ETH Zürich · Institute for Theoretical Physics

Quantum Field Theory

Quantum Field Theory I · Quantum Electrodynamics · Quantum Field Theory II · The Physics of Electroweak Symmetry Breaking · Advanced Field Theory

Babis Anastasiou

Lecture script

babis@phys.ethz.ch

Contents

I Quantum Field Theory I 8

Preface to Part I 9

Conventions for special relativity 10

1 Why quantum field theory? 11

2 Review of principles of classical and quantum mechanics 13
- 2.1 Time evolution in classical mechanics . . . 13
- 2.2 Time evolution in quantum mechanics . . . 17
- 2.3 Conservation and symmetries in classical mechanics . . . 18
- 2.4 Symmetries in quantum mechanics. . . . 21

3 Theory of classical fields 24
- 3.1 Fields from a discretised space (lattice) . . . 24
- 3.2 Euler-Lagrange equations for a classical field from a Lagrangian density 27
- 3.3 Noether's theorem . . . 28
- 3.4 Field Hamiltonian density from discretisation . . . 38
- 3.5 An example: acoustic waves . . . 40

4 Quantisation of the Schrödinger field 42
- 4.1 The Schrödinger equation from a Lagrangian density. . . . 42
- 4.2 Symmetries of the Schrödinger field . . . 44
- 4.3 Quantisation of fields . . . 45
- 4.4 Quantised Schrödinger field . . . 46
- 4.5 Particle states from quantised fields . . . 48
- 4.6 What is the wave-function in the field quantisation formalism? . . . 52

5 The Klein-Gordon field 55
- 5.1 Real Klein-Gordon field . . . 55
- 5.2 Casimir effect: the energy of the vacuum . . . 62
- 5.3 Two real Klein-Gordon fields . . . 65
- 5.4 Conserved charges as generators of symmetry transformations . . . 71
- 5.5 Can the Klein-Gordon field be a wave-function? . . . 71

6 Quantisation of the free electromagnetic field 73
- 6.1 Maxwell equations and Lagrangian formulation . . . 73
- 6.2 Quantisation of the electromagnetic field . . . 77

7 The Dirac equation 81
- 7.1 Mathematical interlude . . . 82
- 7.2 Dirac representation of γ-matrices . . . 83
- 7.3 Traces of γ-matrices . . . 85
- 7.4 γ-matrices as a basis of 4×4 matrices . . . 86
- 7.5 Lagrangian for the Dirac field . . . 87

8 Lorentz symmetry and free fields 89
- 8.1 Field transformations and representations of the Lorentz group . . . 90

8.7 Majorana equation . . . 104

9 Classical solutions of the Dirac equation 106
9.1 Solution in the rest frame . . . 107
9.2 Lorentz boost of the rest-frame Dirac spinor along the z-axis . . . 107
9.3 Solution for an arbitrary vector . . . 110
9.4 A general solution . . . 110

10 Quantisation of the Dirac field 112
10.1 One-particle states . . . 114
10.2 Fermions . . . 118
10.3 Quantum symmetries . . . 119
10.4 Lorentz transformation of the quantised spinor field . . . 119
10.5 Parity . . . 122
10.6 Other discrete symmetries . . . 124

11 Propagation of free particles 125
11.1 Transition amplitude for the Schrödinger field . . . 125
11.2 Transition amplitude for the real Klein-Gordon field . . . 126
11.3 Time ordering and the Feynman-Stückelberg propagator for the real Klein-Gordon field . . . 130
11.4 Feynman propagator for the complex Klein-Gordon field . . . 133
11.5 Feynman propagator for the Dirac field . . . 134
11.6 Feynman propagator for the photon field . . . 135
11.7 Wick's theorem: time ordering, normal ordering and propagation . . . 135

12 Scattering theory (S-matrix) 141
12.1 Propagation in a general field theory . . . 142
12.2 Spectral assumptions in scattering theory . . . 147
12.3 "In" and "Out" states . . . 148
12.4 Scattering matrix elements . . . 150
12.5 S-matrix and Green's functions . . . 151
12.6 The LSZ reduction formula . . . 152
12.7 Truncated Green's functions . . . 155
12.8 Cross-sections* . . . 155

13 Perturbation theory and Feynman diagrams 161
13.1 Time evolution operator in the interaction picture . . . 163
13.2 Field operators in the interacting and free theory . . . 165
13.3 The ground state of the interacting and the free theory . . . 166
13.4 Feynman diagrams for ϕ^4 theory . . . 167
13.5 Feynman rules in momentum space . . . 174
13.6 Truncated Green's functions in perturbation theory . . . 176

II Quantum Electrodynamics 179

Preface to Part II 180

1 Introduction: the Standard Model and gauge symmetry 181
1.1 The Standard Model . . . 181
1.2 The wider programme . . . 182
1.3 QED . . . 183

2 Scattering in QED 187
2.1 The QED Lagrangian . . . 187
2.2 "In" and "out" states . . . 188
2.3 Perturbation theory . . . 188
2.4 Contractions and Wick's theorem . . . 189
2.5 Vacuum-to-vacuum transition . . . 190
2.6 QED Feynman rules . . . 192
2.7 Higher-order terms . . . 193
2.8 External particles . . . 193

3 Dimensional regularisation and one-loop integrals 194
3.1 Why we need a regulator . . . 194
3.2 The simplest loop integral. Wick rotation . . . 195
3.3 Dimensional regularisation. . . . 197
3.4 Expansion in ϵ and scaleless integrals . . . 202
3.5 Feynman parameters . . . 203

4 Techniques for loop integrals 207
4.1 General one-loop integrals with Feynman parameters . . . 207
4.2 Integration by parts . . . 210
4.3 Dimensionally shifted integrals . . . 211
4.4 Tensor integrals and scalar integrals . . . 212

5 Quantum electrodynamics at one loop 217
5.1 Dimensional regularisation for QED . . . 217
5.2 The electron propagator at one loop . . . 219
5.3 The electron propagator at all orders . . . 221
5.4 The photon propagator at one loop . . . 223
5.5 Ward identity* . . . 225
5.6 The photon propagator at all orders . . . 228

6 Renormalisation of QED 229
6.1 Running of the QED coupling constant and the electron mass* . . . 232

7 Higgs boson production at NLO in QCD 233
7.1 Hadron colliders and the factorisation theorem . . . 233
7.2 A minimum of QCD . . . 234
7.3 Leading order: $gg \to h$ through a top-quark loop . . . 236
7.4 The heavy-top limit: an effective Lagrangian . . . 238
7.5 The virtual corrections . . . 239
7.6 The structure of infrared divergences . . . 243
7.7 Real emission: $gg \to h + g$ . . . 244
7.8 Infrared-safe observables and the cancellation of soft poles . . . 246
7.9 Collinear factorisation and the parton densities . . . 248

7.10 The NLO cross-section . . . 249
7.11 The numerical picture . . . 250
7.12 Summary . . . 251

III Quantum Field Theory II **254**

Preface to Part III **255**

1 Path integral quantisation in Quantum Mechanics **256**
Conventions . . . 256
1.1 The propagator . . . 257
1.2 The path integral . . . 259
1.3 An "adventurous" transition . . . 261
1.4 Functional differentiation . . . 264
1.5 Vacuum to vacuum transitions . . . 265
1.6 The simple harmonic oscillator . . . 267

2 Path integrals and scalar fields **271**
2.1 Real Klein–Gordon field . . . 272
2.2 Functional integration and determinants of differential operators . . . 275
2.3 Path integrals and interacting fields . . . 277
2.4 Fermionic path integrals . . . 282

3 Non-abelian gauge theories **288**
3.1 Gauge invariance and QED . . . 288
3.2 Non-abelian (global) $SU(N)$ transformations . . . 290
3.3 Local non-abelian gauge symmetries . . . 292

4 Quantisation of non-abelian gauge theories **297**
4.1 Perturbative QCD . . . 303

5 BRST symmetry **307**
5.1 Application to the free electromagnetic field . . . 312

6 Quantum effective action and the effective potential **315**
6.1 The quantum effective action as a generating functional . . . 316
6.2 The effective potential . . . 322

7 Symmetries of the path integral and the effective action **325**
7.1 Slavnov–Taylor identities . . . 325
7.2 Slavnov–Taylor identities in QED . . . 326
7.3 Symmetry constraints on the effective action . . . 328
7.4 Constraints on the effective action from BRST: the Zinn-Justin equation . . . 330

8 Spontaneous symmetry breaking **333**
8.1 Goldstone theorem . . . 333
8.2 General broken global symmetries . . . 336
8.3 Spontaneous symmetry breaking of local gauge symmetries . . . 338

9 Renormalisation: counting the degree of ultraviolet divergences 342
9.1 Subdivergences . . . 345
9.2 Cancellation of superficial divergences with counterterms . . . 346
9.3 Nested and overlapping divergences . . . 348

10 Proof of renormalisability for non-abelian gauge theories 349

IV The Physics of Electroweak Symmetry Breaking 357

Preface to Part IV 358

1 Recap: the effective action and spontaneous symmetry breaking 359
Conventions . . . 359
1.1 The effective action . . . 359
1.2 Symmetries of the effective action . . . 361
1.3 Spontaneous symmetry breaking . . . 363
1.4 Goldstone bosons . . . 365

2 Goldstone's theorem from the spectral representation 368
2.1 Review: effective action and broken symmetries . . . 368
2.2 Conserved currents and charges . . . 368
2.3 The spectral representation of the commutator . . . 369
2.4 The Goldstone theorem . . . 371
2.5 Properties of the Goldstone bosons . . . 372
2.6 Goldstone boson decay constant and field normalisation . . . 373

3 Amplitudes with Goldstone bosons 374
3.1 Current matrix elements . . . 374
3.2 Diagrams of type III: the Goldstone pole . . . 374
3.3 Diagrams of type I: current on an external line . . . 375
3.4 Diagrams of type II: current on an internal line . . . 376
3.5 The soft Goldstone boson theorem . . . 377

4 The $O(N)$ linear sigma model 378
4.1 An example of a spontaneously broken global symmetry . . . 378
4.2 The mass spectrum . . . 379
4.3 Counting broken generators . . . 379
4.4 Goldstone boson states and the sum rule $F \cdot Z = v$ . . . 380
4.5 Interactions of Goldstone bosons . . . 381

5 Approximate symmetries and pseudo-Goldstone bosons 384
5.1 Exact symmetries and Goldstone's theorem revisited . . . 384
5.2 Explicit breaking by a small perturbation . . . 384
5.3 The vacuum alignment condition . . . 385
5.4 Existence of an aligned vacuum . . . 386
5.5 Finding the vacuum . . . 387
5.6 Pseudo-Goldstone bosons . . . 387

6 Chiral symmetry of QCD **390**
6.1 The $SU(2) \otimes SU(2)$ symmetry of the light quarks . . . 390
6.2 Generators of the symmetry and Lie algebra . . . 391
6.3 Conserved currents . . . 392
6.4 Explicit breaking of the symmetry . . . 393
6.5 Spontaneous breaking and the pions . . . 394
6.6 Weak interactions of pions . . . 395
6.7 Nucleon matrix elements of the axial current and the Goldberger–Treiman relation . . . 396

7 Pions: approximate chiral symmetry and effective field theory **398**
7.1 An approximate $SU(2) \otimes SU(2)$ symmetry . . . 398
7.2 Vacuum alignment in QCD. . . . 399
7.3 Mass of the pion . . . 400
7.4 An effective Lagrangian approach for pion interactions . . . 401
7.5 The $SO(4)$ sigma model in polar variables . . . 401
7.6 Symmetries of the Lagrangian . . . 403
7.7 The non-linear sigma model . . . 405
7.8 Linear versus non-linear representation . . . 405
7.9 Alternative formulation: the U matrix . . . 406
7.10 Goldstone bosons in general . . . 407

8 Non-renormalisable effective theories at one loop **408**
8.1 Effective non-renormalisable theories in physics . . . 408
8.2 The large Higgs mass limit of the Standard Model . . . 410
8.3 The gauged non-linear σ-model . . . 412

9 Oblique corrections and precision tests **415**
9.1 Motivation: the ρ parameter and custodial symmetry . . . 415
9.2 Formalism of oblique corrections . . . 416
9.3 Starred couplings . . . 418
9.4 The S, T, U parameters . . . 421
9.5 Examples . . . 422

10 Little Higgs models **425**
10.1 Goldstone bosons . . . 425
10.2 Introducing gauge interactions . . . 427
10.3 The simplest little Higgs model. . . . 428

V Advanced Field Theory **431**

Preface to Part V **432**

1 Dimensional regularisation and the axial anomaly **433**
Conventions . . . 433
1.1 Ultraviolet and infrared divergences . . . 434
1.2 What is an integral in d dimensions? . . . 435
1.3 The Dirac algebra in d dimensions . . . 436
1.4 Chiral symmetry of massless QED . . . 437

1.5 The axial anomaly in two dimensions . . . 438
1.6 The axial anomaly in four dimensions . . . 442

2 Chiral symmetry of the strong interactions 447
2.1 The chiral symmetries of QCD . . . 447
2.2 The linear sigma model with nucleons . . . 454
2.3 Chiral perturbation theory . . . 458
2.4 The θ vacuum . . . 463
2.5 The axion . . . 468
2.6 Instantons . . . 472

3 Topological aspects of field theory 477
3.1 Topological objects in field theory . . . 477
3.2 Vortices in the abelian Higgs model . . . 481
3.3 The Dirac monopole . . . 483
3.4 The 't Hooft-Polyakov monopole . . . 485
3.5 Phase transitions in the early universe . . . 489
3.6 Baryon and lepton number violation . . . 495
3.7 Sphalerons . . . 496
3.8 Baryogenesis . . . 500

4 (Quantum) fields in curved spacetime 504
4.1 Evolution of the universe . . . 504
4.2 Inflation . . . 507
4.3 Quantum theory of the inflaton . . . 508

A Special relativity 512
A.1 Proper time . . . 512
A.2 Subgroups of Lorentz transformations . . . 513
A.3 Time dilation . . . 515
A.4 Doppler effect . . . 516
A.5 Particle dynamics . . . 516
A.6 Energy and momentum . . . 518
A.7 The inverse of a Lorentz transformation . . . 519
A.8 Vectors and tensors . . . 520
A.9 Currents and densities . . . 521
A.10 The energy-momentum tensor . . . 522
A.11 Relativistic formulation of electrodynamics . . . 525

PART I

Quantum Field Theory I

Preface to Part I

This is the script of the first of a sequence of courses: *Quantum Field Theory I*, *Quantum Electrodynamics*, *Quantum Field Theory II*, *The Physics of Electroweak Symmetry Breaking* and *Advanced Field Theory*. The five parts of this volume follow these courses. Each part is self-contained, with its own chapter and equation numbering; equations of an earlier part are cited with the part number in front, e.g. Eq. (I.5.3).

Part I introduces the quantisation of free fields of spin 0, 1/2 and 1, the S-matrix, perturbation theory and Feynman diagrams. The last chapters of the QFT I course, on loop integrals, quantum electrodynamics at one loop and its renormalisation, are found in Part II, where they are combined with the QED lectures of the Standard Model course into one treatment of quantum electrodynamics. The bibliography at the end of the volume lists the textbooks on which the course is based.

Conventions for special relativity

Our metric convention is

$$g_{\mu\nu} = \mathrm{diag}\,(1, -1, -1, -1)\,. \tag{0.1}$$

A contravariant position four-vector is

$$x^\mu = (x^0, x^1, x^2, x^3) \equiv \left(ct, x, y, z\right) = (ct, \vec{x}). \tag{0.2}$$

A covariant position four-vector is

$$x_\mu = g_{\mu\nu} x^\nu, \tag{0.3}$$

which gives

$$x_\mu = (x_0, x_1, x_2, x_3) \equiv \left(ct, -x, -y, -z\right) = (ct, -\vec{x}). \tag{0.4}$$

Space-time derivatives form four-vectors,

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\right) = \left(\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) = \left(\frac{1}{c}\frac{\partial}{\partial t}, \vec{\nabla}\right), \tag{0.5}$$

$$\partial^\mu \equiv \frac{\partial}{\partial x_\mu} = \left(\frac{\partial}{\partial x_0}, \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right) = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}\right) = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\vec{\nabla}\right). \tag{0.6}$$

The d'Alembert operator, a Lorentz-scalar second-order differential operator, is

$$\partial^2 \equiv \partial_\mu \partial^\mu = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \vec{\nabla}^2. \tag{0.7}$$

1

Why quantum field theory?

The goal of this lecture series is to introduce a synthesis of quantum mechanics and special relativity into a unified theory, the theory of quantised fields. Rendering the two theories consistent with each other was a challenge for the physicists of the last century. Naive generalisations of the Schrödinger equation to incorporate relativity were giving non-physical results, such as particles with negative kinetic energies. Quantum field theory provided the solution to this and other problems.

Quantum field theory allows us to tackle deep questions. What is a particle? Why do particles have spin? Why do particles carry electric charge? What types of charge may exist beyond the electric charge? Why do particles have mass?

Novel phenomena emerge at high energies in collider experiments. As an example, which was studied exhaustively at the Large Electron–Positron (LEP) collider at CERN in Geneva, think of the production of a muon and its anti-particle out of the annihilation of an electron and a positron:

$$e^- + e^+ \to \mu^- + \mu^+. \tag{1.1}$$

Such a common reaction cannot be explained with quantum mechanics as we have known it so far. While we could assign a wave-function to the electron/positron system before the reaction takes place and similarly a different wave-function to the muon/anti-muon system, the Schrödinger equation does not predict that the latter is the evolution of the former.

Quantised fields, on the other hand, allow for the annihilation of particles and the creation of others, as long as this is consistent with symmetries and the corresponding conservation laws. Quantum field theory is a predictive framework. Together with symmetries, it tells us precisely how particles may interact at the shortest distances and highest energies that we have explored so far in nature.

Combined with symmetries, quantum field theory has been a remarkably predictive framework. Once we have an electron, consistency of the theory requires a positron, its anti-particle; the same holds for every other particle. The existence of a W boson required a Z boson, and the theory of the W and Z bosons required a Higgs boson. The strange quark called for a charm quark, and the violation of the symmetry between matter and anti-matter (CP violation) called for a third generation of quarks; the top quark was found almost two decades after the bottom quark, as its partner. Even quarks and leptons are not independent of each other: the consistency of the quantum theory (the cancellation of the so-called anomalies) ties each generation of quarks to a generation of leptons. Quantum field theory is a very tight framework.

Remark. Strictly speaking, some of these "requirements" hold within a given theory rather than absolutely. The Georgi–Glashow $SO(3)$ model (1972) had $W^\pm$ bosons and no Z; it was the discovery of neutral currents (1973) that selected the $SU(2) \times U(1)$ theory, in

which the Z is indeed required. Similarly, unitarity of WW scattering requires new states below about 1 TeV, and the single Higgs boson is the minimal (Standard Model) realisation (Lee, Quigg, Thacker 1977).

The theory is confronted today with measurements of an unprecedented precision. At the Large Hadron Collider (LHC), the rates for producing a Higgs boson are measured and compared with the predictions of quantum field theory, probing reactions which take place within extremely short times and distances. The two agree. Notice, however, that deriving such predictions is far from easy. For the dominant Higgs production mode, the first approximation was computed in the 1970s and it accounts for only about a third of the measured rate. Corrections computed in the 1990s turned out to be even larger than the original estimate, and further corrections were computed in the decades that followed, each one smaller than the previous one. We can write beautiful theories down, but going from a theory to a prediction is a big mathematical challenge. In this course we will learn both the concepts and the tools to make predictions out of them. Without the one, we cannot do anything useful with the other.

How far can we take this predictive power? From the measured masses of the Higgs boson and the top quark, together with other measured parameters, quantum field theory predicts the shape of the potential energy which the Higgs field stores in the whole universe. Notice that this is the energy of the field, which exists and stores energy even if no Higgs particles are present. Depending on the precise values of the parameters, our ground state (vacuum) can be the true minimum of the potential, or the potential can develop a second, deeper minimum. In the latter case, the universe can tunnel to the new vacuum, where all particle masses would be different. The present measurements place us at the border between a stable and a metastable vacuum, with a tunnelling time, in the metastable case, much longer than the age of the universe. A more precise measurement of the top-quark mass will tell us more. It is remarkable that the measurement of a few parameters together with quantum field theory allows for such statements about the fate of the universe.

Remark. This conclusion assumes that the Standard Model remains valid, without new particles or interactions, up to the very high field values (about 10^{10} GeV and above) where the potential turns over; new physics or gravitational effects at such scales can change it. With this assumption, current values of m_t, m_H and α_s place the vacuum in the metastable region (Degrassi et al. 2012; Buttazzo et al. 2013).

2

Review of principles of classical and quantum mechanics

Before we introduce quantum field theory, it will be useful to recall how we described the dynamics of simple mechanical systems in classical and quantum physics. In QFT, we will postulate principles that we have already seen there, such as the principle of least action and canonical quantisation. Conservation theorems derived in classical and quantum mechanics will also apply to QFT.

Quantum field theory is not really a new theory. It is quantum mechanics applied to fields. In classical and quantum mechanics we studied systems with a finite number of degrees of freedom; a field has an infinite number of them, and going from ten coordinates to infinitely many is just taking a limit. We will not attempt to justify the principle of least action itself. It can be exchanged with other starting points; for example, Weinberg in his book does not start from an action but from general requirements on physical scattering processes, and the principle of least action then emerges as a natural consequence. However, it cannot be eliminated completely: some starting axioms are always needed.

2.1 Time evolution in classical mechanics

Consider, for simplicity, a one-dimensional mechanical system whose dynamical behaviour is encoded in a Lagrangian

$$L\left(x(t), \dot{x}(t)\right).$$

The Lagrangian depends on the position and the velocity, which are functions of time. We will focus on energy-conserving systems, in which the Lagrangian acquires all of its time dependence through these functions and has no other explicit time dependence,

$$\frac{\partial L}{\partial t} = 0. \tag{2.1}$$

An explicit time dependence describes an external intervention, for example an experimentalist who switches on a light or "kicks" the system at some moment. Our dynamical variables will be fields and we will aim to describe them, in a maximal way, in the full universe. There is nobody outside the universe to give our fields an explicit time dependence. We let the fields do what the laws of physics dictate, and we restrict ourselves to the above simpler class of systems throughout this course.

The time evolution of the position and the velocity obeys the principle of least action. This states that the action is stationary,

$$\delta S = 0\,, \tag{2.2}$$

under small variations of the physical trajectory

$$x(t) \to x(t) + \delta x(t). \tag{2.3}$$

Eq. (2.2) leads to the Euler-Lagrange differential equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0. \tag{2.4}$$

For example, a particle moving in one dimension under the influence of a potential $V(x)$ has a Lagrangian

$$L = \frac{1}{2}m\dot{x}^2 - V(x), \tag{2.5}$$

and the corresponding Euler-Lagrange equation yields Newton's law

$$m\ddot{x} + \frac{\partial V}{\partial x} = 0. \tag{2.6}$$

An alternative description of the same dynamics can be obtained through the Hamiltonian formalism. We define first the canonical momentum,

$$p = \frac{\partial L}{\partial \dot{x}}, \tag{2.7}$$

which, in general, is a function of both the position and velocity,

$$p = p\left(x(t), \dot{x}(t)\right). \tag{2.8}$$

We now define the Hamiltonian

$$H(x, p) = p\dot{x} - L, \tag{2.9}$$

and pick the pair (x, p) as our independent variables, "inverting" Eq. (2.8), and considering the velocity as a function of position and momentum,

$$\dot{x} = \dot{x}\left(x, p\right). \tag{2.10}$$

Differentiating the Hamiltonian with respect to the position, we have

$$\frac{\partial H}{\partial x} = \ldots = -\dot{p}, \tag{2.11}$$

while differentiating with respect to the momentum we arrive at

$$\frac{\partial H}{\partial p} = \ldots = \dot{x}. \tag{2.12}$$

Therefore, as an alternative to the second-order Euler-Lagrange differential equation, we can describe the dynamical system with the pair of first-order differential equations of Hamilton,

$$\dot{p} = -\frac{\partial H}{\partial x}, \quad \dot{x} = \frac{\partial H}{\partial p}. \tag{2.13}$$

Let us now consider a generic physical quantity,

$$A = A(x, p, t). \tag{2.14}$$

Differentiating with respect to time, we have

$$\frac{dA}{dt} = \frac{\partial A}{\partial p}\dot{p} + \frac{\partial A}{\partial x}\dot{x} + \frac{\partial A}{\partial t}. \tag{2.15}$$

Substituting above the Hamilton equations (2.13), we obtain

$$\frac{dA}{dt} = \{A, H\} + \frac{\partial A}{\partial t}, \tag{2.16}$$

where the Poisson bracket for two functions $f(x,p)$ and $g(x,p)$ is given by

$$\{f, g\} \equiv \frac{\partial f}{\partial x}\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial g}{\partial x}. \tag{2.17}$$

For example, the canonical momentum for a particle in a potential, corresponding to the Lagrangian of Eq. (2.5), is

$$p = \frac{\partial L}{\partial \dot{x}} = m\dot{x}, \tag{2.18}$$

which we invert to read

$$\dot{x} = \frac{p}{m}. \tag{2.19}$$

The Hamiltonian is then computed to be

$$H = p\dot{x} - L = \frac{p^2}{2m} + V(x), \tag{2.20}$$

and Hamilton's equations are

$$\dot{x} = \frac{p}{m}, \quad \dot{p} = -\frac{\partial V}{\partial x}, \tag{2.21}$$

and the general evolution equation reads

$$\frac{dA}{dt} = \left\{A, \frac{p^2}{2m} + V(x)\right\} + \frac{\partial A}{\partial t}. \tag{2.22}$$

We will soon study the theoretical implications of this time-evolution equation for generic physical quantities A, and of the Poisson brackets.

2.1.1 Properties of Poisson brackets

We can derive the following properties of Poisson brackets from their definition in Eq. (2.17).

- Antisymmetry,
$$\{A, B\} = -\{B, A\}, \tag{2.23}$$
- Linearity,
$$\{A, B + C\} = \{A, B\} + \{A, C\}, \tag{2.24}$$
- Product rule,
$$\{A, B\,C\} = B\ \{A, C\} + \{A, B\}\ C, \tag{2.25}$$
- Action on position and momentum,
$$\{x, p\} = 1, \quad \{x, x\} = 0, \quad \{p, p\} = 0. \tag{2.26}$$

In addition, Poisson brackets satisfy the Jacobi identity

$$\{A,\{B,C\}\}+\{B,\{C,A\}\}+\{C,\{A,B\}\}=0. \tag{2.27}$$

We can also show the converse. For the class of functions of position and momentum which have a series representation as in

$$f(x,p)=\sum_{n,m=0}^{\infty} c_{nm}x^n p^m, \tag{2.28}$$

the properties of Eqs. (2.23)–(2.26) lead back to the definition of Eq. (2.17). The proof can be somewhat lengthy but relies on simple induction. The restriction to functions with a power-series representation may look severe; for example, the Coulomb potential $1/r$ is not of this form. However, in field theory, $1/r$ potentials will not be assumed. They will be derived, as a property of the electromagnetic field, whose Hamiltonian belongs to the class of Eq. (2.28).

We first prove inductively that

$$\{x,x^m\}=0, \quad \{p,p^m\}=0, \tag{2.29}$$

and

$$\{x,p^m\}=mp^{m-1}=\frac{\partial}{\partial p}p^m, \quad \{x^m,p\}=mx^{m-1}=\frac{\partial}{\partial x}x^m. \tag{2.30}$$

Then, it follows that for a generic function $f(x,p)$ of the form of Eq. (2.28), the Poisson bracket with respect to position (momentum) is equivalent to the derivative with respect to momentum (position),

$$\{x,f\}=\frac{\partial f}{\partial p}, \quad \{f,p\}=\frac{\partial f}{\partial x}. \tag{2.31}$$

Now, we can show inductively that

$$\{x^n,f(x,p)\}=nx^{n-1}\frac{\partial f}{\partial p}=\frac{\partial x^n}{\partial x}\frac{\partial f}{\partial p}, \quad \{f(x,p),p^n\}=np^{n-1}\frac{\partial f}{\partial x}=\frac{\partial p^n}{\partial p}\frac{\partial f}{\partial x}. \tag{2.32}$$

In Eq. (2.32) and in the antisymmetry, Eq. (2.23), linearity, Eq. (2.24), and product rule, Eq. (2.25), of the Poisson brackets we have all the ingredients needed to return to Eq. (2.17):

$$\begin{aligned}\{f(x,p),g(x,p)\} &= \left\{\sum_{nm}c_{nm}x^n p^m, g(x,p)\right\}=\sum_{nm}c_{nm}\{x^n p^m,g\}\\ &=\sum_{nm}c_{nm}\,x^n\{p^m,g\}+\sum_{nm}c_{nm}\{x^n,g\}p^m=\sum_{nm}c_{nm}\left[-x^n\{g,p^m\}+\{x^n,g\}p^m\right]\\ &=\sum_{nm}c_{nm}\left[-x^n\frac{\partial p^m}{\partial p}\frac{\partial g}{\partial x}+\frac{\partial x^n}{\partial x}p^m\frac{\partial g}{\partial p}\right]=\frac{\partial f}{\partial x}\frac{\partial g}{\partial p}-\frac{\partial f}{\partial p}\frac{\partial g}{\partial x}.\end{aligned} \tag{2.33}$$

2.1.2 A way to think of classical time evolution

We have thus shown that the definition of Poisson brackets in Eq. (2.17) is equivalent to the properties of Eqs. (2.23)–(2.26). We can then choose to formulate time evolution as follows.

The dynamics of a physical system is encoded in its Hamiltonian,

$$H(x_i,p_i),$$

which depends on space coordinates x_i and canonical momenta p_i. A physical quantity $A(x_i, p_i, t)$ evolves in time according to

$$\frac{dA}{dt} = \{A, H\} + \frac{\partial A}{\partial t}, \tag{2.34}$$

and the Poisson brackets are defined to satisfy

$$\{x_i, x_j\} = \{p_i, p_j\} = 0, \quad \{x_i, p_j\} = \delta_{ij}, \tag{2.35}$$

as well as to have the antisymmetry, linearity and product-rule properties of Eqs. (2.23)–(2.25). These statements are all that one needs to describe the dynamical evolution of a classical system.

2.2 Time evolution in quantum mechanics

Physical systems are in quantum states $|\psi\rangle$ of a Hilbert space. Physical quantities, such as the position and momentum, are Hermitian operators acting on this space,

$$\hat{x}_i\,, \hat{p}_i\,, \ldots$$

The mean value of experimental measurements of an observable quantity $O(x_i, p_i)$, which is represented by an operator $\hat{O}$ in Hilbert space, is given by

$$\langle O \rangle = \langle \psi | \hat{O}(\hat{x}_i, \hat{p}_i) | \psi \rangle\,. \tag{2.36}$$

Quantum time evolution has a very similar form to classical evolution, as formulated in Section 2.1.2. A central object in classical evolution is the Poisson bracket of position and momentum. What is the analogous object for the corresponding quantum mechanical operators? Dirac observed that the commutator of two operators,

$$[A, B] \equiv AB - BA,$$

possesses algebraic properties analogous to those of the classical Poisson brackets,

$$[A, B] = -[B, A]\,, \tag{2.37}$$
$$[A, B + C] = [A, B] + [A, C]\,, \tag{2.38}$$
$$[A, BC] = B\,[A, C] + [A, B]\,C\,, \tag{2.39}$$
$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0\,. \tag{2.40}$$

In classical mechanics, the Poisson bracket of a coordinate and its canonical momentum is a constant (unity). We will postulate the same for the commutator of their quantum mechanical operators,

$$\left[\hat{x}, \hat{p}\right] = \text{constant}.$$

The constant must then be imaginary, given that the positions and momenta are Hermitian,

$$\text{constant}^* = \left[\hat{x}, \hat{p}\right]^\dagger = \left[\hat{p}^\dagger, \hat{x}^\dagger\right] = -\left[\hat{x}, \hat{p}\right] = -\text{constant}. \tag{2.41}$$

In analogy to classical mechanics, it is natural to postulate the following commutation relations for positions and momenta,

$$\left[\hat{x}_i, \hat{x}_j\right] = \left[\hat{p}_i, \hat{p}_j\right] = 0\,, \quad \left[\hat{x}_i, \hat{p}_j\right] = i\hbar\delta_{ij}\,. \tag{2.42}$$

Comparing Eq. (2.42) with the Poisson-bracket "postulates" of classical mechanics in Eq. (2.35), we are led to identify the quantum mechanical analogue of the Poisson bracket with the commutator,

$$\{A, B\} \to \frac{1}{i\hbar}\left[\hat{A}, \hat{B}\right]. \tag{2.43}$$

Inspired by this analogy and Eq. (2.34), it seems natural to postulate the following time evolution for the operators of physical quantities $\hat{O}\left(\hat{x}_i, \hat{p}_i\right)$ in quantum mechanics,

$$\frac{d\hat{O}}{dt} = \frac{1}{i\hbar}\left[\hat{O}, \hat{H}\right], \tag{2.44}$$

where $\hat{H}$ is the Hamiltonian operator.

As we know, this is the correct evolution for operators (in the Heisenberg picture) in quantum mechanics. Eq. (2.44) has the solution,

$$\hat{O}(t) = e^{\frac{i}{\hbar}Ht}\hat{O}_0 e^{-\frac{i}{\hbar}Ht}. \tag{2.45}$$

The Heisenberg picture will be more convenient for studying time evolution in QFT. In this picture, the evolution equation is a one-to-one analogue of the classical evolution with Poisson brackets, which is not manifest in the Schrödinger picture. In addition, our operators will be fields. Classically, a field, such as the electromagnetic field, is a function of space and time. When we quantise it, we do not want to depart from this picture: we keep the time dependence in the fields, which are the operators. Eq. (2.45) is of course equivalent to the more commonly used Schrödinger picture. Indeed, we can change picture by shifting time evolution from the operator to the quantum states

$$\left\langle\psi_0\right|\hat{O}(t)\left|\psi_0\right\rangle = \left\langle\psi(t)\right|\hat{O}_0\left|\psi(t)\right\rangle, \tag{2.46}$$

where

$$\left|\psi(t)\right\rangle = e^{-\frac{i}{\hbar}Ht}\left|\psi_0\right\rangle. \tag{2.47}$$

By differentiating the above equation with respect to time, it is easy to see that the time-dependent states (Schrödinger picture) satisfy Schrödinger's equation,

$$i\hbar\frac{\partial}{\partial t}\left|\psi(t)\right\rangle = \hat{H}\left|\psi(t)\right\rangle. \tag{2.48}$$

2.3 Conservation and symmetries in classical mechanics

Symmetries will play a very important role in our introduction to quantum field theory and we will discuss them throughout this course. Let us recall here a few selected results from classical mechanics first. We will limit ourselves, for now, to continuous symmetry transformations of coordinates,

$$x_i \to x_i' = G_{ij}(\theta^a)x_j, \tag{2.49}$$

which leave the Hamiltonian invariant. In the above, θ^a are parameters which define the transformation (for example, for three-dimensional rotations these parameters can be chosen as $\theta^a = \{\omega_{xy}, \omega_{yz}, \omega_{zx}\}$, the rotation angles in the x-y, y-z and z-x planes). We can assume that there exists a set of values of the θ^a parameters for which the transformation leaves the coordinates unchanged (the identity transformation). Conventionally, we calibrate the θ^a's so that

$$G_{ij}(\theta^a)\big|_{\theta^a=0} = \delta_{ij}. \tag{2.50}$$

For small values of the parameters θ^a we can approximate

$$G_{ij}(\theta^a) \approx \delta_{ij} + i\theta^a J^a_{ij}, \tag{2.51}$$

where the matrices J^a_{ij} are the generators of the symmetry on coordinates. Actually, the generators are the only quantities needed in order to produce any symmetry transformation since "large" symmetry transformations can always be broken down into infinitesimal symmetry transformations for which Eq. (2.51) holds. The generators satisfy the Lie algebra

$$\left[J^a, J^b\right] = i f^{abc} J^c, \tag{2.52}$$

where f^{abc} are the structure constants of the group.

The symmetry transformation should preserve Poisson brackets for positions and their corresponding momenta,

$$\left\{x'_i, p'_j\right\} = \delta_{ij}. \tag{2.53}$$

If the momenta transform as

$$p_i \to p'_i = \tilde{G}_{ij} p_j, \tag{2.54}$$

then Eq. (2.53) yields

$$G_{ik}\tilde{G}_{jk} = \delta_{ij}. \tag{2.55}$$

For small transformations, the above equation gives

$$\tilde{G}(\theta^a)_{ij} = \delta_{ij} - i\theta^a {J^a}^T_{ij}. \tag{2.56}$$

In matrix notation, $\tilde{G} = \left(G^{-1}\right)^T$: if coordinates transform as $x \to Gx$, momenta transform as $p \to \left(G^{-1}\right)^T p$. For rotations, G is orthogonal, $G^{-1} = G^T$, and positions and momenta transform in the same way. For a generic linear transformation, however, they do not. The requirement of Eq. (2.53) is not a technicality. Poisson brackets determine time evolution; if they were different for two symmetry-related observers, Newton's law would be different for the two of them. In other words, our symmetry transformations are canonical transformations.

Let us now consider a general quantity $A(x_i, p_i)$. Under an infinitesimal symmetry transformation as described above,

$$x_i \to x'_i = x_i + \delta x_i = x_i + i\theta^a J^a_{ik} x_k, \quad p_i \to p'_i = p_i + \delta p_i = p_i - i\theta^a {J^a}^T_{ik} p_k, \tag{2.57}$$

A changes by

$$A \to A' = A + \delta A, \tag{2.58}$$

with

$$\begin{aligned}
\delta A &= \frac{\partial A}{\partial x_i}\delta x_i + \delta p_i \frac{\partial A}{\partial p_i} = \left\{A, p_i\right\}\delta x_i + \delta p_i \left\{x_i, A\right\} \\
&= \left\{A, p_i\right\}\delta x_i - \delta p_i \left\{A, x_i\right\} = \left\{A, p_i\right\}\delta x_i - \left(-i\theta^a {J^a}^T_{ik} p_k\right)\left\{A, x_i\right\} \\
&= \left\{A, p_i\right\}\delta x_i + p_k \left\{A, \left(i\theta^a {J^a}^T_{ik}\right) x_i\right\} = \left\{A, p_i\right\}\delta x_i + p_k \left\{A, \left(i\theta^a J^a{}_{ki}\right) x_i\right\} \\
&= \left\{A, p_i\right\}\delta x_i + p_k \left\{A, \delta x_k\right\} = \left\{A, p_i \delta x_i\right\}.
\end{aligned} \tag{2.59}$$

We have arrived at an important result for the effect of a symmetry transformation on a generic quantity,

$$\delta A = \left\{A, p_i \delta x_i\right\}. \tag{2.60}$$

For the Hamiltonian, which we have assumed to stay invariant under our symmetry transformations, we have that

$$\delta H = 0 \rightsquigarrow \left\{H, p_i \delta x_i\right\} = 0. \tag{2.61}$$

This leads to the well-known conclusion (Noether's theorem) that there exist conserved quantities as a consequence of the symmetry, namely,

$$\frac{d}{dt}\left(p_i \delta x_i\right) = i\frac{d}{dt}\left(\theta^a J^a_{ik} p_i x_k\right) = 0. \tag{2.62}$$

Since the θ^a's are independent, we obtain that for every symmetry generator there is a conserved quantity

$$J^a_{ik} p_i x_k = \text{constant}, \quad \forall\, a. \tag{2.63}$$

2.3.1 A classical example

As an example, consider a system with N complex coordinates $q_i,\ i = 1, \ldots, N$, which is described by the Lagrangian

$$L = m\sum_{i=1}^{N} \dot{q}^*_i \dot{q}_i - \lambda\left(\sum_{i=1}^{N} q^*_i q_i - v\right)^2, \tag{2.64}$$

One can easily check that the above Lagrangian is invariant under the transformation

$$q_i \to q'_i = U_{ij} q_j, \quad q^*_i \to q^{*\prime}_i = U^*_{ij} q^*_j, \tag{2.65}$$

where the transformation matrix U is unitary,

$$U^*_{ki} U_{kj} = \delta_{ij} \qquad (U^\dagger U = \mathbf{1}). \tag{2.66}$$

Differentiating with respect to time, we obtain the symmetry transformations of velocities,

$$\dot{q}_i \to \dot{q}'_i = U_{ij} \dot{q}_j, \quad \dot{q}^*_i \to \dot{q}^{*\prime}_i = U^*_{ij} \dot{q}^*_j. \tag{2.67}$$

Therefore, the velocities transform exactly as their corresponding coordinates.

We now compute the canonical momenta,

$$p_i = \frac{\partial L}{\partial \dot{q}_i} = m\dot{q}^*_i, \quad p^*_i = \frac{\partial L}{\partial \dot{q}^*_i} = m\dot{q}_i. \tag{2.68}$$

Notice that the canonical momenta are

$$p_i = m\dot{q}^*_i \neq m\dot{q}_i,$$

and they do not transform like their velocities,

$$p_i = m\dot{q}^*_i \to p'_i = m\dot{q}^{*\prime}_i = mU^*_{ij}\dot{q}^*_j = U^*_{ij} p_j = \left[U^\dagger\right]^T_{ij} p_j = \left[U^{-1}\right]^T_{ij} p_j. \tag{2.69}$$

Exercise: *What are the conserved quantities predicted by Noether's theorem, which originate from the $U(N)$ symmetry of this Lagrangian?*

2.4 Symmetries in quantum mechanics

A continuous symmetry transformation acts on quantum states, transforming them as

$$|\psi\rangle \to |\psi'\rangle = U(\theta^a)\,|\psi\rangle\,, \tag{2.70}$$

where θ^a are parameters of the transformation. From Wigner's theorem we know that the transformation matrix U is linear and unitary (rather than antilinear and antiunitary, which is the second allowed option of the theorem). Indeed, continuous symmetry transformations, such as rotations or translations, contain the unit transformation (no transformation) as an element of the group, which is linear and unitary. For small values of the parameters we have that

$$U(\theta^a) \approx \mathbf{1} + i\theta^a t^a, \tag{2.71}$$

where the generators of the symmetry (as represented on Hilbert space) satisfy the Lie algebra

$$\left[t^a, t^b\right] = i f^{abc} t^c. \tag{2.72}$$

We recall that the generators J^a in the "classical mechanics" representation of a symmetry transformation on coordinates, as in Eq. (2.52), are different from the generators of the representation of the symmetry on quantum states, as in Eq. (2.72). Of course, the structure constants f^{abc} are the same, as they are common to the generators of all representations.

As an example, consider rotations. On the coordinates of a vector they act with 3×3 matrices, with generators $J^a_{ij} = -i\epsilon^{aij}$ (for example, the only non-zero entries of J^3 are $J^3_{12} = -J^3_{21} = -i$). The same rotation acts on the states of a spin-$\frac{1}{2}$ system, $|\uparrow\rangle = (1,0)^T$, $|\downarrow\rangle = (0,1)^T$, with 2×2 matrices generated by $t^a = \frac{\sigma^a}{2}$, where σ^a are the Pauli matrices. The two sets of generators are qualitatively different, but both satisfy the Lie algebra with the same structure constants, $f^{abc} = \epsilon^{abc}$. The symmetry is one and the same; it is simply represented on different objects. This is the subject of representation theory, which we will meet again when we study Lorentz transformations.

From unitarity, we can easily derive that the generators are Hermitian,

$$U^\dagger U = \mathbf{1} \rightsquigarrow t^{a\dagger} = t^a, \tag{2.73}$$

which renders them good candidates for operators of physical (measurable) quantities. If we assume that the expectation value of the energy of a system in an arbitrary quantum state is invariant under the symmetry transformation,

$$\langle\psi|\,H\,|\psi\rangle = \langle\psi'|\,H\,|\psi'\rangle\,, \tag{2.74}$$

then we obtain that

$$U^\dagger(\theta^a) H U(\theta^a) = H, \tag{2.75}$$

or, equivalently, that the commutator of the symmetry generators and the Hamiltonian vanishes

$$[t^a, H] = 0. \tag{2.76}$$

On one hand, the above result implies that the generators t^a are quantum mechanical operators of conserved physical quantities,

$$i\hbar \frac{d}{dt}\, t^a = [t^a, H] = 0. \tag{2.77}$$

On the other hand, it implies, in general, that the energy spectrum is degenerate. Consider an energy eigenstate $|E\rangle$,

$$H\ |E\rangle = E\ |E\rangle\,. \tag{2.78}$$

Then, the states $(t^a\,|E\rangle)$ are also energy eigenstates,

$$[t^a, H]\,|E\rangle = 0 \rightsquigarrow H\,(t^a\,|E\rangle) = E\,(t^a\,|E\rangle)\,. \tag{2.79}$$

Remark. A degeneracy follows only if $t^a\,|E\rangle$ is not proportional to $|E\rangle$. For an abelian symmetry (e.g. a single $U(1)$ phase or a single rotation axis) the energy eigenstates can be chosen as eigenstates of t^a and no degeneracy is implied; for a non-abelian symmetry, levels are degenerate multiplets of dimension equal to that of the irreducible representation, which can be one (singlets, e.g. $\ell = 0$ states under rotations). Since the generators commute with the Hamiltonian, we can also find common eigenstates of H and (a commuting subset of) the t^a. Their eigenvalues provide extra labels, beyond the energy, which lift the degeneracy. This will be our programme in quantum field theory: take the Hamiltonian of a system of fields, find its symmetries and their generators, and construct their common eigenstates. The eigenvalues of the generators will become properties of particles; for example, the electric charge of an electron will be the eigenvalue of a symmetry generator, and its momentum the eigenvalue of another. Conversely, if a degeneracy remains after we have used all the labels that we know, it signals a symmetry that we have not yet identified.

Under a symmetry transformation, an expectation value of an operator changes to

$$\langle\psi|\,A\,|\psi\rangle \to \langle\psi'|\,A\,|\psi'\rangle = \langle\psi|\,U^\dagger A U\,|\psi\rangle\,. \tag{2.80}$$

It is then convenient to calculate the effect of a symmetry operator on expectation values or matrix elements by leaving the states unchanged and performing a similarity transformation on operators

$$A \to A' = U^\dagger A U\,, \tag{2.81}$$

which, for small symmetry transformations, gives

$$\delta A = A' - A = i\theta^a\,[A, t^a]\,. \tag{2.82}$$

We would now like to see how continuous symmetry transformations change a generic operator and compare with our classical mechanics intuition. To study the variation, we first need to introduce the notion of a "derivative" for quantum mechanical operators. We can do that by means of commutators with position and momenta. For a generic operator $A(x_i, p_i)$ we define the following derivatives

$$i\hbar\frac{\partial A}{\partial x_i} = \left[A, p_i\right], \tag{2.83}$$

and

$$i\hbar\frac{\partial A}{\partial p_i} = [x_i, A]\,. \tag{2.84}$$

You can check with explicit calculations that the right-hand sides behave, algebraically, like derivatives.

Now let us introduce a convention for writing down an operator which is a function of position and momenta. We can agree to order products with positions to the right and momenta to the left. For example, we will use the commutation relation of position

and momenta to write $A = \hat{p}\hat{x} + i\hbar$ rather than $A = \hat{x}\hat{p}$. With this convention for casting operators, a generic operator $A(x_i, p_i)$, under a symmetry transformation, varies by

$$\begin{aligned} \delta A &= \frac{\partial A}{\partial x_i}\delta x_i + \delta p_i \frac{\partial A}{\partial p_i} \\ &= \frac{1}{i\hbar}\left[A, p_i\right]\delta x_i + \frac{1}{i\hbar}\delta p_i \left[x_i, A\right]. \end{aligned} \tag{2.85}$$

It is now not difficult to repeat the steps in the classical mechanics derivation and to find the quantum mechanics analogue of Eq. (2.60),

$$i\hbar\delta A = \left[A, p_i\delta x_i\right]. \tag{2.86}$$

Analogously to classical mechanics, for $A = H$, we find that

$$\left[H, p_i\delta x_i\right] = 0 \rightsquigarrow \frac{d}{dt}\left(p_i\delta x_i\right) = 0. \tag{2.87}$$

This is (a case of) Noether's theorem in quantum mechanics. Given a symmetry, it informs us which quantum mechanical operators are conserved.

In conclusion, we have seen that by quantising a classical system in a canonical quantisation formalism, we can get guidance about what quantities are conserved from our classical studies of the system. The conserved operators commute with the Hamiltonian. Once we have identified them from symmetries, we can find their common eigenstates with the Hamiltonian.

We will apply exactly the same strategy for systems of an infinite number of generalised coordinates (fields). For free (non-interacting) fields, the common eigenstates of the Hamiltonian and the conserved quantities will be the ground state (vacuum), particle and multi-particle states.

3

Theory of classical fields

We start by introducing a Lagrangian formalism for fields at the classical level. A classical field is a continuous function defined at every point in space-time. For example, the amplitudes of a mechanical wave or of an electromagnetic wave are fields. Fields contain an infinite number of degrees of freedom, as necessitated by their existence at every point in spacetime. However, the Lagrangian formalism for them is a simple extension of, and completely analogous to, the formalism for discrete systems.

In quantum mechanics, the position of a particle is an operator, while time is a label. In quantum field theory, the field becomes the operator, while both the time t and the position $\vec{x}$ are labels (and integration variables of the action). This is a great advantage when we aim for a theory which respects special relativity: Lorentz transformations mix time and space, and with fields we can mix them without worrying that we are mixing operators with labels.

3.1 Fields from a discretised space (lattice)

The Lagrangian formalism can be generalised to fields if we first assume that a field function takes non-zero values only at discrete points of a lattice which spans all space, and then take the "continuum limit" of zero distance between the lattice points.

As a simple example, we shall study the vibrations of a one-dimensional elastic rod. The rod has a mass density μ. For an elastic rod, the force applied on it is proportional to the elongation or the compression per unit length (ξ) of the rod,

$$F = -Y\xi, \tag{3.1}$$

where Y is a constant called Young's modulus.

A rod at rest can be seen as an infinite number of equally spaced particles at rest, all having the same mass m, with a relative distance a. For a small spacing a, we have

$$\mu = \frac{dm}{dx} = \lim_{a\to 0} \frac{m}{a}. \tag{3.2}$$

A vibration is created when these particles are displaced from their positions at rest. We will assume that each particle in the lattice can interact with its immediate neighbours only. When two neighbouring particles are displaced to a larger relative distance, an attractive force tends to bring them back to their resting position, while a repulsive force is developed when they are found at a shorter relative distance. Such a force can be approximated (elastic rod) by considering an elastic spring between the particles, with spring (Hooke's)

constant κ. The force exerted by one spring is

$$\begin{aligned} F &= -\kappa\left(y_{i+1}-y_i\right) = -\left(\kappa a\right)\frac{y_{i+1}-y_i}{a} \\ &= -\left(\kappa a\right)\xi. \end{aligned} \tag{3.3}$$

We can then relate the microscopic spring constant to the macroscopic Young's modulus:

$$Y = \lim_{a\to 0}\left(\kappa a\right). \tag{3.4}$$

The potential energy in the rod is equal to the sum of the potential energies in all springs,

$$V = \sum_i \frac{1}{2}\kappa\Delta y_i^2 = \sum_i \frac{1}{2}\kappa(y_{i+1}-y_i)^2, \tag{3.5}$$

where Δy_i is the expansion or contraction of the spring between particle i and particle $i+1$, and y_i is the displacement of particle i from its position at rest. The kinetic energy of all particles is,

$$T = \sum_i \frac{1}{2}m\dot{y}_i^2. \tag{3.6}$$

The Lagrangian of the system is simply given by,

$$\begin{aligned} & L = T - V \\ \rightsquigarrow\quad & L = \sum_i\left[\frac{1}{2}m\dot{y}_i^2 - \frac{1}{2}\kappa\left(y_{i+1}-y_i\right)^2\right]. \end{aligned} \tag{3.7}$$

We can easily find the equations of motion for each particle k in the discretised rod from the Euler-Lagrange equations, Eq. (2.4),

$$\begin{aligned} & \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}_k}\right) - \frac{\partial L}{\partial y_k} = 0 \\ \rightsquigarrow\quad & \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{y}_k}\right) + \frac{\partial V}{\partial y_k} = 0 \\ \rightsquigarrow\quad & \frac{d}{dt}\left(\frac{\partial \sum_i \frac{1}{2}m\dot{y}_i^2}{\partial \dot{y}_k}\right) + \frac{\partial \sum_i \frac{1}{2}\kappa(y_{i+1}-y_i)^2}{\partial y_k} = 0 \\ \rightsquigarrow\quad & \frac{d}{dt}\left(\frac{\partial \frac{1}{2}m\dot{y}_k^2}{\partial \dot{y}_k}\right) + \frac{\partial\left[\frac{1}{2}\kappa(y_{k+1}-y_k)^2 + \frac{1}{2}\kappa(y_k-y_{k-1})^2\right]}{\partial y_k} = 0 \\ \rightsquigarrow\quad & m\ddot{y}_k + \kappa(y_k - y_{k-1}) - \kappa(y_{k+1}-y_k) = 0. \end{aligned} \tag{3.8}$$

Let us now consider the limit in which the spacing in the discretised rod tends to zero. The vibration amplitude $y_k(t)$ defines a continuous function at each position x in the one-dimensional rod:

$$\begin{aligned} & y_k(t) \to y(x,t) \\ & y_{k+1}(t) \to y(x+a,t) \\ & y_{k-1}(t) \to y(x-a,t) \\ & y_{k+2}(t) \to y(x+2a,t) \\ & y_{k-2}(t) \to y(x-2a,t) \\ & \dots \end{aligned}$$

The equations of motion are written as,

$$\begin{aligned} & m\ddot{y}(x,t) - \kappa\left[\big(y(x+a,t) - y(x,t)\big) - \big(y(x,t) - y(x-a,t)\big)\right] = 0 \\ \rightsquigarrow \quad & \frac{m}{a}\ddot{y}(x,t) - \kappa\left[\frac{y(x+a,t) - y(x,t)}{a} - \frac{y(x,t) - y(x-a,t)}{a}\right] = 0 \\ \rightsquigarrow \quad & \frac{m}{a}\ddot{y}(x,t) - (\kappa a)\,\frac{\frac{y(x+a,t)-y(x,t)}{a} - \frac{y(x,t)-y(x-a,t)}{a}}{a} = 0. \end{aligned} \tag{3.9}$$

Taking the limit $a \to 0$, we obtain the equation of motion for the vibration field $y(x,t)$,

$$\mu\frac{\partial^2 y(x,t)}{\partial t^2} - Y\frac{\partial^2 y(x,t)}{\partial x^2} = 0. \tag{3.10}$$

It is interesting to take the continuum limit in the expression of Eq. (3.7) for the Lagrangian as well. We have,

$$L \;=\; \sum_i a\left[\frac{1}{2}\frac{m}{a}\dot{y}_i^2 - \frac{1}{2}(\kappa a)\left(\frac{y_{i+1} - y_i}{a}\right)^2\right]. \tag{3.11}$$

In the continuum limit, the summation turns into an integration over the position variable x, leading to

$$L = \int dx\left[\frac{1}{2}\mu\left(\frac{\partial y(x,t)}{\partial t}\right)^2 - \frac{1}{2}Y\left(\frac{\partial y(x,t)}{\partial x}\right)^2\right]. \tag{3.12}$$

The action is then written as an integral over time and space,

$$S = \int dt dx \mathcal{L}, \tag{3.13}$$

over a *Lagrangian density*,

$$\mathcal{L} = \frac{1}{2}\mu\left(\frac{\partial y(x,t)}{\partial t}\right)^2 - \frac{1}{2}Y\left(\frac{\partial y(x,t)}{\partial x}\right)^2. \tag{3.14}$$

Can we obtain the equations of motion (Eq. (3.10)) from the action integral over the Lagrangian density of Eq. (3.13) using the principle of least action? If so, we can avoid the cumbersome discretisation derivation and work using a direct formalism for continuum systems. We require that

$$\delta\int dt dx \mathcal{L}\left[y, \frac{\partial y}{\partial t}, \frac{\partial y}{\partial x}\right] = 0, \tag{3.15}$$

if we vary the field $y(x,t)$ from its physical solution,

$$y(x,t) \to y(x,t) + \delta y(x,t). \tag{3.16}$$

We also require that the variation of the field vanishes at the boundaries of the integrations,

$$\delta y(x,t_1) = \delta y(x,t_2) = 0, \quad \delta y(x_1,t) = \delta y(x_2,t) = 0. \tag{3.17}$$

We then have,

$$\begin{aligned} & \delta\int dt dx \mathcal{L}\left[y, \frac{\partial y}{\partial t}, \frac{\partial y}{\partial x}\right] = 0 \\ \rightsquigarrow \quad & \int dt dx\left[\frac{\partial \mathcal{L}}{\partial y}\delta y + \frac{\partial \mathcal{L}}{\partial(\partial_t y)}\delta(\partial_t y) + \frac{\partial \mathcal{L}}{\partial(\partial_x y)}\delta(\partial_x y)\right] = 0. \end{aligned} \tag{3.18}$$

The field variation commutes with space and time derivatives, and we can rewrite the above equation as,

$$\begin{aligned} & \int dtdx \left[\frac{\partial \mathcal{L}}{\partial y}\delta y + \frac{\partial \mathcal{L}}{\partial(\partial_t y)}\partial_t(\delta y) + \frac{\partial \mathcal{L}}{\partial(\partial_x y)}\partial_x(\delta y) \right] = 0 \\ \rightsquigarrow \quad & \int dtdx \left\{ \left[\frac{\partial \mathcal{L}}{\partial y} - \partial_t \frac{\partial \mathcal{L}}{\partial(\partial_t y)} - \partial_x \frac{\partial \mathcal{L}}{\partial(\partial_x y)} \right] \delta y \right. \\ & \left. + \partial_t \left(\frac{\partial \mathcal{L}}{\partial(\partial_t y)}\delta y \right) + \partial_x \left(\frac{\partial \mathcal{L}}{\partial(\partial_x y)}\delta y \right) \right\} = 0. \end{aligned} \tag{3.19}$$

Since the field variation δy is generic and it vanishes at the boundaries, we obtain an Euler-Lagrange differential equation for $y(x,t)$,

$$\frac{\partial \mathcal{L}}{\partial y} - \partial_t \frac{\partial \mathcal{L}}{\partial(\partial_t y)} - \partial_x \frac{\partial \mathcal{L}}{\partial(\partial_x y)} = 0. \tag{3.20}$$

Substituting into this equation the Lagrangian density for the elastic rod, we find the following equation of motion:

$$\mu \frac{\partial^2 y(x,t)}{\partial t^2} - Y \frac{\partial^2 y(x,t)}{\partial x^2} = 0, \tag{3.21}$$

which is the same as Eq. (3.10).

3.2 Euler-Lagrange equations for a classical field from a Lagrangian density

We readily generalise the Lagrangian formalism developed for the example of the one-dimensional elastic rod to fields which are defined in four space-time dimensions $x^\mu = (x^0 = ct, x^1 = x, x^2 = y, x^3 = z)$. In this lecture series, we use natural units, setting $\hbar = c = 1$, and the space-time metric is defined as $g^{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$. We consider a system of N fields $\phi^{(i)}, i = 1,\ldots,N$ (for example, a theory of electrons and photons needs a field for each species, and even photons alone come in different polarisations, i.e. field components; the label i enumerates all fields and field components) with a Lagrangian density which depends only on the fields and their first (for simplicity) space-time derivatives,

$$\mathcal{L} = \mathcal{L}\left[\phi^{(i)}, \partial_\mu \phi^{(i)}\right]. \tag{3.22}$$

The action is given by the integral,

$$S \equiv \int_{t_1}^{t_2} dx_0 \int_V d^3\vec{x}\, \mathcal{L}\left[\phi^{(i)}, \partial_\mu \phi^{(i)}\right], \tag{3.23}$$

and we consider small variations $\phi^{(i)}(\vec{x},t) \to \phi^{(i)}(\vec{x},t) + \delta\phi^{(i)}(\vec{x},t)$ of the fields from their physical values, which vanish at the times t_1, t_2 and on the boundary surface $S(V)$ of the volume V,

$$\delta\phi^{(i)}(\vec{x},t)\big|_{t=t_1,t_2} = \delta\phi^{(i)}(\vec{x},t)\big|_{\vec{x}\in S(V)} = 0, \tag{3.24}$$

but arbitrary otherwise. Applying the variational principle, we obtain

$$\begin{aligned}
& \delta S = 0 \\
\rightsquigarrow \quad & \delta \int d^4x \mathcal{L} = 0 \\
\rightsquigarrow \quad & \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi^{(i)}} \delta\phi^{(i)} + \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi^{(i)}\right)} \delta\left(\partial_\mu \phi^{(i)}\right) \right] = 0 \\
\rightsquigarrow \quad & \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi^{(i)}} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi^{(i)}\right)} \right] \delta\phi^{(i)} + \int d^4x \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi^{(i)}\right)} \delta\phi^{(i)} \right] = 0 \\
\rightsquigarrow \quad & \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi^{(i)}} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi^{(i)}\right)} \right] \delta\phi^{(i)} = 0, \qquad (3.25)
\end{aligned}$$

where we have used the conditions for vanishing field variations at the boundaries. Given that the functional form of the field variation $\delta\phi^{(i)}$ is arbitrary, we obtain the Euler-Lagrange differential equations,

$$\frac{\partial \mathcal{L}}{\partial \phi^{(i)}} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi^{(i)}\right)} = 0. \qquad (3.26)$$

From the above derivation we can easily see that two Lagrangian densities, $\mathcal{L}$ and $\mathcal{L} + \partial_\nu F^\nu(\phi^{(i)}, \partial\phi^{(i)})$, which differ by a total divergence yield identical Euler-Lagrange equations. Indeed the variation of the total divergence term in the action integral is zero,

$$\delta \int_{V,T} d^4x \partial_\nu F^\nu = \delta \int_{S(V,T)} d\sigma_\nu F^\nu = \int_{S(V,T)} d\sigma_\nu \delta F^\nu = 0. \qquad (3.27)$$

In the above we have used Gauss' theorem and that the variation of a smooth function F of the fields and their derivatives vanishes at the boundaries of the action integral.

Remark. The boundary argument as given covers only $F^\nu = F^\nu(\phi^{(i)})$: we have only required $\delta\phi^{(i)}$, not $\delta(\partial_\mu\phi^{(i)})$, to vanish on the boundary. The statement is nevertheless true for $F^\nu(\phi, \partial\phi)$, since the Euler-Lagrange operator annihilates any total divergence identically (one checks it by direct computation; the third-derivative terms cancel). Such an F^ν makes $\mathcal{L}$ depend on second derivatives.

3.3 Noether's theorem

We can prove a very powerful theorem for Lagrangian systems which exhibit symmetries, known as *Noether's theorem*. The theorem states that for every symmetry of the system, there exists a conserved current and a corresponding conserved charge.

3.3.1 Internal field symmetry transformations

Internal symmetry transformations mix fields, or field components, among themselves without touching space-time. This may sound strange at first. What could it mean to mix an electron field with something else? In fact, there is a symmetry of the weak interactions which rotates the (left-handed) electron and its neutrino into each other; it is hidden at low energies, where the two particles have very different masses, but it becomes apparent at high energies, where masses can be neglected. Similarly, quarks come in three "colours", and the physics does not change if we rotate the colours into each other. We will meet such symmetries in a concrete form later in the course.

Consider a Lagrangian density

$$\mathcal{L}\left[\phi^{(i)}, \partial_\nu \phi^{(i)}\right], \tag{3.28}$$

and an infinitesimal symmetry transformation of the fields

$$\phi^{(i)}(x) \to \phi^{(i)}(x) + \delta\phi^{(i)}(x), \quad \partial_\nu \phi^{(i)}(x) \to \partial_\nu \phi^{(i)}(x) + \partial_\nu \delta\phi^{(i)}(x). \tag{3.29}$$

A symmetry transformation leaves the equations of motion unchanged. We call a transformation a symmetry if the Lagrangian density changes at most by a total derivative,

$$\delta\mathcal{L} = \partial_\mu X^\mu. \tag{3.30}$$

Remark. This is the definition of a (Noether) symmetry, not a consequence of the invariance of the equations of motion: a rescaling $\mathcal{L} \to (1+\epsilon)\mathcal{L}$ leaves the equations of motion unchanged but is not of the form $\partial_\mu X^\mu$ and yields no conserved current. The relation $\delta\mathcal{L} = \partial_\mu X^\mu$ must hold identically, i.e. without using the equations of motion. Notice that X^μ is not something that we are free to choose. Given the Lagrangian and the symmetry transformation, we substitute the transformed fields into the Lagrangian and compute the change; if the transformation is indeed a symmetry, the change can only be a total derivative, and this total derivative is fixed (up to the trivial ambiguity discussed next). Most symmetries that we will encounter induce no total derivative at all, $X^\mu = 0$.

Remark. Strictly speaking, X^μ is unique only up to terms $\partial_\nu A^{\mu\nu}$ with $A^{\mu\nu} = -A^{\nu\mu}$, which are identically divergence-free. They change the current J^μ by an "improvement" term but not the conserved charge (cf. the discussion of $\partial_\rho f^{\rho\nu\mu}$ at the end of this section).

The change in the Lagrangian density can be computed as,

$$\begin{aligned}
\delta\mathcal{L} &= \sum_i \left[\frac{\partial\mathcal{L}}{\partial\phi^{(i)}}\delta\phi^{(i)} + \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\partial_\nu\delta\phi^{(i)}\right] \\
&= \sum_i \left\{\delta\phi^{(i)}\left[\frac{\partial\mathcal{L}}{\partial\phi^{(i)}} - \partial_\nu\frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\right] + \partial_\nu\left(\frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\delta\phi^{(i)}\right)\right\} \\
\rightsquigarrow \delta\mathcal{L} &= \sum_i \partial_\nu\left(\frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\delta\phi^{(i)}\right) = \partial_\nu X^\nu.
\end{aligned} \tag{3.31}$$

Therefore the currents

$$J^\nu = \sum_i \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\delta\phi^{(i)} - X^\nu, \tag{3.32}$$

are conserved:

$$\partial_\nu J^\nu = 0. \tag{3.33}$$

The conserved currents J^ν are as many as the number of independent generators for the transformations $\delta\phi^{(i)}$.

To each conserved current corresponds a conserved charge, i.e. a physical quantity

which keeps the same value at all times. Indeed,

$$\begin{aligned} & \partial_\nu J^\nu = 0 \\ \rightsquigarrow \quad & \frac{\partial J^0}{\partial t} + \frac{\partial J^i}{\partial x^i} = 0 \\ \rightsquigarrow \quad & \frac{\partial J^0}{\partial t} + \vec{\nabla} \cdot \vec{J} = 0 \\ \rightsquigarrow \quad & \int d^3\vec{x} \frac{\partial J^0}{\partial t} + \int d^3\vec{x} \vec{\nabla} \cdot \vec{J} = 0 \\ \rightsquigarrow \quad & \frac{\partial Q}{\partial t} = 0 \text{ with } Q = \int d^3\vec{x} J^0. \end{aligned} \tag{3.34}$$

In the last step, the divergence term is turned into a surface integral at infinity, where we assume that the fields vanish. It is therefore important to integrate over all space. If we integrate the charge density over a finite sub-volume, the corresponding charge (for example, momentum) can escape through the boundary and it is not conserved; conservation theorems refer to the total charge everywhere in space. The conserved charges will play a central role after quantisation. They will become operators commuting with the Hamiltonian, and their eigenvalues will label particle states. In this way, internal symmetries will endow particles with an electric charge, a colour charge for quarks, or other types of charge.

3.3.2 Space-time symmetry transformations

We now consider a more complicated version of the theorem, which deals with space-time symmetries. We consider a classical system of N physical fields $\phi^{(i)}$, and we assume that the Lagrangian density does not depend explicitly on space-time coordinates:

$$\mathcal{L} = \mathcal{L}\left[\phi^{(i)}, \frac{\partial}{\partial x^\nu}\phi^{(i)}\right]. \tag{3.35}$$

However, the Lagrangian depends on the space-time coordinates implicitly, through the fields $\phi^{(i)}(x^\mu)$. This is a reasonable assumption if we require that physical laws (the equations of motion) derived from this Lagrangian are valid everywhere and always.

We now require that if we perform a space-time symmetry transformation, such as a translation or a rotation or a boost,

$$x^\mu \to x'^\mu, \tag{3.36}$$

the action remains the same:

$$S = \int_V d^4x \mathcal{L}\left[\phi^{(i)}(x), \partial_\nu \phi^{(i)}(x)\right] = \int_{V'} d^4x' \mathcal{L}\left[\phi'^{(i)}(x'), \frac{\partial}{\partial x^{\nu'}}\phi'^{(i)}(x')\right]. \tag{3.37}$$

In this way, since the action is invariant under the transformation, the variational principle $\delta S = 0$ is automatically satisfied after the transformation if it was satisfied before the transformation. We consider continuous transformations for which there exists a choice of the transformation parameters resulting in the identity transformation, i.e. no transformation. An example is a Lorentz boost with some velocity $\vec{v}$, where for $v = 0$ we have $x^\mu \to x'^\mu = x^\mu$. There are symmetry transformations for which this is not the case. For example, a parity transformation does not have this property, and Noether's theorem is then not applicable. For continuous symmetry transformations, we can consider that they are infinitesimally different from the identity transformation,

$$x_\mu \to x'_\mu = x_\mu + \delta x_\mu, \tag{3.38}$$

with δx_μ very small.

In general, the fields $\phi^{(i)}$ also transform under general space-time transformations, since they may be, for example, vectors, tensors, etc. We write, to a linear approximation,

$$\phi'^{(i)}(x') = \phi^{(i)}(x) + \Delta\phi^{(i)}(x), \tag{3.39}$$

$$\frac{\partial}{\partial x'^{\nu}}\phi'^{(i)}(x') = \frac{\partial}{\partial x^{\nu}}\phi^{(i)}(x) + \Delta\left(\frac{\partial}{\partial x^{\nu}}\phi^{(i)}(x)\right). \tag{3.40}$$

We decompose the variation of a generic function $f(x)$, which can stand for either a field or its derivative $f(x) = \phi^{(i)}(x), \partial_\nu\phi^{(i)}(x)$, into two parts,

$$\Delta f(x) \equiv f'(x') - f(x) = \left(f'(x') - f(x')\right) + \left(f(x') - f(x)\right). \tag{3.41}$$

To understand why both pieces are needed, think of a vector field, such as the vector potential of electrodynamics, and of a rotation. The rotated observer evaluates the field at the rotated point x', but it would be wrong to simply transport the original vector there: the direction of the vector must also be rotated. The difference in the second bracket of the right-hand side describes the change of the field (of its derivative) due to transporting its argument from x to x' without considering the change that the function itself should undergo in its form due to the transformation. In the linear approximation,

$$\delta f(x) = f(x') - f(x) = \left(f(x) + \partial_\mu f\,\delta x^\mu + \ldots\right) - f(x) = \partial_\mu f(x)\,\delta x^\mu + O(\delta^2). \tag{3.42}$$

The remaining change originates from changing the form of the function f. We call this difference *variation at a point*,

$$\delta_* f(x') \equiv f'(x') - f(x'). \tag{3.43}$$

We note that, in the linear approximation, we may evaluate the variation at a point of a field function at either the original or the transformed point, as the difference between the two is of quadratic order. Indeed,

$$\delta_* f(x') = \delta_* f(x) + \left(\partial_\mu\delta_* f(x)\right)\delta x^\mu + \ldots = \delta_* f(x) + O(\delta^2). \tag{3.44}$$

Explicitly, for the transformation of the fields we have

$$\Delta\phi^{(i)} = \delta_*\phi^{(i)}(x) + \delta x^\mu\partial_\mu\phi^{(i)}, \tag{3.45}$$

and, for the transformation of the field derivatives,

$$\Delta\partial_\nu\phi^{(i)} = \partial_\nu\delta_*\phi^{(i)}(x) + \delta x^\mu\partial_\mu\partial_\nu\phi^{(i)}(x). \tag{3.46}$$

The action, after the transformation, becomes

$$\begin{aligned} S &= \int_{V'} d^4x'\,\mathcal{L}\left[\phi'^{(i)}(x'), \partial_\nu\phi'^{(i)}(x')\right] \\ &= \int_V d^4x\,\det\left(\frac{\partial x'_\mu}{\partial x_\nu}\right)\mathcal{L}\left[\phi^{(i)}(x) + \Delta\phi^{(i)}(x), \partial_\nu\phi^{(i)}(x) + \Delta\left(\partial_\nu\phi^{(i)}(x)\right)\right]. \end{aligned} \tag{3.47}$$

The variation of the Lagrangian density due to the symmetry transformation can be computed as,

$$\begin{aligned} \Delta\mathcal{L} &= \mathcal{L}\left[\phi^{(i)}(x) + \Delta\phi^{(i)}(x), \partial_\nu\phi^{(i)}(x) + \Delta\left(\partial_\nu\phi^{(i)}(x)\right)\right] - \mathcal{L}\left[\phi^{(i)}(x), \partial_\nu\phi^{(i)}(x)\right] \\ \rightsquigarrow \Delta\mathcal{L} &= \sum_i\left[\frac{\partial\mathcal{L}}{\partial\phi^{(i)}}\Delta\phi^{(i)} + \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\Delta\left(\partial_\nu\phi^{(i)}\right)\right] + O\left(\left(\Delta\phi^{(i)}\right)^2\right), \end{aligned} \tag{3.48}$$

where we have expanded to first order in the variations.

The Jacobian of our space-time transformation is,

$$\det\left(\frac{\partial x'_\mu}{\partial x_\nu}\right) = \det\left(\frac{\partial\left[x_\mu + \delta x_\mu\right]}{\partial x_\nu}\right) = \det\left(\delta^\nu_\mu + \partial^\nu \delta x_\mu\right). \tag{3.49}$$

We now need the determinant of a four-by-four matrix ($\mu, \nu = 0, 1, 2, 3$). We can verify with an explicit calculation that through order $O(\delta x)$:

$$\det\left(\frac{\partial x'_\mu}{\partial x_\nu}\right) = 1 + \partial^\mu\left(\delta x_\mu\right) + O((\delta x)^2). \tag{3.50}$$

Exercise: *Verify the above explicitly.*

Substituting the expansions of Eq. (3.48) and Eq. (3.50) in the action integral of Eq. (3.47), and keeping terms through $O\left(\Delta\phi^{(i)}\right)$, we find

$$\begin{aligned} S &= S + \int d^4x \left[\Delta\mathcal{L} + \mathcal{L}\partial^\mu \delta x_\mu\right] \\ \rightsquigarrow 0 &= \int d^4x \left\{\sum_i \left[\frac{\partial\mathcal{L}}{\partial\phi^{(i)}}\Delta\phi^{(i)} + \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\Delta\left(\partial_\nu\phi^{(i)}\right)\right] + \mathcal{L}\partial^\mu\delta x_\mu\right\}. \end{aligned} \tag{3.51}$$

Substituting Eq. (3.45) and Eq. (3.46) into Eq. (3.51), we obtain:

$$\begin{aligned} 0 = &\int d^4x \sum_i \left[\frac{\partial\mathcal{L}}{\partial\phi^{(i)}}\delta_*\phi^{(i)} + \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\partial_\nu\left(\delta_*\phi^{(i)}\right)\right] \\ &+ \int d^4x \left\{\mathcal{L}\partial^\mu\delta x_\mu + \delta x^\mu \sum_i \left[\frac{\partial\mathcal{L}}{\partial\phi^{(i)}}\partial_\mu\phi^{(i)} + \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\partial_\mu\partial_\nu\phi^{(i)}\right]\right\}. \end{aligned} \tag{3.52}$$

Applying the chain rule, we see that the term in the curly brackets is a total divergence,

$$\partial_\mu\left(\mathcal{L}\delta x^\mu\right) = \mathcal{L}\partial_\mu\delta x^\mu + \delta x^\mu \sum_i \left[\frac{\partial\mathcal{L}}{\partial\phi^{(i)}}\partial_\mu\phi^{(i)} + \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\partial_\mu\partial_\nu\phi^{(i)}\right]. \tag{3.53}$$

The term in the square brackets of Eq. (3.52) can also be written as a total divergence, applying integration by parts and using the Euler-Lagrange equations,

$$\begin{aligned} &\frac{\partial\mathcal{L}}{\partial\phi^{(i)}}\delta_*\phi^{(i)} + \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\partial_\nu\left(\delta_*\phi^{(i)}\right) \\ = &\left[\frac{\partial\mathcal{L}}{\partial\phi^{(i)}} - \partial_\nu\frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\right]\delta_*\phi^{(i)} + \partial_\nu\left(\frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\delta_*\phi^{(i)}\right) \\ = &\partial_\nu\left(\frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\delta_*\phi^{(i)}\right). \end{aligned} \tag{3.54}$$

Eq. (3.52), with Eq. (3.53) and Eq. (3.54), becomes

$$\begin{aligned} 0 &= \int d^4x \left[\sum_i \partial_\nu\left(\frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\delta_*\phi^{(i)}\right) + \partial_\mu\left(\mathcal{L}\delta x^\mu\right)\right] \\ \rightsquigarrow\ 0 &= \int d^4x\,\partial_\nu\left[\sum_i \frac{\partial\mathcal{L}}{\partial\left(\partial_\nu\phi^{(i)}\right)}\delta_*\phi^{(i)} + \mathcal{L}\delta x^\nu\right]. \end{aligned} \tag{3.55}$$

This is essentially the proof of Noether's theorem. A word of caution: we usually discard total divergences in space-time integrals, assuming that the fields vanish at infinity. Here we must not do this. The principle of least action is valid for any initial and final configuration of the fields, i.e. for an arbitrary integration volume which need not extend to infinity. The volume of integration in Eq. (3.55) is considered to have boundaries which are arbitrary. Then, in order for Eq. (3.55) to hold, the integrand must vanish. This gives:

$$\partial_\nu J^\nu = 0, \tag{3.56}$$

where the conserved currents are:

$$J^\nu \equiv \sum_i \frac{\partial \mathcal{L}}{\partial \left(\partial_\nu \phi^{(i)}\right)} \delta_* \phi^{(i)} + \mathcal{L} \delta x^\nu. \tag{3.57}$$

We have therefore proven that, if the system possesses a symmetry under a space-time transformation, there exist conserved currents.

3.3.3 Energy-momentum tensor

For simplicity, we restrict ourselves to fields $\phi^{(i)}$ which transform as scalars under a space-time transformation, i.e. they do not transform:

$$\phi'^{(i)}(x') = \phi^{(i)}(x) \rightsquigarrow \Delta\phi^{(i)} = 0. \tag{3.58}$$

For this to happen, the variation of a field at a point $\delta_*\phi^{(i)}$ must compensate the variation $\delta\phi^{(i)}$ due to changing the space-time position:

$$\begin{aligned} & \Delta\phi^{(i)} = 0 \rightsquigarrow \delta_*\phi^{(i)} + \delta\phi^{(i)} = 0 \\ \rightsquigarrow \quad & \delta_*\phi^{(i)} = -\delta\phi^{(i)} \rightsquigarrow \delta_*\phi^{(i)} = -\delta x^\mu \partial_\mu \phi^{(i)}. \end{aligned} \tag{3.59}$$

A temperature field is a good example to keep in mind. If we measure 20°C at the point x, a rotated observer must also measure 20°C, but at the rotated point x'; at the point x the new temperature function will in general give a different value. Therefore, even for a scalar field the variation at a point $\delta_*\phi$ is not zero. It is precisely what is needed to compensate the variation due to the transport from x to x'.

Substituting Eq. (3.58) into Eq. (3.57) we obtain

$$\partial_\nu \left(T^{\mu\nu} \delta x_\mu\right) = 0, \tag{3.60}$$

where we have defined:

$$T^{\mu\nu} \equiv \sum_i \frac{\partial \mathcal{L}}{\partial \left(\partial_\nu \phi^{(i)}\right)} \left(\partial^\mu \phi^{(i)}\right) - \mathcal{L} g^{\mu\nu}. \tag{3.61}$$

The tensor $T^{\mu\nu}$ is known as the *energy-momentum tensor* for reasons that will become obvious shortly. The name *stress-energy tensor* is also encountered very often. In summary, a space-time symmetry of a system of scalar fields yields the conserved currents

$$\partial_\nu J^\nu = 0, \qquad J^\nu = \delta x_\mu T^{\mu\nu}. \tag{3.62}$$

We can write as many currents J^ν as there are independent space-time symmetry transformations δx_μ of the physical system.

Translation symmetry transformations

As a first application of Noether's theorem, we find the conserved currents and the corresponding charges for physical systems which are symmetric under space-time translations. An infinitesimal translation transformation is

$$x^\mu \longrightarrow x'^\mu = x^\mu + \epsilon^\mu \rightsquigarrow \delta x^\mu = \epsilon^\mu, \tag{3.63}$$

where ϵ^μ is a small constant four-vector. Which fields transform as scalars under translations? The answer is: all of them. A vector field, for example the electromagnetic vector potential, does not change its orientation if we simply shift it in space or time. Therefore Eq. (3.58) holds for every field under translations, and the energy-momentum tensor of Eq. (3.61) is applicable in general. When we call a field a scalar, a vector or a tensor, we must always say with respect to which transformation; the electromagnetic field is a vector under Lorentz transformations but a scalar under translations.

The corresponding conserved current is

$$J^\nu = T^{\mu\nu}\epsilon_\mu, \tag{3.64}$$

satisfying the continuity equation

$$\begin{aligned} & \partial_\nu J^\nu = 0 \\ \rightsquigarrow \quad & \partial_\nu T^{\mu\nu}\epsilon_\mu = 0. \end{aligned} \tag{3.65}$$

The vector ϵ^μ is small but otherwise arbitrary. The above equation is therefore satisfied in general only if

$$\partial_\nu T^{\mu\nu} = 0, \tag{3.66}$$

for every value of the index $\mu = 0, 1, 2, 3$ separately.

To the four conserved currents correspond four "conserved charges" (Eq. (3.34)),

$$P^\mu = \int d^3\vec{x}\, T^{\mu 0}. \tag{3.67}$$

Specifically, the "time-component" of P^μ is

$$\begin{aligned} P^0 &= \int d^3\vec{x}\, T^{00} \\ &= \int d^3\vec{x} \left[\sum_i \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \phi^{(i)}\right)} \left(\partial_0 \phi^{(i)}\right) - g^{00}\mathcal{L} \right]. \end{aligned} \tag{3.68}$$

In the above, we identify the conjugate momentum of the field $\phi^{(i)}$,

$$\pi^{(i)} = \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \phi^{(i)}\right)}, \tag{3.69}$$

and thus

$$P^0 = \int d^3\vec{x} \left[\sum_i \pi^{(i)} \partial_0 \phi^{(i)} - \mathcal{L} \right]. \tag{3.70}$$

In the square bracket, we recognise a Hamiltonian density,

$$\mathcal{H} = \sum_i \pi^{(i)} \partial_0 \phi^{(i)} - \mathcal{L}. \tag{3.71}$$

The conserved charge is therefore

$$P^0 = \int d^3\vec{x}\,\mathcal{H}, \tag{3.72}$$

i.e. the charge P^0 is the Hamiltonian of the system, and it is conserved. Energy conservation is thus a pure consequence of time-translation symmetry. Similarly, we can identify the charges P^i, $i = 1, 2, 3$ as the momentum components of the system in the three space directions, and they are of course also conserved. This identification is forced by special relativity. P^μ transforms as a four-vector, and a Lorentz transformation mixes its time component, which we have identified with the energy, with its space components. The only consistent interpretation of the space components of a four-vector whose time component is the energy is the momentum.

