. \quad (\text{A.11})$$

Higher-rank tensors can now be built using tensor products of $\mathcal{T}_x\mathcal{M}$ and $\mathcal{T}_x^*\mathcal{M}$. Generally, a *tensor of type (q, p)* at $x \in \mathcal{M}$ is an element of $(\mathcal{T}_x\mathcal{M})^{\otimes q} \otimes (\mathcal{T}_x^*\mathcal{M})^{\otimes p}$. A class of tensors of special importance to both differential and integral calculus on manifolds is obtained by taking the p -fold antisymmetric tensor product of $\mathcal{T}_x^*\mathcal{M}$, where p is a positive integer. The elements of this space, $\Omega_x^p\mathcal{M}$, are called *p -forms* and define antisymmetric multilinear functions on $\mathcal{T}_x\mathcal{M}$. Note that $\Omega_x^1\mathcal{M}$ coincides with $\mathcal{T}_x^*\mathcal{M}$; covectors are indeed often referred to as 1-forms. An antisymmetric tensor product of p covectors $\omega_1, \dots, \omega_p$ is represented compactly by the notation

$$\omega_1 \wedge \cdots \wedge \omega_p \equiv \sum_{\pi} \text{sgn } \pi \, \omega_{\pi(1)} \otimes \cdots \otimes \omega_{\pi(p)} , \quad (\text{A.12})$$

where the sum is over all permutations π of the indices $1, \dots, p$. Using this notation, we can write a general element $\omega \in \Omega_x^p\mathcal{M}$ in the dual coordinate basis as

$$\omega = \frac{1}{p!} \omega_{a_1 \dots a_p} dx^{a_1} \wedge \cdots \wedge dx^{a_p} , \quad \omega_{a_1 \dots a_p} = \omega(\partial_{a_1}, \dots, \partial_{a_p}) . \quad (\text{A.13})$$

The expression for the components of the p -form is a direct generalization of the relation $\omega_a = \omega(\partial_a)$ for 1-forms.

Equation (A.12) is more than just a useful shorthand notation. It can be extended to an operation on forms of arbitrary degree that is linear and associative: the *exterior product*. With the dual coordinate basis expansion (A.13) at hand, linearity and associativity require that any two forms $\omega \in \Omega_x^p\mathcal{M}$ and $\sigma \in \Omega_x^q\mathcal{M}$ are mapped to the $(p+q)$ -form

$$\omega \wedge \sigma = \frac{1}{p!q!} \omega_{a_1 \dots a_p} \sigma_{b_1 \dots b_q} dx^{a_1} \wedge \cdots \wedge dx^{a_p} \wedge dx^{b_1} \wedge \cdots \wedge dx^{b_q} . \quad (\text{A.14})$$

The same result can be alternatively expressed in a coordinate-independent manner by the action on a set of test vectors $\mathbf{v}_1, \dots, \mathbf{v}_{p+q}$ from $\mathcal{T}_x\mathcal{M}$,

$$\begin{aligned} & (\omega \wedge \sigma)(\mathbf{v}_1, \dots, \mathbf{v}_{p+q}) \\ &= \frac{1}{p!q!} \sum_{\pi} \text{sgn } \pi \, \omega(\mathbf{v}_{\pi(1)}, \dots, \mathbf{v}_{\pi(p)}) \sigma(\mathbf{v}_{\pi(p+1)}, \dots, \mathbf{v}_{\pi(p+q)}) , \end{aligned} \quad (\text{A.15})$$

where π now runs over permutations of $1, \dots, p+q$. The antisymmetry of forms as multilinear maps implies that the exterior product is graded-anticommutative,

$$\omega \wedge \sigma = (-1)^{pq} \sigma \wedge \omega , \quad \omega \in \Omega_x^p\mathcal{M} , \, \sigma \in \Omega_x^q\mathcal{M} . \quad (\text{A.16})$$

The direct sum $\bigoplus_{p=0}^n \Omega_x^p\mathcal{M}$ equipped with the exterior product possesses the structure of an associative algebra: the *exterior (Grassmann) algebra*. Here $\Omega_x^0\mathcal{M}$ is identified with \mathbb{R} and the direct sum terminates at $p = n$; by the antisymmetry property there are no p -forms with $p > n$ on an n -dimensional manifold.

Another operation on forms that will prove useful is the *interior product*. This assigns to any vector $\mathbf{v} \in \mathcal{T}_x\mathcal{M}$ a linear map $\iota_{\mathbf{v}} : \Omega_x^{p+1}\mathcal{M} \rightarrow \Omega_x^p\mathcal{M}$ defined by

$$\iota_v \omega(v_1, \dots, v_p) \equiv \omega(v, v_1, \dots, v_p), \quad v_1, \dots, v_p \in \mathcal{T}_x \mathcal{M} \quad (\text{A.17})$$

for any $\omega \in \Omega_x^{p+1} \mathcal{M}$. While this definition is coordinate-independent, the interior product takes a particularly simple form in the component notation where it amounts to contracting the vector v with the $(p+1)$ -form ω , $(\iota_v \omega)_{a_1 \dots a_p} = v^b \omega_{b a_1 \dots a_p}$. Thanks to the antisymmetry of forms, the interior product is anticommutative that is $\iota_u \circ \iota_v = -\iota_v \circ \iota_u$ for any two tangent vectors u, v . Moreover, ι_v with any fixed v is an *antiderivation* of the Grassmann algebra: for any p -form ω and q -form σ one has

$$\iota_v (\omega \wedge \sigma) = (\iota_v \omega) \wedge \sigma + (-1)^p \omega \wedge (\iota_v \sigma). \quad (\text{A.18})$$

In fact, the properties of the interior product are already completely fixed by requiring that it is an antiderivation that reduces to $\iota_v \omega = \omega(v)$ on 1-forms ω .

So far I have limited the discussion in this subsection to tensors defined at a single point $x \in \mathcal{M}$. However, we can build *tensor fields* of arbitrary type (q, p) in complete analogy with the construction of vector fields out of tangent vectors at different points.

Example A.4

A *metric* g is a symmetric tensor field of type $(0, 2)$ such that $g(x) \in \mathcal{T}_x^* \mathcal{M} \odot \mathcal{T}_x^* \mathcal{M}$ is a positive-definite bilinear form for any $x \in \mathcal{M}$. For example, \mathbb{R}^n as a Euclidean space is equipped with the metric $g = \delta_{ab} dx^a \otimes dx^b$ where x^a are the Cartesian coordinates. The metric can of course be converted to any other coordinates by using (A.11). For instance, in \mathbb{R}^3 we may want to switch to the spherical coordinates (r, θ, φ) , defined by $(x, y, z) = (r \sin \theta \cos \varphi, r \sin \theta \sin \varphi, r \cos \theta)$. The metric then assumes the form

$$g = dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2 \theta d\varphi \otimes d\varphi. \quad (\text{A.19})$$

An interested reader will find more about the metric tensor in Sect. A.6.

Promoting a p -form to a tensor field leads to the concept of *differential p -form*.² Just like for vector fields, we have the freedom to choose other bases for differential p -forms than those induced by local coordinates on \mathcal{M} . Thus, a (*local*) *coframe* is a set of differential 1-forms e^{*A} defined on some open set $U \in \mathcal{M}$ such that the set of 1-forms $\{e^{*A}(x)\}_{A=1}^n$ spans a basis of $\Omega_x^1 \mathcal{M}$ at any $x \in U$. Any frame e_A induces a unique dual coframe e^{*A} (and vice versa) by requiring that $e^{*A}(e_B) = \delta_B^A$. All the properties of p -forms reviewed above can be immediately promoted to frame components of differential p -forms by the replacements $\partial_a \rightarrow e_A$, $dx^a \rightarrow e^{*A}$ and $\omega_{a_1 \dots a_p} \rightarrow \omega_{A_1 \dots A_p}$ wherever necessary. The same remark applies to tensor fields of an arbitrary type (q, p) .

² For the sake of brevity, I will mostly drop the adjective “differential” where it is clear from the context that I speak of a tensor field rather than a tensor defined at a single point.

Example A.5

Let us use the spherical coordinates (θ, φ) to parameterize the 2-sphere S^2 . Following Example A.4, this allows us to define a metric on S^2 by restricting (A.19) to fixed radial variable, $r = 1$,

$$g = d\theta \otimes d\theta + \sin^2 \theta d\varphi \otimes d\varphi . \quad (\text{A.20})$$

The vector fields

$$\mathbf{e}_\theta \equiv \partial_\theta , \quad \mathbf{e}_\varphi \equiv \frac{1}{\sin \theta} \partial_\varphi \quad (\text{A.21})$$

then constitute an orthonormal frame, $g(\mathbf{e}_A, \mathbf{e}_B) = \delta_{AB}$, wherever well-defined. The corresponding dual coframe is $\mathbf{e}^{*\theta} = d\theta$ and $\mathbf{e}^{*\varphi} = \sin \theta d\varphi$. Out of these two 1-forms, we can naturally build a 2-form through

$$\omega = \mathbf{e}^{*\theta} \wedge \mathbf{e}^{*\varphi} = \sin \theta d\theta \wedge d\varphi . \quad (\text{A.22})$$

One might be tempted to think of the frame $(\mathbf{e}_\theta, \mathbf{e}_\varphi)$ as defining an area element on the sphere that can be subsequently used for integration. However, as I will explain in Sect. A.7, integration on manifolds is done using differential forms. For the 2-sphere, it is exactly the differential 2-form (A.22) that does the job of defining the proper area element.

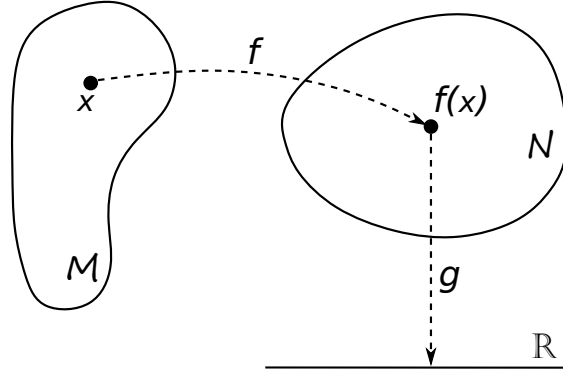
A.3 Maps Between and on Manifolds

In Sect. A.1 we saw how to define a smooth function on a manifold \mathcal{M} as a map from \mathcal{M} to \mathbb{R} . This concept of smoothness is easily generalized to maps between two manifolds \mathcal{M} and \mathcal{N} . For a map $f : \mathcal{M} \rightarrow \mathcal{N}$ that is continuous, it is always possible to choose a chart (U, φ) on \mathcal{M} and a chart (V, χ) on \mathcal{N} such that $f(U) \subset V$. The map is then called smooth if $\chi \circ f \circ \varphi^{-1}$ is infinitely differentiable. A smooth map $f : \mathcal{M} \rightarrow \mathcal{N}$ that is a homeomorphism is called a *diffeomorphism*.

A.3.1 Push-Forward and Pull-Back

Smooth maps between manifolds naturally induce linear maps between the corresponding (co)tangent spaces. Since tangent vectors are differential operators acting on locally defined test functions, we first have to inspect how the map $f : \mathcal{M} \rightarrow \mathcal{N}$ affects the latter. It is clear from Fig. A.2 that by “connecting the arrows,” any test function $g : \mathcal{N} \rightarrow \mathbb{R}$ can be converted into the function $g \circ f : \mathcal{M} \rightarrow \mathbb{R}$. This is enough to connect tangent vectors on the two manifolds. Formally, one starts with a tangent vector $\mathbf{v} \in \mathcal{T}_x \mathcal{M}$ and maps it to a tangent vector $f_* \mathbf{v} \in \mathcal{T}_{f(x)} \mathcal{N}$ such that

Fig. A.2 A smooth map f between manifolds \mathcal{M} and \mathcal{N} , defined locally in the neighborhood of the point $x \in \mathcal{M}$. Composing a test function g , defined locally around the image $f(x) \in \mathcal{N}$, with f allows one to lift the map f to the push-forward $f_* : \mathcal{T}_x \mathcal{M} \rightarrow \mathcal{T}_{f(x)} \mathcal{N}$



$$f_* \mathbf{v}[g] \equiv \mathbf{v}[g \circ f] \quad (\text{A.23})$$

for any test function g defined in a neighborhood of $f(x)$ on \mathcal{N} . The image $f_* \mathbf{v}$ is referred to as the *push-forward* of \mathbf{v} by f . Suppose that we have a set of coordinates x^a in a neighborhood of x on \mathcal{M} and a set of coordinates y^b in a neighborhood of $f(x)$ on \mathcal{N} . It then follows from the chain rule that the coordinate basis vectors at x are mapped to

$$f_* \frac{\partial}{\partial x^a} = \frac{\partial f^b}{\partial x^a} \frac{\partial}{\partial y^b}, \quad (\text{A.24})$$

where f^b is a shorthand notation for the b -th coordinate of $f(x)$. This looks almost like a change of variables. It is therefore worth stressing that the manifolds \mathcal{M} and \mathcal{N} need not have the same dimension, hence the map f need not be invertible.

Example A.6

Consider again the 2-sphere with its spherical coordinates (θ, φ) and define the map $f : S^2 \rightarrow \mathbb{R}^3$ as the natural embedding that takes the point $(\theta, \varphi) \in S^2$ to $(x, y, z) = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta) \in \mathbb{R}^3$. Then (A.24) gives us

$$\begin{aligned} f_* \partial_\theta &= (\cos \theta \cos \varphi) \partial_x + (\cos \theta \sin \varphi) \partial_y - (\sin \theta) \partial_z, \\ f_* \partial_\varphi &= -(\sin \theta \sin \varphi) \partial_x + (\sin \theta \cos \varphi) \partial_y. \end{aligned} \quad (\text{A.25})$$

This “realizes” tangent vectors to the sphere as vectors in \mathbb{R}^3 . The same strategy can be used to give a literal geometric interpretation to tangent vectors on any smooth manifold by embedding the latter in a higher-dimensional Euclidean space.

Once we know how to relate tangent vectors to \mathcal{M} and \mathcal{N} , we can do the same for covectors by using the dual nature of the cotangent space. The “connecting the arrows” strategy hints that the map between the cotangent spaces acts in the opposite direction than f_* . Thus, we assign to $\omega \in \mathcal{T}_{f(x)}^* \mathcal{N}$ a covector $f^* \omega \in \mathcal{T}_x^* \mathcal{M}$ so that

$$f^* \omega(\mathbf{v}) \equiv \omega(f_* \mathbf{v}) \quad \text{for any } \mathbf{v} \in \mathcal{T}_x \mathcal{M}. \quad (\text{A.26})$$

For obvious reasons, the image $f^* \omega$ is referred to as the *pull-back* of ω by f . The chain rule tells us that covectors from the dual coordinate basis are mapped to

$$f^* dy^b = \frac{\partial f^b}{\partial x^a} dx^a, \quad (\text{A.27})$$

which again looks like a mere change of variables but is really much more than that.

By following the same line of reasoning, one can define the push-forward f_* for any tensor of type $(q, 0)$ and the pull-back f^* for any tensor of type $(0, p)$. In general, it is however not possible to define either of the two for tensors of the mixed type (q, p) . In a similar vein, it is generally not possible to promote the point-wise map $f_* : \mathcal{T}_x \mathcal{M} \rightarrow \mathcal{T}_{f(x)} \mathcal{N}$ to a push-forward of vector (or tensor) *fields*. A problem arises when two points $x_1, x_2 \in \mathcal{M}$ have the same image under f . Our prescription (A.23) would then naively try to assign to a single point $f(x_1) = f(x_2)$ two different tangent vectors. No such a problem arises for the pull-back, which can be extended from tensors of type $(0, p)$ to tensor fields for any smooth map f .

Both of the above problems disappear when f is a diffeomorphism. In that case, we have a locally well-defined one-to-one mapping between \mathcal{M} and \mathcal{N} . The push-forward f_* is an isomorphism between the respective tangent spaces, and the pull-back f^* is an isomorphism between the respective cotangent spaces. By combining $f_* : \mathcal{T}_x \mathcal{M} \rightarrow \mathcal{T}_{f(x)} \mathcal{N}$ with $(f^{-1})^* : \mathcal{T}_x^* \mathcal{M} \rightarrow \mathcal{T}_{f(x)}^* \mathcal{N}$, we can map any tensor (field) on \mathcal{M} to a tensor (field) on \mathcal{N} .

Example A.7

In Example A.4, I introduced the Euclidean metric on \mathbb{R}^n , $g = \delta_{ab} dx^a \otimes dx^b$, where x^a are the Cartesian coordinates. This metric can be pulled back to any manifold \mathcal{M} that can be embedded as a hypersurface in \mathbb{R}^n by a smooth map f . The induced metric on \mathcal{M} in chosen local coordinates y^a on \mathcal{M} follows from (A.27),

$$f^* g = \delta_{ab} \frac{\partial f^a}{\partial y^c} \frac{\partial f^b}{\partial y^d} dy^c \otimes dy^d. \quad (\text{A.28})$$

This justifies the previously used expression (A.20) for the metric on the 2-sphere.

In the same manner, one can pull back any differential form defined on \mathbb{R}^n to the embedded manifold \mathcal{M} . For instance, the area 2-form (A.22) on the 2-sphere has an elegant representation in terms of the natural embedding of S^2 in \mathbb{R}^3 . It is left to the reader as an exercise to check that (A.22) is recovered by pulling back the 2-form

$$\frac{1}{2} \mathbf{r} \cdot (\partial_a \mathbf{r} \times \partial_b \mathbf{r}) dx^a \wedge dx^b \quad (\text{A.29})$$

in \mathbb{R}^3 , where $\mathbf{r} = (x, y, z)$ and the dot and cross denote respectively the conventional dot and cross products of 3-vectors. It is common to abuse the notation and represent the area 2-form on S^2 by simply replacing the 3-vector \mathbf{r} above with a unit 3-vector \mathbf{n} that defines the embedding of S^2 in \mathbb{R}^3 . One can then write the area 2-form on S^2 simply as $(1/2)\mathbf{n} \cdot (d\mathbf{n} \times d\mathbf{n})$.