3.3.4 Lorentz symmetry transformations and conserved currents

Lorentz transformations,

$$x'^\mu = \Lambda^\mu_\nu x^\nu, \tag{3.73}$$

preserve the distance

$$ds^2 = g^{\mu\nu} dx_\mu dx_\nu = g^{\mu\nu} dx'_\mu dx'_\nu. \tag{3.74}$$

This leads to

$$\begin{aligned} g^{\mu\nu} dx_\mu dx_\nu &= g^{\mu\nu} \Lambda^\rho_\mu \Lambda^\sigma_\nu dx_\rho dx_\sigma \\ &= g^{\rho\sigma} \Lambda^\mu_\rho \Lambda^\nu_\sigma dx_\mu dx_\nu \\ \rightsquigarrow g^{\mu\nu} &= g^{\rho\sigma} \Lambda^\mu_\rho \Lambda^\nu_\sigma \quad \Longleftrightarrow \quad g^{\mu\nu} = g^{\rho\sigma} \Lambda^\mu_\rho \Lambda^\nu_\sigma. \end{aligned} \tag{3.75}$$

Considering an infinitesimal transformation,

$$\Lambda^\mu_\nu = \delta^\mu_\nu + \omega^\mu_\nu + O(\omega^2), \tag{3.76}$$

we obtain

$$\begin{aligned} g^{\mu\nu} &= g^{\rho\sigma} \Lambda^\mu_\rho \Lambda^\nu_\sigma \\ \rightsquigarrow g^{\mu\nu} &= g^{\rho\sigma} \left(\delta^\mu_\rho + \omega^\mu_\rho + \ldots\right) \left(\delta^\nu_\sigma + \omega^\nu_\sigma + \ldots\right) \\ \rightsquigarrow g^{\mu\nu} &= g^{\mu\nu} + \omega^{\nu\mu} + \omega^{\mu\nu} + O(\omega^2) \\ \rightsquigarrow \omega^{\mu\nu} &= -\omega^{\nu\mu}. \end{aligned} \tag{3.77}$$

The parameters $\omega^{\mu\nu}$ for generating Lorentz transformations are therefore antisymmetric.

Notice that an expansion around the identity transformation is only possible for a subset of Lorentz transformations. Taking $\mu = \nu = 0$ in the defining relation $g^{\mu\nu} = g^{\rho\sigma} \Lambda^\mu_\rho \Lambda^\nu_\sigma$, we find

$$\left(\Lambda^0_0\right)^2 = 1 + \sum_{i=1}^{3} \left(\Lambda^0_i\right)^2 \geq 1, \tag{3.78}$$

so that either $\Lambda^0_0 \geq 1$ or $\Lambda^0_0 \leq -1$; there are no values in between. Transformations with $\Lambda^0_0 \leq -1$ reverse the direction of time. Similarly, the determinant of Λ can be $+1$ or -1, and $\det \Lambda = -1$ involves a reflection. Only the "proper orthochronous" transformations, with $\Lambda^0_0 \geq 1$ and $\det \Lambda = +1$, are connected to the identity transformation, and these are

the ones that we consider here. Reflections of space or time can be performed separately, as discrete transformations.

A transformed space-time coordinate is given by

$$\begin{aligned} x'^{\mu} &= \Lambda^{\mu\rho} x_{\rho} \\ &= \left(g^{\mu\rho} + \omega^{\mu\rho} + \ldots\right) x_{\rho} \\ &= x^{\mu} + \omega^{\mu\rho} x_{\rho} + O(\omega^2) \\ \rightsquigarrow \delta x^{\mu} &\equiv x'^{\mu} - x^{\mu} = \omega^{\mu\rho} x_{\rho}. \end{aligned} \tag{3.79}$$

If Lorentz transformations are a symmetry of a physical system consisting of scalar fields, then the corresponding conserved currents are

$$\begin{aligned} J^{\nu} &= T^{\mu\nu}\delta x_{\mu} = T^{\mu\nu} x^{\rho} \omega_{\mu\rho} = \frac{1}{2} T^{\mu\nu} x^{\rho} \left[\omega_{\mu\rho} - \omega_{\rho\mu}\right] \\ &= \frac{1}{2}\left[T^{\mu\nu} x^{\rho} \omega_{\mu\rho} - T^{\mu\nu} x^{\rho} \omega_{\rho\mu}\right] = \frac{1}{2}\left[T^{\mu\nu} x^{\rho} \omega_{\mu\rho} - T^{\rho\nu} x^{\mu} \omega_{\mu\rho}\right] \\ \rightsquigarrow J^{\nu} &= \frac{1}{2} \omega_{\mu\rho} \mathcal{J}^{\mu\nu\rho} \text{ with } \mathcal{J}^{\mu\nu\rho} \equiv T^{\mu\nu} x^{\rho} - T^{\rho\nu} x^{\mu}. \end{aligned} \tag{3.80}$$

The continuity equations for the conserved currents J^{ν} are,

$$\partial_{\nu} J^{\nu} = 0 \rightsquigarrow \partial_{\nu} \mathcal{J}^{\mu\nu\rho} = 0, \tag{3.81}$$

and the corresponding conserved charges are:

$$M^{\mu\rho} = \int d^3\vec{x}\, \mathcal{J}^{\mu 0\rho} = \int d^3\vec{x} \left[T^{\mu 0} x^{\rho} - T^{\rho 0} x^{\mu}\right]. \tag{3.82}$$

If we associate with $T^{\mu 0}$ the "four-momentum density" $p^{\mu} = T^{\mu 0}$ of the fields, then we can interpret the above conserved charges as the field "four-angular-momentum". For example, for the space components we have

$$M^{12} = \int d^3\vec{x} \left[T^{10} x^2 - T^{20} x^1\right] = -\int d^3\vec{x} \left[\vec{r} \times \vec{p}\,\right]_3 = -L_3. \tag{3.83}$$

Exercise: *Find the conserved currents for a Lagrangian system of*

1. *a vector field A^{μ}*
2. *a tensor field $X^{\mu\nu}$*

if the action is invariant under a Lorentz symmetry transformation.

Exercise: *The charges M^{ij} are the components of the angular momentum. What is the physical meaning of the conserved charges M^{0i}, which correspond to boosts? (Hint: show that their conservation implies $\frac{d}{dt}\int d^3\vec{x}\, x^i\, T^{00} = P^i$, and interpret this result in terms of the centre of energy of the system.)*

Poincaré symmetry transformations

We now require that a physical system of scalar fields possesses both space-time translation and Lorentz transformation symmetry,

$$x^{\mu} \rightarrow x'^{\mu} = \epsilon^{\mu} + \Lambda^{\mu}_{\rho} x^{\rho}. \tag{3.84}$$

The symmetry group of both translations and Lorentz transformations is called the Poincaré group. The conservation equations which we have found above must now hold simultaneously:

$$\partial_\nu T^{\mu\nu} = 0, \text{ due to translation symmetry,} \tag{3.85}$$

$$\partial_\nu \mathcal{J}^{\mu\nu\rho} = 0, \text{ due to Lorentz symmetry.} \tag{3.86}$$

This is only possible if the *energy-momentum tensor* $T^{\mu\nu}$ is symmetric. Indeed,

$$\begin{aligned} 0 &= \partial_\nu \mathcal{J}^{\mu\nu\rho} = \partial_\nu \left[T^{\mu\nu} x^\rho - T^{\rho\nu} x^\mu\right] \\ &= x^\rho \partial_\nu T^{\mu\nu} + T^{\mu\nu} \partial_\nu x^\rho - x^\mu \partial_\nu T^{\rho\nu} - T^{\rho\nu} \partial_\nu x^\mu \\ &= x^\rho 0 + T^{\mu\nu} \delta^\rho_\nu - x^\mu 0 - T^{\rho\nu} \delta^\mu_\nu \\ &= T^{\mu\rho} - T^{\rho\mu} \\ \rightsquigarrow\ & T^{\mu\rho} = T^{\rho\mu}. \end{aligned} \tag{3.87}$$

Both symmetries are needed for this conclusion. A system which is invariant only under translations, such as the non-relativistic Schrödinger field of the next chapter, has a conserved energy-momentum tensor which need not be symmetric.

It is not obvious from its definition,

$$T^{\mu\nu} \equiv \sum_i \frac{\partial \mathcal{L}}{\partial \left(\partial_\nu \phi^{(i)}\right)} \left(\partial^\mu \phi^{(i)}\right) - \mathcal{L} g^{\mu\nu}, \tag{3.88}$$

that the energy-momentum tensor is symmetric: it does not exhibit an explicit symmetry in the indices μ and ν.

Remark. For scalar fields the canonical tensor is in fact symmetric, because Lorentz invariance forces $\mathcal{L}$ to depend on $\partial_\mu \phi$ only through $\partial_\mu \phi\, \partial^\mu \phi$, so that $\partial \mathcal{L} / \partial(\partial_\nu \phi) \propto \partial^\nu \phi$. For vector or spinor fields $\delta_* \phi$ contains an extra spin rotation, the Lorentz current is $\mathcal{J}^{\mu\nu\rho}$ plus a spin term, and the symmetry argument above does not apply to the canonical $T^{\mu\nu}$, which is then in general not symmetric. The improvement below (Belinfante 1940) is what produces a symmetric tensor in that case (see Weinberg, Vol. I, Sect. 7.4).

In any case, the energy-momentum tensor is not uniquely defined. Let us consider a general tensor of the form

$$\partial_\rho f^{\rho\nu\mu},$$

where $f^{\rho\nu\mu}$ is antisymmetric in the indices ρ, ν,

$$f^{\rho\nu\mu} = -f^{\nu\rho\mu}.$$

We can show that such a tensor is a conserved current,

$$\begin{aligned} \partial_\nu \left(\partial_\rho f^{\rho\nu\mu}\right) &= \frac{1}{2}\left[\partial_\nu \partial_\rho f^{\rho\nu\mu} - \partial_\nu \partial_\rho f^{\nu\rho\mu}\right] \quad \text{(antisymmetry)} \\ &= \frac{1}{2}\left[\partial_\nu \partial_\rho f^{\rho\nu\mu} - \partial_\rho \partial_\nu f^{\rho\nu\mu}\right] \quad \text{(relabelling)} \\ &= \frac{1}{2}\left[\partial_\nu \partial_\rho f^{\rho\nu\mu} - \partial_\nu \partial_\rho f^{\rho\nu\mu}\right] = 0. \quad \text{(commuting derivatives)} \end{aligned} \tag{3.89}$$

However, it produces a vanishing conserved charge, as we can verify easily:

$$\begin{aligned} Q &= \int d^3\vec{x}\, \partial_\rho f^{\rho 0\mu} \\ &= \int d^3\vec{x}\, \left(\partial_0 f^{00\mu} + \partial_i f^{i0\mu}\right) \\ &= \int d^3\vec{x}\, \left(\partial_0 0 + \nabla_i f^{i0\mu}\right) \\ &= \int d^3\vec{x}\, \nabla_i f^{i0\mu} \quad \text{(using Gauss' theorem)} \\ &= 0. \end{aligned} \tag{3.90}$$

We can therefore add such a tensor to the energy-momentum tensor,

$$T^{\mu\nu} \to \bar{T}^{\mu\nu} = T^{\mu\nu} + \partial_\rho f^{\rho\nu\mu}, \tag{3.91}$$

without modifying the corresponding conserved charges (energy and momenta). The function $f^{\rho\nu\mu}$ can be chosen so that $\bar{T}^{\mu\nu}$ is manifestly symmetric in the indices μ, ν.

Exercise: *Consider the Lagrangian for the electromagnetic field:*

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{3.92}$$

Calculate the energy-momentum tensor using Eq. (3.88). Is this symmetric? Is this gauge invariant? Find an appropriate way to symmetrise it. Is the result gauge invariant?

3.4 Field Hamiltonian density from discretisation

Noether's theorem for systems with time-translation invariance gave rise to the Hamiltonian as a conserved quantity. We can confirm here that indeed the expression for the Hamiltonian density of Eq. (3.71) can be derived from the familiar definitions for discrete systems in classical mechanics by taking the continuum limit.

For a system with a finite number of degrees of freedom, we find the conjugate momentum of a generalised coordinate by differentiating the Lagrangian with respect to the time derivative of the coordinate:

$$p_j = \frac{\partial L}{\partial \dot{q}_j}, \tag{3.93}$$

The Hamiltonian of the system is then given by,

$$H = \sum_j p_j \dot{q}_j - L. \tag{3.94}$$

Let us now start from the Lagrangian of a field (in one dimension for simplicity),

$$L = \int dx \mathcal{L}, \tag{3.95}$$

which after discretisation becomes

$$L = \sum_i a L_i, \tag{3.96}$$

where L_i is the value of the Lagrangian density at the point i of the lattice (discretised line in our case). The field $q(x,t)$ has the value $q_i(t)$ at the same point. The corresponding conjugate momentum is,

$$p_j = \frac{\partial L}{\partial \dot{q}_j} = \sum_i a \frac{\partial L_i}{\partial \dot{q}_j}. \tag{3.97}$$

The Hamiltonian of the discretised system is then

$$H = \sum_i a \left(\sum_j \dot{q}_j \frac{\partial L_i}{\partial \dot{q}_j} - L_i \right). \tag{3.98}$$

We now assume that the discretised Lagrangian at any point i contains the time derivative of only one degree of freedom, $\dot{q}_i$, and no other (e.g. $\dot{q}_{i+1}, \dot{q}_{i-1}, \ldots$):

$$\frac{\partial L_i}{\partial \dot{q}_j} = \delta_{ij} \frac{\partial L_i}{\partial \dot{q}_i}. \tag{3.99}$$

This was also the assumption that we made for the example of the elastic rod in deriving Euler-Lagrange equations and for all physical systems that we will study. It essentially means that each point in the lattice has its own self-determined kinetic energy. On the other hand, the potential term in L_i can depend on the coordinates of its neighbours, but we will assume that this dependence is in the form of differences, which turn into derivatives at the point i in the continuum limit. The Hamiltonian now becomes,

$$H = \sum_i a \left(\dot{q}_i \frac{\partial L_i}{\partial \dot{q}_i} - L_i \right). \tag{3.100}$$

Finally, by taking the continuum limit $a \to 0$, we obtain a volume integral over a *Hamiltonian density*,

$$H = \int dx \mathcal{H}, \tag{3.101}$$

with

$$\mathcal{H} = \dot{q}(x,t)\pi(x,t) - \mathcal{L}, \tag{3.102}$$

where the conjugate momentum of the field is defined as

$$\pi(x,t) = \frac{\partial \mathcal{L}}{\partial \dot{q}(x,t)}. \tag{3.103}$$

3.4.1 Hamilton equations for fields

Notice that time and space are no longer treated on an equal footing. The Lagrangian density and the action treat them symmetrically, but the Hamiltonian formalism, and the canonical quantisation which will be based on it, single out a time coordinate: momenta are conjugate to time derivatives only. We do not introduce canonical momenta for the space derivatives of the fields. In the discretised picture, a space derivative is a difference of the coordinates at neighbouring lattice points, each of which already has its own canonical momentum; a further momentum for their difference would be superfluous. Time and space nevertheless remain labels, rather than one of them becoming an operator. A formalism in which space and time are treated equally also after quantisation exists; this is the path-integral formulation due to Feynman.

Let us consider a Lagrangian density $\mathcal{L}\left[\phi, \partial_\mu \phi\right]$ for a set of fields $\phi_i(t, \vec{x})$ which are defined in three space dimensions. The conjugate momenta of the fields are defined as:

$$\pi_i(t, \vec{x}) \equiv \frac{\partial \mathcal{L}}{\partial \dot{\phi}_i(t, \vec{x})}. \tag{3.104}$$

For the physical Lagrangian densities that we will consider, the canonical field momenta depend on the fields and their time derivatives (but not their space derivatives):

$$\pi_i(t, \vec{x}) = \pi_i\left[\phi(t, \vec{x}), \dot{\phi}(t, \vec{x})\right]. \tag{3.105}$$

We can assume that we are able to invert the above equation and cast it in the form:

$$\dot{\phi}_i(t, \vec{x}) = \dot{\phi}_i\left[\phi(t, \vec{x}), \pi(t, \vec{x})\right]. \tag{3.106}$$

Generalising the result of Eq. (3.102), the Hamiltonian density is:

$$\mathcal{H}\left[\phi, \partial_j \phi, \pi\right] = \pi_i \dot{\phi}_i[\phi, \pi] - \mathcal{L}\left[\phi, \partial_j \phi, \dot{\phi}\left[\phi, \pi\right]\right], \tag{3.107}$$

with ∂_j, $j = 1, 2, 3$, denoting spatial derivatives. It is now straightforward (**exercise**) to show that

$$\frac{\partial \mathcal{H}}{\partial \phi_i(t, \vec{x})} = -\frac{\partial \mathcal{L}}{\partial \phi_i(t, \vec{x})} \tag{3.108}$$

and

$$\frac{\partial \mathcal{H}}{\partial \pi_i(t, \vec{x})} = \dot{\phi}_i(t, \vec{x}). \tag{3.109}$$

We can now separate the time and space derivatives in the Euler-Lagrange equations:

$$\partial_0 \frac{\partial \mathcal{L}}{\partial \dot{\phi}_i} = \frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_j \frac{\partial \mathcal{L}}{\partial\left(\partial_j \phi_i\right)}. \tag{3.110}$$

Recalling the definition of the canonical momentum, we arrive at Hamilton's equations for fields,

$$\dot{\phi}_i(t, \vec{x}) = \frac{\partial \mathcal{H}}{\partial \pi_i(t, \vec{x})}, \quad \dot{\pi}_i(t, \vec{x}) = -\frac{\partial \mathcal{H}}{\partial \phi_i(t, \vec{x})} + \partial_j \frac{\partial \mathcal{H}}{\partial\left(\partial_j \phi_i\right)}. \tag{3.111}$$

3.5 An example: acoustic waves

As an example, we use an elastic medium (e.g. air) for acoustic waves. The Lagrangian density is a generalisation in three space dimensions of the elastic-rod Lagrangian that we have already studied,

$$\mathcal{L} = \frac{\rho}{2}\left(\frac{\partial y(\vec{r}, t)}{\partial t}\right)^2 - \frac{\rho v_{sound}^2}{2}\left(\vec{\nabla} y(\vec{r}, t)\right)^2, \tag{3.112}$$

with v_{sound} the speed of sound. Setting $v_{sound} = 1$, we write

$$\begin{aligned} \mathcal{L} &= \frac{\rho}{2}\left(\partial_\mu y\right)\left(\partial^\mu y\right) \\ &\equiv \frac{\rho}{2}\left[\left(\partial_{x_0} y\right)^2 - \left(\partial_{x_1} y\right)^2 - \left(\partial_{x_2} y\right)^2 - \left(\partial_{x_3} y\right)^2\right]. \end{aligned} \tag{3.113}$$

We first find the corresponding Euler-Lagrange equations, where we need the derivatives

$$\frac{\partial \mathcal{L}}{\partial y} = 0, \tag{3.114}$$

$$\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu y\right)} = \rho\, \partial^\mu y, \qquad \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu y\right)} = \rho\, \partial_\mu \partial^\mu y = \rho \partial^2 y. \tag{3.115}$$

The equations of motion are

$$\begin{aligned} \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu y\right)} &= \frac{\partial \mathcal{L}}{\partial y} \\ \rightsquigarrow \qquad \partial^2 y &= 0. \end{aligned} \tag{3.116}$$

A solution of the equation of motion is a plane wave,

$$y = e^{ik_\mu x^\mu} \equiv e^{ik\cdot x}, \qquad \text{with} \qquad k_\mu k^\mu = k^2 = 0. \tag{3.117}$$

The conjugate momentum of the field y is,

$$\begin{aligned} \pi &\equiv \frac{\partial \mathcal{L}}{\partial \left(\partial_0 y\right)} = \rho \partial_0 y \\ \rightsquigarrow \quad \partial_0 y &= \frac{\pi}{\rho}. \end{aligned} \tag{3.118}$$

The Hamiltonian density is

$$\begin{aligned} \mathcal{H} &\equiv \pi \partial_0 y - \mathcal{L} \\ \rightsquigarrow \quad \mathcal{H} &= \frac{\pi^2}{2\rho} + \frac{1}{2}\rho \left(\vec{\nabla} y\right)^2. \end{aligned} \tag{3.119}$$

4

Quantisation of the Schrödinger field

We start our study of field quantisation by reproducing the known results of quantum mechanics for a system of many identical non-interacting particles.

4.1 The Schrödinger equation from a Lagrangian density

As a first step, we introduce a Lagrangian that yields the Schrödinger equation for a free particle with the Euler-Lagrange formalism. We can verify that the following density,

$$\mathcal{L} = i\psi^* \frac{\partial \psi}{\partial t} - \frac{\left(\vec{\nabla}\psi^*\right)\left(\vec{\nabla}\psi\right)}{2m}, \tag{4.1}$$

does exactly this. Why this Lagrangian? Writing down a Lagrangian is postulating a physics law; it is our axiom. Here we know the answer that we want to reproduce, the Schrödinger equation, and we guess a Lagrangian which yields it. The Lagrangian is not unique, since we can always add total derivatives, and in general one cannot "derive" it from the equations of motion. In quantum field theory, the most powerful guide for guessing a Lagrangian will be symmetry; for the Schrödinger field we would demand invariance under space-time translations and Galilean transformations. Notice that the Lagrangian depends on two fields, ψ and ψ^*, which we vary independently in the principle of least action. This is a matter of counting degrees of freedom. The wave-function, which we wish to describe classically in this first step, has two real degrees of freedom, its real and its imaginary part. We could write the Lagrangian in terms of these two real fields or, equivalently, in terms of ψ and ψ^*. Either way, we need two Euler-Lagrange equations. ψ and ψ^* turn out to be complex conjugate when they take their physical values, but this is not an assumption that we need to make now. It will be a natural consequence of solving the Euler-Lagrange equations, which are:

$$\begin{aligned} 0 &= \partial_\mu \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \psi\right)} - \frac{\partial \mathcal{L}}{\partial \psi} \\ &= \partial_0 \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \psi\right)} + \partial_i \frac{\partial \mathcal{L}}{\partial\left(\partial_i \psi\right)} - \frac{\partial \mathcal{L}}{\partial \psi} = i\frac{\partial \psi^*}{\partial t} + \frac{\partial_i \partial^i \psi^*}{2m} + 0 \\ \rightsquigarrow 0 &= \left[i\frac{\partial}{\partial t} - \frac{\nabla^2}{2m}\right]\psi^*, \end{aligned} \tag{4.2}$$

and

$$\begin{aligned} 0 &= \partial_\mu \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \psi^*\right)} - \frac{\partial \mathcal{L}}{\partial \psi^*} \\ &= \partial_0 \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \psi^*\right)} + \partial_i \frac{\partial \mathcal{L}}{\partial\left(\partial_i \psi^*\right)} - \frac{\partial \mathcal{L}}{\partial \psi^*} = 0 + \frac{\partial_i \partial^i \psi}{2m} - i\frac{\partial \psi}{\partial t} \\ \rightsquigarrow 0 &= \left[i\frac{\partial}{\partial t} + \frac{\nabla^2}{2m}\right]\psi. \end{aligned} \tag{4.3}$$

Indeed, we obtain the Schrödinger equation for ψ and its complex conjugate for ψ^*. A general solution of the above equations is a superposition of plane-wave solutions,

$$\psi(\vec{x}, t) = \int \frac{d^3\vec{k}}{(2\pi)^3}\, a(\vec{k}) e^{-i\left(\omega_k t - \vec{k}\vec{x}\right)}, \text{ with } \omega_k = \frac{\vec{k}^2}{2m}. \tag{4.4}$$

We can arrive at this solution in a way which does not rely on our quantum mechanics intuition and which we will use for every other field. We write the field as a general four-dimensional Fourier transform, $\psi(x) = \int \frac{d^4k}{(2\pi)^4}\, c(k)\, e^{-i(k_0 t - \vec{k}\cdot\vec{x})}$. The Schrödinger equation then requires $\left(k_0 - \frac{\vec{k}^2}{2m}\right) c(k) = 0$, i.e. $c(k) = 2\pi\, \delta\left(k_0 - \frac{\vec{k}^2}{2m}\right) a(\vec{k})$. The delta function sets k_0 equal to the energy of a free non-relativistic particle, and the k_0 integration yields Eq. (4.4).

It is also useful to compute the conjugate momenta for ψ and ψ^* as well as the Hamiltonian density corresponding to the Schrödinger Lagrangian. We find

$$\pi = \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \psi\right)} = i\psi^*, \tag{4.5}$$

and

$$\pi^* = \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \psi^*\right)} = 0. \tag{4.6}$$

The Hamiltonian density is

$$\begin{aligned} \mathcal{H} &= \pi \partial_0 \psi + \pi^* \partial_0 \psi^* - \mathcal{L} \\ \rightsquigarrow \mathcal{H} &= -\frac{\left(\partial_i \psi^*\right)\left(\partial^i \psi\right)}{2m} = \frac{\left(\vec{\nabla}\psi^*\right)\left(\vec{\nabla}\psi\right)}{2m}. \end{aligned} \tag{4.7}$$

Notice that this Hamiltonian density differs from

$$-\frac{\psi^* \nabla^2 \psi}{2m} = \frac{\left(\vec{\nabla}\psi^*\right)\left(\vec{\nabla}\psi\right)}{2m} - \frac{\vec{\nabla}\left(\psi^* \vec{\nabla}\psi\right)}{2m}, \tag{4.8}$$

by a total divergence which does not contribute to the Hamiltonian integral $H = \int d^3\vec{x}\mathcal{H}$.

Substituting the general solution of Eq. (4.4) in the expression for the Hamiltonian density, Eq. (4.7), we obtain

$$\mathcal{H} = \int \frac{d^3\vec{k}_1}{(2\pi)^3} \frac{d^3\vec{k}_2}{(2\pi)^3} \frac{\vec{k}_1 \vec{k}_2}{2m}\, e^{-i\left[\left(\omega_{k_1} - \omega_{k_2}\right)t - \left(\vec{k}_1 - \vec{k}_2\right)\vec{x}\right]} a^*(k_2) a(k_1). \tag{4.9}$$

The Hamiltonian of the Schrödinger field is,

$$H = \int d^3\vec{x}\, \mathcal{H}, \tag{4.10}$$

and we can perform the volume integration easily using $\int d^3\vec{x} e^{i\vec{k}\cdot\vec{x}} = (2\pi)^3\delta^{(3)}(\vec{k})$. We find the following expression for the Hamiltonian,

$$H = \int \frac{d^3\vec{k}}{(2\pi)^3}\,\omega_k\, a^*(k)a(k). \tag{4.11}$$

The Hamiltonian is time independent, as it should be for a conserved quantity. Its form is suggestive. It is an integral over all momenta of the energy ω_k of a free particle with momentum $\vec{k}$, weighted with the factor $a^*(k)a(k)$. If we had a system of free particles, the total energy would be precisely of this form if $a^*(k)a(k)$ were the density of particles with momentum $\vec{k}$. At this classical level this is only a hint. We will see that quantisation makes this interpretation exact. Notice also that, although $\mathcal{H}$ contains no term with a time derivative of the field, the energy of the system is entirely kinetic.

4.2 Symmetries of the Schrödinger field

Under the transformation

$$\psi(\vec{x},t) \to e^{i\theta}\,\psi(\vec{x},t), \quad \psi^*(\vec{x},t) \to e^{-i\theta}\,\psi^*(\vec{x},t), \tag{4.12}$$

the Schrödinger Lagrangian is invariant,

$$\mathcal{L} \to \mathcal{L}. \tag{4.13}$$

In infinitesimal form, we have

$$\delta\psi = i\theta\psi, \qquad \delta\psi^* = -i\theta\psi^*. \tag{4.14}$$

The phase θ must be a constant. If it depended on space and time, the derivatives in the Lagrangian would act on it and the Lagrangian would no longer be invariant. We can insist on a symmetry with a different phase at every space-time point, but then we are forced to modify the Lagrangian by adding a new vector field. This field can be interpreted as the electromagnetic field, and the modified Lagrangian describes non-relativistic charged particles interacting with it. This is the road that will lead us to quantum electrodynamics.

Noether's theorem tells us that the following current:

$$J^\mu = (J^0, \vec{J}), \tag{4.15}$$

with

$$J^0 = \frac{\partial\mathcal{L}}{\partial(\partial_0\psi)}\delta\psi + \frac{\partial\mathcal{L}}{\partial(\partial_0\psi^*)}\delta\psi^* = -\theta\psi^*\psi, \tag{4.16}$$

and

$$J^i = \frac{\partial\mathcal{L}}{\partial(\partial_i\psi)}\delta\psi + \frac{\partial\mathcal{L}}{\partial(\partial_i\psi^*)}\delta\psi^* = -\theta\frac{i\left[\psi\vec{\nabla}\psi^* - \psi^*\vec{\nabla}\psi\right]}{2m}, \tag{4.17}$$

is conserved:

$$\frac{\partial J^0}{\partial t} + \vec{\nabla}\cdot\vec{J} = 0. \tag{4.18}$$

The conserved charge corresponding to this "phase" symmetry transformation is:

$$N = \int d^3\vec{x} J^0 = \int d^3\vec{x}\psi^*(\vec{x},t)\psi(\vec{x},t), \tag{4.19}$$

where we have dropped an unnecessary $-\theta$ normalisation factor. Substituting the solution of the equations of motion, Eq. (4.4), and performing the integrations over the volume, we find

$$N = \int \frac{d^3\vec{k}}{(2\pi)^3} a^*(\vec{k})a(\vec{k}). \tag{4.20}$$

It satisfies

$$\frac{dN}{dt} = 0. \tag{4.21}$$

If ψ were a wave-function, the conservation of N would be the familiar conservation of the total probability. We insist, however, that ψ is a field. In terms of $a(\vec{k})$, N is again an integral over $a^*(k)a(k)$. With our interpretation of $a^*(k)a(k)$ as a density of particles, N is the total number of particles. This is reasonable: free particles do not interact, so they are neither destroyed nor created.

The Lagrangian is also invariant under space-time translations and space rotations. Symmetry under time translations dictates that the Hamiltonian of Eq. (4.11) is also conserved:

$$\frac{dH}{dt} = 0. \tag{4.22}$$

Symmetry under space translations predicts the existence of a conserved field momentum,

$$P^i = \int d^3\vec{x}\, T^{i0}. \tag{4.23}$$

Explicitly, for the Schrödinger field we find,

$$\vec{P} = \int \frac{d^3\vec{k}}{(2\pi)^3} \vec{k}\, a^*(\vec{k})a(\vec{k}), \tag{4.24}$$

and it is, of course, time independent:

$$\frac{d\vec{P}}{dt} = 0. \tag{4.25}$$

Exercise:

- *Calculate the components of the energy-momentum tensor for the Schrödinger field.*
- *Prove Eq. (4.24) for the total momentum of the Schrödinger field.*
- *Show that the Lagrangian for the Schrödinger field is invariant under space rotations. What are the corresponding conserved charges (angular momentum)? Can you find an expression for the field angular momentum which is manifestly time independent?*
- *Calculate the change of the Lagrangian of the Schrödinger field under an infinitesimal Lorentz boost transformation. Prove that the action is not invariant under boost transformations either.*

4.3 Quantisation of fields

For a system with a finite number of degrees of freedom, $L[q_i, \dot{q}_i]$, the quantisation procedure is the following. For each generalised coordinate, we first find the corresponding conjugate momentum,

$$p_j = \frac{\partial L}{\partial \dot{q}_j}. \tag{4.26}$$

Then we promote all q_i, p_i to operators, which satisfy commutation relations,

$$\left[q_i, p_j\right] = i\delta_{ij}, \tag{4.27}$$

and

$$\left[p_i, p_j\right] = \left[q_i, q_j\right] = 0. \tag{4.28}$$

We now need to generalise this quantisation procedure to systems of fields, where the number of degrees of freedom is infinite. Viewing fields as the continuum limit of a discrete spectrum of degrees of freedom is particularly useful. In this way, we can make the "discrete to continuum" correspondence:

$$\begin{aligned} j &\rightarrow \vec{x}, \\ q_j(t) &\rightarrow \phi(\vec{x}, t), \\ p_j(t) &\rightarrow \pi(\vec{x}, t). \end{aligned} \tag{4.29}$$

We are now in a position to impose quantisation conditions for fields. The canonical prescription for quantising bosonic fields is:

$$\left[\phi(\vec{x}_1, t), \pi(\vec{x}_2, t)\right] = i\delta^{(3)}(\vec{x}_1 - \vec{x}_2), \tag{4.30}$$

and

$$\left[\phi(\vec{x}_1, t), \phi(\vec{x}_2, t)\right] = \left[\pi(\vec{x}_1, t), \pi(\vec{x}_2, t)\right] = 0, \tag{4.31}$$

where the Kronecker symbol of the discrete equations has been replaced by a Dirac delta function. Notice that the quantisation conditions are imposed at equal times, exactly as for $q_i(t)$ and $p_j(t)$ in quantum mechanics. Time is the evolution parameter and not an integration variable; this is why we do not write a covariant-looking factor $\delta(t_1 - t_2)$. Commutators of fields at different times are not imposed; they follow from the dynamics, by evolving the operators in time with the Hamiltonian.

We remark that our quantisation procedure should not be viewed as a result of a mathematical derivation but rather as an axiom that defines our quantum theory. We shall need to revise this procedure a few times, for example when we quantise fermionic fields and gauge fields with self-interactions (like gluons). Two remarks are in order. First, why canonical quantisation? Promoting Poisson brackets to commutators guarantees that operators evolve in time exactly as their classical counterparts, so that classical physics is recovered in the appropriate limit. It is not the only possibility; for example, an anti-commutator also acts algebraically like a derivative, and we will need it for fermions. Second, we have set the constant in the commutator equal to i, i.e. $\hbar = 1$. We could have left an arbitrary real constant there, as we did in quantum mechanics before $\hbar$ was determined from experiment. For free fields such a constant can be absorbed into the normalisation of the conserved quantities and it has no physical consequences; it acquires a meaning once interactions are present.

4.4 Quantised Schrödinger field

We can apply the field quantisation conditions of the previous section (Eqs. (4.30)–(4.31)) to the Schrödinger field $\psi(\vec{x}, t)$ and its conjugate momentum $\pi(\vec{x}, t) = i\psi^*(\vec{x}, t)$. We obtain

$$\begin{aligned} \left[\psi(\vec{x}_1, t), \pi(\vec{x}_2, t)\right] &= i\delta^{(3)}(\vec{x}_1 - \vec{x}_2) \\ \rightsquigarrow \left[\psi(\vec{x}_1, t), \psi^*(\vec{x}_2, t)\right] &= \delta^{(3)}(\vec{x}_1 - \vec{x}_2), \end{aligned} \tag{4.32}$$

and

$$\begin{aligned} \left[\psi(\vec{x}_1,t),\psi(\vec{x}_2,t)\right] &= \left[\pi(\vec{x}_1,t),\pi(\vec{x}_2,t)\right] = 0 \\ \rightsquigarrow \left[\psi(\vec{x}_1,t),\psi(\vec{x}_2,t)\right] &= \left[\psi^*(\vec{x}_1,t),\psi^*(\vec{x}_2,t)\right] = 0. \end{aligned} \tag{4.33}$$

We now run into a problem if we impose a quantisation condition for ψ^* and π^*, since we have found $\pi^* = 0$ in Eq. (4.6). The requirement

$$\delta^{(3)}(\vec{x}_1 - \vec{x}_2) = \left[\psi^*(\vec{x}_1,t),\pi^*(\vec{x}_2,t)\right] = \left[\psi^*(\vec{x}_1,t),0\right] = 0, \tag{4.34}$$

is clearly inconsistent.

To circumvent the problem we exploit the freedom we have in writing down the Lagrangian of the Schrödinger field in Eq. (4.1). A different Lagrangian, which is classically equivalent, yielding identical equations of motion, is

$$\mathcal{L}_{new} = \frac{i}{2}\psi^*\frac{\partial\psi}{\partial t} - \frac{i}{2}\psi\frac{\partial\psi^*}{\partial t} - \frac{\left(\vec{\nabla}\psi^*\right)\left(\vec{\nabla}\psi\right)}{2m}. \tag{4.35}$$

This Lagrangian differs from the one of Eq. (4.1) by a time derivative $\frac{i}{2}\partial_t\left(\psi^*\psi\right)$. It has the advantage of being manifestly real. One might find it more natural to treat ψ and ψ^* completely symmetrically, with a plus sign between the first two terms. However, then these two terms would combine into the total derivative $\frac{i}{2}\partial_t\left(\psi^*\psi\right)$ and no time derivative would be left in the equations of motion. Aesthetics and intuition guide us in guessing Lagrangians, but only some of the beautiful candidates give the right physics.

Exercise: *Verify that this Lagrangian yields the Schrödinger equation as its Euler-Lagrange equation and the conserved quantities of Section 4.2.*

With $\mathcal{L}_{new}$ both π and π^* are non-vanishing, and all commutation relations that we can write lead to one set of three quantisation conditions, which are equivalent to Eqs. (4.32)–(4.33).[1] This example illustrates that Lagrangian densities which are equivalent at the classical level, i.e. yield the same equations of motion, are not all equally suited to the naive canonical quantisation. In our case, $\mathcal{L}_{new}$ can be quantised with the rules of the previous section, while $\mathcal{L}$ cannot.

Remark. $\mathcal{L}$ can nevertheless be quantised consistently. It is first order in time derivatives, so ψ^* is not an independent coordinate: $\pi = i\psi^*$ is itself the momentum conjugate to ψ, and the phase space consists of (ψ,π) only. Imposing $[\psi,\pi] = i\delta^{(3)}$ alone, as in Eq. (4.32), is a consistent quantisation of $\mathcal{L}$ (Faddeev–Jackiw 1988), and it gives the same result as $\mathcal{L}_{new}$ with Dirac brackets. The inconsistency above comes only from treating ψ^* as a further coordinate with its own momentum.

To summarise, the quantisation conditions for the Schrödinger field and its conjugate are:

$$\begin{aligned} \left[\psi(\vec{x}_1,t),\psi(\vec{x}_2,t)\right] = \left[\psi^*(\vec{x}_1,t),\psi^*(\vec{x}_2,t)\right] = 0, \\ \left[\psi(\vec{x}_1,t),\psi^*(\vec{x}_2,t)\right] = \delta^{(3)}\left(\vec{x}_2 - \vec{x}_1\right). \end{aligned} \tag{4.36}$$

[1] To obtain the same quantisation conditions we need to impose $\left[\psi(\vec{x},t),\pi(\vec{y},t)\right] = \frac{i}{2}\delta^{(3)}(\vec{x}-\vec{y})$ and $\left[\psi^*(\vec{x},t),\pi^*(\vec{y},t)\right] = \frac{i}{2}\delta^{(3)}(\vec{x}-\vec{y})$. The factor $\frac{1}{2}$ is not arbitrary: $\mathcal{L}_{new}$ is first order in time derivatives, so the relations $\pi = \frac{i}{2}\psi^*$ and $\pi^* = -\frac{i}{2}\psi$ are constraints among the canonical variables (second-class constraints, in the terminology of Dirac). For such systems the commutators are given by the Dirac brackets, which here reproduce exactly Eqs. (4.32)–(4.33); the naive rule of Eq. (4.30) is not applicable.

From these quantisation conditions for the fields, it is easy to derive the corresponding quantisation conditions for the operators $a(\vec{k})$ and $a^*(\vec{k})$. First, we observe from Eq. (4.4) that the operators $a(\vec{k})$ and $a^*(\vec{k})$ are the Fourier transforms of ψ and ψ^*, respectively,

$$a(\vec{k}) = \int d^3\vec{x}\psi(\vec{x},t)e^{i\left[\omega_k t - \vec{k}\cdot\vec{x}\right]}. \tag{4.37}$$

Exercise: *Verify the above explicitly by substituting the expression of Eq. (4.4) for the field ψ.*

Then, we have

$$\begin{aligned}
&\left[a(\vec{k}_1), a^*(\vec{k}_2)\right] \\
&= \int d^3\vec{x}_1 d^3\vec{x}_2 e^{i\left[\omega_{k_1} t - \vec{k}_1\cdot\vec{x}_1\right]} e^{-i\left[\omega_{k_2} t - \vec{k}_2\cdot\vec{x}_2\right]} \left[\psi(\vec{x}_1,t), \psi^*(\vec{x}_2,t)\right] \\
&= \int d^3\vec{x}_1 d^3\vec{x}_2 e^{i\left[\omega_{k_1} t - \vec{k}_1\cdot\vec{x}_1\right]} e^{-i\left[\omega_{k_2} t - \vec{k}_2\cdot\vec{x}_2\right]} \delta^{(3)}\left(\vec{x}_2 - \vec{x}_1\right) \\
&= \int d^3\vec{x} e^{i\left[\left(\omega_{k_1} - \omega_{k_2}\right)t - \left(\vec{k}_1 - \vec{k}_2\right)\cdot\vec{x}\right]} \\
&\rightsquigarrow \quad \left[a(\vec{k}_1), a^*(\vec{k}_2)\right] = (2\pi)^3 \delta^{(3)}\left(\vec{k}_1 - \vec{k}_2\right).
\end{aligned} \tag{4.38}$$

Similarly (**exercise**), we can prove that

$$\left[a(\vec{k}_1), a(\vec{k}_2)\right] = \left[a^*(\vec{k}_1), a^*(\vec{k}_2)\right] = 0. \tag{4.39}$$

This is the algebra of the raising and lowering operators of the harmonic oscillator, with the Kronecker delta of discrete oscillators replaced by a delta function. The free quantum field is a collection of infinitely many independent harmonic oscillators, one for each value of the momentum $\vec{k}$. Everything that we know from the algebraic solution of the harmonic oscillator, including the existence and uniqueness of its ground state, can be carried over. This will be the case for all free fields that we will study: the Schrödinger field, the Klein-Gordon field, the electromagnetic field and the Dirac field.

Remark. The uniqueness of the ground state deserves a qualification. For a single oscillator the uniqueness of the representation (and of its ground state) is guaranteed by the Stone–von Neumann theorem. For infinitely many oscillators this theorem fails and inequivalent representations exist; the Fock representation built on $a(\vec{k})|0\rangle = 0$ is a choice, which is the physically correct one for free fields (cf. Haag's theorem).

To emphasise that a^* and ψ^* are now operators, we write

$$a^* \to a^\dagger, \quad \psi^* \to \psi^\dagger.$$

4.5 Particle states from quantised fields

Let us now look at the so-called "number density" operator,

$$\mathcal{N}(k) \equiv a^\dagger(k)a(k). \tag{4.40}$$

We can prove that

$$\begin{aligned}
&\left[\mathcal{N}(k), a(p)\right] = \left[a^\dagger(k)a(k), a(p)\right] \\
&= a^\dagger(k)\left[a(k), a(p)\right] + \left[a^\dagger(k), a(p)\right]a(k)
\end{aligned}$$

$$\rightsquigarrow \quad \left[\mathcal{N}(k), a(p)\right] = -(2\pi)^3 \delta^{(3)}(\vec{k} - \vec{p})a(p). \tag{4.41}$$

Also, in the same manner we find that

$$\left[\mathcal{N}(k), a^\dagger(p)\right] = (2\pi)^3\, \delta^{(3)}(\vec{k} - \vec{p}) a^\dagger(p). \tag{4.42}$$

It is also easy to prove (**exercise**) that two number density operators commute with each other,

$$\begin{aligned}
&\left[\mathcal{N}(k), \mathcal{N}(p)\right] = \left[\mathcal{N}(k), a^\dagger(p)a(p)\right] \\
&= \left[\mathcal{N}(k), a^\dagger(p)\right] a(p) + a^\dagger(p) \left[\mathcal{N}(k), a(p)\right] \\
&= (2\pi)^3\, \delta^{(3)}(\vec{k} - \vec{p}) a^\dagger(p)a(p) - (2\pi)^3\, \delta^{(3)}(\vec{k} - \vec{p}) a^\dagger(p)a(p)
\end{aligned}$$

$$\rightsquigarrow \qquad \left[\mathcal{N}(k), \mathcal{N}(p)\right] = 0. \tag{4.43}$$

We recall that the Hamiltonian, Eq. (4.11), is

$$H = \int \frac{d^3\vec{k}}{(2\pi)^3}\, \omega_k\, a^\dagger(k)a(k) \rightsquigarrow H = \int \frac{d^3\vec{k}}{(2\pi)^3}\, \omega_k\, \mathcal{N}(k). \tag{4.44}$$

Obviously, it commutes with the number density operator,

$$\left[\mathcal{N}(p), H\right] = \left[\mathcal{N}(p), \int \frac{d^3\vec{k}}{(2\pi)^3}\, \omega_k\, \mathcal{N}(k)\right] = \int \frac{d^3\vec{k}}{(2\pi)^3}\, \omega_k\, \left[\mathcal{N}(p), \mathcal{N}(k)\right] = 0. \tag{4.45}$$

As a consequence, the eigenstates of the Hamiltonian can be found as eigenstates of the "number density" operator, or the "particle number operator", defined as

$$\hat{N} \equiv \int \frac{d^3\vec{k}}{(2\pi)^3}\, \mathcal{N}(k) = \int \frac{d^3\vec{k}}{(2\pi)^3}\, a^\dagger(k)a(k), \tag{4.46}$$

and which also obviously commutes with the Hamiltonian. This is not a surprise, since we have found earlier that at the classical level this quantity is conserved due to the phase symmetry of the Lagrangian.

From Eqs. (4.41)–(4.42), we derive that

$$\left[\hat{N}, a^\dagger(p)\right] = +a^\dagger(p), \tag{4.47}$$

and

$$\left[\hat{N}, a(p)\right] = -a(p). \tag{4.48}$$

If a state $|n\rangle$ is an eigenstate of the number operator $\hat{N}$ with eigenvalue n,

$$\hat{N}\,|n\rangle = n\,|n\rangle\,, \tag{4.49}$$

then the states $a^\dagger(p)\,|n\rangle$ and $a(p)\,|n\rangle$ are also eigenstates with eigenvalues $n+1$ and $n-1$, respectively. Indeed,

$$\begin{aligned}
&\hat{N}\left(a^\dagger(p)\,|n\rangle\right) = \left\{\left[\hat{N}, a^\dagger(p)\right] + a^\dagger(p)\hat{N}\right\}|n\rangle \\
\rightsquigarrow\ &\hat{N}\left(a^\dagger(p)\,|n\rangle\right) = \left\{+a^\dagger(p) + a^\dagger(p)n\right\}|n\rangle \\
\rightsquigarrow\ &\hat{N}\left(a^\dagger(p)\,|n\rangle\right) = (n+1)\left(a^\dagger(p)\,|n\rangle\right),
\end{aligned} \tag{4.50}$$

and

$$\begin{aligned} & \hat{N}\left(a(p)\left|n\right\rangle\right) = \left\{\left[\hat{N}, a(p)\right] + a(p)\hat{N}\right\}\left|n\right\rangle \\ \rightsquigarrow\ & \hat{N}\left(a(p)\left|n\right\rangle\right) = \left\{-a(p) + a(p)n\right\}\left|n\right\rangle \\ \rightsquigarrow\ & \hat{N}\left(a(p)\left|n\right\rangle\right) = (n-1)\left(a(p)\left|n\right\rangle\right). \end{aligned} \tag{4.51}$$

We can prove that the eigenvalues of the number operator must always be non-negative if the corresponding eigenstates are states of a Hilbert space with a positive norm. Consider the state $\left|\omega\right\rangle = a(p)\left|n\right\rangle$,

$$\begin{aligned} & 0 \leq \|\left|\omega\right\rangle\|^2 = \left\langle\omega\right| \omega\rangle = \left\langle n\right| a^\dagger(p)a(p)\left|n\right\rangle = \left\langle n\right| \mathcal{N}(p)\left|n\right\rangle \\ \rightsquigarrow\ & 0 \leq \left\langle n\right| \left(\int \frac{d^3\vec{p}}{(2\pi)^3}\mathcal{N}(p)\right)\left|n\right\rangle = \left\langle n\right| \hat{N}\left|n\right\rangle = n\left\langle n\right| n\rangle = n\|\left|n\right\rangle\|^2 \\ \rightsquigarrow\ & n \geq 0. \end{aligned} \tag{4.52}$$

This condition is in apparent contradiction with Eq. (4.48), since a repeated application of the operator $a(p)$ on an eigenstate with a positive eigenvalue will eventually yield an eigenstate with a negative eigenvalue. This would be true if no eigenstate were annihilated by the operator $a(p)$. Assume, however, that there is a state $\left|0\right\rangle$ which is annihilated by $a(p)$,

$$a(p)\left|0\right\rangle = 0. \tag{4.53}$$

This state is an eigenstate of the number operator with a zero eigenvalue:

$$\begin{aligned} a(p)\left|0\right\rangle &= 0 \\ \rightsquigarrow a^\dagger(p)a(p)\left|0\right\rangle &= 0\left|0\right\rangle \\ \rightsquigarrow \left(\int \frac{d^3\vec{p}}{(2\pi)^3}a^\dagger(p)a(p)\right)\left|0\right\rangle &= 0\left|0\right\rangle \\ \rightsquigarrow \hat{N}\left|0\right\rangle &= 0\left|0\right\rangle. \end{aligned}$$

Then, Eq. (4.53) only permits us to act on the vacuum state with $a^\dagger$ operators, producing eigenstates of the number operator with non-negative integer eigenvalues. Indeed, if $\hat{N}$ had an eigenvalue which is not an integer, say $n = 0.3$, acting once with $a(p)$ would produce an eigenstate with a negative eigenvalue, which is forbidden. The integers are safe, since the chain stops at the vacuum. We remark that the states $a^\dagger(p)\left|0\right\rangle$ are not normalisable: $\left\langle 0\right| a(p)a^\dagger(p)\left|0\right\rangle = (2\pi)^3\delta^{(3)}(\vec{0}) = V$, the volume of space (see the Casimir effect in the next chapter). This infinite normalisation is common to all states and cancels in physical observables; equivalently, we can work in a finite volume and take the limit $V \to \infty$ at the end. For a general one-particle state

$$\left|1_{\text{particle}}\right\rangle \equiv \int \frac{d^3\vec{p}_1}{(2\pi)^3} f(p_1)a^\dagger(p_1)\left|0\right\rangle, \tag{4.54}$$

we have (**exercise**):

$$\hat{N}\left|1_{\text{particle}}\right\rangle = (+1)\left|1_{\text{particle}}\right\rangle. \tag{4.55}$$

Similarly, for a general state:

$$\left|2_{\text{particles}}\right\rangle \equiv \int \frac{d^3\vec{p}_1}{(2\pi)^3}\frac{d^3\vec{p}_2}{(2\pi)^3} f(p_1, p_2)a^\dagger(p_1)a^\dagger(p_2)\left|0\right\rangle, \tag{4.56}$$

we have (**exercise**):

$$\hat{N}\left|2_{\text{particles}}\right\rangle = (+2)\left|2_{\text{particles}}\right\rangle, \tag{4.57}$$

and so on:

$$\hat{N}\left|m_{\text{particles}}\right\rangle = (+m)\left|m_{\text{particles}}\right\rangle, \tag{4.58}$$

with

$$\left|m_{\text{particles}}\right\rangle \equiv \int \prod_{i=1}^{m}\left(\frac{d^3\vec{p}_i}{(2\pi)^3}a^\dagger(p_i)\right)|0\rangle\, f(p_1,\ldots,p_m). \tag{4.59}$$

Suggestively, we have labelled the eigenstates of the number operator $\hat{N}$ as states of a discrete number of particles. This becomes better justified when we find common eigenstates with the Hamiltonian H and the field momentum operators. The states of Eq. (4.59) with $f(q_1,\ldots,q_m) = \prod_i (2\pi)^3\delta^{(3)}(\vec{q}_i - \vec{p}_i)$, $i = 1\ldots m$:

$$\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle \equiv \left(\prod_{i=1}^{m} a^\dagger(p_i)\right)|0\rangle \tag{4.60}$$

diagonalise the Hamiltonian,

$$H\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle = E\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle, \quad \text{with } E = \sum_{i=1}^{m}\omega_{p_i}. \tag{4.61}$$

Exercise: *Show the above for $m = 1$ and $m = 2$. Convince yourself that it is valid for arbitrary m.*

It is easy to see that a state $\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle$ of Eq. (4.60) is also an eigenstate of the field momentum operator:

$$\vec{P}\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle = \vec{p}_{\text{tot}}\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle, \quad \text{with } \vec{p}_{\text{tot}} = \sum_{i=1}^{m}\vec{p}_i. \tag{4.62}$$

Exercise: *Show the above for $m = 1$ and $m = 2$. Convince yourself that it is valid for arbitrary m.*

To summarise, we have proven that the states of Eq. (4.60) are simultaneous eigenstates of the number operator, the field Hamiltonian operator, and the field momentum operator. Their eigenvalues are the number of particles, the sum of their energies, and the sum of their momenta respectively. We therefore see that the quantisation of the Schrödinger field is a very elegant formalism to describe multi-particle states. This is how we define particles in quantum field theory: we start from fields, we find the conserved energy, momentum and number operators, and we construct their common eigenstates. Notice that all particles created by $a^\dagger$ have the same mass m. A Lagrangian with one field and one mass parameter yields identical particles of a single species; to describe species with different masses we need more fields.

The conservation of the number of particles is special to free theories. In a realistic theory, such as quantum electrodynamics, an electron and a positron can annihilate into photons, which may in turn produce new electron-positron pairs, and so on; the number of particles is not conserved. Quantum electrodynamics still possesses a phase symmetry, but its conserved charge is the electric charge rather than the number of particles. Indeed, in all of the above reactions the total electric charge is zero at every stage.

Exercise: *Show that an n-particle eigenstate of the Hamiltonian for a free Schrödinger field is also an eigenstate of an angular momentum operator.*

Exercise: *We can write the Schrödinger field as*

$$\psi(\vec{x},t) = \rho(\vec{x},t)e^{-i\phi(\vec{x},t)}, \quad \psi(\vec{x},t)^* = \rho(\vec{x},t)e^{i\phi(\vec{x},t)}, \tag{4.63}$$

where ρ, ϕ are real fields.

- *Cast the Lagrangian in terms of the ρ, ϕ fields.*
- *Write down their equations of motion.*
- *What is the Hamiltonian in terms of the ϕ, ρ fields?*
- *Attempt to carry out the quantisation of the theory imposing commutation relations on the ρ, ϕ fields and their conjugate momenta.*

Exercise: *Find the energy, momentum and particle number operator for the following Lagrangian:*

$$\mathcal{L}_{new} = \frac{i}{2}\psi^*\frac{\partial\psi}{\partial t} - \frac{i}{2}\psi\frac{\partial\psi^*}{\partial t} - \frac{\left(\vec{\nabla}\psi^*\right)\left(\vec{\nabla}\psi\right)}{2m} + \lambda\psi^*\psi. \tag{4.64}$$

Show that they commute and find their common eigenstates.

4.6 What is the wave-function in the field quantisation formalism?

In quantum mechanics, the wave function $y(\vec{x},t)$[2] corresponding to a single particle state is the probability amplitude to find a particle at a position $\vec{x}$,

$$y(\vec{x},t) = \left\langle \vec{x} \right| 1_{\text{particle}} \left\rangle \right. . \tag{4.65}$$

For a system of N particles the wave-function is the probability amplitude to find the N particles at positions $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N$,

$$y(\vec{x}_1, \ldots, \vec{x}_N, t) = \left\langle \vec{x}_1, \ldots, \vec{x}_N \right| N_{\text{particles}} \left\rangle \right. . \tag{4.66}$$

The wave-function satisfies the Schrödinger equation, which for free particles is:

$$\left[i\frac{\partial}{\partial t} + \sum_{i=1}^{N} \frac{\vec{\nabla}_{x_i}^2}{2m} \right] y(\vec{x}_1, \ldots, \vec{x}_N, t) = 0. \tag{4.67}$$

Let us now look at a generic N-particle state of the previous section, Eq. (4.59),

$$\left| N_{\text{particles}} \right\rangle = \int \left(\prod_{i=1}^{N} \frac{d^3\vec{p}_i}{(2\pi)^3} a^\dagger(p_i) \right) |0\rangle \, f(p_1, \ldots, p_N), \tag{4.68}$$

where we can express the creation operators $a^\dagger(p_i)$ as Fourier transforms of the field operators $\psi^\dagger(\vec{x},t)$. Substituting Eq. (4.37) into Eq. (4.68), we obtain

$$\left| N_{\text{particles}} \right\rangle = \int \left(\prod_{i=1}^{N} d^3\vec{x}_i \right) |S(x_1, \ldots, x_N)\rangle \, \tilde{f}(x_1, \ldots, x_N, t), \tag{4.69}$$

[2] In this section, we use the letter y to denote the wave-function, and the letter ψ for the Schrödinger field operator.

with,

$$|S(x_1,\ldots,x_N)\rangle = \left(\prod_{i=1}^{N}\psi^\dagger(x_i)\right)|0\rangle\,, \tag{4.70}$$

and

$$\tilde{f}(x_1,\ldots,x_N,t) = \int\prod_{i=1}^{N}\left(\frac{d^3\vec{p}_i}{(2\pi)^3}e^{-i\left(\omega_{p_i}t-\vec{p}_i\vec{x}_i\right)}\right)f(p_1,\ldots,p_N). \tag{4.71}$$

By an explicit calculation (**exercise**), we can verify that

$$\left[i\frac{\partial}{\partial t}+\sum_{i=1}^{N}\frac{\vec{\nabla}^2_{x_i}}{2m}\right]\tilde{f}(\vec{x}_1,\ldots,\vec{x}_N,t)=0, \tag{4.72}$$

i.e. it satisfies the Schrödinger equation. Therefore, the function $\tilde{f}$ is identified with the wave-function $y(x_1,\ldots,x_N,t)$,

$$\tilde{f}(x_1,\ldots,x_N,t) = y(x_1,\ldots,x_N,t) = \left\langle\vec{x}_1,\ldots,\vec{x}_N\right|N_{\text{particles}}\rangle\,. \tag{4.73}$$

Using this observation, we write Eq. (4.69) as

$$\begin{aligned}\left|N_{\text{particles}}\right\rangle &= \int\left(\prod_{i=1}^{N}d^3\vec{x}_i\right)|S(x_1,\ldots,x_N)\rangle\left\langle\vec{x}_1,\ldots,\vec{x}_N\right|N_{\text{particles}}\rangle\,,\\ \rightsquigarrow\left|N_{\text{particles}}\right\rangle &= \left[\int\left(\prod_{i=1}^{N}d^3\vec{x}_i\right)|S(x_1,\ldots,x_N)\rangle\left\langle\vec{x}_1,\ldots,\vec{x}_N\right|\right]\left|N_{\text{particles}}\right\rangle\,.\end{aligned} \tag{4.74}$$

Inside the square bracket we have the unit operator, which leads us to the conclusion that

$$|S(x_1,\ldots,x_N)\rangle = \left|\vec{x}_1,\ldots,\vec{x}_N\right\rangle\,. \tag{4.75}$$

Remark. The last step is heuristic and misses a combinatorial factor. For identical bosons,

$$\left\langle\vec{x}_1,\ldots,\vec{x}_N\right|\vec{y}_1,\ldots,\vec{y}_N\rangle = \sum_\sigma\prod_i\delta^{(3)}(\vec{x}_i-\vec{y}_{\sigma(i)})$$

(sum over the $N!$ permutations σ), so that the unit operator on N-particle states is

$$\frac{1}{N!}\int\prod_i d^3\vec{x}_i\,\left|\vec{x}_1,\ldots,\vec{x}_N\right\rangle\left\langle\vec{x}_1,\ldots,\vec{x}_N\right|,$$

and $\left\langle\vec{x}_1,\ldots,\vec{x}_N\right|N_{\text{particles}}\rangle = N!\,\tilde{f}$ for a symmetric f (only the symmetric part of f contributes). With this normalisation the identification of $\tilde{f}$ with the wave-function holds up to the constant $N!$.

We summarise the main result of this section, which connects the field operator formalism with the traditional wave-function description of non-relativistic quantum mechanics:

- A quantum mechanical state of N particles at positions $\vec{x}_1,\ldots,\vec{x}_N$ is given by the action of field operators on the vacuum state:

$$\left|\vec{x}_1,\ldots,\vec{x}_N\right\rangle = \left(\prod_{i=1}^{N}\psi^\dagger(x_i)\right)|0\rangle\,. \tag{4.76}$$

- A general N-particle state is a superposition,

$$\left|N_{\text{particles}}\right\rangle = \int d^3\vec{x}_1 \ldots d^3\vec{x}_N \tilde{f}(x_1, \ldots, x_N) \left|\vec{x}_1, \ldots, \vec{x}_N\right\rangle . \tag{4.77}$$

- The kernel of the superposition integral, $\tilde{f}(\vec{x}_1, \ldots, \vec{x}_N, t)$, is the wave-function, satisfying the Schrödinger equation

$$\left[i\frac{\partial}{\partial t} + \sum_{i=1}^{N} \frac{\vec{\nabla}^2_{x_i}}{2m} \right] \tilde{f} = 0. \tag{4.78}$$

5

The Klein-Gordon field

In the study of the Schrödinger field, we demonstrated how quantum field theory can be used to construct quantum states of particles with definite momentum and the energy-momentum relation

$$E = \frac{\vec{p}^2}{2m}. \tag{5.1}$$

This is the correct energy-momentum relation in the non-relativistic limit. Would it be possible to modify this procedure in such a way that we obtain the relativistic energy-momentum relation

$$E^2 = \vec{p}^2 + m^2? \tag{5.2}$$

5.1 Real Klein-Gordon field

Let us consider the Lagrangian

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\right)\left(\partial^\mu\phi\right) - \frac{1}{2}m^2\phi^2, \tag{5.3}$$

where ϕ is a real scalar field. This Lagrangian is invariant under Lorentz transformations and therefore consistent with special relativity. We will attempt to write a quantum relativistic theory of free scalar (spin-0) particles by quantising the field ϕ.

5.1.1 Real solution of the Klein-Gordon equation

The Euler-Lagrange equation for the field ϕ of Eq. (5.3) is

$$\partial_\mu \frac{\partial\mathcal{L}}{\partial\left(\partial_\mu\phi\right)} - \frac{\partial\mathcal{L}}{\partial\phi} = 0$$

$$\rightsquigarrow \qquad \left[\partial^2 + m^2\right]\phi(x) = 0. \tag{5.4}$$

This is known as the Klein-Gordon equation. A common mistake, when computing $\frac{\partial\mathcal{L}}{\partial(\partial_\nu\phi)}$ for the first time, is to differentiate only the factor $\partial_\mu\phi$ with the lower index. The factor $\partial^\mu\phi$ depends on the same components (up to signs for the space-like ones) and must be differentiated as well:

$$\frac{\partial}{\partial(\partial_\nu\phi)}\left[\frac{1}{2}g^{\mu\rho}(\partial_\mu\phi)(\partial_\rho\phi)\right] = \frac{1}{2}g^{\mu\rho}\left[\delta^\nu_\mu\partial_\rho\phi + \partial_\mu\phi\,\delta^\nu_\rho\right] = \partial^\nu\phi. \tag{5.5}$$

The two terms are equal and cancel the factor $\frac{1}{2}$ of the Lagrangian. For $m = 0$, Eq. (5.4) has the same form as the Maxwell equations for the vector potential in the Lorenz gauge, which we will meet in the next chapter.

We shall find a general solution of the Klein-Gordon equation for a real field ϕ. We start with the ansatz:

$$\phi(x) = \int \frac{d^4k}{(2\pi)^4} \left[f(k)e^{-ik\cdot x} + f^*(k)e^{+ik\cdot x} \right], \tag{5.6}$$

which is manifestly real. In addition, it is invariant under Lorentz transformations, as anticipated for a scalar field. Substituting into the Klein-Gordon equation, we find

$$\int \frac{d^4k}{(2\pi)^4} \left(m^2 - k^2\right) \left[f(k)e^{-ik\cdot x} + f^*(k)e^{+ik\cdot x} \right] = 0. \tag{5.7}$$

The above is satisfied if

$$f(k) = (2\pi)\delta(k^2 - m^2)c(k), \qquad f^*(k) = (2\pi)\delta(k^2 - m^2)c^*(k). \tag{5.8}$$

A general real solution of the Klein-Gordon equation is therefore

$$\phi(x^\mu) = \int \frac{d^4k}{(2\pi)^3} \delta\left(m^2 - k^2\right) \left[c(k)e^{-ik\cdot x} + c^*(k)e^{+ik\cdot x} \right]. \tag{5.9}$$

The delta function requires that

$$k^2 = k_\mu k^\mu = m^2 \rightsquigarrow k_0 = \pm\omega_k, \quad \text{with } \omega_k = \sqrt{\vec{k}^2 + m^2}. \tag{5.10}$$

The delta function can be written as,

$$\delta(k^2 - m^2) = \delta(k_0^2 - \omega_k^2) = \frac{\delta(k_0 - \omega_k) + \delta(k_0 + \omega_k)}{2\omega_k}. \tag{5.11}$$

We can now perform the k_0 integration in Eq. (5.9), obtaining

$$\begin{aligned} \phi(x^\mu) \quad = \quad & \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \Big[c(\omega_k, \vec{k})e^{-i\left(\omega_k t - \vec{k}\vec{x}\right)} + c(-\omega_k, \vec{k})e^{+i\left(\omega_k t + \vec{k}\vec{x}\right)} \\ & + c^*(\omega_k, \vec{k})e^{+i\left(\omega_k t - \vec{k}\vec{x}\right)} + c^*(-\omega_k, \vec{k})e^{-i\left(\omega_k t + \vec{k}\vec{x}\right)} \Big]. \end{aligned} \tag{5.12}$$

We perform a change of integration variable, $\vec{k} \longrightarrow -\vec{k}$, in the integrals corresponding to the second and fourth terms of the sum inside the square bracket. Then, the four terms can be grouped together into a sum of only two exponentials, and we can write the general solution as:

$$\phi(x^\mu) = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \left[a(\vec{k})e^{-i\left(\omega_k t - \vec{k}\vec{x}\right)} + a^*(\vec{k})e^{i\left(\omega_k t - \vec{k}\vec{x}\right)} \right], \tag{5.13}$$

where $a(\vec{k}) = c(\omega_k, \vec{k}) + c^*(-\omega_k, -\vec{k})$.

We remark that we often write the integration measure in an explicitly Lorentz-invariant form,

$$\int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} = \int \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\Theta(k_0) = \int \frac{d^4k}{(2\pi)^3} \delta^{(+)}(k^2 - m^2), \tag{5.14}$$

with

$$\delta^{(+)}(k^2 - m^2) \equiv \delta(k^2 - m^2)\Theta(k_0). \tag{5.15}$$

The measure is invariant under orthochronous transformations, which do not change the sign of the time component of the integration variable. Indeed, d^4k is invariant under proper Lorentz transformations (the Jacobian is $\det\Lambda = 1$), k^2 is a scalar, and the sign of k_0 is preserved by orthochronous transformations. We will meet this measure again and again in relativistic quantum field theory. Since the measure and the plane waves $e^{\mp ik\cdot x}$ are Lorentz invariant, the coefficients $a(\vec{k})$ inherit the transformation properties of the scalar field: if ϕ is a scalar, then a transforms as a scalar function of the momentum.

5.1.2 Quantisation of the real Klein-Gordon field

The conjugate momentum of the Klein-Gordon field, $\phi(\vec{x},t)$, is

$$\pi = \frac{\partial\mathcal{L}}{\partial(\partial_0\phi)} = \partial_0\phi. \tag{5.16}$$

The Hamiltonian of the Klein-Gordon field is,

$$H = \int d^3\vec{x}\frac{\left(\partial_0\phi\right)^2 + \left(\vec{\nabla}\phi\right)^2 + m^2\phi^2}{2}. \tag{5.17}$$

Substituting into the Hamiltonian integral the solution of Eq. (5.13), and performing the integration over the space volume, we find:

$$H = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}\frac{\omega_k}{2}\left[a^\dagger(k)a(k) + a(k)a^\dagger(k)\right], \tag{5.18}$$

where we have promoted the field to an operator. We now impose commutation relations for the field and its conjugate momentum,

$$\left[\phi(\vec{x}_1,t),\phi(\vec{x}_2,t)\right] = \left[\pi(\vec{x}_1,t),\pi(\vec{x}_2,t)\right] = 0, \tag{5.19}$$

and

$$\left[\phi(\vec{x}_1,t),\pi(\vec{x}_2,t)\right] = i\delta^{(3)}\left(\vec{x}_2 - \vec{x}_1\right). \tag{5.20}$$

With a straightforward calculation, we can verify (**exercise**) that the operators a and $a^\dagger$ are written in terms of the ϕ and π operators as

$$a(k) = i\int d^3\vec{x}e^{i\left(\omega_k t - \vec{k}\vec{x}\right)}\left[\pi(\vec{x},t) - i\omega_k\phi(\vec{x},t)\right], \tag{5.21}$$

and

$$a^\dagger(k) = -i\int d^3\vec{x}e^{-i\left(\omega_k t - \vec{k}\vec{x}\right)}\left[\pi(\vec{x},t) + i\omega_k\phi(\vec{x},t)\right]. \tag{5.22}$$

Then we can compute the commutators that can be formed with a and $a^\dagger$. We find (**exercise**),

$$\left[a(k_1),a(k_2)\right] = \left[a^\dagger(k_1),a^\dagger(k_2)\right] = 0, \tag{5.23}$$

and

$$\left[a(k_1),a^\dagger(k_2)\right] = (2\pi)^3(2\omega_{k_1})\delta^{(3)}\left(\vec{k}_1 - \vec{k}_2\right). \tag{5.24}$$

5.1.3 Particle states for the real Klein-Gordon field

We can now construct the particle number operator,

$$\hat{N} = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} a^\dagger(k)a(k), \tag{5.25}$$

for which we can prove that

$$\left[\hat{N}, a(p)\right] = -a(p), \tag{5.26}$$

and

$$\left[\hat{N}, a^\dagger(p)\right] = +a^\dagger(p). \tag{5.27}$$

Therefore, the operators $a(p)$ and $a^\dagger(p)$ are ladder operators. In particular, if a state $|\mathbf{S}\rangle$ is an eigenstate of the number operator $\hat{N}$ with eigenvalue n, then the states $a(p)|\mathbf{S}\rangle$ and $a^\dagger(p)|\mathbf{S}\rangle$ are also eigenstates with eigenvalues $n-1$ and $n+1$, respectively. The construction of particle states proceeds similarly to the quantisation for the Schrödinger field, after we observe that the field Hamiltonian and the number operator commute (**exercise**):

$$\left[\hat{N}, H\right] = 0. \tag{5.28}$$

The above equations lead to the existence of a vacuum state $|0\rangle$, which is annihilated by the operator $a(k)$,

$$a(k)|0\rangle = 0. \tag{5.29}$$

The vacuum state is an eigenstate of the number operator with a zero eigenvalue,

$$\hat{N}|0\rangle = 0|0\rangle. \tag{5.30}$$

All other states can be produced from the vacuum state by acting on it with *creation* operators $a^\dagger(p_i)$. A generic state

$$|\Psi_m\rangle \equiv \left(\prod_{i=1}^{m} a^\dagger(p_i)\right)|0\rangle, \tag{5.31}$$

contains m particles,

$$\hat{N}|\Psi_m\rangle = m|\Psi_m\rangle. \tag{5.32}$$

5.1.4 Energy of particles and "normal ordering"

We consider the simplest case of a state $|\Psi_m\rangle$,

$$\left|p\right\rangle \equiv a^\dagger(p)|0\rangle, \tag{5.33}$$

which is a one-particle state, satisfying

$$\hat{N}\left|p\right\rangle = 1\left|p\right\rangle. \tag{5.34}$$

We can compute the energy level of the state by acting on it with the field Hamiltonian operator,

$$\begin{aligned}
H\left|p\right\rangle &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \frac{\omega_k}{2} \left\{a^\dagger(k)a(k) + a(k)a^\dagger(k)\right\}\left|p\right\rangle \\
&= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \frac{\omega_k}{2} \left\{2a^\dagger(k)a(k) + \left[a(k), a^\dagger(k)\right]\right\}\left|p\right\rangle \\
&= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \frac{\omega_k}{2} \left\{2a^\dagger(k)a(k) + (2\pi)^3 2\omega_k \delta^{(3)}(0)\right\}\left|p\right\rangle \\
\rightsquigarrow H\left|p\right\rangle &= \left(\omega_p + \delta^{(3)}(0)\left[\int d^3\vec{k}\frac{\omega_k}{2}\right]\right)\left|p\right\rangle.
\end{aligned} \tag{5.35}$$

To our surprise, we find an infinite eigenvalue! This is at first sight very embarrassing. Could we have avoided it by writing $2a^\dagger(k)a(k)$ in the Hamiltonian from the start? Not if we start from the Lagrangian. The term $m^2\phi^2$, for example, is the square of a field containing both a and $a^\dagger$. Its product creates terms aa, $a^\dagger a^\dagger$, $a^\dagger a$ and $aa^\dagger$. The first two are eliminated by the space integration, but the last two remain, in the order dictated by the product. Changing their order changes the Hamiltonian operator. As we will see, the change amounts to a recalibration of the energy, but it is a change nonetheless.

We can make sense of this infinity if we look at the energy of the vacuum state, which contains no particles,

$$\begin{aligned} H|0\rangle &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}\frac{\omega_k}{2}\left\{a^\dagger(k)a(k)+a(k)a^\dagger(k)\right\}|0\rangle \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}\frac{\omega_k}{2}\left\{2a^\dagger(k)a(k)+\left[a(k),a^\dagger(k)\right]\right\}|0\rangle \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}\frac{\omega_k}{2}\left\{2a^\dagger(k)a(k)+(2\pi)^3\,2\omega_k\delta^{(3)}(0)\right\}|0\rangle \\ &\rightsquigarrow\quad H|0\rangle = \delta^{(3)}(0)\left[\int d^3\vec{k}\frac{\omega_k}{2}\right]|0\rangle \\ &\rightsquigarrow\quad \delta^{(3)}(0)\left[\int d^3\vec{k}\frac{\omega_k}{2}\right] = \langle 0|H|0\rangle = E_{\text{vacuum}}. \end{aligned} \tag{5.36}$$

We find that the field Hamiltonian gives an energy eigenvalue for the vacuum state in Eq. (5.36) which is equal to the same infinite constant that appeared in the energy eigenvalue (Eq. (5.35)) of the one-particle state. We can then re-write Eq. (5.35) in an apparently innocent manner,

$$H|p\rangle = \left(\sqrt{\vec{p}^2+m^2}+E_{\text{vacuum}}\right)|p\rangle\,. \tag{5.37}$$

Using the property (**exercise**)

$$\left[H,a^\dagger(p_i)\right] = +\omega(p_i)a^\dagger(p_i), \tag{5.38}$$

we can prove that states with more particles have energy

$$H|\vec{p}_1,\ldots,\vec{p}_m\rangle = \left[E_{\text{vacuum}}+\sum_i\sqrt{\vec{p}_i^2+m^2}\right]|\vec{p}_1,\ldots,\vec{p}_m\rangle\,. \tag{5.39}$$

All Hamiltonian eigenvalues contain an identical infinite constant which is equal to the vacuum energy. However, measurements of absolute energy levels are not possible. The vacuum energy is harmless if we want to measure the energy difference of the states $|\Psi_m\rangle$ from the vacuum. We can then "recalibrate" our energy levels (by an infinite constant) removing from the Hamiltonian operator the energy of the vacuum:

$$H \rightarrow :H: \equiv H - E_{\text{vacuum}} = H - \langle 0|H|0\rangle\,. \tag{5.40}$$

Explicitly, the Hamiltonian operator after we subtract the vacuum constant can be written as

$$:\int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}\frac{\omega_k}{2}\left[a^\dagger(k)a(k)+a(k)a^\dagger(k)\right]: = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}\omega_k\left[a^\dagger(k)a(k)\right]. \tag{5.41}$$

A practical trick to remove from a conserved operator, such as the Hamiltonian, its vacuum expectation value:

$$:O: \equiv O - \langle 0|O|0\rangle\,, \tag{5.42}$$

is to perform what is known as "normal ordering". As you can see in Eq. (5.41), this is achieved by putting creation operators to the left of annihilation operators. With this practical rule,

$$: \left[a^\dagger(k)a(k) + a(k)a^\dagger(k)\right] := a^\dagger(k)a(k) + a^\dagger(k)a(k) = 2a^\dagger(k)a(k). \tag{5.43}$$

This is equivalent to using the commutation relation $[a(k), a^\dagger(k)] = (2\pi)^3 2\omega_k \delta^{(3)}(0)$ in order to put the annihilation operators to the right, and dropping the infinities $\delta^{(3)}(0)$. Normal ordering leaves annihilation operators to the right, and guarantees that the vacuum is automatically annihilated by the modified operator:

$$:O:|0\rangle = 0. \tag{5.44}$$

Normal ordering has a name, which makes it sound important, but it is nothing more than this recalibration. We do not claim that the vacuum energy does not exist; we identify the physical energy of particles with the excess of energy that particle states carry over the vacuum. Nevertheless, the infinite and uncomputable energy of the vacuum is a limitation of quantum field theory, which tells us correctly only energy differences. It is one of the indications that quantum field theory is not the final theory of nature.

5.1.5 Field momentum conservation

Noether's theorem predicts that the field momentum is a conserved operator, due to the symmetry of the Klein-Gordon action under space translations. The momentum operator is given by

$$\begin{aligned}
\vec{P} &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \frac{\vec{k}}{2} \left[a^\dagger(k)a(k) + a(k)a^\dagger(k)\right] \\
&= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \vec{k} \left[a^\dagger(k)a(k)\right] + \frac{\delta^{(3)}(0)}{2} \int d^3\vec{k}\,\vec{k} \\
\rightsquigarrow \vec{P} &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \vec{k} \left[a^\dagger(k)a(k)\right].
\end{aligned} \tag{5.45}$$

Here, the infinity $\delta^{(3)}(0)$ cancels because the momentum integral multiplying the delta-function term vanishes due to the antisymmetry of the integrand under the transformation $\vec{k} \to -\vec{k}$. $\vec{P}$ commutes with both the Hamiltonian and the number operator, with which it has common eigenstates.

The momentum of the vacuum state is zero,

$$\vec{P}\,|0\rangle = 0. \tag{5.46}$$

The $\vec{P}$ eigenvalues of multi-particle states can easily be found using

$$\left[\vec{P}, a^\dagger(p)\right] = +\vec{p}\,a^\dagger(p). \tag{5.47}$$

For example, its eigenvalue for the one-particle state $\left|p\right\rangle = a^\dagger(p)\,|0\rangle$ is

$$\vec{P}\left|p\right\rangle = \left[\vec{P}, a^\dagger(p)\right]|0\rangle = \vec{p}\,a^\dagger(p)\,|0\rangle = \vec{p}\,\left|p\right\rangle. \tag{5.48}$$

For a multi-particle state

$$\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle \equiv \prod_{i=1}^{m} a^\dagger(\vec{p}_i)\left|0\right\rangle, \tag{5.49}$$

we have

$$\vec{P}\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle = \left(\sum_{i=1}^{m}\vec{p}_i\right)\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle. \tag{5.50}$$

5.1.6 Labels of particle states?

The main outcome of quantising the real Klein-Gordon field is a spectrum of states which can be identified as states of many particles with definite number of particles, energy and momentum:

$$\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle \equiv \left(\prod_{i=1}^{m} a^\dagger(p_i)\right)\left|0\right\rangle. \tag{5.51}$$

Such a state describes m identical particles,

$$\hat{N}\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle = m\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle, \tag{5.52}$$

each of them carrying momentum $\vec{p}_i$,

$$\vec{P}\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle = \left(\sum_{i=1}^{m}\vec{p}_i\right)\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle, \tag{5.53}$$

and having a relativistic energy $+\sqrt{\vec{p}_i^{\,2}+m^2}$,

$$H\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle = \left(\sum_{i=1}^{m}\sqrt{\vec{p}_i^{\,2}+m^2}\right)\left|\vec{p}_1,\ldots,\vec{p}_m\right\rangle. \tag{5.54}$$

We have therefore arrived at a consistent combination of quantum mechanics and special relativity, where field quanta give rise to particles with the correct properties as anticipated from relativity.

Notice that all energies are positive. The Klein-Gordon equation, $k^2 = m^2$, admits solutions with $k_0 = -\omega_k$, which are problematic if ϕ is interpreted as a wave-function. For a field, the negative-frequency solutions are not discarded; they are absorbed, after $\vec{k} \to -\vec{k}$, into the $a^\dagger$ term of Eq. (5.13). After quantisation, the Hamiltonian has common eigenstates with the number operator, whose spectrum is bounded from below by the vacuum, and this leaves no room for negative energies. Quantum field theory, as we will see shortly, produces negative charges, but not negative energies. The real Klein-Gordon field is not only a toy model. The Higgs boson is an elementary scalar particle, and the above is a description of free Higgs particles.

Nothing prevents us from creating several particles with the same momentum, e.g. $\left(a^\dagger(p)\right)^n\left|0\right\rangle$. Particles without other degrees of freedom which can be piled up in the same state are bosons; the Klein-Gordon field describes bosons.