A.3.2 Flow of Vector Fields

In Sect. A.2.1, I emphasized that the concept of a tangent vector does not have a literal geometric interpretation unless we embed the manifold \mathcal{M} in a higher-dimensional Euclidean space. It is however possible to visualize a vector field \mathbf{v} through the *flow* it induces on \mathcal{M} . Imagine that the manifold is filled with a liquid, and think of $\mathbf{v}(x)$ as the velocity of the liquid at $x \in \mathcal{M}$. We can then extract information about the vector field from the trajectories of individual particles of the liquid.

Let us now formalize this concept. We need a real parameter $t \in \mathbb{R}$ to measure the “time” for the liquid motion. Next, we think of a single particle trajectory as a curve $\gamma : \mathbb{R} \rightarrow \mathcal{M}$ satisfying the first-order differential equation

$$\frac{df(\gamma(t))}{dt} = \mathbf{v}[f] \Big|_{\gamma(t)} \quad (\text{A.30})$$

for any test function $f : \mathcal{M} \rightarrow \mathbb{R}$. This is just a coordinate-independent version of the relation between the particle’s trajectory and its velocity, $d\gamma^a/dt = v^a(\gamma(t))$. By choosing the initial condition as $\gamma(0) = x$ and varying $x \in \mathcal{M}$, we get a family of trajectories that define a map $\phi_{\mathbf{v},t} : \mathcal{M} \rightarrow \mathcal{M}$. In terms of the liquid analogy, $\phi_{\mathbf{v},t}(x)$ is the position at time t of a particle that was at x at time $t = 0$. For smooth vector fields \mathbf{v} , the flow equation (A.30) is guaranteed to have a (locally) unique solution. Hence the map $\phi_{\mathbf{v},t}$ with fixed \mathbf{v} and t is a diffeomorphism on \mathcal{M} . We can thus use it to push forward tensors from one point of \mathcal{M} to another. This is called *Lie transport*.

Example A.8

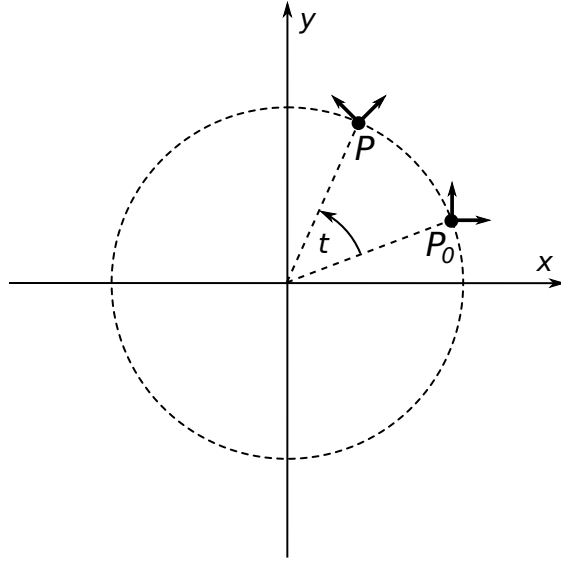
Consider the vector field $\mathbf{v} = -y\partial_x + x\partial_y$ in \mathbb{R}^2 . The flow equation (A.30) is easily solved and gives

$$x(t) = x_0 \cos t - y_0 \sin t, \quad y(t) = x_0 \sin t + y_0 \cos t, \quad (\text{A.31})$$

where $P_0 = (x_0, y_0)$ is the starting point of the flow, see Fig. A.3. From the definition (A.23) of push-forward, or directly from (A.24), we then see that the coordinate basis (∂_x, ∂_y) at P_0 is transported to

$$\phi_{\mathbf{v},t*}\partial_x = (\cos t)\partial_x + (\sin t)\partial_y, \quad \phi_{\mathbf{v},t*}\partial_y = -(\sin t)\partial_x + (\cos t)\partial_y \quad (\text{A.32})$$

Fig. A.3 Illustration of Lie transport of tangent vectors in \mathbb{R}^2 . The Lie transport is generated by the vector field $\mathbf{v} = -y\partial_x + x\partial_y$, one of whose flow lines is indicated by the *dashed circle*. As the point P_0 is transported to P , the attached basis of tangent space at P_0 (displayed as *solid arrows*) is pushed forward to a basis of tangent space at P



at $P = (x(t), y(t))$. The flow generated by \mathbf{v} corresponds to a rotation of \mathbb{R}^2 around the origin by angle t . Indeed, switching to polar coordinates (ϱ, θ) through $(x, y) = (\varrho \cos \theta, \varrho \sin \theta)$ shows that $\mathbf{v} = \partial_\theta$. Upon changing the basis of the tangent space, (A.32) is then recognized as a trivial map between the coordinate bases at P_0 and P , that is $\phi_{\mathbf{v},t*}\partial_\varrho|_{P_0} = \partial_\varrho|_P$ and $\phi_{\mathbf{v},t*}\partial_\theta|_{P_0} = \partial_\theta|_P$. This is a special case of a general rule that when \mathbf{v} is one of the coordinate basis fields, all the coordinate basis vectors at x are transported to the corresponding coordinate basis vectors at $\phi_{\mathbf{v},t}(x)$.

As another example, we may choose $\mathbf{v} = x\partial_x + y\partial_y$. In this case, we find

$$x(t) = x_0 e^t, \quad y(t) = y_0 e^t, \quad (\text{A.33})$$

and accordingly

$$\phi_{\mathbf{v},t*}\partial_x = e^t \partial_x, \quad \phi_{\mathbf{v},t*}\partial_y = e^t \partial_y. \quad (\text{A.34})$$

This is another illustration of Lie transport generated by a coordinate basis field. We just have to notice that in polar coordinates, $\mathbf{v} = \varrho \partial_\varrho = \partial_{\log \varrho}$. Choosing local coordinates as $(\log \varrho, \theta)$ reproduces (A.34) as a trivial transport of $(\partial_{\log \varrho}, \partial_\theta)$ at P_0 to $(\partial_{\log \varrho}, \partial_\theta)$ at P .

Lie transport plays a critical role for the implementation of symmetries on manifolds. Indeed, the map $\phi_{\mathbf{v},t}$ for fixed \mathbf{v} and varying t constitutes a one-parameter group of transformations on \mathcal{M} . The group structure amounts to the obvious properties

$$\phi_{\mathbf{v},t} \circ \phi_{\mathbf{v},s} = \phi_{\mathbf{v},t+s}, \quad \phi_{\mathbf{v},0} = \text{id}, \quad \phi_{\mathbf{v},t}^{-1} = \phi_{\mathbf{v},-t}. \quad (\text{A.35})$$

We can think of this as a “representation” of \mathbb{R} by an additive group of transformations on \mathcal{M} . It is straightforward to generalize this concept to other (finite or Lie) groups. Formally, the *action* of a group G on \mathcal{M} is defined as a set of diffeomorphisms T_g on \mathcal{M} that respect the group structure of G ,³

$$T_{g_1} \circ T_{g_2} = T_{g_1 g_2}, \quad T_e = \text{id}, \quad (T_g)^{-1} = T_{g^{-1}}, \quad g, g_1, g_2 \in G. \quad (\text{A.36})$$

For any Lie group G , we can restrict to a specific one-parameter subgroup. This is mapped by (A.30) to a vector field that represents the corresponding generator of G on \mathcal{M} . Combined with (A.35), we deduce that there is a one-to-one correspondence between vector fields and one-parameter groups of transformations on \mathcal{M} .

A.3.3 Lie Derivative

The practical utility of Lie transport is limited due to the necessity to solve the set of coupled first-order differential equations (A.30) for the flow. It is usually more convenient to consider the variation of tensor fields on \mathcal{M} under infinitesimal transformations generated by \mathbf{v} . This variation is formally encoded in the *Lie derivative*, defined for a given tensor field \mathbf{T} as

$$\mathcal{L}_{\mathbf{v}} \mathbf{T} \equiv - \left. \frac{d(\phi_{\mathbf{v},t} \mathbf{T})}{dt} \right|_{t=0} = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{T} - \phi_{\mathbf{v},\epsilon} \mathbf{T}}{\epsilon}. \quad (\text{A.37})$$

The overall minus sign is a convention arising from the fact that $(\phi_{\mathbf{v},\epsilon} \mathbf{T})(x)$ corresponds to the value of \mathbf{T} at $\phi_{\mathbf{v},-\epsilon}(x)$, pushed forward to x by the map $\phi_{\mathbf{v},\epsilon}$.

Let us see how this works for vector fields. For the moment, it is practical to resort, even if only implicitly, to local coordinates. The vector field \mathbf{v} then becomes a linear differential operator. The flow equation (A.30) in turn has a formal solution for $\phi_{\mathbf{v},t}$ in terms of the (path-ordered) exponential of \mathbf{v} , given implicitly by

$$f(\phi_{\mathbf{v},t}(x)) = e^{\mathbf{v}t} f(x). \quad (\text{A.38})$$

The definition (A.23) of push-forward now tells us that for any test function f and any vector field \mathbf{u} , $\phi_{\mathbf{v},t*} \mathbf{u}[f] = \mathbf{u}[f \circ \phi_{\mathbf{v},t}] = \mathbf{u}[e^{\mathbf{v}t} f]$. If we start with $\mathbf{u}(x)$, this will give us the value of $\phi_{\mathbf{v},t*} \mathbf{u}$ at $\phi_{\mathbf{v},t}(x)$. To find the value of $\phi_{\mathbf{v},t*} \mathbf{u}$ at x , we invert (A.38) to $f(x) = e^{-\mathbf{v}t} f(\phi_{\mathbf{v},t}(x))$. This leads to

$$\phi_{\mathbf{v},t*} \mathbf{u} = e^{-\mathbf{v}t} \mathbf{u} e^{\mathbf{v}t} \quad \text{hence} \quad \mathcal{L}_{\mathbf{v}} \mathbf{u} = [\mathbf{v}, \mathbf{u}], \quad (\text{A.39})$$

where I introduced the commutator of \mathbf{v} and \mathbf{u} as differential operators, also called the *Lie bracket*. Such a commutator is only meaningful when expressed in local coordinates. However, the Lie derivative as defined by (A.37) is coordinate-independent. Hence also the Lie bracket is well-defined independently of the choice of coordinates.

³ The concept of group action is introduced in a more pedestrian manner and used in Chap. 7.

Let us pause to ponder on the significance of the Lie bracket. First, as a formal commutator, this defines a Lie algebra structure on the space of vector fields on \mathcal{M} . The Jacobi identity for this Lie algebra is equivalent to the Leibniz (product) rule for the Lie derivative,

$$\mathcal{L}_u[v, w] = [\mathcal{L}_u v, w] + [v, \mathcal{L}_u w] . \quad (\text{A.40})$$

What if we take a special set of vector fields that are the generators of the action of a Lie group G on \mathcal{M} ? We find, not surprisingly, that the Lie brackets of these fields reproduce the Lie algebra \mathfrak{g} of G . Instead of a formal proof of this statement, let me show a simple illustrative example.

Example A.9

In Example A.8, I introduced a vector field that generates rotations in \mathbb{R}^2 around the origin. This is easily promoted to \mathbb{R}^3 where we have three independent rotations around the Cartesian coordinate axes, generated by

$$v_x = y\partial_z - z\partial_y , \quad v_y = z\partial_x - x\partial_z , \quad v_z = x\partial_y - y\partial_x . \quad (\text{A.41})$$

The Lie brackets of these,

$$[v_x, v_y] = -v_z , \quad [v_y, v_z] = -v_x , \quad [v_z, v_x] = -v_y , \quad (\text{A.42})$$

reproduce the Lie algebra of infinitesimal rotations in \mathbb{R}^3 , $\mathfrak{so}(3)$. Further generalization to \mathbb{R}^n is straightforward and amounts to considering the set of vector fields $v_{ab} \equiv x_a\partial_b - x_b\partial_a$ with $a, b = 1, \dots, n$.

Second, having at hand a coordinate basis, it is easy to give a general expression for the Lie derivative of a vector field in terms of its components,

$$(\mathcal{L}_v u)^a = v^b \partial_b u^a - u^b \partial_b v^a = v[u^a] - u[v^a] . \quad (\text{A.43})$$

The latter expression clarifies the comment on the relation between coordinate basis fields and (local) frames that I made at the end of Sect. A.2.1. Namely, the set of coordinate basis fields $\{\partial_a\}_{a=1}^n$ always has vanishing Lie brackets. Thus the necessary condition for the existence of local coordinates y^A such that a frame $\{e_A\}_{A=1}^n$ can be represented by $e_A = \partial/\partial y^A$ is $[e_A, e_B] = 0$ for all A, B . It turns out that this condition is also sufficient; see Chap. 9 of [5] for a proof.

Let us now return to the general discussion of the Lie derivative. This can be evaluated on other tensor fields by making use of what we already know for vector fields. For example, let us take a 1-form ω and a test vector field u . Using the definition (A.26) of pull-back together with the inverse of (A.38), we get

$$(\phi_{v,t*}\omega)(\phi_{v,t*}u) = (\phi_{v,-t}^*\omega)(\phi_{v,t*}u) = e^{-vt}\omega(u) . \quad (\text{A.44})$$

Combining the general definition (A.37) of the Lie derivative with the Leibniz rule then leads immediately to

$$(\mathcal{L}_v \omega)(u) = v[\omega(u)] - \omega([v, u]) . \quad (\text{A.45})$$

In fact, the same reasoning can be applied without change to any differential p -form. For a set of test vector fields u_1, \dots, u_p , we find

$$\begin{aligned} (\mathcal{L}_v \omega)(u_1, \dots, u_p) &= v[\omega(u_1, \dots, u_p)] \\ &\quad - \sum_{i=1}^p \omega(u_1, \dots, u_{i-1}, [v, u_i], u_{i+1}, \dots, u_p) . \end{aligned} \quad (\text{A.46})$$

For tensors of the general type (q, p) , it is more straightforward to work with components with respect to a coordinate basis. One can think of a tensor T as a multilinear map acting on a set of test vector fields and 1-forms. Combining the Leibniz rule with (A.39) and (A.45) then leads, after some manipulation, to

$$\begin{aligned} (\mathcal{L}_v T)_{mn\dots}^{ab\dots} &= v^l \partial_l T_{mn\dots}^{ab\dots} + T_{ln\dots}^{ab\dots} \partial_m v^l + T_{ml\dots}^{ab\dots} \partial_n v^l + \dots \\ &\quad - T_{mn\dots}^{lb\dots} \partial_l v^a - T_{mn\dots}^{al\dots} \partial_l v^b - \dots . \end{aligned} \quad (\text{A.47})$$

It is common to extend this identity to the special case of $(q, p) = (0, 0)$, corresponding to functions on \mathcal{M} . In this case, the Lie derivative coincides with the previously defined directional derivative, $\mathcal{L}_v f = v[f]$.

Before closing the section, let me mention a useful identity, connecting the Lie derivative of differential forms with the interior product, defined by (A.17). For two vector fields u, v , one finds that

$$\mathcal{L}_u \circ \iota_v - \iota_v \circ \mathcal{L}_u = \iota_{[u, v]} . \quad (\text{A.48})$$

This is another variation on the Leibniz rule, as one can easily see by rewriting it as $\mathcal{L}_u(\iota_v \omega) = \iota_{[u, v]} \omega + \iota_v(\mathcal{L}_u \omega)$ for an arbitrary test p -form ω . The proof proceeds by combining the definition (A.17) of interior product with (A.46).

A.4 Exterior Derivative

Until now, I have tried to treat on more or less equal footing tensor fields of arbitrary type (q, p) . However, as Sect. A.2.2 already hinted, the (differential) p -forms have a special standing among all the tensors. In this section, I will introduce a new differential structure on the Grassmann algebra of differential forms. The central concept thereof is the *exterior derivative*, which is a derivative-like operation that converts a p -form into a $(p+1)$ -form. This is a tool of great importance for much of both differential and integral calculus on manifolds. There are several ways of introducing the exterior derivative with different balance of abstraction, elegance

and technical simplicity. While there does not seem to be a unique, intuitively simple definition, I will at least try to make a brief comparison of the different approaches.