Exercise: *Attempt to quantise the real Klein-Gordon field with anti-commutation relations,* $\left\{a(k_1), a^\dagger(k_2)\right\} = (2\pi)^3\, 2\omega_{k_1}\delta^{(3)}(\vec{k}_1 - \vec{k}_2)$. *Show that the Hamiltonian of Eq. (5.17) then reduces to a constant, so that the states created by* $a^\dagger$ *do not carry the energy of relativistic particles.*

5.2 Casimir effect: the energy of the vacuum

A remarkable consequence of the quantisation of the Klein-Gordon field is that a state with no particles, the vacuum $|0\rangle$, has energy. For the real Klein-Gordon field, for example, we find that the vacuum energy is:

$$E_{\text{vacuum}} = \delta^{(3)}(0) \int d^3\vec{k} \frac{\omega_k}{2}. \tag{5.55}$$

We can associate the infinity $\delta^{(3)}(0)$ with the volume of our physical system. We have considered that the classical solutions for the fields extend to an infinite volume, and we have performed all space integrations from $-\infty$ to ∞. If we were to consider a finite volume, this delta function would be replaced by the volume of integration. Recall that

$$\int d^3\vec{x} e^{-i\vec{k}\cdot\vec{x}} = (2\pi)^3 \delta^{(3)}(\vec{k}). \tag{5.56}$$

For $\vec{k} = 0$, we have

$$V = \int d^3\vec{x} = (2\pi)^3 \delta^{(3)}(\vec{0}), \tag{5.57}$$

and the density of the vacuum energy can be cast in the physically more appealing form,

$$\epsilon = \frac{E_{\text{vacuum}}}{V} = \int \frac{d^3\vec{k}}{(2\pi)^3} \frac{\sqrt{m^2 + \vec{k}^2}}{2}. \tag{5.58}$$

A second source of infinity in the expression for the vacuum energy is the integration over all frequency modes $\omega_k = \sqrt{m^2 + \vec{k}^2}$.

The value of the energy of the vacuum does not enter physical predictions (if one neglects gravitational effects), and we need not worry that it comes out infinite. Gravity is the exception, because gravity couples to all energy, including the energy of the vacuum. In a theory with gravity the action is $S = \int d^4x \sqrt{-\det g_{\mu\nu}}\, \mathcal{L}$, where the metric $g_{\mu\nu}(x)$ is a dynamical field. Shifting $\mathcal{L}$ (or $\mathcal{H}$) by a constant now changes the equations of motion of the metric; such a constant acts as a cosmological constant and can drive an accelerated expansion of the universe, which is indeed observed. Other fields contribute to the vacuum energy as well. Fermionic fields give infinite contributions with the opposite sign; in supersymmetric theories, which have equal numbers of bosonic and fermionic degrees of freedom, these contributions cancel exactly. However, we should be able to see that the vacuum energy changes if, for some reason, the fields vanish in some region of space and/or some frequencies ω_k do not contribute to the vacuum. This can be achieved if the field is forced to satisfy boundary conditions.

Let us assume that the Klein-Gordon field is forced to vanish on the planes $x = 0$ and $x = L$,

$$\phi(t, x = 0, y, z) = \phi(t, x = L, y, z) = 0. \tag{5.59}$$

The general field solution of the Klein-Gordon equation which in addition satisfies the boundary conditions is

$$\phi(x^\mu) = \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right) \int \frac{d^2\vec{k}_\perp}{(2\pi)^2 2\omega_k} \left[a(\vec{k}) e^{-i\left(k^0 t - \vec{k}_\perp \cdot \vec{x}_\perp\right)} + a^\dagger(\vec{k}) e^{i\left(k^0 t - \vec{k}_\perp \cdot \vec{x}_\perp\right)} \right], \tag{5.60}$$

with

$$\omega_k = \sqrt{m^2 + \vec{k}_\perp^2 + \frac{n^2\pi^2}{L^2}}, \tag{5.61}$$

and

$$\vec{x}_\perp = (y,z), \quad \vec{k} = \left(\frac{n\pi}{L}, \vec{k}_\perp\right), \quad \vec{k}_\perp = (k_y, k_z)\,. \tag{5.62}$$

Notice that we only integrate over the directions k_y, k_z of the momenta parallel to the plates, and we sum over the values of k_x, which is discretised.

As in the absence of boundary conditions, the quantisation conditions for the real Klein-Gordon field are given by Eqs. (5.19)–(5.20). For the creation and annihilation operators, this yields (**exercise**) the commutation relations

$$\left[a\left(\frac{n\pi}{L}, \vec{k}_\perp\right), a\left(\frac{r\pi}{L}, \vec{l}_\perp\right)\right] = 0 = \left[a^\dagger\left(\frac{n\pi}{L}, \vec{k}_\perp\right), a^\dagger\left(\frac{r\pi}{L}, \vec{l}_\perp\right)\right], \tag{5.63}$$

$$\left[a\left(\frac{n\pi}{L}, \vec{k}_\perp\right), a^\dagger\left(\frac{r\pi}{L}, \vec{l}_\perp\right)\right] = \frac{2}{L}\delta_{nr}(2\pi)^2(2\omega_k)\delta^{(2)}\left(\vec{k}_\perp - \vec{l}_\perp\right). \tag{5.64}$$

Substituting the solution for the field of Eq. (5.60) and for the corresponding conjugate momentum into the Hamiltonian integral of Eq. (5.17), we find (**exercise**)

$$H = \frac{L}{2}\sum_{n=1}^{\infty}\int \frac{d^2\vec{k}_\perp}{(2\pi)^2\,2\omega_k}\,\frac{\omega_k}{2}\left\{a\left(\frac{n\pi}{L}, \vec{k}_\perp\right)a^\dagger\left(\frac{n\pi}{L}, \vec{k}_\perp\right) + a^\dagger\left(\frac{n\pi}{L}, \vec{k}_\perp\right)a\left(\frac{n\pi}{L}, \vec{k}_\perp\right)\right\}. \tag{5.65}$$

The vacuum energy per unit area of the planes is (**exercise**)

$$\frac{E_{\text{vacuum}}}{S} = \sum_{n=1}^{\infty}\int \frac{d^2\vec{k}_\perp}{(2\pi)^2}\,\frac{\sqrt{\frac{\pi^2 n^2}{L^2} + \vec{k}_\perp^2 + m^2}}{2}. \tag{5.66}$$

In polar coordinates,

$$\frac{E_{\text{vacuum}}}{S} = \sum_{n=1}^{\infty}\int \frac{dk_\perp k_\perp}{(2\pi)}\,\frac{\sqrt{\frac{\pi^2 n^2}{L^2} + k_\perp^2 + m^2}}{2}. \tag{5.67}$$

This integral appears to be divergent in the limit $k_\perp \to \infty$. For simplicity, we will study it in the massless limit, $m = 0$.

Let us now work with a slightly modified integrand,

$$\frac{E_{\text{vacuum}}}{S} = \sum_{n=1}^{\infty}\int \frac{dk_\perp k_\perp^{1-\delta}}{(2\pi)}\,\frac{\sqrt{\frac{\pi^2 n^2}{L^2} + k_\perp^2}}{2}. \tag{5.68}$$

For suitable values of δ the integral is well defined. We shall perform our calculation for such values, and then we shall analytically continue the result to $\delta = 0$.

Remark. Strictly speaking, there is no single δ for which everything converges: the $k_\perp$ integral needs $\delta > 3$ at large $k_\perp$ but $\delta < 2$ at small $k_\perp$, and the sum over n below needs $\delta > 4$. As always in dimensional regularisation, each factor (the Beta function and the ζ-function) is defined in its own convergence region and then continued analytically; the continuations are unique, and the result agrees with that of a physical (e.g. exponential) cutoff once the cutoff-dependent terms proportional to L and independent of L are dropped (see Ref. [9]). This trick has a deeper meaning. Imagine that space-time has D dimensions, one of them time and one the direction perpendicular to the plates. Then $\vec{k}_\perp$ has $D-2$ components, and in spherical coordinates $d^{D-2}\vec{k}_\perp \propto k_\perp^{D-3}dk_\perp$. For $D = 4-\delta$, this is exactly the modified measure $k_\perp^{1-\delta}dk_\perp$. We are continuing analytically in the number of space-time

dimensions. This method, known as **dimensional regularisation**, was developed by 't Hooft and Veltman (and independently by Bollini and Giambiagi) in 1972; it was the key tool in 't Hooft and Veltman's proof of the renormalisability of gauge theories, which earned them the 1999 Nobel prize. It can be applied to every integral of quantum field theory, by performing the action integral in D rather than four dimensions. It does not remove infinities; it quantifies them, as poles in δ. We will make extensive use of it when we study loop integrals in Part II.

We change variables once again,

$$k_\perp = l_\perp \frac{n\pi}{L}, \tag{5.69}$$

and we obtain,

$$\begin{aligned} \frac{E_{\text{vacuum}}}{S} &= \frac{1}{4\pi}\left(\frac{\pi}{L}\right)^{3-\delta}\left(\sum_{n=1}^{\infty} n^{3-\delta}\right)\int_0^\infty dl_\perp l_\perp^{1-\delta}\sqrt{1+l_\perp^2}, \\ \rightsquigarrow \frac{E_{\text{vacuum}}}{S} &= \frac{1}{8\pi}\left(\frac{\pi}{L}\right)^{3-\delta}\left(\sum_{n=1}^{\infty} \frac{1}{n^{-3+\delta}}\right)\int_0^\infty dl_\perp^2 (l_\perp^2)^{-\frac{\delta}{2}}\sqrt{1+l_\perp^2}. \end{aligned} \tag{5.70}$$

We recognise the sum as the ζ-function,

$$\sum_{n=1}^{\infty} \frac{1}{n^{-3+\delta}} = \zeta(\delta - 3). \tag{5.71}$$

Performing the change of variables,

$$l_\perp^2 \to \frac{x}{1-x}, \tag{5.72}$$

we find that the integral is

$$\int_0^\infty dl_\perp^2 (l_\perp^2)^{-\frac{\delta}{2}}\sqrt{1+l_\perp^2} = \int_0^1 dx x^{-\frac{\delta}{2}}(1-x)^{\frac{\delta-5}{2}} = B(1-\delta/2, -3/2+\delta/2), \tag{5.73}$$

where we have recognised the beta function,

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \tag{5.74}$$

The final result reads,

$$\frac{E_{\text{vacuum}}}{S} = \frac{1}{8\pi}\left(\frac{\pi}{L}\right)^{3-\delta} B(1-\delta/2, -3/2+\delta/2)\zeta(\delta-3). \tag{5.75}$$

The limit $\delta \to 0$ exists and it is finite. This is not an accident: the analytic continuation in δ discards the divergent terms, which are scaleless integrals and vanish in dimensional regularisation. These are the bulk vacuum energy density of unbounded space times the volume between the plates, a term proportional to L whose pressure is balanced by the same vacuum outside the plates, and an L-independent surface energy of the plates (see the remarks at the end of this section). The remainder, which falls off like $1/L^3$ and is finite, is physical. We find,

$$\frac{E_{\text{vacuum}}}{S} = -\frac{\pi^2}{1440 L^3}. \tag{5.76}$$

The vacuum energy depends on the distance between the two planes, on which the field vanishes. Can we realise this in an experiment?

The electromagnetic field is zero inside a conductor. If we place two "infinite" conducting sheets parallel to each other at a distance L, then we can reproduce the boundary conditions of the setup that we have just studied. The quantisation of the electromagnetic field is rather more complicated than that of the Klein-Gordon field, but as far as the vacuum energy is concerned, we shall see that they are very similar.

Our analysis leads to an amazing prediction: two electrically neutral conductors attract each other. This is known as the Casimir effect. Casimir's idea was ingenious: change the vacuum, by not allowing the fields to occupy all of space, and make its energy depend on a geometrical parameter which we can vary, here the distance L. Nature then adjusts L to lower the energy, and this manifests itself as a measurable force. The geometry of parallel plates is chosen because we can solve for the field easily; any other geometry also gives rise to a force, which may even be repulsive. Who carries this force? There are no photons in the vacuum; the number of photons and the expectation value of the field there are zero. The field nevertheless exists and fluctuates, and every mode in the vacuum energy integral is a "virtual" excitation of it. We should not think of fields only in terms of particles. For the Higgs field, the distinction is even more striking: it has a non-zero expectation value everywhere in the universe, while no Higgs particles are present. Notice that the energy of the vacuum E_{vacuum} gets smaller when the conducting plates are closer. Therefore, there is an attractive force between them. The force per unit area (pressure or rather anti-pressure) between the two sheets which set a zero boundary condition for a real Klein-Gordon field is

$$\mathcal{F} = -\frac{\partial \frac{E_{\text{vacuum}}}{S}}{\partial L} = -\frac{\pi^2}{480 L^4}. \tag{5.77}$$

For the experimentally relevant case of the electromagnetic field the pressure is twice the answer that we have just found for the real Klein-Gordon field (due to the two physical polarisations of the electromagnetic field). The Casimir effect has been tested experimentally; for a review see Ref. [8] (arXiv:quant-ph/9907076).

Remark. The Casimir force between real metals can also be derived without reference to zero-point energies, as the (retarded) van der Waals force between the charges in the plates; it then vanishes as the fine-structure constant goes to zero, and the result above is the limit of perfectly conducting plates (R. L. Jaffe, Phys. Rev. D **72** (2005) 021301, hep-th/0503158). The Casimir effect therefore does not by itself prove the reality of the vacuum energy, and it says nothing about its gravitational effect.

Some physics subtleties have been "swept under the carpet" in the above calculation by using an all too convenient regularisation method for the integrals, which may be physically misleading. For example, our regularisation of the energy integral subtracts, secretly, an infinity which corresponds to the vacuum energy in the absence of boundary conditions. Indeed, computing the integral of Eq. (5.58) with the regularisation method introduced here and setting $m = 0$ gives a zero, rather than an infinite, answer. Issues of this type will become clearer when we properly introduce the method of dimensional regularisation. For a more rigorous derivation of the Casimir effect with various boundary conditions we refer to Ref. [9] (arXiv:hep-th/9901011).

5.3 Two real Klein-Gordon fields

Our task is to describe all known particles and their interactions. It is then interesting to study the quantisation of a system with more than one field. We can start simply,

describing a system of two Klein-Gordon fields which differ only in their mass parameter,

$$\begin{aligned}\mathcal{L} &= \frac{1}{2}\left(\partial_\mu\phi_1\right)\left(\partial^\mu\phi_1\right) - \frac{m_1^2}{2}\phi_1^2 \\ &+ \frac{1}{2}\left(\partial_\mu\phi_2\right)\left(\partial^\mu\phi_2\right) - \frac{m_2^2}{2}\phi_2^2.\end{aligned} \tag{5.78}$$

This Lagrangian yields two Klein-Gordon equations as equations of motion for the fields ϕ_1 and ϕ_2:

$$\left(\partial^2 + m_1^2\right)\phi_1 = 0, \qquad \left(\partial^2 + m_2^2\right)\phi_2 = 0, \tag{5.79}$$

with solutions

$$\phi_1(x^\mu) = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_1}\left[a_1(\vec{k})e^{-ik\cdot x} + a_1^\dagger(\vec{k})e^{ik\cdot x}\right], \tag{5.80}$$

$$\phi_2(x^\mu) = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_2}\left[a_2(\vec{k})e^{-ik\cdot x} + a_2^\dagger(\vec{k})e^{ik\cdot x}\right], \tag{5.81}$$

and

$$\omega_1 = \sqrt{\vec{k}^2 + m_1^2}, \qquad \omega_2 = \sqrt{\vec{k}^2 + m_2^2}. \tag{5.82}$$

The Hamiltonian H, the field momentum $\vec{P}$ and the number operator $\hat{N}$ can be written as

$$\hat{N} = \hat{N}_1 + \hat{N}_2, \qquad \hat{N}_i = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_i} a_i^\dagger(\vec{k})a_i(\vec{k}), \tag{5.83}$$

$$\vec{P} = \vec{P}_1 + \vec{P}_2, \qquad \vec{P}_i = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_i}\frac{\vec{k}}{2}\left\{a_i^\dagger, a_i\right\}, \tag{5.84}$$

$$H = H_1 + H_2, \qquad H_i = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_i}\frac{\omega_i}{2}\left\{a_i^\dagger, a_i\right\}, \tag{5.85}$$

where the anticommutator of two operators is defined as:

$$\{A, B\} = AB + BA. \tag{5.86}$$

We can construct particle states in the same fashion as with the Lagrangian of just a single Klein-Gordon field. Products of $a_1^\dagger$ operators acting on the vacuum state create relativistic particles with mass m_1, while $a_2^\dagger$ operators create particles with mass m_2. For example, the states

$$|S_1\rangle \equiv a_1^\dagger(p)|0\rangle, \qquad |S_2\rangle \equiv a_2^\dagger(p)|0\rangle \tag{5.87}$$

have

$$\hat{N}|S_i\rangle = +1|S_i\rangle, \tag{5.88}$$

$$\vec{P}|S_i\rangle = +\vec{p}|S_i\rangle, \tag{5.89}$$

$$H|S_1\rangle = \sqrt{\vec{p}^2 + m_1^2}|S_1\rangle \quad \text{and} \quad H|S_2\rangle = \sqrt{\vec{p}^2 + m_2^2}|S_2\rangle. \tag{5.90}$$

They are degenerate in that they are single-particle states with the same momentum $\vec{p}$; however, as long as the masses are different, $m_1 \neq m_2$, these two states can be distinguished by measuring the energy of the particles.

The possibility of $m_1 = m_2$ is interesting; if in a multi-particle state we replace an $a_1^\dagger$ by an $a_2^\dagger$ operator, or vice versa, the new state is degenerate in all three conserved quantities: the number of particles, the energy and the momentum.

5.3.1 Two equal-mass real Klein-Gordon fields

The special case of $m_1 = m_2 = m$ merits a detailed study, because a new rotation symmetry emerges in the space of fields ϕ_1 and ϕ_2. This leads, according to Noether's theorem, to a conserved quantity which can be interpreted as the charge of particles and antiparticles.

We rewrite the Lagrangian in a form where the symmetry is manifest:

$$\mathcal{L} = \frac{1}{2}\left(\partial^\mu \vec{\phi}\right)^T \left(\partial_\mu \vec{\phi}\right) - \frac{m^2}{2}\vec{\phi}^{\,T}\vec{\phi}, \tag{5.91}$$

where we have defined the field "vector"

$$\vec{\phi} \equiv \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \text{ with its transpose } \vec{\phi}^{\,T} = \begin{pmatrix} \phi_1 \, , & \phi_2 \end{pmatrix}. \tag{5.92}$$

Obviously, the Lagrangian is invariant (**exercise**) if we perform an orthogonal transformation,

$$\vec{\phi} \to \vec{\phi}\,' = R\vec{\phi}, \text{ with } R^T = R^{-1}. \tag{5.93}$$

To compute the conserved current we need to find the change $\delta\phi_i$ of the fields under an infinitesimal symmetry transformation. We then expand,

$$R_{ij} = \delta_{ij} + \theta_{ij} + O(\theta^2). \tag{5.94}$$

Orthogonality of R implies that the matrix θ_{ij} is antisymmetric,

$$R^T_{ij} = R^{-1}_{ij} \rightsquigarrow \delta_{ji} + \theta_{ji} = \delta_{ij} - \theta_{ij} \rightsquigarrow \theta_{ij} = -\theta_{ji}. \tag{5.95}$$

The variation of the fields is:

$$\begin{aligned} \phi_1 \quad &\to \quad \phi_1' = R_{1j}\phi_j \\ &\quad\;\; = \left(\delta_{1j} + \theta_{1j}\right)\phi_j = \phi_1 + \theta_{11}\phi_1 + \theta_{12}\phi_2 \\ &\quad\;\; = \phi_1 + \theta_{12}\phi_2 \\ \rightsquigarrow \delta\phi_1 \quad &= \quad \theta_{12}\phi_2, \end{aligned} \tag{5.96}$$

and, similarly,

$$\delta\phi_2 = \theta_{21}\phi_1 = -\theta_{12}\phi_1. \tag{5.97}$$

The Noether current is,

$$\begin{aligned} J^\mu \quad &= \quad \frac{\partial\mathcal{L}}{\partial\left(\partial_\mu\phi_1\right)}\delta\phi_1 + \frac{\partial\mathcal{L}}{\partial\left(\partial_\mu\phi_2\right)}\delta\phi_2 \\ \rightsquigarrow J^\mu \quad &= \quad \theta_{12}\left[\left(\partial^\mu\phi_1\right)\phi_2 - \left(\partial^\mu\phi_2\right)\phi_1\right], \end{aligned} \tag{5.98}$$

and the "charge"

$$Q = \int d^3\vec{x}\left[\left(\partial^0\phi_1\right)\phi_2 - \left(\partial^0\phi_2\right)\phi_1\right], \tag{5.99}$$

is conserved. Substituting in the above expressions the physical solutions for the fields ϕ_i, and performing the $\vec{x}$ integration, we obtain (**exercise**):

$$Q = -i\int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}\left[a_1(k)a_2^\dagger(k) - a_2(k)a_1^\dagger(k)\right]. \tag{5.100}$$

A Noether charge is not defined uniquely. For example, we can multiply it by a constant factor: obviously, if Q is a time-independent (conserved) quantity, then λQ, where λ is a constant number, is also conserved. The normalisation in Eq. (5.100) is convenient, yielding a Hermitian Q (**exercise**),

$$Q^\dagger = Q. \tag{5.101}$$

We now proceed to find the eigenstates of the Q charge operator. This becomes easy if we construct ladder operators for Q. We compute first the commutators of the charge with a_1 and a_2 (**exercise**). We find

$$\left[Q, a_1(p)\right] = -ia_2(p), \tag{5.102}$$

and

$$\left[Q, a_2(p)\right] = +ia_1(p). \tag{5.103}$$

Define the linear combinations,

$$a(p) \equiv \frac{a_1(p) + ia_2(p)}{\sqrt{2}}, \tag{5.104}$$

and

$$b(p) \equiv \frac{a_1(p) - ia_2(p)}{\sqrt{2}}. \tag{5.105}$$

These are indeed ladder operators (**exercise**), satisfying the following commutation relations with the charge Q:

$$\left[Q, a(p)\right] = -a(p), \qquad \left[Q, b(p)\right] = +b(p). \tag{5.106}$$

By taking the Hermitian conjugate of the above equations, and using $Q^\dagger = Q$, we find that

$$\left[Q, a^\dagger(p)\right] = +a^\dagger(p), \qquad \left[Q, b^\dagger(p)\right] = -b^\dagger(p). \tag{5.107}$$

From a state $|S\rangle$ with charge q,

$$Q\,|S\rangle = q\,|S\rangle\,, \tag{5.108}$$

we can obtain states with charges $q \pm 1$ by applying the ladder operators $a^\dagger(p)$ and $b^\dagger(p)$ (**exercise**),

$$Q\left(a^\dagger(p)\,|S\rangle\right) = (q+1)\left(a^\dagger(p)\,|S\rangle\right), \tag{5.109}$$

$$Q\left(b^\dagger(p)\,|S\rangle\right) = (q-1)\left(b^\dagger(p)\,|S\rangle\right). \tag{5.110}$$

We remark that the operators $a^\dagger, b^\dagger$ are also ladder operators for the Hamiltonian H and the field momentum operator $\vec{P}$. This is a consequence of the fact that $a^\dagger$ and $b^\dagger$ are linear combinations of $a_1^\dagger$ and $a_2^\dagger$.

We now construct the common eigenstates of the charge operator Q, the Hamiltonian H, the momentum operator $\vec{P}$ and the number operator $\hat{N}$. We first observe that the vacuum state is also an eigenstate of the charge operator (**exercise**):

$$Q\,|0\rangle = 0. \tag{5.111}$$

Repeated application of the $a^\dagger$ operator on the vacuum builds states of *positive charge*,

$$\left|\Psi_n^{(+)}\right\rangle \equiv \left(\prod_{i=1}^{n} a^\dagger(p_i)\right)|0\rangle\,, \tag{5.112}$$

with

$$\hat{N}\left|\Psi_n^{(+)}\right\rangle = n\left|\Psi_n^{(+)}\right\rangle, \tag{5.113}$$

$$\vec{P}\left|\Psi_n^{(+)}\right\rangle = \left(\sum_{i=1}^{n}\vec{p}_i\right)\left|\Psi_n^{(+)}\right\rangle, \tag{5.114}$$

$$H\left|\Psi_n^{(+)}\right\rangle = \left(\sum_{i=1}^{n}\sqrt{\vec{p}_i{}^2+m^2}\right)\left|\Psi_n^{(+)}\right\rangle, \tag{5.115}$$

$$Q\left|\Psi_n^{(+)}\right\rangle = +n\left|\Psi_n^{(+)}\right\rangle. \tag{5.116}$$

Repeated application of the $b^\dagger$ operator on the vacuum builds states of *negative charge*,

$$\left|\Psi_n^{(-)}\right\rangle \equiv \left(\prod_{i=1}^{n}b^\dagger(p_i)\right)|0\rangle, \tag{5.117}$$

with

$$\hat{N}\left|\Psi_n^{(-)}\right\rangle = n\left|\Psi_n^{(-)}\right\rangle, \tag{5.118}$$

$$\vec{P}\left|\Psi_n^{(-)}\right\rangle = \left(\sum_{i=1}^{n}\vec{p}_i\right)\left|\Psi_n^{(-)}\right\rangle, \tag{5.119}$$

$$H\left|\Psi_n^{(-)}\right\rangle = \left(\sum_{i=1}^{n}\sqrt{\vec{p}_i{}^2+m^2}\right)\left|\Psi_n^{(-)}\right\rangle, \tag{5.120}$$

$$Q\left|\Psi_n^{(-)}\right\rangle = -n\left|\Psi_n^{(-)}\right\rangle. \tag{5.121}$$

To summarise the main results of this section, the mass degeneracy resulted in a new $O(2)$ symmetry of the Lagrangian. This gives rise to a new conserved quantity, the charge Q. A particle state is now characterised by its momentum, the mass of the particle (or equivalently the energy), and its charge which can be either positive or negative. We have just demonstrated a method to describe physical systems of particles which are accompanied by their anti-particles.

Anti-particles are thus an outcome of symmetry, which requires particles and anti-particles of equal mass. The charge Q must not be confused with the number of particles. A state $a^\dagger(p)b^\dagger(q)|0\rangle$, with one particle and one anti-particle, has $\hat{N}=2$ but $Q=0$. The two conserved quantities originate from different symmetries. In the free theory, H, $\vec{P}$ and $\hat{N}$ are invariant under the independent phase changes $a_1 \to e^{i\alpha_1}a_1$, $a_2 \to e^{i\alpha_2}a_2$, which are not visible as simple transformations of the fields; these give the separate conservation of the numbers of the two types of quanta. When we add interactions, for example with photons, a particle and an anti-particle can annihilate and $\hat{N}$ is no longer conserved. The rotation symmetry, however, can be preserved; its charge Q then becomes the electric charge of the particles. Other kinds of charge, such as the colour of quarks or the weak charge, arise in the same way from larger internal symmetries.

5.3.2 Two real Klein-Gordon fields = one complex Klein-Gordon field

We can recast the original expressions for the operators whose eigenvalues characterise particles $(\hat{N},\vec{P},H,Q)$ in terms of the charge ladder operators a,b, rather than a_1,a_2.

Substituting,

$$a_1 = \frac{a+b}{\sqrt{2}}, \qquad a_2 = \frac{a-b}{\sqrt{2}i}, \tag{5.122}$$

we find

$$\hat{N} = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \left[a^\dagger(\vec{k})a(\vec{k}) + b^\dagger(\vec{k})b(\vec{k}) \right], \tag{5.123}$$

$$\vec{P} = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \frac{\vec{k}}{2} \left[\{a^\dagger, a\} + \{b^\dagger, b\} \right], \tag{5.124}$$

$$H = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \frac{\omega_k}{2} \left[\{a^\dagger, a\} + \{b^\dagger, b\} \right]. \tag{5.125}$$

We can now re-write the field operators in terms of a and b. We find that

$$\phi_1 = \frac{\phi + \phi^\dagger}{\sqrt{2}}, \qquad \phi_2 = \frac{\phi - \phi^\dagger}{\sqrt{2}i}, \tag{5.126}$$

with

$$\phi(x) = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \left[a(k)e^{-ik\cdot x} + b^\dagger(k)e^{ik\cdot x} \right], \qquad k^0 = \omega_k = \sqrt{\vec{k}^2 + m^2}. \tag{5.127}$$

We can cast the Lagrangian representing the fields in terms of ϕ and $\phi^\dagger$, rather than ϕ_1 and ϕ_2. We find,

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \vec{\phi}\right)^T \left(\partial^\mu \vec{\phi}\right) - \frac{1}{2}m^2 \vec{\phi}^T \vec{\phi} = \left(\partial_\mu \phi\right)\left(\partial^\mu \phi^\dagger\right) - m^2 \phi\phi^\dagger, \tag{5.128}$$

where

$$\vec{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \text{ and } \phi = \frac{\phi_1 + i\phi_2}{\sqrt{2}}, \quad \phi^\dagger = \frac{\phi_1 - i\phi_2}{\sqrt{2}}. \tag{5.129}$$

The Lagrangian in the complex-field representation makes it more evident that it yields particle and anti-particle states. We could compute the Hamiltonian and momentum operators directly in terms of ϕ and $\phi^\dagger$ arriving at the same expressions as in the representation with two real fields. To compute the charge Q, we would need to identify the symmetry of this new Lagrangian. It is perhaps already obvious that the Lagrangian is invariant under a field phase-redefinition,

$$\phi \to e^{-i\alpha}\phi, \quad \phi^\dagger \to e^{+i\alpha}\phi^\dagger. \tag{5.130}$$

This is the equivalent, in the complex-field representation, of the rotation symmetry transformation that we found earlier. Let us verify this by performing a rotation on the vector $\vec{\phi}$,

$$\begin{aligned}
\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} &\to \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \\
\rightsquigarrow \begin{pmatrix} \phi_1 \\ i\phi_2 \end{pmatrix} &\to \begin{pmatrix} \cos\theta & -i\sin\theta \\ -i\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \phi_1 \\ i\phi_2 \end{pmatrix} \\
\rightsquigarrow \phi_1 + i\phi_2 &\to e^{-i\theta}\left(\phi_1 + i\phi_2\right) \\
\rightsquigarrow \phi &\to e^{-i\theta}\phi.
\end{aligned} \tag{5.131}$$

Exercise: *We have just observed that the $U(1)$ group of unitary transformations of 1×1 complex matrices is isomorphic to the $SO(2)$ group of rotations in a plane. Prove that for the following groups there exists a $2 \to 1$ homomorphic mapping:*

- $SU(2) \to SO(3)$,
- $SU(2) \times SU(2) \to SO(4)$.

5.4 Conserved charges as generators of symmetry transformations

It may occur that we know a conserved charge but we do not know the symmetry transformation from which it originates. It turns out that we can use the conserved charge in order to uncover the unknown symmetry.

Let us assume that a Lagrangian is symmetric under a generic field transformation,

$$\phi_i \to \phi_i + \delta\phi_i, \tag{5.132}$$

leading to a conserved charge

$$Q = \int d^3x \sum_i \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \phi_i\right)} \delta\phi_i = \int d^3x \sum_i \pi_i \delta\phi_i. \tag{5.133}$$

The commutator of the charge operator Q with a field operator ϕ_j is,

$$\begin{aligned}
\left[Q, \phi_j(x)\right] &= \int d^3y \sum_i \left[\pi_i(y)\delta\phi_i(y), \phi_j(x)\right] \\
&= \int d^3y \sum_i \left(\left[\pi_i(y), \phi_j(x)\right] \delta\phi_i(y) + \pi_i(y) \left[\delta\phi_i(y), \phi_j(x)\right]\right) \\
&= \int d^3y \sum_i \left(-i\delta_{ij}\delta^{(3)}(\vec{y} - \vec{x})\delta\phi_j(x) + 0\right) \\
&\rightsquigarrow \left[Q, \phi_j(x)\right] = -i\delta\phi_j(x).
\end{aligned} \tag{5.134}$$

Exercise: *Verify the above equation for the commutators* $\left[Q, \phi\right]$ *and* $\left[Q, \phi^\dagger\right]$ *of the complex Klein-Gordon field.*

Exercise: *In the case of the real Klein-Gordon field we found that the number operator* $\hat{N}$ *is time independent. How does the Klein-Gordon field transform under the corresponding symmetry transformation?*

5.5 Can the Klein-Gordon field be a wave-function?

In the previous chapter, we performed the quantisation of the Schrödinger field. We found that field quantisation, although an elegant formalism, was not indispensable. In fact, we connected the results of quantum field theory with traditional quantum mechanics, in which the Schrödinger "field" is the wave-function. It is worth asking whether field quantisation is truly necessary for a relativistic field, or whether it is just an alternative and perhaps superfluous formalism.

A one-particle wave-function, as a probability amplitude, is in general complex. Let us then interpret the complex Klein-Gordon field as a wave-function. Recall that the Lagrangian is

$$\mathcal{L}_{KG} = \left(\partial_\mu \psi\right)\left(\partial^\mu \psi^*\right) - m^2 \psi\psi^*. \tag{5.135}$$

We denote the field with ψ in order to emphasise our intention to consider it a wave-function. As such, ψ^* is its complex conjugate and not a Hermitian conjugate operator.

The complex Klein-Gordon field Lagrangian and the Schrödinger field Lagrangian,

$$\mathcal{L}_S = \psi^* i\partial_t \psi - \frac{\left|\nabla\psi\right|^2}{2m}, \tag{5.136}$$

are both symmetric under a phase transformation, as we have discussed in detail. In the Schrödinger case, this led to a conserved charge, which we identified with the probability for the particle to be anywhere in space,

$$\text{Total Probability } = \int d^3\vec{x}\left|\psi(\vec{x},t)\right|^2 = \text{constant}. \tag{5.137}$$

Therefore, we could also interpret the quantity,

$$\rho \equiv \left|\psi(\vec{x},t)\right|^2, \tag{5.138}$$

as a probability density.

An attempt to repeat the same steps for the case of the Klein-Gordon field fails badly. The "total probability" integral is now the charge,

$$Q = -i\int d^3\vec{x}\left[\psi(\vec{x},t)\dot{\psi}^*(\vec{x},t) - \dot{\psi}(\vec{x},t)\psi^*(\vec{x},t)\right], \tag{5.139}$$

with a "density"

$$\rho \equiv -i\left[\psi(\vec{x},t)\dot{\psi}^*(\vec{x},t) - \dot{\psi}(\vec{x},t)\psi^*(\vec{x},t)\right], \tag{5.140}$$

which is not positive definite, and thus physically unacceptable.

A second embarrassment of the Klein-Gordon equation as a relativistic wave-function equation is that it predicts particles with negative energy. This may be accepted for a free particle, if we assume that for whatever reason we can only "see" particles with positive energy. However, if we couple these particles to photons, as is required for all particles with electromagnetic interactions, spontaneous emission of photons will turn all particles into negative-energy particles. Clearly, a single-particle wave-function interpretation of the Klein-Gordon field is wrong and it cannot describe physical particles. Quantum field theory is the only consistent approach.

6

Quantisation of the free electromagnetic field

We will now study the quantisation of the electromagnetic field in empty space. One of the predictions of quantum field theory will be that electromagnetic waves are packaged in photons which travel with the speed of light. Imposing quantisation conditions will not be as straightforward as for the Klein-Gordon field, due to the transversality of the polarisation of the photons. We will therefore need to define carefully which states have physical polarisations. In this chapter we will first review the salient features of classical electrodynamics and then proceed to the quantum theory.

6.1 Maxwell equations and Lagrangian formulation

The equations of Maxwell are:

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{6.1}$$

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0, \tag{6.2}$$

$$\vec{\nabla} \cdot \vec{E} = \rho, \tag{6.3}$$

$$\vec{\nabla} \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = \vec{j}. \tag{6.4}$$

We now introduce the four-vector,

$$A^\mu \equiv \left(\phi, \vec{A}\right), \tag{6.5}$$

where the components are chosen such that

$$\vec{B} = \vec{\nabla} \times \vec{A}, \tag{6.6}$$

and

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi. \tag{6.7}$$

Then, the first two Maxwell equations, Eqs. (6.1) and (6.2), are automatically satisfied,

$$\vec{\nabla} \cdot \vec{B} = \vec{\nabla} \cdot \left(\vec{\nabla} \times \vec{A}\right) = 0, \tag{6.8}$$

and

$$\vec{\nabla}\times\left(-\frac{\partial\vec{A}}{\partial t}-\vec{\nabla}\phi\right)+\frac{\partial}{\partial t}\vec{\nabla}\times\vec{A}=\vec{\nabla}\times\left(\vec{\nabla}\phi\right)=\epsilon_{ijk}\partial_i\partial_j\phi=0. \tag{6.9}$$

We now define the tensor,

$$F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu. \tag{6.10}$$

The components of $F^{\mu\nu}$ are,

$$F^{0i}=\partial^0 A^i-\partial^i A^0=\left(\frac{\partial\vec{A}}{\partial t}+\vec{\nabla}\phi\right)^{(i)}=-E^i, \tag{6.11}$$

and

$$F^{ij}=\partial^i A^j-\partial^j A^i=-\epsilon^{ijk}B^k. \tag{6.12}$$

All other components are zero, since $F^{\mu\nu}$ is antisymmetric. In matrix form, we have

$$F^{\mu\nu}=\begin{pmatrix}0 & -E^1 & -E^2 & -E^3\\ E^1 & 0 & -B^3 & B^2\\ E^2 & B^3 & 0 & -B^1\\ E^3 & -B^2 & B^1 & 0\end{pmatrix}. \tag{6.13}$$

This is called the *electromagnetic field tensor*.

We now define the current four-vector,

$$j^\nu\equiv\left(\rho,\vec{j}\right). \tag{6.14}$$

We can easily verify that the remaining Maxwell equations are given by the compact equation,

$$\partial_\mu F^{\mu\nu}=j^\nu. \tag{6.15}$$

For $\nu=0$,

$$\begin{aligned}&\partial_\mu F^{\mu 0}=\rho\\ \rightsquigarrow\quad&\partial_0 F^{00}+\sum_{i=1}^{3}\partial_i F^{i0}=\rho\\ \rightsquigarrow\quad&\vec{\nabla}\cdot\vec{E}=\rho.\end{aligned} \tag{6.16}$$

For $\nu=1$,

$$\begin{aligned}&\partial_\mu F^{\mu 1}=j^1\\ \rightsquigarrow\quad&\partial_0 F^{01}+\sum_{i=1}^{3}\partial_i F^{i1}=j^1\\ \rightsquigarrow\quad&-\frac{\partial E^1}{\partial t}+\frac{\partial B^3}{\partial x_2}-\frac{\partial B^2}{\partial x_3}=j^1,\end{aligned} \tag{6.17}$$

which is the first component of the vector equation

$$-\frac{\partial\vec{E}}{\partial t}+\vec{\nabla}\times\vec{B}=\vec{j}. \tag{6.18}$$

In this form the Lorentz covariance of the Maxwell equations is manifest. Charge density and current form a four-vector: a boost can bring moving charges to rest, turning a current into a static density, and vice versa. Similarly, a Lorentz transformation mixes the scalar and the vector potential. Contracting one index of $F^{\mu\nu}$ with ∂_μ leaves a vector, and both sides of Eq. (6.15) transform in the same way. Historically, the requirement that mechanics shares this symmetry of electrodynamics was one of the roads to special relativity.

In the absence of sources, the Maxwell equations, expressed in terms of the vector potential A^μ, take the form

$$\begin{aligned} & \partial_\mu F^{\mu\nu} = 0 \\ \rightsquigarrow \quad & \partial_\mu \left(\partial^\mu A^\nu - \partial^\nu A^\mu\right) = 0 \\ \rightsquigarrow \quad & \partial^2 A^\nu - \partial^\nu \left(\partial_\mu A^\mu\right) = 0. \end{aligned} \tag{6.19}$$

6.1.1 Classical gauge invariance and gauge-fixing

The Maxwell equations, Eq. (6.15), are invariant under *gauge transformations*, where the vector potential transforms as

$$A^\mu \to A^{\mu\prime} = A^\mu + \partial^\mu \chi, \tag{6.20}$$

or, in components,

$$\phi \to \phi' = \phi + \dot{\chi}, \tag{6.21}$$

$$\vec{A} \to \vec{A}' = \vec{A} - \vec{\nabla}\chi. \tag{6.22}$$

Here, χ is an arbitrary function of the space-time coordinates. Indeed, the field tensor $F^{\mu\nu}$ is invariant under this transformation,

$$\begin{aligned} F^{\mu\nu} \to F^{\mu\nu\prime} &= \partial^\mu A'^\nu - \partial^\nu A'^\mu \\ &= \partial^\mu \left(A^\nu + \partial^\nu \chi\right) - \partial^\nu \left(A^\mu + \partial^\mu \chi\right) \\ &= F^{\mu\nu}. \end{aligned} \tag{6.23}$$

We can often simplify our calculations if we exploit gauge invariance. Suppose that we are given a field A^μ satisfying the Maxwell equations, Eq. (6.19). We can "fix the gauge", i.e. perform a special gauge transformation, $A^\mu \to A^{\mu\prime} = A^\mu + \partial^\mu \chi$, such that the second term in Eq. (6.19) disappears,

$$0 = \partial_\mu A^{\mu\prime} = \partial_\mu A^\mu + \partial^2 \chi \rightsquigarrow \partial^2 \chi = -\partial_\mu A^\mu. \tag{6.24}$$

Such a χ always exists. The equation $\partial^2 \chi = -\partial_\mu A^\mu$ is a wave equation with a known source, analogous to finding the potential of a given charge distribution, and it always has a solution, even if we cannot write it in closed form. This particular choice of gauge is known as the "Lorenz gauge". (It is named after L. Lorenz, but it is often attributed to H. A. Lorentz, which is also fitting since the condition $\partial_\mu A^\mu = 0$ is Lorentz invariant.) In the Lorenz gauge, the Maxwell equations take the very simple form (dropping the prime)

$$\partial^2 A^\mu = j^\mu. \tag{6.25}$$

In the vacuum, $j^\nu = 0$, we have

$$\partial^2 A^\mu = 0. \tag{6.26}$$

This equation is particularly easy to solve in terms of plane waves, and we write the general real classical field solution as an integral,

$$A_\mu(x) = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \sum_{\lambda=0}^{3} \epsilon_\mu^{(\lambda)}(k) \left[a_\lambda(k) e^{-ik\cdot x} + a_\lambda^*(k) e^{ik\cdot x} \right], \tag{6.27}$$

with $k^\mu = \left(\omega_k, \vec{k}\right)$ and $\omega_k = \left|\vec{k}\right|$. We have introduced a basis of four-vectors which determine the direction of the electromagnetic field (polarisation). For example, these vectors can be taken as

$$\epsilon^{(0)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \epsilon^{(1)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \epsilon^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \epsilon^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{6.28}$$

As we can easily verify, the polarisation vectors are normalised as

$$g^{\mu\nu} \epsilon_\mu^{(\lambda)} \epsilon_\nu^{(\lambda')} = g^{\lambda\lambda'}, \quad g_{\lambda\lambda'} \epsilon_\mu^{(\lambda)} \epsilon_\nu^{(\lambda')} = g_{\mu\nu}. \tag{6.29}$$

Exercise: *Verify the above equations.*

The vector potential must be real, since the electric and magnetic fields, and the forces they exert, are real; hence the complex conjugate term in Eq. (6.27). We now face the main challenge of this chapter. The solution contains four polarisations for every momentum, and after quantisation the operators $a_\lambda^\dagger$ will create photons with four polarisations. Physical photons, like classical electromagnetic waves, have only two, transverse, polarisations. The Lorenz condition removes only one of them, since it requires $k \cdot \epsilon = 0$ for the physical combinations. We need a mechanism to remove two of the four polarisations; we will see that it emerges naturally from a careful quantisation.

6.1.2 Lagrangian of the electromagnetic field

The following Lagrangian,

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{6.30}$$

gives rise to the Maxwell equations as classical equations of motion. Besides checking that it gives the Maxwell equations, we can arrive at this Lagrangian from symmetry. Gauge invariance is guaranteed if the Lagrangian is built out of the gauge-invariant $F^{\mu\nu}$, and Lorentz invariance requires a scalar. The simplest choice is $F_{\mu\nu}F^{\mu\nu}$; the factor $-\frac{1}{4}$ is a normalisation which gives directly the correct energy of the field. There is a second candidate, $\epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}$, with $\epsilon^{\mu\nu\rho\sigma}$ the totally antisymmetric tensor. It is a total derivative, $\epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} = 4\,\partial_\mu \left(\epsilon^{\mu\nu\rho\sigma} A_\nu \partial_\rho A_\sigma\right)$, and it does not modify the equations of motion (**exercise**). Explicitly,

$$\mathcal{L} = -\frac{1}{2} \left[\left(\partial_\mu A_\nu\right) \left(\partial^\mu A^\nu\right) - \left(\partial_\mu A_\nu\right) \left(\partial^\nu A^\mu\right) \right]. \tag{6.31}$$

It is easy to verify that the Euler-Lagrange equations are (**exercise**)

$$\begin{aligned} & \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu A_\nu\right)} - \frac{\partial \mathcal{L}}{\partial A_\nu} = 0 \\ \rightsquigarrow \quad & \partial^2 A^\nu - \partial^\nu \left(\partial_\mu A^\mu\right) = 0. \end{aligned} \tag{6.32}$$

6.2 Quantisation of the electromagnetic field

The conjugate momenta of the field components A^μ are

$$\pi^\mu = \frac{\partial \mathcal{L}}{\partial \left(\partial_0 A_\mu\right)} = F^{\mu 0}. \tag{6.33}$$

One would like to impose the bosonic quantisation condition,

$$\left[A^\mu(x), \pi^\nu(x')\right] = i g^{\mu\nu} \delta^{(3)}\left(\vec{x} - \vec{x}'\right). \tag{6.34}$$

Notice the presence of $g^{\mu\nu}$, which is needed in order to maintain covariance on the indices μ, ν of the quantisation condition. We immediately come across a problem. For $\mu = 0$, since

$$\pi^0 = F^{00} = 0,$$

we encounter an inconsistency,

$$0 = \left[A^0(x), \pi^0(x')\right] = i\delta^{(3)}(\vec{x} - \vec{x}') \neq 0. \tag{6.35}$$

Our first attempt at quantising the classical Lagrangian of the electromagnetic field has failed. However, it is possible to quantise a different Lagrangian which nevertheless gives rise to the same physics as the Lagrangian of Eq. (6.30). We can exploit the fact that classical physics is the same in every gauge for A^μ.

Consider a modified Lagrangian,

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2\xi} \left(\partial_\mu A^\mu\right)^2. \tag{6.36}$$

This yields the following Euler-Lagrange equations (which are not invariant under gauge transformations):

$$\partial^2 A^\mu - \left(1 - \frac{1}{\xi}\right) \partial^\mu \left(\partial_\nu A^\nu\right) = 0. \tag{6.37}$$

Notice that for a special value $\xi = 1$ (the "Feynman gauge"), the equations of motion are identical to Maxwell equations, $\partial^2 A^\mu = 0$, in the Lorenz gauge. The parameter ξ is called the gauge-fixing parameter; we shall meet it again in the photon propagator of Chapter 11 and in quantum electrodynamics. For the new Lagrangian and with $\xi = 1$, we find

$$\pi^0 = \frac{\partial \mathcal{L}}{\partial \dot{A}_0} = -\partial_\mu A^\mu, \tag{6.38}$$

which is not zero, as long as we do not impose the Lorenz condition $\partial_\mu A^\mu = 0$. This is at first sight irreconcilable with classical physics; however, there is a way out. We may think of the classical electromagnetic field as the expectation value of a quantum field in a physical quantum state $\left|\psi\right\rangle$, i.e.

$$A^\mu_{\text{classic}} \sim \left\langle\psi\right| A^\mu_{\text{operator}} \left|\psi\right\rangle. \tag{6.39}$$

We must guarantee that the expectation value of $\partial_\mu A^\mu$ on physical states $\left|\psi\right\rangle$ vanishes,

$$\left\langle\psi\right| \partial_\mu A^\mu \left|\psi\right\rangle = 0, \tag{6.40}$$

which may be achieved even though $\partial_\mu A^\mu \neq 0$ as an operator.

We can now carry on with the quantisation of the Lagrangian of Eq. (6.36). The quantisation conditions of Eq. (6.34), with the expressions of the conjugate momenta $\pi^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)}$ substituted, yield

$$\left[\dot{A}_\mu\left(\vec{x},t\right), A_\nu\left(\vec{x}',t\right)\right] = i g_{\mu\nu}\delta^{(3)}\left(\vec{x}-\vec{x}'\right). \tag{6.41}$$

We now substitute the solution of Eq. (6.27) into Eq. (6.41). We obtain,

$$\left[a_\lambda(k), a^\dagger_{\lambda'}(k')\right] = -g^{\lambda\lambda'} 2k_0 (2\pi)^3 \delta^{(3)}\left(\vec{k}-\vec{k}'\right). \tag{6.42}$$

We can safely interpret the $a_i(k)$ and $a_i^\dagger(k)$ for $i = 1,2,3$ as annihilation and creation operators, respectively, of photons with space-like polarisations $\epsilon^\mu_{(i)}$. This is, however, problematic for the operators $a_0(k), a_0^\dagger(k)$, which satisfy

$$\left[a_0(k), a_0^\dagger(k')\right] = -2k_0 (2\pi)^3 \delta^{(3)}\left(\vec{k}-\vec{k}'\right). \tag{6.43}$$

Consider a general one-photon state with a time-like polarisation,

$$|1,\lambda=0\rangle = \int \frac{d^3\vec{k}}{(2\pi)^3(2k_0)} f(k) a_0^\dagger(k)\,|0\rangle\,. \tag{6.44}$$

The norm of the state is,

$$\langle 1,\lambda=0|\,1,\lambda=0\rangle = -\int \frac{d^3\vec{k}}{(2\pi)^3(2k_0)} \left|f(k)\right|^2 \langle 0|\,0\rangle\,, \tag{6.45}$$

which is negative. In general, a state with n_t photons with a time-like polarisation has a norm

$$\langle n_t,\lambda=0|\,n_t,\lambda=0\rangle = (-1)^{n_t}. \tag{6.46}$$

This is a serious problem since it essentially gives rise to negative probabilities for transition amplitudes among states with time-like polarisations.

We need Eq. (6.40) to hold for any physical state. We will now describe how this can be achieved following the prescription of Gupta and Bleuler. We separate the electromagnetic field into a positive and a negative frequency component,

$$A_\mu(x) = A^{(+)}_\mu(x) + A^{(-)}_\mu(x), \tag{6.47}$$

with

$$A^{(+)\mu} = \int \frac{d^3\vec{k}}{(2\pi)^3 2k_0} \sum_\lambda \epsilon^\mu_\lambda a_\lambda(k) e^{-ik\cdot x}, \tag{6.48}$$

and

$$A^{(-)\mu} = \int \frac{d^3\vec{k}}{(2\pi)^3 2k_0} \sum_\lambda \epsilon^\mu_\lambda a^\dagger_\lambda(k) e^{+ik\cdot x}. \tag{6.49}$$

Physical states $\left|\psi_{\text{phys}}\right\rangle$ are required to satisfy

$$\partial_\mu A^{(+)\mu}\left|\psi_{\text{phys}}\right\rangle = 0. \tag{6.50}$$

This is satisfied automatically by the vacuum state, $\left|\psi_{\text{phys}}\right\rangle = |0\rangle$, since $A^{(+)\mu}$ contains only annihilation operators. We then also find that

$$\begin{aligned} & \left\langle\psi_{\text{phys}}\right| \partial_\mu A^\mu \left|\psi_{\text{phys}}\right\rangle \\ = & \left\langle\psi_{\text{phys}}\right| \partial_\mu A^{(+)\mu} \left|\psi_{\text{phys}}\right\rangle + \left\langle\psi_{\text{phys}}\right| \partial_\mu A^{(-)\mu} \left|\psi_{\text{phys}}\right\rangle \\ = & \left\langle\psi_{\text{phys}}\right| \partial_\mu A^{(+)\mu} \left|\psi_{\text{phys}}\right\rangle + \left\langle\psi_{\text{phys}}\right| \partial_\mu A^{(+)\mu} \left|\psi_{\text{phys}}\right\rangle^* \\ = & 2\text{Re} \left\langle\psi_{\text{phys}}\right| \partial_\mu A^{(+)\mu} \left|\psi_{\text{phys}}\right\rangle \\ = & 0. \end{aligned} \tag{6.51}$$

The Gupta-Bleuler condition takes the explicit form

$$\partial_\mu A^{(+)\mu} \left|\psi_{\text{phys}}\right\rangle = (-i) \int \frac{d^3\vec{k}}{(2\pi)^3 2k_0} e^{-ik\cdot x} \sum_{\lambda=0}^{3} k \cdot \epsilon_\lambda(k) a_\lambda(k) \left|\psi_{\text{phys}}\right\rangle = 0. \tag{6.52}$$

It is easy to verify that the above is not satisfied for photon-states with a purely time-like polarisation, $\left|\psi\right\rangle = |1, \lambda = 0\rangle$, as in Eq. (6.44).

What is a physical state for a single photon? Consider the general one-photon state with a momentum $p^\mu = (1, 0, 0, 1)$, choosing its momentum along the z-axis. This state can be written as

$$|\gamma\rangle = \sum_{\lambda=0}^{3} c_\lambda a_\lambda^\dagger(p) |0\rangle , \tag{6.53}$$

where c_λ are arbitrary coefficients determining the mixture of polarisations in the photon physical state. We must have

$$\begin{aligned} 0 &= \partial_\mu A^{(+)\mu} |\gamma\rangle \\ \rightsquigarrow 0 &= \left(\sum_{\lambda\lambda'} c_\lambda p \cdot \epsilon_{\lambda'} g^{\lambda\lambda'} \right) |0\rangle \\ \rightsquigarrow 0 &= \left(\sum_{\lambda\lambda'} c_\lambda p \cdot \epsilon_{\lambda'} g^{\lambda\lambda'} \right). \end{aligned} \tag{6.54}$$

From the expressions for the vector polarisations in Eq. (6.28) and the momentum $p = (1, 0, 0, 1)$ we derive

$$p \cdot \epsilon_1 = p \cdot \epsilon_2 = 0, \tag{6.55}$$

and

$$p \cdot \epsilon_0 = -p \cdot \epsilon_3. \tag{6.56}$$

Then Eq. (6.54) becomes

$$c_0 + c_3 = 0 \rightsquigarrow c_0 = -c_3. \tag{6.57}$$

Therefore, a component of the photon state with time-like polarisation must be accompanied by an "opposite" component with longitudinal polarisation. A general physical photon state with momentum $\vec{p}$ is thus of the form

$$\left|\gamma, \vec{p}\right\rangle = \left(c \left[a_0^\dagger(p) - a_3^\dagger(p) \right] + c_1 a_1^\dagger(p) + c_2 a_2^\dagger(p) \right) |0\rangle . \tag{6.58}$$

Notice that the norm of the state is

$$\begin{aligned}\left\langle \gamma,\vec{p}\right|\left.\gamma,\vec{p}\right\rangle &= |c|^2\left(\left\langle \vec{p},\lambda=0\right|\left.\vec{p},\lambda=0\right\rangle+\left\langle \vec{p},\lambda=3\right|\left.\vec{p},\lambda=3\right\rangle\right)\\ &\quad+|c_1|^2\left\langle \vec{p},\lambda=1\right|\left.\vec{p},\lambda=1\right\rangle+|c_2|^2\left\langle \vec{p},\lambda=2\right|\left.\vec{p},\lambda=2\right\rangle\\ &= \left\{|c|^2(-1+1)+|c_1|^2 1+|c_2|^2 1\right\}(2\pi)^3 2|\vec{p}|\delta^{(3)}(0)\\ &= \left(|c_1|^2+|c_2|^2\right)(2\pi)^3 2|\vec{p}|\delta^{(3)}(0). \end{aligned} \tag{6.59}$$

It receives contributions only from the transverse polarisations, since the contributions of the time-like and longitudinal components cancel each other.[1] Let us calculate the energy in a one-photon monochromatic physical state. After a calculation which is by now standard, we find that the Hamiltonian is given by

$$H=\int\frac{d^3\vec{k}}{(2\pi)^3 2k_0}k_0\left[\sum_{\lambda=1}^{3}a_\lambda^\dagger(k)a_\lambda(k)-a_0^\dagger(k)a_0(k)\right], \tag{6.60}$$

with $k_0=\left|\vec{k}\right|>0$. We can easily prove that $\left|\gamma,\vec{p}\right\rangle$ is an eigenstate of the Hamiltonian:

$$H\left|\gamma,\vec{p}\right\rangle=|\vec{p}|\left|\gamma,\vec{p}\right\rangle. \tag{6.61}$$

We also find

$$\langle\gamma|H|\gamma\rangle=|\vec{p}|\langle\gamma|\,\gamma\rangle, \tag{6.62}$$

The energy of the state is

$$E_{|\gamma,\vec{p}\rangle}=\frac{\langle\gamma|H|\gamma\rangle}{\langle\gamma|\,\gamma\rangle}=|\vec{p}|. \tag{6.63}$$

The contributions of the time-like and longitudinal components cancel each other in all physical quantities. We shall see realistic examples of this cancellation for scattering amplitudes in quantum electrodynamics.

Remark. The admixture $c\left[a_0^\dagger(p)-a_3^\dagger(p)\right]|0\rangle$ is a state of zero norm, orthogonal to all physical states. Two physical states which differ by such a component give identical expectation values of all gauge-invariant observables; the expectation value of A^μ changes only by a gauge transformation $\partial^\mu\chi$. The physical Hilbert space is therefore the space of equivalence classes of states satisfying Eq. (6.50), which has a positive-definite norm and contains exactly the two transverse polarisations (see C. Itzykson and J.-B. Zuber, *Quantum Field Theory*, Sec. 3-2).

[1] The infinite normalisation ($\delta^{(3)}(0)$) is an artefact of considering a monochromatic state. It is finite for realistic states, which are superpositions of photons, $|\gamma\rangle=\int\frac{d^3\vec{p}}{2p_0(2\pi)^3}f(\vec{p})\left|\gamma,\vec{p}\right\rangle$.

7

The Dirac equation

Dirac identified the source of the problems in interpreting the Klein-Gordon equation as a wave-function equation in the fact that it is a second-order differential equation in time. The correspondence

$$i\partial_t \to E$$

results in a second-order polynomial equation for the energy of a free particle, which of course has both a positive and a negative solution. Dirac's idea was then to find a linear equation whose differential operator represents a sort of "square root" of the differential operator in the Klein-Gordon equation. Explicitly, Dirac's linear equation is

$$\left(i\gamma^\mu\partial_\mu - m\right)\psi = 0. \tag{7.1}$$

Taking the square root of a sum of second derivatives is an old mathematical problem. Hamilton discovered in 1843 the quaternions, three anti-commuting "imaginary units" $\hat{\imath}, \hat{\jmath}, \hat{k}$ with $\hat{\imath}^2 = \hat{\jmath}^2 = \hat{k}^2 = -1$. With them,

$$\left(\hat{\imath}\partial_x + \hat{\jmath}\partial_y + \hat{k}\partial_z\right)^2 = -\left(\partial_x^2 + \partial_y^2 + \partial_z^2\right),$$

i.e. a square root of minus the Laplacian. Clifford generalised this construction in 1878 to an arbitrary number of dimensions and to other signatures. In 1928 this was known only to a few specialists, and Dirac had to re-invent it for the four-dimensional space-time with the signature $(+,-,-,-)$.

Multiplying the Dirac equation by the differential operator $\left(-i\gamma^\mu\partial_\mu - m\right)$ from the left, we obtain

$$\begin{aligned}
&\rightsquigarrow \quad \left(-i\gamma^\nu\partial_\nu - m\right)\left(i\gamma^\mu\partial_\mu - m\right)\psi = 0\\
&\rightsquigarrow \quad \left(\gamma^\mu\gamma^\nu\partial_\mu\partial_\nu + m^2\right)\psi = 0\\
&\rightsquigarrow \quad \left(\frac{1}{2}\gamma^\mu\gamma^\nu\partial_\mu\partial_\nu + \frac{1}{2}\gamma^\nu\gamma^\mu\partial_\nu\partial_\mu + m^2\right)\psi = 0\\
&\rightsquigarrow \quad \left(\frac{1}{2}\gamma^\mu\gamma^\nu\partial_\mu\partial_\nu + \frac{1}{2}\gamma^\nu\gamma^\mu\partial_\mu\partial_\nu + m^2\right)\psi = 0\\
&\rightsquigarrow \quad \left(\frac{1}{2}\left\{\gamma^\mu,\gamma^\nu\right\}\partial_\mu\partial_\nu + m^2\right)\psi = 0,
\end{aligned} \tag{7.2}$$

which for

$$\left\{\gamma^\mu,\gamma^\nu\right\} = 2g^{\mu\nu}, \tag{7.3}$$

gives the Klein-Gordon equation.

In other words, if we can construct objects γ^μ which satisfy the algebra of Eq. (7.3), known as the Clifford algebra, then every solution $\psi(\vec{x}, t)$ of the Dirac equation will also be a solution of the Klein-Gordon equation.

The objects γ^μ are anti-commuting and they cannot be just numbers. They have to be matrices. The smallest dimension for which we can find four anti-commuting matrices $\gamma^0, \gamma^1, \gamma^2, \gamma^3$ is four. In the following we shall construct the so-called "gamma matrices" in four dimensions.

7.1 Mathematical interlude

It is useful to review here the properties of the 2×2 Pauli matrices, which we shall use as building blocks for constructing γ-matrices by taking their Kronecker product.

7.1.1 Pauli matrices and their properties

The Pauli matrices are,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = -i \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{7.4}$$

The Pauli matrices satisfy the following commutation and anti-commutation relations,

$$\left[\sigma_i, \sigma_j\right] = 2i\epsilon_{ijk}\sigma_k \qquad (\epsilon_{123} = 1), \tag{7.5}$$

and

$$\left\{\sigma_i, \sigma_j\right\} = 2\delta_{ij} I_{2\times 2}. \tag{7.6}$$

Adding the two equations together, we find that the product of two Pauli matrices is,

$$\sigma_i \sigma_j = \delta_{ij} I_{2\times 2} + i\epsilon_{ijk}\sigma_k. \tag{7.7}$$

For example,

$$\sigma_i^2 = I_{2\times 2}, \quad \sigma_1\sigma_2 = i\sigma_3, \quad \sigma_3\sigma_2 = -i\sigma_1, \quad \ldots \tag{7.8}$$

The determinant and trace of Pauli matrices are,

$$\det(\sigma_i) = -1, \tag{7.9}$$

and

$$\mathrm{Tr}(\sigma_i) = 0. \tag{7.10}$$

As can be explicitly verified, the Pauli matrices are Hermitian,

$$\sigma_i^\dagger = \sigma_i. \tag{7.11}$$

The Pauli matrices, together with the unit matrix $I_{2\times 2}$, constitute a basis of general 2×2 matrices $M_{2\times 2}$. We write

$$M_{2\times 2} = a_0 I_{2\times 2} + \sum_{i=1}^{3} a_i \sigma_i. \tag{7.12}$$

From the above decomposition we have,

$$\mathrm{Tr}(M) = a_0 \mathrm{Tr}(I_{2\times 2}) + \sum_{i=1}^{3} a_i \mathrm{Tr}(\sigma_i) = 2a_0, \tag{7.13}$$

and (**exercise**),

$$\mathrm{Tr}\,(\sigma_k M) = 2a_k. \tag{7.14}$$

We can therefore write,

$$M_{2\times 2} = \frac{\mathrm{Tr}\,(M)}{2} I_{2\times 2} + \sum_{i=1}^{3} \frac{\mathrm{Tr}\,(M\sigma_i)}{2} \sigma_i. \tag{7.15}$$

7.1.2 Kronecker product of 2×2 matrices

Let us define the Kronecker product of two 2×2 matrices

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, \tag{7.16}$$

as

$$A \otimes B \equiv \begin{pmatrix} b_{11}A & b_{12}A \\ b_{21}A & b_{22}A \end{pmatrix} = \begin{pmatrix} b_{11}a_{11} & b_{11}a_{12} & b_{12}a_{11} & b_{12}a_{12} \\ b_{11}a_{21} & b_{11}a_{22} & b_{12}a_{21} & b_{12}a_{22} \\ b_{21}a_{11} & b_{21}a_{12} & b_{22}a_{11} & b_{22}a_{12} \\ b_{21}a_{21} & b_{21}a_{22} & b_{22}a_{21} & b_{22}a_{22} \end{pmatrix}. \tag{7.17}$$

The following properties can be proved (**exercise**) explicitly,

$$\det\,(A \otimes B) = (\det A)^2\,(\det B)^2\,, \tag{7.18}$$

$$\mathrm{Tr}\,(A \otimes B) = \mathrm{Tr}\,(A)\,\mathrm{Tr}\,(B)\,, \tag{7.19}$$

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger, \tag{7.20}$$

$$(A + C) \otimes B = A \otimes B + C \otimes B, \tag{7.21}$$

$$A \otimes (B + C) = A \otimes B + A \otimes C, \tag{7.22}$$

and

$$(A_1 \otimes B_1)\,(A_2 \otimes B_2) = A_1 A_2 \otimes B_1 B_2. \tag{7.23}$$

7.2 Dirac representation of γ-matrices

The matrices

$$\gamma^0 = I_{2\times 2} \otimes \sigma_3, \tag{7.24}$$

and

$$\gamma^j = \sigma_j \otimes (i\sigma_2)\,, \qquad (j = 1, 2, 3), \tag{7.25}$$

satisfy the Clifford algebra,

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} I_{4\times 4}. \tag{7.26}$$

We find,

$$\begin{aligned} \left(\gamma^0\right)^2 &= (I_{2\times 2} \otimes \sigma_3)\,(I_{2\times 2} \otimes \sigma_3) \\ &= I_{2\times 2}^2 \otimes \sigma_3^2 = I_{2\times 2} \otimes I_{2\times 2} = I_{4\times 4} \\ \rightsquigarrow \{\gamma^0, \gamma^0\} &= 2g^{00} I_{4\times 4}. \end{aligned} \tag{7.27}$$

Also, for the anti-commutator $\{\gamma^j, \gamma^k\}$ with $j, k = 1, 2, 3$ we have

$$\begin{aligned}
\{\gamma^j, \gamma^k\} &= \gamma^j\gamma^k + \gamma^k\gamma^j \\
&= \left(\sigma_j \otimes (i\sigma_2)\right)\left(\sigma_k \otimes (i\sigma_2)\right) + \left(\sigma_k \otimes (i\sigma_2)\right)\left(\sigma_j \otimes (i\sigma_2)\right) \\
&= \sigma_j\sigma_k \otimes (i\sigma_2)^2 + \sigma_k\sigma_j \otimes (i\sigma_2)^2 \\
&= \{\sigma_j, \sigma_k\} \otimes (-I_{2\times 2}) = \left(2\delta_{jk} I_{2\times 2}\right) \otimes (-I_{2\times 2}) \\
\rightsquigarrow \{\gamma^j, \gamma^k\} &= -2\delta^{jk} I_{4\times 4} \\
\rightsquigarrow \{\gamma^j, \gamma^k\} &= 2g^{jk} I_{4\times 4}.
\end{aligned} \tag{7.28}$$

Finally, the anti-commutator $\{\gamma^0, \gamma^j\}$ with $j = 1, 2, 3$ is

$$\begin{aligned}
\{\gamma^0, \gamma^j\} &= \gamma^0\gamma^j + \gamma^j\gamma^0 \\
&= (I_{2\times 2} \otimes \sigma_3)\left(\sigma_j \otimes (i\sigma_2)\right) + \left(\sigma_j \otimes (i\sigma_2)\right)(I_{2\times 2} \otimes \sigma_3) \\
&= \sigma_j \otimes (i\sigma_2\sigma_3) + \sigma_j \otimes (i\sigma_3\sigma_2) \\
&= \sigma_j \otimes (i\{\sigma_2, \sigma_3\}) \\
&= \sigma_j \otimes \left(i2\delta^{23} I_{2\times 2}\right) = \sigma_j \otimes 0_{2\times 2} = 0_{4\times 4} \\
\rightsquigarrow \{\gamma^0, \gamma^j\} &= 2g^{0j} I_{4\times 4}.
\end{aligned} \tag{7.29}$$

We have therefore found a set of γ-matrices satisfying the Clifford algebra. It is instructive to look at the same problem in a two-dimensional space-time, with coordinates (t, x). There we need only two matrices, γ^0 and γ^1, and 2×2 matrices suffice,

$$\gamma^0 = \sigma_3, \qquad \gamma^1 = i\sigma_2, \qquad \{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} I_{2\times 2}.$$

Their product, $\gamma^0\gamma^1 = \sigma_1$, anti-commutes with both of them. Notice that the second factor of the Kronecker products in our four-dimensional $\gamma^0 = I_{2\times 2} \otimes \sigma_3$ and $\gamma^j = \sigma_j \otimes (i\sigma_2)$ is exactly one of these two-dimensional γ-matrices, while the Pauli matrices in the first factor take care of the three space directions. The construction continues to higher even dimensions: the γ-matrices are 2×2 in two, 4×4 in four, and 8×8 in six space-time dimensions.

This set is not unique. Let us define new matrices,

$$\gamma^{\mu\prime} = M_{4\times 4}^{-1} \gamma^\mu M_{4\times 4}, \tag{7.30}$$

with $M_{4\times 4}$ any 4×4 matrix with an inverse. Then,

$$\begin{aligned}
\{\gamma^{\mu\prime}, \gamma^{\nu\prime}\} &= M_{4\times 4}^{-1} \{\gamma^\mu, \gamma^\nu\} M_{4\times 4} \\
&= 2g^{\mu\nu} M_{4\times 4}^{-1} I_{4\times 4} M_{4\times 4} \\
&= 2g^{\mu\nu} I_{4\times 4}.
\end{aligned} \tag{7.31}$$

The matrices $\gamma^{\mu\prime}$ satisfy the same algebra, furnishing a different representation.

Let us now compute the Hermitian conjugate of the γ-matrices in the Dirac representation. We have

$$\gamma^{0\dagger} = (I_{2\times 2} \otimes \sigma_3)^\dagger = I_{2\times 2}^\dagger \otimes \sigma_3^\dagger = I_{2\times 2} \otimes \sigma_3 = \gamma^0 = \gamma^0\gamma^0\gamma^0, \tag{7.32}$$

and

$$\gamma^{j\dagger} = \left(\sigma_j \otimes (i\sigma_2)\right)^\dagger = \sigma_j^\dagger \otimes (-i\sigma_2^\dagger) = \sigma_j \otimes (-i\sigma_2) = -\gamma^j = -\gamma^0\gamma^0\gamma^j = \gamma^0\gamma^j\gamma^0. \tag{7.33}$$

In all cases, we have found

$$\gamma^{\mu\dagger} = \gamma^0\gamma^\mu\gamma^0, \tag{7.34}$$

which can also be written as,

$$\begin{aligned}\gamma^0\gamma^\mu\gamma^0 &= \gamma^0\left\{\gamma^\mu,\gamma^0\right\} - \left(\gamma^0\right)^2\gamma^\mu \\ &= 2g^{\mu 0}\gamma^0 - \gamma^\mu.\end{aligned} \tag{7.35}$$

Explicitly, γ-matrices are Hermitian for $\mu = 0$,

$$\gamma^{0\dagger} = \gamma^0, \tag{7.36}$$

while they are anti-Hermitian for $\mu = j = 1,2,3$,

$$\gamma^{j\dagger} = -\gamma^j. \tag{7.37}$$

Eq. (7.34) is of course valid for every other representation of gamma matrices which can be obtained from the Dirac representation via a unitary similarity transformation.

7.3 Traces of γ-matrices

We first observe that the trace of a single gamma matrix is zero:

$$\mathrm{Tr}(\gamma^\mu) = 0. \tag{7.38}$$

This is an immediate consequence of the fact that gamma matrices are constructed out of Pauli matrices in the Dirac representation. But the result is more general. We can prove that γ-matrices are traceless in every representation. Let us start with

$$\mathrm{Tr}\left(\gamma^\mu\gamma^\nu\gamma^\mu\right) = \mathrm{Tr}\left(\gamma^\nu(\gamma^\mu)^2\right) = \mathrm{Tr}\left(\gamma^\nu g^{\mu\mu}\right) = g^{\mu\mu}\mathrm{Tr}\left(\gamma^\nu\right), \tag{7.39}$$

with no summation over the index μ. This is also equal to

$$\mathrm{Tr}\left(\gamma^\mu\gamma^\nu\gamma^\mu\right) = \mathrm{Tr}\left(\gamma^\mu\left\{\gamma^\nu,\gamma^\mu\right\}\right) - \mathrm{Tr}\left((\gamma^\mu)^2\gamma^\nu\right) = 2g^{\mu\nu}\mathrm{Tr}\left(\gamma^\mu\right) - g^{\mu\mu}\mathrm{Tr}\left(\gamma^\nu\right). \tag{7.40}$$

For $\mu \neq \nu$, we then have

$$g^{\mu\mu}\mathrm{Tr}\left(\gamma^\nu\right) = -g^{\mu\mu}\mathrm{Tr}\left(\gamma^\nu\right) \rightsquigarrow \mathrm{Tr}(\gamma^\nu) = 0. \tag{7.41}$$

Exercise: Prove that

1.
$$\mathrm{Tr}(\gamma^\mu\gamma^\nu) = 4g^{\mu\nu}, \tag{7.42}$$

2.
$$\mathrm{Tr}(\gamma^\mu\gamma^\nu\gamma^\rho) = 0. \tag{7.43}$$

3. the trace of a product of an odd number of gamma matrices is zero.

4.
$$\mathrm{Tr}(\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma) = 4\left(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\rho}\right), \tag{7.44}$$

5. Write a computer program to verify that

$$\begin{aligned}
&\frac{1}{4}\mathrm{Tr}(\gamma^{\mu_1} \dots \gamma^{\mu_6}) = \\
&g^{\mu_1\mu_2} g^{\mu_3\mu_4} g^{\mu_5\mu_6} - g^{\mu_1\mu_2} g^{\mu_3\mu_5} g^{\mu_4\mu_6} + g^{\mu_1\mu_2} g^{\mu_3\mu_6} g^{\mu_4\mu_5} \\
&- g^{\mu_1\mu_3} g^{\mu_2\mu_4} g^{\mu_5\mu_6} + g^{\mu_1\mu_3} g^{\mu_2\mu_5} g^{\mu_4\mu_6} - g^{\mu_1\mu_3} g^{\mu_2\mu_6} g^{\mu_4\mu_5} \\
&+ g^{\mu_1\mu_4} g^{\mu_2\mu_3} g^{\mu_5\mu_6} - g^{\mu_1\mu_4} g^{\mu_2\mu_5} g^{\mu_3\mu_6} + g^{\mu_1\mu_4} g^{\mu_2\mu_6} g^{\mu_3\mu_5} \\
&- g^{\mu_1\mu_5} g^{\mu_2\mu_3} g^{\mu_4\mu_6} + g^{\mu_1\mu_5} g^{\mu_2\mu_4} g^{\mu_3\mu_6} - g^{\mu_1\mu_5} g^{\mu_2\mu_6} g^{\mu_3\mu_4} \\
&+ g^{\mu_1\mu_6} g^{\mu_2\mu_3} g^{\mu_4\mu_5} - g^{\mu_1\mu_6} g^{\mu_2\mu_4} g^{\mu_3\mu_5} + g^{\mu_1\mu_6} g^{\mu_2\mu_5} g^{\mu_3\mu_4}.
\end{aligned} \tag{7.45}$$

Why do we care so much about traces of γ-matrices? We will see that they are the cornerstone for computing amplitudes and cross-sections of every process with spin-$\frac{1}{2}$ particles, for example in quantum electrodynamics. Notice that all trace identities are independent of the representation. Under $\gamma^{\mu\prime} = M^{-1}\gamma^{\mu}M$ the cyclic property of the trace gives $\mathrm{Tr}(\gamma^{\mu_1\prime} \dots \gamma^{\mu_n\prime}) = \mathrm{Tr}(M^{-1}\gamma^{\mu_1} \dots \gamma^{\mu_n}M) = \mathrm{Tr}(\gamma^{\mu_1} \dots \gamma^{\mu_n})$. The results also have a simple pattern. The trace of an even number of γ-matrices is a sum over all the ways to pair the indices into metric tensors. Each term carries the sign of the permutation which brings the indices $\mu\rho\nu\sigma, \mu\sigma\nu\rho, \dots$ from their pairing back to the original order. For four matrices, $g^{\mu\nu}g^{\rho\sigma}$ and $g^{\mu\sigma}g^{\nu\rho}$ come with a + sign (even permutations of $\mu\nu\rho\sigma$) and $g^{\mu\rho}g^{\nu\sigma}$ with a − sign (odd permutation). The number of terms grows quickly (15 for six matrices), and in practice such traces are computed with a computer.

7.4 γ-matrices as a basis of 4×4 matrices

The Clifford algebra restricts the number of independent matrices that we can construct by taking products or combinations of products of γ-matrices in a given representation. First, it imposes that the square of a γ-matrix is proportional to the unit 4×4 matrix,

$$(\gamma^{\mu})^2 = \frac{1}{2}\{\gamma^{\mu}, \gamma^{\mu}\} = g^{\mu\mu}I_{4\times4} = \begin{cases} +I_{4\times4}, & \mu = 0 \\ -I_{4\times4}, & \mu = j = 1,2,3 \end{cases} \tag{7.46}$$

This limits the number of terms in independent products of γ-matrices to at most four.

- The only independent product of four γ-matrices is

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3. \tag{7.47}$$

  Exercise: *Show that you can write the matrix γ_5 as*

$$\gamma_5 = -\frac{i}{4!}\epsilon_{\mu\nu\rho\sigma}\gamma^{\mu}\gamma^{\nu}\gamma^{\rho}\gamma^{\sigma}, \qquad (\epsilon^{0123} = -\epsilon_{0123} = 1). \tag{7.48}$$

- There are four independent products of three γ-matrices, $\gamma^0\gamma^1\gamma^2$, $\gamma^0\gamma^1\gamma^3$, $\gamma^0\gamma^2\gamma^3$ and $\gamma^1\gamma^2\gamma^3$. These can be written as a linear combination of

$$\gamma^5\gamma^{\mu}, \tag{7.49}$$

  and single γ-matrices, by using the Clifford algebra.

- Due to the Clifford algebra, only the six antisymmetric combinations of products of two γ-matrices are independent. These are the combinations

$$\sigma^{01},\ \sigma^{02},\ \sigma^{03},\ \sigma^{12},\ \sigma^{13},\ \sigma^{23},$$

  with

$$\sigma^{\mu\nu} = \frac{i}{2}[\gamma^{\mu}, \gamma^{\nu}]. \tag{7.50}$$

In the Dirac representation, γ_5 takes the form

$$\begin{aligned} \gamma_5 &= i\gamma^0\gamma^1\gamma^2\gamma^3 \\ &= i\,(I_{2\times2}\otimes\sigma_3)\,(\sigma_1\otimes(i\sigma_2))\,(\sigma_2\otimes(i\sigma_2))\,(\sigma_3\otimes(i\sigma_2)) \\ &= i\,(\sigma_1\sigma_2\sigma_3)\otimes(-i\sigma_3\sigma_2) \\ &= i\left(i\sigma_3^2\right)\otimes(-\sigma_1) = I_{2\times2}\otimes\sigma_1 \\ \rightsquigarrow \gamma_5 &= \begin{pmatrix} 0 & I_{2\times2} \\ I_{2\times2} & 0 \end{pmatrix}. \end{aligned} \tag{7.51}$$

Exercise: *Show that, in the Dirac representation,*

$$\sigma^{0j} = i\begin{pmatrix} 0_{2\times2} & \sigma_j \\ \sigma_j & 0_{2\times2} \end{pmatrix}, \text{ and } \sigma^{jk} = \epsilon_{jkl}\begin{pmatrix} \sigma_l & 0_{2\times2} \\ 0_{2\times2} & \sigma_l \end{pmatrix}, \tag{7.52}$$

with $j, k = 1, 2, 3$.

A general 4×4 matrix has 16 independent components. From a representation of γ-matrices we can construct 15 independent 4×4 matrices. Together with the unit matrix $I_{4\times4}$, they constitute a basis. In other words, a general 4×4 matrix M can be written as

$$M_{4\times4} = a_0 I_{4\times4} + a_\mu\gamma^\mu + a_5\gamma_5 + a_{5\mu}\gamma_5\gamma^\mu + a_{\mu\nu}\sigma^{\mu\nu}. \tag{7.53}$$

Exercise: *Find the coefficients* $a_0, a_\mu, a_5, a_{5\mu}, a_{\mu\nu}$ *for a given* 4×4 *matrix* M.

The counting of the basis elements is not accidental. An independent product of γ-matrices contains each of the four matrices $\gamma^0, \gamma^1, \gamma^2, \gamma^3$ at most once. The number of products with k different matrices is the binomial coefficient $\binom{4}{k}$, which gives $1, 4, 6, 4, 1$ elements for $k = 0, 1, 2, 3, 4$, and in total $2^4 = 16$. For finding the coefficients in the exercise above, think of the trace as a scalar product in the space of 4×4 matrices: multiply Eq. (7.53) by one of the basis elements and take the trace, in the same way as we did for the 2×2 matrices with the Pauli matrices.

7.5 Lagrangian for the Dirac field

After this survey of the properties of γ-matrices, we return to the Dirac equation,

$$(i\partial\!\!\!/ - mI_{4\times4})\,\psi = 0. \tag{7.54}$$

Here we have introduced the "slash" notation,

$$a\!\!\!/ \equiv \gamma^\mu a_\mu, \qquad \partial\!\!\!/ \equiv \gamma^\mu\partial_\mu. \tag{7.55}$$

The Dirac differential operator $i\partial\!\!\!/ - m$ is a 4×4 matrix. The object ψ is a four-component "Dirac spinor" and it has its own Lorentz transformation, which we shall derive in detail later. In components,

$$\psi(\vec{x}, t) \equiv \psi_a(\vec{x}, t) = \begin{pmatrix} \psi_1(\vec{x}, t) \\ \psi_2(\vec{x}, t) \\ \psi_3(\vec{x}, t) \\ \psi_4(\vec{x}, t) \end{pmatrix}. \tag{7.56}$$

It is very important not to confuse a spinor ψ with a four-vector, which transforms in a truly different way under Lorentz transformations.

The price for a first-order equation is that one field ϕ has been replaced by four complex fields $\psi_1, \ldots, \psi_4$. What is their meaning? We will see after quantisation that two of them

describe a spin-$\frac{1}{2}$ particle, with its spin up or down, and the other two its anti-particle, again with two spin states. Dirac's original goal, getting rid of the negative-energy solutions, was not achieved; they are still there, as for the Klein-Gordon equation. Instead, the Dirac equation brought in spin and it predicted anti-matter. Dirac first suggested that the partner of the electron with the opposite charge could be the proton, the only positively charged particle known at the time. However, the theory requires the two particles to have the same mass, and the positron was discovered by Anderson in 1932.