Consider a differential p -form ω and a set of $p + 1$ test vector fields, \mathbf{v}_i . We can define the exterior derivative of ω as the unique $(p + 1)$ -form $d\omega$ such that

$$\begin{aligned} d\omega(\mathbf{v}_1, \dots, \mathbf{v}_{p+1}) = & \sum_{i=1}^{p+1} (-1)^{i+1} \mathbf{v}_i [\omega(\mathbf{v}_1, \dots, \hat{\mathbf{v}}_i, \dots, \mathbf{v}_{p+1})] \\ & + \sum_{i < j} (-1)^{i+j} \omega([\mathbf{v}_i, \mathbf{v}_j], \mathbf{v}_1, \dots, \hat{\mathbf{v}}_i, \dots, \hat{\mathbf{v}}_j, \dots, \mathbf{v}_{p+1}) \end{aligned} \quad (\text{A.49})$$

for any choice of test fields, where a hat denotes an argument that is to be omitted. This definition is manifestly independent of the choice of local coordinates, but not exactly intuitive. Let us therefore try to unfold its content in some simple cases. First, for 0-forms, that is functions on \mathcal{M} , (A.49) gives simply $df(\mathbf{v}) = \mathbf{v}[f]$. This agrees with our previous definition (A.10) of the differential of a function. Second, for a 1-form ω and a 2-form Ω we get respectively

$$\begin{aligned} d\omega(\mathbf{u}, \mathbf{v}) &= \mathbf{u}[\omega(\mathbf{v})] - \mathbf{v}[\omega(\mathbf{u})] - \omega([\mathbf{u}, \mathbf{v}]) , \\ d\Omega(\mathbf{u}, \mathbf{v}, \mathbf{w}) &= \mathbf{u}[\Omega(\mathbf{v}, \mathbf{w})] + \mathbf{v}[\Omega(\mathbf{w}, \mathbf{u})] + \mathbf{w}[\Omega(\mathbf{u}, \mathbf{v})] \\ &\quad - \Omega([\mathbf{u}, \mathbf{v}], \mathbf{w}) - \Omega([\mathbf{v}, \mathbf{w}], \mathbf{u}) - \Omega([\mathbf{w}, \mathbf{u}], \mathbf{v}) , \end{aligned} \quad (\text{A.50})$$

for arbitrary test vector fields $\mathbf{u}, \mathbf{v}, \mathbf{w}$. It is becoming clear that for any p -form ω , $d\omega$ only contains two types of contributions; the summation in (A.49) is needed to ensure that $d\omega$ maintains full antisymmetry as a multilinear map on vector fields. It is possible to show (see Chap. 36 of [6]) that (A.49) generalizes the concept of a directional derivative from functions to p -forms. This accounts for the first line of (A.49). The second line thereof makes $d\omega$ local so that $d\omega(\mathbf{v}_1, \dots, \mathbf{v}_{p+1})|_x$ only depends on the values of the test vector fields at the point x , not on their derivatives.

By picking the test vector fields \mathbf{v}_i from a coordinate basis and using the component expansion of p -forms (A.13), we find

$$d\omega = \frac{1}{p!} (\partial_b \omega_{a_1 \dots a_p}) dx^b \wedge dx^{a_1} \wedge \dots \wedge dx^{a_p} . \quad (\text{A.51})$$

This is in practice the easiest way to calculate the exterior derivative of a differential form. It makes manifest the fundamental property of the exterior derivative that

$$d^2\omega \equiv d(d\omega) = 0 \quad (\text{A.52})$$

for any p -form ω . Moreover, it can be used as an alternative definition to (A.49) that is technically simple and quite intuitive. The downside of this approach is that it is not obvious that an object defined by (A.51) actually exists, independently of the choice of local coordinates. To prove this requires some extra effort.

Example A.10

In case ω is a 0-form, that is a function on \mathcal{M} , the components of $d\omega$ are $(d\omega)_a = \partial_a \omega$. This generalizes the gradient of functions in \mathbb{R}^n known from ordinary vector calculus. If, on the other hand, ω is a 1-form, it follows at once from (A.51) that the components of $d\omega$ are

$$(d\omega)_{ab} = \partial_a \omega_b - \partial_b \omega_a . \quad (\text{A.53})$$

This appears to generalize the concept of a curl of a vector field in \mathbb{R}^3 . The problem with this interpretation is that neither ω nor $d\omega$ is a vector field. We will see in Sect. A.6.5 how to circumvent this issue.

Antisymmetric derivatives such as (A.53) are common in physics. If $A = A_a dx^a$ represents the potential of an electromagnetic field, then (A.53) are the components of the corresponding field-strength tensor, $F = dA$. In this case, the fundamental property (A.52) of exterior derivative encodes the familiar Bianchi identity for the electromagnetic field,

$$\partial_c F_{ab} + \partial_a F_{bc} + \partial_b F_{ca} = 0 . \quad (\text{A.54})$$

The exterior derivative interconnects nicely with other operations on differential forms that we have met. It is for example easy to prove using the coordinate definition (A.51) that the exterior derivative is an antiderivation of the Grassmann algebra of differential forms. Thus, for any p -form ω and any q -form σ ,

$$d(\omega \wedge \sigma) = d\omega \wedge \sigma + (-1)^p \omega \wedge d\sigma . \quad (\text{A.55})$$

The coordinate-free definition (A.49), on the other hand, makes it easier to show that the exterior derivative “commutes” with pull-back and hence with Lie derivative. This means that for any smooth map $f : \mathcal{M} \rightarrow \mathcal{N}$, differential form ω on \mathcal{N} and vector field \mathbf{v} on \mathcal{N} we have

$$d(f^* \omega) = f^*(d\omega) , \quad d(\mathcal{L}_{\mathbf{v}} \omega) = \mathcal{L}_{\mathbf{v}}(d\omega) . \quad (\text{A.56})$$

The interior product, Lie derivative and exterior derivative constitute three derivative-like operations on differential forms. The first decreases the form degree by one, the second leaves it unchanged and the last increases it by one. There is a simple identity that brings them all together which, perhaps for its intrinsic beauty, is usually referred to as “Cartan’s magic formula,”

$$\mathcal{L}_{\mathbf{v}} = \mathbf{i}_{\mathbf{v}} \circ d + d \circ \mathbf{i}_{\mathbf{v}} \quad \text{or} \quad \mathcal{L}_{\mathbf{v}} \omega = \mathbf{i}_{\mathbf{v}}(d\omega) + d(\mathbf{i}_{\mathbf{v}} \omega) , \quad (\text{A.57})$$

for any p -form ω and vector field \mathbf{v} . The proof is a straightforward combination of the coordinate-free definition (A.49), the Lie derivative of differential forms (A.46) and the definition (A.17) of interior product. Note how the Cartan magic formula together with $d \circ d = 0$ immediately implies the second of the identities in (A.56).

Example A.11

Here is a somewhat more advanced example, closely related to the subject of this book. Take a Lie group G and introduce the following object, called the *Maurer–Cartan (MC) form*,

$$\omega \equiv -ig^{-1}dg \equiv \omega^A Q_A, \quad g \in G, \quad (\text{A.58})$$

where Q_A are the generators of the Lie algebra \mathfrak{g} of G . In order to give a meaning to this definition, one should think of $g \in G$ in terms of a set of functions on G , returning the matrix elements of g in some faithful representation of G . The identities (7.30) and (7.31) of the main text of this book guarantee that ω takes values in the Lie algebra \mathfrak{g} . Hence the components ω^A constitute a set of well-defined 1-forms on G .

Given that the dimension of G as a manifold equals that of its Lie algebra, the 1-forms ω^A furnish a globally well-defined coframe on G . The exterior derivative $d\omega^A$ should therefore be a linear combination of $\omega^B \wedge \omega^C$. Indeed, a short calculation shows that the expression of $d\omega^A$ in terms of $\omega^B \wedge \omega^C$ reflects the algebraic structure of G as a Lie group,

$$\begin{aligned} d\omega &= -idg^{-1} \wedge dg = ig^{-1}dg \wedge g^{-1}dg = -i\omega \wedge \omega, \\ d\omega^A &= \frac{1}{2}f_{BC}^A \omega^B \wedge \omega^C, \end{aligned} \quad (\text{A.59})$$

where f_{BC}^A are the structure constants of G . This is known as the MC equation.

Before closing the quick overview of exterior calculus, let me mention in passing another way to define the exterior derivative. This might please a reader who prefers a more axiomatic approach. Namely, the exterior derivative turns out to be a unique antiderivation of the Grassmann algebra that reduces to the ordinary differential on 0-forms and satisfies the fundamental property (A.52). For a reader that might prefer a more geometric definition of exterior derivative, there is yet another possibility. This is intimately related to integration on manifolds and I will return to it in Sect. A.7.

A.5 Affine Connection

In order to develop a full-fledged differential calculus on manifolds, we have to be able to differentiate not only functions, but also vector and tensor fields. In spite of the terminology, none of the two previously introduced “derivatives” is suitable for the purpose. First, the Lie derivative \mathcal{L}_v actually does not have the properties we would expect from a derivative in that it is not local in v . The value of $\mathcal{L}_v T$ for a given tensor field T at $x \in \mathcal{M}$ depends not only on $v(x)$ but also on its derivatives, as is clear from (A.47). Second, the exterior derivative is only defined for differential

forms and not for other tensors. To have a well-defined notion of derivative for tensor fields of arbitrary type, we need additional structure on the manifold.

In order to see what the required structure might be, let us consider the simplest case of a test vector field \mathbf{u} . We might think of first finding its components u^A in a chosen local frame \mathbf{e}_A and then taking their directional derivative $\mathbf{v}[u^A]$ with respect to a fixed vector field \mathbf{v} . That is however problematic. The choice of frame \mathbf{e}_A is arbitrary and can be changed locally. This is a smoking gun of what in field theory is called “gauge freedom.” The directional derivative of \mathbf{u} must be a well-defined vector field independent of such a “choice of gauge.” This is not the case of $\mathbf{v}[u^A]\mathbf{e}_A$, so $\mathbf{v}[u^A]$ cannot be the components of a well-defined vector field. Following the analogy with physics, the problem can be fixed by introducing a gauge field.

A.5.1 Covariant Derivative

Let us now formalize the above observation. The basis $\mathbf{e}_A(x)$ of $\mathcal{T}_x\mathcal{M}$ can be changed by any nonsingular linear transformation that varies (smoothly) with x . Upon such a transformation, the basis vectors and components of a tangent vector \mathbf{u} change to

$$\tilde{\mathbf{e}}_A = \mathbf{e}_B (P^{-1})^B_A, \quad \tilde{u}^A = P^A_B u^B, \quad P \in \text{GL}(n). \quad (\text{A.60})$$

The group $\text{GL}(n)$ is our gauge group; in differential geometry, the term *structure group* is common. Any geometrically well-defined object on the manifold \mathcal{M} must be invariant under gauge transformations of the type (A.60). It is known from field theory that in order to have a well-defined notion of derivative of fields, one needs a gauge field, or gauge connection. In differential geometry, this object is called *affine connection*. In a given frame, it constitutes a collection of 1-forms Ω^A_B . Treating these as the matrix elements of a matrix-valued 1-form Ω , the action of the basis transformation (A.60) on the affine connection can be defined compactly as

$$\tilde{\Omega} = P\Omega P^{-1} + P dP^{-1}, \quad P \in \text{GL}(n). \quad (\text{A.61})$$

This is analogous to a (non-Abelian) gauge transformation familiar from field theory.

The collection of 1-forms Ω^A_B carries one vector and one covector frame index. In spite of the suggestive notation, the object $\mathbf{e}_A \otimes \mathbf{e}^{*B} \otimes \Omega^A_B$ is not a well-defined tensor field of type $(1, 2)$ on \mathcal{M} . This is a consequence of the second term in (A.61). The affine connection is therefore inseparably tied to the chosen local frame.

Following the analogy with field theory, we now define a directional *covariant derivative* $\nabla_{\mathbf{v}}\mathbf{u}$ of the vector field \mathbf{u} along \mathbf{v} component-wise as

$$(\nabla_{\mathbf{v}}\mathbf{u})^A \equiv \mathbf{v}[u^A] + \Omega^A_B(\mathbf{v})u^B. \quad (\text{A.62})$$

It is easy to check that the components $(\nabla_{\mathbf{v}}\mathbf{u})^A$ transform under the local change of basis (A.60) just like u^A . Hence the directional covariant derivative $\nabla_{\mathbf{v}}\mathbf{u} \equiv (\nabla_{\mathbf{v}}\mathbf{u})^A \mathbf{e}_A$ is a well-defined vector field on \mathcal{M} independent of the choice of frame. Moreover, $\nabla_{\mathbf{v}}\mathbf{u}$ is manifestly linear and local in \mathbf{v} . We can therefore separate \mathbf{v} out and only keep a tensor object that encodes the variation of \mathbf{u} with respect to the local frame. This is done by setting $(\nabla_{\mathbf{v}}\mathbf{u})^A = v^B (\nabla_{\mathbf{e}_B}\mathbf{u})^A \equiv v^B (\nabla_B \mathbf{u})^A$, where

$$(\nabla_B \mathbf{u})^A = \mathbf{e}_B[u^A] + \Gamma_{BC}^A u^C, \quad \Gamma_{BC}^A \equiv \Omega_C^A(\mathbf{e}_B) = (\nabla_B \mathbf{e}_C)^A, \quad (\text{A.63})$$

and Γ_{BC}^A are the *Christoffel symbols*. These are independent of \mathbf{u} and characterize the geometry of the manifold and of the local frame.

The field theory analogy allows us to write down at once the covariant derivative of a tensor field of any type (q, p) . Different types of tensors transform under different representations of the structure group, induced by the fundamental representation (A.60) acting on vectors. We can thus reuse (A.62) by inserting the appropriate representation of the affine connection. For instance, tensor fields of type $(0, 0)$, that is functions on \mathcal{M} , are blind to the change of local frame. This corresponds to the trivial representation of $\text{GL}(n)$. The covariant derivative then reduces to $\nabla_{\mathbf{v}}f = \mathbf{v}[f]$. For a generic tensor \mathbf{T} of type (q, p) we find

$$\begin{aligned} (\nabla_{\mathbf{v}}\mathbf{T})_{KL\dots}^{AB\dots} &= \mathbf{v}[\mathbf{T}_{KL\dots}^{AB\dots}] + \Omega_I^A(\mathbf{v})\mathbf{T}_{KL\dots}^{IB\dots} + \Omega_I^B(\mathbf{v})\mathbf{T}_{KL\dots}^{AI\dots} + \dots \\ &\quad - \mathbf{T}_{IL\dots}^{AB\dots}\Omega_K^I(\mathbf{v}) - \mathbf{T}_{KI\dots}^{AB\dots}\Omega_L^I(\mathbf{v}) - \dots \end{aligned} \quad (\text{A.64})$$

This can also be rewritten in a manner independent of the vector field \mathbf{v} , following the example of (A.63),

$$\begin{aligned} (\nabla_C \mathbf{T})_{KL\dots}^{AB\dots} &= \mathbf{e}_C[\mathbf{T}_{KL\dots}^{AB\dots}] + \Gamma_{CI}^A \mathbf{T}_{KL\dots}^{IB\dots} + \Gamma_{CI}^B \mathbf{T}_{KL\dots}^{AI\dots} + \dots \\ &\quad - \mathbf{T}_{IL\dots}^{AB\dots}\Gamma_{CK}^I - \mathbf{T}_{KI\dots}^{AB\dots}\Gamma_{CL}^I - \dots \end{aligned} \quad (\text{A.65})$$

Example A.12

The natural choice of \mathbf{e}_A in \mathbb{R}^n is the coordinate basis in the Cartesian coordinates. In this basis, we can set $\Omega = 0$ so that $\nabla_{\mathbf{v}}\mathbf{u} = \mathbf{v}[u^A]\mathbf{e}_A$. This expresses the fact known from elementary vector calculus that \mathbf{e}_A is a globally well-defined, constant basis of vectors. Switching to another local frame $\tilde{\mathbf{e}}_A$ via (A.60) induces a nontrivial connection, $\tilde{\Omega} = P dP^{-1}$. This characterizes the “variation” of the new frame throughout \mathbb{R}^n . A short manipulation leads to an expression for the Christoffel symbols in terms of the derivatives of P ,

$$\begin{aligned} \tilde{\Gamma}_{BC}^A &= -(dPP^{-1})_C^A(\tilde{\mathbf{e}}_B) = -(P^{-1})^E{}_C \tilde{\mathbf{e}}_B[P^A{}_E] \\ &= -(P^{-1})^D{}_B (P^{-1})^E{}_C \mathbf{e}_D[P^A{}_E] = -(P^{-1})^D{}_B (P^{-1})^E{}_C \partial_D P^A{}_E. \end{aligned} \quad (\text{A.66})$$

Equations (A.63) and (A.65) then allow us to calculate derivatives of vectors and tensors in \mathbb{R}^n in terms of their components in arbitrary curvilinear coordinates.

A.5.2 Curvature and Torsion

We can pursue the analogy with field theory even further. Therein, one uses the gauge connection to construct the corresponding field strength which encodes information about the gauge field in a gauge-covariant fashion. In the geometric language, this leads to the *curvature 2-form* on the manifold \mathcal{M} ,

$$R^A{}_B \equiv d\Omega^A{}_B + \Omega^A{}_C \wedge \Omega^C{}_B . \quad (\text{A.67})$$

The matrix elements $R^A{}_B$ can be put together into a single object, $R \equiv e_A \otimes e^{*B} \otimes R^A{}_B$. Unlike the matrix-valued connection 1-form Ω , this is a well-defined tensor field on \mathcal{M} , of type $(1, 3)$, independent of the choice of local frame. It therefore carries information about the intrinsic geometry of the manifold. The curvature 2-form is related to the commutator of directional covariant derivatives via

$$(\nabla_u \nabla_v w - \nabla_v \nabla_u w - \nabla_{[u, v]} w)^A = R^A{}_B(u, v) w^B , \quad (\text{A.68})$$

where u , v and w are vector fields. This identity follows directly from the definition (A.62) of covariant derivative upon using the first line of (A.50). Using (A.64), it can moreover be immediately generalized to the commutator of covariant derivatives of an arbitrary tensor field T ,

$$\begin{aligned} ([\nabla_u, \nabla_v] T - \nabla_{[u, v]} T)^{AB\cdots}_{KL\cdots} &= R^A{}_I(u, v) T^{IB\cdots}_{KL\cdots} + R^B{}_I(u, v) T^{AI\cdots}_{KL\cdots} + \cdots \\ &\quad - T^{AB\cdots}_{IL\cdots} R^I{}_K(u, v) - T^{AB\cdots}_{KI\cdots} R^I{}_L(u, v) - \cdots . \end{aligned} \quad (\text{A.69})$$

The affine connection also gives rise to another well-defined tensor field that characterizes the geometry of the manifold. Namely, the elements of the dual coframe e^{*A} transform under (A.60) identically to the components of a vector. We can then define a covariant exterior derivative of the coframe in parallel with (A.62),

$$T^A \equiv de^{*A} + \Omega^A{}_B \wedge e^{*B} . \quad (\text{A.70})$$

This is dubbed the *torsion 2-form*. It is a collection of 2-forms whose components transform under the structure group as a vector. Hence $T \equiv e_A \otimes T^A$ is a well-defined tensor of type $(1, 2)$. The torsion 2-form satisfies an identity similar to (A.68),

$$(\nabla_u v - \nabla_v u - [u, v])^A = T^A(u, v) . \quad (\text{A.71})$$

Owing to the fact that $d \circ d = 0$, one may expect to find constraints on the curvature and torsion 2-forms by taking another (covariant) exterior derivative. Indeed, a short manipulation based on the definitions (A.67) and (A.70) leads to the so-called *Bianchi identities*,

$$\begin{aligned} dR^A{}_B + \Omega^A{}_C \wedge R^C{}_B - R^A{}_C \wedge \Omega^C{}_B &= 0, \\ dT^A + \Omega^A{}_B \wedge T^B &= R^A{}_B \wedge e^{*B}. \end{aligned} \quad (\text{A.72})$$

Example A.13

Using the MC form introduced in Example A.11, we can get insight into the geometry of Lie groups. The simplest way to introduce an affine connection on a Lie group G is to set $\Omega^A{}_B = 0$ in the coframe ω^A defined by (A.58). The definition (A.67) then tells us that the curvature 2-form vanishes. The torsion 2-form does not, though. From the definition (A.70) and the MC equation (A.59), we get that $T^A = (1/2)f^A_{BC}\omega^B \wedge \omega^C$, where f^A_{BC} are the structure constants of G .