We can now take the Hermitian conjugate of the Dirac equation,

$$\begin{aligned} & (i\partial\!\!\!/ - mI_{4\times4})\,\psi = 0 \\ \rightsquigarrow\quad & -i\left(\gamma^\mu\partial_\mu\psi\right)^\dagger - m\psi^\dagger = 0 \\ \rightsquigarrow\quad & -i\left(\partial_\mu\psi^\dagger\right)\gamma^{\mu\dagger} - m\psi^\dagger = 0 \\ \rightsquigarrow\quad & -i\left(\partial_\mu\psi^\dagger\right)\gamma^0\gamma^\mu\gamma^0 - m\psi^\dagger = 0. \end{aligned} \tag{7.57}$$

Multiplying this equation by γ^0 from the right, we obtain

$$\begin{aligned} & -i\left(\partial_\mu\psi^\dagger\right)\gamma^0\gamma^\mu(\gamma^0)^2 - m\psi^\dagger\gamma^0 = 0 \\ \rightsquigarrow\quad & -i\left(\partial_\mu(\psi^\dagger\gamma^0)\right)\gamma^\mu - m(\psi^\dagger\gamma^0) = 0. \end{aligned} \tag{7.58}$$

We now define

$$\bar\psi \equiv \psi^\dagger\gamma^0. \tag{7.59}$$

Then, the Hermitian conjugate of the Dirac equation reads

$$-i\partial_\mu\bar\psi\gamma^\mu - m\bar\psi = 0. \tag{7.60}$$

We can obtain the Dirac equation and its conjugate from a Lagrangian density,

$$\mathcal{L} = \bar\psi\,(i\partial\!\!\!/ - m)\,\psi. \tag{7.61}$$

Exercise: *Derive the Euler-Lagrange equations for the spinor field components ψ_a and $\bar\psi_a$.*

A hint: the four components ψ_a are complex, so we have eight real degrees of freedom. Instead of their real and imaginary parts, we can use the four ψ_a and the four $\bar\psi_a$ as independent variables (or, equivalently, ψ_a and ψ_a^*). In the form of Eq. (7.61) the derivative acts only on ψ and no derivative of $\bar\psi$ appears. The Euler-Lagrange equations for $\bar\psi_a$ are then simply $\partial\mathcal{L}/\partial\bar\psi_a = 0$, which is the Dirac equation. For the Euler-Lagrange equations of ψ_a you need to differentiate with respect to $\partial_\mu\psi_a$, and you will obtain the conjugate Dirac equation.

The Lagrangian is a very important piece of information if we want to quantise the Dirac field ψ. We will postpone the quantisation of the Dirac field until after we gain a deeper understanding of the origin of the Lagrangian, and the Lorentz transformation properties of spinor fields.

8

Lorentz symmetry and free fields

A fundamental requirement for physical laws is that they must be valid for all relativistic observers. Assume that the fields $\phi_i(x^\mu)$ describe a physical system in a certain coordinate reference frame. Consider now a different relativistic observer. Space-time coordinates in the frame of the two observers are related by a Lorentz transformation,

$$x^\mu \to x^{\mu\prime} = \Lambda^\mu_\nu x^\nu. \tag{8.1}$$

The new observer will have to transform appropriately both fields and coordinates when computing the Lagrangian,

$$\phi_i(x) \to \phi'_i(x'), \qquad \mathcal{L}(x) \to \mathcal{L}'(x'). \tag{8.2}$$

We will require that the action is the same for both observers. In this way, if it is an extremum for the physical values of $\phi_i(x)$ in the first reference frame, it will also be an extremum for the physical values of the field $\phi'_i(x')$ in the second relativistic frame. For proper Lorentz transformations ($\det\left(\Lambda^\mu_\nu\right) = 1$) we have

$$d^4x' = d^4x \det\left(\frac{\partial x^{\mu\prime}}{\partial x^\nu}\right) = d^4x \det\left(\Lambda^\mu_\nu\right) = d^4x. \tag{8.3}$$

Then

$$S = \int d^4x \mathcal{L}(x) = \int d^4x \mathcal{L}'(x') = S'. \tag{8.4}$$

Therefore, the action is relativistically invariant if the Lagrangian density is a scalar, i.e. it does not transform under a Lorentz transformation:

$$\mathcal{L}'(x') = \mathcal{L}(x), \qquad x^\mu \to x^{\mu\prime} = \Lambda^\mu_\nu x^\nu. \tag{8.5}$$

How many such Lagrangian densities can we find? Obviously, Lorentz symmetry is not sufficient on its own to determine all physical laws. We can restrict the problem to Lagrangians which describe free relativistic particles, which do not interact with each other. The Lagrangians that we are after may contain only one parameter, which is the mass m of the free particles. In addition, the equations of motion and field quantisation should lead to the correct relativistic energy-momentum relation,

$$p^2 = m^2. \tag{8.6}$$

A Lagrangian of free particles must contain derivatives of the fields $\partial_\mu\phi_i$. They are needed to generate the momentum $\partial_\mu\phi_i \to p_\mu$ in the energy-momentum relation. We include a generic label i, allowing for the possibility of a particle with a mass m and multiple degrees of freedom, such as charge, spin, etc. Our question has now become more specific: *How can we combine the fields ϕ_i and their derivatives $\partial_\mu\phi_i$ into a scalar Lagrangian with only one mass parameter?* We need two ingredients:

- The transformation of derivatives $\partial_\mu \to \partial'_\mu$ under $x^\mu \to \Lambda^\mu_\nu x^\nu$,
- The fields transformations $\phi_i(x) \to \phi'_i(x')$ under $x^\mu \to \Lambda^\mu_\nu x^\nu$.

It is straightforward to find the transformation of derivatives,

$$\begin{aligned} \partial_\mu \to \partial'_\mu &= \frac{\partial}{\partial x^{\mu\prime}} = \frac{\partial x^\nu}{\partial x^{\mu\prime}} \frac{\partial}{\partial x^\nu} \\ \rightsquigarrow \partial_\mu \to \partial'_\mu &= \left(\Lambda^{-1}\right)^\nu_{\ \mu} \partial_\nu . \end{aligned} \tag{8.7}$$

Similarly,

$$\partial^\mu \to \partial^{\mu\prime} = \Lambda^\mu_\nu \partial^\nu . \tag{8.8}$$

8.1 Field transformations and representations of the Lorentz group

In this section, we will examine the allowed possibilities for linear field transformations,

$$\phi_i(x) \to \phi'_i(x') = M(\Lambda)_{ij} \phi_j(x), \tag{8.9}$$

under a space-time Lorentz transformation

$$x^\mu \to \Lambda^\mu_\nu x^\nu . \tag{8.10}$$

Remark. Why only linear transformations? Non-linearly realised symmetries do exist, and they cannot in general be made linear without adding fields: the pions of a spontaneously broken chiral symmetry transform non-linearly (Callan, Coleman, Wess, Zumino 1969; Weinberg, QFT II, Ch. 19), and the BRST transformation of Part III is non-linear in the fields. For the Lorentz group acting on fields, however, linear transformations are all we need (see Weinberg, Vol. I, Chapter 5).

Lorentz transformations form a group, i.e. if we perform two successive Lorentz transformations,

$$x \to x' = \Lambda x,$$

and

$$x' \to x'' = \Lambda' x',$$

this is equivalent to performing a direct Lorentz transformation

$$x \to x'' = \Lambda'' x,$$

with

$$\Lambda'' = \Lambda' \Lambda . \tag{8.11}$$

How about the matrices $M(\Lambda)_{ij}$? Consider again the transformation $x \to x'' = \Lambda'' x$, where the field is transformed as

$$\phi(x) \to \phi''(x'') = M(\Lambda'')\phi(x). \tag{8.12}$$

We can arrive at the frame x'' by performing successive Lorentz transformations, $x \to x' = \Lambda x \to x'' = \Lambda' x' = \Lambda' \Lambda x$. Then we have that,

$$\phi''(x'') = M(\Lambda')\phi'(x') = M(\Lambda')M(\Lambda)\phi(x). \tag{8.13}$$

For the last two equations to be consistent with each other, the field transformations $M(\Lambda)$ must obey:

$$M(\Lambda'\Lambda) = M(\Lambda')M(\Lambda). \tag{8.14}$$

In group theory terminology, this means that the matrices $M(\Lambda)$ furnish a representation of the Lorentz group. Field Lorentz transformations are therefore not random, but they can be found if we find all (finite-dimensional) representations of the Lorentz group.

8.1.1 Scalar representation $M(\Lambda) = 1$

The simplest representation $M(\Lambda)$ is the scalar representation, where a field $\phi(x)$ does not change under Lorentz transformations,

$$\phi(x) \to \phi'(x') = \phi(x). \tag{8.15}$$

In a Lagrangian of a scalar field and its derivatives, the transformations of the derivatives must cancel among themselves. A single derivative term,

$$\partial_\mu \phi,$$

is not a scalar. The minimum number of derivatives that are needed for a scalar is two,

$$\left(\partial_\mu \phi\right)\left(\partial^\mu \phi\right), \qquad \phi \partial^2 \phi.$$

The second term is actually equivalent to the first up to a total divergence. Let us assume that a Lagrangian contains this term,

$$\mathcal{L} = \left(\partial_\mu \phi\right)\left(\partial^\mu \phi\right) + \dots \tag{8.16}$$

Then we can find the mass dimension of the field ϕ, by requiring that the action is dimensionless (in natural units, $c = \hbar = 1$),

$$[\text{mass}]^0 = [S] = \left[\int d^4x \mathcal{L}\right] = [x]^4 [\mathcal{L}] = [\text{mass}]^{-4} [\mathcal{L}], \tag{8.17}$$

which means that all Lagrangian densities must have mass dimension 4,

$$[\mathcal{L}] = [\text{mass}]^4. \tag{8.18}$$

From this we find that a scalar field ϕ has mass dimension one,

$$\begin{aligned} & \left[\left(\partial_\mu \phi\right)\left(\partial^\mu \phi\right)\right] = [\text{mass}]^4 \\ \rightsquigarrow \quad & \left[\partial_\mu\right]^2 \left[\phi\right]^2 = [\text{mass}]^4 \\ \rightsquigarrow \quad & [\text{mass}]^2 \left[\phi\right]^2 = [\text{mass}]^4 \\ \rightsquigarrow \quad & \left[\phi\right] = [\text{mass}]^1. \end{aligned} \tag{8.19}$$

A relativistically invariant Lagrangian may include terms with no derivatives, of the type

$$\phi^n.$$

Notice that the Klein-Gordon Lagrangian

$$\mathcal{L}_{\text{Klein–Gordon}} = \frac{1}{2}\left(\partial_\mu \phi\right)\left(\partial^\mu \phi\right) - \frac{m^2}{2}\phi^2, \tag{8.20}$$

is, indeed, invariant under Lorentz transformations. In addition, the Klein-Gordon Lagrangian has the reflection symmetry $\phi \leftrightarrow -\phi$.

8.1.2 Vector representation $M(\Lambda) = \Lambda$

Another representation of the Lorentz group which is easy to figure out is the vector representation, corresponding to

$$M(\Lambda^\mu_{\ \nu}) = \Lambda^\mu_{\ \nu}. \tag{8.21}$$

A field in the vector representation transforms as,

$$A^\mu(x) \to A^{\mu\prime}(x') = \Lambda^\mu_{\ \nu} A^\nu(x), \qquad x^{\mu\prime} = \Lambda^\mu_{\ \nu} x^\nu. \tag{8.22}$$

Terms in which all Lorentz indices are contracted are invariant under Lorentz transformations.

Exercise: *Show that the following terms are invariant under Lorentz transformations:*

$$\left(\partial_\mu A^\mu\right), \quad \left(\partial_\mu A_\nu\right)\left(\partial^\mu A^\nu\right), \quad \left(\partial_\mu A_\nu\right)\left(\partial^\nu A^\mu\right). \tag{8.23}$$

Lorentz invariance is not the only symmetry that can be required for free fields. For example, the Lagrangian for the electromagnetic field is

$$\mathcal{L}_{\text{EM}} = -\frac{1}{2}\left[\left(\partial_\mu A_\nu\right)\left(\partial^\mu A^\nu\right) - \left(\partial_\mu A_\nu\right)\left(\partial^\nu A^\mu\right)\right], \tag{8.24}$$

where only a certain combination of two dimension-four operators with two derivatives is present in the Lagrangian. This is because of the additional symmetry under gauge transformations, which further constrains the form of the Lagrangian.

8.2 Generators of field representations of the Lorentz group

Are the scalar and the vector representations the only representations that we can find? How do we find the rest of them? This question simplifies if we look at small Lorentz transformations and their properties.

A small Lorentz transformation is,

$$\Lambda^\mu_{\ \nu} = \delta^\mu_{\ \nu} + \omega^\mu_{\ \nu}, \tag{8.25}$$

where the parameters of the transformation satisfy[1]

$$\omega_{\mu\nu} = -\omega_{\nu\mu}. \tag{8.26}$$

The anti-symmetry is manifest only when both indices are down (or both up). Raising one index with the metric can produce a sign. For a small boost along the x-axis, $\omega_{01} = -\omega_{10}$, we find $\omega^0_{\ 1} = g^{00}\omega_{01} = \omega_{01}$ and $\omega^1_{\ 0} = g^{11}\omega_{10} = \omega_{01}$. The matrix $\omega^\mu_{\ \nu}$ of a boost is therefore symmetric, while for a rotation, e.g. $\omega_{12} = -\omega_{21}$, it is anti-symmetric, $\omega^1_{\ 2} = -\omega_{12} = -\omega^2_{\ 1}$. This is the reason why we prefer to write the parameters with lower indices, $\omega_{\mu\nu}$, in the expansions which follow. Any $n \times n$ matrix representation of this transformation on fields,

$$M_{ab}\left(\Lambda^\mu_{\ \nu}\right) = M_{ab}\left(\delta^\mu_{\ \nu} + \omega^\mu_{\ \nu}\right), \qquad a, b = 1, \ldots, n, \tag{8.27}$$

can be expanded in the small parameters $\omega^\mu_{\ \nu}$,

$$M_{ab}\left(\Lambda^\mu_{\ \nu}\right) = M_{ab}\left(\delta^\mu_{\ \nu}\right) + \sum_{\mu<\nu} \omega_{\mu\nu} \left.\frac{\partial M_{ab}(\Lambda)}{\partial \omega_{\mu\nu}}\right|_{\omega=0} + O(\omega^2). \tag{8.28}$$

[1] Notice that generally $\omega^\mu_{\ \nu} \neq -\omega^\nu_{\ \mu}$. Rather, Eq. (8.26) implies that $\omega^\mu_{\ \nu} = -\omega_\nu^{\ \mu}$.

In the above, we restrict the summation to $\mu < \nu$ so that we do not include twice the contribution of the same parameter $\omega_{\mu\nu}$ via its antisymmetric form $\omega_{\nu\mu}$. The n-dimensional representation $M_{ab}\left(\delta^{\mu}_{\nu}\right)$ of the identity transformation is the $n \times n$ unit matrix,

$$M_{ab}\left(\delta^{\mu}_{\nu}\right) = I_{n\times n}. \tag{8.29}$$

The derivatives of the representation matrices M with respect to the small transformation parameters define the so-called generators of the representation,

$$\left(J^{\mu\nu}_{M}\right)_{ab} \equiv i \left.\frac{\partial M_{ab}(\Lambda^{\mu}_{\nu})}{\partial \omega_{\mu\nu}}\right|_{\omega=0}. \tag{8.30}$$

The generators form a basis, such that the representation of every small Lorentz transformation can be written as a *linear* combination of the generators plus the unit,

$$M_{ab}\left(\Lambda\right) = I_{n\times n} - i \sum_{\mu<\nu} \omega_{\mu\nu} J^{\mu\nu}_{M}. \tag{8.31}$$

Finite Lorentz transformations are also generated by exponentiating linear combinations of the generators. If Λ^{μ}_{ν} represents large boosts and/or rotations, it cannot be expanded directly in small parameters. Then, we can use the representation property,

$$M(\Lambda) = M(\Lambda_1)\ldots M(\Lambda_N), \qquad \text{with } \Lambda = \Lambda_1\Lambda_2\ldots\Lambda_N, \tag{8.32}$$

and construct these large Lorentz transformations exactly from an infinite product of infinitesimal Lorentz transformations with small parameters $\omega_{\mu\nu}$. In the end, every Lorentz transformation can be written in the exponential form,

$$M(\Lambda) = e^{-i\sum_{\mu<\nu}\omega_{\mu\nu}J^{\mu\nu}_{M}}. \tag{8.33}$$

The generators of a group representation are all we ever need in order to determine the representation completely.

8.2.1 Generators of the scalar representation

Let us now study the generators of the scalar representation of the Lorentz symmetry transformation. In this representation, under a Lorentz transformation

$$x^{\mu} \longrightarrow x^{\mu\prime} = \Lambda^{\mu}_{\nu} x^{\nu}, \tag{8.34}$$

a scalar field remains invariant,

$$\phi'(x^{\mu\prime}) = \phi(x^{\mu}) \rightsquigarrow \phi'(\Lambda^{\mu}_{\nu}x^{\nu}) = \phi(x^{\mu}). \tag{8.35}$$

The complete transformation of a scalar field is rather trivial. Recall that a small symmetry transformation of the field can be decomposed into two parts, the "variation at a point" and the variation from one point to another one (of a very close proximity in the linear approximation),

$$\Delta\phi(x) \equiv \phi'(x^{\mu\prime}) - \phi(x^{\mu}) = \delta_*\phi(x) + \delta\phi(x) = 0, \tag{8.36}$$

where

$$\delta\phi(x) = \phi(\Lambda x) - \phi(x), \tag{8.37}$$

and

$$\delta_*\phi(x) = \phi'(x) - \phi(x). \tag{8.38}$$

We can compute the variation at a point, $\delta_*\phi$, as follows. Setting $x \to \Lambda^{-1}x$ in Eq. (8.35), we have

$$\phi'(x^\mu) = \phi(\Lambda^{-1\mu}{}_\nu x^\nu), \tag{8.39}$$

which for small Lorentz transformations becomes

$$\begin{aligned}
\phi'(x^\mu) &= \phi(x^\mu - \omega^\mu{}_\nu x^\nu) \\
&= \phi(x^\mu) - \omega_{\mu\nu} x^\nu \partial^\mu \phi(x^\mu) + O(\omega^2) \\
&= \phi(x^\mu) - \frac{1}{2}\left(\omega_{\mu\nu} - \omega_{\nu\mu}\right) x^\nu \partial^\mu \phi(x^\mu) + O(\omega^2) \\
&= \phi(x^\mu) + \omega_{\mu\nu}\frac{1}{2}\left(x^\mu \partial^\nu - x^\nu \partial^\mu\right)\phi(x^\mu) + O(\omega^2) \\
\rightsquigarrow \phi'(x) &= \left(1 - \frac{i}{2}\omega_{\mu\nu} J^{\mu\nu}_{\text{scalar}}\right)\phi(x),
\end{aligned} \tag{8.40}$$

with the generators of the Lorentz group in the scalar representation being

$$J^{\mu\nu}_{\text{scalar}} = i\left(x^\mu \partial^\nu - x^\nu \partial^\mu\right). \tag{8.41}$$

Notice which transformation these generators describe. The complete transformation of a scalar field, $\phi'(x') = \phi(x)$, is trivial; its representation matrix is $M(\Lambda) = 1$ and its generators vanish. The two pieces of it, the variation at a point $\delta_*\phi$ and the variation from one point to another, cancel each other, but each of them is non-trivial. The generators $J^{\mu\nu}_{\text{scalar}}$ are differential operators which describe the variation at a point. This is the piece that enters Noether's theorem, and it gives rise to the conserved angular momentum (and the conserved charges of boosts) of the field. In the quantum theory it will be the orbital angular momentum.

An explicit calculation (**exercise**) yields the identity,

$$\begin{aligned}
&\left[J^{\mu_1\mu_2}_{\text{scalar}}, J^{\mu_3\mu_4}_{\text{scalar}}\right] = \\
&i\left(g^{\mu_2\mu_3} J^{\mu_1\mu_4}_{\text{scalar}} + g^{\mu_1\mu_4} J^{\mu_2\mu_3}_{\text{scalar}} - g^{\mu_2\mu_4} J^{\mu_1\mu_3}_{\text{scalar}} - g^{\mu_1\mu_3} J^{\mu_2\mu_4}_{\text{scalar}}\right).
\end{aligned} \tag{8.42}$$

The above commutation relations of the generators form what is known as the "Lie algebra" of the Lorentz group. Here, we derived it from the scalar field representation of the symmetry transformation. However, as we shall see explicitly, it is satisfied by the generators of every other representation of the symmetry group.

Finally, let us remark that, by multiplying an infinite number of small transformations, Eq. (8.40) turns, thanks to Eq. (8.42), into an exponentiated form. We have,

$$\phi'(x) = e^{-\frac{i}{2}\omega_{\mu\nu} J^{\mu\nu}_{\text{scalar}}}\phi(x). \tag{8.43}$$

8.2.2 Generators of the vector representation

We can see that the Lie algebra is common to all representations by repeating the same analysis for the generators of the vector representation. A vector field $A^\alpha(x^\mu)$ transforms as

$$\begin{aligned}
A^\alpha(x^\rho) &\to A^{\alpha\prime}(x^{\rho\prime}) = \Lambda^\alpha_\beta A^\beta(x^\rho) \\
&= \left(\delta^\alpha_\beta + \omega^\alpha_\beta\right) A^\beta\left(x^\rho\right) + O(\omega^2) \\
&= A^\alpha\left(x^\rho\right) + \omega^\alpha_\beta A^\beta\left(x^\rho\right) + O(\omega^2) \\
\rightsquigarrow A^{\alpha\prime}(x^{\rho\prime}) &= \left(\delta^\alpha_\beta - \frac{i}{2}\omega_{\mu\nu}\left(J^{\mu\nu}_{\text{vector}}\right)^\alpha{}_\beta\right) A^\beta(x^\rho),
\end{aligned} \tag{8.44}$$

with

$$\left(J^{\mu\nu}_{\text{vector}}\right)_{\alpha\beta} = i\left(\delta^{\mu}_{\alpha}\delta^{\nu}_{\beta} - \delta^{\nu}_{\alpha}\delta^{\mu}_{\beta}\right). \tag{8.45}$$

We can now compute explicitly (**exercise**) the commutators of the generators of the vector field representation,

$$\left[J^{\mu_1\mu_2}_{\text{vector}}, J^{\mu_3\mu_4}_{\text{vector}}\right]_{\alpha\beta} = \\ i\left[g^{\mu_2\mu_3}\left(J^{\mu_1\mu_4}_{\text{vector}}\right)_{\alpha\beta} + g^{\mu_1\mu_4}\left(J^{\mu_2\mu_3}_{\text{vector}}\right)_{\alpha\beta} - g^{\mu_2\mu_4}\left(J^{\mu_1\mu_3}_{\text{vector}}\right)_{\alpha\beta} - g^{\mu_1\mu_3}\left(J^{\mu_2\mu_4}_{\text{vector}}\right)_{\alpha\beta}\right].$$

We see indeed that the commutators of the vector representation satisfy the same Lie algebra as in the scalar representation (Eq. (8.42)). This is of course what we expect from the theory of Lie groups, which we will review in the following subsection.

A Lorentz transformation on a vector field changes its orientation by mixing the indices of A^{α}. The generators of the transformation for the vector orientation are given in Eq. (8.45). In exponentiated form, which can also be applied to large transformations, we have

$$A^{\alpha\prime}(x') = e^{-\frac{i}{2}\omega_{\mu\nu}\left(J^{\mu\nu}_{\text{vector}}\right)^{\alpha}_{\beta}} A^{\beta}(x). \tag{8.46}$$

Exercise: *Prove that a Lorentz transformation of a vector field "at a point" is*

$$A^{\alpha\prime}(x) = e^{-\frac{i}{2}\omega_{\mu\nu}\left[\left(J^{\mu\nu}_{\text{vector}}\right)^{\alpha}_{\beta} + \delta^{\alpha}_{\beta} J^{\mu\nu}_{\text{scalar}}\right]} A^{\beta}(x). \tag{8.47}$$

8.2.3 Lie algebra of continuous groups

Our discussion of the Lorentz symmetry is more general and it applies to any continuous symmetry transformation

$$T(\theta^a),$$

parametrised by a set of continuous parameters

$$\{\theta^a\}, \quad a = 1, \ldots, N.$$

The symmetry transformations form a Lie group and the product of two such transformations is also a symmetry transformation belonging to the group:

$$T(\theta^a_1)T(\theta^a_2) = T\left(f^a(\theta^a_1, \theta^a_2)\right). \tag{8.48}$$

We arrange that the unit element of the symmetry group corresponds to all parameters being zero, $\theta^a = 0$,

$$T(\theta^a = 0) = 1. \tag{8.49}$$

The product rule of Eq. (8.48) and our definition of the unit transformation of Eq. (8.49) yield:

$$f^a(0, \theta^a) = f^a(\theta^a, 0) = \theta^a. \tag{8.50}$$

Restricting ourselves to small transformations around the identity transformation, we perform a Taylor expansion around $\theta^a = 0$:

$$f^a(\theta^a_1, \theta^a_2) = f^a(0,0) + \frac{\partial f^a}{\partial \theta^b_1}\theta^b_1 + \frac{\partial f^a}{\partial \theta^b_2}\theta^b_2 \\ + \frac{1}{2}\frac{\partial^2 f^a}{\partial \theta^b_1 \partial \theta^c_1}\theta^b_1\theta^c_1 + \frac{1}{2}\frac{\partial^2 f^a}{\partial \theta^b_2 \partial \theta^c_2}\theta^b_2\theta^c_2 + \frac{\partial^2 f^a}{\partial \theta^b_1 \partial \theta^c_2}\theta^b_1\theta^c_2 + \ldots \tag{8.51}$$

Imposing the requirement of Eq. (8.50) leads to the form:

$$f^a(\theta_1^a, \theta_2^a) = \theta_1^a + \theta_2^a + f^a_{bc}\theta_1^b\theta_2^c + \dots \tag{8.52}$$

An example helps to see what the function f^a is. If we rotate by 5 degrees around the $z-$axis and then by 6 degrees around the same axis, the combined transformation is a rotation by 11 degrees: for rotations around a fixed axis, f^a is simply additive. Two rotations around different axes, however, combine into a rotation around a third axis, with an angle which is not the sum of the two. Eq. (8.52) tells us that, for small parameters, every continuous symmetry transformation is additive at the linear order. The specific features of the group enter first at the quadratic order, through the constants f^a_{bc}.

For an ordinary representation $U(T)$ of the symmetry group, we have

$$U(T(\theta_1^a))U(T(\theta_2^a)) = U\left(T(f^a(\theta_1^a, \theta_2^a))\right). \tag{8.53}$$

For small parameters θ^a, we can expand the representation of the transformation as follows:

$$U(T(\theta^a)) = 1 + i\theta^a t_a + \frac{1}{2}\theta^b\theta^c t_{bc} + \dots, \tag{8.54}$$

where $1, t_a, t_{bc}$ in the above equation are matrices of the same dimensionality as $U(T(\theta^a))$ and the matrix t_{bc} is symmetric in b and c:

$$t_{bc} = t_{cb}. \tag{8.55}$$

Expanding the terms in Eq. (8.53) with Eq. (8.54), we have

$$\begin{aligned}&\left(1 + i\theta_1^a t_a + \frac{1}{2}\theta_1^b\theta_1^c t_{bc} + \dots\right)\left(1 + i\theta_2^a t_a + \frac{1}{2}\theta_2^b\theta_2^c t_{bc} + \dots\right)\\ =&1 + i\left(\theta_1^a + \theta_2^a + f^a_{bc}\theta_1^b\theta_2^c + \dots\right)t_a + \frac{1}{2}\left(\theta_1^b + \theta_2^b + \dots\right)\left(\theta_1^c + \theta_2^c + \dots\right)t_{bc}.\end{aligned}$$

Matching the powers of θ_1^a, θ_2^a, we find that

$$t_{bc} = -t_b t_c - i f^a_{bc} t_a. \tag{8.56}$$

From the symmetry of t_{bc}, we obtain:

$$0 = t_{bc} - t_{cb} = -(t_b t_c - t_c t_b) - i(f^a_{bc} - f^a_{cb})t_a. \tag{8.57}$$

This leads to the commutation relation for the matrices t^a:

$$[t_b, t_c] = iC^a_{bc} t_a, \tag{8.58}$$

where the constants

$$C^a_{bc} = f^a_{cb} - f^a_{bc} \tag{8.59}$$

are antisymmetric under $b \leftrightarrow c$ and are called the *structure constants* of the Lie group. As you may observe, the structure constants do not depend on the specific representation $U(T(\theta^a))$. Instead, the matrices t_a are specific to the representation and have its dimensionality. These matrices are known as the *generators* of the representation. Knowing the generators, we can construct the representation of an arbitrary symmetry transformation by using the product property of Eq. (8.53) to assemble large symmetry transformations from many small ones where the expansion of Eq. (8.54) is valid. The common commutation relation

of Eq. (8.58) satisfied by the generators of all representations of the symmetry group is known as a *Lie algebra*.

Requiring that the representation of a symmetry transformation is unitary, leads to the conclusion that the generators of the transformation are Hermitian operators:

$$UU^\dagger = 1 \rightsquigarrow t_a^\dagger = t_a. \tag{8.60}$$

Therefore, generators are good candidates for physical observables.

Remark. For the Lorentz group, this statement must be applied with care. The Lorentz group is non-compact, and its non-trivial finite-dimensional representations (vector, spinor, ...) are not unitary: the boost generators are then not Hermitian, e.g. $S^{0j\dagger} = -S^{0j}$ for $S^{\mu\nu} = \frac{i}{4}[\gamma^\mu, \gamma^\nu]$. The unitary representations acting on the Hilbert space of particle states are infinite-dimensional (see Weinberg, Vol. I, Sects. 2.4–2.5).

In the case of the Lorentz group, we have:

$$\left[J_M^{\mu\nu}, J_M^{\rho\sigma}\right] = iC^{\alpha\beta}_{\mu\nu\rho\sigma} J_M^{\alpha\beta}. \tag{8.61}$$

The coefficients $C^{\alpha\beta}_{\mu\nu\rho\sigma}$ are the structure constants of the Lorentz group.

We now have a better strategy in order to find representations of the Lorentz group:

- First, we will use any of the representations that we already know, the scalar or vector representation, in order to find the structure constants $C^{\alpha\beta}_{\mu\nu\rho\sigma}$ of the Lorentz group.
- Then we shall attempt to find matrices $J^{\mu\nu}_{n\times n}$ for each finite dimension n which satisfy the Lie algebra of Eq. (8.61). If we succeed, then we have found an $n\times n$ representation of the generators, and by exponentiation, an $n \times n$ representation of the group.

8.3 Spinor representation

We can find all possible generators for the finite-dimensional representations of the Lorentz group in a systematic manner. Of particular interest to us is the Dirac or spinor representation of the Lorentz group, which has the same dimensionality 4×4 as the vector representation. We can verify, after some algebra, that the 4×4 matrices,

$$S^{\mu\nu} = \frac{i}{4}\left[\gamma^\mu, \gamma^\nu\right], \tag{8.62}$$

satisfy the same commutation relation as in Eq. (8.42). Explicitly,

$$\begin{aligned}&[S^{\mu_1\mu_2}, S^{\mu_3\mu_4}] = \\ &\ i\left(g^{\mu_2\mu_3}S^{\mu_1\mu_4} + g^{\mu_1\mu_4}S^{\mu_2\mu_3} - g^{\mu_2\mu_4}S^{\mu_1\mu_3} - g^{\mu_1\mu_3}S^{\mu_2\mu_4}\right).\end{aligned} \tag{8.63}$$

Therefore, they are generators of a new representation for the Lorentz group, which is neither the scalar nor the vector representation. We call this new representation the "spinor" representation. A spinor field

$$\psi(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \\ \psi_3(x) \\ \psi_4(x) \end{pmatrix}, \tag{8.64}$$

transforms as

$$\psi'(x') = \Lambda_{\frac{1}{2}}\psi(x), \tag{8.65}$$

with a finite spinor transformation being

$$\Lambda_{\frac{1}{2}} = e^{-\frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}}. \tag{8.66}$$

We find (**exercise**) that the commutator of the generator $S^{\mu\nu}$ with a γ-matrix is

$$[\gamma^\mu, S^{\rho\sigma}] = i\left(g^{\mu\rho}\gamma^\sigma - g^{\mu\sigma}\gamma^\rho\right), \tag{8.67}$$

or, in the more suggestive form,

$$[\gamma^\mu, S^{\rho\sigma}] = \left(J^{\rho\sigma}_{\text{vector}}\right)^\mu_{\ \nu}\gamma^\nu. \tag{8.68}$$

The above equation can be exponentiated into a general result,

$$\Lambda^{-1}_{\frac{1}{2}}\gamma^\mu\Lambda_{\frac{1}{2}} = \Lambda^\mu_{\ \nu}\gamma^\nu. \tag{8.69}$$

Exercise: *Prove Eq. (8.69) in the case of very small Lorentz transformations, by expanding both sides of the equation in the transformation parameters $\omega_{\mu\nu}$ and using Eq. (8.68).*

In the same manner, we verify that

$$\Lambda_{\frac{1}{2}}\gamma^\mu\Lambda^{-1}_{\frac{1}{2}} = (\Lambda^{-1})^\mu_{\ \nu}\gamma^\nu. \tag{8.70}$$

8.4 Lorentz invariance of the Dirac Lagrangian

It is easy to verify that both terms, $\bar\psi\psi$ and $\bar\psi\partial\!\!\!/\psi$, in the Dirac Lagrangian are invariant under Lorentz transformations. As in many previous instances, we shall use the mathematical properties of Lie groups, which allow us to find transformation rules for representations for small Lorentz transformations, and simply exponentiate the results at the end in order to write down the general transformations.

Starting from the transformation of a spinor field, we have:

$$\begin{aligned}
\psi &\to \Lambda_{\frac{1}{2}}\psi \\
\rightsquigarrow \psi^\dagger &\to \psi^\dagger\Lambda^\dagger_{\frac{1}{2}} \\
&\simeq \psi^\dagger\left(1 - \frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}\right)^\dagger \\
&\simeq \psi^\dagger\left(1 + \frac{i}{2}\omega_{\mu\nu}S^{\dagger\mu\nu}\right).
\end{aligned} \tag{8.71}$$

We also have,

$$\begin{aligned}
S^{\dagger\mu\nu} = \left(\frac{i}{4}[\gamma^\mu, \gamma^\nu]\right)^\dagger &= -\frac{i}{4}\left[\gamma^{\nu\dagger}, \gamma^{\mu\dagger}\right] = \frac{i}{4}\left[\gamma^0\gamma^\mu\gamma^0, \gamma^0\gamma^\nu\gamma^0\right] \\
&\rightsquigarrow S^{\mu\nu\dagger} = \gamma^0 S^{\mu\nu}\gamma^0.
\end{aligned} \tag{8.72}$$

Thus, we have for the transformation of the conjugate spinor field,

$$\psi^\dagger \to \psi^\dagger\gamma^0\left(1 + \frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}\right)\gamma^0, \tag{8.73}$$

and for finite Lorentz transformations,

$$\psi^\dagger \to \psi^\dagger \gamma^0 \Lambda_{\frac{1}{2}}^{-1} \gamma^0. \tag{8.74}$$

We observe that $\psi^\dagger$ does not really transform with the inverse transformation of ψ. However, $\bar{\psi}$ does. Multiplying both sides of the above transformation rule by γ^0 from the right, we obtain

$$\begin{aligned} \psi^\dagger \gamma^0 &\to \psi^\dagger \gamma^0 \Lambda_{\frac{1}{2}}^{-1} \left(\gamma^0\right)^2 \\ \rightsquigarrow \bar{\psi} &\to \bar{\psi} \Lambda_{\frac{1}{2}}^{-1}. \end{aligned} \tag{8.75}$$

It is now obvious that the mass term in the Dirac Lagrangian is Lorentz invariant,

$$m\bar{\psi}\psi \to m\bar{\psi}\Lambda_{\frac{1}{2}}^{-1}\Lambda_{\frac{1}{2}}\psi = m\bar{\psi}\psi. \tag{8.76}$$

The derivative term of the Dirac Lagrangian is also Lorentz invariant,

$$\begin{aligned} \bar{\psi}\gamma^\mu \partial_\mu \psi &\to \bar{\psi}\Lambda_{\frac{1}{2}}^{-1}\gamma^\mu \left(\Lambda^{-1}\right)_\mu^{\ \nu} \Lambda_{\frac{1}{2}} \partial_\nu \psi \\ &= \bar{\psi}\left(\Lambda_{\frac{1}{2}}^{-1}\gamma^\mu \Lambda_{\frac{1}{2}}\right)\left(\Lambda^{-1}\right)_\mu^{\ \nu} \partial_\nu \psi \\ &= \bar{\psi}\gamma^\rho \Lambda_\rho^{\ \mu}\left(\Lambda^{-1}\right)_\mu^{\ \nu} \partial_\nu \psi \\ &= \bar{\psi}\gamma^\rho \delta_\rho^{\ \nu} \partial_\nu \psi \\ &= \bar{\psi}\partial\!\!\!/\psi. \end{aligned} \tag{8.77}$$

We see that the Dirac Lagrangian contains only the two simplest scalar terms that may be formed out of fields transforming in the spinor representation of the Lorentz group. The relative sign between these two terms is easy to determine by requiring that the solutions of the equations of motion also satisfy the Klein-Gordon equation, which gives the relativistically correct energy-momentum relation, $p^2 = m^2$.

8.5 General representations of the Lorentz group

Why should we care about finding more representations? The question is deeper than it looks: what is a particle? We think of elementary particles as point-like objects which nevertheless carry properties, such as mass and spin. Since the work of Wigner, these properties are understood as labels of the irreducible representations of the symmetry group of space-time, the Poincaré group (Lorentz transformations together with translations in space and time). The translations are generated by the momentum operator P^μ, and $P^2 = m^2$ is the same in every frame; for a massive particle we can go to its rest frame and the mass labels the representation. The Lorentz transformations add the information about spin. Asking which other representations of the Lorentz group exist is therefore asking which other kinds of particles, besides spin-0 and spin-1 bosons and spin-$\frac{1}{2}$ fermions, a relativistic theory can describe. A systematic construction along these lines can be found in Weinberg, Vol. I, Chapter 2.

The scalar, vector, and spinor representations are the simplest representations of the Lorentz group. However, there is an infinite number of $n \times n$ representations. All of them can be found easily by observing that the Lie algebra of the Lorentz group is the same as

for an $SU(2) \times SU(2)$ group. Let us consider the generators in an arbitrary representation, which satisfy the commutation relations

$$[J^{\mu_1\mu_2}, J^{\mu_3\mu_4}] = i\left(g^{\mu_2\mu_3}J^{\mu_1\mu_4} + g^{\mu_1\mu_4}J^{\mu_2\mu_3} - g^{\mu_2\mu_4}J^{\mu_1\mu_3} - g^{\mu_1\mu_3}J^{\mu_2\mu_4}\right).$$

We now define the linear combinations,

$$K_i \equiv J^{0i}, \tag{8.78}$$

and

$$L_i \equiv \frac{1}{2}\epsilon_{ijk}J^{jk}. \tag{8.79}$$

We can replace the basis of the six independent generators $J^{\mu\nu}$ (recall that they are antisymmetric) by the linear combinations $\{L_i, K_i\}$. The generators L_i are generators of rotations around the i-th axis, and the generators K_i are generators of boosts in the i-th direction (**exercise**: demonstrate this for the generators in the vector representation). Rotations are a subgroup of the Lorentz group, which satisfy an $SU(2)$ algebra. We find the commutation relations,

$$\left[L_i, L_j\right] = i\epsilon_{ijk}L_k. \tag{8.80}$$

However, boosts do not form a subgroup of the Lorentz group. Two boosts in two different directions are not a boost in a third direction, but rather a rotation. We find that the commutators of boost generators are,

$$\left[K_i, K_j\right] = -i\epsilon_{ijk}L_k. \tag{8.81}$$

The commutator of a boost and a rotation generator is,

$$\left[L_i, K_j\right] = i\epsilon_{ijk}K_k. \tag{8.82}$$

Define now the linear combinations,

$$J_i^{\pm} \equiv \frac{L_i \pm iK_i}{2}. \tag{8.83}$$

Then the Lie algebra breaks into two independent $SU(2)$ algebras,

$$\left[J_i^+, J_j^+\right] = i\epsilon_{ijk}J_k^+, \tag{8.84}$$

$$\left[J_i^-, J_j^-\right] = i\epsilon_{ijk}J_k^-, \tag{8.85}$$

and

$$\left[J_i^+, J_j^-\right] = 0. \tag{8.86}$$

Remark. Since $J_i^{\pm}$ contain the factor i, this is a statement about the complexified Lie algebra, $so(3,1)_{\mathbb{C}} \simeq su(2)_{\mathbb{C}} \oplus su(2)_{\mathbb{C}}$. The real Lorentz algebra is $so(3,1) \simeq sl(2,\mathbb{C})$, not $su(2) \oplus su(2) \simeq so(4)$; for Hermitian L_i, K_i one has $J_i^{\pm\dagger} = J_i^{\mp}$, so complex conjugation exchanges the two factors, which is why $(j^+, j^-) \leftrightarrow (j^-, j^+)$ under conjugation (and parity). The labels (j^+, j^-) classify all finite-dimensional irreducible representations (see Weinberg, Vol. I, Sect. 5.6).

We have already learned about the representations of the $SU(2)$ group in quantum mechanics. These can be labelled according to a "spin" number,

$$j = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots$$

Each representation acts on $(2j+1)$ objects ψ_m with

$$m = -j, -j+1, \dots, j-1, j, \tag{8.87}$$

transforming them into each other under $SU(2)$.

A general representation of the Lorentz group acts on $(2j^+ + 1)(2j^- + 1)$ objects ψ_{m^+,m^-}, and is labelled by the numbers $\left(j^+, j^-\right)$, e.g.

$$(0,0), (\frac{1}{2},0), (0,\frac{1}{2}), (0,1), (\frac{1}{2},\frac{1}{2}), \dots. \tag{8.88}$$

The massive Dirac field transforms in the mixed representation,

$$(\frac{1}{2},0) \oplus (0,\frac{1}{2}).$$

8.6 Weyl spinors

As we have already noted, γ-matrices are not unique, and there are many equivalent representations. We introduce here a representation which better displays the decomposition of the Lorentz group into two $SU(2)$ subgroups. This is called the chiral or Weyl representation, defined as

$$\gamma^0 = I_{2\times2} \otimes \sigma_1, \qquad \gamma^i = \sigma_i \otimes (i\sigma_2). \tag{8.89}$$

As an exercise, you can prove that indeed this set of gamma-matrices satisfies the Clifford algebra,

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} I_{4\times4}. \tag{8.90}$$

In the chiral representation, the generators of the spinorial Lorentz group representation

$$S^{\mu\nu} = \frac{i}{4}\left[\gamma^\mu, \gamma^\nu\right], \tag{8.91}$$

take a block-diagonal form,

$$S^{0j} = -\frac{i}{2}\sigma_j \otimes \sigma_3 = -\frac{i}{2}\begin{pmatrix} \sigma_j & 0_{2\times2} \\ 0_{2\times2} & -\sigma_j \end{pmatrix}, \tag{8.92}$$

and

$$S^{jk} = \frac{1}{2}\epsilon_{jkl}\sigma_l \otimes I_{2\times2} = \frac{1}{2}\epsilon_{jkl}\begin{pmatrix} \sigma_l & 0_{2\times2} \\ 0_{2\times2} & \sigma_l \end{pmatrix}. \tag{8.93}$$

A Lorentz transformation for a spinor is

$$\Lambda_{\frac{1}{2}} = e^{-\frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}}, \tag{8.94}$$

and the exponent takes the explicit form,

$$\begin{aligned} -\frac{i}{2}\omega_{\mu\nu}S^{\mu\nu} &= -\frac{i}{2}\omega_{0j}S^{0j} - \frac{i}{2}\omega_{j0}S^{j0} - \frac{i}{2}\omega_{jk}S^{jk} \\ &= -\vec{\beta}\cdot\left(\frac{\vec{\sigma}}{2}\otimes\sigma_3\right) - i\vec{\theta}\cdot\left(\frac{\vec{\sigma}}{2}\otimes I_{2\times2}\right), \end{aligned} \tag{8.95}$$

with

$$\beta_j = \omega_{0j} \text{ and } \theta_l = \frac{1}{2}\epsilon_{jkl}\omega^{jk}. \tag{8.96}$$

Under a small Lorentz transformation, a spinor ψ transforms as

$$\begin{aligned} \psi \to \psi' &= e^{-\frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}}\psi \\ &\simeq \left(1 - \frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}\right)\psi \\ &= \begin{pmatrix} \left[1 - \frac{i}{2}\left(\vec{\theta} - i\vec{\beta}\right)\cdot\vec{\sigma}\right]\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \\ \\ \left[1 - \frac{i}{2}\left(\vec{\theta} + i\vec{\beta}\right)\cdot\vec{\sigma}\right]\begin{pmatrix}\psi_3\\ \psi_4\end{pmatrix} \end{pmatrix}. \end{aligned} \tag{8.97}$$

The two-dimensional objects,

$$\psi_L \equiv \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \text{ and } \psi_R \equiv \begin{pmatrix}\psi_3\\ \psi_4\end{pmatrix}, \tag{8.98}$$

are called "Weyl spinors" and transform independently under a Lorentz transformation,

$$\psi_L \to \psi_L' = \left[1 - \frac{i}{2}\left(\vec{\theta} - i\vec{\beta}\right)\cdot\vec{\sigma}\right]\psi_L, \tag{8.99}$$

and

$$\psi_R \to \psi_R' = \left[1 - \frac{i}{2}\left(\vec{\theta} + i\vec{\beta}\right)\cdot\vec{\sigma}\right]\psi_R. \tag{8.100}$$

In exponential form,

$$\psi_L \to \psi_L' = \Lambda_L\psi_L \text{ and } \psi_R \to \psi_R' = \Lambda_R\psi_R, \tag{8.101}$$

with

$$\Lambda_L = e^{-\frac{i}{2}\left(\vec{\theta} - i\vec{\beta}\right)\cdot\vec{\sigma}}, \tag{8.102}$$

and

$$\Lambda_R = e^{-\frac{i}{2}\left(\vec{\theta} + i\vec{\beta}\right)\cdot\vec{\sigma}}. \tag{8.103}$$

We now write the Dirac equation in terms of Weyl spinors ψ_L and ψ_R,

$$\begin{aligned} 0 &= (i\partial\!\!\!/ - m)\,\psi \\ &= \left(i\gamma^0\partial_0 + i\gamma^j\partial_j - m1_{4\times4}\right)\psi \\ &= \left(i(I_{2\times2}\otimes\sigma_1)\partial_0 + i\left(\vec{\sigma}\otimes(i\sigma_2)\right)\vec{\nabla} - m1_{2\times2}\otimes 1_{2\times2}\right)\psi \\ \\ &\rightsquigarrow \begin{pmatrix} -m & i\left(\partial_0 + \vec{\sigma}\cdot\vec{\nabla}\right) \\ i\left(\partial_0 - \vec{\sigma}\cdot\vec{\nabla}\right) & -m \end{pmatrix}\begin{pmatrix}\psi_L\\ \psi_R\end{pmatrix} = 0. \end{aligned} \tag{8.104}$$

The mass m mixes the two Weyl spinors ψ_L, ψ_R. However, for a massless particle the Dirac equation is diagonalised into two independent equations for the Weyl spinors,

$$\partial_0\psi_R = -\vec{\sigma}\cdot\vec{\nabla}\psi_R, \tag{8.105}$$

and

$$\partial_0 \psi_L = \vec{\sigma} \cdot \vec{\nabla} \psi_L. \tag{8.106}$$

These are known as the Weyl equations. For plane-wave solutions (which correspond to particles of definite momentum),

$$\psi_{L,R}(\vec{x}, t) = u_{L,R} e^{-ip\cdot x}, \tag{8.107}$$

the Weyl equations turn into

$$\frac{\vec{p} \cdot \vec{\sigma}}{|\vec{p}|} \psi_L = -\psi_L, \tag{8.108}$$

and

$$\frac{\vec{p} \cdot \vec{\sigma}}{|\vec{p}|} \psi_R = +\psi_R. \tag{8.109}$$

Therefore, the $\psi_{L,R}$ plane-wave solutions are eigenstates of the "helicity operator"

$$\hat{h} \equiv \frac{\vec{p} \cdot \vec{\sigma}}{|\vec{p}|} = \hat{p} \cdot \vec{\sigma}, \tag{8.110}$$

which measures the projection of the spin of the particle on the direction of the motion. The spin of the "right-handed" spinors ψ_R is aligned with the momentum of the particle, while it points in the opposite direction for "left-handed" spinors ψ_L. Notice that only massless particles can have definite helicity, while for massive particles the value of the helicity is frame-dependent. The reason is simple. A massive particle moves slower than light. We can then boost to a frame which moves faster than the particle in the same direction. In that frame the momentum of the particle points in the opposite direction, while its spin is unchanged, and its helicity has flipped sign. For a massless particle no such frame exists. This is also why a massive spin-$\frac{1}{2}$ particle needs both ψ_L and ψ_R. A second reason, which we will meet in the next chapters, is parity: a parity transformation exchanges ψ_L and ψ_R, and a parity-invariant theory, such as quantum electrodynamics where left-handed and right-handed electrons interact with the photon in the same way, requires both of them. Packing them together into a Dirac spinor is then the natural choice. Conversely, when the masses can be neglected, as for quarks and leptons at very high energies, it is often much simpler to compute with the two-component Weyl spinors; this is the basis of the so-called helicity amplitude methods.

Let us now return to the general case of massive particles. We have:

$$\bar{\psi} = \psi^\dagger \gamma^0 = \left(\psi_L^\dagger, \psi_R^\dagger\right) \begin{pmatrix} 0_{2\times 2} & I_{2\times 2} \\ I_{2\times 2} & 0_{2\times 2} \end{pmatrix} = \left(\psi_R^\dagger, \psi_L^\dagger\right). \tag{8.111}$$

Then, the Dirac Lagrangian in terms of Weyl spinors becomes (**exercise**):

$$\begin{aligned} \mathcal{L}_{Dirac} &= \bar{\psi}\left(i\partial\!\!\!/ - m\right)\psi \\ &= i\psi_L^\dagger \left(\partial_0 - \vec{\sigma}\cdot\vec{\nabla}\right)\psi_L + i\psi_R^\dagger \left(\partial_0 + \vec{\sigma}\cdot\vec{\nabla}\right)\psi_R - m\left(\psi_L^\dagger \psi_R + \psi_R^\dagger \psi_L\right). \end{aligned} \tag{8.112}$$

We observe that it is only the mass term which mixes the left-handed and right-handed spinors. For a massless theory, the left-handed and right-handed fields are oblivious of each other.

Finally, we note (**exercise**) that the Dirac Lagrangian is symmetric under a $U(1)$ phase transformation,

$$\psi \to \psi' = e^{i\theta}\psi \leftrightarrow \psi_{L,R} \to \psi'_{L,R} = e^{i\theta}\psi_{L,R}. \tag{8.113}$$

8.7 Majorana equation

Exercise: *Prove the identity*

$$\sigma_2 \vec{\sigma}^* = -\vec{\sigma}\sigma_2. \tag{8.114}$$

Then, the spinor

$$\psi_L^c \equiv \sigma_2 \psi_L^*, \tag{8.115}$$

transforms as,

$$\begin{aligned} \sigma_2\psi_L^* &\rightarrow \sigma_2(1 - \frac{i}{2}\left(\vec{\theta} - i\vec{\beta}\right)\cdot\vec{\sigma})^*\psi_L^* \\ &= \sigma_2(1 + \frac{i}{2}\left(\vec{\theta} + i\vec{\beta}\right)\cdot\vec{\sigma}^*)\psi_L^* \\ &= (1 - \frac{i}{2}\left(\vec{\theta} + i\vec{\beta}\right)\cdot\vec{\sigma})\sigma_2\psi_L^* \\ &\rightsquigarrow \psi_L^c \rightarrow \Lambda_R\psi_L^c. \end{aligned} \tag{8.116}$$

Similarly,

$$\psi_R^c \rightarrow \Lambda_L\psi_R^c. \tag{8.117}$$

We found a somewhat surprising result, that the spinor ψ_L^c (which is made out of the components of a left-handed Weyl spinor) transforms as a right-handed Weyl-spinor. The Dirac Lagrangian for a massive spin-$\frac{1}{2}$ fermion is therefore not the only one which is Lorentz invariant. We could write a Lagrangian made exclusively out of left-handed spinors, replacing ψ_R with ψ_L^c in the mass term of the Dirac Lagrangian of Eq. (8.112). The Lagrangian we obtain in this way is due to Majorana:

$$\begin{aligned} \mathcal{L}_{Maj} &= i\psi_L^\dagger\left(\partial_0 - \vec{\sigma}\cdot\vec{\nabla}\right)\psi_L - \frac{m}{2}\left(\psi_L^\dagger\psi_L^c + \psi_L^{c\dagger}\psi_L\right) \\ &= i\psi_L^\dagger\left(\partial_0 - \vec{\sigma}\cdot\vec{\nabla}\right)\psi_L - \frac{m}{2}\left[\psi_L^T\sigma_2\psi_L + \psi_L^\dagger\sigma_2\psi_L^*\right]. \end{aligned} \tag{8.118}$$

Exercise: *What is the Majorana Lagrangian for a right-handed Weyl spinor?*

Notice that, unlike the Dirac Lagrangian, the Majorana Lagrangian is not invariant under a $U(1)$ transformation. In particular, under

$$\psi_L \rightarrow e^{i\theta}\psi_L,$$

the mass term depends on θ, since

$$\psi_L^T(i\sigma_2)\psi_L \rightarrow e^{2i\theta}\psi_L^T(i\sigma_2)\psi_L, \quad \psi_L^\dagger(i\sigma_2)\psi_L^* \rightarrow e^{-2i\theta}\psi_L^\dagger(i\sigma_2)\psi_L^*.$$

The absence of a $U(1)$ symmetry means that there is no conserved quantity which we can assign to the electric charge. Therefore, the Majorana equation could not be used to describe spin-$\frac{1}{2}$ electrically charged particles, such as electrons.

For a long time, it was believed that only the Dirac equation had a physical realisation and that the Majorana equation was a mere theoretical possibility. However, it has now been shown experimentally that neutrinos, which have no electric charge and no electromagnetic interactions, have a mass. It is believed that the Majorana equation can be used to give a mass contribution to neutrinos. It is still an open experimental question whether the mass of the neutrinos is "Dirac" or "Majorana".

Let us now focus on the mass term of the Lagrangian. If we write

$$\psi_L = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \tag{8.119}$$

then

$$\psi_L^c = \sigma_2 \psi_L^* = -i \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \psi_1^* \\ \psi_2^* \end{pmatrix} = i \begin{pmatrix} -\psi_2^* \\ \psi_1^* \end{pmatrix}, \tag{8.120}$$

and then

$$\psi_L^\dagger \psi_L^c = i \left(-\psi_1^* \psi_2^* + \psi_2^* \psi_1^* \right). \tag{8.121}$$

If ψ_1, ψ_2 are customary classical fields which commute with each other, then the above expression is zero and therefore the mass term does not exist. However, there is another possibility. If we impose that the ψ_i's are anticommuting (so-called *Grassmann variables*), then

$$\psi_1^* \psi_2^* = -\psi_2^* \psi_1^*, \tag{8.122}$$

and the mass term, $\psi_L^\dagger \psi_L^c = -2i\psi_1^* \psi_2^*$, is non-zero. We will see that the quantisation of fermion fields indeed requires anticommuting fields.

8.7.1 The Majorana equation of motion

The equations of motion of the Majorana Lagrangian give the Majorana equation,

$$i \left(\partial_0 - \vec{\sigma} \cdot \vec{\nabla} \right) \psi_L = m \psi_L^c. \tag{8.123}$$

Exercise: *Prove that in the massless limit:*

$$\hat{h} \psi_L^c = +\psi_L^c, \quad \hat{h} \psi_R^c = -\psi_R^c.$$

9

Classical solutions of the Dirac equation

In this chapter, we shall solve classically the equations of motion for the Dirac field, i.e. the classical Dirac equation. As we have already demonstrated, a field ψ which satisfies the Dirac equation also satisfies the Klein-Gordon equation. So, it must admit a plane-wave solution. We consider first the general case

$$\psi(x) = u(p)e^{-ip\cdot x}, \tag{9.1}$$

with

$$p^2 = m^2. \tag{9.2}$$

For now, we shall restrict ourselves to the cases with positive energy solutions, $p^0 > 0$. Substituting into the Dirac equation, we obtain additional restrictions for the spinor $u(p)$,

$$(i\partial\!\!\!/ - m)\, u(p)e^{-ip\cdot x} = 0$$

$$\rightsquigarrow \quad (p\!\!\!/ - m)\, u(p) = 0. \tag{9.3}$$

Using the Weyl representation for the γ-matrices, $\gamma^0 = I_{2\times 2} \otimes \sigma_1$, $\gamma^j = \sigma_j \otimes (i\sigma_2)$, we obtain

$$\begin{pmatrix} -mI_{2\times 2} & p^0 I_{2\times 2} - \vec{\sigma}\cdot\vec{p} \\ p^0 I_{2\times 2} + \vec{\sigma}\cdot\vec{p} & -mI_{2\times 2} \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{9.4}$$

where we write the four-component spinor $u(p)$ in terms of two Weyl spinors ξ, η with two components each:

$$u(p) \equiv \begin{pmatrix} \xi \\ \eta \end{pmatrix}. \tag{9.5}$$

It is of course possible to find solutions of the above equation directly. However, we may also want to use some cleverness, and exploit the fact that the field ψ transforms as a spinor. Consequently,

$$\psi'(x') = \Lambda_{\frac{1}{2}}\psi(x), \tag{9.6}$$

and equivalently

$$\begin{aligned} u'(p')e^{-ip'\cdot x'} &= \Lambda_{\frac{1}{2}} u(p)e^{-ip\cdot x} \\ \rightsquigarrow u'(\Lambda p)e^{-i(\Lambda p)\cdot(\Lambda x)} &= \Lambda_{\frac{1}{2}} u(p)e^{-ip\cdot x} \\ \rightsquigarrow u'(\Lambda p)e^{-ip\cdot x} &= \Lambda_{\frac{1}{2}} u(p)e^{-ip\cdot x} \\ u'(\Lambda p) &= \Lambda_{\frac{1}{2}} u(p), \end{aligned} \tag{9.7}$$

i.e. $u(p)$ also transforms as a spinor under a Lorentz transformation,

$$p^\mu \to p^{\mu\prime} = \Lambda^\mu_{\ \nu} p^\nu. \tag{9.8}$$

The strategy for finding a general solution $u(p)$ is then simple. We first solve the Dirac equation for an especially simple value of the momentum p^μ, and then we perform a Lorentz transformation to obtain the result for a general value of the momentum.

9.1 Solution in the rest frame

The Dirac equation (Eq. 9.4) assumes its simplest form in the rest frame,

$$p^\mu = m(1, \vec{0}), \tag{9.9}$$

yielding

$$\begin{pmatrix} -mI_{2\times2} & mI_{2\times2} \\ mI_{2\times2} & -mI_{2\times2} \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{9.10}$$

and, equivalently,

$$\xi = \eta. \tag{9.11}$$

The general solution for the spinor u in the rest frame is,

$$u\left(p^\mu = (m, \vec{0})\right) = \sqrt{m} \begin{pmatrix} \xi \\ \xi \end{pmatrix}. \tag{9.12}$$

The factor $\sqrt{m}$ is a convenient normalisation for future purposes, and the Weyl spinor ξ is arbitrary. We can choose it to be a linear combination,

$$\xi = \sum_{s=+,-} c_s \xi^s, \tag{9.13}$$

with

$$\xi^+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \text{ and } \xi^- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{9.14}$$

Notice that this basis diagonalises σ_3, with

$$\sigma_3 \xi^\pm = \pm \xi^\pm. \tag{9.15}$$

From this solution of the Dirac equation in the rest frame, we can go to any other momentum value with a boost.

9.2 Lorentz boost of the rest-frame Dirac spinor along the z-axis

First, we boost any vector along the z-axis with a very small Lorentz transformation,

$$p^\mu \to p^{\mu\prime} = \Lambda^\mu_{\ \nu} p^\nu \simeq p^\mu + \omega^\mu_{\ \nu} p^\nu, \tag{9.16}$$

where,

$$\omega_{\mu\nu} = 0, \text{ except } \omega_{03} = -\omega_{30} = \Upsilon. \tag{9.17}$$

Notice that,

$$\omega^0_{\ 3} = g^{00}\omega_{03} = \Upsilon, \qquad \omega^3_{\ 0} = g^{33}\omega_{30} = -(-\Upsilon) = \Upsilon. \tag{9.18}$$

The transformed momenta are,

$$p^{0\prime} = p^0 + \Upsilon p^3, p^{1\prime} = p^1, p^{2\prime} = p^2, p^{3\prime} = p^3 + \Upsilon p^0, \tag{9.19}$$

which is conveniently written as,

$$\begin{pmatrix} p^{0\prime} \\ p^{3\prime} \end{pmatrix} = (I_{2\times 2} + \Upsilon\sigma_1)\begin{pmatrix} p^0 \\ p^3 \end{pmatrix}. \tag{9.20}$$

For a finite boost of a vector along the z-direction, we have

$$\begin{pmatrix} p^{0\prime} \\ p^{3\prime} \end{pmatrix} = e^{\Upsilon\sigma_1}\begin{pmatrix} p^0 \\ p^3 \end{pmatrix}. \tag{9.21}$$

We can compute the exponential easily, noting that $\sigma_1^2 = I_{2\times 2}$,

$$\begin{aligned} e^{\Upsilon\sigma_1} &= \sum_{n=0}^{\infty}\frac{\Upsilon^n}{n!}\sigma_1^n \\ &= \sum_{k=0}^{\infty}\frac{\Upsilon^{2k}}{(2k)!}\sigma_1^{2k} + \sum_{k=0}^{\infty}\frac{\Upsilon^{2k+1}}{(2k+1)!}\sigma_1^{2k+1} \\ &= I_{2\times 2}\sum_{k=0}^{\infty}\frac{\Upsilon^{2k}}{(2k)!} + \sigma_1\sum_{k=0}^{\infty}\frac{\Upsilon^{2k+1}}{(2k+1)!} \\ \rightsquigarrow e^{\Upsilon\sigma_1} &= I_{2\times 2}\cosh\Upsilon + \sigma_1\sinh\Upsilon. \end{aligned} \tag{9.22}$$

We apply this transformation for the momentum vector at rest,

$$\begin{aligned} \begin{pmatrix} m \\ 0 \end{pmatrix} \to \begin{pmatrix} E \\ p^3 \end{pmatrix} &= (I_{2\times 2}\cosh\Upsilon + \sigma_1\sinh\Upsilon)\begin{pmatrix} m \\ 0 \end{pmatrix} \\ \rightsquigarrow \begin{pmatrix} E \\ p^3 \end{pmatrix} &= \begin{pmatrix} m\cosh\Upsilon \\ m\sinh\Upsilon \end{pmatrix}. \end{aligned} \tag{9.23}$$

We can then invert the above equations and compute the parameter Υ which generates the Lorentz boost, known as *rapidity*, in terms of the components of the boosted momentum. We find (**exercise**),

$$e^{\Upsilon} = \frac{E+p^3}{m}, \quad e^{-\Upsilon} = \frac{E-p^3}{m}, \qquad \Upsilon = \frac{1}{2}\ln\frac{E+p^3}{E-p^3}. \tag{9.24}$$

Let us now find the transformation of a spinor, under the same Lorentz boost. The

transformation matrix is,

$$\begin{aligned}
\Lambda_{\frac{1}{2}} &= e^{-\frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}} = e^{-\frac{i}{2}\omega_{03}S^{03}-\frac{i}{2}\omega_{30}S^{30}} = e^{-\frac{Y}{2}\sigma_3\otimes\sigma_3} \\
&= \sum_{n=0}^{\infty}\frac{\left(-\frac{Y}{2}\right)^n}{n!}\left(\sigma_3\otimes\sigma_3\right)^n \\
&= \sum_{n=0}^{\infty}\frac{\left(-\frac{Y}{2}\right)^n}{n!}\sigma_3^n\otimes\sigma_3^n \\
&= \sum_{n=\text{odd}}^{\infty}\frac{\left(-\frac{Y}{2}\right)^n}{n!}\sigma_3^n\otimes\sigma_3^n + \sum_{n=\text{even}}^{\infty}\frac{\left(-\frac{Y}{2}\right)^n}{n!}\sigma_3^n\otimes\sigma_3^n \\
&= \sum_{n=\text{odd}}^{\infty}\frac{\left(-\frac{Y}{2}\right)^n}{n!}\sigma_3\otimes\sigma_3 + \sum_{n=\text{even}}^{\infty}\frac{\left(-\frac{Y}{2}\right)^n}{n!}I_{2\times2}\otimes I_{2\times2} \\
\rightsquigarrow \Lambda_{\frac{1}{2}} &= \cosh\left(\frac{Y}{2}\right)I_{2\times2}\otimes I_{2\times2} - \sinh\left(\frac{Y}{2}\right)\sigma_3\otimes\sigma_3. \qquad (9.25)
\end{aligned}$$

Applying this spin boost transformation to our solution of the Dirac equation in the rest frame, we obtain

$$\begin{aligned}
u(p) &= \Lambda_{\frac{1}{2}}\sqrt{m}\begin{pmatrix}\xi^s\\ \xi^s\end{pmatrix} \\
\rightsquigarrow u^s(p) &= \begin{pmatrix}\left[\sqrt{E+p^3}\frac{(1-\sigma_3)}{2}+\sqrt{E-p^3}\frac{(1+\sigma_3)}{2}\right]\xi^s \\ \left[\sqrt{E+p^3}\frac{(1+\sigma_3)}{2}+\sqrt{E-p^3}\frac{(1-\sigma_3)}{2}\right]\xi^s\end{pmatrix}, \qquad (9.26)
\end{aligned}$$

where we have used the expression for the rapidity, derived in Eq. 9.24. We observe that,

$$\left(\frac{1\pm\sigma_3}{2}\right)^2 = \frac{1\pm\sigma_3}{2}, \qquad (9.27)$$

or, equivalently,

$$\frac{1\pm\sigma_3}{2} = \sqrt{\frac{1\pm\sigma_3}{2}}, \qquad (9.28)$$

and that

$$\frac{1+\sigma_3}{2}\cdot\frac{1-\sigma_3}{2} = 0. \qquad (9.29)$$

We can then collect all terms in Eq. 9.26 under a common square root,

$$u^s(p) = \begin{pmatrix}\sqrt{(E+p^3)\frac{(1-\sigma_3)}{2}+(E-p^3)\frac{(1+\sigma_3)}{2}}\,\xi^s \\ \sqrt{(E+p^3)\frac{(1+\sigma_3)}{2}+(E-p^3)\frac{(1-\sigma_3)}{2}}\,\xi^s\end{pmatrix}, \qquad (9.30)$$

which simplifies to

$$u^s(p) = \begin{pmatrix}\sqrt{E-p^3\sigma_3}\,\xi^s \\ \sqrt{E+p^3\sigma_3}\,\xi^s\end{pmatrix}. \qquad (9.31)$$

This is the general solution for the spinor $u(p)$, when $p^\mu = (E, 0, 0, p^3)$.

9.3 Solution for an arbitrary vector

We can obtain the spinor solution $u(p)$ of the Dirac equation for an arbitrary vector p^μ, by performing a rotation of the solution that we found for $p^\mu = (E, 0, 0, p^3)$. In fact, things are simpler now and there is not much work to do, if we can write down the previous result in a covariant form.

We define,

$$\sigma^\mu \equiv \left(1, \vec{\sigma}\right), \tag{9.32}$$

and

$$\bar{\sigma}^\mu \equiv \left(1, -\vec{\sigma}\right). \tag{9.33}$$

Then the expression for the spinor $u^s(p)$, Eq. 9.31, is written as

$$u^s(p) = \begin{pmatrix} \sqrt{\sigma \cdot p}\, \xi^s \\ \sqrt{\bar{\sigma} \cdot p}\, \xi^s \end{pmatrix}. \tag{9.34}$$

A word of caution about the notation. $\sqrt{\sigma \cdot p}$ is the square root of the 2×2 matrix $\sigma \cdot p = \sigma^\mu p_\mu = E - \vec{\sigma} \cdot \vec{p}$, i.e. the matrix with positive eigenvalues whose square is $\sigma \cdot p$. It is **not** obtained by taking the square root of each matrix element. The two coincide only when $\sigma \cdot p$ is diagonal, as for $\vec{p}$ along the z-axis in Eq. 9.31. Using $(\sigma \cdot p)^2 = 2E\,(\sigma \cdot p) - m^2$, we can verify that, explicitly,

$$\sqrt{\sigma \cdot p} = \frac{\sigma \cdot p + m}{\sqrt{2(E+m)}}, \qquad \sqrt{\bar{\sigma} \cdot p} = \frac{\bar{\sigma} \cdot p + m}{\sqrt{2(E+m)}}.$$

You can verify (**exercise**) that the spinor of Eq. 9.34 is indeed a solution of the Dirac equation for an arbitrary vector p^μ. For this, you will need the generally useful identity,

$$(\sigma \cdot p)(\bar{\sigma} \cdot p) = p^2 = m^2, \tag{9.35}$$

and to rewrite the Dirac differential operator as

$$\not{p} - m = \begin{pmatrix} -m & p \cdot \sigma \\ p \cdot \bar{\sigma} & -m \end{pmatrix} \tag{9.36}$$

in Eq. 9.4.

9.4 A general solution

We have found a general solution for the Dirac equation, of the form

$$\psi(x) = u^s(p) e^{-ip\cdot x}, \text{ with } p^0 > 0. \tag{9.37}$$

Another possibility is,

$$\psi(x) = v(p) e^{+ip\cdot x}, \text{ with } p^0 > 0. \tag{9.38}$$

Substituting into the Dirac equation we find that the spinor $v^s(p)$ satisfies

$$\left(\not{p} + m\right) v(p) = 0. \tag{9.39}$$

Repeating the same steps as for $u(p)$, we find this equation is satisfied by spinors

$$v^s(p) = \begin{pmatrix} \sqrt{\sigma \cdot p}\, \xi^s \\ -\sqrt{\bar{\sigma} \cdot p}\, \xi^s \end{pmatrix}. \tag{9.40}$$

A general solution for the classical Dirac equation is,

$$\psi(x) = \sum_s \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \left[a_s(p) u^s(p) e^{-ip\cdot x} + b_s^*(p) v^s(p) e^{+ip\cdot x} \right], \tag{9.41}$$

where the $a_s(p)$ and $b_s(p)$ are arbitrary coefficients, and

$$\omega_p = p^0 = \sqrt{\vec{p}^2 + m^2}. \tag{9.42}$$

We have not discussed explicitly solutions with $p^0 < 0$. At the classical level, these are included in the above integral, as you can easily verify by repeating here the analysis of Section 5.1.1. Dirac interpreted these negative energy solutions with an ingenious theory (Dirac sea), which led to the discovery of antimatter. We do not need to bother with this interpretation, except out of historical interest. As with the Klein-Gordon equation, field quantisation consistently gives rise to physical states which have a positive energy.

10

Quantisation of the Dirac field

To quantise the Dirac field, we promote the field to an operator,

$$\psi(\vec{x},t) = \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \sum_s \left[a_s(p)u_s(p)e^{-ip\cdot x} + b_s^\dagger(p)v_s(p)e^{ip\cdot x}\right]. \tag{10.1}$$

The conjugate momentum of the field is,

$$\pi(\vec{x},t) = \frac{\partial \mathcal{L}}{\partial \dot{\psi}} = i\psi^\dagger(\vec{x},t), \tag{10.2}$$

where the Lagrangian of the Dirac field is,

$$\mathcal{L} = \bar{\psi}\left(i\partial\!\!\!/ - m\right)\psi. \tag{10.3}$$

The conjugate field is given by,

$$\psi^\dagger(\vec{x},t) = \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \sum_s \left[a_s^\dagger(p)u_s^\dagger(p)e^{+ip\cdot x} + b_s(p)v_s^\dagger(p)e^{-ip\cdot x}\right]. \tag{10.4}$$

We also find,

$$\bar{\psi}(\vec{x},t) \equiv \psi^\dagger(\vec{x},t)\gamma^0 = \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \sum_s \left[a_s^\dagger(p)\bar{u}_s(p)e^{+ip\cdot x} + b_s(p)\bar{v}_s(p)e^{-ip\cdot x}\right]. \tag{10.5}$$

In terms of ψ and $\bar{\psi}$, the operators a_s, b_s are written as,

$$a_s(p) = \int d^3\vec{x}\,\bar{u}_s(p)e^{ip\cdot x}\gamma^0\psi(x), \tag{10.6}$$

and

$$b_s(p) = \int d^3\vec{x}\,\bar{\psi}(x)\gamma^0 e^{+ip\cdot x}v_s(p). \tag{10.7}$$

To prove the above you will need the orthogonality relations of the spinors, $u_r^\dagger(p)u_s(p) = v_r^\dagger(p)v_s(p) = 2\omega_p\delta_{rs}$ and $u_r^\dagger(\vec{p})v_s(-\vec{p}) = 0$ (with $\vec{p} \to -\vec{p}$ at fixed ω_p). We shall also need the spinor summations,

$$\sum_s u_s(p)\bar{u}_s(p) = \not{p} + m, \text{ and } \sum_s v_s(p)\bar{v}_s(p) = \not{p} - m, \tag{10.8}$$

which can be derived easily from the explicit expressions for $u_s(p), v_s(p)$ of the previous chapter.

Using Noether's theorem, we can compute the Hamiltonian of the Dirac field, which is

$$H = \int \frac{d^3\vec{k}}{(2\pi)^3(2\omega_k)}\omega_k \sum_s \left[a_s^\dagger(k)a_s(k) - b_s(k)b_s^\dagger(k)\right]. \tag{10.9}$$

In the above, we have been careful to preserve the order of operators as they appear.

We can now proceed with the quantisation procedure. It is left as an exercise to show that imposing commutation relations leads to particle states $b_s^\dagger(p)\,|0\rangle$ which have a negative energy. This is due to the minus sign in the second term of the Hamiltonian integral in Eq. 10.9. To obtain positive energy eigenvalues for such states, we are forced to quantise by requiring anti-commutation relations,

$$\left\{\psi(\vec{x}_1,t),\pi(\vec{x}_2,t)\right\} = i\delta^{(3)}(\vec{x}_1-\vec{x}_2), \tag{10.10}$$

and

$$\left\{\psi(\vec{x}_1,t),\psi(\vec{x}_2,t)\right\} = \left\{\pi(\vec{x}_1,t),\pi(\vec{x}_2,t)\right\} = 0. \tag{10.11}$$

For the operators a_s, b_s this results in the anticommutation relations,

$$\left\{a_s(\vec{p}_1),a_{s'}^\dagger(\vec{p}_2)\right\} = \left\{b_s(\vec{p}_1),b_{s'}^\dagger(\vec{p}_2)\right\} = \delta_{s's}(2\pi)^3(2\omega_{p_1})\delta^{(3)}(\vec{p}_2-\vec{p}_1), \tag{10.12}$$

$$\left\{a_s(\vec{p}_1),a_{s'}(\vec{p}_2)\right\} = \left\{b_s(\vec{p}_1),b_{s'}(\vec{p}_2)\right\} = 0, \tag{10.13}$$

$$\left\{a_s^\dagger(\vec{p}_1),a_{s'}^\dagger(\vec{p}_2)\right\} = \left\{b_s^\dagger(\vec{p}_1),b_{s'}^\dagger(\vec{p}_2)\right\} = 0, \tag{10.14}$$

$$\left\{a_s(\vec{p}_1),b_{s'}^\dagger(\vec{p}_2)\right\} = \left\{b_s(\vec{p}_1),a_{s'}^\dagger(\vec{p}_2)\right\} = 0. \tag{10.15}$$

Anti-commutators replace commutators only in the quantisation conditions for the fermion fields. The time evolution of an operator is still given by the Heisenberg equation with a **commutator**, $\partial_t\psi = i\left[H,\psi\right]$, since the Hamiltonian is bilinear in the fermion fields. Commutators with such bilinears are computed easily with the identity

$$[AB,C] = A\,\{B,C\} - \{A,C\}\,B,$$

which uses only the anti-commutation relations. Using the anticommutation relation of Eq. 10.12, we cast the Hamiltonian of Eq. 10.9 as,

$$H = \int \frac{d^3\vec{k}}{(2\pi)^3(2\omega_k)}\omega_k \sum_s \left[a_s^\dagger(k)a_s(k) + b_s^\dagger(k)b_s(k)\right] + \langle 0|\,H\,|0\rangle\,, \tag{10.16}$$

with

$$\langle 0|\,H\,|0\rangle = -2\,\delta^{(3)}(0)\int d^3\vec{k}\,\omega_k, \tag{10.17}$$

where the factor 2 comes from the sum over the two spin states. **Exercise:** *Notice that the sign of the vacuum energy is now negative.*

1. *What would be the Casimir force in a gedanken experiment where we could impose that the Dirac field vanishes on two parallel plates?*

2. *In supersymmetric theories, for each fermion field there exists a partner boson field with the same mass. What is the Casimir force?*

As for the Klein-Gordon field, we can redefine the field Hamiltonian to have zero vacuum expectation value,

$$: H := H - \langle 0| H |0\rangle . \tag{10.18}$$

The subtraction of the vacuum energy is implemented at the operator level with normal ordering, which in the case of anticommuting fields requires an additional minus sign,

$$: b_s(k) b_s^\dagger(k) := -b_s^\dagger(k) b_s(k). \tag{10.19}$$

Again with the use of Noether's theorem, we find that the momentum is conserved and takes the form,

$$\vec{P} = \int \frac{d^3\vec{k}}{(2\pi)^3 (2\omega_k)} \vec{k} \sum_s \left[a_s^\dagger(k) a_s(k) + b_s^\dagger(k) b_s(k) \right] . \tag{10.20}$$

10.1 One-particle states

It is left as an exercise to repeat the steps that we followed in the case of the Schrödinger and the Klein-Gordon fields, and to prove that

- there exists a vacuum state with

$$a_s(p) |0\rangle = b_s(p) |0\rangle = 0, \tag{10.21}$$

- the operators $a_s(p), b_s(p), a_s^\dagger(p), b_s^\dagger(p)$ are ladder operators.

Notice that the annihilation operators acting on the vacuum give the number zero, and not the vacuum state $|0\rangle$, which has norm one. This is a frequent mistake.

We can build two types of one-particle states, using either the $a_s^\dagger$ creation operator

$$\left| p, s \right\rangle_a \equiv a_s^\dagger(p) |0\rangle , \tag{10.22}$$

or the $b_s^\dagger$ creation operator

$$\left| p, s \right\rangle_b \equiv b_s^\dagger(p) |0\rangle . \tag{10.23}$$

These states are degenerate eigenstates of the Hamiltonian and momentum operators. We define the field momentum four-vector,

$$P^\mu \equiv (: H :, \vec{P}). \tag{10.24}$$

Then

$$P^\mu \left| p, s \right\rangle_a = p^\mu \left| p, s \right\rangle_a , \tag{10.25}$$

and

$$P^\mu \left| p, s \right\rangle_b = p^\mu \left| p, s \right\rangle_b , \tag{10.26}$$

with

$$p^\mu = \left(\sqrt{\vec{p}^2 + m^2}, \vec{p} \right) . \tag{10.27}$$

These are states of positive energy, with the correct relativistic energy-momentum relation,

$$p^2 = m^2. \tag{10.28}$$

10.1.1 Particles and anti-particles

The Dirac Lagrangian is symmetric under the $U(1)$ transformation,

$$\psi(x) \to \psi'(x) = e^{i\theta}\psi(x), \qquad \psi^\dagger(x) \to \psi^{\dagger\prime}(x) = e^{-i\theta}\psi^\dagger(x). \tag{10.29}$$

For small values of θ, we find

$$\delta\psi = \psi' - \psi = i\theta\psi, \tag{10.30}$$

and

$$\delta\psi^\dagger = \psi^{\dagger\prime} - \psi^\dagger = -i\theta\psi^\dagger, \tag{10.31}$$

or, equivalently,

$$\delta\bar{\psi} = \bar{\psi}' - \bar{\psi} = -i\theta\bar{\psi}. \tag{10.32}$$

Using Noether's theorem, we find that the current,

$$\begin{aligned} J^\mu &= \delta\bar{\psi}\frac{\partial\mathcal{L}}{\partial(\partial_\mu\bar{\psi})} + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\psi)}\delta\psi, \\ J^\mu &= \bar{\psi}\gamma^\mu\psi, \end{aligned} \tag{10.33}$$

is conserved, $\partial_\mu J^\mu = 0$. The corresponding conserved charge is,

$$Q = \int d^3\vec{x} J^0 = \int d^3\vec{x}\bar{\psi}\gamma^0\psi = \int d^3\vec{x}\psi^\dagger\psi. \tag{10.34}$$

We can express Q in terms of creation and annihilation operators,

$$Q = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \sum_s \left[a_s^\dagger(k)a_s(k) + b_s(k)b_s^\dagger(k)\right]. \tag{10.35}$$

We calibrate the charge of the vacuum state to be zero with normal ordering,

$$\begin{aligned} :Q: &= Q - \langle 0|Q|0\rangle \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \sum_s \left[: a_s^\dagger(k)a_s(k) : + : b_s(k)b_s^\dagger(k) :\right] \\ :Q: &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \sum_s \left[a_s^\dagger(k)a_s(k) - b_s^\dagger(k)b_s(k)\right]. \end{aligned} \tag{10.36}$$

The states $|p,s\rangle_a$ and $|p,s\rangle_b$ are non-degenerate eigenstates of the charge operator. We find,

$$:Q:|p,s\rangle_a = (+1)|p,s\rangle_a, \tag{10.37}$$

and

$$:Q:|p,s\rangle_b = (-1)|p,s\rangle_b. \tag{10.38}$$

Therefore the operators $a_s^\dagger(p)$ give rise to particles with momentum p^μ and charge $(+1)$, while the operators $b_s^\dagger(p)$ give rise to anti-particles with the same momentum p^μ and opposite charge (-1). To emphasise this, we shall display explicitly the charge of one-particle states,

$$|p,s,+\rangle \equiv |p,s\rangle_a = a_s^\dagger(p)|0\rangle, \text{ and } |p,s,-\rangle \equiv |p,s\rangle_b = b_s^\dagger(p)|0\rangle. \tag{10.39}$$

At this stage Q is only the conserved charge of a global $U(1)$ symmetry. When we couple the Dirac field to the photon in quantum electrodynamics, we will see that Q (multiplied by the elementary charge) is the electric charge. The electron and the positron are created by the two different operators of the same field; they have the same mass and opposite charges, and the same holds for a quark with charge $\frac{2}{3}$ and its anti-quark with charge $-\frac{2}{3}$. Anti-particles are not a privilege of fermions: the complex Klein-Gordon field has them too. For a real field, such as the photon field, the particle is its own anti-particle.