Making Ω trivial is however not the only choice of connection we can make. Given the structure that we have available on the Lie group G , we can try for instance

$$\lambda \Omega^A{}_B \equiv \lambda f^A_{BC} \omega^C, \quad (\text{A.73})$$

where $\lambda \in \mathbb{R}$ is a parameter. This class of connections includes the previous trivial case as $\lambda = 0$. Using the MC equation (A.59), one now finds that

$$\begin{aligned} \lambda R^A{}_B &= \left(\frac{\lambda}{2} f^A_{BE} f^E_{CD} + \lambda^2 f^A_{EC} f^E_{BD} \right) \omega^C \wedge \omega^D, \\ \lambda T^A &= \left(\frac{1}{2} - \lambda \right) f^A_{BC} \omega^B \wedge \omega^C. \end{aligned} \quad (\text{A.74})$$

If we do not impose any constraints on the structure constants f^A_{BC} , we can generally make the curvature 2-form vanish only by returning to the special case $\lambda = 0$. On the other hand, we can make the torsion 2-form vanish by setting $\lambda = 1/2$. We can however not make both 2-forms vanish simultaneously. A generic non-Abelian Lie group therefore possesses nontrivial geometry as a manifold.

Our original motivation for introducing the affine connection was to take account of the freedom to choose a local frame at will. This is largely responsible for the fact that I have phrased this section entirely in a coordinate-free language so far. It is however possible to reformulate the definitions and properties of covariant derivative, curvature and torsion in terms of local coordinates. This leads to minor simplification of some of the equations in this section. Thus, the action of the basis vector e_C in (A.65) becomes a mere partial derivative and (A.65) turns into

$$\begin{aligned} (\nabla_c T)^{ab\dots}_{kl\dots} &= \partial_c T^{ab\dots}_{kl\dots} + \Gamma^a_{ci} T^{ib\dots}_{kl\dots} + \Gamma^b_{ci} T^{ai\dots}_{kl\dots} + \dots \\ &\quad - T^{ab\dots}_{il\dots} \Gamma^i_{ck} - T^{ab\dots}_{ki\dots} \Gamma^i_{cl} - \dots. \end{aligned} \quad (\text{A.75})$$

The Christoffel symbol herein is defined by a coordinate-basis equivalent of (A.63),

$$\Gamma^a_{bc} = (\nabla_b \partial_c)^a \quad \text{or} \quad \nabla_b \partial_c = \Gamma^a_{bc} \partial_a. \quad (\text{A.76})$$

Together with (A.70), this shows that the torsion tensor corresponds to the antisymmetric part of the Christoffel symbol in a coordinate basis,

$$\Gamma_{bc}^a - \Gamma_{cb}^a = (\nabla_b \partial_c - \nabla_c \partial_b)^a = T^a(\partial_b, \partial_c) \equiv T_{bc}^a . \quad (\text{A.77})$$

Finally, the left-hand side of the identity (A.68) for the curvature tensor reduces to a mere commutator of covariant derivatives, hence

$$(\nabla_b \nabla_c \mathbf{u} - \nabla_c \nabla_b \mathbf{u})^a = R^a_d(\partial_b, \partial_c) u^d \equiv R^a_{dbc} u^d . \quad (\text{A.78})$$

Likewise, (A.69) gives the commutator of covariant derivatives of an arbitrary tensor field in the coordinate basis,

$$\begin{aligned} (\nabla_c \nabla_d \mathbb{T} - \nabla_d \nabla_c \mathbb{T})^{ab\cdots}_{kl\cdots} &= R^a_{icd} \mathbb{T}^{ib\cdots}_{kl\cdots} + R^b_{icd} \mathbb{T}^{ai\cdots}_{kl\cdots} + \cdots \\ &\quad - \mathbb{T}^{ab\cdots}_{il\cdots} R^i_{kcd} - \mathbb{T}^{ab\cdots}_{ki\cdots} R^i_{lcd} - \cdots . \end{aligned} \quad (\text{A.79})$$

A.6 Riemannian Geometry

My exposition of differential geometry has so far been largely ahistorical. I made a conscious effort to only introduce the structure that was strictly necessary in order to develop the elementary concepts of calculus on manifolds. Thus, it only requires the basic differential structure on a manifold to define tensor fields, their Lie derivative, and the exterior calculus of differential forms. The first time that a new, additional structure was needed was in the previous section. There, we saw how the concept of affine connection naturally arises from the need to compute (covariant) derivatives of tensor fields.

However, as the Greek origin of the word “geometry” suggests, much of the early progress in the subject was driven by the practical problem of measuring distance. The concept of distance is very natural when the manifold in question represents space or spacetime in physics. Yet, even more abstract realizations of manifolds often possess a quantitative notion of closeness that is more refined than mere topology. The purpose of this section is to introduce the structure needed to measure distance on a manifold. Despite being anchored in geometry, we shall see that this will also give us further useful tools for our quest to develop calculus on manifolds.

A.6.1 Riemannian Metric

Since I have already mentioned the concept of metric before, let me start right away with the formal definition. A *Riemannian metric* g on a manifold \mathcal{M} is a tensor field of type $(0, 2)$ such that for any $x \in \mathcal{M}$, $g(x)$ is a symmetric, positive-definite bilinear form on $\mathcal{T}_x \mathcal{M}$. See Example A.4 and Example A.5. A manifold endowed with a Riemannian metric is called a *Riemannian manifold*. It is common to represent the

metric by its components in a given coframe or dual coordinate basis,

$$g = g_{AB} \mathbf{e}^{*A} \otimes \mathbf{e}^{*B} = g_{ab} dx^a \otimes dx^b, \quad (\text{A.80})$$

where $g_{AB} = g(\mathbf{e}_A, \mathbf{e}_B)$ and $g_{ab} = g(\partial_a, \partial_b)$. A Riemannian metric gives by construction rise to an inner product, both locally of tangent vectors and of vector fields. It is practically convenient to have a shorthand notation for this inner product,

$$\langle \mathbf{u}, \mathbf{v} \rangle \equiv g(\mathbf{u}, \mathbf{v}) = g_{AB} u^A v^B, \quad (\text{A.81})$$

where \mathbf{u}, \mathbf{v} are arbitrary vector fields.

For some applications, the axioms for Riemannian metric are too restrictive. By removing the requirement of positive-definiteness and insisting merely that g as a bilinear form is nondegenerate, one obtains a generalization called *pseudo-Riemannian metric*. Likewise, manifolds endowed with a pseudo-Riemannian metric are known as *pseudo-Riemannian manifolds*. This broader concept finds an important application in the general theory of relativity, which assumes the physical spacetime to be a *Lorentzian manifold*. The latter possesses by definition a metric with signature $(-, +, +, \dots)$ or $(+, -, -, \dots)$, depending on the convention adopted. Most of the contents of this section apply, with occasional minor adjustments, to Riemannian and pseudo-Riemannian manifolds alike. However, within the context of the book, pseudo-Riemannian manifolds play a little role. I will therefore content myself with this remark of caution and focus solely on Riemannian manifolds in the following.

Before illustrating the utility of the metric, it is worthwhile to contemplate its very definition. How is this actually related to the distance of two different points on the manifold? First, the metric by definition acts point-wise on tangent vectors. Second, the tangent vectors themselves were defined as differential operators on local test functions. To connect the abstract formalism to the down-to-earth problem of measuring distance, it helps to recall the concept of flow generated by the vector field \mathbf{v} , introduced in Sect. A.3.2. One can interpret $\langle \mathbf{v}, \mathbf{v} \rangle$ as the squared length of the vector tangent to the curve (trajectory) $\gamma(t)$ defined by (A.30). In the liquid analogy of the flow generated by \mathbf{v} , $\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ becomes the speed of the liquid particle. The distance of points $x_{1,2} = \gamma(t_{1,2})$ measured along the curve $\gamma(t)$ is then naturally defined by integration,

$$d_\gamma(x_1, x_2) \equiv \int_{t_1}^{t_2} \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \Big|_{\gamma(t)} dt. \quad (\text{A.82})$$

This distance is independent of the choice of parameterization of the trajectory, that is of the speed at which the trajectory is traversed by the liquid particle. It is therefore an intrinsic property of the curve γ as a set of points on \mathcal{M} .

Now that we know how to measure the length of a curve on \mathcal{M} , we define the distance of points $x_1, x_2 \in \mathcal{M}$ as the minimum possible value of $d_\gamma(x_1, x_2)$. This is a problem in variational calculus that leads to a second-order differential equation for $\gamma(t)$ (see Sect. 11.5.1 of [1]),

$$\frac{d^2 \gamma^a}{dt^2} + \hat{\Gamma}_{bc}^a \frac{d\gamma^b}{dt} \frac{d\gamma^c}{dt} = 0 . \quad (\text{A.83})$$

Here the set of coefficients $\hat{\Gamma}_{bc}^a$ is determined by the Riemannian metric,

$$\hat{\Gamma}_{bc}^a \equiv \frac{1}{2} g^{ad} (\partial_b g_{dc} + \partial_c g_{bd} - \partial_d g_{bc}) , \quad (\text{A.84})$$

where g^{ab} are the matrix elements of the inverse of g_{ab} . Equation (A.83) is known as the *geodesic equation*. The resemblance between its left-hand side and the covariant derivative $\nabla_{\mathbf{v}} \mathbf{v}$ with $v^a = d\gamma^a/dt$ is not accidental. The coefficients (A.84) are the Christoffel symbols of a special connection on the Riemannian manifold \mathcal{M} , called the *Levi-Civita (LC) connection*. The fact that $\hat{\Gamma}_{bc}^a$ is symmetric in its lower indices implies by means of (A.77) that the LC connection has vanishing torsion. A given Riemannian manifold can support many different affine connections. The LC connection is however unique in that it is torsion-free and makes the metric of the manifold covariantly constant (see Sect. 7.4 of [2]),

$$(\hat{\nabla}_c g)_{ab} = \partial_c g_{ab} - g_{db} \hat{\Gamma}_{ca}^d - g_{ad} \hat{\Gamma}_{cb}^d = 0 . \quad (\text{A.85})$$

Example A.14

The origin of the LC connection can be understood in elementary terms. Imagine that the manifold \mathcal{M} is smoothly embedded in \mathbb{R}^n for some $n > \dim \mathcal{M}$. We can then rely on the geometric intuition whereby tangent vectors to \mathcal{M} are “realized” as vectors in \mathbb{R}^n by the embedding $f : \mathcal{M} \rightarrow \mathbb{R}^n$, as in Example A.6. I will take the liberty to identify tangent vectors to \mathcal{M} with their images under the push-forward f_* .

Suppose we want to define the derivative of a vector field \mathbf{u} on \mathcal{M} along the vector field \mathbf{v} . To do so, we need to compare $\mathbf{u}(x)$ at a given point $x \in \mathcal{M}$ with $\mathbf{u}(x + \epsilon \mathbf{v})$, where the shorthand notation $x + \epsilon \mathbf{v}$ indicates the point on \mathcal{M} with coordinates $x^a + \epsilon v^a(x)$. We cannot do this directly since the tangent spaces to \mathcal{M} at x and $x + \epsilon \mathbf{v}$ are different. What we will do instead is take the vector $\mathbf{u}(x)$, transport it as a constant vector in \mathbb{R}^n to $x + \epsilon \mathbf{v}$ and then project it back to the hypersurface \mathcal{M} . See Fig. A.4 for an illustration of the procedure.

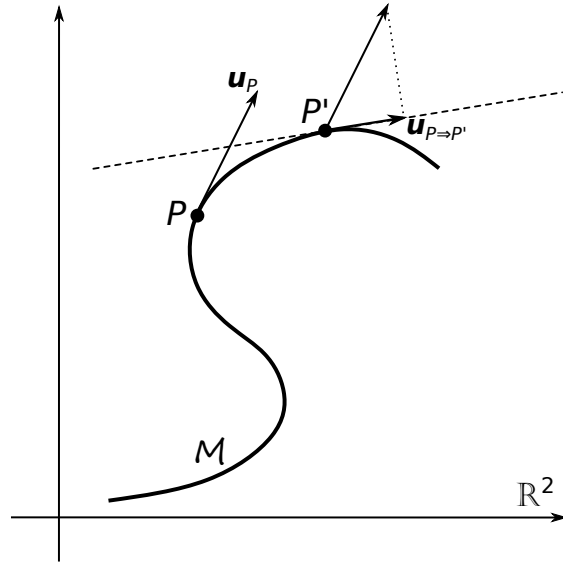
It is easiest to use a frame \mathbf{e}_A on \mathcal{M} that is orthonormal with respect to the Euclidean metric in \mathbb{R}^n . The projection to \mathcal{M} is then accomplished by taking the inner product of $\mathbf{u}(x)$ and $\mathbf{e}_A(x + \epsilon \mathbf{v})$. This leads to a definition of covariant derivative on \mathcal{M} as

$$(\nabla_{\mathbf{v}} \mathbf{u})^A(x) = \lim_{\epsilon \rightarrow 0} \frac{u^A(x + \epsilon \mathbf{v}) - \langle \mathbf{e}_A(x + \epsilon \mathbf{v}), \mathbf{u}(x) \rangle}{\epsilon} . \quad (\text{A.86})$$

Comparing this to (A.63) allows us to identify the corresponding Christoffel symbols,

$$\Gamma_{BC}^A = -\langle \partial_B \mathbf{e}_A, \mathbf{e}_C \rangle = \langle \mathbf{e}_A, \partial_B \mathbf{e}_C \rangle , \quad (\text{A.87})$$

Fig. A.4 Comparison of tangent vectors to \mathcal{M} at different points P and P' is possible if the manifold is embedded in \mathbb{R}^n ; here shown for simplicity for one-dimensional \mathcal{M} embedded in \mathbb{R}^2 . A tangent vector at $P \in \mathcal{M}$, \mathbf{u}_P , is transported to P' as a constant vector in \mathbb{R}^2 . Subsequently, it is projected to the tangent space to \mathcal{M} at P' , resulting in the vector $\mathbf{u}_{P \Rightarrow P'}$. This can now be compared to any other tangent vector at P' . The distance of points P and P' is exaggerated for clarity; for the definition of covariant derivative, it should actually be infinitesimal



where $\partial_B e_A(x)$ is defined as $\lim_{\epsilon \rightarrow 0} [e_A(x + \epsilon e_B) - e_A(x)] / \epsilon$. We have thus managed to define an induced affine connection on \mathcal{M} in a way that encodes the variation of the orthonormal frame e_A as a set of vectors in \mathbb{R}^n . It is an easy exercise to check using (A.71) that this affine connection is torsion-free. Moreover, with the induced metric on \mathcal{M} given by $g_{AB} = \langle e_A, e_B \rangle = \delta_{AB}$, one readily checks using (A.65) that the connection (A.87) makes the metric covariantly constant. Using the embedding in \mathbb{R}^n , we have therefore found the (unique) LC connection on \mathcal{M} .

As a side remark, recall that the comparison of tangent vectors at different points on \mathcal{M} was made possible by “transporting” them as constant vectors in \mathbb{R}^n . This intuition lies behind an often-used geometric picture of affine connection in terms of *parallel transport*. In general, a tensor field T is said to be parallel-transported along the vector field \mathbf{v} if $\nabla_{\mathbf{v}} T = 0$. The geodesic equation (A.83) then describes curves whose tangent vector is parallel-transported along the curve. The solutions of (A.83) thus have the dual interpretation as the shortest (or more precisely extremal) or the “straightest” paths on \mathcal{M} . See for example Sect. 7.2 of [2] for more details.