10.1.2 Particles and anti-particles of spin-$\frac{1}{2}$

We have demonstrated in previous chapters that the Dirac Lagrangian is invariant under Lorentz transformations. The spinor field transforms "at a point" as,

$$\psi(x) \to \psi'(x) = \Lambda_{\frac{1}{2}}\psi(\Lambda^{-1}x). \tag{10.40}$$

The variation of the field "at a point" is

$$\begin{aligned}
\delta_*\psi &\equiv \psi'(x) - \psi(x) \\
&= \Lambda_{\frac{1}{2}}\psi(\Lambda^{-1}x) - \psi(x) \\
&= \left(1 - \frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}\right)\psi(x^\mu - \omega^\mu_{\ \nu}x^\nu) - \psi(x^\mu) + O(\omega^2) \\
&= \psi(x^\mu - \omega^\mu_{\ \nu}x^\nu) - \psi(x^\mu) - \frac{i}{2}\omega_{\mu\nu}S^{\mu\nu}\psi(x^\mu - \omega^\mu_{\ \nu}x^\nu) + O(\omega^2).
\end{aligned} \tag{10.41}$$

We can expand,

$$\psi(x^\mu - \omega^\mu_{\ \nu}x^\nu) = \psi(x^\mu) - \omega_{\mu\nu}x^\nu\partial^\mu\psi = \psi(x) - i\frac{\omega_{\mu\nu}}{2}J^{\mu\nu}_{\text{scalar}}\psi(x), \tag{10.42}$$

where we recall that

$$J^{\mu\nu}_{\text{scalar}} = i\left(x^\mu\partial^\nu - x^\nu\partial^\mu\right). \tag{10.43}$$

Eq. 10.41 becomes,

$$\delta_*\psi = -\frac{i}{2}\omega_{\mu\nu}\left(J^{\mu\nu}_{\text{scalar}} + S^{\mu\nu}\right)\psi. \tag{10.44}$$

Noether's theorem, Eq. 3.55, leads to the result that the quantity

$$J^{\mu\nu} = -i\int d^3\vec{x}\,\frac{\partial\mathcal{L}}{\partial(\partial_0\psi)}\left(J^{\mu\nu}_{\text{scalar}} + S^{\mu\nu}\right)\psi, \tag{10.45}$$

is conserved. Explicitly,

$$J^{\mu\nu} = \int d^3\vec{x}\,\bar{\psi}\gamma^0\left(J^{\mu\nu}_{\text{scalar}} + S^{\mu\nu}\right)\psi, \tag{10.46}$$

and, equivalently,

$$J^{\mu\nu} = \int d^3\vec{x}\,\psi^\dagger\left(J^{\mu\nu}_{\text{scalar}} + S^{\mu\nu}\right)\psi. \tag{10.47}$$

This is very close to what we find for the total angular momentum of a scalar field, but not exactly the same. The term with $S^{\mu\nu}$ is novel, and it arises here for the first time. This term is responsible for giving spin, an intrinsic angular momentum, to Dirac particle states.

How can we figure out if a particle has spin? We can do the following test. We put the particle at rest, and measure its total angular momentum. If we find it zero, then the

particle has an angular momentum only when it moves and therefore has no intrinsic spin. If it is not zero although the particle has zero momentum, then it is all due to the intrinsic spin of the particle.

First, we would like to use normal ordering, as with every conserved quantity, to make sure that the vacuum has a zero total angular momentum:

$$: J^{\mu\nu} : |0\rangle = 0. \tag{10.48}$$

Let us consider a state of a positively charged Dirac particle with zero momentum,

$$|0, s, +\rangle = a_s^\dagger(0)\,|0\rangle\,. \tag{10.49}$$

Does it have any angular momentum? The action of the total angular momentum on the state is,

$$: J^{\mu\nu} : |0, s, +\rangle = \left\{: J^{\mu\nu} :, a_s^\dagger(0)\right\}|0\rangle - a_s^\dagger(0) : J^{\mu\nu} : |0\rangle = \left\{: J^{\mu\nu} :, a_s^\dagger(0)\right\}|0\rangle\,. \tag{10.50}$$

We now specialise to a rotation around the z-axis,

$$\omega_{12} = -\omega_{21} = \theta, \text{ otherwise } \omega_{\mu\nu} = 0. \tag{10.51}$$

The algebra is shorter than it looks if we think first about which terms of $: J^{12} :$ can contribute. Inserting the expansions of ψ and $\psi^\dagger$ into Eq. 10.47, we obtain terms with the operators $a^\dagger a$ and $b^\dagger b$ (after normal ordering) and terms with ba and $a^\dagger b^\dagger$. The latter annihilate or create a particle-antiparticle pair and change the energy; they cannot contribute, since the conserved $: J^{12} :$ commutes with the Hamiltonian. Acting on $a_s^\dagger(0)\,|0\rangle$, every operator $b_r(\vec{k})$, and every $a_r(\vec{k})$ with $\vec{k} \neq 0$, can be anti-commuted to the right and annihilates the vacuum. Only the terms with $a_r^\dagger(0)\,a_{r'}(0)$ survive. For zero momentum the orbital part, $J^{12}_{\text{scalar}} = i\left(x^1\partial^2 - x^2\partial^1\right)$, acting on $e^{\pm ip\cdot x}$ with $\vec{p} = 0$, gives zero: a particle at rest has no orbital angular momentum. We are left with the spin part, which in the Weyl representation is $S^{12} = \frac{1}{2}\,\mathrm{diag}\,(\sigma_3, \sigma_3)$, i.e. $\frac{1}{2}\sigma_3$ acting on both two-component blocks of $u_r(0) = \sqrt{m}\,(\xi^r, \xi^r)^T$. The z-axis is not special. By choosing the basis $\xi^{(1)}, \xi^{(2)}$ of eigenvectors of σ_3 we have chosen the z-axis as the quantisation axis of the spin. Only one component of the angular momentum can be diagonal together with the Hamiltonian and the momentum, since rotations around different axes do not commute.

After some algebra, we find that

$$\begin{aligned} J^3\,|0, s, +\rangle \;&\equiv\; : J^{12} : |0, s, +\rangle = \left\{: J^{12} :, a_s^\dagger(0)\right\}|0\rangle \\ &=\; \sum_r \left(\xi^{r\dagger}\frac{\sigma_3}{2}\xi^s\right)|0, r, +\rangle \\ &=\; \xi^{(1)\dagger}\frac{\sigma_3}{2}\xi^s\,|0, (1), +\rangle + \xi^{(2)\dagger}\frac{\sigma_3}{2}\xi^s\,|0, (2), +\rangle\,, \end{aligned} \tag{10.52}$$

where we choose the basis

$$\xi^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \text{ and } \xi^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{10.53}$$

which diagonalises σ_3,

$$\sigma_3\xi^{(1)} = +\xi^{(1)}, \quad \sigma_3\xi^{(2)} = -\xi^{(2)}. \tag{10.54}$$

Then we have,

$$J^3\,|0, s, +\rangle = \left(\delta_{s,(1)}\frac{1}{2} - \delta_{s,(2)}\frac{1}{2}\right)|0, s, +\rangle\,. \tag{10.55}$$

We now have a physical interpretation of the quantum number s. A particle at rest with charge + and $s = (1)$ has total angular momentum $+\frac{1}{2}$ in the z-direction,

$$J^3 |0,(1),+\rangle = +\frac{1}{2} |0,(1),+\rangle . \tag{10.56}$$

A particle at rest with charge + and $s = (2)$ has total angular momentum $-\frac{1}{2}$ in the z-direction. We have then discovered that a general superposition $\sum_s c_s \left|p,s,+\right\rangle$ describes a particle with total spin-$\frac{1}{2}$. Repeating the same steps, we also conclude that a general anti-particle state $\sum_s c_s \left|p,s,-\right\rangle$ is also a state with spin-$\frac{1}{2}$.

In summary, the quantisation of the Dirac field yields quantum states with spin-$\frac{1}{2}$. Spin emerged simply as a component of the total angular momentum, due to the Dirac fields transforming in the spinorial representation of the Lorentz group. This theory also predicts the existence of both matter and anti-matter. Of course, the theory is highly successful. The electron and positron are states of the quantised Dirac field. The same is valid for other known spin-$\frac{1}{2}$ particles, the muon and tau leptons, the quarks and their antiparticles.

10.2 Fermions

Particles of the Dirac field are fermions. A state with two positively charged particles is (in the second line we use Eq. 10.14)

$$\begin{aligned} \left|p_1,s_1,+;p_2,s_2,+\right\rangle &\equiv a^\dagger_{s_1}(p_1)a^\dagger_{s_2}(p_2)|0\rangle \\ \rightsquigarrow \left|p_1,s_1,+;p_2,s_2,+\right\rangle &= \frac{1}{2}\left[a^\dagger_{s_1}(p_1)a^\dagger_{s_2}(p_2) - a^\dagger_{s_2}(p_2)a^\dagger_{s_1}(p_1)\right]|0\rangle . \end{aligned} \tag{10.57}$$

The state where all quantum numbers of the two particles are identical $p_1 = p_2$ and $s_1 = s_2$ is forbidden,

$$\left|p,s,+;p,s,+\right\rangle = \frac{1}{2}\left[a^\dagger_s(p)a^\dagger_s(p) - a^\dagger_s(p)a^\dagger_s(p)\right]|0\rangle = 0, \tag{10.58}$$

which is Pauli's exclusion principle for fermions. Using the anti-commutation of creation operators we can write a state of N identical particles or anti-particles as,

$$\left|p_1,s_1,\pm;p_2,s_2,\pm;\ldots;p_N,s_N,\pm\right\rangle = \frac{1}{N!}\sum_{\sigma\in S_N} \mathrm{sgn}(\sigma) a^\dagger_{s_{\sigma(1)}}(p_{\sigma(1)})a^\dagger_{s_{\sigma(2)}}(p_{\sigma(2)})\ldots a^\dagger_{s_{\sigma(N)}}(p_{\sigma(N)})|0\rangle , \tag{10.59}$$

which is manifestly anti-symmetric.

The quantisation of the Dirac field explains another mystery of quantum mechanics, namely that elementary particles with spin-$\frac{1}{2}$ are fermions. Notice the logic. We did not impose anti-commutation relations because we wanted fermions. We were forced to impose them in order to have a Hamiltonian which is bounded from below, and the exclusion principle came out as a consequence. For the scalar and the photon fields the creation operators commute and the particles are bosons. This is a special case of the spin-statistics theorem (Pauli, 1940): fields of integer spin must be quantised with commutators and describe bosons, while fields of half-integer spin must be quantised with anti-commutators and describe fermions.

10.3 Quantum symmetries

There is a powerful theorem, due to Wigner, which states that quantum representations of symmetry transformations are either unitary and linear

$$\langle U\psi|\, U\chi\rangle = \langle\psi|\,\chi\rangle\,, \quad U(a_1\,|\psi_1\rangle + a_2\,|\psi_2\rangle) = a_1 U\,|\psi_1\rangle + a_2 U\,|\psi_2\rangle\,, \tag{10.60}$$

or anti-unitary

$$\langle U\psi|\, U\chi\rangle = \langle\chi|\,\psi\rangle\,, \quad U(a_1\,|\psi_1\rangle + a_2\,|\psi_2\rangle) = a_1^* U\,|\psi_1\rangle + a_2^* U\,|\psi_2\rangle\,. \tag{10.61}$$

In this way, the probability for a transition from a state $|\phi\rangle$ to a state $|\psi\rangle$ is preserved under a symmetry transformation. For example, for an antiunitary transformation:

$$P(|\phi\rangle \to |\psi\rangle) = |\langle\psi|\,\phi\rangle|^2 \to |\langle U\psi|\, U\phi\rangle|^2 = |\langle\phi|\,\psi\rangle|^2 = P(|\phi\rangle \to |\psi\rangle). \tag{10.62}$$

The identity transformation $U = 1$ is unitary. All transformations which are connected to the identity with a continuous parameter (such as Lorentz and gauge symmetry transformations) are therefore unitary. Antilinear transformations concern discrete symmetries and in particular the ones which involve time reversal.

Requiring that the normalisation of a state is invariant under a symmetry transformation, we obtain that

$$\big||\psi\rangle\big|^2 = \big|U\,|\psi\rangle\big|^2 \rightsquigarrow U^\dagger = U^{-1}. \tag{10.63}$$

Measurable quantities are obtained from matrix elements of the type

$$\langle \mathrm{STATE}_1|\,\hat{O}\,|\mathrm{STATE}_2\rangle\,. \tag{10.64}$$

For them to be invariant under a symmetry transformation, we must have that if

$$|\mathrm{STATE}\rangle \to |\mathrm{STATE}'\rangle = U\,|\mathrm{STATE}\rangle\,,$$

then

$$\hat{O} \to \hat{O}' = U\hat{O}U^{-1}. \tag{10.65}$$

10.4 Lorentz transformation of the quantised spinor field

A spinor field transforms classically as

$$\psi(x) \to \psi'(x') = \Lambda_{\frac{1}{2}}\psi(x). \tag{10.66}$$

What is the transformation of the corresponding quantum field and the quantum states? It does not have to be the same as for the classical field, and in fact it is not. Relativistic invariance at the classical level is cast as the requirement that the Lagrangian is a scalar,

$$\mathcal{L}'(x') = \mathcal{L}(x).$$

This is sufficient to guarantee that observers in different reference frames find the same minimum value for the action. However, this is not the same criterion that we must introduce for Lorentz invariance of the quantum system.

We consider a quantum one-particle state, e.g.

$$|p, s, +\rangle\,.$$

Under a Lorentz transformation Λ, a state transforms as

$$\left|p,s,+\right\rangle \rightarrow U(\Lambda)\left|p,s,+\right\rangle . \tag{10.67}$$

What is the new state, i.e. what is the transformation representation $U(\Lambda)$? It is logical to require that the action of $U(\Lambda)$ yields a particle state with a momentum $\Lambda p \equiv \Lambda^{\mu}_{\nu} p^{\nu}$,

$$\left|p,s,+\right\rangle \rightarrow U(\Lambda)\left|p,s,+\right\rangle = N(\Lambda,p)\left|\Lambda p,s,+\right\rangle . \tag{10.68}$$

We express Eq. 10.68 in terms of creation operators,

$$\begin{aligned} U(\Lambda)\left|p,s,+\right\rangle &= N(\Lambda,p)\left|\Lambda p,s,+\right\rangle \\ \rightsquigarrow U(\Lambda)a_s^{\dagger}(p)\left|0\right\rangle &= N(\Lambda,p)a_s^{\dagger}(\Lambda p)U(\Lambda)\left|0\right\rangle , \end{aligned} \tag{10.69}$$

where $N(\Lambda,p)$ is a normalisation factor that we need to determine. Conjugating Eq. 10.69 we find that,

$$\left\langle 0\right| U(\Lambda)a_s(p)U^{-1}(\Lambda) = N(\Lambda,p)^* \left\langle 0\right| a_s(\Lambda p). \tag{10.70}$$

Particle states are orthogonal,

$$\begin{aligned} \left\langle p_1,s_1,+\right| p_2,s_2,+\right\rangle &= \left\langle 0\right| a_{s_1}(p_1)a_{s_2}^{\dagger}(p_2)\left|0\right\rangle \\ &= \left\langle 0\right| \left\{a_{s_1}(p_1),a_{s_2}^{\dagger}(p_2)\right\}\left|0\right\rangle - \left\langle 0\right| a_{s_2}^{\dagger}(p_2)a_{s_1}(p_1)\left|0\right\rangle \\ &= (2\pi)^3 2\omega_{p_1}\delta^{(3)}(p_2-p_1)\delta_{s_1s_2}\left\langle 0\right| 0\right\rangle - 0 \\ \\ \rightsquigarrow \left\langle p_1,s_1,+\right| p_2,s_2,+\right\rangle &= (2\pi)^3 2\omega_{p_1}\delta^{(3)}(p_2-p_1)\delta_{s_1s_2}. \end{aligned} \tag{10.71}$$

Two different particles in a reference frame should remain distinct in any other reference frame. If we perform a Lorentz transformation, Eq. 10.68, we have

$$\begin{aligned} \left\langle p_1,s_1,+\right| U^{-1}U\left|p_2,s_2,+\right\rangle &= \left\langle 0\right| a_{s_1}(p_1)U^{-1}Ua_{s_2}^{\dagger}(p_2)\left|0\right\rangle \\ &= \left\langle 0\right| Ua_{s_1}(p_1)U^{-1}Ua_{s_2}^{\dagger}(p_2)U\left|0\right\rangle \\ &= \left|N(\Lambda,p_1)\right|^2 (2\pi)^3 2\omega_{\Lambda p_1}\delta^{(3)}(\Lambda p_2-\Lambda p_1)\delta_{s_1s_2}. \end{aligned} \tag{10.72}$$

We can now prove that

$$2\omega_{\Lambda p_1}\delta^{(3)}(\Lambda p_1-\Lambda p_2) = 2\omega_{p_1}\delta^{(3)}(p_1-p_2). \tag{10.73}$$

Proof: Consider a scalar function $f(\vec{p})$. Then,

$$f(\vec{p}\,') = \int \frac{d^3\vec{p}}{2\omega_p} 2\omega_p \delta^{(3)}(\vec{p}-\vec{p}\,')f(\vec{p}). \tag{10.74}$$

The measure

$$\int \frac{d^3\vec{p}}{2\omega_p} = \int d^4p\,\delta(p^2-m^2)\Theta\left(p^0>0\right)$$

is invariant under Lorentz transformations. Given that f is a scalar, we must have that

$$\omega_p\delta^{(3)}(\vec{p}-\vec{p}\,')$$

is also a scalar, which proves Eq. 10.73. The orthogonality of the states is therefore preserved under Lorentz transformations. Eq. 10.72 becomes

$$\left\langle p_1,s_1,+\right| U^{-1}U\left|p_2,s_2,+\right\rangle = \left|N(\Lambda,p_1)\right|^2 (2\pi)^3 2\omega_{p_1}\delta^{(3)}(p_2-p_1)\delta_{s_1s_2}, \tag{10.75}$$

which, for $N(\Lambda,p)=1$, guarantees that

$$\left\langle p_1,s_1,+\right| U^{-1}U\left|p_2,s_2,+\right\rangle = \left\langle p_1,s_1,+\right| p_2,s_2,+\right\rangle . \tag{10.76}$$

10.4.1 Transformation of the quantised Dirac field

We now proceed heuristically in order to derive the transformation properties of the Dirac field. A more rigorous and detailed exposition can be found in Weinberg Vol. I (chapters 2 and 5). Motivated by Eq. 10.69, we infer that the transformation of the creation and annihilation operators is:

$$U(\Lambda)a_s^\dagger(p)U^{-1}(\Lambda) = a_s^\dagger(\Lambda p), \tag{10.77}$$
$$U(\Lambda)a_s(p)U^{-1}(\Lambda) = a_s(\Lambda p), \tag{10.78}$$
$$U(\Lambda)b_s^\dagger(p)U^{-1}(\Lambda) = b_s^\dagger(\Lambda p), \tag{10.79}$$
$$U(\Lambda)b_s(p)U^{-1}(\Lambda) = b_s(\Lambda p). \tag{10.80}$$

What is then the transformed quantum field $\psi'(x')$,

$$\psi(x) \to \psi'(x') = U(\Lambda)\psi(x)U^{-1}(\Lambda)? \tag{10.81}$$

We can compute the right-hand side, knowing the transformation of the creation and annihilation operators from Eq. 10.77,

$$\begin{aligned} U\psi(x)U^{-1} &= \int \frac{d^3k}{(2\pi)^3 2\omega_k} \sum_s U\left[a_s(k)u(k)e^{-ik\cdot x} + b_s^\dagger(k)v(k)e^{+ik\cdot x}\right]U^{-1} \\ &= \int \frac{d^3k}{(2\pi)^3 2\omega_k} \sum_s \left[a_s(\Lambda k)u(k)e^{-ik\cdot x} + b_s^\dagger(\Lambda k)v(k)e^{+ik\cdot x}\right]. \end{aligned} \tag{10.82}$$

Now we use that

$$\frac{d^3k}{(2\pi)^3 2\omega_k} = \frac{d^3\Lambda k}{(2\pi)^3 2\omega_{\Lambda k}}, \quad k\cdot x = \Lambda k \cdot \Lambda x, \tag{10.83}$$

and

$$u(k) = \Lambda_{\frac{1}{2}}^{-1}u(\Lambda k), \quad \text{and } v(k) = \Lambda_{\frac{1}{2}}^{-1}v(\Lambda k), \tag{10.84}$$

which yields, setting $p = \Lambda k$,

$$\begin{aligned} U\psi(x)U^{-1} &= \Lambda_{\frac{1}{2}}^{-1}\int \frac{d^3p}{(2\pi)^3 2\omega_p} \sum_s \left[a_s(p)u(p)e^{-ip\cdot\Lambda x} + b_s^\dagger(p)v(p)e^{+ip\cdot\Lambda x}\right] \\ &= \Lambda_{\frac{1}{2}}^{-1}\psi(\Lambda x). \end{aligned} \tag{10.85}$$

We have found the transformation of the quantum Dirac field. It reads,

$$\psi(x) \to \psi'(x') = U(\Lambda)\psi(x)U^{-1}(\Lambda) = \Lambda_{\frac{1}{2}}^{-1}\psi(\Lambda x). \tag{10.86}$$

This is different from the representation of the classical field: $\psi'(x') = \Lambda_{\frac{1}{2}}\psi(x)$. Therefore, the classical field and the quantum operator have inverse Lorentz transformations. *Remark.* The appearance of the inverse is, however, a matter of bookkeeping, not of physics. States transform with U and operators with $U\hat{O}U^{-1}$ (active transformation of the states), and the result is equivalent to $U^{-1}(\Lambda)\psi(x)U(\Lambda) = \Lambda_{\frac{1}{2}}\psi(\Lambda^{-1}x)$, which is exactly the classical law $\psi'(x) = \Lambda_{\frac{1}{2}}\psi(\Lambda^{-1}x)$; the quantum field transforms as the classical field does (Peskin-Schroeder Sec. 3.5, Weinberg I Sec. 5.1). Moreover, the relations $Ua_s(p)U^{-1} = a_s(\Lambda p)$ and $u_s(k) = \Lambda_{\frac{1}{2}}^{-1}u_s(\Lambda k)$ hold with the same label s only for rotation-free (pure boost) bookkeeping; in general the spin labels are mixed by the Wigner rotation of the little group, $\Lambda_{\frac{1}{2}}u_s(k) = \sum_{s'} u_{s'}(\Lambda k)D_{s's}(W)$, and the same matrix appears in Ua_sU^{-1}, so that the final result is unchanged (Weinberg I, Secs. 2.5, 5.5).

10.5 Parity

We now derive the transformation properties of a spinor field under parity. This is a discrete transformation,

$$(t, \vec{x}) \to (t, -\vec{x}). \tag{10.87}$$

Under parity, the momentum of a particle transforms as,

$$\vec{p} \to -\vec{p}. \tag{10.88}$$

The representation of the parity transformation on a one-particle state should be, for example,

$$U_{\mathcal{P}} \left|\vec{p}, s, +\right\rangle = \lambda \left|-\vec{p}, s, +\right\rangle, \tag{10.89}$$

where λ is such that

$$1 = \left\langle \vec{p}, s, +\right| U_{\mathcal{P}}^{\dagger} U_{\mathcal{P}} \left|\vec{p}, s, +\right\rangle = |\lambda|^2 \left\langle -\vec{p}, s, +\right| \left.-\vec{p}, s, +\right\rangle \rightsquigarrow |\lambda|^2 = 1. \tag{10.90}$$

Notice that the spin label s is not changed. Parity reverses the momentum, but the spin is an angular momentum, which behaves like $\vec{L} = \vec{x} \times \vec{p}$; under $\vec{x} \to -\vec{x}$ and $\vec{p} \to -\vec{p}$ it remains the same. As a consequence, parity flips the helicity of a particle. The transformation of the states yields the following transformation for the creation and annihilation operators,

$$U_{\mathcal{P}} a_s(\vec{p}) U_{\mathcal{P}}^{-1} = \lambda_a a_s(-\vec{p}), \tag{10.91}$$

and

$$U_{\mathcal{P}} b_s(\vec{p}) U_{\mathcal{P}}^{-1} = \lambda_b b_s(-\vec{p}). \tag{10.92}$$

Under a parity transformation, a Dirac quantum field transforms as

$$\begin{aligned} U_{\mathcal{P}} \psi(\vec{x}, t) U_{\mathcal{P}}^{-1} &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \sum_s U_{\mathcal{P}} \left[a_s(\vec{k}) u(\vec{k}) e^{-ik\cdot x} + b_s^{\dagger}(\vec{k}) v(\vec{k}) e^{+ik\cdot x} \right] U_{\mathcal{P}}^{-1} \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \sum_s \left[\lambda_a a_s(-\vec{k}) u(\vec{k}) e^{-ik\cdot x} + \lambda_b^* b_s^{\dagger}(-\vec{k}) v(\vec{k}) e^{+ik\cdot x} \right] \end{aligned} \tag{10.93}$$

Now we use that the measure is invariant under a parity transformation,

$$\frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} = \frac{d^3(-\vec{k})}{(2\pi)^3 2\omega_{-k}}, \tag{10.94}$$

and we rewrite,

$$k \cdot x = Et - \vec{k}\vec{x} = Et - (-\vec{k})(-\vec{x}). \tag{10.95}$$

Eq. 10.93 becomes,

$$\begin{aligned} &U_{\mathcal{P}} \psi(\vec{x}, t) U_{\mathcal{P}}^{-1} = \\ &\int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \sum_s \left[\lambda_a a_s(\vec{p}) u(-\vec{p}) e^{-i\left(Et - \vec{p}\cdot(-x)\right)} + \lambda_b^* b_s^{\dagger}(\vec{p}) v(-\vec{p}) e^{+i\left(Et - \vec{p}\cdot(-x)\right)} \right], \end{aligned} \tag{10.96}$$

where we set $\vec{p} = -\vec{k}$. Recall now that, in the Weyl representation,

$$u_s(p) = \begin{pmatrix} \sqrt{E - \vec{p}\vec{\sigma}}\, \xi^s \\ \sqrt{E + \vec{p}\vec{\sigma}}\, \xi^s \end{pmatrix} \text{ and } \gamma^0 = \begin{pmatrix} 0 & I_{2\times 2} \\ I_{2\times 2} & 0 \end{pmatrix}. \tag{10.97}$$

We then have

$$\gamma^0 u_s(\vec{p}) = u_s(-\vec{p}), \tag{10.98}$$

and similarly,

$$\gamma^0 v_s(\vec{p}) = -v_s(-\vec{p}). \tag{10.99}$$

We can now fix the arbitrary normalisation phases λ_a, λ_b to

$$\lambda_a = -\lambda_b^* = 1. \tag{10.100}$$

Then, Eq. 10.96 yields a simple parity transformation for the field,

$$U_{\mathcal{P}} \psi(\vec{x}, t) U_{\mathcal{P}}^{-1} = \gamma^0 \psi(-\vec{x}, t). \tag{10.101}$$

It is easy to show that we can take $U_{\mathcal{P}}$ to be its own inverse. We have,

$$\begin{aligned} U_{\mathcal{P}} \psi(-\vec{x}, t) U_{\mathcal{P}}^{-1} &= \gamma^0 \psi(\vec{x}, t) \\ \rightsquigarrow \psi(-\vec{x}, t) &= \gamma^0 U_{\mathcal{P}}^{-1} \psi(\vec{x}, t) U_{\mathcal{P}} \\ \rightsquigarrow \gamma^0 \psi(-\vec{x}, t) &= (\gamma^0)^2 U_{\mathcal{P}}^{-1} \psi(\vec{x}, t) U_{\mathcal{P}} \\ \\ \rightsquigarrow U_{\mathcal{P}} \psi(\vec{x}, t) U_{\mathcal{P}}^{-1} &= U_{\mathcal{P}}^{-1} \psi(\vec{x}, t) U_{\mathcal{P}}, \end{aligned} \tag{10.102}$$

which is satisfied for

$$U_{\mathcal{P}} = U_{\mathcal{P}}^{-1}. \tag{10.103}$$

We can study the transformation of composite operators made up of two Dirac fields. The combination $\bar{\psi}\psi$ which transforms as a scalar under Lorentz transformations, transforms under parity as a scalar too,

$$\begin{aligned} \bar{\psi}(t, \vec{x}) \psi(t, \vec{x}) \quad &\rightarrow \quad \bar{\psi}'(t, -\vec{x}) \psi'(t, -\vec{x}) = U_{\mathcal{P}} \bar{\psi}(t, \vec{x}) U_{\mathcal{P}} U_{\mathcal{P}} \psi(t, \vec{x}) U_{\mathcal{P}} \\ &= \quad \bar{\psi}(t, -\vec{x}) \gamma^0 \gamma^0 \psi(t, -\vec{x}) \\ &= \quad \bar{\psi}(t, -\vec{x}) \psi(t, -\vec{x}). \end{aligned} \tag{10.104}$$

However, there is a combination $\bar{\psi}\gamma^5\psi$ which transforms as a scalar under Lorentz transformations (**exercise: prove that it is indeed so**), but transforms differently under parity,

$$\begin{aligned} \bar{\psi}(t, \vec{x}) \gamma^5 \psi(t, \vec{x}) \quad &\rightarrow \quad \bar{\psi}'(t, -\vec{x}) \gamma^5 \psi'(t, -\vec{x}) = U_{\mathcal{P}} \bar{\psi}(t, \vec{x}) U_{\mathcal{P}} \gamma^5 U_{\mathcal{P}} \psi(t, \vec{x}) U_{\mathcal{P}} \\ &= \quad \bar{\psi}(t, -\vec{x}) \gamma^0 \gamma^5 \gamma^0 \psi(t, -\vec{x}) \\ &= \quad -\bar{\psi}(t, -\vec{x}) \gamma^5 \gamma^0 \gamma^0 \psi(t, -\vec{x}) \\ &= \quad -\bar{\psi}(t, -\vec{x}) \gamma^5 \psi(t, -\vec{x}). \end{aligned} \tag{10.105}$$

This is called a "pseudoscalar".

The combination $\bar{\psi}\gamma^\mu\psi$ which transforms as a vector under Lorentz transformations (**exercise**), also transforms as a vector under parity,

$$\begin{aligned} \bar{\psi}(t, \vec{x}) \gamma^\mu \psi(t, \vec{x}) \quad &\rightarrow \quad \bar{\psi}'(t, -\vec{x}) \gamma^\mu \psi'(t, -\vec{x}) = U_{\mathcal{P}} \bar{\psi}(t, \vec{x}) U_{\mathcal{P}} \gamma^\mu U_{\mathcal{P}} \psi(t, \vec{x}) U_{\mathcal{P}} \\ &= \quad \bar{\psi}(t, -\vec{x}) \gamma^0 \gamma^\mu \gamma^0 \psi(t, -\vec{x}) \\ &= \quad \bar{\psi}(t, -\vec{x}) \left(2 g^{0\mu} \gamma^0 - \gamma^\mu (\gamma^0)^2\right) \psi(t, -\vec{x}) \\ &= \quad \left(2 g^{0\mu} - 1\right) \bar{\psi}(t, -\vec{x}) \gamma^\mu \psi(t, -\vec{x}). \end{aligned} \tag{10.106}$$

For $\mu = 0$,

$$\bar{\psi}(t,\vec{x})\gamma^0\psi(t,\vec{x}) \to \bar{\psi}(t,-\vec{x})\gamma^0\psi(t,-\vec{x}). \tag{10.107}$$

For $\mu = i = 1,2,3$,

$$\bar{\psi}(t,\vec{x})\gamma^i\psi(t,\vec{x}) \to -\bar{\psi}(t,-\vec{x})\gamma^i\psi(t,-\vec{x}). \tag{10.108}$$

This is exactly how a vector $x^\mu = (t,\vec{x}) \to (t,-\vec{x})$ transforms under parity.

The combination $\bar{\psi}\gamma^\mu\gamma^5\psi$ which transforms as a vector under Lorentz transformations (**exercise: prove that it is indeed so**), transforms oppositely under parity,

$$\begin{aligned}
\bar{\psi}(t,\vec{x})\gamma^\mu\gamma^5\psi(t,\vec{x}) &\to \bar{\psi}'(t,-\vec{x})\gamma^\mu\gamma^5\psi'(t,-\vec{x}) = U_{\mathcal{P}}\bar{\psi}(t,\vec{x})U_{\mathcal{P}}\gamma^\mu\gamma^5U_{\mathcal{P}}\psi(t,\vec{x})U_{\mathcal{P}} \\
&= \bar{\psi}(t,-\vec{x})\gamma^0\gamma^\mu\gamma^5\gamma^0\psi(t,-\vec{x}) \\
&= -\bar{\psi}(t,-\vec{x})\gamma^0\gamma^\mu\gamma^0\gamma^5\psi(t,-\vec{x}) \\
&= -\bar{\psi}(t,-\vec{x})\left(2g^{0\mu}\gamma^0 - \gamma^\mu(\gamma^0)^2\right)\gamma^5\psi(t,-\vec{x}) \\
&= -\left(2g^{0\mu}-1\right)\bar{\psi}(t,-\vec{x})\gamma^\mu\gamma^5\psi(t,-\vec{x}).
\end{aligned} \tag{10.109}$$

This is called a "pseudovector" or axial vector.

The Dirac Lagrangian,

$$i\bar{\psi}\partial\!\!\!/\psi - m\bar{\psi}\psi,$$

is composed of true scalar terms, and it transforms as a scalar under parity.

Exercise: Prove that the classical Dirac field has the same parity transformation as the quantum Dirac field.

10.6 Other discrete symmetries

Exercise: Prove the following:

1. **Under time-reversal** $(t,\vec{x}) \to (-t,\vec{x})$**,**

$$U_{\mathcal{T}}a_s(\vec{p})U_{\mathcal{T}} = a_{-s}(-\vec{p}), \qquad \xi^{-s} = -i\sigma_2(\xi^s)^*, \tag{10.110}$$

and

$$U_{\mathcal{T}}\psi(t,\vec{x})U_{\mathcal{T}} = \left(-\gamma^1\gamma^3\right)\psi(-t,\vec{x}). \tag{10.111}$$

2. **Under charge conjugation, exchanging a particle with its antiparticle,**

$$U_C a_s(\vec{p})U_C = b_s(\vec{p}), \qquad U_C b_s(\vec{p})U_C = a_s(\vec{p}), \tag{10.112}$$

and

$$U_C\psi(t,\vec{x})U_C = -i(\bar{\psi}(t,\vec{x})\gamma^0\gamma^2)^T. \tag{10.113}$$

11

Propagation of free particles

In previous chapters we studied the particle states from the quantisation of free fields, such as the Schrödinger, the Klein-Gordon, the spinor, and the electromagnetic field. We are now ready to study the simplest transition amplitude, for the transition of a free particle from a space-time position $y^\mu \equiv (y^0, \vec{y})$ to a position $x^\mu \equiv (x^0, \vec{x})$, with $x^0 > y^0$. The transition amplitude is given by,

$$M_{y\to x} = \langle x|\, y\rangle\big|_{x^0>y^0}\,, \tag{11.1}$$

where $|x\rangle\,, \big|y\big\rangle$ are states of a single particle at positions x, y, respectively.

Free particles cannot do much; the best they can do is to go from one point to another. Nevertheless, these elementary transitions will be the building blocks for everything that follows. Interactions will happen at points in space-time, and in between the particles will propagate as free particles. The question we ask here is pure quantum mechanics: if a particle is at y, what is the probability amplitude that a detector placed at x finds it there? What quantum field theory provides is the position states themselves. A field contains a creation operator which, acting on the vacuum, creates a state of definite momentum; the field superposes these states with plane waves, i.e. it Fourier transforms them, and the Fourier transform of momentum eigenstates is a position eigenstate. In the canonical quantisation that we follow, we derive the transition amplitudes from the fields. In the path-integral quantisation, which we discuss in Quantum Field Theory II, one can take the opposite route and consider the transition amplitudes as the fundamental objects from which everything else is derived.

11.1 Transition amplitude for the Schrödinger field

We shall compute the transition amplitude for the case of the Schrödinger field. This will give us an opportunity to compare with other theories, which are relativistic, and appreciate the merits of the latter. We recall that the Schrödinger field is given by,

$$\psi(x^0, \vec{x}) = \int \frac{d^3\vec{k}}{(2\pi)^3} a(k) e^{-i\left(E_k x^0 - \vec{k}\cdot\vec{x}\right)}, \tag{11.2}$$

where

$$E_k = \frac{\vec{k}^2}{2m}, \tag{11.3}$$

and the operator $a(k)$ satisfies the quantisation condition,

$$\left[a(k), a^\dagger(k')\right] = (2\pi)^3 \delta^{(3)}(\vec{k} - \vec{k}'). \tag{11.4}$$

A state of a particle at a position x is given by the action of a field on the vacuum (see Section 4.6),

$$|x\rangle = \psi^\dagger(x^0, \vec{x})\,|0\rangle\,, \qquad \big|y\big\rangle = \psi^\dagger(y^0, \vec{y})\,|0\rangle\,. \tag{11.5}$$

The transition amplitude of Eq. 11.1 is then,

$$\begin{aligned} M_{y\to x} &= \langle 0|\,\psi(x^0, \vec{x})\psi^\dagger(y^0, \vec{y})\,|0\rangle \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3} e^{-i\left(E_k\Delta t - \vec{k}\cdot\Delta\vec{x}\right)}, \end{aligned} \tag{11.6}$$

with

$$\Delta t = x^0 - y^0, \qquad \Delta\vec{x} = \vec{x} - \vec{y}. \tag{11.7}$$

We can compute the above integral exactly. We first change to spherical coordinates,

$$d^3\vec{k} = dk\, k^2\, d\cos\theta\, d\phi, \tag{11.8}$$

and write

$$\vec{k}\cdot\Delta\vec{x} = k\,\Delta x\,\cos\theta. \tag{11.9}$$

Then it is easy to perform successively the integrations in ϕ, $\cos\theta$ and k. The result reads,

$$M_{y\to x} = \left(\frac{m}{2\pi i\Delta t}\right)^{\frac{3}{2}} e^{+i\Delta t\left[\frac{1}{2}m\left(\frac{\Delta x}{\Delta t}\right)^2\right]}. \tag{11.10}$$

The phase has a beautiful interpretation. For a free particle the Lagrangian is the kinetic energy, $L = \frac{1}{2}mv^2$, and for a particle which moves uniformly from y to x, with $v = \Delta x/\Delta t$, the classical action is $S_{\rm cl} = \Delta t\,\frac{1}{2}m\,(\Delta x/\Delta t)^2$. The amplitude is therefore $\propto e^{iS_{\rm cl}}$. This observation inspired Feynman: he composed large transitions out of many small ones and arrived at a new formulation of quantum mechanics, in terms of a sum over all paths weighted with e^{iS}, the path integral.

Exercise: *Derive the same amplitude in ordinary quantum mechanics, as* $\big\langle\vec{x}\big|\,e^{-iH\Delta t}\,\big|\vec{y}\big\rangle$ *with* $H = \vec{p}^{\,2}/(2m)$. The amplitude is an oscillatory function which is defined for arbitrary values of the particle speed $\frac{\Delta x}{\Delta t}$, even when it exceeds the speed of light. Particles which are quanta of the Schrödinger field can propagate to space-time points anywhere outside the light-cone, which is in blatant disagreement with special relativity.

11.2 Transition amplitude for the real Klein-Gordon field

We write the real Klein-Gordon field as a sum of a "positive frequency" and a "negative frequency" term,

$$\phi(x) = \phi^{(+)}(x) + \phi^{(-)}(x), \tag{11.11}$$

with

$$\phi^{(+)}(x) = \int \frac{d^4k}{(2\pi)^3}\delta(k^2 - m^2)\Theta(k^0)e^{-ik\cdot x}a(k), \tag{11.12}$$

and

$$\phi^{(-)}(x) = \int \frac{d^4k}{(2\pi)^3}\delta(k^2 - m^2)\Theta(k^0)e^{+ik\cdot x}a^\dagger(k). \tag{11.13}$$

We notice that,

$$\phi^{(+)}(x)\,|0\rangle = 0, \tag{11.14}$$

and

$$\langle 0|\, \phi^{(-)}(x) = 0. \tag{11.15}$$

The operator $a(k)$ satisfies the quantisation condition,

$$\left[a(k), a^{\dagger}(k')\right] = (2\pi)^3 2\omega_k \delta^{(3)}(\vec{k} - \vec{k}'), \tag{11.16}$$

and

$$\omega_k = \sqrt{\vec{k}^2 + m^2}. \tag{11.17}$$

The amplitude for the transition $y^\mu \to x^\mu$, with $x^0 > y^0$, is then

$$\begin{aligned}
M_{y\to x} &= \langle 0|\, \phi^{\dagger}(x)\phi(y)\, |0\rangle \\
&= \langle 0|\, \phi(x)\phi(y)\, |0\rangle \\
&= \langle 0|\, \left(\phi^{(+)}(x) + \phi^{(-)}(x)\right)\left(\phi^{(+)}(y) + \phi^{(-)}(y)\right) |0\rangle \\
&= \langle 0|\, \phi^{(+)}(x)\phi^{(+)}(y) + \phi^{(+)}(x)\phi^{(-)}(y) + \phi^{(-)}(x)\phi^{(+)}(y) + \phi^{(-)}(x)\phi^{(-)}(y)\, |0\rangle \\
&= \langle 0|\, \phi^{(+)}(x)\phi^{(-)}(y)\, |0\rangle \\
&= \langle 0|\, \left[\phi^{(+)}(x), \phi^{(-)}(y)\right] + \phi^{(-)}(x)\phi^{(+)}(y)\, |0\rangle \\
&= \langle 0|\, \left[\phi^{(+)}(x), \phi^{(-)}(y)\right] |0\rangle\, .
\end{aligned} \tag{11.18}$$

The commutator gives,

$$\left[\phi^{(+)}(x), \phi^{(-)}(y)\right] = \int \frac{d^4k}{(2\pi)^3} \Theta(k^0) e^{-ik(x-y)} \delta(k^2 - m^2), \tag{11.19}$$

which is a c-number. We then have for the transition amplitude,

$$M_{y\to x} = \langle 0|\, \left[\phi^{(+)}(x), \phi^{(-)}(y)\right] |0\rangle = \left[\phi^{(+)}(x), \phi^{(-)}(y)\right] \langle 0|\, 0\rangle = \left[\phi^{(+)}(x), \phi^{(-)}(y)\right], \tag{11.20}$$

and explicitly,

$$M_{y\to x} = \int \frac{d^4k}{(2\pi)^3} \Theta(k^0) e^{-ik(x-y)} \delta(k^2 - m^2). \tag{11.21}$$

We now integrate out the energy component k^0 and use spherical coordinates for $\vec{k}$, yielding:

$$M_{y\to x} = \int_0^\infty \frac{dk}{(2\pi)^2} \int_{-1}^{1} d\cos\theta \frac{k^2}{2\sqrt{k^2 + m^2}} e^{-i\sqrt{k^2+m^2}\Delta t + ik\Delta x \cos\theta}, \tag{11.22}$$

with $\Delta t = x^0 - y^0$ and $\Delta x = |\vec{x} - \vec{y}|$. We can also perform the integration over the angle θ. This yields:

$$\begin{aligned}
M_{y\to x} &= \int_0^\infty \frac{dk}{(2\pi)^2} \frac{k^2}{2\sqrt{k^2 + m^2}} e^{-i\sqrt{k^2+m^2}\Delta t} \frac{e^{+ik\Delta x} - e^{-ik\Delta x}}{ik\Delta x} \\
&= -\frac{i}{2(2\pi)^2 \Delta x} \int_{-\infty}^{\infty} dk \frac{k}{\sqrt{k^2 + m^2}} e^{-i\sqrt{k^2+m^2}\Delta t + ik\Delta x}.
\end{aligned} \tag{11.23}$$

Notice how the integration over the magnitude $k \geq 0$ became an integration over the whole real axis. The angular integration produced two terms, with $e^{+ik\Delta x}$ and $e^{-ik\Delta x}$. Changing $k \to -k$ in the second term turns it into the first one with $-\infty < k \leq 0$, and the two terms combine into a single integral.

Exercise: *Evaluate the integral of Eq. 11.23 numerically, for a grid of values of Δt and Δx. You will need a method suited to oscillating integrands. Verify that the amplitude oscillates for time-like transitions and decreases rapidly for space-like transitions, the faster the further away from the light-cone.*

We would like to investigate whether the above expression for the transition amplitude respects the expectation from special relativity, namely that particles are restricted to transitions within the light-cone. It is now a bit more tedious to evaluate the integral of Eq. 11.23. For simplicity, we shall evaluate it for two special transitions,

- **Transition A:** where we take $\Delta x > 0$ and $\Delta t = 0$, which is a space-like transition (outside the light-cone).
- **Transition B:** where we take $\Delta x = 0$ and $\Delta t > 0$, which is a time-like transition (within the light-cone).

Space-like transition A

In this case, the transition amplitude becomes

$$M_{y\to x} = \frac{-i}{2(2\pi)^2(\Delta x)} \int_{-\infty}^{+\infty} dk \frac{k e^{ik(\Delta x)}}{\sqrt{k^2+m^2}}. \tag{11.24}$$

The above integral has branch cuts at $k = \pm im$, as in Figure 11.1. For $(\Delta x) > 0$ we

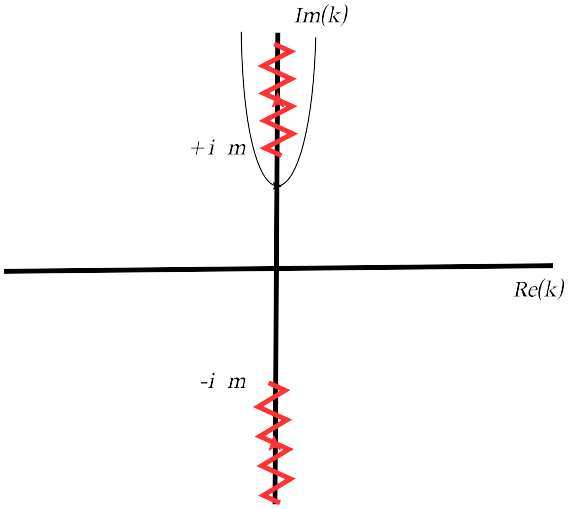


Figure 11.1. Branch cuts of the transition amplitude integral for a space-like transition.

can wrap the contour of integration around the upper branch cut. Why the upper and not the lower one? Deforming the contour is allowed only if no singularity is crossed and if the contributions far away from the real axis vanish. For $k = i\rho$ with $\rho > 0$, the exponential becomes $e^{ik\Delta x} = e^{-\rho\Delta x}$, which is suppressed; in the lower half-plane it would grow exponentially. There is a subtlety which we should not hide. The factor $k/\sqrt{k^2+m^2}$ tends to one as $|k| \to \infty$ and does not help the convergence. The non-decreasing piece amounts to $\int_{-\infty}^{\infty} dk\, e^{ik\Delta x} = 2\pi\delta(\Delta x)$, which vanishes for $\Delta x > 0$; for space-like separations our manipulation is therefore legitimate. However, it signals that the amplitude is singular when the two points coincide. Indeed, for $x = y$ we find $\langle 0|\,\phi(x)\phi(x)\,|0\rangle = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}$, which diverges at large $|\vec{k}|$. Large momenta probe short distances, and quantum field

theory gives infinite answers when we ask about the behaviour at a single point. We will meet such "ultraviolet" infinities again in loop integrals, and we will learn to treat them with regularisation (e.g. computing in $d < 4$ space-time dimensions) and renormalisation.

The integral then becomes, after setting $k = i\rho$,

$$M_{y\to x} = \frac{1}{4\pi^2(\Delta x)} \int_m^\infty d\rho \frac{\rho e^{-\rho(\Delta x)}}{\sqrt{\rho^2 - m^2}}, \tag{11.25}$$

or equivalently,

$$\begin{aligned} M_{y\to x} &= \frac{1}{4\pi^2(\Delta x)} \int_m^\infty d\rho \frac{e^{-\rho(\Delta x)}}{\sqrt{1 - \frac{m^2}{\rho^2}}} \\ &= \frac{1}{4\pi^2(\Delta x)} \int_m^\infty d\rho e^{-\rho(\Delta x)} \sum_{n=0}^\infty \frac{\left(\frac{1}{2}, n\right)}{n!} m^{2n} \rho^{-2n}. \end{aligned} \tag{11.26}$$

We can perform the ρ integration in terms of the incomplete Gamma function,

$$\Gamma(a, z) = \int_z^\infty dt e^{-t} t^{a-1}, \tag{11.27}$$

which, as $z \to \infty$, has the asymptotic behaviour:

$$\Gamma(a, z) \sim z^{a-1} e^{-z}. \tag{11.28}$$

Asymptotically, for a large space-like interval $\Delta x \to \infty$, the transition integral becomes

$$\lim_{(\Delta x)\to\infty} M_{y\to x} \sim e^{-m(\Delta x)}. \tag{11.29}$$

We find that the transition amplitude vanishes exponentially the further away we move from the light-cone ($\Delta x = 0$). Is this a contradiction with special relativity? It is not, but the reason deserves to be stated carefully. The amplitude $\langle 0| \phi(x)\phi(y) |0\rangle$ is not an observable. What relativity forbids is that a measurement performed at y influences the outcome of a measurement performed at x when the two points are separated by a space-like interval, and in a quantum theory this is guaranteed if the corresponding operators commute. Let us therefore compute the commutator of two field operators,

$$\begin{aligned} \left[\phi(x), \phi(y)\right] &= \left[\phi^{(+)}(x), \phi^{(-)}(y)\right] - \left[\phi^{(+)}(y), \phi^{(-)}(x)\right] \\ &= \int \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\Theta(k^0) \left[e^{-ik(x-y)} - e^{+ik(x-y)}\right], \end{aligned} \tag{11.30}$$

where we have used Eq. 11.21 twice. Both terms are Lorentz invariant, since the integration measure is invariant under orthochronous Lorentz transformations (Section 5.1.1). For a space-like separation, $(x-y)^2 < 0$, there exists a Lorentz transformation which maps $(x-y)$ to $-(x-y)$; applying it to the first term turns it into the second, and the commutator vanishes:

$$\left[\phi(x), \phi(y)\right] = 0, \qquad (x-y)^2 < 0. \tag{11.31}$$

For a time-like separation no such transformation exists, and the commutator does not vanish. Measurements of the field (and of any local observable built from it, such as the energy density) at space-like separated points therefore never interfere with each other,

even though the transition amplitude $M_{y\to x}$ is not zero there. This property is called microcausality, and we shall find it in all the field theories of this course.

The non-vanishing of the amplitude outside the light-cone has a heuristic interpretation. Essentially, a particle can escape from the light-cone by a distance Δx which is not observable to us due to the uncertainty principle. The transition amplitude is only significant for

$$m(\Delta x) \sim 1, \tag{11.32}$$

which is the uncertainty we expect for determining its position for a momentum uncertainty $\Delta p \sim m$. Essentially, to observe that a particle of a certain mass has escaped the light-cone, we must first measure the momentum of the particle with an accuracy as good as its mass (in order to identify the particle from its mass). Then, the uncertainty in its position will be as large as the distance over which we expect a finite probability for the particle to escape from the light-cone. Numerically, the suppression is strong: for $m\,\Delta x \sim 10$ the amplitude is already reduced by a factor $e^{-10} \approx 5 \times 10^{-5}$. Such short excursions matter for particles which are created and absorbed again without ever being detected, the so-called **virtual particles**. They need not satisfy $p^2 = m^2$. For example, an electron and a positron of very high energy can be treated as massless; they cannot fuse into a real photon, since their total momentum has $(p_1 + p_2)^2 = 2p_1 \cdot p_2 > 0$ (for non-collinear momenta) while a real photon has $p^2 = 0$. The photon which they produce, and which later creates another pair, is virtual. When we include interactions, we will see that the propagation of virtual particles is described by the propagators which we construct in this chapter.

Time-like transition B

For large times, $\Delta t \to \infty$, we can prove (**exercise**) that

$$\lim_{\Delta t\to\infty} M_{y\to x} \sim e^{-im\Delta t}. \tag{11.33}$$

The transition amplitude for a particle to remain within the light-cone is finite.

11.3 Time ordering and the Feynman-Stückelberg propagator for the real Klein-Gordon field

In order for the "correlation function",

$$\langle 0|\,\phi(x)\phi(y)\,|0\rangle\,,$$

to be a transition amplitude where a particle is created at point y^μ and it is destroyed at a point x^μ, the time x^0 must occur after y^0, i.e. $x^0 > y^0$. These are the transitions which have a physical realisation. Keeping track of the time variable is cumbersome in a relativistic theory where space and time coordinates get mixed up with Lorentz transformations. We need a simple method to do so.

To see why this matters, let us anticipate how propagation enters a full physical process. Later we will add to the Lagrangian interaction terms, for example a term cubic in the field, and particles will be able to decay or to fuse. Consider two particles A and B in the initial state and two particles C and D in the final state, all of the same type in a theory with a single field ϕ. One way for this to happen is that A emits a particle V at a point y and becomes D; V propagates from y to x, where it is absorbed by B, which becomes C. If a quantum mechanical process happens in stages, its amplitude is the product of the amplitudes of the stages. We cannot know where the emission and the absorption

took place, and by the superposition principle we must integrate over all positions x and y. Causality requires that the emission happens before the absorption. Schematically,

$$M_1 = \int d^4x\, d^4y\; g_{BVC}(x)\, \Theta(x^0 - y^0) M_{y\to x}\, g_{AVD}(y),$$

where the factors g denote the (yet unknown) amplitudes for the elementary interactions. But the same final state can also be reached if B emits V at x and becomes C, while V propagates from x to y and is absorbed by A, which becomes D:

$$M_2 = \int d^4x\, d^4y\; g_{AVD}(y)\, \Theta(y^0 - x^0) M_{x\to y}\, g_{BVC}(x).$$

It will turn out that the interaction factors are the same, irrespective of whether a particle is emitted or absorbed. The two processes always come together, and we will always need the sum

$$M_1 + M_2 = \int d^4x\, d^4y\; g_{BVC}(x)\, g_{AVD}(y) \left[\Theta(x^0 - y^0) M_{y\to x} + \Theta(y^0 - x^0) M_{x\to y}\right].$$

Notice that in both terms the particle propagates forward in time. This is the case for all transitions in quantum field theory, for particles and for anti-particles. We will derive these rules from perturbation theory later.

It has been noticed that there is a simple way to demand a physical flow of time in transition amplitudes (propagators) when they are appropriately combined. Consider the Feynman-Stückelberg propagator which is defined to account for both cases, $x^0 > y^0$ and $x^0 < y^0$ in a transition between two arbitrary space-time points x^μ, y^μ:

$$\begin{aligned} \langle 0|\, T\phi(x)\phi(y)\, |0\rangle &= \Theta(x^0 - y^0)\, \langle 0|\, \phi(x)\phi(y)\, |0\rangle \\ &\quad + \Theta(y^0 - x^0)\, \langle 0|\, \phi(y)\phi(x)\, |0\rangle\,. \end{aligned} \tag{11.34}$$

The "time-ordering" symbol T is defined to order the operators following it from the later to the earlier times. The Feynman-Stückelberg propagator has a very nice integral representation. Explicitly, from the above definition and the result of Eq. 11.23, we find

$$\begin{aligned} \langle 0|\, T\phi(x)\phi(y)\, |0\rangle &= \int \frac{d^3\vec{k}}{(2\pi)^3} \frac{1}{2\sqrt{\vec{k}^2 + m^2}} \left[\Theta(x^0 - y^0) e^{-ik\cdot(x-y)} + \Theta(y^0 - x^0) e^{-ik\cdot(y-x)}\right] \\ &= \lim_{\delta\to 0} \int_{-\infty}^{+\infty} \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\delta} e^{-ik\cdot(x-y)}. \end{aligned} \tag{11.35}$$

We can verify that the four-dimensional integral of the last line is equal to the line before explicitly, by using Cauchy's theorem. In Fig. 11.2 we plot the poles of the integrand in the complex plane of the energy k^0. These are located at $k^0 = -\omega + i\delta$ and $k^0 = \omega - i\delta$, with $\omega = \sqrt{\vec{k}^2 + m^2}$. (Strictly, expanding $\sqrt{\omega^2 - i\delta}$ for small δ gives $\omega - i\delta/(2\omega)$. Since $\delta \to 0^+$ at the end, only the sign of the imaginary part matters, and we can rename $\delta/(2\omega) \to \delta$.) At $k^0 = \pm i\infty$, the exponential in the integrand behaves as,

$$e^{-ik\cdot(x-y)} \to e^{+i\vec{k}(\vec{x}-\vec{y})} e^{\pm\infty(x^0-y^0)}. \tag{11.36}$$

If $x^0 > y^0$ we can close the contour of integration to the lower complex half-plane, guaranteeing that the integrand vanishes at $-i\infty$. If $x^0 < y^0$ we must close the contour to

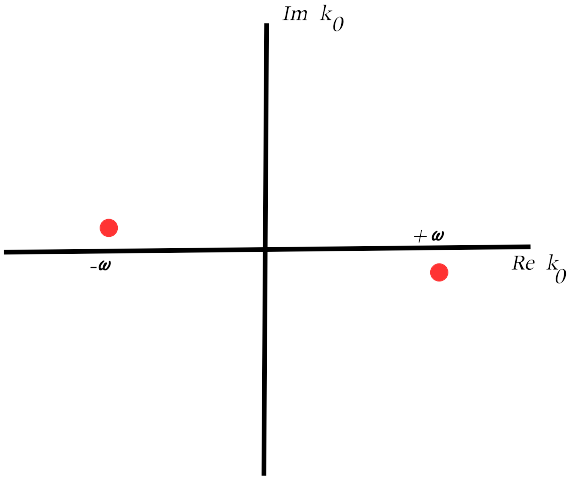


Figure 11.2. Poles of the integrand for the Feynman-Stückelberg propagator.

the upper half-plane. In each case, we pick up the residue of one of the poles, at $k^0 = \pm\omega$, recovering the two terms of the first line in Eq. 11.35.

The expression for the Feynman propagator,

$$\langle 0|\, T\phi(x)\phi(y)\,|0\rangle = \lim_{\delta\to 0}\int_{-\infty}^{+\infty}\frac{d^4k}{(2\pi)^4}\frac{i}{k^2-m^2+i\delta}e^{-ik\cdot(x-y)}, \tag{11.37}$$

is of paramount importance for the computation of generic transition amplitudes in quantum field theory and it is simple. It accounts for the correct time-flow with the $i\delta$ *prescription* in the denominator. Is this only a cosmetic rewriting, with the Θ functions hidden in the k^0 integration? No. Recall the integrations over the positions x and y of the interactions in the process which we discussed above. With the Θ functions, the time integrations are restricted and they are tedious to perform. In the four-dimensional form, x and y appear only in the exponentials $e^{-ik\cdot(x-y)}$, and the unrestricted integrations over all of space-time simply produce δ-functions of the momenta, i.e. momentum conservation at each interaction point. This is what makes Feynman diagrams so efficient. The propagator was introduced independently by Feynman and by the Swiss physicist Ernst Stückelberg, whose work came at the same time or even somewhat earlier. An important property of the Feynman propagator is that it is a **Green's function** of the Klein-Gordon equation. We can act on it with the differential operator of the Klein-Gordon equation,

$$\begin{aligned}
\left(\partial^2+m^2\right)\langle 0|\, T\phi(x)\phi(y)\,|0\rangle &= \left(\partial^2+m^2\right)\int_{-\infty}^{+\infty}\frac{d^4k}{(2\pi)^4}\frac{i}{k^2-m^2+i\delta}e^{-ik\cdot(x-y)} \\
&= -i\int_{-\infty}^{+\infty}\frac{d^4k}{(2\pi)^4}e^{-ik\cdot(x-y)} \\
\rightsquigarrow \left(\partial^2+m^2\right)\langle 0|\, T\phi(x)\phi(y)\,|0\rangle &= -i\delta^4(x-y).
\end{aligned} \tag{11.38}$$

11.4 Feynman propagator for the complex Klein-Gordon field

We will now compute the transition amplitude, or the Feynman propagator, for the complex Klein-Gordon field,

$$\phi(\vec{x},t) = \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}\left[a(k)e^{-ik\cdot x} + b^\dagger(k)e^{ik\cdot x}\right]. \tag{11.39}$$

Notice that the Hermitian conjugate of the field acting on the vacuum creates a particle with charge +1 at a certain position x,

$$\left|\text{particle at position } x\right\rangle = \phi^\dagger(\vec{x},t)\,|0\rangle\,, \tag{11.40}$$

while a field acting on the vacuum creates an anti-particle with charge −1,

$$\left|\text{anti-particle at position } x\right\rangle = \phi(\vec{x},t)\,|0\rangle\,. \tag{11.41}$$

Consider the correlation function,

$$\langle 0|\,\phi(y)\phi^\dagger(x)\,|0\rangle\,, \tag{11.42}$$

which we can interpret as a causal sequence where a particle is first created at $x^\mu \equiv (x^0,\vec{x})$ and it is destroyed at $y^\mu \equiv (y^0,\vec{y})$. We must then have $y^0 > x^0$. Similarly, the correlation function

$$\langle 0|\,\phi^\dagger(x)\phi(y)\,|0\rangle \tag{11.43}$$

can be interpreted as an anti-particle which is first created at $y^\mu \equiv (y^0,\vec{y})$ and it is destroyed at $x^\mu \equiv (x^0,\vec{x})$. We must then have $x^0 > y^0$.

We define the Feynman propagator to account for the two equivalent possibilities for the forward-in-time propagation of either a particle or an anti-particle,

$$\begin{aligned}&\langle 0|\,T\phi(x)\phi^\dagger(y)\,|0\rangle\\ &= \Theta(x^0-y^0)\,\langle 0|\,\phi(x)\phi^\dagger(y)\,|0\rangle + \Theta(y^0-x^0)\,\langle 0|\,\phi^\dagger(y)\phi(x)\,|0\rangle\,.\end{aligned} \tag{11.44}$$

We can repeat the same computation as for the Feynman propagator of the real Klein-Gordon field. The result is identical:

$$\langle 0|\,T\phi(x)\phi^\dagger(y)\,|0\rangle \;=\; \lim_{\delta\to 0}\int_{-\infty}^{+\infty}\frac{d^4k}{(2\pi)^4}\frac{i}{k^2-m^2+i\delta}e^{-ik\cdot(x-y)}. \tag{11.45}$$

The physical picture is as for the real field. Suppose that our charged particles can interact with photons, and consider two photons A and B in the initial state. Photon A may convert into a particle-antiparticle pair at y, and the antiparticle propagates to x where it is absorbed by photon B. Alternatively, photon B converts into a pair at x, and it is now the particle which propagates from x to y, where it is absorbed by photon A. The Feynman propagator combines the propagation of a particle in one time ordering with the propagation of an anti-particle in the other. In both cases the propagation is forward in time. You may have heard the picture of an anti-particle as a particle moving backwards in time. This is not what happens in quantum field theory: particles and anti-particles both propagate forward in time, and the time-ordered product collects the two possibilities.

11.5 Feynman propagator for the Dirac field

We now recall the expression of the Dirac field operator,

$$\psi(\vec{x},t) = \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \sum_s \left[a_s(p) u_s(p) e^{-ip\cdot x} + b_s^\dagger(p) v_s(p) e^{ip\cdot x} \right]. \tag{11.46}$$

The operators $a_s^\dagger(\vec{p})$ acting on the vacuum create particles of momentum $\vec{p}$ with a spin index s, while the operators $b_s^\dagger(\vec{p})$ create antiparticles. The state

$$\bar{\psi}(x^\mu)\,|0\rangle = \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \sum_s \bar{u}_s(p) e^{+ip\cdot x} a_s^\dagger(p)\,|0\rangle\,,$$

is the Fourier transform of one-particle momentum eigenstates. Such states can therefore be interpreted as particle states at a certain position x^μ. Let us now compute the correlation function of two Dirac spinor fields,

$$\begin{aligned}
\langle 0|\,\psi_a(x)\bar{\psi}_b(y)\,|0\rangle &= \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \sum_s u_a^s(p)\bar{u}_b^s(p) e^{-ip\cdot(x-y)} \\
&= \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \left(\not{p} + m\right)_{ab} e^{-ip\cdot(x-y)} \\
\rightsquigarrow \langle 0|\,\psi_a(x)\bar{\psi}_b(y)\,|0\rangle &= \left(i\gamma^\mu \frac{\partial}{\partial x^\mu} + m\mathbf{1} \right)_{ab} \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} e^{-ip\cdot(x-y)}.
\end{aligned} \tag{11.47}$$

Similarly, we find (**exercise**)

$$\langle 0|\,\bar{\psi}_b(y)\psi_a(x)\,|0\rangle = -\left(i\gamma^\mu \frac{\partial}{\partial x^\mu} + m\mathbf{1} \right)_{ab} \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} e^{-ip\cdot(y-x)}. \tag{11.48}$$

Notice the overall minus sign for the transition amplitude of an anti-particle with respect to the corresponding amplitude in the case of a particle.

We can combine the two forward-in-time propagations by including this relative minus sign into the definition of the time ordering for fermions,

$$\langle 0|\,T\psi_a(x)\bar{\psi}_b(y)\,|0\rangle = \Theta(x^0-y^0)\,\langle 0|\,\psi_a(x)\bar{\psi}_b(y)\,|0\rangle - \Theta(y^0-x^0)\,\langle 0|\,\bar{\psi}_b(y)\psi_a(x)\,|0\rangle\,. \tag{11.49}$$

From the above, we find

$$\begin{aligned}
\langle 0|\,T\psi_a(x)\bar{\psi}_b(y)\,|0\rangle &= \left(i\gamma^\mu \frac{\partial}{\partial x^\mu} + m\mathbf{1} \right)_{ab} \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \left[\Theta(y^0-x^0) e^{-ip\cdot(y-x)} \right. \\
&\qquad \left. + \Theta(x^0-y^0) e^{-ip\cdot(x-y)} \right] \\
&= \left(i\gamma^\mu \frac{\partial}{\partial x^\mu} + m\mathbf{1} \right)_{ab} \int_{-\infty}^{+\infty} \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\delta} e^{-ik\cdot(x-y)},
\end{aligned} \tag{11.50}$$

which yields for the fermion Feynman propagator the integral representation,

$$\langle 0|\,T\psi_a(x)\bar{\psi}_b(y)\,|0\rangle = \int_{-\infty}^{+\infty} \frac{d^4k}{(2\pi)^4} \frac{i\,(\not{k} + m\mathbf{1})_{ab}}{k^2 - m^2 + i\delta} e^{-ik\cdot(x-y)}. \tag{11.51}$$

Exercise: Show that

$$\langle 0|\,T\psi_a(x)\psi_b(y)\,|0\rangle = \langle 0|\,T\bar{\psi}_a(x)\bar{\psi}_b(y)\,|0\rangle = 0. \tag{11.52}$$

11.6 Feynman propagator for the photon field

The photon Feynman propagator is defined as,

$$\begin{aligned}\langle 0|\, TA^{\mu}(x)A^{\nu}(y)\,|0\rangle &= \langle 0|\, A^{\mu}(x)A^{\nu}(y)\,|0\rangle\,\Theta(x^0-y^0)\\ &\quad+\langle 0|\, A^{\nu}(y)A^{\mu}(x)\,|0\rangle\,\Theta(y^0-x^0).\end{aligned}\tag{11.53}$$

This is worked out in the exercises. We find that

$$\langle 0|\, TA^{\mu}(x)A^{\nu}(y)\,|0\rangle = \int \frac{d^4p}{(2\pi)^4}\frac{-i\left[g^{\mu\nu}-(1-\xi)\frac{p^{\mu}p^{\nu}}{p^2}\right]}{p^2+i\delta}e^{-ip(x-y)}.\tag{11.54}$$

11.7 Wick's theorem: time ordering, normal ordering and propagation

We decompose the free scalar field ϕ into a term with a creation operator and a term with an annihilation operator,

$$\phi(x)=\phi_+(x)+\phi_-(x),\qquad \text{(as in Section 11.2, + labels the positive-frequency part)}\tag{11.55}$$

with

$$\phi_+(x)=\int\frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}e^{-ik\cdot x}a(k),\tag{11.56}$$

and

$$\phi_-(x)=\int\frac{d^3\vec{k}}{(2\pi)^3 2\omega_k}e^{+ik\cdot x}a^{\dagger}(k).\tag{11.57}$$

The time-ordered product of two scalar fields is,

$$\begin{aligned}&T\left\{\phi(x)\phi(y)\right\}=\Theta(x^0-y^0)\phi(x)\phi(y)+\left(x^{\mu}\leftrightarrow y^{\mu}\right)\\
=\;&\Theta(x^0-y^0)\Big\{\phi_-(x)\phi_-(y)+\phi_+(x)\phi_+(y)+\phi_-(x)\phi_+(y)\\
&+\phi_+(x)\phi_-(y)\Big\}+\left(x^{\mu}\leftrightarrow y^{\mu}\right)\\
=\;&\Theta(x^0-y^0)\Big\{\phi_-(x)\phi_-(y)+\phi_+(x)\phi_+(y)+\phi_-(x)\phi_+(y)\\
&+\phi_-(y)\phi_+(x)+\left[\phi_+(x),\phi_-(y)\right]\Big\}+\left(x^{\mu}\leftrightarrow y^{\mu}\right)\\
=\;&\left[\Theta(x^0-y^0)+\Theta(y^0-x^0)\right]\Big\{\phi_-(x)\phi_-(y)+\phi_+(x)\phi_+(y)\\
&+\phi_-(x)\phi_+(y)+\phi_-(y)\phi_+(x)\Big\}\\
&+\left[\phi_+(x),\phi_-(y)\right]\Theta(x^0-y^0)+\left[\phi_+(y),\phi_-(x)\right]\Theta(y^0-x^0).\end{aligned}\tag{11.58}$$

The sum of the Θ functions in the first term is equal to one. We observe that in the curly bracket all field products appear with creation operators preceding annihilation operators.

The curly bracket is then just the normal-ordered product of the two field operators,

$$\left\{ \dots \right\} = : \phi(x)\phi(y) : \ . \tag{11.59}$$

The sum of the two terms in the last line is a known object, the Feynman-Stückelberg propagator, which is not an operator but a c-number. Using the relation

$$[a(k), a^\dagger(k')] = (2\pi)^3 2\omega_k \delta^{(3)}(\vec{k} - \vec{k}'), \tag{11.60}$$

we find that

$$\left[\phi_+(x), \phi_-(y)\right] = \int \frac{d^3k}{(2\pi)^3 2\omega_k} e^{-ik(x-y)}. \tag{11.61}$$

Combining the two terms together, we have

$$\begin{aligned} & \left[\phi_+(x), \phi_-(y)\right] \Theta\left(x^0 - y^0\right) + \left[\phi_+(y), \phi_-(x)\right] \Theta\left(y^0 - x^0\right) \\ = & \int \frac{d^3k}{(2\pi)^3 2\omega_k} \left[e^{-ik(x-y)} \Theta(x^0 - y^0) + e^{-ik(y-x)} \Theta(y^0 - x^0)\right] \\ = & \langle 0 | T\left\{\phi(x)\phi(y)\right\} | 0 \rangle \, . \end{aligned} \tag{11.62}$$

We can then write that,

$$T\left\{\phi(x)\phi(y)\right\} =: \phi(x)\phi(y) : + \langle 0 | T\left\{\phi(x)\phi(y)\right\} | 0 \rangle \, . \tag{11.63}$$

We have proved that the time-ordered product of two free-field operators is the normal-ordered product plus a c-number function, which is the propagator of the two fields.

This result generalises easily to the time-ordered product of an arbitrary number of field operators. For three fields we have

$$\begin{aligned} T\left\{\phi(x_1)\phi(x_2)\phi(x_3)\right\} = & : \phi(x_1)\phi(x_2)\phi(x_3) : \\ & + \langle 0 | T\left\{\phi(x_1)\phi(x_2)\right\} | 0 \rangle \, \phi(x_3) \\ & + \langle 0 | T\left\{\phi(x_1)\phi(x_3)\right\} | 0 \rangle \, \phi(x_2) \\ & + \langle 0 | T\left\{\phi(x_2)\phi(x_3)\right\} | 0 \rangle \, \phi(x_1). \end{aligned} \tag{11.64}$$

For four fields we have

$$\begin{aligned} T\left\{\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\right\} = & : \phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4) : \\ & + \langle 0 | T\left\{\phi(x_1)\phi(x_2)\right\} | 0 \rangle : \phi(x_3)\phi(x_4) : \\ & + \langle 0 | T\left\{\phi(x_1)\phi(x_3)\right\} | 0 \rangle : \phi(x_2)\phi(x_4) : \\ & + \langle 0 | T\left\{\phi(x_1)\phi(x_4)\right\} | 0 \rangle : \phi(x_2)\phi(x_3) : \\ & + \langle 0 | T\left\{\phi(x_1)\phi(x_2)\right\} | 0 \rangle \langle 0 | T\left\{\phi(x_3)\phi(x_4)\right\} | 0 \rangle \\ & + \langle 0 | T\left\{\phi(x_1)\phi(x_3)\right\} | 0 \rangle \langle 0 | T\left\{\phi(x_2)\phi(x_4)\right\} | 0 \rangle \\ & + \langle 0 | T\left\{\phi(x_1)\phi(x_4)\right\} | 0 \rangle \langle 0 | T\left\{\phi(x_2)\phi(x_3)\right\} | 0 \rangle \, . \end{aligned} \tag{11.65}$$

In general, Wick's theorem states that

$$\begin{aligned} & T\left\{\phi(x_1) \dots \phi(x_n)\right\} \\ = & : \phi(x_1) \dots \phi(x_n) + \text{ all contractions} : , \end{aligned} \tag{11.66}$$

where a "contraction" means to replace one or more pairs of fields with their propagator. The theorem can be proved easily by induction. The notation deserves a comment. Each contraction is a c-number, a function of two space-time points and not an operator, and it is taken out of the normal-ordering symbol, as in Eqs. 11.64–11.65; only the uncontracted fields remain normal ordered. The mechanics behind the theorem is simple: we decompose every field into its creation and annihilation parts, and we move all creation parts to the left of all annihilation parts. Every time we exchange the order of two such parts we pick up a commutator, which is a c-number and gives rise to a propagator.

Notice that the normal-ordered products in the expressions produced with Wick's theorem vanish when sandwiched between vacuum states,

$$\langle 0| : \prod_j \phi(x_j) : |0\rangle = 0, \tag{11.67}$$

given that creation operators are placed before annihilation operators in the normal ordering, and

$$\langle 0| a^\dagger(k) = a(k) |0\rangle = 0.$$

From Eqs. 11.63–11.65 we derive the tautology

$$\langle 0| T\left\{\phi(x)\phi(y)\right\} |0\rangle = \langle 0| T\left\{\phi(x)\phi(y)\right\} |0\rangle\,, \tag{11.68}$$

and the more informative equations,

$$\langle 0| T\left\{\phi(x_1)\phi(x_2)\phi(x_3)\right\} |0\rangle \;\; = \;\; 0, \tag{11.69}$$

and

$$\begin{aligned}\langle 0| T\left\{\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\right\} |0\rangle \;\; &= \;\; \langle 0| T\left\{\phi(x_1)\phi(x_2)\right\} |0\rangle \langle 0| T\left\{\phi(x_3)\phi(x_4)\right\} |0\rangle \\ &+ \;\; \langle 0| T\left\{\phi(x_1)\phi(x_3)\right\} |0\rangle \langle 0| T\left\{\phi(x_2)\phi(x_4)\right\} |0\rangle \\ &+ \;\; \langle 0| T\left\{\phi(x_1)\phi(x_4)\right\} |0\rangle \langle 0| T\left\{\phi(x_2)\phi(x_3)\right\} |0\rangle\end{aligned} \tag{11.70}$$

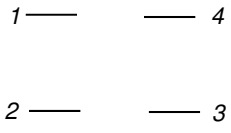
These equations have a transparent physical meaning. In a free theory particles can only propagate. Eq. 11.70 sums over all the ways in which two free particles can travel among the four points: from x_1 to x_2 and independently from x_3 to x_4, or x_1 with x_3 and x_2 with x_4, or x_1 with x_4 and x_2 with x_3, each propagation in the direction dictated by the time ordering. Since the two particles do not talk to each other, the amplitude for each possibility is the product of the two propagation amplitudes. With $2n$ points we obtain in the same way all the possibilities for n free particles. With an odd number of points, Eq. 11.69, one point is always left alone and the result is zero. A particle created at that point would have nowhere to go. Notice that it cannot simply stay where it is: a particle at rest still propagates in time, i.e. to another space-time point. The vacuum expectation value of a time-ordered product of field operators therefore encodes, in a single compact object, everything that free particles can do.

Such equations admit a graphical representation. Let us represent the "two-point" correlation function with a straight line,

$$\langle 0| T\left\{\phi(x_1)\phi(x_2)\right\} |0\rangle = \; 1 \text{———} 2. \tag{11.71}$$

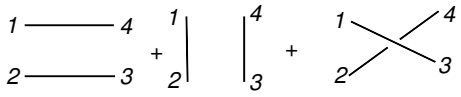
The four-point function is represented as

$$\langle 0|\, T\left\{\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\right\}|0\rangle =$$



contracted

$$=$$



(11.72)

It is then very easy to apply Wick's theorem pictorially, simply by drawing all possible pairings of the points which appear in $\langle 0|\, T\left\{\phi(x_1)\phi(x_2)\ldots\right\}|0\rangle$.