The metric obviously provides us with a crucial piece of information about the intrinsic geometry of a Riemannian manifold. It is nevertheless also invaluable for the calculus on manifolds, as it allows us to relate tensors of different types. It is known from elementary linear algebra that a vector space such as $\mathcal{T}_x \mathcal{M}$ for fixed $x \in \mathcal{M}$ and its dual, $\mathcal{T}_x^* \mathcal{M}$, have the same dimension. As such, they are necessarily isomorphic. There is, however, no natural way to set up the isomorphism without first fixing a basis in the two spaces. Here the metric comes to rescue. It allows us to define two mutually inverse *musical isomorphisms*, $\flat : \mathcal{T}_x \mathcal{M} \rightarrow \mathcal{T}_x^* \mathcal{M}$ and $\sharp : \mathcal{T}_x^* \mathcal{M} \rightarrow \mathcal{T}_x \mathcal{M}$,

and by extension isomorphisms between vector fields and 1-forms. Thus, any vector field \mathbf{v} is mapped to its *flat* \mathbf{v}^\flat , which is a 1-form such that

$$\mathbf{v}^\flat(\mathbf{u}) \equiv \langle \mathbf{v}, \mathbf{u} \rangle \quad \text{for any vector field } \mathbf{u} . \quad (\text{A.88})$$

Vice versa, any differential 1-form ω is mapped to its *sharp* ω^\sharp , which is a vector field such that

$$\langle \omega^\sharp, \mathbf{v} \rangle \equiv \omega(\mathbf{v}) \quad \text{for any vector field } \mathbf{v} . \quad (\text{A.89})$$

Most theoretical physicists are familiar with these maps in their component forms,

$$v_A^\flat = g_{AB} v^B , \quad \omega^{\sharp A} = g^{AB} \omega_B . \quad (\text{A.90})$$

These correspond to the usual lowering and raising of indices by the metric and its inverse. The generalization of these operations to tensors of other types is straightforward and I will not dwell on details.

It is common in the physics literature to denote a tensor and its image under \flat or \sharp by the same symbols, and only distinguish them by the positions of their indices. In a restricted sense, it can even be meaningful to identify a vector with its image under \flat or a 1-form with its image under \sharp . Suppose that we choose an *orthonormal frame* on \mathcal{M} , that is a frame \mathbf{e}_A such that $\langle \mathbf{e}_A, \mathbf{e}_B \rangle = \delta_{AB}$. Then it follows from (A.88) that \flat maps this basis to the corresponding dual basis, $\mathbf{e}_A^\flat = \mathbf{e}^{*A}$. The same correspondence applies to the inverse isomorphism, $(\mathbf{e}^{*A})^\sharp = \mathbf{e}_A$. By linearity, the components of any vector then remain unaffected by \flat , and the components of any 1-form remain unaffected by \sharp .

A.6.2 Isometries of Riemannian Metric

The whole book is devoted to symmetries and their consequences in quantum field theory. But what do we mean by a symmetry in case of a manifold? I already pointed out in Sect. A.3.2 that a group G can act on a manifold \mathcal{M} via a set of transformations (diffeomorphisms) on \mathcal{M} . However, not every transformation on \mathcal{M} is a symmetry. For that, the transformation should in some sense respect the structure of the manifold. The most common type of symmetry one meets in differential geometry is one that preserves the Riemannian metric of the manifold.

Formally, a diffeomorphism $f : \mathcal{M} \rightarrow \mathcal{M}$ is called an *isometry* if it leaves the Riemannian metric g on \mathcal{M} invariant, $f^*g = g$. Equivalently, one may require that the inner product of test vector fields \mathbf{u}, \mathbf{v} is unaffected by push-forward by f ,⁴

⁴ This definition can at once be generalized to a *local isometry*, which is a smooth but not necessarily invertible map between two different manifolds, $f : \mathcal{M} \rightarrow \mathcal{N}$, satisfying (A.91). (The

$$\langle f_*\mathbf{u}, f_*\mathbf{v} \rangle|_{f(x)} = \langle \mathbf{u}, \mathbf{v} \rangle|_x \quad \text{for any } x \in \mathcal{M}. \quad (\text{A.91})$$

The set of isometries of a Riemannian manifold forms a group G under map composition. This group may be both discrete, or even trivial, and continuous. It is however generally not easy to determine the isometry group G of a given manifold \mathcal{M} directly, unless one has some additional insight into the structure of \mathcal{M} .

The task is much easier when G is a Lie group. We can then focus on symmetries of the metric under infinitesimal transformations. In line with the discussion in Sect. A.3.3, this amounts to solving the condition $\mathcal{L}_\xi g = 0$, where ξ is a vector field representing a generator of the isometry group on \mathcal{M} . This is known as the *Killing vector field*, or more briefly *Killing vector*. The condition on ξ is equivalent to a set of homogeneous linear partial differential equations for the components of ξ . These are usually much easier to solve than (A.91). In practice, it is often convenient to use an alternative formulation of the invariance condition, based on the LC connection instead of the Lie derivative,

$$\mathcal{L}_\xi g = 0 \quad \Leftrightarrow \quad (\hat{\nabla}_a \xi^b)_b + (\hat{\nabla}_b \xi^b)_a = 0. \quad (\text{A.92})$$

One then first solves for the 1-form ξ^b and subsequently recovers the Killing vector field ξ by acting on ξ^b with \sharp .

Example A.15

For illustration, let us find the continuous isometries of the Euclidean metric, $g = \delta_{ab} dx^a \otimes dx^b$, in \mathbb{R}^n . Dropping for simplicity the superscript b on ξ , and using the fact that the LC connection of the Euclidean metric is trivial in Cartesian coordinates, (A.92) becomes

$$\partial_a \xi_b + \partial_b \xi_a = 0. \quad (\text{A.93})$$

Hence $\partial_a \xi_b$ must be antisymmetric in its two indices. Using this repeatedly together with the symmetry of second partial derivatives, we infer

$$\partial_a \partial_b \xi_c = -\partial_a \partial_c \xi_b = \partial_b \partial_c \xi_a = -\partial_b \partial_a \xi_c = -\partial_a \partial_b \xi_c. \quad (\text{A.94})$$

It follows that all second partial derivatives of ξ_a must vanish. This conclusion has a neat generalization to an arbitrary Riemannian manifold. Following essentially the same reasoning, one finds that

$$\hat{\nabla}_a \hat{\nabla}_b \xi_c = \xi_d R^d_{abc} \quad \text{or} \quad \hat{\nabla}_a \hat{\nabla}_b \xi^c = R^c_{bad} \xi^d, \quad (\text{A.95})$$

where R^d_{abc} is the curvature tensor of the LC connection. The outer covariant derivative in (A.95) is defined by treating $\hat{\nabla}_b \xi_c$ as a rank-2 tensor.

Together with (A.93), the vanishing of second partial derivatives of ξ_a implies the most general solution of the *Killing equation* (A.92) for the Euclidean metric,

inner products on the two sides of the equation are then defined with the respective metrics on \mathcal{M} and \mathcal{N} .) In this case, it is of course not appropriate to speak of a symmetry of either \mathcal{M} or \mathcal{N} .

$$\xi_a(x) = \alpha_a + \beta_{ab}x^b . \quad (\text{A.96})$$

Here α_a is a set of constants and β_{ab} a constant antisymmetric matrix. This corresponds to a combination of infinitesimal translations and rotations, which together generate the Euclidean group $\text{ISO}(n)$.

The Killing equation only gives us information about the Lie algebra of the isometry group G . The reconstruction of the whole group thus suffers from the usual ambiguities of the correspondence between Lie groups and Lie algebras. In addition, (A.92) is based on local information about the metric structure of the manifold \mathcal{M} . The global topology of \mathcal{M} may select only a subalgebra of solutions to (A.92). The following example, adapted from Sect. 4.6 of [3], illustrates this subtlety.

Example A.16

The set $\Gamma \equiv \{(c_1, \dots, c_n) \mid c_a \in \mathbb{Z}\}$ defines a lattice of points in \mathbb{R}^n with integer coordinates. The quotient set \mathbb{R}^n/Γ is the n -dimensional torus, T^n . In local Cartesian coordinates, the Euclidean metric on T^n takes the same form as in \mathbb{R}^n , $g = \delta_{ab}dx^a \otimes dx^b$. The Killing equation (A.92) therefore produces the same Lie algebra of local solutions as in Example A.15. However, not all the solutions (A.96) are well-defined 1-forms on T^n . In fact, the matrix β_{ab} has to vanish; the torus does not have continuous rotation symmetries. Moreover, the continuous translation symmetry of T^n differs from that of \mathbb{R}^n . It is $U(1)^{\times n}$ for the torus but \mathbb{R}^n for the Euclidean space. In addition, the torus has a discrete group of rotation isometries, which corresponds to the point group of the lattice Γ . Such discrete isometries are of course not captured by the Killing equation (A.92) at all.

A.6.3 Symmetries of Curvature Tensor

Before developing further the basics of Riemannian geometry, I will digress to put together some useful properties of the curvature tensor R^A_{BCD} . This is defined in any local frame by $R^A_B(\mathbf{u}, \mathbf{v}) = R^A_{BCD}u^Cv^D$, where \mathbf{u}, \mathbf{v} are arbitrary vector fields. The first property that follows immediately from the definition of R^A_B as a 2-form is the antisymmetry,

$$R^A_B(\mathbf{u}, \mathbf{v}) = -R^A_B(\mathbf{v}, \mathbf{u}) \quad \text{or} \quad R^A_{BCD} = -R^A_{BDC} . \quad (\text{A.97})$$

This is valid for the curvature tensor based on any connection on the manifold.

Next comes a pair of relations that require the connection to be torsion-free. However, it does not have to be metric-compatible, nor does the manifold have to be endowed with a metric to start with. For vanishing torsion, the second of the Bianchi identities (A.72) reduces to $R^A_B \wedge e^{*B} = 0$. This translates to

$$\begin{aligned}
R^A{}_B(\mathbf{u}, \mathbf{v})w^B + R^A{}_B(\mathbf{v}, \mathbf{w})u^B + R^A{}_B(\mathbf{w}, \mathbf{u})v^B &= 0, \\
R^A{}_{BCD} + R^A{}_{CDB} + R^A{}_{DBC} &= 0,
\end{aligned} \tag{A.98}$$

which is incidentally also known as the (first or algebraic) Bianchi identity. Likewise, the first line of (A.72) leads to

$$\begin{aligned}
(\nabla_{\mathbf{u}} R)^A{}_B(\mathbf{v}, \mathbf{w}) + (\nabla_{\mathbf{v}} R)^A{}_B(\mathbf{w}, \mathbf{u}) + (\nabla_{\mathbf{w}} R)^A{}_B(\mathbf{u}, \mathbf{v}) &= 0, \\
(\nabla_E R)^A{}_{BCD} + (\nabla_C R)^A{}_{BDE} + (\nabla_D R)^A{}_{BEC} &= 0,
\end{aligned} \tag{A.99}$$

known as the second or differential Bianchi identity. To prove this in a coordinate-free fashion requires certain amount of manipulation combining (A.72) with (A.50) and the identity $\nabla_{\mathbf{u}} \mathbf{v} - \nabla_{\mathbf{v}} \mathbf{u} = [\mathbf{u}, \mathbf{v}]$, which follows from (A.71) for vanishing torsion.

In presence of a Riemannian metric, the curvature tensor of the LC connection (the *Riemann curvature tensor*) has further properties that make it particularly symmetric. The starting point for uncovering these properties is the relation

$$\mathbf{u}[\langle \mathbf{v}, \mathbf{w} \rangle] = \hat{\nabla}_{\mathbf{u}} \langle \mathbf{v}, \mathbf{w} \rangle = \langle \hat{\nabla}_{\mathbf{u}} \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{v}, \hat{\nabla}_{\mathbf{u}} \mathbf{w} \rangle. \tag{A.100}$$

This follows at once from the Leibniz rule for the covariant derivative and the fact that the LC connection is metric-compatible. By a repeated application of (A.100), it is straightforward to show using (A.68) that

$$\langle \mathbf{z}, R(\mathbf{u}, \mathbf{v})\mathbf{w} \rangle = -\langle \mathbf{w}, R(\mathbf{u}, \mathbf{v})\mathbf{z} \rangle. \tag{A.101}$$

Here $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}$ are arbitrary vector fields and $R(\mathbf{u}, \mathbf{v})\mathbf{w}$ is a shorthand notation for $e_A R^A{}_B(\mathbf{u}, \mathbf{v})w^B$. Upon lowering the first index of the curvature tensor with the metric, (A.101) acquires a particularly simple component form,

$$R_{ABCD} = -R_{BACD}. \tag{A.102}$$

Finally, combining (A.102) with the algebraic Bianchi identity (A.98) shows that the type-(0, 4) tensor R_{ABCD} is symmetric under swapping its two pairs of indices,

$$\langle \mathbf{z}, R(\mathbf{u}, \mathbf{v})\mathbf{w} \rangle = \langle \mathbf{v}, R(\mathbf{w}, \mathbf{z})\mathbf{u} \rangle \quad \text{or} \quad R_{ABCD} = R_{CDAB}. \tag{A.103}$$

A.6.4 Geodesic Normal Coordinates

While the general language of differential geometry is largely coordinate-free, it is often convenient to resort to specific coordinates tailored to the problem at hand. For instance, in Euclidean space, the global Cartesian coordinates play a privileged role. On a general Riemannian manifold, one can similarly introduce coordinates x^a that are “locally Cartesian” (see Chap. 5 of [7] for details). These are defined around an arbitrarily chosen point $x_0 \in \mathcal{M}$ and have the following properties. First, the point x_0 itself is mapped to the origin, that is $x_0^a = 0$. Second, the Riemannian metric

at x_0 reduces to the standard Euclidean metric, $g_{ab}(x_0) = \delta_{ab}$. (Any other choice amounts to a mere change of basis of $\mathcal{T}_{x_0}^*\mathcal{M}$.) Third, all the first derivatives $\partial_c g_{ab}$ of the metric, and thus also the Christoffel symbols $\hat{\Gamma}_{bc}^a$ of the LC connection, vanish at x_0 . Finally, geodesics, that is solutions to (A.83), passing through x_0 are straight lines $\gamma^a(t) = v^a t$ with some fixed vector $v \in \mathcal{T}_{x_0}\mathcal{M}$.

Such *normal coordinates* can be constructed explicitly, if somewhat formally, in a neighborhood of x_0 using the last of the properties listed above. One chooses arbitrarily a basis e_a of the tangent space $\mathcal{T}_{x_0}\mathcal{M}$. For any vector $v \equiv v^a e_a \in \mathcal{T}_{x_0}\mathcal{M}$, one then denotes as γ_v the geodesic satisfying the initial conditions $\gamma^a(0) = 0$ and $d\gamma^a(0)/dt = v^a$. The normal coordinates of the point $\gamma_v(t)$ are defined as $x^a = v^a t$.

The normal coordinates make it easy to check some differential-geometric identities involving covariant derivatives with respect to the LC connection. For instance, (A.47) is immediately seen to be equivalent to

$$(\mathcal{L}_v \mathbb{T})_{mn\dots}^{ab\dots} = v^l (\hat{\nabla}_l \mathbb{T})_{mn\dots}^{ab\dots} + \mathbb{T}_{ln\dots}^{ab\dots} (\hat{\nabla}_m v)^l + \mathbb{T}_{ml\dots}^{ab\dots} (\hat{\nabla}_n v)^l + \dots \\ - \mathbb{T}_{mn\dots}^{lb\dots} (\hat{\nabla}_l v)^a - \mathbb{T}_{mn\dots}^{al\dots} (\hat{\nabla}_l v)^b - \dots \quad (\text{A.104})$$

Indeed, at any chosen point x_0 , the normal coordinates make the Christoffel symbols vanish so that (A.104) agrees with (A.47). Since both sides of (A.104) are well-defined tensors on \mathcal{M} , the identity must then hold for any choice of local coordinates.

Example A.17

For an illustration, consider the Lie derivative of a Riemannian metric,

$$(\mathcal{L}_v g)_{ab} = g_{cb} (\hat{\nabla}_a v)^c + g_{ac} (\hat{\nabla}_b v)^c. \quad (\text{A.105})$$

Here the first term in (A.104) drops owing to the fact that the metric is covariantly constant. This immediately recovers the equivalence of the two formulations of the Killing equation in (A.92).

Perhaps even more importantly, normal coordinates allow us to systematically compute $g_{ab}(x)$ as a power series in x^a and the Riemann curvature tensor and its covariant derivatives at x_0 . Up to fourth order in the coordinates, the series reads

$$g_{ab}(x) = g_{ab}(0) - \frac{1}{3} x^c x^d R_{acbd}(0) - \frac{1}{6} x^c x^d x^e (\hat{\nabla}_e R)_{acbd}(0) \\ - \frac{1}{20} x^c x^d x^e x^f (\hat{\nabla}_e \hat{\nabla}_f R)_{acbd}(0) + \frac{2}{45} x^c x^d x^e x^f R_{acgd}(0) R_{ebf}^g(0) + \dots, \quad (\text{A.106})$$

where I lowered the first index of R^a_{bcd} using the metric. The derivation of (A.106) is technical and I will therefore skip details. An interested reader will find a justification and a compilation of further results in an accessible form in [8].

A *locally symmetric space* is a special type of a Riemannian manifold for which the first covariant derivative of the Riemann curvature tensor (hence also all higher

covariant derivatives) vanishes. See Chap. 10, in particular Theorem 10.19, of [7]. For such manifolds, $g_{ab}(x)$ can be expressed in a closed form in terms of the curvature tensor at x_0 . Again skipping details, the final result reads

$$g_{ab}(x) = g_{ac}(0) \left[\frac{\sin^2 \sqrt{\hat{R}(x)}}{\hat{R}(x)} \right]_b^c = g_{ac}(0) \sum_{k=0}^{\infty} \frac{(-1)^k 2^{2k+1}}{(2k+2)!} [\hat{R}(x)^k]_b^c, \quad (\text{A.107})$$

where $\hat{R}^a_b(x) \equiv R^a_{cbd}(0)x^c x^d$.

A.6.5 Hodge Star

Using the \sharp isomorphism, the inner product induced by the metric can be extended from vector fields to 1-forms. Thus, for any two 1-forms ω and σ , we set

$$\langle \omega, \sigma \rangle \equiv \langle \omega^\sharp, \sigma^\sharp \rangle = g_{AB} \omega^\sharp{}^A \sigma^\sharp{}^B = g^{AB} \omega_A \sigma_B. \quad (\text{A.108})$$

This in turn extends to two differential forms ω, σ of an arbitrary (equal) degree p ,

$$\langle \omega, \sigma \rangle \equiv \frac{1}{p!} g^{A_1 B_1} \dots g^{A_p B_p} \omega_{A_1 \dots A_p} \sigma_{B_1 \dots B_p}. \quad (\text{A.109})$$

To see what this inner product of differential forms might be good for, we first pick an orthonormal frame e_A . Then we define a new top-dimensional form (n -form) in terms of the corresponding dual coframe,

$$\text{vol} \equiv e^{*1} \wedge \dots \wedge e^{*n}. \quad (\text{A.110})$$

For any other local frame, $\tilde{e}_A = e_B (P^{-1})^B_A$ with $P \in \text{GL}(n)$, we get $\text{vol}(\tilde{e}_1, \dots, \tilde{e}_n) = \det P^{-1}$. This corresponds to the (oriented) volume of the parallelepiped defined by the vectors \tilde{e}_A . The n -form (A.110) is known as the *volume form*, hence the notation “vol.” Provided that the two local frames have the same orientation, that is $\det P$ is positive, the volume form can be expressed in the new dual coframe \tilde{e}^{*A} as

$$\text{vol} = \det P^{-1} \tilde{e}^{*1} \wedge \dots \wedge \tilde{e}^{*n} = \sqrt{\tilde{g}} \tilde{e}^{*1} \wedge \dots \wedge \tilde{e}^{*n}. \quad (\text{A.111})$$

Here \tilde{g} is the determinant of the matrix representation of the metric in the new coframe, $\tilde{g}_{AB} = (P^{-1T} P^{-1})_{AB}$.

The frame e_A is in general defined only locally. We thus need to apply the definition (A.110) to an atlas of open sets that covers the manifold \mathcal{M} . In order that the volume form is well-defined and nonvanishing on the whole of \mathcal{M} , we have to make sure that local frames on mutually overlapping charts have the

same orientation. The existence of such an atlas and a set of equally oriented local frames is the defining property of an *orientable manifold*.

I now get to the main idea of this subsection. At any point $x \in \mathcal{M}$, the spaces $\Omega_x^p \mathcal{M}$ and $\Omega_x^{n-p} \mathcal{M}$ of p -forms and $(n-p)$ -forms have the same dimension, $\binom{n}{p}$, and thus are isomorphic. In order to define a natural isomorphism between these spaces, and by extension between differential p -forms and $(n-p)$ -forms, we need both the inner product (A.109) and the volume form (A.110). A given p -form ω is then mapped to the unique $(n-p)$ -form $\star \omega$ such that

$$\sigma \wedge (\star \omega) = \langle \sigma, \omega \rangle \text{vol} \quad (\text{A.112})$$

for any test p -form σ . The inner product is needed to make (A.112) linear in both ω and σ . Likewise, the volume form serves as a reference n -form to compare $\sigma \wedge (\star \omega)$ to. The linear operator \star is known as the *Hodge star*, and $\star \omega$ is referred to as the *Hodge dual* of ω . In the special case of $p = 0$, (A.112) gives a simple coordinate-free definition of the volume form, $\text{vol} = \star 1$.

The definition (A.112) is coordinate-free and elegant, but also not very transparent. For practical purposes, it is almost always more convenient to use instead the component expression for the Hodge dual,

$$\star \omega = \frac{\sqrt{g}}{p!(n-p)!} g^{A_1 B_1} \dots g^{A_p B_p} \omega_{A_1 \dots A_p} \varepsilon_{B_1 \dots B_n} e^{*B_{p+1}} \wedge \dots \wedge e^{*B_n}. \quad (\text{A.113})$$

Here $\varepsilon_{A_1 \dots A_n}$ is the fully antisymmetric LC symbol, equal to +1 when A_1, \dots, A_n is an even permutation of $1, \dots, n$, to -1 when A_1, \dots, A_n is an odd permutation of $1, \dots, n$, and to 0 otherwise. It is not to be confused with the *Levi-Civita tensor*, $\sqrt{g} \varepsilon_{A_1 \dots A_n}$, which by (A.111) gives the components of the volume form.

It is clear that by applying the Hodge star \star twice, we map a p -form ω to another p -form, $\star \star \omega$. What is less clear but easy to check is that the two forms are equal up to a sign,⁵

$$\star \star \omega = (-1)^{p(n-p)} \omega, \quad \text{or equivalently} \quad \star^{-1} \omega = (-1)^{p(n-p)} \star \omega. \quad (\text{A.114})$$

Being an isomorphism on the Grassmann algebra of differential forms, the Hodge star can be used to conjugate other linear maps defined on forms. The most important example of such a conjugate operator is the *codifferential* δ . This is defined on p -forms by conjugating d with an extra conventional minus sign,

$$\delta \omega \equiv (-1)^p \star^{-1} d \star \omega = (-1)^{np+n+1} \star d \star \omega. \quad (\text{A.115})$$

It follows from the conjugation and from (A.52) that $\delta \circ \delta = 0$. However, unlike d itself, the codifferential is not an antiderivation of the Grassmann algebra.

⁵ Here and occasionally in the following, I drop the \circ symbol indicating composition of operators to prevent the equations from becoming awkward.

It is possible to conjugate in the same way the interior product (A.17). This does not lead to a new interesting operator though. Indeed, combining (A.17) with (A.113), it is straightforward to check that for any p -form ω and vector field \mathbf{v} ,

$$(-1)^p \star^{-1} \iota_{\mathbf{v}} \star \omega = \mathbf{v}^b \wedge \omega . \quad (\text{A.116})$$

Finally, one can similarly conjugate the Lie derivative $\mathcal{L}_{\mathbf{v}}$. Thanks to the Cartan magic formula (A.57), the result is fixed by the conjugation of d and $\iota_{\mathbf{v}}$. It is however not particularly illuminating and I will therefore not show further details.

Let us now get back to the codifferential. This together with d furnishes a powerful toolkit that generalizes much of the differential structure familiar from ordinary vector calculus. The exterior derivative itself generalizes the notion of a gradient of scalar fields in \mathbb{R}^n and curl of vector fields in \mathbb{R}^3 , as noted in Example A.10. This leads to a far-reaching generalization of integral theorems in vector calculus that I will address in Sect. A.7. In \mathbb{R}^n we also have the divergence operator that turns a vector field into a scalar. We can construct such a map on any orientable manifold \mathcal{M} by combining the flat \flat and the codifferential. Thus, for any vector field \mathbf{v} ,

$$\text{div } \mathbf{v} \equiv -\delta \mathbf{v}^b = \frac{1}{\sqrt{g}} \partial_a (\sqrt{g} v^a) = (\hat{\nabla}_a \mathbf{v})^a , \quad (\text{A.117})$$

where g now stands for the determinant of the matrix representation of the metric in the dual coordinate basis dx^a . In the last expression in (A.117), the LC connection (A.84) is implicitly assumed. The next-to-last expression in (A.117) is a special case of the component form of the codifferential of a p -form,

$$(\delta \omega)_{a_2 \dots a_p} = -\frac{1}{\sqrt{g}} \partial_{b_1} [\sqrt{g} (\omega^\sharp)^{b_1 \dots b_p}] g_{a_2 b_2} \dots g_{a_p b_p} . \quad (\text{A.118})$$

This follows directly from (A.113) and (A.115) upon some manipulation. Last but not least, there is another, rather elegant way of computing the divergence of a vector field through $(\text{div } \mathbf{v}) \text{ vol} = \mathcal{L}_{\mathbf{v}} \text{ vol}$. This is an infinitesimal version of the familiar fact from multivariate calculus that upon a transformation of coordinates, the volume measure in an integral changes by the Jacobian of the transformation.

Example A.18

The component expression (A.118) makes it easy to see why the codifferential is not an antiderivation of the Grassmann algebra. Should δ satisfy an identity analogous to (A.55), we would have, for instance, for any 0-form f and p -form ω that

$$\delta(f \wedge \omega) = \delta f \wedge \omega + f \wedge \delta \omega = f \wedge \delta \omega . \quad (\text{A.119})$$

Here I used the fact that $\delta f = 0$ for any 0-form f , which is a consequence of the fact that d vanishes on any n -form. But (A.119) cannot be correct, since (A.118) makes it clear that $\delta(f \wedge \omega)$ must depend on both f and its derivatives.

Example A.19

The Noether current of a continuous symmetry of a physical system is a divergence-free vector field \mathbf{J} , that is $\delta \mathbf{J}^b = 0$. The latter condition is naturally generalized to $\delta J = 0$ for any “ p -form current” J . Such a generalized conservation law is via Noether’s theorem associated with a so-called $(p-1)$ -form symmetry [9]. Additional insight follows from writing the conservation condition $\delta J = 0$ equivalently as $d \star J = 0$. Namely, if J is a conserved p -form current and K a conserved q -form current, then the exterior derivative of the $(2n - p - q)$ -form $\star J \wedge \star K$ vanishes thanks to (A.55). Hence $\star(\star J \wedge \star K)$ is a $(p + q - n)$ -form current that is also conserved. In this way, we can use already known generalized conservation laws to construct new ones as long as $p + q > n$ [10].

Finally, by combining d and δ , we can construct a “second-order differential operator” which maps p -forms to p -forms. While it would in principle be possible to consider separately $d \circ \delta$ and $\delta \circ d$, it is their sum that is of special interest,

$$\Delta \equiv (d + \delta)^2 = d \circ \delta + \delta \circ d . \quad (\text{A.120})$$

It follows at once from (A.117) that on 0-forms this reduces to

$$\Delta f = -\frac{1}{\sqrt{g}} \partial_a (\sqrt{g} g^{ab} \partial_b f) = -g^{ab} \hat{\nabla}_a \hat{\nabla}_b f , \quad (\text{A.121})$$

where the second expression again requires the LC connection. Thus, Δ is a generalization of the Laplace operator to differential forms on orientable Riemannian manifolds. It is known as the *Laplace–de Rham operator*, or alternatively as the *Hodge Laplacian*.

A.7 Integration on Manifolds

So far I have discussed solely the differential structure on manifolds and related concepts. It is finally time to address the other essential ingredient of calculus: integration. Following the analogy with multivariate calculus in \mathbb{R}^n , one might want to integrate functions on a given manifold \mathcal{M} . Without the privilege of having globally well-defined Cartesian coordinates, it however turns out more natural to integrate forms. I will show how to do this in the next subsection. The abstract notion of integration of forms can nevertheless be bypassed in case the manifold possesses a metric structure. It is then possible to sweep the differential form nature of integration under the rug. This results into a definition of integration in terms of a volume measure determined by the metric, which is the subject of Sect. A.7.2.

A.7.1 Orientable Manifolds

A well-defined concept of integration on a manifold must have a geometric meaning independent of the choice of coordinates. It makes no sense to just pick random coordinates x^a and define an integral of a function $f : \mathcal{M} \rightarrow \mathbb{R}$ via something like

$$\int_{\mathcal{M}} f \equiv \int_{\varphi(\mathcal{M})} (f \circ \varphi^{-1})(x) d^n x \quad (\text{wrong}), \quad (\text{A.122})$$

where $\varphi : \mathcal{M} \rightarrow \mathbb{R}^n$ is the (possibly only locally defined) coordinate map. This naive attempt cannot be fixed even by adding an integration measure that would correctly account for changes of coordinates. The problem is that without further structure on the manifold, there is no unique measure we could use for the purpose.

It turns out that differential forms have all the ingredients we need. Let us consider an n -form ω on an n -dimensional manifold \mathcal{M} . In chosen local coordinates, it is given by (A.13),

$$\omega = \frac{1}{n!} \omega_{a_1 \dots a_n} dx^{a_1} \wedge \dots \wedge dx^{a_n} = \omega_{1 \dots n} dx^1 \wedge \dots \wedge dx^n. \quad (\text{A.123})$$

Under the change of coordinates from x^a to $\tilde{x}^a(x)$, the single independent component $\omega_{1 \dots n}(x)$ transforms by the inverse of the Jacobian of the coordinate transformation, $\tilde{\omega}_{1 \dots n}(\tilde{x}) = |\partial \tilde{x} / \partial x|^{-1} \omega_{1 \dots n}(x)$. It therefore appears that all we have to do is strip off the factor $dx^1 \wedge \dots \wedge dx^n$ in (A.123) and integrate $\omega_{1 \dots n}(x)$ as a function on \mathbb{R}^n . Upon a change of coordinates, the transformation of $\omega_{1 \dots n}$ will exactly cancel the Jacobian generated by the Euclidean volume measure $d^n x$.

There are two flies in the ointment though. First, a closer look reveals that while $\omega_{1 \dots n}$ indeed changes by the inverse of the Jacobian of the coordinate transformation, the Euclidean measure $d^n x$ changes by its absolute value. Another manifestation of the same problem is that $d^n x$ is invariant under any permutation of the coordinates, whereas $dx^1 \wedge \dots \wedge dx^n$ and thus $\omega_{1 \dots n}$ changes sign under odd permutations. Yet in other words, $(dx^1 \wedge \dots \wedge dx^n)(v_1, \dots, v_n)$ returns the *oriented* volume of the parallelepiped defined by the test vectors v_1, \dots, v_n in \mathbb{R}^n . The second, related problem is that we rarely have a set of globally defined coordinates on the manifold. It is mandatory to ensure that we can consistently sew together integrals of ω over individual coordinate patches. This requires that the coordinate systems in overlapping patches have equal orientation. For that to be possible at all, the manifold \mathcal{M} must be orientable. Even on orientable manifolds, we still have the freedom to choose global orientation. Fixing the orientation requires that we restrict coordinate transformations to those with a positive Jacobian. At the end of the day, the integral of an n -form will be defined up to a sign, depending on the choice of global orientation.

The problem of how to sew together integrals of ω over individual coordinate patches requires more careful consideration. The idea is to “triangulate” the manifold in a way that each cell of the triangulation lies in a single coordinate patch. To implement this strategy, we start by defining what such an elementary cell corresponds to in \mathbb{R}^n . A p -simplex (P_0, \dots, P_p) , where $0 \leq p \leq n$, is the convex envelope of

(that is the smallest convex set containing) the set of points $P_0, \dots, P_p \in \mathbb{R}^n$. For the triangulation approach to integration to be successful, we next need to be able to put together two or more simplexes. This is done by defining a set, C_p , whose elements, called *p-chains*, are formal linear combinations of different *p-simplexes*.

The introduction of *p-simplexes* was motivated by the desire to triangulate a given manifold \mathcal{M} . It would therefore appear most natural to only consider linear combinations of simplexes with coefficients ± 1 . It will however turn out convenient to allow for arbitrary real coefficients, thus making C_p a real vector space. The advantage of doing so will become apparent in Sect. A.8 where I introduce the dual structure to *p-chains* in terms of differential forms.

For integration of differential forms, we need to define an orientation of the *p-simplex*. We can do this by identifying $(P_{\pi(0)}, \dots, P_{\pi(p)})$ with $(\text{sgn } \pi)(P_0, \dots, P_p)$ for any permutation π on the indices $0, \dots, p$. With this convention, we can now also define an oriented boundary of a *p-simplex*. This is formalized by the *boundary operator*, which is a linear map $\partial_p : C_p \rightarrow C_{p-1}$ such that

$$\partial_p(P_0, \dots, P_p) \equiv \sum_{i=0}^p (-1)^i (P_0, \dots, P_{i-1}, P_{i+1}, \dots, P_p) . \quad (\text{A.124})$$

The fundamental property of the boundary operator is that when applied twice in succession, it gives zero, $\partial_{p-1} \circ \partial_p = 0$.

Eventually, we would like to construct a triangulation of a manifold \mathcal{M} by lifting simplexes from \mathbb{R}^n to \mathcal{M} . To that end, it is sufficient to consider just one, *standard p-simplex* s_p in \mathbb{R}^n . This is defined to have P_0 at the origin and all the other vertices on the positive Cartesian semiaxes at unit distance from the origin. The Cartesian coordinates of the vertex P_i are thus $P_i^a = \delta_i^a$. Now we can take any smooth map $f : s_p \rightarrow \mathcal{M}$. Its image, $\sigma_p \equiv f(s_p)$, is called a *singular p-simplex* on \mathcal{M} . By extension, a *singular p-chain* c on \mathcal{M} is a formal linear combination,

$$c = \sum_i c_i \sigma_p^i = \sum_i c_i f^i(s_p) \quad \text{where } c_i \in \mathbb{R} , \quad (\text{A.125})$$

and σ_p^i are singular *p-simplexes*. Finally, the boundary of a singular *p-simplex* is defined by lifting the boundary of s_p with the same map f , $\partial_p \sigma_p \equiv f(\partial_p s_p)$.⁶

After the rather long series of formal steps, we are finally ready to define integration on the manifold \mathcal{M} . In fact, we can do even more than we originally anticipated, namely integrate *p-forms* on any singular *p-chain* on \mathcal{M} such as (A.125), with any $0 \leq p \leq n$. The task is first reduced to individual singular *p-simplexes* by linearity. Then, the integral is pulled back to \mathbb{R}^p using the map defining the singular *p-simplex*. At the end of the day, one sets for any *p-form* ω and singular *p-chain* c

⁶ Here I am tacitly extending the action of f from the standard *p-simplex* to linear combinations of *p-simplexes* in \mathbb{R}^n by linearity.

$$\int_c \omega \equiv c_i \int_{\sigma_p^i} \omega, \quad \int_{\sigma_p^i} \omega \equiv \int_{s_p} f^{i*} \omega \equiv \int_{s_p} (f^{i*} \omega)_{1\dots p}(x) \, d^p x. \quad (\text{A.126})$$

To achieve the original goal to define integration of forms on manifolds, it remains to clarify how to represent an n -dimensional manifold \mathcal{M} as a singular n -chain. There is obviously considerable freedom in doing so, but to identify \mathcal{M} with an n -chain $c = \sum_i c_i \sigma_n^i$, the latter must satisfy several conditions. First, the simplexes σ_n^i must cover the whole of \mathcal{M} without overlaps. Second, each σ_n^i should lie in a single coordinate patch. Third, if the local coordinate system x^a in which σ_n^i is mapped to the standard n -simplex s_n is compatible with the global orientation of the manifold \mathcal{M} , we should set $c_i = 1$. If the orientation is opposite, we have to set $c_i = -1$. In practice, all this is easier than it might look, as the following example shows.