11.7.1 Wick's theorem for Dirac fermion fields*

Consider a free Dirac field,

$$\psi_b(x) = \psi_b^-(x) + \psi_b^+(x), \tag{11.73}$$

where the index b is a spinor index, and

$$\psi_b^+(x) = \sum_s \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} u_s^b(k) a_s(k) e^{-ik\cdot x}, \tag{11.74}$$

$$\psi_b^-(x) = \sum_s \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} v_s^b(k) b_s^\dagger(k) e^{+ik\cdot x}, \tag{11.75}$$

are the terms with annihilation and creation operators respectively. The spin-sum is over all the solutions s of the Dirac equation for the spinors u, v. We recall the spin-sum identities,

$$\sum_s u_s^a(p)\bar{u}_s^b(p) = \left(\not{p} + m\mathbf{1}\right)_{ab}, \tag{11.76}$$

$$\sum_s v_s^a(p)\bar{v}_s^b(p) = \left(\not{p} - m\mathbf{1}\right)_{ab}. \tag{11.77}$$

Let us now work out the time-ordered product

$$T\psi_a(x)\bar{\psi}_b(y).$$

Following the steps of the previous section and recalling the minus signs in the definitions of the normal ordering and the time-ordering, we find that (**exercise**):

$$\begin{aligned} T\psi_a(x)\bar{\psi}_b(y) &= :\psi_a(x)\bar{\psi}_b(y): + \Theta(x^0 - y^0)\left\{\psi_a^+(x), \overline{\psi_b^+}(y)\right\} \\ &\quad - \Theta(y^0 - x^0)\left\{\overline{\psi_b^-}(y), \psi_a^-(x)\right\}. \end{aligned} \tag{11.78}$$

The anticommutators above are complex numbers and, without any harm, we can take them to be equal to their vacuum expectation values,

$$\begin{aligned} \left\{\psi_a^+(x), \overline{\psi_b^+}(y)\right\} &= \langle 0|\left\{\psi_a^+(x), \overline{\psi_b^+}(y)\right\}|0\rangle \\ &= \langle 0|\,\psi_a^+(x)\overline{\psi_b^+}(y)\,|0\rangle = \langle 0|\,\psi_a(x)\bar{\psi}_b(y)\,|0\rangle, \end{aligned} \tag{11.79}$$

and similarly,

$$\left\{\overline{\psi_b^-}(y), \psi_a^-(x)\right\} = \langle 0|\,\bar{\psi}_b(y)\psi_a(x)\,|0\rangle. \tag{11.80}$$

We can then cast Eq. 11.78 in the form

$$T\psi_a(x)\bar{\psi}_b(y) =: \psi_a(x)\bar{\psi}_b(y) : + \langle 0|\, T\psi_a(x)\bar{\psi}_b(y)\,|0\rangle\,. \tag{11.81}$$

Let us now study the vacuum expectation value of products of four fields. To facilitate the manipulations we shall introduce the shorthand notation:

$$\psi_{a_i}(x_i) \equiv i, \quad \langle 0|\ldots|0\rangle \equiv \langle\ldots\rangle\,. \tag{11.82}$$

Assuming that $x_1^0 > x_2^0 > x_3^0 > x_4^0$, we find after a little algebra

$$\begin{aligned}
\left\langle T\bar{1}2\bar{3}4\right\rangle &= \left\langle \bar{1}2\bar{3}4\right\rangle \\
&= \left\langle \bar{1}_-(2_+ + 2_-)(\bar{3}_+ + \bar{3}_-)4_-\right\rangle \\
&= \left\langle \left[\left\{\bar{1}_-, 2_-\right\} + \bar{1}_-2_+\right]\left[\left\{\bar{3}_-, 4_-\right\} + \bar{3}_+4_-\right]\right\rangle \\
&= \left\langle \bar{1}2\right\rangle\left\langle \bar{3}4\right\rangle + \left\langle \bar{1}_-2_+\bar{3}_+4_-\right\rangle \\
&= \left\langle \bar{1}2\right\rangle\left\langle \bar{3}4\right\rangle + \left\langle \bar{1}_-\left\{2_+, \bar{3}_+\right\}4_-\right\rangle - \left\langle \bar{1}_-\bar{3}_+2_+4_-\right\rangle \\
&= \left\langle \bar{1}2\right\rangle\left\langle \bar{3}4\right\rangle + \left\langle \bar{1}_-4_-\right\rangle\left\{2_+, \bar{3}_+\right\} + \left\langle \bar{3}_+\bar{1}_-2_+4_-\right\rangle \\
&= \left\langle \bar{1}2\right\rangle\left\langle \bar{3}4\right\rangle + \left\langle \bar{1}4\right\rangle\left\langle 2\bar{3}\right\rangle + 0 \\
&= \left\langle T\bar{1}2\right\rangle\left\langle T\bar{3}4\right\rangle + \left\langle T\bar{1}4\right\rangle\left\langle T2\bar{3}\right\rangle\,.
\end{aligned} \tag{11.83}$$

Let us now assume that $x_1^0 > x_2^0 > x_4^0 > x_3^0$. We have

$$\begin{aligned}
\left\langle T\bar{1}2\bar{3}4\right\rangle &= -\left\langle \bar{1}24\bar{3}\right\rangle \\
&\quad\ldots \\
&= -\left(\left\langle \bar{1}2\right\rangle\left\langle 4\bar{3}\right\rangle - \left\langle \bar{1}4\right\rangle\left\langle 2\bar{3}\right\rangle\right) \\
&= -\left(-\left\langle T\bar{1}2\right\rangle\left\langle T\bar{3}4\right\rangle - \left\langle T\bar{1}4\right\rangle\left\langle T2\bar{3}\right\rangle\right) \\
&= \left\langle T\bar{1}2\right\rangle\left\langle T\bar{3}4\right\rangle + \left\langle T\bar{1}4\right\rangle\left\langle T2\bar{3}\right\rangle\,.
\end{aligned} \tag{11.84}$$

You can verify explicitly that all possible time-orderings give the same result, i.e. the result of Eq. 11.83. We can compute similarly other expectation values of time-ordered products of fermion fields. For example,

$$\left\langle T\bar{1}\bar{2}34\right\rangle = -\left\langle T\bar{1}3\right\rangle\left\langle T\bar{2}4\right\rangle + \left\langle T\bar{1}4\right\rangle\left\langle T\bar{2}3\right\rangle\,. \tag{11.85}$$

The construction of expressions such as Eqs. 11.83–11.85 is algorithmic. Consider the expectation value of a generic time-ordered product of fermionic fields:

$$\left\langle T f_1 f_2 \ldots f_n\right\rangle\,.$$

In the above, f_i can be either ψ or $\bar{\psi}$. Now take the first field in the product and "contract" it (form a propagator) with every other field:

$$\begin{aligned}
\left\langle T f_1 f_2 \ldots f_n\right\rangle \quad = \quad & \left(+\left\langle T f_1 f_2\right\rangle\right)\left\langle T f_3 f_4 f_5 \ldots f_n\right\rangle \\
& + \left(-\left\langle T f_1 f_3\right\rangle\right)\left\langle T f_2 f_4 f_5 \ldots f_n\right\rangle \\
& + \left(+\left\langle T f_1 f_4\right\rangle\right)\left\langle T f_2 f_3 f_5 \ldots f_n\right\rangle \\
& + \ldots
\end{aligned} \tag{11.86}$$

The sign of the contracted terms on the right-hand side can be positive or negative. It is equal to $(-1)^{n_P}$ where n_P is the number of permutations which are needed in order to

bring the two contracted fields which form a propagator in adjacent positions, starting from the ordering in the expression on the left-hand side. Some of these contractions may vanish, since $\langle ji \rangle = \langle \bar{i}\bar{j} \rangle = 0$ and only $\langle \bar{i}j \rangle = -\langle j\bar{i} \rangle \neq 0$. Eq. 11.86 can be applied recursively until no more propagators can be formed. For example, for the expectation value of the time-ordered product of six fields we get:

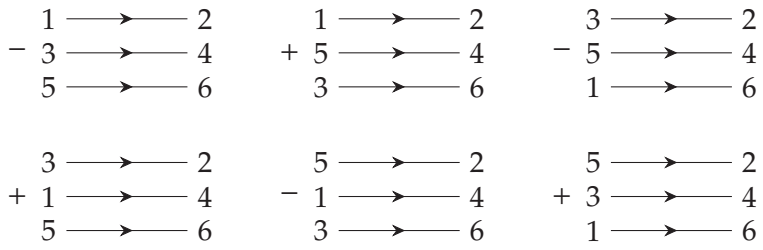
$$\begin{aligned}
\langle T\bar{1}2\bar{3}4\bar{5}6 \rangle &= +\langle T\bar{1}2 \rangle \langle T\bar{3}4\bar{5}6 \rangle + \langle T\bar{1}4 \rangle \langle T2\bar{3}\bar{5}6 \rangle + \langle T\bar{1}6 \rangle \langle T2\bar{3}4\bar{5} \rangle \\
&= +\langle T\bar{1}2 \rangle \left(\langle T\bar{3}4 \rangle \langle T\bar{5}6 \rangle + \langle T\bar{3}6 \rangle \langle T4\bar{5} \rangle \right) \\
&\quad + \langle T\bar{1}4 \rangle \left(\langle T2\bar{3} \rangle \langle T\bar{5}6 \rangle - \langle T2\bar{5} \rangle \langle T\bar{3}6 \rangle \right) \\
&\quad + \langle T\bar{1}6 \rangle \left(\langle T2\bar{3} \rangle \langle T4\bar{5} \rangle + \langle T2\bar{5} \rangle \langle T\bar{3}4 \rangle \right) \\
&= -\langle T2\bar{1} \rangle \langle T4\bar{3} \rangle \langle T6\bar{5} \rangle + \langle T2\bar{1} \rangle \langle T4\bar{5} \rangle \langle T6\bar{3} \rangle \\
&\quad + \langle T2\bar{3} \rangle \langle T4\bar{1} \rangle \langle T6\bar{5} \rangle - \langle T2\bar{3} \rangle \langle T4\bar{5} \rangle \langle T6\bar{1} \rangle \\
&\quad + \langle T2\bar{5} \rangle \langle T4\bar{3} \rangle \langle T6\bar{1} \rangle - \langle T2\bar{5} \rangle \langle T4\bar{1} \rangle \langle T6\bar{3} \rangle .
\end{aligned} \tag{11.87}$$

We can represent Eq. 11.87 graphically. Let us define the Dirac fermion propagator as an arrowed line:

$$\langle i\bar{j} \rangle \equiv \langle 0 | T\psi_{a_i}(x_i)\bar{\psi}_{a_j}(x_j) | 0 \rangle \equiv \quad j \longrightarrow i \tag{11.88}$$

The arrow leaves the field $\bar{\psi}(x_j)$, where the particle is created, and points to the field $\psi(x_i)$, where it is annihilated: it follows the flow of fermion number. This is the convention of Peskin and Schroeder, and it is used for all fermion lines in this book. Then, Eq. 11.87 can be depicted as:

$\langle \bar{1}2\bar{3}4\bar{5}6 \rangle =$



(11.89)

This accounts for all the ways in which we can connect the six points with propagators (bearing in mind that only ψ and $\bar{\psi}$ can be contracted, while two ψ's or two $\bar{\psi}$'s cannot). The sign of each graph is determined by the number of permutations which are required in order to bring the contracted fields next to each other, starting from the ordering on the left-hand side of the equation.

A useful application of what we have learnt is in computing probability amplitudes in perturbative computations in quantum electrodynamics (QED) and in the Standard Model of particle physics. One encounters expectation values of the type:

$$\frac{1}{n!} \int d^4x_1 \dots d^4x_n \ \langle TV(x_1) \dots V(x_n) \rangle , \tag{11.90}$$

with

$$V(x) \equiv \bar{\psi}(x)\Gamma\psi(x) = \Gamma_{ab}\bar{\psi}_a(x)\psi_b(x). \tag{11.91}$$

Γ_{ab} is a generic 4×4 matrix which can of course be written in the basis 1, γ^μ, γ_5, $\gamma_5\gamma^\mu$, $[\gamma^\mu, \gamma^\nu]$.

12

Scattering theory (S-matrix)

We have studied the quantum field theories of free elementary particles with spin-0 (Klein-Gordon field), spin-$\frac{1}{2}$ (Dirac spinor field), and spin-1 (electromagnetic photon field). In all these cases, we could identify a field four-momentum $P^\mu = \left(H, \vec{P}\right)$ which was a conserved quantity. It was also relatively easy to find a complete set of eigenstates for P^μ. In all these "free field theories" the Hamiltonian eigenstates happened to be, in addition, eigenstates of a time-independent "number operator". States with definite energy and momentum were, therefore, also states of a conserved number of indestructible particles.[1]

In reality, however, particles interact and may also be destroyed or created in a scattering process. Realistic field theories contain additional terms in their Lagrangian which do not make it possible to find a constant particle-number operator. In a free field theory, conservation of energy and momentum is a consequence of the fact that the theory is invariant under time and space translations. This is a property which will also hold for interacting field theories. Conservation of the particle number was, on the other hand, only a consequence of the "accident" that a free field theory is rather simple. This "accident" does not occur any more in an interacting field theory, which we expect to describe successfully particle creation and annihilation.

In this chapter we will consider interacting field theories in general, without specifying the details of their Lagrangian. We will only require that they yield physical laws which are invariant under space-time translations, and therefore there is a conserved field-momentum operator. We will also make one further assumption, which we will explain later in detail, that particle interactions happen at short distances, and that particles which are far apart behave as if they were free. These two considerations will lead us to general expressions for the probability amplitude of a scattering process.

Once the Lagrangian contains terms of higher than quadratic order in the fields, the Euler-Lagrange equations become non-linear and the plane-wave solutions of the free theories are no longer available. We then have three tools at our disposal. The first is a full computer simulation of the theory on a space-time lattice (*lattice quantum field theory*). These are among the most expensive computations that can be performed today, and they are feasible only for a limited set of quantities. The second is symmetry. Symmetries do not solve a theory, but they restrict severely the form of its solutions. The third is perturbation theory, which relies on the luck that the additional terms in the Lagrangian come with a small parameter. In this chapter we will see how far we can go with Poincaré symmetry and a few physical assumptions alone. Perturbation theory is the subject of the next chapter.

[1] In mathematical parlance, the Hilbert space of states had a Fock-space representation.

12.1 Propagation in a general field theory

We start our analysis by studying the propagator of fields in a general theory, as we have specified it above. We shall also assume that there is a ground state, which we denote by

$$|\Omega\rangle .$$

For simplicity we consider here the propagator of a scalar (interacting) field $\phi(x)$, which we write as

$$\langle\Omega| T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle = \langle\Omega| T\left\{\phi(x)\mathbf{1}\phi(y)\right\}|\Omega\rangle . \tag{12.1}$$

Let us form a unit operator $\mathbf{1}$ out of a complete set of states. Space-time translation invariance dictates that $P^\mu \equiv (H, \vec{P})$, the field four-momentum of the Lagrangian system, is a conserved quantity. Notice that we do not need the explicit expression for P^μ; Noether's theorem guarantees that such an operator does exist. We denote by $\left|p\right\rangle$ the corresponding eigenstates of the momentum operator,

$$P^\mu \left|p\right\rangle = p^\mu \left|p\right\rangle , \tag{12.2}$$

or, in components,

$$H\left|p\right\rangle = E_p \left|p\right\rangle , \qquad \vec{P}\left|p\right\rangle = \vec{p}\left|p\right\rangle , \tag{12.3}$$

with

$$p^\mu = \left(E_p, \vec{p}\right) . \tag{12.4}$$

As we have remarked, the states $\left|p\right\rangle$ are not necessarily states with a definite number of particles. To simplify the discussion, we also assume here that no momentum eigenvalue is light-like, i.e. $p^2 \neq 0$. Then, for each state $\left|p\right\rangle$ with momentum p^μ, we can perform a Lorentz boost which renders $\vec{p} = 0$. Conversely, all states $\left|p\right\rangle$ can be produced by applying Lorentz transformations to states $|\lambda_0\rangle$ with

$$\vec{P}\,|\lambda_0\rangle = 0. \tag{12.5}$$

Let $\left|\lambda_p\right\rangle$ be the state obtained by boosting $|\lambda_0\rangle$. Then we can write

$$\mathbf{1} = |\Omega\rangle\langle\Omega| + \sum_\lambda \int \frac{d^3\vec{p}}{(2\pi)^3 2E_p} \left|\lambda_p\right\rangle \left\langle\lambda_p\right| . \tag{12.6}$$

The sum in λ runs over all possible states $|\lambda_0\rangle$ with zero space momentum. The measure $\frac{d^3\vec{p}}{(2\pi)^3 2E_p}$ is not a matter of convention. Under a Lorentz transformation the states transform with a unitary operator U, and the unit operator must remain the unit operator, $U\mathbf{1}U^{-1} = \mathbf{1}$; otherwise expectation values would change from one frame to another. This holds only if the integration measure is Lorentz invariant, which is the case for $\frac{d^3\vec{p}}{2E_p} = d^4p\,\delta(p^2 - m_\lambda^2)\,\Theta(p^0)$. We should also not think of the states $\left|\lambda_p\right\rangle$ as single-particle states. We can talk about particles only when there is a number operator which commutes with the Hamiltonian, and Poincaré symmetry does not provide one. For example, in a collision at the LHC we may identify gluons approaching each other from far apart, but when they are very close together the system cannot be described as a definite number of gluons; it is nevertheless a state with definite total energy and momentum. Up to degeneracies, which require additional symmetries in order to be resolved, the label λ simply enumerates the values

m_λ of the energy of the states at rest. Let us now assume that $x^0 > y^0$. The propagator is then written as

$$\begin{aligned} & \langle\Omega| T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle\Big|_{x^0>y^0} = \\ & \langle\Omega|\phi(x)\left(|\Omega\rangle\langle\Omega| + \sum_\lambda \int \frac{d^3\vec{p}}{(2\pi)^3 2E_p}\left|\lambda_p\right\rangle\left\langle\lambda_p\right|\right)\phi(y)|\Omega\rangle \\ = \; & \langle\Omega|\phi(x)|\Omega\rangle\langle\Omega|\phi(y)|\Omega\rangle \\ + \; & \sum_\lambda \int \frac{d^3\vec{p}}{(2\pi)^3 2E_p}\langle\Omega|\phi(x)\left|\lambda_p\right\rangle\left\langle\lambda_p\right|\phi(y)|\Omega\rangle\,. \end{aligned} \tag{12.7}$$

Consider now the matrix element

$$\langle\Omega|\phi(x)\left|\lambda_p\right\rangle,$$

and recall that the system is invariant under space-time translations $x^\mu \to x^\mu + \epsilon^\mu$. The field four-momentum is the generator of space-time translations. If we know the value of the quantum field at a point y^μ, then we can find its value at a different space-time point by applying

$$\phi(x^\mu) = e^{iP\cdot(x-y)}\phi(y)e^{-iP\cdot(x-y)}. \tag{12.8}$$

Exercise: *Prove the above using the theorem of Section 5.4.*
It is very convenient to express the field in the matrix element in terms of the value of the quantum field at the origin,

$$\langle\Omega|\phi(x)\left|\lambda_p\right\rangle \;=\; \langle\Omega|e^{iP\cdot x}\phi(0)e^{-iP\cdot x}\left|\lambda_p\right\rangle. \tag{12.9}$$

The ground state is invariant under space-time translations,

$$e^{-iP\cdot x}|\Omega\rangle = |\Omega\rangle\,. \tag{12.10}$$

Also, a state $\left|\lambda_p\right\rangle$ has a definite momentum $p^\mu = \left(E_p, \vec{p}\right)$. Thus,

$$e^{-iP\cdot x}\left|\lambda_p\right\rangle = \left|\lambda_p\right\rangle e^{-ip\cdot x}\Big|_{p^0=E_p}. \tag{12.11}$$

We can then write

$$\langle\Omega|\phi(x)\left|\lambda_p\right\rangle = \langle\Omega|\phi(0)\left|\lambda_p\right\rangle e^{-ip\cdot x}\Big|_{p^0=E_p}. \tag{12.12}$$

We now recall that the state $\left|\lambda_p\right\rangle$ is produced from a state with zero space momentum by a Lorentz boost,

$$\left|\lambda_p\right\rangle = U\left(\Lambda_p\right)|\lambda_0\rangle\,. \tag{12.13}$$

We can also exploit that the vacuum state is invariant under Lorentz transformations,

$$U\left(\Lambda_p\right)|\Omega\rangle = U^{-1}\left(\Lambda_p\right)|\Omega\rangle = |\Omega\rangle\,, \tag{12.14}$$

and cast Eq. 12.12 as

$$\langle\Omega|\phi(x)\left|\lambda_p\right\rangle = \langle\Omega|U^{-1}\left(\Lambda_p\right)\phi(0)U\left(\Lambda_p\right)|\lambda_0\rangle\, e^{-ip\cdot x}\Big|_{p^0=E_p}. \tag{12.15}$$

For a scalar field,

$$U\left(\Lambda_p\right)\phi(x^\mu)U^{-1}\left(\Lambda_p\right) = \phi\left(\Lambda^\mu_\nu x^\nu\right), \tag{12.16}$$

which for a position at the origin $x^\mu = 0$ yields

$$U\left(\Lambda_p\right)\phi(0)U^{-1}\left(\Lambda_p\right) = U^{-1}\left(\Lambda_p\right)\phi(0)U\left(\Lambda_p\right) = \phi(0). \tag{12.17}$$

We then find that the space-time dependence of the matrix element in Eq. 12.15 is a simple exponential,

$$\langle\Omega|\,\phi(x)\,\left|\lambda_p\right\rangle = \langle\Omega|\,\phi(0)\,|\lambda_0\rangle\; e^{-ip\cdot x}\Big|_{p^0=E_p}. \tag{12.18}$$

Similarly,

$$\left\langle\lambda_p\right|\phi(y)\,|\Omega\rangle = \langle\lambda_0|\,\phi(0)\,|\Omega\rangle\; e^{+ip\cdot y}\Big|_{p^0=E_p}. \tag{12.19}$$

Finally,

$$\langle\Omega|\,\phi(y)\,|\Omega\rangle = \langle\Omega|\,e^{iP\cdot y}\phi(0)e^{-iP\cdot y}\,|\Omega\rangle = \langle\Omega|\,\phi(0)\,|\Omega\rangle\,. \tag{12.20}$$

Using Eqs. 12.18–12.20, we find that the propagator in Eq. 12.7 becomes

$$\begin{aligned}\langle\Omega|\,T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle\Big|_{x^0>y^0} &= \langle\Omega|\,\phi(0)\,|\Omega\rangle^2 \\ +\quad \sum_\lambda \left|\langle\Omega|\,\phi(0)\,|\lambda_0\rangle\right|^2 &\int \frac{d^3\vec{p}}{(2\pi)^3 2E_p}\; e^{-ip\cdot(x-y)}\Big|_{p^0=E_p}.\end{aligned} \tag{12.21}$$

In the same fashion, we obtain for the second time-ordering possibility

$$\begin{aligned}\langle\Omega|\,T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle\Big|_{y^0>x^0} &= \langle\Omega|\,\phi(0)\,|\Omega\rangle^2 \\ +\quad \sum_\lambda \left|\langle\Omega|\,\phi(0)\,|\lambda_0\rangle\right|^2 &\int \frac{d^3\vec{p}}{(2\pi)^3 2E_p}\; e^{-ip\cdot(y-x)}\Big|_{p^0=E_p}.\end{aligned} \tag{12.22}$$

The propagator is

$$\begin{aligned}\langle\Omega|\,T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle \;=\; &\langle\Omega|\,T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle\Big|_{x^0>y^0}\,\Theta(x^0-y^0) \\ &+\langle\Omega|\,T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle\Big|_{y^0>x^0}\,\Theta(y^0-x^0),\end{aligned} \tag{12.23}$$

which yields

$$\begin{aligned}\langle\Omega|\,T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle &= \langle\Omega|\,\phi(0)\,|\Omega\rangle^2 + \sum_\lambda \left|\langle\Omega|\,\phi(0)\,|\lambda_0\rangle\right|^2 \times \\ \times \int \frac{d^3\vec{p}}{(2\pi)^3 2E_p} &\left[e^{-ip\cdot(x-y)}\Theta(x^0-y^0) + e^{-ip\cdot(y-x)}\Theta(y^0-x^0)\right]_{p^0=E_p}.\end{aligned} \tag{12.24}$$

We recognise that the above integral is the Feynman-Stückelberg propagator of the free scalar field theory, Eq. 11.35. We denote by m_λ^2 the eigenvalue of the squared field-momentum P^2 corresponding to the state $|\lambda_0\rangle$,

$$P^2\,|\lambda_0\rangle = m_\lambda^2\,|\lambda_0\rangle\,. \tag{12.25}$$

Then, for the boosted state $\left|\lambda_p\right\rangle$, we have

$$P^2\left|\lambda_p\right\rangle = \left(H^2 - \vec{P}^2\right)\left|\lambda_p\right\rangle = \left(E_p^2 - \vec{p}^2\right)\left|\lambda_p\right\rangle. \tag{12.26}$$

Given that $\left|\lambda_p\right\rangle = U(\Lambda_p)\,|\lambda_0\rangle$, and P^2 is invariant under a Lorentz transformation, we obtain that

$$E_p(\lambda)^2 - \vec{p}^2 = m_\lambda^2. \tag{12.27}$$

Then, using the result of Eq. 11.35, we write

$$\langle\Omega|\, T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle = \langle\Omega|\,\phi(0)\,|\Omega\rangle^2 + \sum_\lambda \left|\langle\Omega|\,\phi(0)\,|\lambda_0\rangle\right|^2 \int \frac{d^4p}{(2\pi)^4}\frac{i}{p^2-m_\lambda^2+i\delta}e^{-ip\cdot(x-y)}. \tag{12.28}$$

This is a general result for the propagator of any scalar field in any quantum field theory which is symmetric under Poincaré (Lorentz and translation) transformations.

Let us now assume that there exists a density of states $|\lambda_0\rangle$, with $P^2|\lambda_0\rangle = m_\lambda^2|\lambda_0\rangle$. We can then replace the sum in the above expression with an integral,

$$\langle\Omega|\, T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle = \langle\Omega|\,\phi(0)\,|\Omega\rangle^2 + \int dM^2\rho(M^2)\left|\langle\Omega|\,\phi(0)\,|\lambda_0(M)\rangle\right|^2 \int \frac{d^4p}{(2\pi)^4}\frac{i}{p^2-M^2+i\delta}e^{-ip\cdot(x-y)}. \tag{12.29}$$

This is the Källén–Lehmann representation for the scalar field propagator. It is an amazing result for its simplicity, given its generality. Up to constants ($\langle\Omega|\,\phi(0)\,|\Omega\rangle^2$, $\left|\langle\Omega|\,\phi(0)\,|\lambda_0\rangle\right|^2$), the propagator is determined fully by the density of the energy eigenstates with zero space-momentum. It was derived independently by G. Källén (1952) and H. Lehmann (1954).

12.1.1 A special case: free scalar field theory

Our result for the Feynman-Stückelberg propagator can now be rederived as a special case of the Källén–Lehmann representation of Eq. 12.29 when considering the free Klein-Gordon field. In this case, we are able to find a simple expression for the quantum field $\phi(x)$,

$$\phi(x) = \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p}\left[a(p)e^{-ip\cdot x} + a^\dagger(p)e^{+ip\cdot x}\right], \tag{12.30}$$

with $\omega_p = \sqrt{\vec{p}^2+m^2}$. The vacuum state is

$$|\Omega\rangle = |0\rangle, \tag{12.31}$$

and it is annihilated by the ladder operator $a(p)$,

$$a(p)\,|0\rangle = \langle 0|\,a^\dagger(p) = 0. \tag{12.32}$$

We can easily see that

$$\langle 0|\,\phi(0)\,|0\rangle = 0. \tag{12.33}$$

We also need to compute the density of states $|\lambda_0\rangle$, with $H|\lambda_0\rangle = m_\lambda|\lambda_0\rangle$ and $\vec{P}|\lambda_0\rangle = 0$. In the case of the free Klein-Gordon field we can identify a number of particles for each eigenstate of the P^μ operator. We can build states of higher and higher energy with $\vec{p}=0$ by applying repeatedly the creation operator $a^\dagger(0)$ on the vacuum. We have

$$\begin{aligned} H\left(a^\dagger(0)\,|0\rangle\right) &= m\left(a^\dagger(0)\,|0\rangle\right), \\ H\left(a^\dagger(0)^2\,|0\rangle\right) &= 2m\left(a^\dagger(0)^2\,|0\rangle\right), \\ &\cdots \\ H\left(a^\dagger(0)^n\,|0\rangle\right) &= nm\left(a^\dagger(0)^n\,|0\rangle\right), \\ &\cdots \end{aligned}$$

The density of these states is then

$$\rho(M^2) = \sum_{n=1}^{\infty} \delta(M^2 - n^2\, m^2). \tag{12.34}$$

The states $a^\dagger(0)^n\,|0\rangle$ are, however, not the only zero-momentum states of the free theory. Two particles with momenta $\vec{k}$ and $-\vec{k}$ also have $\vec{P} = 0$ and $M = 2\sqrt{\vec{k}^2 + m^2}$, so already the free theory has a continuum of multi-particle states starting at $M^2 = 4m^2$; the density above is therefore incomplete. This does not affect the result below, because $\phi(0)$ connects the vacuum only to one-particle states, so only the $\delta(M^2 - m^2)$ term survives (Peskin–Schroeder, Sec. 7.1). The same remark explains the threshold $4m^2$ of the continuum in the interacting case: it is kinematic (relative motion of two particles), not caused by the interaction potential.

Finally, we require the constants

$$\langle 0|\,\phi(0)\,|\lambda_0\rangle \quad = \quad \langle 0|\,\phi(0)\Big(a^\dagger(0)^n\,|0\rangle\Big) = \delta_{n,1}. \tag{12.35}$$

Exercise: Prove the above.

We now substitute Eqs. 12.33, 12.34 and 12.35 into Eq. 12.29. We recover

$$\langle 0|\,T\left\{\phi(x)\phi(y)\right\}|0\rangle = \int \frac{d^4p}{(2\pi)^4}\frac{i}{p^2 - m^2 + i\delta}e^{-ip\cdot(x-y)}, \tag{12.36}$$

which is the familiar Feynman-Stückelberg propagator for the free Klein-Gordon field.

12.1.2 "Typical" interacting scalar field theory

It is not possible to compute the density $\rho(M^2)$ of zero-momentum ($\vec{p} = 0$) states $|\lambda_0\rangle$ in realistic field theories, where the Lagrangian contains potential terms in addition to the ones of the free scalar Klein-Gordon field. We can make, however, some reasonable assumptions which agree with observations on the systems that we would like to describe.

We have emphasised several times that in interacting theories it is not, in general, possible to identify a number of particles corresponding to a certain energy-momentum eigenstate. Nevertheless, we do observe particles which act as if they were free and isolated. Such a state is characterised by the mass m of the particle. If we now consider a state of two "approximately free" particles, the energy of the state when the combined momentum of the two particles is $\vec{p} = 0$ is also "approximately" twice the mass of one particle, $2\,m$. However, their interaction potential will not allow for this exact value, and, in general, their energy will be larger. States other than one-particle states form a continuum of masses M^2 which is larger than the minimum energy (corresponding to free particles) required to get two particles in the same state. An exception to this occurs when the two particles form a bound state, where their binding energy is negative and their effective mass is generally smaller than $2\,m$.

Although a very important issue, we will not analyse bound-state problems in this course. Then, a typical density is

$$\rho\left(M^2\right) = \delta(M^2 - m^2_{1-\text{particle}}) + \Theta(M^2 - 4m^2_{1-\text{particle}})\rho_{\text{cont}}(M^2), \tag{12.37}$$

where ρ_{cont} is a continuous function. Substituting this density functional form into the general expression of Eq. 12.29 for the scalar propagator we find,

$$\begin{aligned}
&\langle\Omega|\, T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle = \\
&\left|\langle\Omega|\,\phi(0)\left|1_{\text{particle}}\right\rangle\right|^2 \int \frac{d^4p}{(2\pi)^4}\frac{i}{p^2-m^2_{1_{\text{particle}}}+i\delta}e^{-ip\cdot(x-y)} \\
&+\int dM^2\rho_{\text{cont}}(M^2)\left|\langle\Omega|\,\phi(0)\,|\lambda_0(M)\rangle\right|^2\int\frac{d^4p}{(2\pi)^4}\frac{i}{p^2-M^2+i\delta}e^{-ip\cdot(x-y)} \\
&\quad+\langle\Omega|\,\phi(0)\,|\Omega\rangle^2\,.
\end{aligned} \tag{12.38}$$

The above result is very important: it states that one-particle states are responsible for poles in the propagator of an interacting field. If a theory has states which behave as isolated particles, then we expect to find poles in the propagator for squared momenta $p^2 = m^2_{1_{\text{particle}}}$ which correspond to the physical mass of the particles. In practice, it is then worth remembering that the scalar propagator is

$$\langle\Omega|\, T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle = \int \frac{d^4p}{(2\pi)^4}\frac{i\tilde{Z}_\phi e^{-ip\cdot(x-y)}}{p^2-m^2_{\text{phys}}+i\delta} + \text{continuum}, \tag{12.39}$$

where

$$\tilde{Z}_\phi = \left|\langle\Omega|\,\phi(0)\left|1_{\text{particle}}\right\rangle\right|^2, \tag{12.40}$$

is a constant, and m_{phys} is the physical mass of the particle.

Notice that m_{phys} is **not** the mass parameter m that we write in the Lagrangian. It depends on the parameters of the Lagrangian, but in general it is not equal to any of them. Similarly, $\tilde{Z}_\phi$ need not be one. Only in the free theory, as we have seen above, $\tilde{Z}_\phi = 1$ and $m_{\text{phys}} = m$. The continuum, on the other hand, does not give rise to poles. The integration over M^2 with a smooth density smears the pole $1/(p^2 - M^2)$ into a branch cut. For example, for a constant ρ_{cont} up to some large value Λ^2,

$$\int_{4m^2}^{\Lambda^2}\frac{dM^2}{p^2-M^2+i\delta} = \log\frac{4m^2-p^2-i\delta}{\Lambda^2-p^2-i\delta},$$

with $m \equiv m_{1_{\text{particle}}}$. This has no pole; it develops instead an imaginary part $(-i\pi)$ for $4m^2 < p^2 < \Lambda^2$. We will see that only the poles contribute to scattering amplitudes. Eq. 12.39 also tells us how we recognise a particle in practice. Suppose that we produce something at a point y and detect it at a point x. The Fourier transform of the propagation amplitude becomes the largest for p^2 close to m^2_{phys}, where it has a pole. Observing such a pole is what we mean by discovering a particle of mass m_{phys}.

12.2 Spectral assumptions in scattering theory

We are now ready to study using quantum field theory the scattering of elementary particles. Scattering processes occur in Lagrangian systems with interaction potentials, i.e. additional terms in the Lagrangian other than the ones that we have found in the quantisation of free fields. It has never been possible to find an exact solution for the Hamiltonian of any field theory which describes realistic particle interactions in four space-time dimensions. However, it is possible under some reasonable assumptions to develop a theory for scattering probability amplitudes. For simplicity, we shall carry out

our analysis for the case of interacting scalar fields. This is sufficient to demonstrate the salient features of the scattering theory.

Our basic set of assumptions is the following:

- The theory has a ground state $|\Omega\rangle$.
- There are "first excited" states $\left|p\right\rangle$ with

$$P^2 \left|p\right\rangle = m^2_{\text{phys}} \left|p\right\rangle , \tag{12.41}$$

  corresponding intuitively to "single-particle" states. We also assume that $m^2_{\text{phys}} \geq 0$, which means that there are no tachyons (particles which travel faster than light). (In 2011 the OPERA experiment reported neutrinos travelling faster than light. The measurement turned out to be faulty, and so far there is no evidence for tachyons in nature.)
- All remaining eigenstates $\left|q\right\rangle$ of the field momentum operator have

$$P^2 \left|q\right\rangle = M^2 \left|q\right\rangle , \tag{12.42}$$

  with

$$M^2 \geq 4m^2_{\text{phys}}. \tag{12.43}$$

  These states form a continuum, and they correspond, intuitively, to multi-particle excitations.
- The expectation value of the field operator in the ground state is zero,

$$\langle\Omega| \phi(x) |\Omega\rangle = 0. \tag{12.44}$$

Roughly, $\phi(x)|\Omega\rangle$ should correspond to a one-particle excited state. A single particle should not disappear spontaneously into a state with no particles $|\Omega\rangle$. The corresponding probability amplitude should therefore vanish,

$$\left\langle\Omega\right| \mathbf{1}_{\text{particle}}\left\rangle \sim \langle\Omega| \phi(x) |\Omega\rangle = 0.\right.$$

This argument should only be taken as a heuristic. Strictly speaking, the overlap $\left\langle\Omega\right| \mathbf{1}_{\text{particle}}\left\rangle\right.$ vanishes automatically, since both are eigenstates of P^μ with different eigenvalues. The condition $\langle\Omega| \phi(x) |\Omega\rangle = 0$ is a separate requirement: it ensures that the asymptotic field creates particles from the vacuum rather than having a constant piece, and it can always be achieved by the shift $\phi \to \phi - \langle\Omega| \phi |\Omega\rangle$ (Peskin–Schroeder, Sec. 7.2; Srednicki, Ch. 5).

12.3 "In" and "Out" states

As we have discussed earlier, states in interacting field theories no longer describe systems with a necessarily constant number of particles. Free field theories have a Hilbert space of states with definite particle number, but Hamiltonian eigenstates in interacting theories do not. Nevertheless, we do observe particles which behave as free (e.g. free electrons and photons) when they are sufficiently separated. By the uncertainty principle, such states describing particles which are isolated at certain times must be wave-packets.

Two classes of wave-packets are interesting in order to describe the scattering of freely moving particles:

- wave-packets which are isolated at a time very far in the past, $t = -\infty$, but may be overlapping at some finite time t;
- wave-packets which are isolated in the far future, $t = +\infty$, but may be overlapping at some finite time t.

We should remember that in our development of the field theory formalism, we are using the Heisenberg picture for time evolution. In this picture, field operators depend on time, while states are time-independent.[2] Time-independent states are perhaps counter-intuitive, and can mislead us into believing that they describe non-evolving systems. But this is of course not true. Measurements of expectation values of field operators on a state are time-dependent, and in this sense, states encapsulate the time-evolution of the physical system that they describe. A useful analogy is a DVD of a film. The disc itself never changes, but the film unfolds in time when we play it. Similarly, a Heisenberg state contains the whole history of a scattering event, which becomes visible when we act on it with time-dependent operators. We do not need to write the whole story on the label of the disc, since the evolution is dictated by the Lagrangian and not by a scriptwriter. It suffices to label the state by what the system looks like at the very beginning, or, alternatively, by what it looks like at the very end. These two labelling conventions lead to the "in" and "out" states that we define now. We define a state as an $|\text{in}\rangle$ state if it describes a system of freely moving particles in the very far past, which may or may not be interacting at a later time. We define a state as an $|\text{out}\rangle$ state if it describes a system of particles which may or may not be interacting at a finite time, but which are freely moving in the far future.

Consider an $|\text{in}\rangle$ state which describes a single particle with momentum p in the far past. We will assume that free particles are indestructible (stable) in the absence of other particles. Then this state will also describe a free single particle in the far future, and it will also be an $|\text{out}\rangle$ state:

$$\left|\text{in, single} - \text{particle, } p\right\rangle = \left|\text{out, single} - \text{particle, } p\right\rangle . \tag{12.45}$$

More complicated "in" and "out" states containing many free particles in the very far past or future respectively, allow for particle interactions at a finite time. What are the possible "in" and "out" states? We will make the following assumption:
"*Every possible state is a linear combination of either* $|\text{in}\rangle$ *or* $|\text{out}\rangle$ *states.*"
In physical terms, this means that every physical system will behave as a system of freely moving particles which are getting increasingly isolated if we wait long enough. Conversely, if we look further back in the far past, every physical system originates from a system of freely moving isolated particles.

The above assumption is in accordance with the fact that particle interactions occur when particles are brought very close together. Short-range interactions cannot keep particles together perpetually. There is a strikingly important "exception" to this assumption: elementary particles may form bound states. In practice, we know from experiment which bound states may be formed (hadrons, atoms, molecules, etc.). These are stable, and we can account for them by including them in a generalised definition of $|\text{in}\rangle$ and $|\text{out}\rangle$ states. Only stable particles and stable bound states (e.g. the proton, the electron, the photon,

[2] For example, in the free scalar field case the field has a solution which is explicitly time-dependent,

$$\phi(x) = \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p} \left[a(p)e^{-ip\cdot x} + a^\dagger(p)e^{+ip\cdot x}\right],$$

while states, $|0\rangle$, $\left|p\right\rangle = a^\dagger(p)\,|0\rangle$, . . ., are manifestly time-independent.

atoms) appear as asymptotic "in" and "out" states. Most hadrons, and the W, Z and Higgs bosons, are unstable; they are not asymptotic states but show up as resonances in the scattering of stable particles. In QCD, quarks and gluons never appear as asymptotic states at all (confinement); the asymptotic states are colour-neutral hadrons (Weinberg, QFT I, Ch. 3).

We adopt the following notation. We denote by

$$\left|\{p_i\}, \text{in}\right\rangle$$

an "in" state where the i^{th} particle has momentum p_i. We denote by

$$\left|\{k_j\}, \text{out}\right\rangle$$

an "out" state where the j^{th} particle has momentum k_j. These states are complete. Every state can be written as a linear combination of either of them, and we have

$$\sum \left|\{p_i\}, \text{in}\right\rangle \left\langle \text{in}, \{p_i\}\right| = \sum \left|\{k_j\}, \text{out}\right\rangle \left\langle \text{out}, \{k_j\}\right| = \mathbf{1}. \tag{12.46}$$

12.4 Scattering matrix elements

Obviously, an $\left|\{p_i\}, \text{in}\right\rangle$ state can be written as a superposition of $\left|\{k_j\}, \text{out}\right\rangle$ states. This physically means that a system of originally free particles will evolve to a system of ultimately free particles. If these particles never interact, the "in" state is the same as the "out" state. If they do, then we may have isolated particles after the interaction with different momenta or even a different number of them. We write

$$\left|\{p_i\}, \text{in}\right\rangle = \sum_{\{k_j\}} c\left(\{k_j\}, \{p_i\}\right) \left|\{k_j\}, \text{out}\right\rangle . \tag{12.47}$$

The probability amplitude for the transition of a system of $i = 1 \ldots N$ incoming particles with momenta $\{p_i\}$ to a system of $j = 1 \ldots M$ outgoing particles with momenta $\{k_j\}$ is known as the "scattering matrix element",

$$S_{ji} \equiv c\left(\{k_j\}, \{p_i\}\right) = \left\langle \text{out}, \{k_j\}\right| \left\{p_i\}, \text{in}\right\rangle . \tag{12.48}$$

The scattering probability is

$$P\left(\{p_i\} \to \{k_j\}\right) = \left|S_{ji}\right|^2 = \left|\left\langle \text{out}, \{k_j\}\right| \left\{p_i\}, \text{in}\right\rangle\right|^2 . \tag{12.49}$$

The scattering matrix elements constitute the S-matrix, which is a matrix acting on the space of "in" states and transforms them into "out" states,

$$S \equiv S_{ij}, \qquad |\alpha, \text{in}\rangle = \sum_\beta S_{\beta\alpha} \left|\beta, \text{out}\right\rangle . \tag{12.50}$$

The S-matrix is unitary,

$$\begin{aligned} \left(S^\dagger S\right)_{\alpha\gamma} = \sum_\beta S^\dagger_{\alpha\beta} S_{\beta\gamma} &= \sum_\beta \langle \alpha, \text{in}| \text{ out}, \beta\rangle \left\langle \beta, \text{out}\right| \gamma, \text{in}\rangle \\ &= \langle \alpha, \text{in}| \left(\sum_\beta \left|\beta, \text{out}\right\rangle \left\langle \text{out}, \beta\right| \right) |\gamma, \text{in}\rangle = \delta_{\alpha\gamma}. \end{aligned} \tag{12.51}$$

We are usually not interested in computing the probability of not having a scattering. The matrix elements of the S-matrix which are relevant for a true scattering are the non-diagonal ones. We then define the so-called "transition matrix" T, which is given by

$$S \equiv \mathbf{1} + \mathbf{i}\mathbf{T}. \tag{12.52}$$

In the rest of this chapter we shall prove a fundamental relation of S-matrix elements to expectation values of field operators in the ground state of the interacting theory.

12.5 S-matrix and Green's functions

In the last section, we established a few properties of the S-matrix which, at first sight, offer little help in computing it explicitly. It is possible, however, to extract S-matrix elements from Green's functions, which are defined as

$$G(x_1, x_2, \ldots, x_N) \equiv \langle\Omega| T\left\{\phi(x_1)\phi(x_2)\ldots\phi(x_N)\right\}|\Omega\rangle . \tag{12.53}$$

In the simplest case of one space-time point,

$$\langle\Omega|\phi(x)|\Omega\rangle = 0, \tag{12.54}$$

we have assumed that the corresponding Green's function vanishes, as part of our spectral assumptions for an interacting field theory.

The two-point Green's function is our familiar propagator, for which we have found the Källén–Lehmann representation

$$G(x,y) = \int \frac{d^4p}{(2\pi)^4} \frac{i\tilde{Z}_\phi e^{-ip\cdot(x-y)}}{p^2 - m^2_{\text{phys}} + i\delta} + \text{continuum}. \tag{12.55}$$

Consider the explicit case with $x^0 > \tau > -\tau > y^0$ with $\tau \to \infty$, for two space-time points where one is very far in the future and the second very far in the past. Then,

$$\begin{aligned}
G(x,y)\big|_{\substack{x^0\to+\infty \\ y^0\to-\infty}} &= \langle\Omega| T\left\{\phi(x)\phi(y)\right\}|\Omega\rangle \\
&= \langle\Omega|\left\{\phi(x)\phi(y)\right\}|\Omega\rangle \\
&= \langle\Omega|\left\{\phi(x)\mathbf{1}\,\mathbf{1}\,\phi(y)\right\}|\Omega\rangle \\
&= \langle\Omega|\phi(x)\left(\sum_{\text{out}} |\text{out}\rangle\langle\text{out}|\right)\left(\sum_{\text{in}} |\text{in}\rangle\langle\text{in}|\right)\phi(y)|\Omega\rangle \\
\rightsquigarrow\ G(x,y)\big|_{\substack{x^0\to+\infty \\ y^0\to-\infty}} &= \sum_{\text{in}}\sum_{\text{out}} \langle\Omega|\phi(x)|\text{out}\rangle\langle\text{out}|\,\text{in}\rangle\langle\text{in}|\phi(y)|\Omega\rangle .
\end{aligned} \tag{12.56}$$

We have defined an $|\text{out}\rangle$ state as any state which behaves as a state of any number of free isolated particles in the far future. The sum over all "out" states then corresponds to

$$\begin{aligned}
&\sum_{out} \langle\Omega|\phi(x)|\text{out}\rangle\langle\text{out}| \to \langle\Omega|\phi(x)|\Omega\rangle\langle\Omega| \\
+ &\int \frac{d^3\vec{p}_1}{(2\pi)^3 2\omega_{p_1}} \langle\Omega|\phi(x)\left|p_1;\text{out}\right\rangle\left\langle p_1;\text{out}\right| \\
+ &\int \frac{d^3\vec{p}_1}{(2\pi)^3 2\omega_{p_1}} \frac{d^3\vec{p}_2}{(2\pi)^3 2\omega_{p_2}} \langle\Omega|\phi(x)\left|p_1,p_2;\text{out}\right\rangle\left\langle p_1,p_2;\text{out}\right| \\
+ &\ \ldots .
\end{aligned} \tag{12.57}$$

We now make use of the property of "out" states that they behave as states of free particles at large times $x^0 > \tau$, and compute $\langle\Omega|\,\phi(x)\,|\text{out}\rangle$ in free field theory. In free field theory, the state $\phi(x)\,|\Omega\rangle$ corresponds to a state with exactly one particle, and it has no overlap with states of a different particle multiplicity. We then have

$$\langle\Omega|\,\phi(x)\,|\Omega\rangle = \langle\Omega|\,\phi(x)\,\big|p_1,p_2;\text{out}\big\rangle = \langle\Omega|\,\phi(x)\,\big|p_1,p_2,p_3;\text{out}\big\rangle = \ldots = 0, \tag{12.58}$$

and only

$$\langle\Omega|\,\phi(x)\,\big|p_1;\text{out}\big\rangle \neq 0. \tag{12.59}$$

We have already computed this non-vanishing matrix element up to a constant. From Eqs. 12.18 and 12.40 we have

$$\langle\Omega|\,\phi(x)\,\big|p;\text{out}\big\rangle = \sqrt{\tilde{Z}_\phi}\,e^{-ip\cdot x}. \tag{12.60}$$

We compute the sum over "in" states in Eq. 12.56 in an identical fashion. Then we arrive at a simple result,

$$G(x,y)\Big|_{\substack{x^0\to+\infty\\ y^0\to-\infty}} = \tilde{Z}_\phi \int \frac{d^3\vec{p}}{(2\pi)^3 2\omega_p}\frac{d^3\vec{q}}{(2\pi)^3 2\omega_q} e^{-i(p\cdot x - q\cdot y)}\,\big\langle p;\text{out}\big|\,q;\text{in}\big\rangle\,. \tag{12.61}$$

Eq. 12.61 relates the Green's function for two points well separated in time to the transition amplitude from an "in" single-particle state to an "out" single-particle state. How about more complicated Green's functions and transition amplitudes? We can repeat the same procedure for a general Green's function

$$G(x_1,\ldots,y_1\ldots)\Big|_{\substack{x_i^0\to+\infty\\ y_i^0\to-\infty}} = \langle\Omega|\,T\left\{\phi(x_1)\ldots\phi(y_1)\ldots\right\}|\Omega\rangle_{\substack{x_i^0\to+\infty\\ y_i^0\to-\infty}}\,, \tag{12.62}$$

with N_{out} points at far-future times x_i^0 and N_{in} points y_j at far-past times. We find that

$$\begin{aligned} G(x_1,\ldots,y_1\ldots)\Big|_{\substack{x_i^0\to+\infty\\ y_i^0\to-\infty}} &= \int \left(\prod_{i=1}^{N_{\text{out}}}\frac{d^3\vec{p}_i\;e^{-ip_i\cdot x_i}}{(2\pi)^3 2\omega_{p_i}}\right)\left(\prod_{j=1}^{N_{\text{in}}}\frac{d^3\vec{q}_j\;e^{+iq_j\cdot y_j}}{(2\pi)^3 2\omega_{q_j}}\right) \\ &\times\tilde{Z}_\phi^{\frac{N_{\text{in}}+N_{\text{out}}}{2}}\,\big\langle p_1,\ldots,p_{N_{\text{out}}};\text{out}\big|\,q_1,\ldots q_{N_{\text{in}}};\text{in}\big\rangle\,. \end{aligned} \tag{12.63}$$

We arrive at a very useful result. Green's functions with $N_{\text{in}} + N_{\text{out}}$ points, with N_{in} of them chosen in the far past and N_{out} in the far future, are, up to an overall constant, some sort of a Fourier transform of S-matrix elements for the scattering of N_{in} particles to N_{out} particles. In the next section we shall use Eq. 12.63 to compute the matrix elements.

12.6 The LSZ reduction formula

We will compute scattering matrix elements by inverting Eq. 12.63. This requires a type of Fourier transformation on the space-time coordinates x_i^μ and y_j^μ. The integration over the space components, $\vec{x}_i$ and $\vec{y}_j$, can be unrestricted (from $-\infty$ to $+\infty$), but we need to be careful concerning the time integrations, since we have made assumptions about x_i^0 being far-future and y_j^0 far-past times.

Let us try to integrate x_i^μ and y_j^μ over the maximum space-time region where we do not contradict our time-ordering assumptions. Consider the integral

$$I\left[\{k_i\},\{l_j\}\right] \equiv \left(\prod_i \int_\tau^\infty dx_i^0 \int d^3\vec{x}_i e^{ik_i\cdot x_i}\right)\left(\prod_j \int_{-\infty}^{-\tau} dy_j^0 \int d^3\vec{y}_j e^{-il_j\cdot y_j}\right) \times G(x_1,\ldots,y_1,\ldots). \tag{12.64}$$

We now substitute the expression for the Green's function that we have found in Eq. 12.63, and perform the space integrations using the identity

$$\int d^3\vec{x}\, e^{i\vec{k}\cdot\vec{x}} = (2\pi)^3\delta^{(3)}(\vec{k}). \tag{12.65}$$

The integral of Eq. 12.64 becomes

$$I\left[\{k_i\},\{l_j\}\right] \equiv \left(\prod_i \int_\tau^\infty dx_i^0 \frac{e^{i(k_i^0-\omega(\vec{k}_i))x_i^0}}{2\omega(\vec{k}_i)}\right)\left(\prod_j \int_{-\infty}^{-\tau} dy_j^0 \frac{e^{-i(l_j^0-\omega(\vec{l}_j))y_j^0}}{2\omega(\vec{l}_j)}\right) \times \tilde{Z}_\phi^{\frac{N_{\text{in}}+N_{\text{out}}}{2}} \left\langle k_1,\ldots,k_{N_{\text{out}}};\text{out}\right| l_1,\ldots l_{N_{\text{in}}};\text{in}\rangle\,. \tag{12.66}$$

The integrals over the time variables are also easy to perform,

$$\int_\tau^\infty dt\, e^{it\left(l^0-\omega(\vec{l})\right)} = \frac{-i}{l^0-\omega(\vec{l})}\left[\lim_{T\to\infty} e^{iT\left(l^0-\omega(\vec{l})\right)} - e^{i\tau\left(l^0-\omega(\vec{l})\right)}\right]. \tag{12.67}$$

The above is problematic due to the term which requires a limit at infinity that is not well defined. Suppose now that we modify the time integration in our Fourier transformations and make it slightly complex:

$$\lim_{\delta\to 0}\int_\tau^\infty dt\, e^{it\left(l^0(1+i\delta)-\omega(\vec{l})\right)} = \frac{i}{l^0-\omega(\vec{l})} e^{i\tau\left(l^0-\omega(\vec{l})\right)}. \tag{12.68}$$

Then, we can write

$$I\left[\{k_i\},\{l_j\}\right] \equiv \left(\prod_i \frac{i}{(k_i^0-\omega(\vec{k}_i))2\omega(\vec{k}_i)} e^{i\tau\left(k_i^0-\omega(\vec{k}_i)\right)}\right) \times\left(\prod_j \frac{i}{(l_j^0-\omega(\vec{l}_j))2\omega(\vec{l}_j)} e^{i\tau\left(l_j^0-\omega(\vec{l}_j)\right)}\right) \times \tilde{Z}_\phi^{\frac{N_{\text{in}}+N_{\text{out}}}{2}} \left\langle k_1,\ldots,k_{N_{\text{out}}};\text{out}\right| l_1,\ldots l_{N_{\text{in}}};\text{in}\rangle\,. \tag{12.69}$$

We have now managed to relate S-matrix elements for the transition of a system with N_{in} particles with momenta $\{l_j^\mu\}$ to a system with N_{out} particles with momenta $\{k_i\}$ to Fourier-type integrals over Green's functions.

Let us analyse Eq. 12.69 further. We observe that a "Fourier transformed" Green's function on the lhs has poles at

$$l_i^0 = \omega(\vec{l}_i) = \sqrt{\vec{l}_i^2 + m_{\text{phys}}^2},$$

due to every particle in the "in" and "out" states. We also notice that the poles are independent of the time τ, since the τ-dependent exponential becomes one at the pole position $l_0 = \omega$. Indeed, we can expand

$$\frac{e^{i\tau(l_0-\omega)}}{2\omega(l_0-\omega)} = \frac{1}{2\omega(l_0-\omega)} + \frac{i\tau}{2\omega} + O((l_0-\omega)\tau^2). \tag{12.70}$$

This means that, had we decided to put no restrictions on the time integrations for $I\left[\{k_i\},\{l_j\}\right]$, the coefficients of the poles (their residues) would be unchanged. We can then write

$$\begin{aligned} & \left(\text{ Regular at } l_i^0, k_j^0 \to \omega(\vec{l}_i), \omega(\vec{k}_j)\right) \\ + & \left(\prod \frac{i\tilde{Z}_\phi^{\frac{1}{2}}}{2\omega(\vec{l}_i)(l_i^0-\omega(\vec{l}_i))}\right)\left(\prod \frac{i\tilde{Z}_\phi^{\frac{1}{2}}}{2\omega(\vec{k}_j)(k_j^0-\omega(\vec{k}_j))}\right)\left\langle k_1,\ldots,k_{N_{\text{out}}};\text{out}\right|\left. l_1,\ldots l_{N_{\text{in}}};\text{in}\right\rangle \\ = & \ \tilde{G}\left(k_1,\ldots,k_{N_{\text{out}}};l_1,\ldots l_{N_{\text{in}}}\right), \end{aligned} \tag{12.71}$$

where the Green's function $\tilde{G}$ in momentum space is defined as the ordinary Fourier transform of the Green's function in x-space,

$$\begin{aligned} \tilde{G}\left(k_1,\ldots,k_{N_{\text{out}}};l_1,\ldots l_{N_{\text{in}}}\right) = & \int_{-\infty}^{+\infty} d^4x_1\ldots d^4x_{N_{\text{out}}}d^4y_1\ldots d^4y_{N_{\text{in}}}e^{i\left[\sum_i x_i\cdot k_i-\sum_j y_j\cdot l_j\right]} \\ & \times\langle\Omega|\,T\left\{\phi(x_1)\ldots\phi(x_{N_{\text{out}}})\phi(y_1)\ldots\phi(y_{N_{\text{in}}})\right\}|\Omega\rangle\,. \end{aligned} \tag{12.72}$$

In the above integration, the exponential factors must be slightly deformed[3] in the complex plane,

$$e^{ik\cdot x} \to e^{ik^0x^0(1+i\delta)-i\vec{k}\cdot\vec{x}}. \tag{12.75}$$

We now multiply Eq. 12.71 by a factor

$$\frac{l^2-m_{\text{phys}}^2}{i\sqrt{\tilde{Z}_\phi}}$$

for each "in" state and "out" state particle, and take the limit $l^2-m_{\text{phys}}^2 = (l^0-\omega)(l^0+\omega) \to 0$. The regular term on the lhs of the equation vanishes in this limit. We are left with the LSZ reduction formula,

$$\begin{aligned} \left\langle k_1,\ldots,k_{N_{\text{out}}};\text{out}\right|\left. l_1,\ldots l_{N_{\text{in}}};\text{in}\right\rangle = & \lim_{\substack{l_i^2=m_{\text{phys}}^2\\ k_j^2=m_{\text{phys}}^2}}\left[\prod_i \frac{l_i^2-m_{\text{phys}}^2}{i\sqrt{\tilde{Z}_\phi}}\right]\left[\prod_j \frac{k_j^2-m_{\text{phys}}^2}{i\sqrt{\tilde{Z}_\phi}}\right] \\ & \times\tilde{G}\left(k_1,\ldots,k_{N_{\text{out}}};l_1,\ldots l_{N_{\text{in}}}\right), \end{aligned} \tag{12.76}$$

[3] Alternatively, we can leave the exponentials intact and perform a real time-integration in Eq. 12.72, but then we have to evaluate the kernel Green's function at deformed space-time points,

$$\begin{aligned} \tilde{G}\left(k_1,\ldots,k_{N_{\text{out}}};l_1,\ldots l_{N_{\text{in}}}\right) = & \int_{-\infty}^{+\infty} d^4x_1\ldots d^4x_{N_{\text{out}}}d^4y_1\ldots d^4y_{N_{\text{in}}}e^{i\left[\sum_i x_i\cdot k_i-\sum_j y_j\cdot l_j\right]} \\ & \times\langle\Omega|\,T\left\{\phi(\tilde{x}_1)\ldots\phi(\tilde{x}_{N_{\text{out}}})\phi(\tilde{y}_1)\ldots\phi(\tilde{y}_{N_{\text{in}}})\right\}|\Omega\rangle\,, \end{aligned} \tag{12.73}$$

where

$$\tilde{x}^\mu = (x^0(1-i\delta),\vec{x}). \tag{12.74}$$

which is due to Lehmann, Symanzik and Zimmermann.

Further reading: the asymptotic limits $t \to \pm\infty$ underlying the LSZ formula are *weak* limits: they hold for matrix elements of the fields between normalisable states (wave packets), not as operator identities for the fields themselves. A careful modern derivation along these lines is given in J. Collins, "A new approach to the LSZ reduction formula", arXiv:1904.10923 (2019).

12.7 Truncated Green's functions

Let us recall here the Källén–Lehmann representation of the propagator of Eq. 12.39, which we can invert by applying a Fourier transformation. We find

$$\tilde{G}(p) = \frac{i\tilde{Z}_\phi}{p^2 - m^2_{\text{phys}}} + \left(\text{Regular at } p^2 \to m^2_{\text{phys}}\right), \tag{12.77}$$

where we have defined the propagator in momentum space as

$$\tilde{G}(p) \equiv \int d^4x\, e^{ip\cdot x} \langle\Omega| T\left\{\phi(x)\phi(0)\right\} |\Omega\rangle . \tag{12.78}$$

We can then cast the S-matrix from the LSZ formula as

$$\left\langle k_1, \ldots, k_{N_{\text{out}}}; \text{out}\right| \left. l_1, \ldots l_{N_{\text{in}}}; \text{in}\right\rangle = \tilde{Z}_\phi^{\frac{N_{\text{in}}+N_{\text{out}}}{2}} \left.\frac{\tilde{G}\left(k_1, \ldots, k_{N_{\text{out}}}; l_1, \ldots l_{N_{\text{in}}}\right)}{\left(\prod_i \tilde{G}(k_i)\right)\left(\prod_j \tilde{G}(l_j)\right)}\right|_{\substack{k_i^2 \to m^2_{\text{phys}} \\ l_j^2 \to m^2_{\text{phys}}}} . \tag{12.79}$$

A Green's function which is divided by a propagator for each external particle in the "in" and "out" states is called truncated:

$$\tilde{G}_{\text{trunc}}\left(k_1, \ldots, k_{N_{\text{out}}}; l_1, \ldots l_{N_{\text{in}}}\right) \equiv \frac{\tilde{G}\left(k_1, \ldots, k_{N_{\text{out}}}; l_1, \ldots l_{N_{\text{in}}}\right)}{\left(\prod_i \tilde{G}(k_i)\right)\left(\prod_j \tilde{G}(l_j)\right)}. \tag{12.80}$$

The LSZ formula takes the form

$$\begin{aligned}\left\langle k_1, \ldots, k_{N_{\text{out}}}; \text{out}\right| \left. l_1, \ldots l_{N_{\text{in}}}; \text{in}\right\rangle = \tilde{Z}_\phi^{\frac{N_{\text{in}}+N_{\text{out}}}{2}} \\ \left.\tilde{G}_{\text{trunc}}\left(k_1, \ldots, k_{N_{\text{out}}}; l_1, \ldots l_{N_{\text{in}}}\right)\right|_{\substack{k_i^2 \to m^2_{\text{phys}} \\ l_j^2 \to m^2_{\text{phys}}}} .\end{aligned} \tag{12.81}$$

We now have an expression for the probability amplitudes for the scattering of free particles in quantum field theory in terms of truncated Green's functions. As we have discussed already, there is no realistic field theory in four dimensions where all such probabilities can be computed exactly. In the next chapter, we shall resort to perturbation theory, which turns out to be an amazingly powerful tool.

12.8 Cross-sections*

What do we measure in a scattering experiment? We do not measure probability amplitudes. At a collider we count events: how many times two beams of particles produce a certain final state per unit time, or, for an unstable particle, how many of them decay per unit time.

For example, the experiments at the Large Hadron Collider measure the **cross-sections** for the production of Higgs bosons and compare them with the predictions of quantum field theory. In this section we will connect such measurable quantities, **cross-sections** and **decay rates**, with the S-matrix elements that we can now compute with the LSZ formula and perturbation theory.

12.8.1 Transition probabilities and the normalisation of states

Recall that our one-particle states are normalised as (see the commutation relations of the Klein-Gordon field)

$$\langle p | \, q \rangle = 2\omega_p \, (2\pi)^3 \delta^{(3)}(\vec{p} - \vec{q}), \qquad \omega_p = \sqrt{\vec{p}^{\,2} + m^2}. \tag{12.82}$$

The scattering probability $|S_{ji}|^2$ that we wrote down earlier in this chapter tacitly assumed states normalised to one. For our states, the probability for a transition from an initial state $|i\rangle$ to a final state $|f\rangle$ is

$$P(i \to f) = \frac{\left| \langle f | \, S \, | i \rangle \right|^2}{\langle f | \, f \rangle \, \langle i | \, i \rangle}. \tag{12.83}$$

We have divided by the norms of the states, since our states are not normalised to one. However, the norm of a momentum eigenstate is infinite,

$$\langle p | \, p \rangle = 2\omega_p \, (2\pi)^3 \delta^{(3)}(\vec{0}) = \infty \, !$$

This is not a mathematical accident. A state of definite momentum is spread over the whole space and over all times. Two such particles always "overlap" and they "collide" forever. Physical particles in a beam are wave-packets, which are localised in space and have a small spread of momenta. The rigorous way is to build the initial state out of normalisable wave-packets, compute the probability, and take the limit of sharply peaked momenta at the end (see Peskin-Schroeder, Sec. 4.5). We will follow a shortcut, which gives the same answer. We put the system in a large box of volume V and let it evolve for a large time T. Then the δ-functions at zero momentum become

$$(2\pi)^3 \delta^{(3)}(\vec{0}) = \int d^3\vec{x} \, e^{i\vec{0}\cdot\vec{x}} = V, \qquad (2\pi)^4 \delta^{(4)}(0) = \int d^4x \, e^{i0\cdot x} = VT. \tag{12.84}$$

The norm of a one-particle state, $\langle p | \, p \rangle = 2\omega_p V$, is now finite. We will take $V, T \to \infty$ at the end, and physical quantities should not depend on V and T.

We are interested in a true scattering, $|f\rangle \neq |i\rangle$, for which only the transition matrix in $S = \mathbf{1} + iT$ contributes. Space-time translation invariance guarantees that the total momentum is conserved. We can therefore always factor out a δ-function and define the **scattering amplitude** $\mathcal{M}$,

$$\langle f | \, iT \, | i \rangle = i \, \mathcal{M}(i \to f) \, (2\pi)^4 \delta^{(4)} \left(P_f - P_i \right), \tag{12.85}$$

where P_i and P_f are the total four-momenta of the initial and final states. With the LSZ formula of Eq. 12.81, $i\mathcal{M}$ is the truncated Green's function, times $\tilde{Z}_\phi^{(N_{\text{in}}+N_{\text{out}})/2}$, with the overall $(2\pi)^4\delta^{(4)}$ removed. In perturbation theory, it is given by the sum of connected and amputated Feynman diagrams of the next chapter, without the overall momentum-conserving δ-function.

The square of the amplitude contains the square of a δ-function. Using Eq. 12.84, we have

$$\left[(2\pi)^4\delta^{(4)}\left(P_f - P_i\right)\right]^2 = (2\pi)^4\delta^{(4)}\left(P_f - P_i\right)\,(2\pi)^4\delta^{(4)}(0) = (2\pi)^4\delta^{(4)}\left(P_f - P_i\right)\,VT. \quad (12.86)$$

Consider now N_{in} particles with momenta $p_1, \ldots, p_{N_{\text{in}}}$ scattering into n particles with momenta $k_1, \ldots, k_n$. In a box, the momenta are discrete, and the number of states with momentum in a cell $d^3\vec{k}$ is $V d^3\vec{k}/(2\pi)^3$. Summing Eq. 12.83 over all final states in the cells $d^3\vec{k}_1 \ldots d^3\vec{k}_n$, we obtain

$$\begin{aligned} dP &= \frac{|\mathcal{M}|^2\,(2\pi)^4\delta^{(4)}\left(P_f - P_i\right)\,VT}{\prod_{a=1}^{N_{\text{in}}}\left(2\omega_{p_a}V\right)}\prod_{j=1}^{n}\frac{V d^3\vec{k}_j}{(2\pi)^3\,2\omega_{k_j}V} \\ &= \frac{T\,V^{1-N_{\text{in}}}}{\prod_{a=1}^{N_{\text{in}}}2\omega_{p_a}}\,|\mathcal{M}|^2\,d\Phi_n. \end{aligned} \quad (12.87)$$

Notice that the volume factors of the final-state particles have cancelled. We have defined the **n-body phase space**,

$$d\Phi_n = \left(\prod_{j=1}^{n}\frac{d^3\vec{k}_j}{(2\pi)^3\,2\omega_{k_j}}\right)(2\pi)^4\delta^{(4)}\left(p_1 + \ldots + p_{N_{\text{in}}} - k_1 - \ldots - k_n\right). \quad (12.88)$$

It is Lorentz invariant, since each one-particle measure can be written as

$$\frac{d^3\vec{k}}{(2\pi)^3\,2\omega_k} = \frac{d^4k}{(2\pi)^4}\,2\pi\,\delta\left(k^2 - m^2\right)\theta(k^0), \quad (12.89)$$

and d^4k, $\delta(k^2 - m^2)$, $\theta(k^0)$ and $\delta^{(4)}$ are all invariant under (orthochronous) Lorentz transformations. The probability of Eq. 12.87 grows linearly with the time T. The meaningful quantity is the transition rate: the probability per unit time, dP/T.

12.8.2 Decay rates

For one particle A in the initial state, $N_{\text{in}} = 1$, the volume drops out of Eq. 12.87. The probability per unit time that A decays into the n particles is the **differential decay rate**,

$$d\Gamma = \frac{1}{2E_A}\,|\mathcal{M}(A \to k_1 \ldots k_n)|^2\,d\Phi_n. \quad (12.90)$$

In the rest frame of A, $E_A = m_A$. In a frame where A moves, $E_A = \gamma m_A$ and the rate is smaller by the factor $1/\gamma$. This is the time dilation of special relativity! The total decay rate is the sum over all final states (decay channels) which are allowed, $\Gamma = \sum_{\text{channels}}\int d\Gamma$, and the lifetime of the particle in its rest frame is $\tau = 1/\Gamma$.

Strictly speaking, an unstable particle is not an asymptotic "in" state, as we remarked earlier in this chapter. Eq. 12.90 makes sense for a particle which lives long enough, $\Gamma \ll m_A$, so that it can be treated as a free particle before it decays.

12.8.3 Cross-sections

For two particles A and B in the initial state, $N_{\text{in}} = 2$, Eq. 12.87 gives the rate

$$\frac{dP}{T} = \frac{1}{V}\,\frac{|\mathcal{M}|^2}{2E_A\,2E_B}\,d\Phi_n. \quad (12.91)$$

It depends on the volume, and this is reasonable. We have one particle A and one particle B in the box, and the larger the box, the less likely it is that they meet. What is independent of the box? Think of B as a target and of A as a projectile. The density of A in the box is $1/V$, and in the rest frame of B it moves with velocity $|\vec{v}_A - \vec{v}_B|$. The flux of A particles onto B, i.e. the number of particles crossing a unit area per unit time, is then $|\vec{v}_A - \vec{v}_B|/V$. We define the **cross-section** σ as the rate divided by the incoming flux,

$$\frac{dP}{T} \equiv d\sigma \times \frac{|\vec{v}_A - \vec{v}_B|}{V}. \tag{12.92}$$

It has the dimension of an area. Classically, it is the area of the target which the projectile must hit in order to scatter. We find

$$d\sigma = \frac{|\mathcal{M}(AB \to k_1 \dots k_n)|^2}{2E_A\, 2E_B\, |\vec{v}_A - \vec{v}_B|}\, d\Phi_n. \tag{12.93}$$

The factor in the denominator is the **flux factor**. For collinear beams, $\vec{p}_A \parallel \vec{p}_B$ (as in every collider, or for a target at rest), we can write it in a Lorentz-invariant form. With $\vec{v} = \vec{p}/E$, we have

$$E_A E_B |\vec{v}_A - \vec{v}_B| = \left|E_B\, \vec{p}_A - E_A\, \vec{p}_B\right| = \sqrt{(p_A \cdot p_B)^2 - m_A^2 m_B^2}. \tag{12.94}$$

We then write the cross-section as

$$d\sigma = \frac{|\mathcal{M}|^2}{4\sqrt{(p_A \cdot p_B)^2 - m_A^2 m_B^2}}\, d\Phi_n, \tag{12.95}$$

which is manifestly invariant under boosts along the beam axis. For non-collinear momenta, Eq. 12.95 is the definition of the cross-section that is used in practice. It corresponds to Eq. 12.93 with $|\vec{v}_A - \vec{v}_B|$ replaced by the so-called Møller velocity,

$$E_A E_B\, v_{\text{Møl}} \equiv \sqrt{(p_A \cdot p_B)^2 - m_A^2 m_B^2}, \qquad v_{\text{Møl}} = \sqrt{\left(\vec{v}_A - \vec{v}_B\right)^2 - \left(\vec{v}_A \times \vec{v}_B\right)^2}. \tag{12.96}$$

In terms of $s = (p_A + p_B)^2$, we have $p_A \cdot p_B = (s - m_A^2 - m_B^2)/2$ and the flux factor reads

$$4\sqrt{(p_A \cdot p_B)^2 - m_A^2 m_B^2} = 2\sqrt{\lambda(s, m_A^2, m_B^2)} = \begin{cases} 4\,|\vec{p}_{\text{cm}}|\sqrt{s}, & \text{(centre-of-mass frame)} \\ 4\, m_B\, |\vec{p}_A^{\,\text{lab}}|, & \text{(laboratory frame, } B \text{ at rest)} \\ 2s, & (m_A = m_B = 0) \end{cases} \tag{12.97}$$

where $\lambda(x, y, z) = x^2 + y^2 + z^2 - 2xy - 2yz - 2zx$ is the Källén function and $\vec{p}_{\text{cm}}$ is the momentum of either incoming particle in the centre-of-mass frame.

Exercise: *Prove the three forms of the flux factor in Eq. 12.97, and the expression of the Møller velocity in terms of $\vec{v}_A$ and $\vec{v}_B$.*

12.8.4 Two-body phase space

The simplest and most frequent case is a final state with two particles, of masses m_1 and m_2. In the centre-of-mass frame, $\vec{p}_A + \vec{p}_B = 0$ and $E_A + E_B = \sqrt{s}$. We perform the $d^3\vec{k}_2$ integration with the spatial δ-function, which sets $\vec{k}_2 = -\vec{k}_1$. Writing $d^3\vec{k}_1 = |\vec{k}|^2 d|\vec{k}|\, d\Omega$, we have

$$d\Phi_2 = \frac{|\vec{k}|^2 d|\vec{k}|\, d\Omega}{(2\pi)^3\, 4\omega_1\omega_2}\, 2\pi\, \delta\left(\sqrt{s} - \omega_1 - \omega_2\right), \qquad \omega_{1,2} = \sqrt{\vec{k}^2 + m_{1,2}^2}. \tag{12.98}$$

Since

$$\frac{d(\omega_1+\omega_2)}{d|\vec{k}|} = \frac{|\vec{k}|}{\omega_1} + \frac{|\vec{k}|}{\omega_2} = \frac{|\vec{k}|\,(\omega_1+\omega_2)}{\omega_1\omega_2},$$

the energy δ-function yields

$$d\Phi_2 = \frac{|\vec{k}_f|}{16\pi^2\sqrt{s}}\,d\Omega, \qquad |\vec{k}_f| = \frac{\sqrt{\lambda(s,m_1^2,m_2^2)}}{2\sqrt{s}}, \tag{12.99}$$

where $d\Omega$ is the solid angle of particle 1. For massless particles, $|\vec{k}_f| = \sqrt{s}/2$, and the total two-body phase space is $\Phi_2 = 1/(8\pi)$. Combining Eq. 12.99 with the flux factor of Eq. 12.97, we obtain for $2 \to 2$ scattering in the centre-of-mass frame

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{cm}} = \frac{1}{64\pi^2 s}\,\frac{|\vec{k}_f|}{|\vec{p}_{\text{cm}}|}\,|\mathcal{M}|^2. \tag{12.100}$$

For a decay $A \to 1+2$ in the rest frame of A, we only need to replace $\sqrt{s} \to m_A$:

$$\frac{d\Gamma}{d\Omega} = \frac{|\vec{k}_f|}{32\pi^2 m_A^2}\,|\mathcal{M}|^2. \tag{12.101}$$

12.8.5 Identical particles in the final state

The states $|k_1, k_2\rangle$ and $|k_2, k_1\rangle$ of two identical bosons are the **same** state, since the creation operators commute. When we integrate $d\Phi_n$ over all momenta of n identical particles, we count every physical final state $n!$ times. Therefore, a total cross-section or a total decay rate must be multiplied by a **symmetry factor**

$$\sigma = \frac{1}{n!}\int d\sigma, \qquad \text{or more generally} \qquad \prod_{\text{species}} \frac{1}{n_j!}, \tag{12.102}$$

where n_j is the number of identical particles of species j in the final state. For two identical particles we can, equivalently, integrate over half of the solid angle only. The differential cross-section $d\sigma/d\Omega$ of Eq. 12.100 needs no such factor: it is the rate for finding **a** particle in the direction $d\Omega$, and a detector at this angle counts either of the two.

12.8.6 Units

In natural units, a cross-section has dimension $[\text{mass}]^{-2}$ and a decay rate has dimension [mass]. To compare with experiment we restore $\hbar$ and c with

$$\hbar c = 0.1973\ \text{GeV fm}, \qquad (\hbar c)^2 = 0.3894\ \text{mb GeV}^2, \qquad \hbar = 6.582\times10^{-25}\ \text{GeV s}, \tag{12.103}$$

where $1\,\text{b} = 10^{-28}\,\text{m}^2$ (a "barn") and $1\,\text{mb} = 0.1\,\text{fm}^2$. A cross-section of $1\,\text{GeV}^{-2}$ is 0.3894 mb, and a decay rate of 1 GeV corresponds to a lifetime of 6.6×10^{-25} s. Typical cross-sections at the Large Hadron Collider are measured in picobarns (pb) and femtobarns (fb).

12.8.7 Example: $2 \to 2$ scattering in ϕ^4 theory

Let us compute the cross-section for $\phi(p_1)\,\phi(p_2) \to \phi(p_3)\,\phi(p_4)$ in the theory with $\mathcal{L}_{\rm int} = -\frac{\lambda}{4!}\phi^4$, at the lowest order in λ. We will see in the next chapter that the only connected and amputated diagram at this order is a single vertex, which gives the truncated Green's function $(-i\lambda)(2\pi)^4\delta^{(4)}(p_1+p_2-p_3-p_4)$. At this order $\tilde{Z}_\phi = 1$ and $m_{\rm phys} = m$. Comparing with Eq. 12.85, we read off

$$i\mathcal{M} = -i\lambda \qquad \rightsquigarrow \qquad \mathcal{M} = -\lambda + O(\lambda^2). \tag{12.104}$$

The amplitude is a constant. All masses are equal, $|\vec{k}_f| = |\vec{p}_{\rm cm}|$, and Eq. 12.100 gives an isotropic differential cross-section,

$$\left(\frac{d\sigma}{d\Omega}\right)_{\rm cm} = \frac{\lambda^2}{64\pi^2 s}. \tag{12.105}$$

The two particles in the final state are identical. The total cross-section is, with the symmetry factor $1/2!$,

$$\sigma(\phi\phi \to \phi\phi) = \frac{1}{2}\int d\Omega\, \frac{\lambda^2}{64\pi^2 s} = \frac{1}{2}\,\frac{4\pi\lambda^2}{64\pi^2 s} = \frac{\lambda^2}{32\pi s}. \tag{12.106}$$

This result is exact in the masses at this order, for any $s \geq 4m^2$. For example, for $\lambda = 1$ and $\sqrt{s} = 10$ GeV, we find $\sigma = 1/(3200\pi)\,\text{GeV}^{-2} \approx 3.9 \times 10^{-5}$ mb ≈ 39 nb. Notice that the cross-section decreases as $1/s$ at high energies. This is dictated by dimensional analysis, since λ is dimensionless and s is the only other scale at high energies.

12.8.8 Example: the decay of a scalar into two scalars

Consider two real scalar fields, Φ with mass M and χ with mass m, and a cubic interaction

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\Phi)(\partial^\mu\Phi) - \frac{1}{2}M^2\Phi^2 + \frac{1}{2}(\partial_\mu\chi)(\partial^\mu\chi) - \frac{1}{2}m^2\chi^2 - \frac{g}{2}\,\Phi\chi^2. \tag{12.107}$$

The coupling g has the dimension of a mass. Following the same steps as for the ϕ^4 vertex in the next chapter, the vertex with one Φ and two χ lines is $-ig$: the factor $1/2$ is cancelled by the 2! ways to contract the two χ fields with the two external χ particles. At lowest order, $\mathcal{M}(\Phi \to \chi\chi) = -g$. If $M > 2m$, the decay is kinematically allowed, and in the rest frame of Φ the momentum of each χ is

$$|\vec{k}_f| = \frac{M}{2}\sqrt{1 - \frac{4m^2}{M^2}}.$$

From Eq. 12.101, with the symmetry factor $1/2!$ for the two identical χ particles, we find

$$\Gamma(\Phi \to \chi\chi) = \frac{1}{2}\,4\pi\,\frac{|\vec{k}_f|}{32\pi^2 M^2}\,g^2 = \frac{g^2}{32\pi M}\sqrt{1 - \frac{4m^2}{M^2}}. \tag{12.108}$$

The factor $\sqrt{1-4m^2/M^2}$ is a "phase-space suppression". The decay rate vanishes at the threshold $M = 2m$, where the two χ particles are produced at rest and the phase space shrinks to a point.

Exercise: *Derive the Feynman rule $-ig$ for the $\Phi\chi\chi$ vertex by expanding*

$$\langle 0|\,T\{\chi(x_1)\chi(x_2)\Phi(y)\,e^{-i\int d^4z\,\frac{g}{2}\Phi\chi^2}\}\,|0\rangle$$

to first order in g. Then compute the lifetime of Φ for $M = 1$ GeV, $m = 0.1$ GeV and $g = 0.1$ GeV, in seconds.

13

Perturbation theory and Feynman diagrams

We have been able to solve exactly Hamiltonian systems for the simplest quantum field theories which describe free particles. For these, we could find all eigenstates and eigenvalues of the Hamiltonian. However, in general, when interactions among particles are present, we must resort to perturbation theory.

As a concrete example, we consider a simple new term in the Lagrangian of a real Klein-Gordon field,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4. \tag{13.1}$$

The Hamiltonian density is

$$\mathcal{H} = \pi\dot{\phi} - \mathcal{L}, \quad \pi = \frac{\partial \mathcal{L}}{\partial \dot{\phi}}. \tag{13.2}$$

The Hamiltonian is then found to be

$$H = H_0 + H', \tag{13.3}$$

where

$$H_0 = \int d^3\vec{x} \frac{\pi^2 + \left(\vec{\nabla}\phi\right)^2 + m^2\phi^2}{2}, \tag{13.4}$$

is the Hamiltonian of a free real Klein-Gordon field, and

$$H' = \int d^3\vec{x} \frac{\lambda}{4!}\phi^4, \tag{13.5}$$

is a new term.

For a generic $\lambda \neq 0$, it is not possible to find an exact expression for the quantum field ϕ. In this chapter, we will learn how to construct a solution using perturbation theory, by expanding around the known result for $\lambda = 0$.

The words "free" and "interacting" are physical rather than mathematical labels; "free" really means "solvable". Apart from special cases in lower space-time dimensions, the Lagrangians that we know how to solve exactly are quadratic in the fields. Their Euler-Lagrange equations are linear and they describe free particles. A term such as ϕ^4 makes the equations of motion non-linear, and this is the origin of interactions. Notice also that λ is dimensionless. In natural units the action is dimensionless and d^4x has mass dimension -4. The kinetic term $\frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi)$ then requires that ϕ has mass dimension 1,

and $\int d^4x\, \lambda\phi^4$ is dimensionless only if λ is. It is therefore meaningful to ask whether λ is small compared to one.

It is only necessary to know the field $\phi(\vec{x}, t)$ at a specific time value $t = \tau$. We can use symmetry under time translations in order to determine the field at any other moment t. The generator of space-time translations $x^\mu \longrightarrow x^\mu + \epsilon^\mu$ is the field momentum $P^\mu = \left(H, \vec{P}\right)$. Specifically for time translations, the generator is the Hamiltonian H. Fields at two different times are related via

$$\phi(\vec{x}, t) = e^{+iH(t-\tau)}\phi(\vec{x}, \tau)e^{-iH(t-\tau)} \tag{13.6}$$

(this is an exponentiated form of an equation such as Eq. 5.134, with $Q = P^\mu \delta x_\mu$, $\delta x_i = 0$). As we may have noticed, field operators evolve with time, while states do not. This is usually termed the Heisenberg picture.

Let us now compute the time evolution of the field $\phi(\vec{x}, \tau)$ in the hypothetical case that there is no interaction ($\lambda = 0$). We find that

$$\phi_I(\vec{x}, t) = e^{+iH_0(t-\tau)}\phi(\vec{x}, \tau)e^{-iH_0(t-\tau)}, \tag{13.7}$$

and, solving for $\phi(\vec{x}, \tau)$, we find

$$\phi(\vec{x}, \tau) = e^{-iH_0(t-\tau)}\phi_I(\vec{x}, t)e^{+iH_0(t-\tau)}. \tag{13.8}$$

The field $\phi_I(\vec{x}, t)$ is the field in the "interaction picture". The fields in the interaction picture and the Heisenberg picture are identical if $H = H_0$.

Combining Eqs. 13.6 and 13.8, we find the identity

$$\phi(\vec{x}, t) = U^{-1}(t, \tau)\phi_I(\vec{x}, t)U(t, \tau), \tag{13.9}$$

where we have defined the time-evolution operator in the interaction picture,

$$U(t, \tau) \equiv e^{+iH_0(t-\tau)}e^{-iH(t-\tau)}. \tag{13.10}$$

We now make a very important assumption. We require that there is a time, for example in the far past, for which the field of the full Hamiltonian H is a solution of the free Hamiltonian H_0. This is a reasonable approximation if we can identify a time when particles are far from each other and their interaction can be neglected. We can formally implement this by requiring, for example, that the interaction switches on at a certain time.

This assumption is very natural for the experiments that we wish to describe. In a scattering experiment we prepare beams of particles, for example electrons or photons, which are far apart and behave as free. We bring them together, they interact, and the products fly apart towards detectors far away, where they again behave as free particles. At times when the particles do not appear to interact, we expect the field to be very close to a free field. The same assumption also marks the limits of perturbation theory. It is useful when the full theory differs little from the solvable free theory, with particles propagating freely most of the time and interacting only here and there. If the solutions of the full theory are drastically different from the free ones, we cannot expect to describe them with an expansion around the free theory.

Given the existence of such a special time, the interaction-picture field $\phi_I(\vec{x}, t)$ will continue to be a solution of the free Hamiltonian at any time t. Then, Eq. 13.9 is a transformation of the field ϕ_I in the free theory (H_0) to the field ϕ in the full "interacting" theory (H).

13.1 Time evolution operator in the interaction picture

The calculation of the field operator in the full theory proceeds through the evaluation of the time-evolution operator $U(t,\tau)$, given that we can determine the field operator in the free theory.

We shall first derive a differential equation for $U(t,\tau)$. Differentiating with respect to time, we have

$$\begin{aligned} i\frac{\partial U(t,\tau)}{\partial t} &= i\left[iH_0 e^{iH_0(t-\tau)}e^{-iH(t-\tau)} - ie^{iH_0(t-\tau)}He^{-iH(t-\tau)}\right] \\ &= e^{iH_0(t-\tau)}(H-H_0)e^{-iH(t-\tau)} \\ &= e^{iH_0(t-\tau)}(H-H_0)e^{-iH_0(t-\tau)}e^{iH_0(t-\tau)}e^{-iH(t-\tau)}. \end{aligned} \tag{13.11}$$

We write the above in the form

$$i\frac{\partial U(t,\tau)}{\partial t} = V_I(t-\tau)U(t,\tau), \tag{13.12}$$

with the interaction potential defined as

$$V_I(t-\tau) \equiv e^{iH_0(t-\tau)}(H-H_0)e^{-iH_0(t-\tau)}. \tag{13.13}$$

Notice how we have used the Hamiltonian. We do not use it through Hamilton's equations, but as the Noether charge of time translations: it generates time evolution and it is conserved. Therefore, we can evaluate H and H_0 with the fields at any time we like, in particular at the special time τ where $\phi(\vec{x},\tau) = \phi_I(\vec{x},\tau)$. Then $H - H_0 = \int d^3\vec{x}\frac{\lambda}{4!}\phi_I(\vec{x},\tau)^4$ is a functional of the free field, and the conjugation with $e^{iH_0(t-\tau)}$ in Eq. 13.13 translates it in time with the free Hamiltonian,

$$V_I(t-\tau) = \int d^3\vec{x}\,\frac{\lambda}{4!}\,\phi_I(\vec{x},t)^4.$$

The interaction potential is thus built entirely out of the free field, which we know explicitly in terms of creation and annihilation operators.

We can cast the general solution of the differential equation Eq. 13.12 as a time-ordered exponential (as with an ordinary Schrödinger equation). Integrating both sides of the equation with respect to time, we obtain

$$U(t,\tau) = U(\tau,\tau) + \frac{1}{i}\int_\tau^t dt_1 V_I(t_1-\tau)U(t_1,\tau). \tag{13.14}$$

Let us rewrite the above equation, replacing t with t_1 and t_1 with t_2. We have

$$U(t_1,\tau) = U(\tau,\tau) + \frac{1}{i}\int_\tau^{t_1} dt_2 V_I(t_2-\tau)U(t_2,\tau). \tag{13.15}$$

We can substitute Eq. 13.15 into Eq. 13.14, obtaining

$$\begin{aligned} U(t,\tau) &= \left[1 + \frac{1}{i}\int_\tau^t dt_1 V_I(t_1-\tau)\right]U(\tau,\tau) \\ &+ \left(\frac{1}{i}\right)^2 \int_\tau^t dt_1 \int_\tau^{t_1} dt_2 V_I(t_1-\tau)V_I(t_2-\tau)U(t_2,\tau). \end{aligned} \tag{13.16}$$

Obviously, we can repeat inserting Eq. 13.14 into itself as many times as we wish. After an infinite number of iterations we obtain

$$\begin{aligned} U(t,\tau) &= \left[1+\sum_{n=1}^{\infty}\left(\frac{1}{i}\right)^n \int_\tau^t dt_1 \int_\tau^{t_1} dt_2 \ldots \int_\tau^{t_{n-1}} dt_n \right. \\ &\left. V_I(t_1-\tau)V_I(t_2-\tau)\ldots V_I(t_n-\tau)\right] U(\tau,\tau). \end{aligned} \tag{13.17}$$

Time ordering

Consider one of the simplest integrals in the above series,

$$I[t,\tau] = \int_\tau^t dt_1 \int_\tau^{t_1} dt_2 V_I(t_1-\tau)V_I(t_2-\tau). \tag{13.18}$$

It can be written as

$$I[t,\tau] = \int_\tau^t dt_1 \int_\tau^{t} dt_2 V_I(t_1-\tau)V_I(t_2-\tau)\Theta(t_1-t_2). \tag{13.19}$$

By changing integration variables, $t_1 \longrightarrow t_2$ and $t_2 \longrightarrow t_1$, we obtain

$$I[t,\tau] = \int_\tau^t dt_1 \int_\tau^{t} dt_2 V_I(t_2-\tau)V_I(t_1-\tau)\Theta(t_2-t_1). \tag{13.20}$$

Adding Eqs. 13.19 and 13.20, we obtain

$$\int_\tau^t dt_1 \int_\tau^{t_1} dt_2 V_I(t_1-\tau)V_I(t_2-\tau) = \frac{1}{2!}\int_\tau^t dt_1 dt_2 T\left\{V_I(t_1-\tau)V_I(t_2-\tau)\right\}, \tag{13.21}$$

where the symbol $T\{\ldots\}$ is our familiar time-ordering symbol, ordering operators from the largest to the smallest times,

$$\begin{aligned} T\left\{V_I(t_1-\tau)V_I(t_2-\tau)\right\} &= V_I(t_1-\tau)V_I(t_2-\tau)\Theta(t_1-t_2) \\ &+V_I(t_2-\tau)V_I(t_1-\tau)\Theta(t_2-t_1). \end{aligned} \tag{13.22}$$

It is not hard to convince ourselves that, in general,

$$\begin{aligned} &\int_\tau^t dt_1 V_I(t_1-\tau)\int_\tau^{t_1} dt_2 V_I(t_2-\tau)\ldots\int_\tau^{t_{n-1}} dt_n V_I(t_n-\tau) \\ &= \frac{1}{n!}\int_\tau^t dt_1\ldots dt_n\, T\left\{V_I(t_1-\tau)\ldots V_I(t_n-\tau)\right\}. \end{aligned} \tag{13.23}$$

Time-ordered exponential

We define a time-ordered exponentiated integral as the time-ordering of its Taylor series expansion,

$$\begin{aligned} Te^{\int_\tau^t dt' A(t')} &\equiv T\sum_{n=0}^{\infty}\frac{\left[\int_\tau^t dt' A(t')\right]^n}{n!} \\ &= 1+\sum_{n=1}^{\infty}\frac{1}{n!}\int_\tau^t dt_1\ldots dt_n T\left\{A(t_1)\ldots A(t_n)\right\}. \end{aligned} \tag{13.24}$$

We can now write a compact expression for the time-evolution operator of Eq. 13.17 as a time-ordered exponential, by using Eq. 13.23 and the definition of Eq. 13.24. We obtain

$$U(t,\tau) = Te^{-i\int_\tau^t dt' V_I(t'-\tau)} U(\tau,\tau). \tag{13.25}$$

To determine the operator $U(t,\tau)$ at an arbitrary time, we require a boundary value $U(\tau,\tau)$ at a time τ. This is identically the unit operator,

$$U(\tau,\tau) = 1. \tag{13.26}$$

Recall that τ is selected to be a time in the far past, when indeed

$$\phi(\vec{x},\tau) = \phi_I(\vec{x},\tau). \tag{13.27}$$

13.2 Field operators in the interacting and free theory

Let us summarise here what we have achieved.

- The field in the full theory $\phi(\vec{x},t)$ is related to the field in the free theory $\phi_I(\vec{x},t)$ at any time t, via a simple relation,

$$\phi(\vec{x},t) = U(t,\tau)^{-1}\phi_I(\vec{x},t)U(t,\tau). \tag{13.28}$$

- The time-evolution operator $U(t,\tau)$ is a time-ordered exponential, given by Eq. 13.25.

A time-ordered product of fields in the full theory can be written as

$$\begin{aligned} &T\left\{\phi(x_1)\phi(x_2)\dots\phi(x_n)\right\} = \\ &T\left\{U^{-1}(x_1^0,\tau)\phi_I(x_1)\tilde{U}(x_1^0,x_2^0)\phi_I(x_2)\tilde{U}(x_2^0,x_3^0)\dots\tilde{U}(x_{n-1}^0,x_n^0)\phi_I(x_n)U(x_n^0,\tau)\right\}, \end{aligned} \tag{13.29}$$

where we define the operator

$$\tilde{U}(t_2,t_1) \equiv U(t_2,\tau)U^{-1}(t_1,\tau). \tag{13.30}$$

The above operator is independent of the reference time τ. It is given by (**exercise**):

$$\tilde{U}(t_2,t_1) = Te^{-i\int_{t_1}^{t_2} dt' V_I(t')}, \tag{13.31}$$

where $t_2 > t_1$. We can also prove (**exercise**) that

$$\tilde{U}(t_3,t_2)\tilde{U}(t_2,t_1) = \tilde{U}(t_3,t_1), \tag{13.32}$$

with $t_3 > t_2 > t_1$. The inverse is then

$$\tilde{U}(t_2,t_1)\tilde{U}(t_1,t_2) = 1 \rightsquigarrow \tilde{U}^{-1}(t_2,t_1) = \tilde{U}(t_1,t_2) = \tilde{U}^\dagger(t_2,t_1). \tag{13.33}$$

13.3 The ground state of the interacting and the free theory

Our goal is to develop a formalism for the evaluation of Green's functions

$$\langle\Omega|\, T\left\{\phi(x_i)\dots\phi(x_n)\right\}|\Omega\rangle\,,$$

in the interacting theory. As we have seen, their Fourier transforms give us, with the LSZ reduction formula, the probability amplitudes for physical scattering processes. We recall that the points x_i^μ in the above need to have time components which are slightly complex (Eq. 12.74).

From the discussion of the previous section we have that

$$\begin{aligned}&\langle\Omega|\, T\left\{\phi(x_i)\dots\phi(x_n)\right\}|\Omega\rangle = \\ &\langle\Omega|\, T\left\{U^{-1}(x_1^0,\tau)\tilde{U}(x_1^0,x_n^0)\phi_I(x_i)\dots\phi_I(x_n)U(x_n^0,\tau)\right\}|\Omega\rangle\,,\end{aligned} \tag{13.34}$$

where the time ordering allowed us to combine together all the $\tilde{U}$ operators.

In general, we anticipate that the ground state $|\Omega\rangle$ of the interacting theory $H = H_0 + H'$ is a different state from the vacuum state $|0\rangle$ of the free theory H_0. Let us assume that the Hamiltonian H has a spectrum $|\psi_n\rangle$ with

$$H\,|\psi_n\rangle = E_n\,|\psi_n\rangle\,, \tag{13.35}$$

and $|\psi_0\rangle \equiv |\Omega\rangle$. Eigenstates of the full Hamiltonian form a complete set,

$$\mathbf{1} = |\Omega\rangle\langle\Omega| + \sum_{n\neq 0}|\psi_n\rangle\langle\psi_n|\,. \tag{13.36}$$

Let us now act with an operator on the vacuum state $|0\rangle$ of the free theory H_0,

$$e^{-iH(\tau+T)(1-i\delta)}e^{+iH_0(\tau+T)(1-i\delta)}\,|0\rangle\,,$$

choosing T a large time in the future and δ a very small "damping" parameter, $\delta \to 0_+$. This is in anticipation of needing to evaluate Green's functions with slightly complex times. We have

$$\begin{aligned}e^{-iH(\tau+T)(1-i\delta)}e^{+iH_0(\tau+T)(1-i\delta)}\,|0\rangle &= e^{-iH(\tau+T)(1-i\delta)}e^{+i0(\tau+T)(1-i\delta)}\,|0\rangle \\ &= e^{-iH(\tau+T)(1-i\delta)}\,|0\rangle\,.\end{aligned} \tag{13.37}$$

We now use the completeness of the Hamiltonian eigenstates of the full theory,

$$\begin{aligned}&e^{-iH(\tau+T)(1-i\delta)}e^{+iH_0(\tau+T)(1-i\delta)}\,|0\rangle = e^{-iH(\tau+T)(1-i\delta)}\mathbf{1}\,|0\rangle \\ =\;& e^{-iH(\tau+T)(1-i\delta)}\left(|\Omega\rangle\langle\Omega| + \sum_{n\neq 0}|\psi_n\rangle\langle\psi_n|\right)|0\rangle \\ =\;& |\Omega\rangle\, e^{-iE_0(\tau+T)-E_0(\tau+T)\delta}\,\langle\Omega|\,0\rangle \\ +\;& \sum_{n\neq 0}|\psi_n\rangle\, e^{-iE_n(\tau+T)-E_n(\tau+T)\delta}\,\langle\psi_n|\,0\rangle\,.\end{aligned} \tag{13.38}$$

Assuming a hierarchy $E_n > E_0$ for the first energy level of the full Hamiltonian with respect to the ground energy, the factor $e^{-E_n\delta(\tau+T)}$ vanishes faster than $e^{-E_0\delta(\tau+T)}$, as $T \to \infty$ at fixed $\delta > 0$ (the limit $\delta \to 0_+$ is taken only at the end). In this limit, we can write a simple relation between the ground state in the full theory and the vacuum in the free theory,

$$e^{-iH(\tau+T)(1-i\delta)}e^{+iH_0(\tau+T)(1-i\delta)}\,|0\rangle = |\Omega\rangle\, e^{-iE_0(1-i\delta)(\tau+T)}\,\langle\Omega|\,0\rangle\,. \tag{13.39}$$

We now observe that the operator on the left side of the above equation is nothing else than $U^{-1}(-T(1-i\delta), \tau(1-i\delta))$. We then have

$$|\Omega\rangle = \mathcal{N} U^{-1}\left(-T(1-i\delta), \tau(1-i\delta)\right)|0\rangle\,, \tag{13.40}$$

$$\langle\Omega| = \mathcal{N}^* \langle 0| U\left(T(1-i\delta), \tau(1-i\delta)\right), \tag{13.41}$$

where $\mathcal{N}$ is a normalisation constant, which we determine from the normalisation condition $\langle\Omega|\,\Omega\rangle = 1$,

$$|\mathcal{N}|^{-2} = \langle 0|\tilde{U}\left(T(1-i\delta), -T(1-i\delta)\right)|0\rangle\,. \tag{13.42}$$

Substituting into Eq. 13.34, we have that

$$\begin{aligned}
&\langle\Omega|\, T\left\{\phi(x_i)\ldots\phi(x_n)\right\}|\Omega\rangle = |\mathcal{N}|^2 \\
&\langle 0|\, U(T,\tau) \\
&T\left\{U^{-1}(x_1^0,\tau)\tilde{U}(x_1^0,x_n^0)\phi_I(x_i)\ldots\phi_I(x_n)U(x_n^0,\tau)\right\} \\
&U^{-1}(-T,\tau)\,|0\rangle\,.
\end{aligned} \tag{13.43}$$

Given that the times satisfy $T > x_i^0 > -T$, we can put all operators under the time-ordering symbol. We then obtain the final result for the Green's function,

$$\begin{aligned}
\langle\Omega|\, T\left\{\phi(x_i)\ldots\phi(x_n)\right\}|\Omega\rangle = &\lim_{\substack{\delta\to 0\\ T\to\infty}} \\
&\frac{\langle 0|\, T\phi_I(x_1)\ldots\phi_I(x_n) e^{-i\int_{-T}^{T} dt' V_I(t')}|0\rangle}{\langle 0|\, T e^{-i\int_{-T}^{T} dt' V_I(t')}|0\rangle},
\end{aligned} \tag{13.44}$$

where the points x_i^μ are taken at slightly complex times $x_i^0(1-i\delta)$. Notice what we have achieved. On the lhs we require a Green's function for the fields $\phi(x)$ of the full theory. On the rhs this is expressed purely in terms of fields ϕ_I and the vacuum state $|0\rangle$ in the free field theory. Eq. 13.44 is known as the **Gell-Mann–Low formula**. Notice the physical content of the step which related $|\Omega\rangle$ to $|0\rangle$. If we follow the ground state of the full theory to the very far past or future, where particles behave as free, it becomes the vacuum of the free theory, up to a normalisation constant. We may regard this as an additional physical assumption, analogous to the one we made for the fields. The slightly complex times $T(1-i\delta)$ implement it mathematically, by suppressing all excited states of the full Hamiltonian.

Eq. 13.44 is an exact result, and it will serve as our basis for perturbation theory.

The exponential can be written as a function of the perturbation Lagrangian with free fields. For our example interaction

$$e^{-i\int_{-\infty}^{\infty} dt' V_I(t')} = e^{-i\int d^4x \frac{\lambda}{4!}\phi_I(x)^4}, \tag{13.45}$$

and it can be expanded as a Taylor series in λ.

13.4 Feynman diagrams for ϕ^4 theory

In the previous section, we found that Green's functions can be cast in a form suitable for applying the method of perturbation theory. For the example Hamiltonian of Eq. 13.5, we

can write the result as

$$\langle\Omega|\,T\left\{\phi(x_1)\ldots\phi(x_n)\right\}|\Omega\rangle = \lim_{\substack{\delta\to 0\\ T\to\infty}} \frac{\langle 0|\,T\phi_I(x_1)\ldots\phi_I(x_n)\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{-i\lambda}{4!}\int d^4x\phi_I(x)^4\right)^n|0\rangle}{\langle 0|\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{-i\lambda}{4!}\int d^4x\phi_I(x)^4\right)^n|0\rangle}. \tag{13.46}$$

To evaluate the terms on the right-hand side of Eq. 13.46, we need to be able to compute expectation values in the free-field theory of time-ordered products of field operators. This can be achieved by means of Wick's theorem of Section 11.7.

The pictorial application of Wick's theorem leads to the representation of Green's functions in the full theory as a perturbative expansion in terms of Feynman diagrams. As a concrete example, we shall consider the two-point function in the full theory, through order $O(\lambda)$ in the coupling parameter λ. From Eq. 13.46 we have

$$\begin{aligned}
&\langle\Omega|\,T\left\{\phi(x_1)\phi(x_2)\right\}|\Omega\rangle\\
&= \frac{\langle 0|\,T\phi_I(x_1)\phi_I(x_2)\left(1+\frac{(-i\lambda)}{4!}\int d^4x\phi_I(x)^4\right)|0\rangle + O(\lambda^2)}{\langle 0|\,T\left(1+\frac{(-i\lambda)}{4!}\int d^4x\phi_I(x)^4\right)|0\rangle + O(\lambda^2)}\\
&= \langle 0|\,T\left\{\phi_I(x_1)\phi_I(x_2)\right\}|0\rangle + \frac{(-i\lambda)}{4!}\int d^4x\,\langle 0|\,T\left\{\phi_I(x_1)\phi_I(x_2)\phi_I(x)^4\right\}|0\rangle\\
&-\frac{(-i\lambda)}{4!}\langle 0|\,T\left\{\phi_I(x_1)\phi_I(x_2)\right\}|0\rangle\int d^4x\,\langle 0|\,\phi_I(x)^4|0\rangle + O(\lambda^2).
\end{aligned} \tag{13.47}$$

Let us compute pictorially the second term in the expansion,

$$\frac{(-i\lambda)}{4!}\int d^4x\,\langle 0|\,T\left\{\phi_I(x_1)\phi_I(x_2)\phi_I(x)^4\right\}|0\rangle\,.$$

The corresponding Green's function is equal to the sum of the contractions of the graph in Fig. 13.1. We find two different types of contractions which lead to two different "Feynman

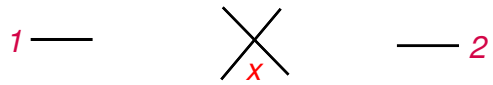
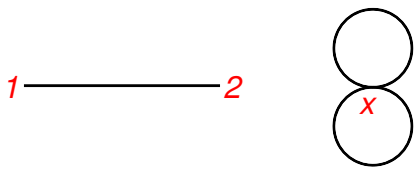


Figure 13.1. The term $\langle 0|\,T\left\{\phi_I(x_1)\phi_I(x_2)\phi_I(x)^4\right\}|0\rangle$ as the sum of all contractions of the graph above.

diagrams".

- **Diagram I:** The two fields defined at the "external" points x_1 and x_2 get contracted with each other. The four fields defined at the "internal" point x are then contracted among themselves.

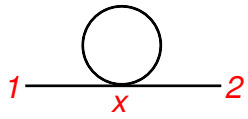


This Feynman diagram is classified as "disconnected", meaning that the internal point x is not connected to any of the external points x_1 and x_2.

This configuration occurs in more than one way. The external fields at x_1 and x_2 can be contracted in 1 way. Take now one of the four fields at the internal point x. This field can be contracted with any of the remaining internal fields in 3 ways. Finally, the remaining two internal fields can be contracted in 1 way only. Therefore, this Feynman diagram appears

$$1 \times 3 \times 1 \text{ times.}$$

- **Diagram II:** The two external fields defined at x_1 and x_2 are contracted with internal fields at x.



This is a "connected" diagram, meaning that internal points are connected to external points by following the lines of the graph.

We can now compute the multiplicity of the Feynman diagram. There are 4 ways that we can contract the field at the first external point with one of the fields at the internal point x. The second external point can then be contracted to another internal field in three ways. Finally, two remaining internal fields can be contracted with each other in one way. Therefore, this Feynman diagram occurs

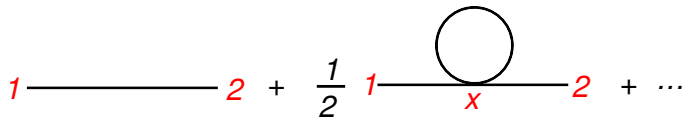
$$4 \times 3 \times 1 = \frac{4!}{2} \text{ times.}$$

Now we examine the last term

$$-\frac{(-i\lambda)}{4!} \langle 0| T\left\{\phi_I(x_1)\phi_I(x_2)\right\} |0\rangle \int d^4x \langle 0| \phi_I(x)^4 |0\rangle$$

in Eq. 13.47, which is the contribution of the denominator in Eq. 13.46 to the Taylor expansion in λ. It is easy to convince ourselves that this term yields the same Feynman diagram as the disconnected **Diagram I**. The two external fields are contracted together. The four internal fields are also contracted among themselves. However, this term carries an overall minus sign. As a result, the disconnected Feynman diagram drops out from the result. This observation holds at all orders in perturbation theory. *Disconnected Feynman diagrams do not contribute to the perturbative expansion of Green's functions.*[1] They factorise in the numerator of Eq. 13.46 and cancel exactly against the denominator.

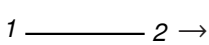
In summary, the perturbative expansion of the two-point Green's function through order $O(\lambda)$ can be represented very simply with the following Feynman diagrams:

$\langle \Omega| T\left\{\phi(x_1)\phi(x_2)\right\} |\Omega\rangle =$  (13.48)

We can obtain a concrete mathematical expression from the above diagrams by following very simple rules, pioneered by Feynman:

[1] We prove this statement for the two-point function in Section 13.4.1. The general proof is easier with the path-integral formalism of QFT II.

1. We associate a propagator to each line,



$$\rightarrow \langle 0|\, T\left\{\phi_I(x_1)\phi_I(x_2)\right\}|0\rangle\,. \tag{13.49}$$

2. We associate a factor of $(-i\lambda)$ and a space-time integration with each vertex,



$$\rightarrow (-i\lambda)\int d^4x. \tag{13.50}$$

The integration over the position of a vertex is the superposition principle at work. An interaction which we do not observe at the external points could have happened anywhere in space-time, and we must add up the amplitudes for all possibilities. The structure of the vertices can be read off directly from the interaction terms of the Lagrangian. In ϕ^4 theory four lines meet at a vertex; with a ϕ^3 term we would also have vertices where three lines meet. Every vertex brings one power of the coupling. If λ is small, diagrams with fewer vertices dominate, and we can decide how many orders to compute according to the precision that we wish to achieve. We are fortunate that the strengths of the interactions in the realistic quantum field theories are small, at least in some range of energies.

These rules give us a simple pictorial representation in terms of Feynman diagrams of the perturbative series for any Green's function $G(x_1, x_2, \ldots, x_N)$. We draw all possible graphs with N external points and with no more vertices than the maximum order λ^n in the perturbative expansion that we require. Then we translate the Feynman diagrams into mathematical expressions using the rules above. One difficulty in this procedure is to determine the combinatorial rational factor that multiplies the diagram, as for example the factor $1/2$ in the second diagram of Eq. 13.48. This number can always be obtained by following the procedure that led to Eq. 13.48, which is a rather brute-force method. For all practical purposes, this method is sufficient, and for more complicated cases we can easily program it in a computer code. The combinatorial factor, also known as the symmetry factor, can be determined cleverly as the inverse of the number of independent exchange symmetry operations of the Feynman diagram (exchanging vertices and propagators) which leave it intact. This is a nice method for everyone who is confident in spotting all symmetries without double counting equivalent ones. It is, however, prone to human errors.

13.4.1 Why disconnected graphs drop out

The statement above, that disconnected Feynman diagrams cancel against the denominator of Eq. 13.46, can be proven to all orders without path integrals by a direct counting of Wick contractions. We do this here for the two-point function in ϕ^4 theory; the argument generalises with no essential change.

We want to show that in

$$\langle\Omega|\, T\left\{\phi(x_1)\phi(x_2)\right\}|\Omega\rangle = \frac{\langle 0|\, T\,\phi_I(x_1)\phi_I(x_2)\, e^{-i\int d^4z\,\frac{\lambda}{4!}\phi_I^4(z)}\,|0\rangle}{\langle 0|\, T\, e^{-i\int d^4z\,\frac{\lambda}{4!}\phi_I^4(z)}\,|0\rangle} \tag{13.51}$$

the disconnected Feynman diagrams drop out. Let us start with the numerator,

$$\langle 0|\, T\,\phi_I(x_1)\phi_I(x_2)\, e^{\int [dz]\,\phi_I^4(z)}\,|0\rangle\,, \qquad \text{with} \quad [dz] = -\frac{i\lambda}{4!}\,d^4z. \tag{13.52}$$

We denote the external fields by dots, $\dot{x}_1 \equiv \phi_I(x_1)$, $\dot{x}_2 \equiv \phi_I(x_2)$, and the interaction by a box,

$$\square \equiv -i \int d^4z \, \frac{\lambda}{4!} \, \phi_I^4(z) = \int [dz] \, \phi_I^4(z), \tag{13.53}$$

so that the numerator reads $\langle 0| T \, \dot{x}_1 \, \dot{x}_2 \, e^{\square} |0\rangle$, with

$$e^{\square} = 1 + \square + \frac{1}{2!}\square\square + \frac{1}{3!}\square\square\square + \ldots + \frac{1}{n!}\underbrace{\square\cdots\square}_{n \text{ times}} + \ldots . \tag{13.54}$$

So we have to consider all contractions in

$$\mathcal{A} \equiv \langle 0| T \, \dot{x}_1 \, \dot{x}_2 \left(1 + \square + \frac{1}{2!}\square\square + \frac{1}{3!}\square\square\square + \ldots\right) |0\rangle . \tag{13.55}$$

Let us classify the contractions according to whether x_1 and x_2 connect to 0, 1 or 2 internal boxes, and how many internal boxes are also connected to x_1 and x_2:

- $\mathcal{A}_0$: contractions where x_1 and x_2 do not connect internally;
- $\mathcal{A}_1$: contractions where x_1 and x_2 connect to the same internal box;
- $\mathcal{A}_{2,1}, \mathcal{A}_{2,2}, \mathcal{A}_{2,3}, \ldots$: contractions where x_1 and x_2 connect to two internal boxes which are linked directly ($\mathcal{A}_{2,1}$) or through 1, 2, ... other internal boxes ($\mathcal{A}_{2,2}$, $\mathcal{A}_{2,3}$, ...); in $\mathcal{A}_{2,m}$ the component containing x_1, x_2 has $m+1$ boxes.

Since x_1 and x_2 each carry a single field, this classification is exhaustive. Then

$$\mathcal{A} = \mathcal{A}_0 + \mathcal{A}_1 + \mathcal{A}_{2,1} + \mathcal{A}_{2,2} + \mathcal{A}_{2,3} + \ldots . \tag{13.56}$$

$\mathcal{A}_0$. Here x_1 is contracted with x_2 and all boxes contract among themselves:

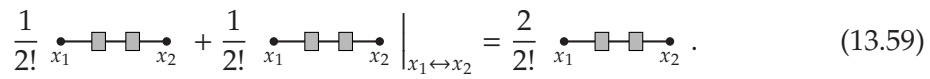
$$\begin{aligned} \mathcal{A}_0 &= \langle 0| T \, \phi_I(x_1)\phi_I(x_2) |0\rangle \; \langle 0| T \left(1 + \square + \frac{\square\square}{2!} + \ldots\right) |0\rangle \\ &= \langle 0| T \, \phi_I(x_1)\phi_I(x_2) |0\rangle \; \langle 0| T \exp(\square) |0\rangle \\ &= \quad \underset{x_1}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\underset{x_2}{\bullet} \quad \langle 0| T \exp(\square) |0\rangle . \end{aligned} \tag{13.57}$$

$\mathcal{A}_1$. Both x_1 and x_2 attach to the same box. At order n the box they attach to can be any of the n identical boxes in $\frac{1}{n!}\square^n$, and the remaining $n-1$ boxes contract among themselves:

$$\begin{aligned} \mathcal{A}_1 &= \quad \underset{x_1}{\bullet}\!\!-\square-\!\!\underset{x_2}{\bullet} \quad \Big[1 + \frac{2}{2!}\square + \frac{3}{3!}\square\square \\ &\qquad\qquad + \ldots + \frac{n}{n!}\underbrace{\square\cdots\square}_{n-1} + \ldots \Big]_{\text{vac}} \\ &= \quad \underset{x_1}{\bullet}\!\!-\square-\!\!\underset{x_2}{\bullet} \quad \langle 0| T \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \underbrace{\square\cdots\square}_{n-1} |0\rangle = \underset{x_1}{\bullet}\!\!-\square-\!\!\underset{x_2}{\bullet} \quad \langle 0| T \exp(\square) |0\rangle \\ &= \frac{1}{2} \quad \underset{x_1}{\bullet}\!\!\!\bigcirc\!\!\!\underset{x_2}{\bullet} \quad \langle 0| T \exp(\square) |0\rangle , \end{aligned} \tag{13.58}$$

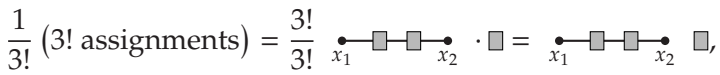
where $[\ldots]_{\text{vac}}$ means that the boxes inside the bracket contract only among themselves, and $\underset{x_1}{\bullet}\!\!-\square-\!\!\underset{x_2}{\bullet}$ denotes x_1 and x_2 attached to one box whose remaining two fields are contracted together. The factor $\frac{1}{2}$ is the symmetry factor of Diagram II above (4×3 ways to attach x_1, x_2, divided by the 4! in $[dz]$).

$\mathcal{A}_{2,1}$**..** Now x_1 and x_2 attach to two different boxes which are connected to each other. This needs at least two boxes (the terms 1 and $\square$ in $e^{\square}$ do not contribute). At order $n = 2$ we may attach x_1 to the first box and x_2 to the second, or the other way round. Both give the same Feynman diagram:



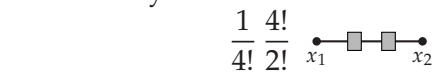
(13.59)

At order $n = 3$, x_1 attaches to any of the three boxes, x_2 to any of the two remaining ones, and the third box is a vacuum bubble:



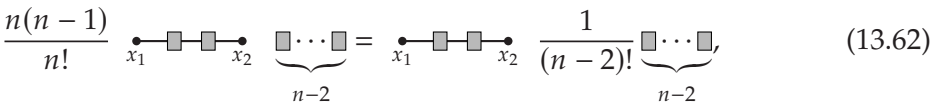
(13.60)

and at $n = 4$ in the same way



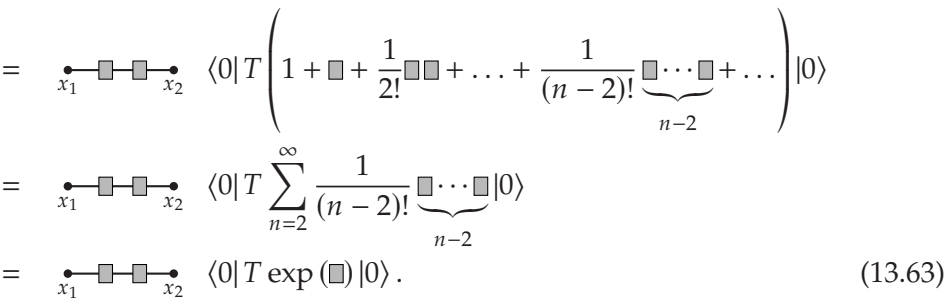
(13.61)

For general n the two boxes linked to x_1, x_2 are chosen in $n(n-1)$ ways,



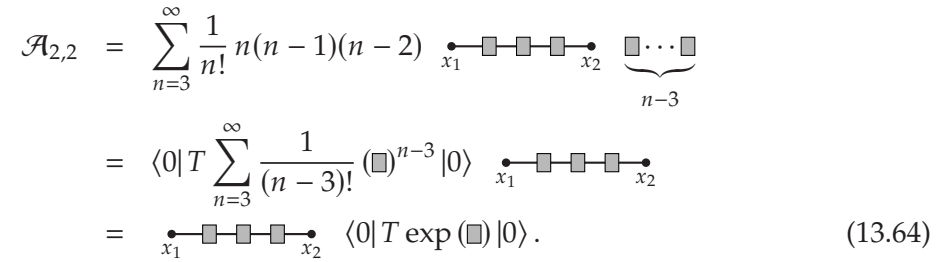
(13.62)

and thus



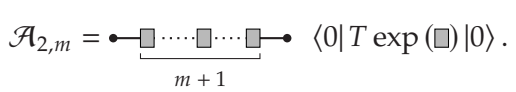
(13.63)

$\mathcal{A}_{2,2}$**..** Here three boxes are linked in the component of x_1, x_2; they can be chosen in $n(n-1)(n-2)$ ways:



(13.64)

$\mathcal{A}_{2,m}$**..** In the same way we can show that



(13.65)

Indeed, at order n the box attached to x_1 can be chosen in n ways, the box attached to x_2 in $n-1$ ways, and the $m-1$ intermediate boxes of the chain in $(n-2)(n-3)\cdots(n-m)$ ways,

while the remaining $n-m-1$ boxes form vacuum bubbles:

$$\frac{1}{n!}\,\frac{n!}{(n-2)!}\,\frac{(n-2)!}{(n-m-1)!}\;\underbrace{\bullet\!-\!\square\cdots\square\cdots\square\!-\!\bullet}_{m+1}\;\overbrace{\square\cdots\square}^{n-m-1} = \underbrace{\bullet\!-\!\square\cdots\square\cdots\square\!-\!\bullet}_{m+1}\;\frac{1}{(n-m-1)!}\overbrace{\square\cdots\square}^{n-m-1}. \tag{13.66}$$

Summing over $n \geq m+1$ again reproduces $\langle 0|\,T\exp(\square)\,|0\rangle$.

Thus,

$$\begin{aligned}\mathcal{A} &= \mathcal{A}_0 + \mathcal{A}_1 + \mathcal{A}_{2,1} + \mathcal{A}_{2,2} + \mathcal{A}_{2,3} + \ldots + \mathcal{A}_{2,m} + \ldots \\ &= \underset{x_1\quad x_2}{\bullet\!-\!\bullet}\;\langle 0|\,Te^{\square}\,|0\rangle + \underset{x_1\quad\quad x_2}{\bullet\!-\!\square\!-\!\bullet}\;\langle 0|\,Te^{\square}\,|0\rangle \\ &\quad + \underset{x_1\quad\quad\quad x_2}{\bullet\!-\!\square\!-\!\square\!-\!\bullet}\;\langle 0|\,Te^{\square}\,|0\rangle + \ldots + \underbrace{\bullet\!-\!\square\cdots\square\cdots\square\!-\!\bullet}_{m+1}\;\langle 0|\,Te^{\square}\,|0\rangle + \ldots,\end{aligned} \tag{13.67}$$

$$\begin{aligned}\Rightarrow\quad \langle 0|\,T\,\dot{x}_1\,\dot{x}_2\,\exp(\square)\,|0\rangle &= \Big(\underset{x_1\quad x_2}{\bullet\!-\!\bullet} + \underset{x_1\quad\quad x_2}{\bullet\!-\!\square\!-\!\bullet} + \underset{x_1\quad\quad\quad x_2}{\bullet\!-\!\square\!-\!\square\!-\!\bullet} \\ &\qquad + \ldots + \underbrace{\bullet\!-\!\square\cdots\square\cdots\square\!-\!\bullet}_{m+1} + \ldots\Big) \\ &\quad\cdot\,\langle 0|\,T\exp(\square)\,|0\rangle,\end{aligned} \tag{13.68}$$

$$\Rightarrow\quad \frac{\langle 0|\,T\,\dot{x}_1\,\dot{x}_2\,\exp(\square)\,|0\rangle}{\langle 0|\,T\exp(\square)\,|0\rangle} = \underbrace{\underset{x_1\quad x_2}{\bullet\!-\!\bullet} + \underset{x_1\quad\quad x_2}{\bullet\!-\!\square\!-\!\bullet} + \ldots + \underbrace{\bullet\!-\!\square\cdots\square\cdots\square\!-\!\bullet}_{m+1} + \ldots}_{\text{sum of connected graphs only!}} \tag{13.69}$$

Remarks: (i) In the pictures above a "chain" x_1–□–$\cdots$–□–x_2 stands for *any* connected graph with the given number of vertices that contains x_1 and x_2 (e.g. $\mathcal{A}_{2,1}$ contains the sunset graph, where the two vertices are joined by three propagators, as well as two tadpole insertions joined by one propagator); the legs of the vertices are not drawn. The counting does not depend on the topology: of the n identical boxes in $\frac{1}{n!}\square^n$, the k boxes belonging to the connected component can be chosen in $\binom{n}{k}$ ways, and $\frac{1}{n!}\binom{n}{k} = \frac{1}{k!}\,\frac{1}{(n-k)!}$: the connected component keeps its own $\frac{1}{k!}\square^k$, and the rest is the order-$(n-k)$ term of $\langle 0|\,Te^{\square}\,|0\rangle$. (For a chain, the $\frac{1}{k!}$ is cancelled by the $k!$ ways of placing the k chosen boxes along the chain, which is the ordered counting $n(n-1)\cdots(n-k+1) = \binom{n}{k}k!$ used above.) (ii) What drops out are the *vacuum bubbles*, i.e. components with no external point. For an N-point function with $N \geq 4$, products of connected pieces each containing external points (e.g. x_1–x_2 times x_3–x_4) do *not* drop out; they are removed only later, by truncation and LSZ; the argument above, applied to the vacuum part, goes through unchanged. (iii) It is tempting to identify $\langle 0|\,Te^{\square}\,|0\rangle$ with $\langle\Omega|\,\Omega\rangle$. With the normalisation $\langle\Omega|\,\Omega\rangle = 1$ used in this chapter this is not correct: by Eq. 13.39, $\langle 0|\,Te^{\square}\,|0\rangle \to |\langle\Omega|\,0\rangle|^2\,e^{-2iE_0T(1-i\delta)}$ (the integration runs over $-T < z^0 < T$), a phase and a modulus which, since the vacuum energy E_0 is proportional to the spatial volume, are both exponential in the space-time volume VT; equivalently, the sum of vacuum bubbles is proportional to VT, and $\langle 0|\,Te^{\square}\,|0\rangle$ is the exponential of the sum of connected bubbles. Neither numerator nor denominator has a limit as $T \to \infty$; only the ratio in Eq. 13.69 does, and it is this ratio that the factorisation proven above controls, order by order in λ.

13.5 Feynman rules in momentum space

The LSZ reduction formula expresses scattering amplitudes as truncated Green's functions in momentum space. We can develop simple Feynman rules for computing them directly.

We start with the Fourier transform of the propagator from the origin to a point x,

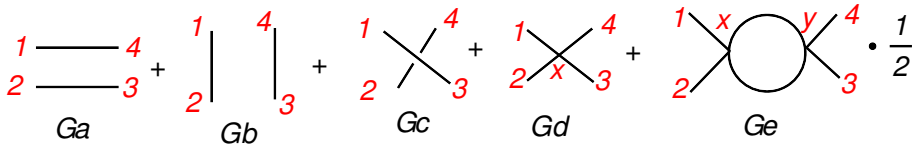
$$\tilde{G}(p) = \int d^4x\, e^{ip\cdot x} \langle\Omega| T\left\{\phi(x)\phi(0)\right\} |\Omega\rangle\,. \tag{13.70}$$

At leading order in perturbation theory we have

$$\begin{aligned}
\tilde{G}(p) &= \int d^4x\, e^{ip\cdot x} \langle 0| T\left\{\phi_I(x)\phi_I(0)\right\} |0\rangle + O(\lambda) \\
&= \int d^4x\, e^{ip\cdot x} \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\delta} e^{-ik\cdot x} + O(\lambda) \\
&= \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\delta} \int d^4x\, e^{i(p-k)\cdot x} + O(\lambda) \\
&= \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\delta} (2\pi)^4 \delta^{(4)}(p-k) + O(\lambda) \\
\rightsquigarrow \tilde{G}(p) &= \frac{i}{p^2 - m^2 + i\delta} + O(\lambda).
\end{aligned} \tag{13.71}$$

Let us now look at the scattering of four particles. The Green's function we require is

$G(x_1, x_2, x_3, x_4) =$



$+\ldots$ (13.72)

Let us now compute the Green's function in momentum space,

$$\tilde{G}(p_3, p_4; p_1, p_2) = \int d^4x_1\, d^4x_2\, d^4x_3\, d^4x_4\, e^{-i(x_1\cdot p_1 + x_2\cdot p_2 - x_3\cdot p_3 - x_4\cdot p_4)} G(x_1, x_2, x_3, x_4), \tag{13.73}$$

where we consider p_1, p_2 in the initial state and p_3, p_4 in the final state. We write

$$\tilde{G}(p_3, p_4; p_1, p_2) = G_a + G_b + G_c + G_d + G_e + \ldots, \tag{13.74}$$

where $G_{a,b,c}$ are the contributions of the first three diagrams and G_d, G_e are the contributions of the fourth and fifth diagrams at order $O(\lambda)$ and $O(\lambda^2)$, respectively. The first three diagrams give

$$\begin{aligned}
G_a + G_b + G_c &= (2\pi)^4\delta^{(4)}(p_1 - p_4) \frac{i}{p_1^2 - m^2 + i\delta} (2\pi)^4\delta^{(4)}(p_2 - p_3) \frac{i}{p_2^2 - m^2 + i\delta} \\
&+ (2\pi)^4\delta^{(4)}(p_1 - p_3) \frac{i}{p_1^2 - m^2 + i\delta} (2\pi)^4\delta^{(4)}(p_2 - p_4) \frac{i}{p_2^2 - m^2 + i\delta} \\
&+ (2\pi)^4\delta^{(4)}(p_1 + p_2) \frac{i}{p_1^2 - m^2 + i\delta} (2\pi)^4\delta^{(4)}(p_3 + p_4) \frac{i}{p_3^2 - m^2 + i\delta}
\end{aligned} \tag{13.75}$$

The fourth diagram is converted into a mathematical expression following again the Feynman rules of the previous section. We have

$$\begin{aligned} G_d &= \int d^4x_1 d^4x_2 d^4x_3 d^4x_4 e^{-i\left(x_1\cdot p_1 + x_2\cdot p_2 - x_3\cdot p_3 - x_4\cdot p_4\right)} \\ &\times (-i\lambda)\int d^4x \qquad (\text{ vertex }) \\ &\times \langle 0|T\left\{\phi_I(x_1)\phi_I(x)\right\}|0\rangle \qquad (\text{ contraction of } x_1 \text{ and } x\ ) \\ &\times \langle 0|T\left\{\phi_I(x_2)\phi_I(x)\right\}|0\rangle \qquad (\text{ contraction of } x_2 \text{ and } x\ ) \\ &\times \langle 0|T\left\{\phi_I(x_3)\phi_I(x)\right\}|0\rangle \qquad (\text{ contraction of } x_3 \text{ and } x\ ) \\ &\times \langle 0|T\left\{\phi_I(x_4)\phi_I(x)\right\}|0\rangle \qquad (\text{ contraction of } x_4 \text{ and } x\ ). \end{aligned} \tag{13.76}$$

Substituting the momentum-space expression for each of the propagators and performing the x and x_i integrations, we find

$$G_d = (-i\lambda)\frac{i}{p_1^2 - m^2 + i\delta}\frac{i}{p_2^2 - m^2 + i\delta}\frac{i}{p_3^2 - m^2 + i\delta}\frac{i}{p_4^2 - m^2 + i\delta}(2\pi)^4\delta^{(4)}(p_1 + p_2 - p_3 - p_4). \tag{13.77}$$

The diagram G_e is a "loop diagram". Let us write the corresponding expression, using the Feynman rules in position space. We have

$$\begin{aligned} G_e &= \int d^4x_1 d^4x_2 d^4x_3 d^4x_4 e^{-i\left(x_1\cdot p_1 + x_2\cdot p_2 - x_3\cdot p_3 - x_4\cdot p_4\right)} \\ &\times \frac{1}{2}(-i\lambda)^2\int d^4x d^4y \qquad (\text{ vertices }) \\ &\times \langle 0|T\left\{\phi_I(x_1)\phi_I(x)\right\}|0\rangle \qquad (\text{ contraction of } x_1 \text{ and } x\ ) \\ &\times \langle 0|T\left\{\phi_I(x_2)\phi_I(x)\right\}|0\rangle \qquad (\text{ contraction of } x_2 \text{ and } x\ ) \\ &\times \langle 0|T\left\{\phi_I(x_3)\phi_I(y)\right\}|0\rangle \qquad (\text{ contraction of } x_3 \text{ and } y\ ) \\ &\times \langle 0|T\left\{\phi_I(x_4)\phi_I(y)\right\}|0\rangle \qquad (\text{ contraction of } x_4 \text{ and } y\ ) \\ &\times \langle 0|T\left\{\phi_I(x)\phi_I(y)\right\}|0\rangle^2 \qquad (\text{ two contractions of } x \text{ and } y\ ). \end{aligned} \tag{13.78}$$

Performing all x-space integrations, we find

$$\begin{aligned} G_e &= \frac{i}{p_1^2 - m^2 + i\delta}\frac{i}{p_2^2 - m^2 + i\delta}\frac{i}{p_3^2 - m^2 + i\delta}\frac{i}{p_4^2 - m^2 + i\delta} \\ &\times(2\pi)^4\delta^{(4)}(p_1 + p_2 - p_3 - p_4) \\ &\times\frac{1}{2}(-i\lambda)^2 \\ &\times\int_{-\infty}^{+\infty}\frac{d^4k}{(2\pi)^4}\frac{i}{k^2 - m^2 + i\delta}\frac{i}{(k + p_1 + p_2)^2 - m^2 + i\delta}. \end{aligned} \tag{13.79}$$

Observe our final expressions for $G_{a,b,c}$, G_d and G_e. We can formulate easy rules (Feynman rules) to produce them from the Feynman diagrams:

- At every vertex we pick up a factor

$$(-i\lambda).$$

- With each propagator comes a factor

$$\frac{i}{p^2 - m^2 + i\delta}.$$

- At each vertex, we have delta functions guaranteeing that momentum is conserved. As a result, for all particles that are connected with each other, there is an overall factor

$$(2\pi)^4 \delta^{(4)} \left(\sum_{in} p_{in} - \sum_{out} p_{out} \right),$$

  which forces the sum of incoming momenta in a connected diagram to be equal to the sum of the outgoing momenta.

- The loop in diagram G_e introduces an integration over a "loop momentum" with a measure

$$\int_{-\infty}^{+\infty} \frac{d^4k}{(2\pi)^4}.$$

Two further points complete this list. (i) Each diagram is multiplied by its symmetry factor; for G_e it is $\frac{1}{2}$, as drawn in the figure: there are $2 \cdot (4 \cdot 3)^2 \cdot 2 = 576$ Wick contractions against $2!\,(4!)^2 = 1152$ from the expansion. (ii) At order λ^2 there are three one-loop diagrams of this type, in the s-, t- and u-channels, with the loop carrying $p_1 + p_2$, $p_1 - p_3$ and $p_1 - p_4$ respectively, each with factor $\frac{1}{2}$; G_e is the s-channel one (Peskin–Schroeder, Sec. 10.2). The integration over the loop momentum has again its origin in the superposition principle. We know the momenta of the incoming and outgoing particles, but the momentum circulating in the loop is never observed and we must sum over all its values. Notice that these integrations extend to infinitely large momenta. It is rather ambitious to believe that our theory describes correctly what happens at infinite energies, and indeed loop integrals are often divergent. For example, rescaling $k \to \Lambda k$ with Λ large, the measure d^4k scales as Λ^4 and the two propagators of G_e as Λ^{-4}; the integral diverges logarithmically. We will learn how to deal with such infinities in Part II.

13.6 Truncated Green's functions in perturbation theory

We are interested in computing amplitudes for physical scattering processes, as given by the LSZ formula. For the scattering of two particles with momenta p_1, p_2 to two particles with momenta p_3, p_4 we then need a truncated Green's function

$$\langle p_3, p_4 | \, p_1, p_2 \rangle = \tilde{Z}_\phi^2 \tilde{G}_{trunc}(p_3, p_4; p_1, p_2) \Big|_{p_i^2 = m^2_{\text{phys}}}, \tag{13.80}$$

computed for squared external momenta equal to the squared physical mass of the particles. The truncated Green's function is defined as

$$\tilde{G}_{trunc}(p_3, p_4; p_1, p_2) = \frac{\tilde{G}(p_3, p_4; p_1, p_2)}{\tilde{G}(p_1)\tilde{G}(p_2)\tilde{G}(p_3)\tilde{G}(p_4)}, \tag{13.81}$$

where the Green's function in momentum space is divided by the propagator of each of the particles in the initial and final states.

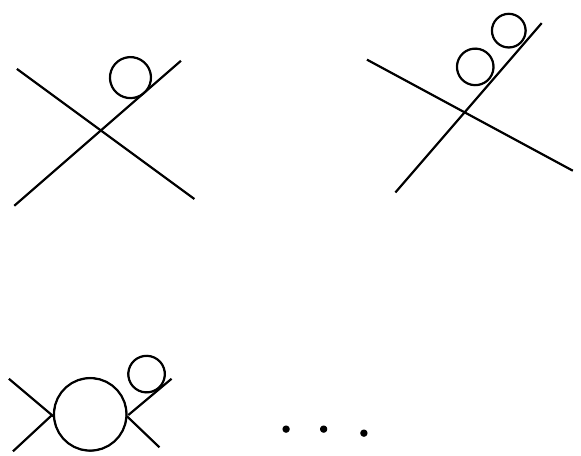


Figure 13.2. Feynman diagrams which are corrections to external lines and do not contribute to physical scattering amplitudes.

The Källén–Lehmann representation for the pole of the propagator is

$$\tilde{G}(p) = \frac{i\tilde{Z}_\phi}{p^2 - m^2_{\text{phys}}} + \text{regular terms.} \tag{13.82}$$

In the previous section we computed the perturbative result,

$$\tilde{G}(p) = \frac{i}{p^2 - m^2} + \mathcal{O}(\lambda). \tag{13.83}$$

At leading order in perturbation theory, we can then identify the physical mass of a particle m_{phys} with the mass parameter m of our Lagrangian, and determine the normalisation $\tilde{Z}_\phi = 1$.

We then find that the diagrams G_a, G_b, G_c, which do not connect all external particles with each other, do not contribute to the scattering amplitude $\langle p_3, p_4 | \, p_1, p_2 \rangle$. We find for example that, at leading order in λ,

$$\frac{G_a}{\tilde{G}(p_1)\tilde{G}(p_2)\tilde{G}(p_3)\tilde{G}(p_4)} \sim (p_3^2 - m^2)(p_4^2 - m^2)\delta(p_4 - p_1)\delta(p_3 - p_2)\Big|_{p_i^2 = m^2} = 0. \tag{13.84}$$

This "0" should be read with care. The diagrams G_a and G_c describe particles passing each other without scattering; with properly normalised wave packets they reproduce the identity part of $S = \mathbf{1} + iT$, $\langle p_3, p_4 | \, p_1, p_2 \rangle \supset (2\pi)^6 \, 4E_1E_2 \, [\delta^3(\vec{p}_1 - \vec{p}_3)\delta^3(\vec{p}_2 - \vec{p}_4) + (3 \leftrightarrow 4)]$, which is not visible in the naive on-shell limit (a product of a vanishing factor and a δ-function at the same point). What is true is that they do not contribute to the transition matrix T, which is given by the connected, amputated diagrams (Srednicki, Ch. 5 and 10; Weinberg QFT I, Sec. 4.3). The diagrams G_d and G_e connect all external particles and have poles $\frac{i}{p_i^2 - m^2}$ for each one of them. These contribute to the scattering amplitude.

To perform the division by the external propagators, as demanded by the LSZ formula, we need two more rules in our set of Feynman rules.

- Lines starting from an external point contribute a factor 1 (and not $i/(p^2 - m^2)$).

We must be, though, a bit more careful. Eq. 13.83 has perturbative corrections of order λ. We need to divide our Feynman diagrams by the full perturbative expansion of the propagator and not just by the leading-order contribution.

We can account for this division entirely with one more clever trick.

- Throw away all Feynman diagrams which "correct" external lines.

For example, the diagrams of Fig. 13.2 should all be disregarded when computing truncated Green's functions for physical scattering amplitudes.

With these rules we can compute any scattering amplitude of the ϕ^4 theory in perturbation theory. However, many Feynman diagrams contain closed loops, such as the second diagram of Eq. 13.48, which require integrations over unconstrained loop momenta. Many of these integrals are divergent! We will learn how to compute loop integrals, and how to make sense of their infinities, in Part II. There we develop perturbation theory for quantum electrodynamics, the quantum field theory of electrons and photons, up to one loop and its renormalisation.

PART II

Quantum Electrodynamics

Preface to Part II

This part takes the first steps from the free fields of Part I towards the Standard Model as a quantum field theory, with quantum electrodynamics as the guiding example. It joins two sources: the last chapters of the QFT I course (loop integrals, QED at one loop, renormalisation) and the QED lectures of the Standard Model course. Both describe QED; here they are merged into one treatment. It is meant to be read between Parts I and III; the path-integral methods of Part III can be studied in parallel.

Chapter 1 recalls the particle content of the Standard Model and derives the QED Lagrangian from two principles, local gauge invariance and locality, together with the counting of operator dimensions. Chapter 2 expands the S-matrix in perturbation theory and uses Wick's theorem to obtain the Feynman rules of QED, with the vacuum-to-vacuum amplitude as a worked example.

Chapters 3 and 4 are devoted to loop integrals. We see why loop integrals need a regulator, introduce dimensional regularisation, evaluate the tadpole and the one-loop bubble, and develop the Feynman-parameter representation of one-loop integrals. We then learn how to relate integrals to each other, with integration by parts, dimensional shifts and the reduction of tensor integrals to scalar ones.

Chapter 5 implements dimensional regularisation at the level of the QED Lagrangian and computes the one-loop corrections to the electron and photon propagators; we sum them to all orders, determine the electron mass, and derive the Ward identity. Chapter 6 removes the ultraviolet divergences by renormalisation of the fields, the mass and the coupling in the $\overline{\text{MS}}$ scheme, and derives the running of the QED coupling.

Chapter 7 puts all these tools to work in a complete next-to-leading-order calculation in QCD: the production of a Higgs boson in gluon fusion, $gg \to h$, at a hadron collider. In the limit of a heavy top quark the calculation becomes a one-loop computation in an effective theory. It shows, in a realistic setting, how ultraviolet divergences are removed by renormalisation, how soft divergences cancel between virtual and real corrections, and how the remaining collinear divergences are absorbed into the parton densities.

Exercises are interspersed in the text; they are an essential part of the material.

1

Introduction: the Standard Model and gauge symmetry

1.1 The Standard Model

The Standard Model is the theory of the known elementary particles.

- It is a local non-abelian gauge theory.
- Its symmetry group is $SU(3)_c \times SU(2)_L \times U(1)_Y$; the electroweak part is $SU(2)_L \times U(1)_Y$.
- Matter particles (fundamental representation):

$$\text{Leptons:} \quad \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \qquad e_R,$$

$$\text{Quarks:} \quad \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \qquad u_R,\ d_R,$$

  plus two more families.

- Gauge bosons (adjoint representation): $W_\mu^{1,2,3}$, B_μ and the gluons. After symmetry breaking the electroweak bosons appear as the mass eigenstates

$$A^\mu, \quad Z^\mu, \quad W^{+\mu}, \quad W^{-\mu}.$$

- Higgs boson (symmetry breaking): H.

Only the left-handed fields form $SU(2)_L$ doublets, ordered with the $T_3 = +\frac{1}{2}$ member on top; e_R, u_R, d_R are $SU(2)_L$ singlets. A right-handed neutrino ν_R is not part of the minimal Standard Model. The photon A^μ and the Z^μ are mixtures of W_μ^3 and B_μ (see the PDG review "Electroweak model and constraints on new physics").

Aim.. Understand the Standard Model as a quantum field theory → make predictions.

Prerequisites.. Introductory QFT (Part I); the path-integral methods of Part III can be studied in parallel.

1.2 The wider programme

A complete course on the Standard Model as a quantum field theory covers the topics below. This part treats the first two of them for QED: the Feynman rules of QED, the tools for one-loop integrals in dimensional regularisation, the electron and photon propagators at one loop, the Ward identity, and the renormalisation of QED with the running of its coupling. It closes with a complete next-to-leading-order calculation in QCD, Higgs boson production in gluon fusion, in which ultraviolet renormalisation, infrared divergences and factorisation all appear. Many of the other topics are developed in Parts III and IV.

- QED: Lagrangian, quantisation, derivation of Feynman rules, Ward identity and quantum symmetries of QED.
- Loop corrections, UV divergences, renormalisation of QED, renormalisation group evolution in QED.
- QCD as a gauge theory, short description of quantisation (details in Part III), Feynman rules, basic processes at tree level, renormalisation, running of the strong coupling, asymptotic freedom.
- $e^+e^- \to$ hadrons, optical theorem.
- Landau singularities, infrared divergences, strategy of regions, "cancellation" of infrared divergences.
- Factorisation, resummation.
- Non-perturbative aspects.
- Spontaneous symmetry breaking, abelian Higgs model as a quantum theory, R_ξ gauge fixing, unitary gauge, symmetries and "Ward" identities.
- Renormalisation of the abelian Higgs model.
- Standard Model at tree level.
- Dirac and Majorana fermions.
- Feynman rules with Majorana fermions.
- Neutrinos, neutrino oscillations, see-saw mechanism.
- Flavour physics and precision tests from loop-generated processes.
- Electroweak precision tests from lepton colliders.
- Higgs boson physics, production and decay.
- Extensions of the Standard Model and effective field theory methods.

1.3 QED

One of the most amazing successes of quantum field theory is the precision with which it describes the interaction of radiation and matter. The rules governing these interactions are very simple, and are given by the theory of Quantum Electrodynamics (QED). What is also astonishing is that QED emerges naturally by combining two very simple principles: symmetry and locality. The QED Lagrangian is symmetric under phase-redefinitions ($U(1)$ transformations) of the electron field which are local, i.e. we can choose the symmetry transformation parameters differently at each space-time point.

Local symmetry transformations govern all the quantum field theories which describe the three forces of nature other than gravity: QED, QCD and the Standard Model of electroweak interactions. These theories describe physical phenomena extremely well within the experimentally accessible accuracy.

1.3.1 Gauge invariance

We consider a free electron field:

$$\mathcal{L} = \bar{\psi}(x)\,(i\partial\!\!\!/ - m)\,\psi(x). \tag{1.1}$$

As we have discussed already in Part I, this Lagrangian is invariant under a $U(1)$ transformation,

$$\psi(x) \to \psi'(x) = \exp(ie\theta)\psi(x), \tag{1.2}$$

if the phase θ is global, i.e. it is chosen to be the same at every point in space-time $\left(\frac{\partial\theta}{\partial x} = 0\right)$.

Let us now take a bold step and ask whether we can have a different phase θ at every space-time point x^μ,

$$\psi(x) \to \psi'(x) = U(x)\,\psi(x), \qquad U(x) = \exp\left(ie\theta(x)\right). \tag{1.3}$$

The Lagrangian is no longer invariant,

$$\bar{\psi}\partial\!\!\!/\psi \to \bar{\psi}\partial\!\!\!/\psi + \bar{\psi}e^{-ie\theta}\left[\partial\!\!\!/ e^{ie\theta}\right]\psi. \tag{1.4}$$

The problem is that the space-time derivative does not transform simply under the local $U(1)$ transformation any more,

$$\partial\!\!\!/\psi \to e^{ie\theta}\partial\!\!\!/\psi + \text{ something else }.$$

If we insist on locality, we shall need to modify the derivative so that it transforms more conveniently. We will look for a new derivative which, under a *local* $U(1)$ transformation, transforms as

$$\not{D}\psi \to U(x)\not{D}\psi. \tag{1.5}$$

The simplest modification we can think of is to add to the space-time derivative a space-time function,

$$\partial^\mu \to D^\mu = \partial^\mu + \text{function}^\mu, \tag{1.6}$$

which transforms in the same way as the space-time derivative under Lorentz transformations, i.e. it transforms as a vector. A function of space-time, in field theory, is nothing else but a field. We are then proposing to modify the definition of a derivative by adding to it a vector field. We write a "covariant derivative",

$$D_\mu = \partial_\mu - ieA_\mu(x), \tag{1.7}$$

where the $-ie$ factor is conventional.

The vector field $A^\mu(x)$ must have a very specific transformation, which we can find by requiring that the modified derivative transforms simply. We require that

$$\begin{aligned} & D_\mu\psi(x) \to D'_\mu\psi' = U(x)D_\mu\psi \\ \rightsquigarrow\ & \left(\partial_\mu - ieA'_\mu\right)\left(U(x)\psi\right) = U(x)\left(\partial_\mu - ieA_\mu\right)\psi \\ \rightsquigarrow\ & U(x)\partial_\mu\psi + \left[\partial_\mu U(x)\right]\psi - ieA'_\mu U(x)\psi = U(x)\partial_\mu\psi - ieA_\mu U(x)\psi \\ \rightsquigarrow\ & A'_\mu = A_\mu - \frac{i}{e}U^{-1}(x)\partial_\mu U(x). \end{aligned} \tag{1.8}$$

This is an astonishing result. We find that the vector field transforms as,

$$A'_\mu = A_\mu + \partial_\mu\theta, \tag{1.9}$$

which is a gauge transformation, the transformation of the photon field. In other words, if we would like a Lagrangian of electrons to be invariant under local $U(1)$ transformations, then the same Lagrangian must also describe a photon field.

The covariant derivative transforms as:

$$\begin{aligned} & D_\mu \to D'_\mu = \partial_\mu - ieA'_\mu \\ = \ & \partial_\mu - ie\left(A_\mu - \frac{i}{e}U^{-1}\partial_\mu U\right) \\ = \ & \partial_\mu - ieA_\mu - U^{-1}\left(\partial_\mu U\right) \\ = \ & \partial_\mu - ieA_\mu + U\left(\partial_\mu U^{-1}\right) \\ = \ & U(x)\left(\partial_\mu - ieA_\mu\right)U^{-1}(x). \end{aligned} \tag{1.10}$$

Therefore:

$$D_\mu \to D'_\mu = U(x)D_\mu U^{-1}(x). \tag{1.11}$$

Exercise: *Verify the last step of the derivation above: the covariant derivative acts as an operator on everything to its right, and $U\,\partial_\mu\left(U^{-1}f\right) = \partial_\mu f + U\left(\partial_\mu U^{-1}\right)f$ for any function f.*

We can now replace the free Lagrangian of the spin-1/2 field with a new Lagrangian which is symmetric locally,

$$\begin{aligned} & \mathcal{L} = \bar\psi\left[i\not{D} - m\right]\psi \\ & \to \bar\psi U^{-1}U\left[i\not{D} - m\right]U^{-1}U\psi \\ & = \bar\psi\left[i\not{D} - m\right]\psi. \end{aligned}$$

If A_μ is a physical field, we need to introduce a kinetic term in the Lagrangian for it. We will insist on constructing a fully gauge-invariant Lagrangian. For this purpose, we can use the covariant derivative as a building block. Consider the gauge transformation of the product of two covariant derivatives:

$$\begin{aligned} & D_\mu D_\nu \to D'_\mu D'_\nu = UD_\mu U^{-1}UD_\nu U^{-1} \\ & = UD_\mu D_\nu U^{-1}. \end{aligned}$$

This is not a gauge-invariant object. Now look at the commutator:

$$\begin{aligned} & \left[D_\mu, D_\nu\right] \to \left[D'_\mu, D'_\nu\right] \\ & = U\left[D_\mu, D_\nu\right]U^{-1}. \end{aligned} \tag{1.12}$$

This is gauge invariant. To convince ourselves we write the commutator explicitly:

$$\begin{aligned} & \left[D_\mu, D_\nu\right] = \left(\partial_\mu - ieA_\mu\right)\left(\partial_\nu - ieA_\nu\right) - \left[\mu \leftrightarrow \nu\right] \\ = \;& \partial_\mu\partial_\nu - ie\left(\partial_\mu A_\nu\right) - ieA_\nu\partial_\mu - ieA_\mu\partial_\nu + (ie)^2 A_\mu A_\nu - \left[\mu \leftrightarrow \nu\right] \\ = \;& -ie\left[\partial_\mu A_\nu - \partial_\nu A_\mu\right]. \end{aligned} \tag{1.13}$$

The commutator contains no derivative operators any more; it is a multiplicative function of x, which commutes with $U(x)$. Inserting Eq. 1.13 into Eq. 1.12, we find therefore that the commutator of covariant derivatives (in the abelian case) is gauge invariant,

$$\left[D_\mu, D_\nu\right] \to U(x)\left[D_\mu, D_\nu\right]U^{-1}(x) = \left[D_\mu, D_\nu\right]. \tag{1.14}$$

We have also found that it is proportional to the field strength tensor of the gauge (photon) field:

$$F_{\mu\nu} = \frac{i}{e}\left[D_\mu, D_\nu\right] = \partial_\mu A_\nu - \partial_\nu A_\mu, \qquad F_{\mu\nu} \to F'_{\mu\nu} = F_{\mu\nu} \quad \text{(gauge invariant)}. \tag{1.15}$$

More terms are allowed by the symmetry in the Lagrangian. With $F_{\mu\nu}$ we can form scalar products, e.g.

$$F_{\mu\nu}F^{\mu\nu} \propto \vec{E}^{\,2} - \vec{B}^{\,2}, \qquad \epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma} \propto \vec{E}\cdot\vec{B}, \qquad F^{\mu\nu}F_{\nu\rho}F^{\rho}{}_{\mu}, \quad \ldots \tag{1.16}$$

- Two dimension-4 operators: $F_{\mu\nu}F^{\mu\nu}$ and $\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma}$. The latter can be written as a total derivative.
- Higher-dimensional operators, e.g.

$$\frac{1}{\Lambda}\bar{\psi}\sigma^{\mu\nu}\psi F_{\mu\nu}, \qquad \frac{1}{\Lambda^4}\left(F_{\mu\nu}F^{\mu\nu}\right)^2, \tag{1.17}$$

  require a "mass scale" Λ. It is set by heavier particles which are not part of the theory: for the QED of electrons, the muon mass and the hadronic scale, and ultimately the W-boson mass m_W of the electroweak sector.

Operators with fermions start at dimension 5 (the Pauli term above), hence the $O(1/\Lambda)$ in eq. (1.18) below. A pure-photon operator such as $F_{\mu\nu}F^{\nu\rho}F_{\rho}{}^{\mu}$, which one might write down first, vanishes identically for an abelian field strength (it is the trace of the cube of an antisymmetric matrix) and is also C-odd. The lowest pure-photon operators beyond F^2 and $F\tilde{F}$ are of dimension 8, $(F_{\mu\nu}F^{\mu\nu})^2$ and $(F_{\mu\nu}\tilde{F}^{\mu\nu})^2$ (Euler–Heisenberg).

Exercise: *Show that $\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma}$ can be written as a total derivative.*

We now have invariant terms for a Lagrangian with an "electron" and a "photon" field. At low energies, compared with the scale Λ, the classical Lagrangian for QED reads

$$\mathcal{L} = \bar{\psi}\left(i\not{D} - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + O\left(\frac{1}{\Lambda}\right). \tag{1.18}$$

This Lagrangian accounts for the majority of the phenomena that we experience in nature. It is a beautiful synthesis of locality and symmetry! In what follows we drop the operators suppressed by $1/\Lambda$.

- It is gauge invariant.
- It is renormalisable, i.e. UV infinities can be absorbed into field, mass and coupling redefinitions. We will see how this works at one loop in Chapter 6.

- It is an interacting theory. The Lagrangian contains a $\bar{\psi}\not{A}\psi$ term, a photon–electron–positron interaction (the positron is the anti-electron).
- The theory cannot be solved exactly. But we can use perturbation theory around the $e = 0$ limit (zero electric charge limit). This is the subject of the next chapter.

Exercise: *Find the Noether current and conserved charge due to the invariance under the $U(1)$ gauge transformation of the QED Lagrangian.*

2

Scattering in QED

In this chapter we derive the Feynman rules of QED from the perturbative expansion of the S-matrix, using Wick's theorem.

2.1 The QED Lagrangian

Let us take the QED Lagrangian of Eq. 1.18 and substitute the expression for the covariant derivative. We have,

$$\begin{aligned} \mathcal{L} &= \bar{\psi}\left[i\partial\!\!\!/ - m\right]\psi \qquad \text{(free electron field)} \\ &+ -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \qquad \text{(free photon field)} \\ &+ e\bar{\psi}A\!\!\!/\psi \qquad \text{(electron-photon interaction).} \end{aligned} \tag{2.1}$$

The last term, which couples the electron and photon fields, differentiates the QED Lagrangian from the Lagrangian of free electron and photon fields. Can we solve for the energy eigenstates of the QED Hamiltonian? No, unless we resort to perturbation theory. The strength of the photon-electron interaction is very small. We can consider the last term a small perturbation, and expand probability amplitudes around $e = 0$.

For $e = 0$, we should recover the results from the quantisation of the free photon and free electron fields of Part I. We then encounter the same problems as when we quantised the free electromagnetic field, which required us to fix the gauge for the field A^μ. We then add to the QED Lagrangian a term,

$$\mathcal{L}_{\text{gauge-fix}} = -\frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2 . \tag{2.2}$$

In this way, we can at least be sure that the theory is correctly defined for $e = 0$. It is not obvious that this modification will solve the problem of obtaining a consistently quantised theory at higher orders in perturbation theory. It actually does! But the justification of this statement will be a very important topic of study in QFT II (Part III). As a primitive diagnostic, we shall consider ξ an arbitrary parameter in our calculations. We will then check that higher-order corrections do not modify it.

The QED Lagrangian is then

$$\mathcal{L}_{\text{QED}} = \underbrace{\mathcal{L}_e + \mathcal{L}_\gamma}_{\text{free}} + \mathcal{L}_{\text{interaction}}, \tag{2.3}$$

where

$$\mathcal{L}_e = \bar{\psi}\left[i\partial\!\!\!/ - m\right]\psi, \tag{2.4}$$

$$\mathcal{L}_\gamma = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2, \tag{2.5}$$

$$\mathcal{L}_{\text{interaction}} = +e\,\bar{\psi}A\!\!\!/\psi. \tag{2.6}$$

The interaction term mixes photons, electrons and positrons. We can develop a formalism for perturbation theory similar to the one for a scalar field theory in Part I. This leads to the derivation of Feynman rules for the S-matrix elements of QED transition amplitudes, which we carry out in the rest of this chapter.

2.2 "In" and "out" states

Quantum states in QED describe free particles at $t = -\infty$ or $t = +\infty$, i.e. interactions should happen for a short time and particles move freely most of the time. We need the matrix element

$$\langle \text{out}|\ \text{in}\rangle\,, \tag{2.7}$$

where $|\text{in}\rangle$ is a state where the particles are free at $t = -\infty$ and $|\text{out}\rangle$ is a state where the particles are free at $t = +\infty$. $|\text{in}\rangle$ and $|\text{out}\rangle$ are eigenstates of the full QED Hamiltonian (cf. the chapter on scattering theory in Part I).

2.3 Perturbation theory

In perturbation theory,

$$\langle \text{out}|\ \text{in}\rangle = \langle \text{final}|\, S\, |\text{initial}\rangle\,, \tag{2.8}$$

where $|\text{initial}\rangle$ is an eigenstate of $H_e + H_\gamma \equiv H_{\text{free}}$ and describes the same physics as $|\text{in}\rangle$ at $t = -\infty$; $|\text{final}\rangle$ is an eigenstate of $H_e + H_\gamma \equiv H_{\text{free}}$ as well and describes the same physics as $|\text{out}\rangle$ at $t = +\infty$. The S-matrix is

$$S = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int d^4x_1 \ldots d^4x_n\, T\left(H_{\text{int}}(x_1)\ldots H_{\text{int}}(x_n)\right), \tag{2.9}$$

where T orders operators from later to earlier times, and

$$H_{\text{int}}(x) \equiv -\mathcal{L}_{\text{interaction}}(x) = -e\,\bar{\psi}(x)A\!\!\!/(x)\psi(x), \tag{2.10}$$

with the fields in the interaction picture.

- For boson fields

$$TA^\mu(x)A^\nu(y) \equiv A^\mu(x)A^\nu(y)\,\theta(x^0 - y^0) + A^\nu(y)A^\mu(x)\,\theta(y^0 - x^0). \tag{2.11}$$

- For fermion fields

$$T\psi_a(x)\bar{\psi}_b(y) = \psi_a(x)\bar{\psi}_b(y)\,\theta(x^0 - y^0) - \bar{\psi}_b(y)\psi_a(x)\,\theta(y^0 - x^0). \tag{2.12}$$

To ensure that $\langle 0|\, H\, |0\rangle = 0$ we need to use normal ordering,

$$H \to :H: \quad \rightsquigarrow \quad H_{\text{int}} \to :H_{\text{int}}:. \tag{2.13}$$

Thus,

$$S = 1 + ie \int d^4x \, \left[:\bar\psi(x) \not{A}(x) \psi(x):\right] + \frac{(ie)^2}{2!} \int d^4x \, d^4y \, T\left[:\bar\psi(x) \not{A}(x) \psi(x): :\bar\psi(y) \not{A}(y) \psi(y):\right] + \ldots \tag{2.14}$$

2.4 Contractions and Wick's theorem

Contraction.. We define a contraction as

$$T(AB) - :AB: \equiv \overbrace{A\, B}. \tag{2.15}$$

In QED (see Section 11.7 of Part I)

$$\overbrace{A^\mu(x_1)\, A^\nu(x_2)} = -iD^{\mu\nu}(x_1 - x_2), \tag{2.16}$$

$$\overbrace{\psi_a(x_1)\, \bar\psi_b(x_2)} = iS_{F,ab}(x_1 - x_2), \tag{2.17}$$

$$\overbrace{\bar\psi_b(x_2)\, \psi_a(x_1)} = -iS_{F,ab}(x_1 - x_2), \tag{2.18}$$

where the photon and fermion propagators are

$$S_{F,ab}(x-y) = \int \frac{d^4p}{(2\pi)^4} \frac{(\not{p} + m)_{ab}}{p^2 - m^2 + i\delta} e^{-ip\cdot(x-y)}, \qquad \delta \to 0, \tag{2.19}$$

$$D^{\mu\nu}(x-y) = \int \frac{d^4p}{(2\pi)^4} \frac{\left[g^{\mu\nu} - (1-\xi)\frac{p^\mu p^\nu}{p^2}\right]}{p^2 + i\delta} e^{-ip\cdot(x-y)}. \tag{2.20}$$

These are the Feynman propagators of the free fields which we have computed in Part I. All other contractions vanish:

$$\overbrace{\psi_a\, \psi_b} = \overbrace{\bar\psi_a\, \bar\psi_b} = \overbrace{\psi\, A} = \ldots = 0. \tag{2.21}$$

Wick's theorem..

1.

$$\begin{aligned} T(ABCD) = {} & :ABCD: \\ & + :\overbrace{A\,B}\,C\,D: + :\overbrace{A\,B\,C\,D}: + :\overbrace{A\,B\,C\,D}: \\ & + :A\,\overbrace{B\,C}\,D: + :A\,\overbrace{B\,C\,D}: + :A\,B\,\overbrace{C\,D}: \\ & + :\overbrace{A\,B}\,\overbrace{C\,D}: + :\overbrace{A\,\overbrace{B\,C}\,D}: + :\overbrace{A\,B\,\overbrace{C\,D}}: \end{aligned} \tag{2.22}$$

2.

$$T(A:BCD:) = :ABCD: + :\overbrace{A\,B}\,C\,D: + :\overbrace{A\,B\,C\,D}: + :\overbrace{A\,B\,C\,D}: \tag{2.23}$$

i.e. contractions among fields inside the same normal-ordered product do not appear.
Remark. As written, 1 and 2 hold for bosonic fields. For fermionic fields each term carries the sign $(-1)^P$ of the permutation P that brings the contracted fields next to each other (and the uncontracted ones into the order written inside $:\ldots:$); e.g. the term with contractions AC and BD in 1 has a factor -1 when all four are fermion fields (Peskin–Schroeder, Sec. 4.7).

Exercise: *Verify 1 and 2 for fermion and photon fields.*

2.4.1 A vanishing contraction

We have stated that in

$$:\bar\psi \not{A} \psi: \ldots :\bar\psi \not{A} \psi:$$

contractions inside a normal-ordered product do not need to be taken into account. One may object that a Feynman diagram of this type nevertheless exists:

$$:\bar\psi \not{A} \psi: \ldots \propto \int_{-\infty}^{+\infty} \frac{d^4k}{(2\pi)^4} \frac{\mathrm{tr}\left((\not{k}+m)\gamma^\mu\right)}{k^2-m^2} \ldots . \tag{2.24}$$

This Feynman diagram is zero. The first term $(\not{k})$ in the numerator is odd under $k \leftrightarrow -k$ and therefore vanishes upon integration. The second term gives

$$\propto m \underbrace{\mathrm{tr}(\gamma^\mu)}_{=0} \tag{2.25}$$

and vanishes since γ-matrices are traceless.

2.5 Vacuum-to-vacuum transition

Let us compute

$$\begin{aligned}\langle 0|\, S\, |0\rangle = \underbrace{\langle 0|\, 0\rangle}_{1} &+ ie\, \langle 0|\, T\, \bar\psi(x)\not{A}(x)\psi(x)\, |0\rangle \\ &- \frac{e^2}{2!}\, \langle 0|\, T :\bar\psi(x)\not{A}(x)\psi(x):\bar\psi(y)\not{A}(y)\psi(y):|0\rangle + \ldots\end{aligned} \tag{2.26}$$

- The second term vanishes:

$$\begin{aligned}\langle 0|\, T\, \bar\psi_a(x)\not{A}_{ab}(x)\psi_b(x)\, |0\rangle &= \langle 0|\, \bar\psi_a(x)\, \not{A}_{ab}(x)\, \psi_b(x)\, |0\rangle \\ &= \langle 0|\, \not{A}_{ab}(x)\, |0\rangle \left(-iS_{F,ba}(x-x)\right),\end{aligned} \tag{2.27}$$

  where a, b are spinor indices, $\not{A}_{ab} \equiv A_\mu \gamma^\mu_{ab}$. But

$$\langle 0|\, A\, |0\rangle = \langle 0|\, \big(\underbrace{A^{(+)}}_{\to 0} + \underbrace{A^{(-)}}_{\to 0}\big)\, |0\rangle = 0. \tag{2.28}$$

- The last term gives

$$\begin{aligned}&-\frac{e^2}{2}\, \langle 0|\, T :\bar\psi \not{A}\psi:\bar\psi\not{A}\psi:|0\rangle \\ &= -\frac{e^2}{2}\, \langle 0|\, T :\bar\psi_a A^\mu \psi_b:\bar\psi_c A^\nu \psi_d:|0\rangle\, (\gamma_\mu)_{ab}(\gamma_\nu)_{cd}.\end{aligned} \tag{2.29}$$

  The only terms which survive the vacuum expectation value are the fully contracted ones:

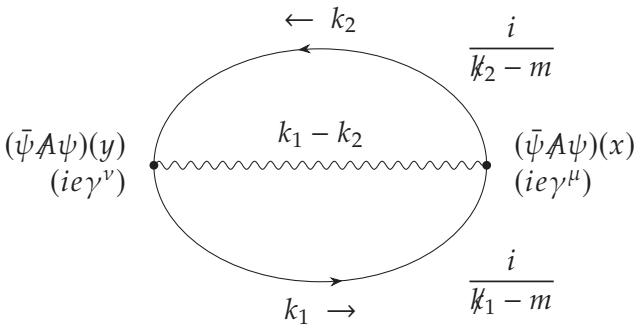
$$\begin{aligned}&= -\frac{e^2}{2}\, \langle 0|\, \big(\bar\psi_a\, A^\mu\, \psi_b\big)_x \big(\bar\psi_c\, A^\nu\, \psi_d\big)_y\, |0\rangle\, (\gamma_\mu)_{ab}(\gamma_\nu)_{cd} \\ &= -\frac{e^2}{2}\, \left(-iD^{\mu\nu}(x-y)\right) \left(iS_{F,bc}(x-y)\right) \left(-iS_{F,da}(y-x)\right) (\gamma_\mu)_{ab}(\gamma_\nu)_{cd} \\ &= \frac{(-1)}{2}\, (ie\gamma_\mu)_{ab}\, iS_{F,bc}(x-y)\, (ie\gamma_\nu)_{cd}\, iS_{F,da}(y-x) \left[-iD^{\mu\nu}(x-y)\right].\end{aligned} \tag{2.30}$$

Let us now integrate over x and y:

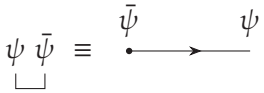
$$\begin{aligned}
&\frac{(-1)}{2}\int d^4x\, d^4y\ \mathrm{tr}\left[(ie\gamma_\mu)\, iS_F(x-y)(ie\gamma_\nu)\, iS_F(y-x)\right]\left(-iD^{\mu\nu}(x-y)\right)\\
&=\frac{(-1)}{2}\int d^4x\, d^4y\ \mathrm{tr}\Bigg[(ie\gamma_\mu)\int\frac{d^4k_1}{(2\pi)^4}\,e^{-ik_1\cdot(x-y)}\,\frac{i(\not{k}_1+m)}{k_1^2-m^2+i\delta}\\
&\qquad\cdot(ie\gamma_\nu)\int\frac{d^4k_2}{(2\pi)^4}\,e^{-ik_2\cdot(y-x)}\,\frac{i(\not{k}_2+m)}{k_2^2-m^2+i\delta}\Bigg]\\
&\qquad\cdot\int\frac{d^4k_3}{(2\pi)^4}\,e^{-ik_3\cdot(x-y)}\,\frac{i\left[-g^{\mu\nu}+(1-\xi)\frac{k_3^\mu k_3^\nu}{k_3^2}\right]}{k_3^2+i\delta}\\
&=\frac{(-1)}{2}\left(\int d^4x\right)\int\frac{d^4k_1}{(2\pi)^4}\frac{d^4k_2}{(2\pi)^4}\frac{d^4k_3}{(2\pi)^4}\,(2\pi)^4\delta^{(4)}(k_1+k_3-k_2)\\
&\qquad\cdot\mathrm{tr}\left[(ie\gamma_\mu)\frac{i(\not{k}_1+m)}{k_1^2-m^2}(ie\gamma_\nu)\frac{i(\not{k}_2+m)}{k_2^2-m^2}\right]\frac{i\left[-g^{\mu\nu}+(1-\xi)\frac{k_3^\mu k_3^\nu}{k_3^2}\right]}{k_3^2+i\delta}\\
&=\frac{(-1)}{2}\,(\text{Volume})\cdot\left[\int\frac{d^4k_1}{(2\pi)^4}\frac{d^4k_2}{(2\pi)^4}\,\mathrm{tr}\left[(ie\gamma_\mu)\left[\frac{i(\not{k}_1+m)}{k_1^2-m^2}\right](ie\gamma_\nu)\left[\frac{i(\not{k}_2+m)}{k_2^2-m^2}\right]\right]\right.\\
&\qquad\left.\cdot\left[\frac{-i\left[g^{\mu\nu}-(1-\xi)\frac{(k_2-k_1)^\mu(k_2-k_1)^\nu}{(k_1-k_2)^2}\right]}{(k_2-k_1)^2}\right]\right].
\end{aligned}\tag{2.31}$$

2.5.1 Graphical representation

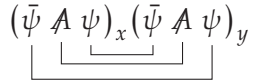
Let us look at this graphically:



Fermion arrow:



Recall:



2.6 QED Feynman rules

We can now read off the Feynman rules of QED. Essentially, we can picture QED processes as photons and electrons propagating in between interactions, where a photon is absorbed or emitted by electrons. For every photon absorption or emission we need a factor of e, which is a number representing the strength of the interaction. We then have to include all possibilities for a scattering process to happen, with as many interaction vertices as the maximum order in our perturbative series. The rules are:

- For the propagation of a photon with momentum p, assign the factor



$$\equiv -iD^0_{\mu\nu}(p) = \frac{i}{p^2 + i\delta}\left[-g_{\mu\nu} + (1-\xi)\frac{p_\mu p_\nu}{p^2 + i\delta}\right]. \tag{2.32}$$

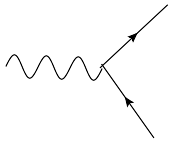
  $D^0_{\mu\nu}(p)$ is the Fourier transform of $D_{\mu\nu}(x-y)$ in Eq. 2.20.