Example A.20

The spherical coordinates (θ, φ) on S^2 take values from the open rectangle $(0, \pi) \times (0, 2\pi)$ in \mathbb{R}^2 and cover the whole of S^2 except for one selected meridian. The integral of the 2-form $\omega = \sin \theta \, d\theta \wedge d\varphi$ (A.22) then equals

$$\int_{S^2} \omega = \int_0^{2\pi} d\varphi \int_0^\pi \sin \theta \, d\theta = 4\pi, \quad (\text{A.127})$$

which is the area of a unit sphere. This example shows that it is not strictly necessary to actually triangulate the integration domain. It is sufficient to dissect it into open sets, each of which is mapped by a single coordinate chart to a polytope in \mathbb{R}^n . The subsequent triangulation of the polytopes is a formal step that need not be carried out explicitly. The example also demonstrates that we can get away with a single coordinate chart even if it does not cover the whole manifold. It is sufficient to cover it up to a subset of lower dimension.

With all the necessary definitions out of the way, let me proceed directly to the most important result of integral calculus on manifolds: the *Stokes theorem*. Let c be a singular $(p+1)$ -chain on an orientable manifold \mathcal{M} , and let ω be a p -form. Then

$$\int_c d\omega = \int_{\partial_{p+1} c} \omega. \quad (\text{A.128})$$

For a proof, see for instance Sect. 6.1 of [2]. The Stokes theorem puts under the same umbrella, and further generalizes, a number of well-known results from ordinary calculus and vector calculus. These include the fundamental theorem of calculus in \mathbb{R} , the gradient theorem for line integrals in \mathbb{R}^n , the Green theorem for area integrals in \mathbb{R}^2 , the eponymous Stokes theorem for surface integrals in \mathbb{R}^3 , and the Gauss theorem for volume integrals in \mathbb{R}^n .

There is an intriguing connection between the Stokes theorem and a geometric definition of the exterior derivative that I alluded to at the end of Sect. A.4. Namely,

the exterior derivative of a p -form ω can be defined in a coordinate-free manner through integration of ω over an “infinitesimal” p -chain that is the boundary of some $(p+1)$ -chain. This generalizes the usual definition of the divergence of a vector field in \mathbb{R}^3 using its flux through the surface of an infinitesimal volume element. The key problem is then to assert the consistency of such a definition; see Sect. 36 of [11] for details. In this sense, the exterior derivative can be defined exactly so that the Stokes theorem (A.128) holds. Rather than a nontrivial result of integral calculus, the theorem may then be viewed as arising from remarkable ingenuity in defining the right differential structure on the manifold.

Example A.21

In (A.126), the integral of a p -form ω is obtained by pulling ω back to \mathbb{R}^p using the map $f : s_p \rightarrow \mathcal{M}$ defining the singular p -simplex σ_p . This idea can be generalized to maps between any two manifolds (see Sect. 7.8 of [3]). Consider two manifolds \mathcal{M} and \mathcal{N} and a smooth map $f : \mathcal{M} \rightarrow \mathcal{N}$. Then, for any p -form ω on \mathcal{N} and a singular p -chain c on \mathcal{M} , we have

$$\int_{f(c)} \omega = \int_c f^* \omega . \quad (\text{A.129})$$

Such integration of differential forms via pull-back is common in field theory. In this context, f can be thought of as a field defined on the domain \mathcal{M} and taking values in the target manifold \mathcal{N} .

Suppose now that \mathcal{M} is an n -dimensional manifold without boundary and ω is an n -form on \mathcal{N} such that $d\omega = 0$. Then $\int_{\mathcal{M}} f^* \omega$ is a *topological invariant*, that is, it does not change value under a smooth deformation of f . To see why, think of the deformation $f \rightarrow \tilde{f}$ as a family of functions f_t where $t \in [0, 1]$ is a deformation parameter and $f_0 = f, f_1 = \tilde{f}$. Formally, f_t defines a smooth map from $\tilde{\mathcal{M}} \equiv [0, 1] \times \mathcal{M}$ to \mathcal{N} . Since \mathcal{M} itself does not have boundary, we can view $\tilde{\mathcal{M}}$ as a cylinder whose “bases” at $t = 0$ and $t = 1$ are diffeomorphic to \mathcal{M} . Using the Stokes theorem, we then find

$$\begin{aligned} \int_{\mathcal{M}} \tilde{f}^* \omega - \int_{\mathcal{M}} f^* \omega &= \int_{\mathcal{M}} f_1^* \omega - \int_{\mathcal{M}} f_0^* \omega \\ &= \int_{\partial_{n+1} \tilde{\mathcal{M}}} f_t^* \omega = \int_{\mathcal{M}} d(f_t^* \omega) = 0 . \end{aligned} \quad (\text{A.130})$$

In the last step, I used the fact that exterior derivative commutes with pull-back (A.56) and the assumption that $d\omega = 0$. The invariance of $\int_{\mathcal{M}} f^* \omega$ under smooth deformations of f can also be understood from the left-hand side of (A.129). Here the integrand ω is fixed, but the integration domain $f(\mathcal{M})$ is deformed. Upon tuning the parameter t from 0 to 1, $f_t(\mathcal{M})$ sweeps a singular $(n+1)$ -chain in \mathcal{N} with boundaries $f_0(\mathcal{M})$ and $f_1(\mathcal{M})$. Invoking the Stokes theorem and the assumption $d\omega = 0$ then again asserts the invariance of $\int_{f(\mathcal{M})} \omega$ under smooth deformations of f .

A.7.2 Riemannian Manifolds

On a first encounter, the differential form approach to integration on manifolds may seem a bit abstract. There is however a conceptually simpler alternative in case the manifold \mathcal{M} is endowed with a Riemannian metric. As explained in Sect. A.6.5, the metric induces a natural volume form (A.110) on the manifold. For any smooth function $f : \mathcal{M} \rightarrow \mathbb{R}$, $f \text{ vol}$ is then a well-defined n -form on \mathcal{M} , which can be used to define the integral of f over \mathcal{M} ,

$$\int_{\mathcal{M}} f \equiv \int_{\mathcal{M}} f \text{ vol} = \int_{\mathcal{M}} f(x) \sqrt{g(x)} d^n x . \quad (\text{A.131})$$

The last, coordinate expression follows from (A.111) and is often used in physics to perform integration in curved space(time)s. The path leading to this expression can also be viewed as a natural consequence of Hodge duality, which maps the 0-form f to the n -form $\star f = f \star 1 = f \text{ vol}$.

In mathematics, it is common to define a volume form on an n -dimensional manifold \mathcal{M} as any n -form that is nonvanishing everywhere on \mathcal{M} . One then proves that the existence of a volume form is equivalent to the requirement that \mathcal{M} be orientable (see Sect. 6.3 of [3]). The definition of a volume form I gave in Sect. A.6.5 is more restrictive in that it is tied to the metric structure on \mathcal{M} . Such volume form is normalized to unity in an orthonormal frame. It can thus be thought of as defining on \mathcal{M} locally a generalization of the usual volume measure in \mathbb{R}^n .

Let me append two remarks of caution for more rigorously-minded readers. First, the coordinates may be defined only locally. In such a case, one has to partition the manifold and apply (A.131) separately to each coordinate patch. Second, the expression (A.131) may appear less general than (A.126), which is well-defined for a p -form and a singular p -chain of any degree $0 \leq p \leq n$. It is however easy to generalize (A.131) to integration of functions on any submanifold \mathcal{N} of \mathcal{M} . All one has to do is pull the metric g on \mathcal{M} back to \mathcal{N} using the embedding of \mathcal{N} in \mathcal{M} . The integral over \mathcal{N} is then defined as in (A.131) using the induced metric on \mathcal{N} .

Example A.22

I will illustrate the use of (A.131) by calculating the length of a curve on a manifold. The curve is defined by a smooth map $\gamma : \mathbb{R} \rightarrow \mathcal{M}$ and we are interested in the length of the segment $\Gamma \equiv \{\gamma(t) \mid t \in [t_1, t_2]\}$. Treating t as a coordinate on the curve, we pull the metric g on \mathcal{M} back to Γ via

$$\gamma^* g = g_{ab} \frac{d\gamma^a}{dt} \frac{d\gamma^b}{dt} dt \otimes dt . \quad (\text{A.132})$$

The length of the segment of interest then follows as

$$\int_{\Gamma} 1 = \int_{t_1}^{t_2} \sqrt{g_{ab} \frac{d\gamma^a}{dt} \frac{d\gamma^b}{dt}} dt . \quad (\text{A.133})$$

This agrees with the distance along the curve (A.82), previously postulated ad hoc.

A.8 Homology and Cohomology

Most of our discussion of differential geometry so far was restricted to local properties of manifolds. However, we have already seen how the global topology of the manifold \mathcal{M} may place restrictions on otherwise locally defined structures. For instance, there may be no globally well-defined frame on \mathcal{M} . Also, the global topology of \mathcal{M} may forbid some solutions of the Killing equation and thus reduce the isometry group of \mathcal{M} ; see Example A.16.

Topology is a vast area of mathematics and I have no intention to cover even its very basics in a self-contained manner. Instead, I will conclude the appendix with a brief introduction to some aspects of topology, pertinent to the subject of this book. As stressed in Sect. 15.3, spontaneous symmetry breaking is intimately related to the presence of defects in the ordered medium. The classification of defects was historically the first major application of topology to physics. The classic paper by Mermin [12] gives a physicist-friendly introduction to the relevant branch of topology, called *homotopy theory*. See also Chap. 4 of [2] for a somewhat more mathematical exposition. In the present book, homotopy theory plays a marginal role. A reader that is altogether unfamiliar with the subject will get away with the basic definition of homotopy groups, found early on in the above references. What I will cover here are elements of *de Rham cohomology*. This neatly connects the global topology of the manifold \mathcal{M} to the analytic properties of differential forms on \mathcal{M} . Cohomology theory plays an important role in the construction of actions for systems with spontaneously broken symmetry. It is therefore highly relevant for Parts III and IV of the book.

A.8.1 Singular Homology

Let me start somewhat indirectly by recalling the concept of a singular chain, introduced in Sect. A.7.1. The real vector spaces of singular p -chains on \mathcal{M} , denoted from now on more precisely as $C_p(\mathcal{M}, \mathbb{R})$, form a sequence under the action of the boundary operator (A.124),

$$\cdots \xrightarrow{\partial_{p+2}} C_{p+1}(\mathcal{M}, \mathbb{R}) \xrightarrow{\partial_{p+1}} C_p(\mathcal{M}, \mathbb{R}) \xrightarrow{\partial_p} C_{p-1}(\mathcal{M}, \mathbb{R}) \xrightarrow{\partial_{p-1}} \cdots . \quad (\text{A.134})$$

The spaces $C_p(\mathcal{M}, \mathbb{R})$ are infinite-dimensional. It is however possible to identify certain finite-dimensional spaces descending from $C_p(\mathcal{M}, \mathbb{R})$ that carry information about the topology of \mathcal{M} . To that end, we first introduce the subspace of *p-cycles*, $Z_p(\mathcal{M}, \mathbb{R}) \equiv \ker \partial_p$, that is those *p*-chains that have no boundary. Likewise, the subspace of *p-boundaries*, $B_p(\mathcal{M}, \mathbb{R}) \equiv \text{im } \partial_{p+1}$, consists of those *p*-chains that are boundaries of some $(p+1)$ -chains. Due to the fundamental property $\partial_p \circ \partial_{p+1} = 0$, any *p*-boundary is a *p*-cycle. In other words, $B_p(\mathcal{M}, \mathbb{R})$ is a subspace of $Z_p(\mathcal{M}, \mathbb{R})$. The quotient space,

$$H_p(\mathcal{M}, \mathbb{R}) \equiv Z_p(\mathcal{M}, \mathbb{R}) / B_p(\mathcal{M}, \mathbb{R}) , \quad 0 \leq p \leq n , \quad (\text{A.135})$$

is called the *p-th homology group* of \mathcal{M} . Miraculously, this space has a finite dimension although both $Z_p(\mathcal{M}, \mathbb{R})$ and $B_p(\mathcal{M}, \mathbb{R})$ are infinite-dimensional. Roughly speaking, $H_p(\mathcal{M}, \mathbb{R})$ is a formal vector space spanned on mutually inequivalent *p*-dimensional submanifolds of \mathcal{M} that have no boundary and are themselves not the boundary of any $(p+1)$ -dimensional submanifold of \mathcal{M} .

Example A.23

To see how the homology groups relate to the topology of \mathcal{M} , it is good to have a look at some very simple examples. First, the homology groups of the 2-sphere are

$$H_0(S^2, \mathbb{R}) \simeq \mathbb{R} , \quad H_1(S^2, \mathbb{R}) \simeq 0 , \quad H_2(S^2, \mathbb{R}) \simeq \mathbb{R} . \quad (\text{A.136})$$

The $H_0(S^2, \mathbb{R})$ group reflects the fact that S^2 is connected. Namely, 0-chains on S^2 are points on S^2 and their formal linear combinations. For any two points $P_1, P_2 \in S^2$, the formal difference $P_1 - P_2$ is the boundary of a 1-chain (curve) connecting the points. Hence any 0-chain equals, up to adding a 0-boundary, a multiple of a single point on S^2 . At the same time, any 0-chain is trivially a 0-cycle. It follows that $H_0(S^2, \mathbb{R})$ is one-dimensional. An obvious extension of the argument shows that in general, $H_0(\mathcal{M}, \mathbb{R}) \simeq \mathbb{R}^k$, where k is the number of connected components of \mathcal{M} .

The triviality of $H_1(S^2, \mathbb{R})$ reflects the fact that there is no 1-cycle (closed curve) on S^2 that would not be the boundary of some 2-chain (area) on S^2 . Finally, as to $H_2(S^2, \mathbb{R})$, there is obviously one 2-cycle on S^2 that is not a 2-boundary, namely the sphere itself. The fact that $H_2(S^2, \mathbb{R}) \simeq \mathbb{R}$ formalizes the observation that any 2-cycle on S^2 can be deformed to S^2 or its multiple. I will add without proof that the above results for the homology groups of the 2-sphere readily generalize to the *n*-sphere with any positive *n*,

$$H_0(S^n, \mathbb{R}) \simeq H_n(S^n, \mathbb{R}) \simeq \mathbb{R} , \quad H_p(S^n, \mathbb{R}) \simeq 0 \quad \text{for any } 0 < p < n . \quad (\text{A.137})$$

Let us look at another simple example, the two-dimensional torus T^2 , for which

$$H_0(T^2, \mathbb{R}) \simeq \mathbb{R} , \quad H_1(T^2, \mathbb{R}) \simeq \mathbb{R}^2 , \quad H_2(T^2, \mathbb{R}) \simeq \mathbb{R} . \quad (\text{A.138})$$

The reasoning behind the zeroth and second homology groups is the same as for the 2-sphere. The two independent generators of $H_1(T^2, \mathbb{R})$ arise from the two possibilities

to wind a circle around the torus so that its “inside” does not lie on the torus. Just like for the sphere, (A.138) is but a special case of a more general statement valid (without proof) for any n -torus with positive n ,

$$H_p(T^n, \mathbb{R}) \simeq \mathbb{R}^{\binom{n}{p}} \quad \text{for any } 0 \leq p \leq n. \quad (\text{A.139})$$

As a side remark, the dimensions of the homology groups can be combined into a single quantity characterizing the topology of the manifold: the *Euler characteristic*,

$$\chi(\mathcal{M}) \equiv \sum_{p=0}^n (-1)^p \dim H_p(\mathcal{M}, \mathbb{R}). \quad (\text{A.140})$$

For the n -sphere and the n -torus, we have $\chi(S^n) = 1 + (-1)^n$ and $\chi(T^n) = 0$.

The above example provides basic geometric intuition behind the concept of *singular homology*. By no means does it explain how to actually find the homology groups for a manifold that is less trivial to visualize than a sphere. The problem how to compute homology groups belongs to algebraic topology. In physics-oriented literature, one usually encounters the older concept of *simplicial homology*. This makes it possible, although not straightforward, to find the homology groups by a “triangulation” of the manifold. An interested reader will find more details in Chap. 3 of [2]. I will instead proceed directly to the connection of homology to the Grassmann algebra of differential forms on \mathcal{M} .

A.8.2 De Rham Cohomology

How can the topological notion of homology be connected to differential forms? The basic idea is that any p -form ω on \mathcal{M} can be thought of as a linear function on singular p -chains on \mathcal{M} ,

$$\omega : c \rightarrow (c, \omega) \equiv \int_c \omega, \quad c \in C_p(\mathcal{M}, \mathbb{R}). \quad (\text{A.141})$$

Thus, ω can be identified with an element of the dual space $C_p^*(\mathcal{M}, \mathbb{R})$ of $C_p(\mathcal{M}, \mathbb{R})$. Moreover, we can set up a structure on differential forms that is “dual” to (A.134),

$$\dots \xleftarrow{d} \Omega^{p+1} \mathcal{M} \xleftarrow{d} \Omega^p \mathcal{M} \xleftarrow{d} \Omega^{p-1} \mathcal{M} \xleftarrow{d} \dots, \quad (\text{A.142})$$

where $\Omega^p \mathcal{M}$ denotes the space of differential p -forms on \mathcal{M} . Following the analogy with (A.134), we then define the space of p -cocycles, $Z^p(\mathcal{M}) \equiv \ker d|_{\Omega^p \mathcal{M}}$, as consisting of p -forms with vanishing exterior derivative. Such forms are usually called *closed*. Likewise, the space of p -coboundaries, $B^p(\mathcal{M}) \equiv \text{im } d|_{\Omega^{p-1} \mathcal{M}}$, consists of those p -forms that equal the exterior derivative of some $(p-1)$ -form. Such

forms are called *exact*. Due to the fundamental property (A.52), every exact form is automatically closed, hence $B^p(\mathcal{M})$ is a subspace of $Z^p(\mathcal{M})$. The quotient space,

$$H^p(\mathcal{M}) \equiv Z^p(\mathcal{M})/B^p(\mathcal{M}), \quad 0 \leq p \leq n, \quad (\text{A.143})$$

is known as the *p-th de Rham cohomology group* of \mathcal{M} . According to the *Poincaré lemma*, every closed p -form ω can be locally written as the exterior derivative of a $(p-1)$ -form σ , $\omega = d\sigma$; see Chap. 9 of [3] for a proof and numerous examples. The de Rham cohomology classifies obstructions to the global validity of the Poincaré lemma. It therefore gives us access to global topological information about the manifold \mathcal{M} .

To see the connection between de Rham cohomology and (singular) homology, consider a p -cycle c and a closed p -form ω . The Stokes theorem (A.128) guarantees that the integral of ω over c does not change if we shift ω by an exact p -form, or c by a p -boundary. In other words, for any $(p-1)$ -form σ and any $(p+1)$ -chain d ,

$$\begin{aligned} \int_{c+\partial d} (\omega + d\sigma) &= \int_c \omega + \int_c d\sigma + \int_{\partial d} \omega + \int_{\partial d} d\sigma \\ &= \int_c \omega + \int_{\partial c} \sigma + \int_d d\omega + \int_d d(d\sigma) = \int_c \omega, \end{aligned} \quad (\text{A.144})$$

where I for simplicity dropped the subscript on the boundary operator. What this says is that the integral $\int_c \omega$ only depends on the equivalence class of c as an element of $H_p(\mathcal{M}, \mathbb{R})$, and on the equivalence class of ω as an element of $H^p(\mathcal{M})$. The pairing (c, ω) , defined by (A.141), therefore represents a bilinear function on $H_p(\mathcal{M}, \mathbb{R}) \times H^p(\mathcal{M})$. The *de Rham theorem* asserts that when the manifold \mathcal{M} is compact, this bilinear map is nondegenerate. This means that (c, ω) vanishes for all closed p -forms ω if and only if c is a p -boundary, and for all p -cycles c if and only if ω is exact. It follows at once that the spaces $H_p(\mathcal{M}, \mathbb{R})$ and $H^p(\mathcal{M})$ are isomorphic. The bilinear map (c, ω) establishes a duality between (singular) homology and (de Rham) cohomology.

The statement of duality is highly nontrivial. It only applies to the quotient spaces $H_p(\mathcal{M}, \mathbb{R})$ and $H^p(\mathcal{M})$ and not separately to the spaces of p -cycles and p -cocycles (closed p -forms), or p -boundaries and p -coboundaries (exact p -forms). Luckily, out of all the different classes of p -chains, it is the homology group $H_p(\mathcal{M}, \mathbb{R})$ that is of greatest interest. The de Rham theorem then allows us to compute the homology groups of the manifold by means of the exterior calculus of differential forms.

Example A.24

There are no (-1) -forms on any manifold, and hence no exact 0-forms. Thus, $H^0(\mathcal{M})$ consists of all closed 0-forms. The condition $df = 0$ for $f \in \Omega^0 \mathcal{M}$ that should be

closed is satisfied if and only if f is a piecewise-constant function, that is a constant function on every connected component of \mathcal{M} . As a consequence, $H^0(\mathcal{M})$ counts the connected components of \mathcal{M} . This is in accord with our previous result for $H_0(\mathcal{M}, \mathbb{R})$ and the de Rham theorem.

For an illustration of the cohomology groups of higher degree, let us consider the 2-sphere S^2 . We know from Example A.23 and the de Rham theorem that $H^2(S^2) \simeq H_2(S^2, \mathbb{R}) \simeq \mathbb{R}$, which guarantees the existence of one linearly independent 2-form that has a nonzero integral over S^2 . This can be naturally chosen as the volume form (A.22). On the other hand, the first de Rham cohomology group of S^2 is trivial since $H_1(S^2, \mathbb{R})$ is. We conclude that on the 2-sphere, the Poincaré lemma holds globally for 1-forms; any closed 1-form is necessarily exact.

Somewhat more interesting than the 2-sphere is the 2-torus T^2 . Here we deduce from Example A.23 that $H^1(T^2) \simeq H_1(T^2, \mathbb{R}) \simeq \mathbb{R}^2$. It is easy to guess what the two independent nontrivial closed 1-forms might be. Just think of T^2 as the Cartesian product $S^1 \times S^1$. This gives natural angular coordinates on T^2 , $\theta^{1,2} \in (0, 2\pi)$. The forms $d\theta^1, d\theta^2$ are closed but not exact since the angular coordinates are not globally well-defined. A different way to view this is to think of the basis of $H_1(T^2, \mathbb{R})$, c_a , as two independent circles that wind around the torus. Then $\omega^a \equiv d\theta^a/(2\pi)$ is the corresponding dual basis of $H^1(T^2)$ in the sense that $(c_a, \omega^b) = \delta_a^b$. As to $H^2(T^2) \simeq H_2(T^2, \mathbb{R}) \simeq \mathbb{R}$, it is likewise easy to guess that its single generator can be taken as $d\theta^1 \wedge d\theta^2$, that is the volume form on the torus.

It is straightforward to generalize the above observation to the n -torus with any positive n . All we need to do is think of T^n as $(S^1)^{\times n}$ and introduce a set of n angular variables θ^a , $a = 1, \dots, n$. We then expect that the p -th de Rham cohomology group of T^n is generated by the set of p -forms of the type $d\theta^{a_1} \wedge \dots \wedge d\theta^{a_p}$, where a_1, \dots, a_p is a set of different indices. This indicates that $H^p(T^n) \simeq \mathbb{R}^{\binom{n}{p}}$, which allows us to understand the corresponding result (A.139) for homology.

Example A.25

Example A.21 shows that closed forms can be used to construct topological invariants. One considers a smooth map $f : \mathcal{M} \rightarrow \mathcal{N}$ defined on the n -dimensional manifold \mathcal{M} and taking values from the target manifold \mathcal{N} . If \mathcal{M} has no boundary and ω is a closed n -form on \mathcal{N} , then the integral $\int_{\mathcal{M}} f^* \omega$ does not change under smooth deformations of f . But if ω is exact, then so is $f^* \omega$ and its integral over \mathcal{M} necessarily vanishes by the Stokes theorem. Nontrivial topological invariants can therefore only arise from nontrivial cohomology equivalence classes on \mathcal{N} .

A simple example of a topological invariant is provided by the volume form (A.22) on S^2 . Let us consider a function defined on \mathbb{R}^2 and taking values on S^2 . Following the notation of Example A.7, I will denote the function using a unit vector, $\mathbf{n} : \mathbb{R}^2 \rightarrow S^2$. In physics, \mathbf{n} may represent for instance a configuration of a two-dimensional (anti)ferromagnet. Suppose now that our function tends to a constant at infinity. We can then treat the infinity as a single point, which effectively turns \mathbb{R}^2 into a

compact manifold without boundary. (It is common to say that we have topologically compactified \mathbb{R}^2 to S^2 .) Using the expression for the volume form on S^2 in terms of \mathbf{n} and pulling it back to \mathbb{R}^2 , we get the topological invariant

$$w[\mathbf{n}] = \frac{1}{4\pi} \int_{\mathbb{R}^2} \mathbf{n} \cdot (\partial_x \mathbf{n} \times \partial_y \mathbf{n}) \, dx \, dy, \quad (\text{A.145})$$

where x, y are the Cartesian coordinates in \mathbb{R}^2 . The normalization factor $1/(4\pi)$ is conventional and ensures that $w[\mathbf{n}]$ only takes integer values. This is the *Brouwer degree* of \mathbf{n} , treated as a map $S^2 \rightarrow S^2$.

Let me conclude the appendix with a lightning summary of some further interesting properties of de Rham cohomology. A curious reader will find more details in Sect. 6.4 of [2]. For an n -dimensional compact manifold \mathcal{M} , let us define a pairing of a p -form ω and an $(n-p)$ -form σ as the bilinear map

$$(\omega, \sigma) \rightarrow \int_{\mathcal{M}} \omega \wedge \sigma. \quad (\text{A.146})$$

It is easy to see that if \mathcal{M} has no boundary and both ω, σ are closed, then the result of the integration depends only on the respective equivalence classes of ω, σ as elements of $H^p(\mathcal{M})$ and $H^{n-p}(\mathcal{M})$. Moreover, (A.146) as a bilinear map $H^p(\mathcal{M}) \times H^{n-p}(\mathcal{M}) \rightarrow \mathbb{R}$ turns out to be nondegenerate. It follows that the spaces $H^p(\mathcal{M})$ and $H^{n-p}(\mathcal{M})$ are isomorphic. This fact is known as the *Poincaré duality*.

Example A.26

We already saw the isomorphism $H^p(\mathcal{M}) \simeq H^{n-p}(\mathcal{M})$ at work in Example A.24. The Poincaré duality however also allows us to draw some generally valid conclusions. For instance, for any compact connected manifold \mathcal{M} without boundary, it tells us that $H^n(\mathcal{M}) \simeq H^0(\mathcal{M}) \simeq \mathbb{R}$. Let us unpack the meaning of this statement in the special case that the manifold \mathcal{M} is Riemannian. Then any n -form ω is represented as $\omega = \star f$ in terms of some function f on \mathcal{M} . The representative of the single nontrivial equivalence class of n -cycles can be taken as \mathcal{M} itself. By the de Rham theorem, ω is therefore exact if and only if $\int_{\mathcal{M}} f = 0$. The exactness implies that $\omega = \star f = d\sigma$ for some $(n-1)$ -form σ . This leads to

$$f = \star d\sigma = (-1)^n \delta \star \sigma = (-1)^{n-1} \operatorname{div}(\star \sigma)^\sharp, \quad (\text{A.147})$$

where I used (A.114) plus the definitions of codifferential and divergence. In plain terms, the function f is a divergence of some vector field if and only if the integral of f over \mathcal{M} vanishes. This generalizes a statement frequently used in physics.

Sometimes, the manifold in question has the structure of a Cartesian product, $\mathcal{M} \times \mathcal{N}$. Any differential form ω on \mathcal{M} can be lifted to $\mathcal{M} \times \mathcal{N}$. Formally, this is

done using the projection $\pi_M : M \times N \rightarrow M$ which maps a point $(x, y) \in M \times N$ to $x \in M$. Then, for any $\omega \in \Omega^q M$, $\pi_M^* \omega$ is a q -form on $M \times N$. Analogously, forms on N are lifted to $M \times N$ using the projection $\pi_N : M \times N \rightarrow N$. Now for any closed q -form ω on M and any closed r -form σ on N , $\pi_M^* \omega \wedge \pi_N^* \sigma$ is a closed $(q+r)$ -form on $M \times N$. Moreover, the cohomology equivalence class of $\pi_M^* \omega \wedge \pi_N^* \sigma$ depends only on those of ω and σ . We can thus obtain some information about de Rham cohomology groups of $M \times N$ if we know those of M and N . What is less trivial is that all cohomology generators on $M \times N$ can in fact be recovered in this way. This is expressed by the *Künneth formula*

$$H^p(M \times N) \simeq \bigoplus_{q+r=p} [H^q(M) \otimes H^r(N)], \quad (\text{A.148})$$

where the \otimes symbol denotes tensor product of vector spaces.

Example A.27

The de Rham cohomology groups of the circle, S^1 , are $H^0(S^1) \simeq H^1(S^1) \simeq \mathbb{R}$. The former follows from the fact that S^1 is connected, and the latter most easily from Poincaré duality. Using the Künneth formula (A.148), we can now recursively construct all de Rham cohomology groups of the n -torus. Let us see how this works for the 2-torus, $T^2 \simeq S^1 \times S^1$. The zeroth cohomology group is of course \mathbb{R} since the torus is connected. The first cohomology group is

$$H^1(S^1 \times S^1) \simeq [H^0(S^1) \otimes H^1(S^1)] \oplus [H^1(S^1) \otimes H^0(S^1)] \simeq \mathbb{R} \oplus \mathbb{R} \simeq \mathbb{R}^2. \quad (\text{A.149})$$

The second cohomology group is similarly

$$H^2(S^1 \times S^1) \simeq H^1(S^1) \otimes H^1(S^1) \simeq \mathbb{R} \otimes \mathbb{R} \simeq \mathbb{R}. \quad (\text{A.150})$$

With the same reasoning, one then proves by induction that the de Rham cohomology groups of the n -torus are $H^p(T^n) \simeq H^p((S^1)^{\times n}) \simeq \mathbb{R}^{\binom{n}{p}}$, as guessed in Example A.24.

There is a simple relation between de Rham cohomology groups and homotopy groups that can be utilized to deduce the former from the latter. The precise statement is provided by the *Hurewicz theorem* (see Sect. 22.3 of [4]). Let M be a compact manifold and let $\pi_p(M)$, $p > 1$ be the first nonvanishing homotopy group. Then $H^p(M)$ is the first nonvanishing de Rham cohomology group (with nonzero degree) and $H^p(M) \simeq \pi_p(M)$, if the latter is treated as a real vector space.⁷ With the additional assumption that $\pi_1(M)$ is Abelian, the same statement holds also for

⁷ This is a slight abuse of terminology. The Hurewicz theorem as presented in [4] really connects $\pi_p(M)$ as an additive Abelian group to the homology group $H_p(M, \mathbb{Z})$ with *integer* coefficients. When translating $H_p(M, \mathbb{Z})$ to $H_p(M, \mathbb{R})$, which is by the de Rham theorem in turn isomorphic to $H^p(M)$, any finite part of $H_p(M, \mathbb{Z})$ is lost.

$p = 1$. Let me illustrate the use of this theorem on an example of great relevance for the classification of effective actions in Chap. 8.

Example A.28

With a bit of homotopy theory, we can find the second de Rham cohomology group of the homogeneous space G/H , where G and H are Lie groups. What we need from homotopy theory is the exact sequence of homotopy groups (see Sect. 9 of [12]),

$$\cdots \rightarrow \pi_k(G) \rightarrow \pi_k(G/H) \rightarrow \pi_{k-1}(H) \rightarrow \pi_{k-1}(G) \rightarrow \cdots \quad (\text{A.151})$$

This implies that if both $\pi_k(G)$ and $\pi_{k-1}(G)$ are trivial, then $\pi_k(G/H) \simeq \pi_{k-1}(H)$. Let us assume that G is compact and simply connected and H is connected. Then $\pi_1(G) \simeq \pi_0(G) \simeq \{e\}$, and hence $\pi_1(G/H) \simeq \pi_0(H) \simeq \{e\}$. The homogeneous space is itself simply connected.

Next, we use the fact that $\pi_2(G)$ is trivial for any compact connected Lie group (Appendix B of [13]). Hence $\pi_2(G/H) \simeq \pi_1(H)$. But we have already established that G/H is simply connected. The Hurewicz theorem then tells us that $H^1(G/H) \simeq 0$ and $H^2(G/H) \simeq \pi_2(G/H) \simeq \pi_1(H)$. To summarize, let G be a compact and simply connected Lie group, and let H be its connected subgroup. Then the second de Rham cohomology group of the homogeneous space G/H is isomorphic to $\pi_1(H)$ as a real vector space. Since H is by assumption compact, this means that the generators of the second de Rham cohomology group of G/H are in a one-to-one correspondence with $U(1)$ factors of H [14].

References

1. M. Stone, P. Goldbart, *Mathematics for Physics: A Guided Tour for Graduate Students* (Cambridge University Press, Cambridge, UK, 2009)
2. M. Nakahara, *Geometry, Topology and Physics* (Institute of Physics Publishing, Bristol, UK, 2003)
3. M. Fecko, *Differential Geometry and Lie Groups for Physicists* (Cambridge University Press, Cambridge, UK, 2011)
4. T. Frankel, *The Geometry of Physics: An Introduction* (Cambridge University Press, Cambridge, UK, 2012)
5. J.M. Lee, *Introduction to Smooth Manifolds* (Springer Science+Business Media, New York, 2013)
6. T. Needham, *Visual Differential Geometry and Forms* (Princeton University Press, Princeton, US, 2021)
7. J.M. Lee, *Introduction to Riemannian Manifolds* (Springer International Publishing, Cham, Switzerland, 2018)
8. L. Brewin, *Class. Quant. Grav.* **26**, 175017 (2009). DOI [10.1088/0264-9381/26/17/175017](https://doi.org/10.1088/0264-9381/26/17/175017)
9. D. Gaiotto, A. Kapustin, N. Seiberg, B. Willett, *JHEP* **02**, 172 (2015). DOI [10.1007/JHEP02\(2015\)172](https://doi.org/10.1007/JHEP02(2015)172)
10. T. Brauner, *JHEP* **04**, 045 (2021). DOI [10.1007/JHEP04\(2021\)045](https://doi.org/10.1007/JHEP04(2021)045)
11. V.I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer-Verlag, New York, 1989)

12. N.D. Mermin, Rev. Mod. Phys. **51**, 591 (1979). DOI [10.1103/RevModPhys.51.591](https://doi.org/10.1103/RevModPhys.51.591)
13. S. Weinberg, *The Quantum Theory of Fields*, vol. II (Cambridge University Press, Cambridge, UK, 1996)
14. E. D'Hoker, Nucl. Phys. **B451**, 725 (1995). DOI [10.1016/0550-3213\(95\)00265-T](https://doi.org/10.1016/0550-3213(95)00265-T)