- For the propagation of an electron with momentum p along the arrow, assign the factor

$$\equiv iS_0(p) = \frac{i}{\not{p} - m + i\delta}, \tag{2.33}$$

  the Fourier transform of $iS_F(x-y)$ in Eq. 2.19.

- For photon absorption or emission, assign a factor



$$+ ie\gamma^\mu. \tag{2.34}$$

In addition,

- For each loop in a Feynman diagram, we perform an integration over the loop momentum

$$\int \frac{d^d k}{(2\pi)^d}, \tag{2.35}$$

  where d is the space-time dimension. For the moment $d = 4$; in Chapter 3 the dimension will become our regulator for divergent loop integrals.

- For each fermion loop we multiply with a factor

$$(-1). \tag{2.36}$$

- We divide each Feynman diagram by a "symmetry factor", accounting for the equivalent field permutations. The vacuum diagram above has the symmetry factor $\frac{1}{2}$.

Remark. The vertex is $+ie\gamma^\mu$ because $D_\mu = \partial_\mu - ieA_\mu$ gives $\mathcal{L}_{\text{int}} = +e\bar{\psi}\not{A}\psi$. The familiar $-ie\gamma^\mu$ (Peskin–Schroeder) belongs to $D_\mu = \partial_\mu + ieA_\mu$; the sign of e is a convention and drops out of all observables, but the vertex and the covariant derivative must be used consistently. Here the momentum k of a fermion propagator flows along the arrow (from $\bar{\psi}$ to ψ).

2.7 Higher-order terms

At third order the vacuum-to-vacuum amplitude receives no contribution, since an odd number of photon fields cannot be fully contracted,

$$\frac{(ie)^3}{3!}\int d^4x_1\, d^4x_2\, d^4x_3\; T\left[(\bar\psi \not{A}\psi)_{x_1}(\bar\psi \not{A}\psi)_{x_2}(\bar\psi \not{A}\psi)_{x_3}\right] \;\to\; 0 \tag{2.37}$$

and at fourth order

$$\frac{(ie)^4}{4!}\int d^4x_1 \ldots d^4x_4\; T\left[(\bar\psi \not{A}\psi)_{x_1}(\bar\psi \not{A}\psi)_{x_2}\ldots(\bar\psi \not{A}\psi)_{x_4}\right]. \tag{2.38}$$

Exercise: *Check that the same Feynman rules apply for some of the non-zero contractions.*
Exercise: Calculate the photon–electron–positron vertex. Take

$$|\text{in}\rangle \equiv \left|\gamma, p_1\right\rangle, \qquad |\text{out}\rangle \equiv \left|e^-, p_2;\, e^+, p_3\right\rangle, \tag{2.39}$$

and compute

$$\left\langle e^-, p_2;\, e^+, p_3\right| S \left|\gamma, p_1\right\rangle = \ldots, \qquad \left\langle \gamma, p_1\right| S \left|e^-, p_2;\, e^+, p_3\right\rangle \tag{2.40}$$

up to the first non-vanishing term. The result will be proportional to

$$\delta(p_1 - p_2 - p_3). \tag{2.41}$$

Is this condition ever satisfied for physical momenta?

2.8 External particles

The LSZ reduction formula of Part I is slightly modified for fermions and photons, where the constants $\tilde{Z}^{1/2}$ are replaced by $\tilde{Z}^{1/2}$ times a "spin factor" which is different from one. For truncated external lines, we have the rules:

- multiply with a factor
  $$\bar u(p),$$
  for each outgoing fermion,
- a factor
  $$v(p),$$
  for each outgoing anti-fermion,
- a factor
  $$u(p),$$
  for each incoming fermion,
- a factor
  $$\bar v(p),$$
  for each incoming anti-fermion,
- and a factor
  $$\epsilon^\mu(p)$$
  for each incoming external photon, and $\epsilon^{*\mu}(p)$ for each outgoing photon.

Exercise: *Derive the above Feynman rules for incoming and outgoing particles, by contracting the fields of S with the creation and annihilation operators of the "in" and "out" states, as in the previous exercise.*

3

Dimensional regularisation and one-loop integrals

In order to compute the perturbative expansion of any Green's function, we must perform unrestricted integrations over the momenta of particles circulating in loops of Feynman diagrams. The exact evaluation of loop integrals is a difficult and often prohibitive task. Many loop integrals are divergent! This is a very unpleasant surprise if we aim towards a realistic description of physical phenomena with finite probabilities. The problem of divergences is a very serious one, and it jeopardises the mathematical foundation of the perturbative method.

3.1 Why we need a regulator

Perturbation theory in QED and more generally in quantum field theories generates loop diagrams, which require an integration from $-\infty$ to $+\infty$ for every component of the loop momenta. For example, at one loop for $e^+e^- \to e^+e^-$ in QED we find diagrams such as those of Fig. 3.1.

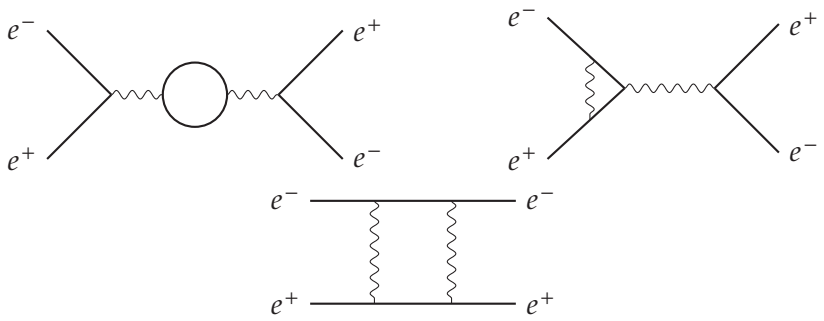


Figure 3.1. Examples of one-loop diagrams for $e^+e^- \to e^+e^-$ in QED: a vacuum-polarisation insertion, a vertex correction and a box diagram.

It turns out that individual diagrams, or individual terms in some diagrams, are integrals which are **infinite** in $d = 4$ dimensions. However, the sum of all diagrams and ultraviolet counterterms due to the renormalisation procedure can be well defined and **finite**. In order to be able to compute the renormalised amplitude, we need a method to render every integral in it well defined. Such a method is a *regularisation method*. Let us first see the problem in the simplest example.

3.2 The simplest loop integral. Wick rotation

Let us consider the simplest loop integral in field theory,

$$I \equiv \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m^2 + i\delta}. \tag{3.1}$$

This is a formidable integral, given that it requires four integrations, one for each space-time dimension. We notice that only the combination $k^2 = k_0^2 - \vec{k}^2$ appears in the integrand. The integral would be easily solvable had we had a Euclidean metric for $k_E^2 = k_0^2 + \vec{k}^2$, by using spherical coordinates in four dimensions. With a Minkowski metric, we ought to treat the time integration specially. We write

$$I = \lim_{\delta \to 0_+} \int_{-\infty}^{+\infty} \frac{d^3\vec{k}}{(2\pi)^4} \int_{-\infty}^{+\infty} dk_0 \frac{1}{(k_0 - \omega_k + i\delta)(k_0 + \omega_k - i\delta)}, \tag{3.2}$$

with

$$\omega_k = \sqrt{\vec{k}^2 + m^2}. \tag{3.3}$$

We write explicitly the "$i\delta$" part of the denominator since it is crucial for the correct evaluation of the integral. The integrand has two poles, at $k_0 = \omega_k - i\delta$ and $k_0 = -\omega_k + i\delta$, which are found in the lower-right and upper-left quarter planes in the $(\Re k_0, \Im k_0)$ plane (see Fig. 3.2). We can perform the so-called Wick rotation, rotating the integration axis by

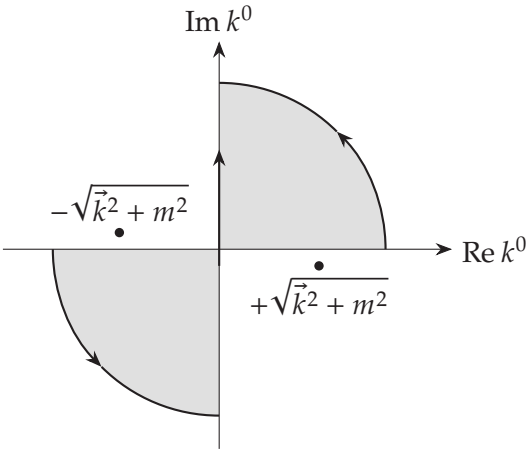


Figure 3.2. Wick rotation: the integrand has poles at $k_0 = \pm(\omega_k - i\delta)$ and no poles in the upper-right and lower-left (shaded) quadrants of the complex k_0 plane. We can rotate the integration axis by 90° and integrate along the imaginary axis, $k_0 \in [-i\infty, +i\infty]$.

90° and integrating over the imaginary axis,

$$I = \int_{-\infty}^{+\infty} \frac{d^3\vec{k}}{(2\pi)^4} \int_{-i\infty}^{+i\infty} dk_0 \frac{1}{k_0^2 - \vec{k}^2 - m^2 + i\delta}. \tag{3.4}$$

The rotation does not encounter any poles, and given that the integrand vanishes fast enough as $|\vec{k}|, |k_0| \to \infty$ (for some value of the dimension, as we will see shortly), the two integrations are equivalent. Setting

$$k_0 = ik_0',$$

the integral becomes

$$I = i \int_{-\infty}^{+\infty} \frac{d^3\vec{k}}{(2\pi)^4} \int_{-\infty}^{+\infty} dk_0' \frac{1}{-(k_0')^2 - \vec{k}^2 - m^2 + i\delta}. \tag{3.5}$$

Notice that after Wick's rotation we can consider the four-vector $k_E \equiv (k_0', \vec{k})$ in Euclidean space, and express k_E in four-dimensional spherical coordinates:

$$d^4k_E = d|k|\,|k|^3 d\Omega_4. \tag{3.6}$$

The integral then becomes

$$I = -i \frac{\Omega_4}{(2\pi)^4} \int_0^{\infty} d|k| \frac{|k|^3}{|k|^2 + m^2 - i\delta}. \tag{3.7}$$

Notice that the integral diverges as $|k| \to \infty$, which is an embarrassment. We can understand the origin of this divergence. Comparing with the Feynman-Stückelberg propagator of Eq. 11.35 of Part I, written as a four-dimensional momentum integral, we see that

$$I = -i \langle 0 | T \left\{ \phi(x)\phi(x) \right\} | 0 \rangle .$$

I is the amplitude for a particle to propagate from a point x back to the same point; it is the loop of the second diagram in Eq. 13.48 of Part I. Plane waves with very large momenta $|k|$ probe very short space-time separations, and the divergence at $|k| \to \infty$ reflects that we are asking about propagation over a vanishing distance.

Infinities were historically the biggest puzzle in the development of quantum field theory. But we now know what to do with them for many Lagrangian systems which are realistic descriptions of nature. We will learn later that we can "sweep infinities under the carpet", exploiting certain "freedoms" that we have. We are allowed to absorb infinities in the normalisation constants $\tilde{Z}_\phi$ of the LSZ formula, or in the definition of physical mass and coupling parameters in terms of the corresponding parameters appearing in the Lagrangian. These are very few "freedoms", and it is amazing that they suffice to solve the problem.

Infinities are not new in physics. Coulomb's law gives an infinite potential energy for two point charges at zero distance. In an atom, quantum mechanics prevents the electron from falling onto the nucleus, and the classical infinity turns out to be a sign that classical physics is incomplete at short distances. In the same spirit, the divergences of loop integrals, which originate from arbitrarily large momenta, may be a message that quantum field theory is not the final theory at extremely high energies. Fortunately, we do not need to know this final theory. Physics at extremely high energies, i.e. at extremely short time scales, should not affect in a noticeable way the dynamics that we observe at our energies; it can only affect the values of constants of nature, such as masses and coupling constants. This is precisely what renormalisation exploits.

The method of "sweeping infinities under the carpet" is called renormalisation. A prerequisite to it is to quantify the infinities with the help of a regulator. We shall introduce a parameter which renders the integrations well defined and finite except at a certain limit. For example, we can restrict the integration in Eq. 3.7 below a certain cut-off value Λ for $|k|$,

$$I(\Lambda) \equiv -i \frac{\Omega_4}{(2\pi)^4} \int_0^{\Lambda} d|k| \frac{|k|^3}{|k|^2 + m^2 - i\delta} = -i \frac{\Omega_4}{(2\pi)^4} \left[\frac{\Lambda^2}{2} + \dots \right]. \tag{3.8}$$

The regularised integral $I(\Lambda)$ is equal to the original integral I in the limit of an infinite Λ,

$$\lim_{\Lambda\to\infty} I(\Lambda) = I. \tag{3.9}$$

What is the benefit of introducing a regularisation method? It will give us the possibility to compute the divergent parts of all loop integrals that enter the evaluation of S-matrix elements for values of the regulator Λ where they are all well-defined. Then, we can perform "renormalisation" and express the parameters of the Lagrangian in terms of physical mass and coupling parameters. Then we shall take the limit of the regulator which corresponds to the original loop integrals. It will turn out that for many field theories the final renormalised S-matrix elements are finite.

3.3 Dimensional regularisation

A method to regularise loop integrals is not unique. A well-chosen regularisation procedure can be very beneficial for practical computations, since the difficulty in carrying out the integrations may vary significantly. What is more important, some regularisation methods may violate some of the symmetries of the theory. For example, a "cut-off" regularisation violates Lorentz invariance. When symmetries are broken due to the regularisation method, additional effort may be required in order to relate S-matrix elements to physical observables.

The most elegant regularisation method known to date was developed by 't Hooft and Veltman in the 1970s. It is called dimensional regularisation. 't Hooft and Veltman were awarded the Nobel prize in physics in 1999 for elucidating the quantum structure of the electroweak interactions, a work in which dimensional regularisation was an essential tool. It has revolutionised the evaluation of loop integrals for its simplicity. It is also known to preserve most of the symmetries which are found in physical Lagrangian systems.

In dimensional regularisation, the regulator is the number of space-time dimensions d. We shall explain how this works by revisiting the simplest example of a loop integral in Eq. 3.7. Let us now compute this integral in an arbitrary number of dimensions,

$$I_d = \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2 - m^2 + i\delta}, \tag{3.10}$$

which after Wick's rotation becomes

$$I_d = i \int \frac{d^d k}{(2\pi)^d} \frac{1}{-k^2 - m^2 + i\delta}, \tag{3.11}$$

and

$$k^2 = k_0^2 + \vec{k}^2. \tag{3.12}$$

In spherical coordinates,

$$I_d = \frac{i\Omega_d}{(2\pi)^d} \int_0^\infty dk \frac{k^{d-1}}{-k^2 - m^2 + i\delta}. \tag{3.13}$$

For infinitely large loop momenta it behaves as

$$I_d \to -\frac{i\Omega_d}{(2\pi)^d} \int^\infty dk\, k^{d-3}. \tag{3.14}$$

We observe that the integral has a finite ultraviolet limit ($k \to \infty$) for values of the dimension $d < 2$. In dimensional regularisation we compute the integral for a generic value of the dimension d for which it is well-defined. At the end of our computation, we shall perform an "analytic continuation" to a physical dimension value $d = 4$.

For physics predictions, we are interested in the values of loop integrals as Laurent expansions around the physical value of the dimension parameter, $d = 4$. It is customary to write

$$d = 4 - 2\epsilon, \tag{3.15}$$

and expand around $\epsilon = 0$. The $(-)$ sign is because we usually think of the dimension parameter as being smaller than four, in order to make an integral convergent in the UV limit $|k| \to \infty$. The factor of 2 in front of ϵ is convenient because the combination $\frac{d}{2}$ typically emerges as a result of loop integrations. With this one regulator, all integrals are well defined. Infinities take the form of $1/\epsilon^n$ poles ($n \in \mathbb{N}$). These poles cancel in the sum of a physical renormalised observable,

$$O_{\text{physical}} = \Big(\sum_i \underbrace{I_i}_{\substack{\text{integrals,} \\ 1/\epsilon \text{ divergent}}} + \text{counterterms} \Big) = \text{Finite} + O(\epsilon), \tag{3.16}$$

so we are allowed to take the $\epsilon \to 0$ limit at the end of our computation. **Remark.** The same ϵ also regulates infrared (soft and collinear) divergences of massless particles. Those poles are *not* removed by ultraviolet counterterms: they cancel only for infrared-safe observables once real-emission contributions (and, for initial-state collinear poles, PDF/mass factorisation) are added (Bloch–Nordsieck, Kinoshita–Lee–Nauenberg; see Peskin–Schroeder sec. 6.5, Weinberg I ch. 13). We will see this at work in Chapter 7.

3.3.1 Angular integrations

For $d = 2$, the solid angle is

$$d\Omega_2 = d\phi,$$

for $d = 3$, we have

$$d\Omega_3 = \sin\theta d\theta d\phi,$$

and so on. We are used to defining solid angles for an integer number of dimensions. What is the value of the solid angle Ω_d for an arbitrary real-valued dimension d? Our generalisation of the dimension from integer to real values for the solid angle is mathematically similar to the generalisation of the factorial to the complex-valued Γ function,

$$n! \to \Gamma(z),$$

and relies on finding a defining recurrence relation.

Let us take a $d-$dimensional vector

$$\vec{k}^{(d)} = \left(\vec{k}^{(d-1)}, k_d \right), \tag{3.17}$$

where k_d is the component corresponding to the d^{th} dimension. We shall assume that we know how to express the $\vec{k}^{(d-1)}$ in terms of $(d-1)$ polar coordinates, i.e. a radial coordinate $r_{(d-1)}$ and $(d-2)$ angular coordinates. Then

$$\int_{-\infty}^{+\infty} dk_1 \dots dk_d = \int_{-\infty}^{+\infty} dk_d \int_0^{\infty} dr_{(d-1)} r_{(d-1)}^{d-2} \int d\Omega_{d-1}. \tag{3.18}$$

The $d-$dimensional vector is now expressed in cylindrical coordinates,

$$\vec{k}^{(d)} = \left(r_{(d-1)} \hat{u}^{(d-1)} \left(\theta_1, \theta_2, \ldots \theta_{d-2} \right), k_d \right), \tag{3.19}$$

where $\hat{u}^{(d-1)}$ is a unit vector. Now we change variables to polar coordinates in $d-$dimensions by performing the transformation

$$r_{(d-1)} = l \sin\theta_{d-1}, \qquad k_d = l \cos\theta_{d-1}. \tag{3.20}$$

This gives for the integration measure $\int d^d k$,

$$\int_{-\infty}^{+\infty} dk_d \int_0^{\infty} dr_{(d-1)} r_{(d-1)}^{d-2} \int d\Omega_{d-1} = \int_0^{\infty} dl\, l^{d-1} \int_0^{\pi} d\theta_{d-1} \sin^{d-2}\theta_{d-1} \int d\Omega_{d-1}. \tag{3.21}$$

We have then arrived at the recurrence relation

$$\int d\Omega_d = \int_0^{\pi} d\theta_{d-1} \sin^{d-2}\theta_{d-1} \int d\Omega_{d-1}. \tag{3.22}$$

Recall that for $d = 2$,

$$\int d\Omega_2 = \int_0^{2\pi} d\theta_1. \tag{3.23}$$

Using this as a boundary condition for Eq. 3.22, we obtain the solid angle in arbitrary integer dimensions,

$$\int d\Omega_d = \int_0^{\pi} d\theta_{d-1} \sin^{d-2}\theta_{d-1} \ldots \int_0^{\pi} d\theta_2 \sin\theta_2 \int_0^{2\pi} d\theta_1. \tag{3.24}$$

All integrals can be performed easily using

$$\begin{aligned} \int_0^{\pi} d\theta \sin^x\theta &= \int_{-1}^{1} d(\cos\theta)(1-\cos^2\theta)^{\frac{x-1}{2}} \\ &= \int_0^1 da\, a^{\frac{1}{2}-1}(1-a)^{\frac{x+1}{2}-1} \qquad (a = \cos^2\theta) \\ \rightsquigarrow \int_0^{\pi} d\theta \sin^x\theta &= B\left(\frac{1}{2}, \frac{1+x}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{x+1}{2}\right)}{\Gamma\left(\frac{x+2}{2}\right)}. \end{aligned} \tag{3.25}$$

The solid angle in d dimensions is then

$$\Omega_d = 2\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}. \tag{3.26}$$

On the rhs we have a function which is defined, as we shall discuss immediately, for arbitrary complex values of the dimension parameter d: $\Gamma(d/2)$ has poles at non-positive even integers d, where Ω_d simply vanishes. We have thus derived a generalisation of the solid angle to an arbitrary such value of d.

A quicker derivation of Eq. 3.26 proceeds as follows. Consider the Gaussian integral

$$\int_{-\infty}^{\infty} dx\, e^{-x^2} = \sqrt{\pi}. \tag{3.27}$$

For d such integrals we have

$$\pi^{\frac{d}{2}} = \int d^d x\, e^{-\sum_{i=1}^d x_i^2} = \int d\Omega_d \int_0^\infty dx\, x^{d-1} e^{-x^2}. \tag{3.28}$$

Setting $t = x^2$, we find

$$\pi^{\frac{d}{2}} = \frac{1}{2}\Omega_d \int_0^\infty dt\, t^{\frac{d}{2}-1} e^{-t} = \frac{1}{2}\Omega_d \Gamma\left(\frac{d}{2}\right). \tag{3.29}$$

3.3.2 Properties of the Gamma function

In the above, the Beta function is defined as

$$B(x,y) = \int_0^1 d\xi\, \xi^{x-1}(1-\xi)^{y-1} = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \tag{3.30}$$

The Γ-function is the generalisation of the factorial. It is defined through the integral representation

$$\Gamma(z) \equiv \int_0^\infty dx\, x^{z-1} e^{-x}. \tag{3.31}$$

By using integration by parts, we can easily show that

$$\Gamma(z+1) = z\Gamma(z), \tag{3.32}$$

which is the same recursion relation as for the factorial of integers, $(n+1)! = (n+1)\,n!$. For $z = n$ a positive integer, we have

$$\Gamma(n) = (n-1)!. \tag{3.33}$$

The function $\Gamma(z)$ is analytic except for $z = 0, -1, -2, \ldots$. From the integral representation we conclude that the Γ function is convergent for $\mathrm{Re}\, z > 0$. But the recursion relation of Eq. 3.32 defines it for all complex values of z, with the exception of non-positive integers.

A useful value is

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \tag{3.34}$$

We can also prove the identity

$$\Gamma(x)\Gamma\left(\frac{1}{2}\right) = 2^{x-1}\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right). \tag{3.35}$$

We shall often need to expand a Gamma function around an integer value for its argument. We have

$$\begin{aligned}
\Gamma(1+\epsilon) &= \int_0^\infty dx\, x^\epsilon e^{-x} \\
&= \int_0^\infty dx\, e^{\epsilon \ln x} e^{-x} \\
&= \int_0^\infty dx \left(1 + \epsilon \ln x + \frac{\epsilon^2}{2}\ln^2 x + \ldots\right) e^{-x} \\
\rightsquigarrow \Gamma(1+\epsilon) &= 1 - \epsilon\gamma + \epsilon^2\left(\frac{\gamma^2}{2} + \frac{\pi^2}{12}\right) + O(\epsilon^3).
\end{aligned} \tag{3.36}$$

The Euler gamma constant is defined as

$$\gamma \equiv -\int_0^\infty dx\, e^{-x} \log(x) = 0.5772156649\ldots. \tag{3.37}$$

To higher orders, it is convenient to factor out $e^{-\gamma\epsilon}$:

$$\Gamma(1+\epsilon) = e^{-\gamma\epsilon}\left[1 + \frac{\pi^2}{12}\epsilon^2 - \frac{\zeta_3}{3}\epsilon^3 + \frac{\pi^4}{160}\epsilon^4 - \left(\frac{\zeta_5}{5} + \frac{\zeta_3\pi^2}{36}\right)\epsilon^5 + O(\epsilon^6)\right], \tag{3.38}$$

where

$$\zeta_n \equiv \sum_{k=1}^{\infty} \frac{1}{k^n} \tag{3.39}$$

are the values of the Riemann zeta function at integer arguments,

$$\zeta_2 = \frac{\pi^2}{6}, \qquad \zeta_3 \simeq 1.202056903\ldots, \qquad \zeta_4 = \frac{\pi^4}{90}, \qquad \zeta_5 = 1.036927755\ldots, \qquad \ldots \tag{3.40}$$

3.3.3 Radial integrations

We now continue with the remaining integration of Eq. 3.14 over the magnitude of the Euclidean four-momentum. We have

$$I_d = -\frac{i}{2}\frac{\Omega_d}{(2\pi)^d}\int dk^2 \frac{(k^2)^{\frac{d}{2}-1}}{k^2 + m^2 - i\delta}. \tag{3.41}$$

We perform the change of variables

$$k^2 \to \frac{x}{1-x}\left(m^2 - i\delta\right), \tag{3.42}$$

which yields

$$\begin{aligned} I_d &= -\frac{i}{2}\frac{\Omega_d}{(2\pi)^d}\left(m^2 - i\delta\right)^{\frac{d}{2}-1}\int_0^1 dx\, x^{\frac{d}{2}-1}(1-x)^{-\frac{d}{2}} \\ &= -\frac{i}{2}\frac{\Omega_d}{(2\pi)^d}\left(m^2 - i\delta\right)^{\frac{d}{2}-1} B\left(\frac{d}{2}, 1-\frac{d}{2}\right) \\ &= -\frac{i}{2}\frac{\Omega_d}{(2\pi)^d}\left(m^2 - i\delta\right)^{\frac{d}{2}-1}\frac{\Gamma\left(\frac{d}{2}\right)\Gamma\left(1-\frac{d}{2}\right)}{\Gamma(1)}. \end{aligned} \tag{3.43}$$

Substituting the solid angle of Eq. 3.26, the $\Gamma\left(\frac{d}{2}\right)$ cancels and we are left with a factor $\pi^{\frac{d}{2}}/(2\pi)^d$. It is customary to absorb this factor in the normalisation of the integration measure. We will write all our loop integrals with the measure

$$\int \frac{d^dk}{(2\pi)^d} = \frac{i}{(4\pi)^{\frac{d}{2}}}\int \frac{d^dk}{i\pi^{\frac{d}{2}}}, \tag{3.44}$$

which gives the final result

$$\int \frac{d^dk}{i\pi^{\frac{d}{2}}}\frac{1}{k^2 - m^2 + i\delta} = -\Gamma\left(1-\frac{d}{2}\right)\left(m^2 - i\delta\right)^{\frac{d}{2}-1}. \tag{3.45}$$

Notice that the result has poles at $d = 2, 4, 6, \ldots$, where $\Gamma\left(1-\frac{d}{2}\right)$ is infinite, but it is well-defined for any other value of d.

Exercise: Prove that

$$I[\nu] \equiv \int \frac{d^dk}{i\pi^{\frac{d}{2}}}\frac{1}{[k^2 - m^2 + i\delta]^\nu} = (-1)^\nu \frac{\Gamma\left(\nu - \frac{d}{2}\right)}{\Gamma(\nu)}\left(m^2 - i\delta\right)^{\frac{d}{2}-\nu}. \tag{3.46}$$

3.4 Expansion in ϵ and scaleless integrals

It is instructive to compare the result of Eq. 3.45 with what we expect from counting powers of $|k|$ in $\int d|k|\,|k|^{d-1}/|k|^2$ at large $|k|$. The integral diverges quadratically in $d = 4$ and logarithmically in $d = 2$; correspondingly, $\Gamma(-1)$ and $\Gamma(0)$ are infinite. In $d = 1$ the integral converges, and indeed $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ is finite: the formula reproduces the elementary result $\int_{-\infty}^{+\infty} \frac{dk_0}{2\pi} \frac{1}{k_0^2 - m^2 + i\delta} = -\frac{i}{2m}$. (In $d = 3$ power counting gives a linear divergence, while $\Gamma\left(-\frac{1}{2}\right)$ is finite; dimensional regularisation produces no poles for such odd-dimensional power divergences.) Setting $d = 4 - 2\epsilon$ and using Eq. 3.32,

$$\Gamma(-1+\epsilon) = \frac{\Gamma(\epsilon)}{\epsilon - 1} = -\frac{1}{\epsilon} + \gamma - 1 + O(\epsilon).$$

The quadratically divergent integral of Eq. 3.7 has become a pole $\frac{1}{\epsilon}$ with a definite coefficient and sign. We have not removed the infinity, but we can now quantify it and manipulate it mathematically.

Let us make this more precise. For $\nu = 1$ and $d = 4 - 2\epsilon$, we have

$$\begin{aligned} I[1] &= -(m^2 - i\delta)^{1-\epsilon} \frac{\Gamma(\epsilon - 1)}{\Gamma(1)} = -(m^2 - i\delta)^{1-\epsilon} \frac{\Gamma(\epsilon)}{\epsilon - 1} = -(m^2 - i\delta)^{1-\epsilon} \frac{\Gamma(1+\epsilon)}{(\epsilon - 1)\,\epsilon} \\ &= \frac{1}{\epsilon} (m^2 - i\delta)^{1-\epsilon} \frac{\Gamma(1+\epsilon)}{1 - \epsilon}. \end{aligned} \tag{3.47}$$

So the integral is divergent as $\epsilon \to 0$. We can perform a Taylor expansion around $\epsilon \simeq 0$ (corresponding to 4 dimensions) with the help of the expansion of the Γ-function in Eq. 3.38. We find

$$I[1] \equiv \int \frac{d^d k}{i\pi^{d/2}} \frac{1}{k^2 - m^2 + i\delta} = e^{-\gamma\epsilon}\, m^2 \left[\frac{1}{\epsilon} + 1 - \log\left(m^2 - i\delta\right) + O(\epsilon)\right]. \tag{3.48}$$

The factor $e^{-\gamma\epsilon}$ accompanies every one-loop integral. It is a matter of convention whether we expand it or keep it as an overall factor; we will see in Chapter 6 that it can be absorbed in the definition of the renormalised coupling.

3.4.1 Scaleless integrals

An interesting case is when $m^2 = 0$. In dimensional regularisation, where the dimension parameter is considered non-integer, we have that

$$\lim_{m \to 0} m^d = 0^d = 0. \tag{3.49}$$

We then have that **integrals with no mass scales vanish**. Indeed, consider the integral of Eq. 3.46 in the limit $m \to 0$, which lacks any mass scale. We have

$$\begin{aligned} \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{[k^2 + i\delta]^\nu} &= \lim_{m \to 0} \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{[k^2 - m^2 + i\delta]^\nu} \\ &= \lim_{m \to 0} (-1)^\nu \frac{\Gamma\left(\nu - \frac{d}{2}\right)}{\Gamma(\nu)} \underbrace{\left(m^2\right)^{\frac{d}{2} - \nu}}_{\text{vanishes in the limit}} = 0 \end{aligned} \tag{3.50}$$

(i.e. in dimensional regularisation we take the $d = 4$ limit only at the end). **Remark.** The limit $(m^2)^{\frac{d}{2} - \nu} \to 0$ holds only for $\mathrm{Re}\, d > 2\nu$, where the massless integral is UV divergent;

for $\mathrm{Re}\, d < 2\nu$ the limit diverges instead (IR divergence). There is no d at which $\int d^d k\, (k^2)^{-\nu}$ converges, and setting it to zero is a *definition*, consistent with linearity and scaling of the dimensionally continued integral, and consistent because the UV and IR poles cancel: in $d = 4 - 2\epsilon$, e.g. $\int d^d k/(k^2)^2 \propto \frac{1}{\epsilon_{\mathrm{UV}}} - \frac{1}{\epsilon_{\mathrm{IR}}} = 0$ (cf. Collins, *Renormalization*, ch. 4; Smirnov, *Analytic Tools for Feynman Integrals*, sec. 1.2).

3.4.2 The $i\delta$ prescription and logarithms

Finally, it will be important to emphasise the role of the $i\delta$ part of the propagator. The series expansion of Eq. 3.48 produces a $\log\left(m^2 - i\delta\right)$. The $i\delta$ term ($\delta \to 0^+$) gives a proper definition of the logarithm $\log(m^2 - i\delta)$ for negative values of m^2 (e.g. when the mass corresponds to a scattering invariant mass, where both positive and negative values are allowed). In the absence of the $i\delta$ regulator, we have for $m^2 < 0$

$$\begin{aligned} \log\left(m^2\right) &= \log\left(-|m^2|\right) = \log\left(e^{i\pi n}|m^2|\right) \\ &= \log|m^2| + i\pi n, \qquad n \in \mathbb{Z}, \end{aligned} \tag{3.51}$$

and the convention of which branch to choose is not defined. For example,

$$\log|m^2| + i\pi \qquad \text{or} \qquad \log|m^2| - i\pi \tag{3.52}$$

are equally justified. But in the presence of the $i\delta$ term, we have (Fig. 3.3)

$$\log\left(-|m^2| - i\delta\right) = \log|m^2| + \log e^{-i\pi} = \log|m^2| - i\pi. \tag{3.53}$$

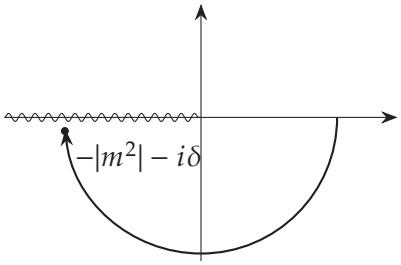


Figure 3.3. The branch cut of the logarithm along the negative real axis; the point $-|m^2| - i\delta$ is reached from the positive real axis through the lower half-plane, giving the phase $e^{-i\pi}$.

3.5 Feynman parameters

Consider a more complicated integral. For example,

$$I_2 = \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{\left[k^2 - m^2 + i\delta\right]\left[(k+p)^2 - m^2 + i\delta\right]}. \tag{3.54}$$

The integrand now has four poles, at

$$k_0 = \pm\left(\sqrt{\vec{k}^2 + m^2} - i\delta\right), \quad -p_0 \pm \left(\sqrt{(\vec{k}+\vec{p})^2 + m^2} - i\delta\right). \tag{3.55}$$

It is now more difficult to perform the integration over k_0, as well as the angular integrations since the denominator depends explicitly on angles,

$$(k+p)^2 - m^2 = k^2 + p^2 - m^2 + 2k_0 p_0 - 2|\vec{k}||\vec{p}|\cos\theta. \tag{3.56}$$

The method of Feynman parameters allows us to integrate out the loop momentum k easily, applying Wick's rotation and integrating over angles in exactly the same manner as for the simplest integral of Eq. 3.46. The price that we have to pay is to introduce new integration variables over "Feynman parameters". It is easy to prove that

$$\frac{1}{AB} = \int_0^1 dx \frac{1}{[xA + (1-x)B]^2}. \tag{3.57}$$

We can use such an identity in order to combine the denominators of loop integrals into one term. For example, the integral of Eq. 3.54 can be written as

$$\begin{aligned} I_2 &= \int_0^1 dx \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{\left[x\left((k+p)^2 - m^2\right) + (1-x)\left(k^2 - m^2\right) + i\delta\right]^2} \\ &= \int_0^1 dx \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{\left[k^2 - m^2 + 2xk\cdot p + xp^2 + i\delta\right]^2} \\ &= \int_0^1 dx \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{\left[(k+px)^2 - m^2 + x(1-x)p^2 + i\delta\right]^2}. \end{aligned} \tag{3.58}$$

We perform a shift in the integration variable k, usually called the "loop momentum", defining

$$K^\mu = k^\mu + xp^\mu. \tag{3.59}$$

We then have

$$I_2 = \int_0^1 dx \int \frac{d^d K}{i\pi^{\frac{d}{2}}} \frac{1}{\left[K^2 - \left(m^2 - x(1-x)p^2 - i\delta\right)\right]^2}. \tag{3.60}$$

The integral over K^μ can be solved using the general result of Eq. 3.46,

$$I_2 = \Gamma\left(2 - \frac{d}{2}\right) \int_0^1 dx \left(m^2 - x(1-x)p^2 - i\delta\right)^{\frac{d}{2}-2}. \tag{3.61}$$

The above integral is only one-dimensional, but we still have to integrate over the Feynman parameter. With $d = 4 - 2\epsilon$, the integral of Eq. 3.61 becomes

$$\begin{aligned} I_2 &= \Gamma(\epsilon) \int_0^1 dx \left(m^2 - x(1-x)p^2 - i\delta\right)^{-\epsilon} \\ &= \frac{1}{\epsilon} - \gamma - \int_0^1 dx \log\left[m^2 - x(1-x)p^2 - i\delta\right] + O(\epsilon). \end{aligned} \tag{3.62}$$

Exercise: Solve the above integral, assuming that $p^2 < 4m^2$. For $p^2 > 4m^2$ the integral develops an imaginary part. Compute it, by performing an analytic continuation of the result for $p^2 < 4m^2$.

Here the $i\delta$ of the Feynman propagators is essential. For $p^2 > 4m^2$ the argument of the logarithm becomes negative in a part of the x integration, and without $i\delta$ we would not know on which side of the branch cut to evaluate it. With $i\delta$ we have unambiguously $\log(-|A| - i\delta) = \log|A| - i\pi$, as in Eq. 3.53. The Feynman prescription gives a unique meaning to all multi-valued functions which appear in the results of loop integrals.

The procedure of combining denominators into a single term using Feynman parameters and completing the square in the loop momentum can be performed in general. For an arbitrary number of denominators in a loop integral we can use the formula

$$\frac{1}{A_1^{\nu_1}\dots A_n^{\nu_n}} = \frac{\Gamma(\nu_1+\dots\nu_n)}{\prod\Gamma(\nu_i)}\int_0^1 dx_1\dots dx_n\,\delta\left(1-\sum_i x_i\right)\frac{x_1^{\nu_1-1}\dots x_n^{\nu_n-1}}{[\sum x_iA_i]^{\sum\nu_i}}. \tag{3.63}$$

Proof: Let us perform the change of variables

$$x_i = \frac{a_i}{1+a_n} \tag{3.64}$$

on the rhs of Eq. 3.63. The delta function becomes

$$\delta\left(1-\sum_{i=1}^n\frac{a_i}{1+a_n}\right) = \delta\left(\frac{1-\sum_{i=1}^{n-1}a_i}{1+a_n}\right) = (1+a_n)\delta\left(1-\sum_{i=1}^{n-1}a_i\right). \tag{3.65}$$

The parameters a_i are constrained to be in the interval $[0,1]$, due to the δ function for $i=1\dots n-1$, while the parameter a_n ranges between $a_n=0$ ($x_n=0$) and $a_n=\infty$ ($x_n=1$). The Jacobian of the transformation is

$$dx_i = \frac{da_i}{1+a_n},\quad i\neq n, \tag{3.66}$$

and

$$dx_n = \frac{da_n}{(1+a_n)^2}. \tag{3.67}$$

All factors conspire in Eq. 3.63, so that the integrand is almost the same as the original, replacing x_i with a_i. The only difference is that the delta function does not contain a_n anymore, and that the range of this variable is from 0 to infinity. Explicitly,

$$\begin{aligned}\text{rhs of Eq. 3.63} &= \frac{\Gamma(\nu_1+\dots\nu_n)}{\prod\Gamma(\nu_i)}\\ &\times\int_0^\infty da_n\int_0^1 da_1\dots da_{n-1}\,\delta\left(1-\sum_{i=1}^{n-1}a_i\right)\frac{a_1^{\nu_1-1}\dots a_n^{\nu_n-1}}{[\sum a_iA_i]^{\sum\nu_i}}.\end{aligned} \tag{3.68}$$

We now change variables once more, to

$$a_n = \frac{x}{1-x}\frac{\sum_{i=1}^{n-1}a_iA_i}{A_n}, \tag{3.69}$$

and perform the integration over x, which ranges from 0 to 1. The result is

$$\begin{aligned}\text{rhs of Eq. 3.63} &= \frac{1}{A_n^{\nu_n}}\times\\ &\times\frac{\Gamma(\nu_1+\dots\nu_{n-1})}{\Gamma(\nu_1)\dots\Gamma(\nu_{n-1})}\int_0^1 dx_1\dots dx_{n-1}\,\delta\left(1-\sum_{i=1}^{n-1}x_i\right)\frac{x_1^{\nu_1-1}\dots x_{n-1}^{\nu_{n-1}-1}}{\left[\sum_{i=1}^{n-1}x_iA_i\right]^{\sum_{i=1}^{n-1}\nu_i}}.\end{aligned} \tag{3.70}$$

We have now factorised the term $1/A_n^{\nu_n}$ out of an integral, which is now cast in the same form as the original but with $n-1$ Feynman parameters. Obviously, we can repeat the

same procedure as many times as needed in order to factorise all $\frac{1}{A_i^{\nu_i}}$ terms, as on the lhs of Eq. 3.63.

For $A_1 = \cdots = A_n = 1$ the formula gives a useful integral over the Feynman parameters,

$$\int dx_1 \cdots dx_n \; x_1^{\nu_1 - 1} \cdots x_n^{\nu_n - 1} \, \delta\Big(1 - \sum_i x_i\Big) = \frac{\Gamma(\nu_1) \cdots \Gamma(\nu_n)}{\Gamma(\nu_1 + \cdots + \nu_n)}, \tag{3.71}$$

which for $n = 2$ reproduces the Beta-function integral.

Feynman-parameter integrals are generalised hypergeometric functions, which are only partially understood in the mathematical science. Said differently, the mathematics of hypergeometric functions is far from sufficient in order to tackle the Feynman integrals that appear in the perturbative expansion of Green's functions. This limits computations to simple cases. Nevertheless, even with the mathematical understanding of loop integrations being in its infancy, we can still derive rather precise theoretical predictions for S-matrix elements of interesting scattering processes. The situation has, however, improved enormously over the last decades. All one-loop integrals are known in closed form (in $d = 4$ in terms of logarithms and dilogarithms, 't Hooft and Veltman 1979), and since then multi-loop integrals have been computed systematically with integration-by-parts reduction (Chetyrkin and Tkachov 1981; Laporta 2000) and differential equations in canonical form (Henn 2013); most results are multiple polylogarithms, with elliptic and more general functions appearing at higher loops (see Weinzierl, *Feynman Integrals*, Springer 2022). In the next chapter we will apply Feynman parameters to a general one-loop integral, and we will learn techniques which relate different loop integrals to each other, so that only a few of them need to be computed explicitly.

4

Techniques for loop integrals

4.1 General one-loop integrals with Feynman parameters

Consider a general one-loop integral (Fig. 4.1),

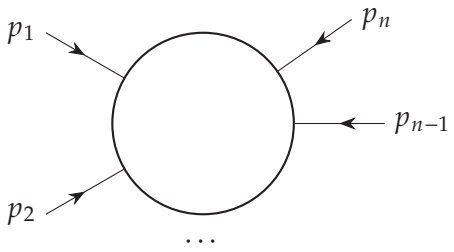


Figure 4.1. A general one-loop graph with n external legs, all momenta incoming.

where all momenta are incoming and

$$\sum_i p_i = p_1 + \cdots + p_n = 0 \qquad \text{(momentum conservation)}. \tag{4.1}$$

The scalar integral corresponding to such a graph (ignoring numerators, and assigning a factor of 1 for vertices, a factor $\frac{1}{k^2-m^2}$ for each internal line and a factor of 1 for each external leg) is

$$I_n = \int \frac{d^d k}{i\pi^{d/2}} \frac{1}{A_1^{\nu_1} \cdots A_n^{\nu_n}}, \tag{4.2}$$

with

$$\begin{aligned} A_1 &= k^2 - m_1^2, \\ A_2 &= (k+p_1)^2 - m_2^2, \\ &\vdots \\ A_n &= (k+p_1+\cdots+p_{n-1})^2 - m_n^2. \end{aligned} \tag{4.3}$$

The integral I_2 of Eq. 3.54 is the special case $n = 2$, $\nu_1 = \nu_2 = 1$, $m_1 = m_2 = m$. For short, let us define

$$\begin{aligned} q_1 &= 0 = \textstyle\sum_i p_i, \\ q_2 &= p_1, \\ q_3 &= p_1 + p_2, \\ &\vdots \\ q_n &= p_1 + p_2 + \cdots + p_{n-1}; \end{aligned} \tag{4.4}$$

then, with the Feynman-parameter formula of Eq. 3.63, we can write

$$\begin{aligned} I_n &= \int \frac{d^dk}{i\pi^{d/2}} \prod_{i=1}^{n} \frac{1}{\left[(k+q_i)^2 - m_i^2\right]^{\nu_i}} \\ &= \frac{\Gamma(\nu_1 + \cdots + \nu_n)}{\Gamma(\nu_1)\cdots\Gamma(\nu_n)} \int dx_1 \cdots dx_n \; x_1^{\nu_1 - 1} \cdots x_n^{\nu_n - 1} \, \delta\Big(1 - \sum_i x_i\Big) \\ &\quad \times \int \frac{d^dk}{i\pi^{d/2}} \frac{1}{\left[\sum_i x_i (k+q_i)^2 - \sum_i x_i m_i^2\right]^{\sum_i \nu_i}}. \end{aligned} \tag{4.5}$$

The sum in the denominator is

$$\begin{aligned} \sum_i x_i (k+q_i)^2 &= \underbrace{\sum_i x_i}_{=1} \; k^2 + 2k \cdot \Big(\sum_i x_i q_i\Big) + \sum_i x_i q_i^2 \\ &= \Big(k + \sum_i x_i q_i\Big)^2 + \underbrace{\sum_i x_i q_i^2 - \Big(\sum_i x_i q_i\Big)^2}_{Q}. \end{aligned} \tag{4.6}$$

The Q-term is

$$\begin{aligned} Q &\equiv \Big(\sum_i x_i q_i^2\Big) - \sum_{i,j} x_i x_j \, q_i \cdot q_j \\ &= \sum_i x_i q_i^2 + \frac{1}{2} \sum_{i,j} x_i x_j \left[(q_i - q_j)^2 - q_i^2 - q_j^2\right] \\ &= \sum_i x_i q_i^2 - \frac{1}{2}\Big(\sum_i x_i q_i^2\Big) \underbrace{\Big(\sum_j x_j\Big)}_{1} - \frac{1}{2}\Big(\sum_j x_j q_j^2\Big) \underbrace{\Big(\sum_i x_i\Big)}_{1} + \frac{1}{2} \sum_{i,j} x_i x_j (q_i - q_j)^2 \end{aligned}$$

$$\implies \quad Q = \sum_{i<j} x_i x_j \, (q_i - q_j)^2. \tag{4.7}$$

Performing a shift,

$$k^\mu \to k^\mu - \sum_i x_i q_i^\mu, \tag{4.8}$$

brings the loop integral into the same form as the tadpole, with $m^2 = \sum_i x_i m_i^2 - Q$. The k-integration is then given by Eq. 3.46. We therefore find, for the general one-loop scalar integral with n propagators,

$$I_n = (-1)^N \frac{\Gamma\left(N - \frac{d}{2}\right)}{\Gamma(\nu_1)\cdots\Gamma(\nu_n)} \int_0^1 \frac{dx_1 \cdots dx_n \; x_1^{\nu_1 - 1} \cdots x_n^{\nu_n - 1} \, \delta(1 - x_1 - x_2 - \ldots - x_n)}{\left[x_1 m_1^2 + \ldots + x_n m_n^2 - \sum_{i<j} x_i x_j (q_i - q_j)^2 - i\delta\right]^{N - \frac{d}{2}}}, \tag{4.9}$$

where

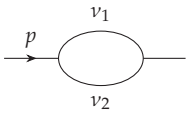
$$N = \nu_1 + \nu_2 + \ldots + \nu_n. \tag{4.10}$$

Examples:

- For $n = 1$, this should reproduce the tadpole integral of eq. (3.46). Indeed:

$$I_1 = (-1)^{\nu_1} \frac{\Gamma\left(\nu_1 - \frac{d}{2}\right)}{\Gamma(\nu_1)} \int_0^1 \frac{dx_1\, \delta(1-x_1)}{\left[x_1 m_1^2 - i\delta\right]^{\nu_1 - \frac{d}{2}}} = (-1)^{\nu_1} \frac{\Gamma\left(\nu_1 - \frac{d}{2}\right)}{\Gamma(\nu_1)} \left(m_1^2 - i\delta\right)^{\frac{d}{2} - \nu_1}. \quad (4.11)$$

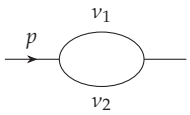
- Let us consider now a "massless" scalar two-propagator integral:



$$= (-1)^{\nu_1+\nu_2} \frac{\Gamma\left(\nu_1 + \nu_2 - \frac{d}{2}\right)}{\Gamma(\nu_1)\Gamma(\nu_2)} \times \int_0^1 \frac{dx_1\, dx_2\, x_1^{\nu_1-1} x_2^{\nu_2-1}\, \delta(1-x_1-x_2)}{\left[-x_1 x_2\, p^2 - i\delta\right]^{\nu_1+\nu_2-\frac{d}{2}}}$$

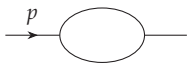
$$= (-1)^{\nu_1+\nu_2} \frac{\Gamma\left(\nu_1 + \nu_2 - \frac{d}{2}\right)}{\Gamma(\nu_1)\Gamma(\nu_2)} \left(-p^2 - i\delta\right)^{\frac{d}{2} - \nu_1 - \nu_2} \times \int_0^1 dx_1\, dx_2\, \delta(1-x_1-x_2)\, x_1^{\frac{d}{2}-\nu_2-1}\, x_2^{\frac{d}{2}-\nu_1-1}. \quad (4.12)$$

It is now easy to perform the Feynman-parameter integration, which yields a B-function. We find



$$= (-1)^{\nu_1+\nu_2} \frac{\Gamma\left(\nu_1 + \nu_2 - \frac{d}{2}\right)}{\Gamma(\nu_1)\Gamma(\nu_2)} \frac{\Gamma\left(\frac{d}{2} - \nu_2\right)\Gamma\left(\frac{d}{2} - \nu_1\right)}{\Gamma(d - \nu_1 - \nu_2)} \cdot \left(-p^2 - i\delta\right)^{\frac{d}{2} - \nu_1 - \nu_2}. \quad (4.13)$$

  - For $\nu_1 = \nu_2 = 1$ and $d = 4 - 2\epsilon$ we have:



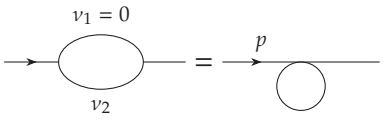
$$= \frac{\Gamma(\epsilon)\,\Gamma^2(1-\epsilon)}{\Gamma(2-2\epsilon)} \left(-p^2 - i\delta\right)^{-\epsilon} = \frac{1}{\epsilon(1-2\epsilon)} \frac{\Gamma(1+\epsilon)\,\Gamma^2(1-\epsilon)}{\Gamma(1-2\epsilon)} \left(-p^2 - i\delta\right)^{-\epsilon} \equiv \frac{c_\Gamma}{\epsilon(1-2\epsilon)} \left(-p^2 - i\delta\right)^{-\epsilon}, \quad (4.14)$$

    where

$$c_\Gamma = \frac{\Gamma(1+\epsilon)\Gamma(1-\epsilon)^2}{\Gamma(1-2\epsilon)} = e^{-\gamma\epsilon} \left[1 - \frac{\pi^2}{12}\epsilon^2 + O(\epsilon^3)\right]. \quad (4.15)$$

    This is also a divergent integral.

  - For $\nu_1 = 0$, we have



(scaleless integral)

$$= (-1)^{\nu_2} \frac{\Gamma\left(\nu_2 - \frac{d}{2}\right)}{\Gamma(0)\,\Gamma(\nu_2)} \frac{\Gamma\left(\frac{d}{2} - \nu_2\right)\Gamma\left(\frac{d}{2}\right)}{\Gamma(d - \nu_2)} \left(-p^2 - i\delta\right)^{\frac{d}{2} - \nu_2}. \quad (4.16)$$

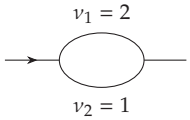
    We can write

$$\frac{1}{\Gamma(0)} = \frac{0}{\Gamma(1+0)} = \frac{0}{\Gamma(1)} = 0. \quad (4.17)$$

    Thus, as expected, the integral vanishes (cf. eq. (3.50)).

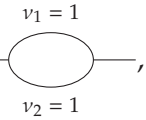
4.2 Integration by parts

Let us also consider the bubble with one squared propagator:



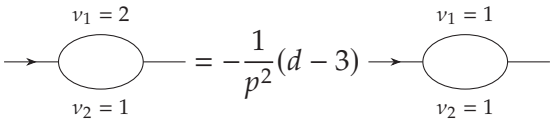
$$= \frac{(-1)^3\,\Gamma\left(3-\frac{d}{2}\right)}{\Gamma(2)\,\Gamma(1)}\,\frac{\Gamma\left(\frac{d}{2}-2\right)\Gamma\left(\frac{d}{2}-1\right)}{\Gamma(d-3)}\left(-p^2\right)^{\frac{d}{2}-3}$$

$$= \frac{1}{p^2}\left(2-\frac{d}{2}\right)\frac{\Gamma\left(2-\frac{d}{2}\right)}{\Gamma(1)^2}\,\frac{\Gamma\left(\frac{d}{2}-1\right)^2}{\Gamma(d-2)}\,\frac{(d-3)}{\left(\frac{d}{2}-2\right)}\left(-p^2\right)^{\frac{d}{2}-2}$$

$$= -\frac{1}{p^2}(d-3)$$ 

, (4.18)

i.e.



(4.19)

The explicit results reveal an algebraic identity between the two integrals. If we can derive such an identity directly, we do not need to compute both integrals separately; it suffices to compute just one. Such an identity can be obtained with integration by parts. Indeed:

$$\begin{aligned} 0 &= \int_{-\infty}^{+\infty} d^dk\,\frac{\partial}{\partial k^\mu}\,\frac{k^\mu}{(k^2)\,(k+p)^2} \\ &= \int_{-\infty}^{+\infty} \frac{d^dk}{(k^2)\,(k+p)^2}\left[k^\mu\left(-\frac{2k_\mu}{k^2}-\frac{2(k+p)_\mu}{(k+p)^2}\right)+d\right] \\ \Rightarrow\quad 0 &= \int \frac{d^dk}{(k^2)\,(k+p)^2}\left\{d-2-\frac{2}{(k+p)^2}\left(k^2+k\cdot p\right)\right\}. \end{aligned} \tag{4.20}$$

We can write

$$k\cdot p = \frac{(k+p)^2-k^2-p^2}{2} \quad\Rightarrow\quad k^2+k\cdot p = \frac{(k+p)^2+k^2-p^2}{2}. \tag{4.21}$$

Thus,

$$0 = \int \frac{d^dk}{k^2\,(k+p)^2}\left\{d-3-\frac{k^2}{(k+p)^2}+\frac{p^2}{(k+p)^2}\right\}, \tag{4.22}$$

where the term $k^2/(k+p)^2$ gives a scaleless integral, which vanishes. Thus

$$\int \frac{d^dk}{k^2\left[(k+p)^2\right]^2} = -\frac{1}{p^2}(d-3)\int \frac{d^dk}{(k^2)\,(k+p)^2}, \tag{4.23}$$

which is the algebraic relation which we observed in the solutions of the two integrals.

Exercise: Derive the integration-by-parts identities

$$0 = \int d^dk\,\frac{\partial}{\partial k^\mu}\,\frac{k^\mu}{(k^2)^{\nu_1}\left[(k+p)^2\right]^{\nu_2}} \qquad \text{and} \qquad 0 = \int d^dk\,\frac{\partial}{\partial k^\mu}\,\frac{k^\mu+p^\mu}{(k^2)^{\nu_1}\left[(k+p)^2\right]^{\nu_2}}. \tag{4.24}$$

You should find something like

$$\begin{aligned} p^2 \nu_1 \, \mathbf{1}^+ + (d - 2\nu_2 - \nu_1) - \nu_1 \, \mathbf{1}^+ \mathbf{2}^- = 0, \\ p^2 \nu_2 \, \mathbf{2}^+ + (d - 2\nu_1 - \nu_2) - \nu_2 \, \mathbf{2}^+ \mathbf{1}^- = 0, \end{aligned} \tag{4.25}$$

where $\mathbf{2}^\pm$ raises/lowers ν_2 to $\nu_2 \pm 1$ and $\mathbf{1}^\pm$ raises/lowers ν_1 to $\nu_1 \pm 1$ (acting on the integral $I[\nu_1, \nu_2]$). Write a small program which computes recursively all $I[\nu_1, \nu_2]$ integrals algebraically in terms of the $I[1,1]$ "master integral".

4.3 Dimensionally shifted integrals

Let us now return to the generic Feynman parameterisation of one-loop integrals, eq. (4.9),

$$I[\nu_1, \ldots, \nu_n; d] \equiv (-1)^{\nu_1 + \ldots + \nu_n} \frac{\Gamma\left(\nu_1 + \ldots + \nu_n - \frac{d}{2}\right)}{\Gamma(\nu_1) \cdots \Gamma(\nu_n)} \int d^n \vec{x} \; \delta\Big(1 - \sum_i x_i\Big) \frac{x_1^{\nu_1 - 1} \cdots x_n^{\nu_n - 1}}{[\cdots]^{\nu_1 + \ldots + \nu_n - \frac{d}{2}}}. \tag{4.26}$$

Let us now change the power of x_1 from ν_1 to $\nu_1 + 1$. Then,

$$\begin{aligned} & (-1)^{\nu_1 + \ldots + \nu_n} \frac{\Gamma\left(\nu_1 + \ldots + \nu_n - \frac{d}{2}\right)}{\Gamma(\nu_1) \cdots \Gamma(\nu_n)} \int d^n \vec{x} \; \delta\Big(1 - \sum_i x_i\Big) \frac{\left(x_1 \cdot x_1^{\nu_1 - 1}\right) \cdots x_n^{\nu_n - 1}}{\left[\sum_i x_i m_i^2 - Q\right]^{\nu_1 + \ldots + \nu_n - \frac{d}{2}}} \\ & = -\nu_1 \, (-1)^{(\nu_1 + 1) + \nu_2 + \ldots + \nu_n} \frac{\Gamma\left((\nu_1 + 1) + \nu_2 + \ldots + \nu_n - \frac{d+2}{2}\right)}{\Gamma(\nu_1 + 1) \cdots \Gamma(\nu_n)} \\ & \quad \cdot \int d^n \vec{x} \; \delta\Big(1 - \sum_i x_i\Big) \frac{x_1^{(\nu_1 + 1) - 1} \cdots x_n^{\nu_n - 1}}{[\cdots]^{(\nu_1 + 1) + \nu_2 + \ldots + \nu_n - \frac{d+2}{2}}} \\ & = -\nu_1 \, \mathbf{1}^+ \mathbf{d}^+, \end{aligned} \tag{4.27}$$

where $\mathbf{1}^+$ increases the power ν_1 by 1, $\nu_1 \to \nu_1 + 1$, and $\mathbf{d}^+$ increases the dimension by 2: $d \to d + 2$. Therefore, inserting a Feynman parameter amounts to

$$x_i \;\longrightarrow\; -\nu_i \, \mathbf{i}^+ \mathbf{d}^+. \tag{4.28}$$

But $\sum_i x_i = 1$ (δ-function). Thus:

$$-\Big(\sum_i \nu_i \, \mathbf{i}^+\Big) \mathbf{d}^+ = 1 \quad \text{(the original integral)}. \tag{4.29}$$

Example: Define

$$I^d[\nu_1, \nu_2] \equiv \int \frac{d^d k}{i\pi^{d/2}} \frac{1}{(k^2)^{\nu_1} \left[(k+p)^2\right]^{\nu_2}}. \tag{4.30}$$

Then

$$I^d[\nu_1, \nu_2] = -\nu_1 \, I^{d+2}[\nu_1 + 1, \nu_2] - \nu_2 \, I^{d+2}[\nu_1, \nu_2 + 1]. \tag{4.31}$$

For $\nu_1 = \nu_2 = 1$:

$$\begin{aligned} I^d[1,1] &= -I^{d+2}[2,1] - I^{d+2}[1,2] \\ &= -2 \, I^{d+2}[2,1] \\ &\overset{\text{IBP}}{=} -2 \left(-\frac{1}{p^2}\right) (d + 2 - 3) \, I^{d+2}[1,1] \\ \Rightarrow \quad I^d[1,1] &= \frac{2}{p^2}(d-1) \, I^{d+2}[1,1] \quad \Rightarrow \quad \frac{I^{d+2}[1,1]}{I^d[1,1]} = \frac{p^2}{2(d-1)}. \end{aligned} \tag{4.32}$$

This is indeed a correct identity. From eq. (4.13),

$$I^d[1,1] = \Gamma\left(2-\tfrac{d}{2}\right)\frac{\Gamma\left(\frac{d}{2}-1\right)^2}{\Gamma(d-2)}\left(-p^2\right)^{\frac{d}{2}-2}, \tag{4.33}$$

$$I^{d+2}[1,1] = \Gamma\left(1-\tfrac{d}{2}\right)\frac{\Gamma\left(\frac{d}{2}\right)^2}{\Gamma(d)}\left(-p^2\right)^{\frac{d}{2}-1}, \tag{4.34}$$

so that

$$\frac{I^{d+2}[1,1]}{I^d[1,1]} = -p^2\,\frac{1}{\left(1-\frac{d}{2}\right)}\,\frac{\left(\frac{d}{2}-1\right)^2}{(d-1)(d-2)} = \frac{p^2}{2}\,\frac{(d-2)}{(d-1)(d-2)} = \frac{p^2}{2}\,\frac{1}{(d-1)}. \qquad \checkmark \tag{4.35}$$

4.4 Tensor integrals and scalar integrals

In computing QED scattering amplitudes, we need to calculate loop integrals where tensors of the loop momentum appear in the numerator. These arise from the Feynman rule for the fermion propagator, which is written as

$$\frac{i}{\not{k}-m} = \frac{i\left(\gamma^\mu k_\mu + m\right)}{k^2-m^2}. \tag{4.36}$$

Particles propagating in loops, called virtual particles, generate loop momenta in the numerators of loop integrands. We will denote a one-loop integral with a numerator $N(k)$ by

$$I^d[\nu_1,\nu_2;N(k)] \equiv \int \frac{d^dk}{i\pi^{d/2}}\,\frac{N(k)}{(k^2)^{\nu_1}\left[(k+p)^2\right]^{\nu_2}}, \tag{4.37}$$

where the $+i\delta$ term in the denominators is implicit and we take for simplicity $m=0$. We drop the superscript d when no confusion can arise, and $I[1,1] \equiv I[1,1;1]$ is the scalar bubble of Eq. 4.14. All relations below are linear, so they hold equally for the measure $\int d^dk$ or $\int \frac{d^dk}{(2\pi)^d}$.

4.4.1 The Passarino–Veltman reduction

At one loop, we can always express tensor integrals in terms of scalar integrals. The technique is known as the Passarino–Veltman technique and we can illustrate it with a couple of examples. Consider the following tensor one-loop integral,

$$I\left[1,1;k^\mu\right] \equiv \int \frac{d^dk}{i\pi^{d/2}}\,\frac{k^\mu}{k^2(k+p)^2}. \tag{4.38}$$

The integrand is a rank-one tensor in the loop momentum. Due to Lorentz covariance, the result of the integration will also be a rank-one tensor. We can then write,

$$I\left[1,1;k^\mu\right] = Ap^\mu, \tag{4.39}$$

where on the rhs we have written the most general rank-one tensor that we can construct from the momenta in the integral other than the loop momentum. The coefficient A is as yet undetermined. We can express it in terms of integrals which have scalar numerators. We multiply Eq. 4.39 with p_μ and rearrange,

$$A = \frac{1}{p^2} I\left[1,1;k\cdot p\right]. \tag{4.40}$$

Notice that the scalar product in the numerator of the integral above can be expressed in terms of the denominators of the integral,

$$k \cdot p = \frac{1}{2}(k+p)^2 - \frac{k^2}{2} - \frac{p^2}{2}. \tag{4.41}$$

We then have that,

$$A = \frac{1}{2}\left[-\int \frac{d^d k}{i\pi^{d/2}} \frac{1}{k^2(k+p)^2} + \frac{1}{p^2}\int \frac{d^d k}{i\pi^{d/2}} \frac{1}{k^2} - \frac{1}{p^2}\int \frac{d^d k}{i\pi^{d/2}} \frac{1}{(k+p)^2} \right]. \tag{4.42}$$

Notice that we only find integrals with a constant numerator, which is independent of the loop momentum. The last two integrals are identical, as we can see by performing the shift $k \to k+p$ in one of them. In this particular example, they cancel against each other (they are also scaleless, and each of them vanishes), and we have

$$I[1,1;k^\mu] = \int \frac{d^d k}{i\pi^{d/2}} \frac{k^\mu}{k^2(k+p)^2} = -\frac{p^\mu}{2}\int \frac{d^d k}{i\pi^{d/2}} \frac{1}{k^2(k+p)^2} = -\frac{p^\mu}{2} I[1,1]. \tag{4.43}$$

Consider now a more complicated example,

$$I\left[1,1;k^\mu k^\nu\right] \equiv \int \frac{d^d k}{i\pi^{d/2}} \frac{k^\mu k^\nu}{k^2(k+p)^2}. \tag{4.44}$$

The most general rank-two tensor that we can write using the external momenta of the integral is,

$$I\left[1,1;k^\mu k^\nu\right] = A_1 g^{\mu\nu} + A_2 \frac{p^\mu p^\nu}{p^2}. \tag{4.45}$$

Contracting with $g_{\mu\nu}$ and $\frac{p_\mu p_\nu}{p^2}$, we obtain the system of equations,

$$\begin{pmatrix} g^{\mu\nu}g_{\mu\nu} & \frac{p^\mu p^\nu}{p^2} g_{\mu\nu} \\ g^{\mu\nu}\frac{p_\mu p_\nu}{p^2} & \frac{p^\mu p^\nu}{p^2}\frac{p_\mu p_\nu}{p^2} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} I\left[1,1;k^2\right] \\ \frac{I\left[1,1;(k\cdot p)^2\right]}{p^2} \end{pmatrix}. \tag{4.46}$$

In arbitrary dimensions,

$$g_{\mu\nu}g^{\mu\nu} = d = 4 - 2\epsilon,$$

and the above system of equations becomes,

$$\begin{pmatrix} 4-2\epsilon & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} I\left[1,1;k^2\right] \\ \frac{I\left[1,1;(k\cdot p)^2\right]}{p^2} \end{pmatrix}, \tag{4.47}$$

with the solution

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \frac{1}{3-2\epsilon} \begin{pmatrix} 1 & -1 \\ -1 & 4-2\epsilon \end{pmatrix} \begin{pmatrix} I\left[1,1;k^2\right] \\ \frac{I\left[1,1;(k\cdot p)^2\right]}{p^2} \end{pmatrix}. \tag{4.48}$$

The two integrals on the rhs can be expressed in terms of two scalar "master integrals",

$$I_a \equiv \int \frac{d^d k}{i\pi^{d/2}} \frac{1}{k^2}, \tag{4.49}$$

and

$$I_b \equiv \int \frac{d^d k}{i\pi^{d/2}} \frac{1}{k^2(k+p)^2} = I[1,1]. \tag{4.50}$$

Specifically, we have

$$\begin{aligned} I\left[1,1;k^2\right] &= \int \frac{d^d k}{i\pi^{d/2}} \frac{k^2}{k^2(k+p)^2} = \int \frac{d^d k}{i\pi^{d/2}} \frac{1}{(k+p)^2} \\ &= \int \frac{d^d (k+p)}{i\pi^{d/2}} \frac{1}{(k+p)^2} = \int \frac{d^d k}{i\pi^{d/2}} \frac{1}{k^2} \\ &= I_a. \end{aligned} \tag{4.51}$$

I_a is a scaleless integral and, as we have seen in Eq. 3.50, it is zero in dimensional regularisation. We leave it as an exercise to prove that

$$I\left[1,1;(k\cdot p)^2\right] = \frac{p^2}{2} I_a + \frac{(p^2)^2}{4} I_b. \tag{4.52}$$

With $I_a = 0$ we find $A_1 = -\frac{p^2}{4(d-1)} I_b$ and $A_2 = \frac{d\,p^2}{4(d-1)} I_b$, i.e.

$$I[1,1;k^\mu k^\nu] = -\frac{I[1,1]}{4(d-1)} p^2 \left[g^{\mu\nu} - d\,\frac{p^\mu p^\nu}{p^2}\right] = \frac{I[1,1]}{4(d-1)} \left[d\,p^\mu p^\nu - p^2 g^{\mu\nu}\right]. \tag{4.53}$$

A quicker route to the same result uses the vanishing of scaleless integrals from the start. Write

$$I[1,1;k^\mu k^\nu] = A\,p^\mu p^\nu + B\,p^2 g^{\mu\nu}. \tag{4.54}$$

Contracting with $g_{\mu\nu}$:

$$0 = \underbrace{I[1,1;k^2]}_{\text{scaleless}} = (A + Bd)\,p^2 \quad \Rightarrow \quad A = -Bd, \tag{4.55}$$

and thus

$$I[1,1;k^\mu k^\nu] = p^2 \left[g^{\mu\nu} - d\,\frac{p^\mu p^\nu}{p^2}\right] C. \tag{4.56}$$

Contracting with p_μ, and using that $(k\cdot p)\,k^\nu \to -\frac{p^2}{2} k^\nu$ under the integral (the $(k+p)^2$ and k^2 pieces give scaleless integrals) together with Eq. 4.43:

$$\begin{aligned} I[1,1;(k\cdot p)\,k^\nu] &= p^2 p^\nu (1-d)\,C \\ \Rightarrow \quad -\frac{p^2}{2} I[1,1;k^\nu] &= p^2 p^\nu (1-d)\,C \\ \Rightarrow \quad +\frac{p^\nu}{4} I[1,1] &= p^\nu (1-d)\,C \\ \Rightarrow \quad C &= -\frac{I[1,1]}{4(d-1)}, \end{aligned} \tag{4.57}$$

in agreement with Eq. 4.53.

We can express all one-loop tensor integrals in terms of four master integrals,

$$M_A \equiv \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{k^2 - m_1^2 + i\delta}, \tag{4.58}$$

$$M_B \equiv \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{\left(k^2 - m_1^2 + i\delta\right)\left[(k+p)^2 - m_2^2 + i\delta\right]}, \tag{4.59}$$

$$M_C \equiv \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{\left(k^2 - m_1^2 + i\delta\right)\left[(k+p_1)^2 - m_2^2 + i\delta\right]\left[(k+p_2)^2 - m_3^2 + i\delta\right]}, \tag{4.60}$$

and

$$M_D \equiv \int \frac{d^d k}{i\pi^{\frac{d}{2}}} \times \\ \times \frac{1}{\left(k^2 - m_1^2 + i\delta\right)\left[(k+p_1)^2 - m_2^2 + i\delta\right]\left[(k+p_2)^2 - m_3^2 + i\delta\right]\left[(k+p_3)^2 - m_4^2 + i\delta\right]}. \tag{4.61}$$

Exercise: Express the following tensor integrals in terms of master integrals:

$$\int d^d k \frac{k^\mu}{k^2 - m^2},$$
$$\int d^d k \frac{k^\mu k^\nu}{k^2 - m^2},$$
$$\int d^d k \frac{k^\mu}{(k^2 - m_1^2)\left[(k+p)^2 - m_2^2\right]},$$
$$\int d^d k \frac{k^\mu k^\nu}{(k^2 - m_1^2)\left[(k+p)^2 - m_2^2\right]}.$$

The above master integrals have been known analytically for many years. Their computation is involved for the general case of arbitrary masses m_i and momenta p_i. We shall need the following two special cases, which we have already computed with Feynman parameters,

$$\int \frac{d^d k}{i\pi^{\frac{d}{2}}} \frac{1}{k^2 - m^2} = -\Gamma(-1+\epsilon)\left(m^2\right)^{1-\epsilon} \tag{4.62}$$

(Eq. 3.45), and the massless bubble of Eq. 4.14.

4.4.2 Tensor integrals with Feynman parameters

Feynman parameters offer an alternative route, which does not require us to solve a system of equations. Let us consider how the tensor integral

$$\int \frac{d^d k}{i\pi^{d/2}} \frac{k^\mu}{(k^2)^{\nu_1}\left[(k+p)^2\right]^{\nu_2}} \tag{4.63}$$

can be written in terms of scalar integrals. Introducing Feynman parameters and shifting the loop momentum, $k^\mu \to k^\mu - \sum_i x_i q_i^\mu$ (here $q_1 = 0$, $q_2 = p$), the term linear in the new k integrates to zero and we obtain

$$\int \frac{d^d k}{i\pi^{d/2}} \frac{k^\mu}{(k^2)^{\nu_1}\left[(k+p)^2\right]^{\nu_2}} = \cdots = \int dx_1\, dx_2 \frac{x_1^{\nu_1 - 1} x_2^{\nu_2 - 1}\left(-x_2\, p^\mu\right)}{[\cdots]^{\cdots}} \cdots = (+\nu_2\, \mathbf{2}^+ \mathbf{d}^+)\, p^\mu, \tag{4.64}$$

using Eq. 4.28. Hence

$$I^d[\nu_1, \nu_2; k^\mu] = \nu_2\, I^{d+2}[\nu_1, \nu_2 + 1]\, p^\mu. \tag{4.65}$$

For $\nu_1 = \nu_2 = 1$, with the dimension-shift relation of the previous section,

$$I^d[1,1;k^\mu] \quad = \quad p^\mu\, I^{d+2}[1,2] = p^\mu \left(-\frac{1}{2}\, I^d[1,1]\right), \tag{4.66}$$

which reproduces Eq. 4.43.

Check: Let us contract both sides with p_μ. Then

$$I^d[1,1;(k\cdot p)] = -\frac{p^2}{2}\, I^d[1,1]. \tag{4.67}$$

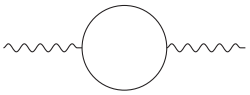
On the left-hand side, $k\cdot p = \frac{1}{2}\left\{(k+p)^2 - k^2 - p^2\right\}$; the first two terms cancel a propagator and give scaleless integrals, which vanish. Thus

$$I^d\left[1,1;-\frac{p^2}{2}\right] = -\frac{p^2}{2}\, I^d[1,1], \tag{4.68}$$

as it should.

Exercise: *Calculate in the same way, with Feynman parameters and dimension shifts, the second-rank tensor $I[1,1;k^\mu k^\nu]$, and check your result against Eq. 4.53.*

Exercise: *Calculate the one-loop photon propagator in terms of $I[1,1]$ (ignore the mass of the electron):*



$$= I[1,1]\cdot\, ? \tag{4.69}$$

Compare your result with Section 5.4 of the next chapter.

5

Quantum electrodynamics at one loop

We have now collected all the tools that we need in order to compute quantum corrections in QED: the Feynman rules of Chapter 2, dimensional regularisation, and the reduction of one-loop integrals to a few master integrals. In this chapter we will first set up dimensional regularisation for the QED Lagrangian. We will then compute the one-loop corrections to the electron and the photon propagators, and we will sum them to all orders in perturbation theory. Along the way we will find that the photon remains massless and that its propagator receives corrections only in its transverse part. The reason is a symmetry, which is expressed in the Ward identity.

5.1 Dimensional regularisation for QED

QED is plagued by divergent loop integrals. We shall use dimensional regularisation as our regularisation method. In this section, we discuss some new features that arise in dimensional regularisation in gauge theories. It is a hard problem in general to prove that a regularisation method is appropriate for a given quantum field theory. What is a consistent way to take $4 \to d$?

First, we consider the QED action integral in arbitrary d dimensions,

$$S = \int d^d x \mathcal{L}_{QED}. \tag{5.1}$$

The action has no mass dimension,

$$[S] = \left[d^d x \mathcal{L}_{QED}\right] = [mass]^0 . \tag{5.2}$$

Since the space-time coordinates have inverse mass dimension, we find that all terms in the Lagrangian must have mass dimensionality d,

$$\left[\mathcal{L}_{QED}\right] = [mass]^d . \tag{5.3}$$

From the kinetic terms of the photon and fermion fields (or, equivalently, from the mass term $\int d^d x\, m\bar{\psi}\psi$, which must be dimensionless) we can determine that their mass dimensionalities must be

$$\left[(\partial_\mu A_\nu)^2\right] = [mass]^d \rightsquigarrow [A_\nu] = [mass]^{\frac{d}{2}-1} = [mass]^{1-\epsilon} , \tag{5.4}$$

and

$$\left[\bar{\psi}\partial\!\!\!/\psi\right] = [mass]^d \rightsquigarrow \left[\psi\right] = \left[\bar{\psi}\right] = [mass]^{\frac{d-1}{2}} = [mass]^{\frac{3}{2}-\epsilon} , \tag{5.5}$$

where in the last steps we have set $d = 4 - 2\epsilon$. From the gauge fixing term, we find that the parameter ξ

$$[\xi] = [mass]^0 \,, \tag{5.6}$$

is dimensionless. Finally, the interaction term gives,

$$\left[e\bar{\psi}\not{A}\psi\right] = [mass]^d \rightsquigarrow [e] = [mass]^{\frac{4-d}{2}} \,. \tag{5.7}$$

The coupling constant e is a dimensionful parameter in dimensions other than four. In dimensional regularisation, we take $d = 4 - 2\epsilon$, which translates into

$$\left[\int d^d x \; e\,\bar{\psi}\not{A}\psi\right] = [e] + 3 - 2\epsilon + 1 - \epsilon - (4 - 2\epsilon) = 0 \quad \rightsquigarrow \quad [e] = [mass]^{\epsilon} \,. \tag{5.8}$$

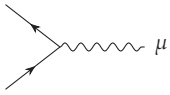
Thus, the electric charge needs to get a mass dimension "ϵ". We can then substitute, in the Lagrangian and in the Feynman rule for photon absorption or emission,

$$e \to e\mu^{\epsilon} \,, \tag{5.9}$$

introducing an arbitrary mass scale μ in the Lagrangian. The QED Lagrangian in d dimensions then reads

$$\mathcal{L} = \bar{\psi}\left[i\not{\partial} - m\right]\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2 \underbrace{+ \left(e\mu^{\epsilon}\right)\bar{\psi}\not{A}\psi}_{\mathcal{L}_{\text{int}}} \,, \tag{5.10}$$

and in our Feynman rules we have to modify the photon-fermion vertex:



$$= +ie\mu^{\epsilon}\gamma^{\mu} \,. \tag{5.11}$$

Remark. With the convention $D_\mu = \partial_\mu + ieA_\mu$ (e.g. Peskin–Schroeder) one has instead $\mathcal{L}_{\text{int}} = -e\mu^{\epsilon}\bar{\psi}\not{A}\psi$ and the vertex $-ie\mu^{\epsilon}\gamma^{\mu}$. We use $D_\mu = \partial_\mu - ieA_\mu$ throughout, as in Chapters 1 and 2; physical results depend only on e^2.

The mass parameter μ is an arbitrary scale. **Physical observables should not depend on the scale.** Naively, given that we shall take the limit $\epsilon \to 0$ after renormalisation, the arbitrary scale μ seems innocuous. In practice, we cannot get rid of μ easily, unless we are able to compute the perturbative series to all orders. Loop integrals will produce poles in ϵ of the form,

$$\frac{\mu^{\epsilon}}{\epsilon} = \frac{1}{\epsilon} + \log(\mu) + O(\epsilon). \tag{5.12}$$

The procedure of renormalisation will eliminate the poles in ϵ from the expressions for scattering matrix elements. However, the $\log(\mu)$ terms can only be eliminated by means of a resummation of all perturbative orders.

5.1.1 Gamma matrices in dimensional regularisation

How do we treat γ matrices in dimensional regularisation? In the Clifford algebra we must decide what should be taken in 4 dimensions and what in d dimensions. There are many choices for this, of varying success. The prescription that is followed in conventional dimensional regularisation, and which is safe in QED, is

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}\mathbf{1}_{4\times 4}, \tag{5.13}$$

where the metric is taken in $d = 4 - 2\epsilon$ dimensions,

$$g^{\mu}_{\mu} = 4 - 2\epsilon, \tag{5.14}$$

and the gamma matrices are four-by-four matrices,

$$\mathrm{tr}\,\mathbf{1}_{4\times 4} = 4. \tag{5.15}$$

Let us do some simple calculations with this algebra.

$$\gamma^{\mu}\gamma_{\mu} = \frac{1}{2}\left\{\gamma^{\mu}, \gamma_{\mu}\right\} = g^{\mu}_{\mu}\mathbf{1}_{4\times 4} = (4-2\epsilon)\mathbf{1}_{4\times 4}. \tag{5.16}$$

Also,

$$\begin{aligned} \gamma^{\mu}\gamma^{\nu}\gamma_{\mu} &= \{\gamma^{\mu}, \gamma^{\nu}\}\,\gamma_{\mu} - \gamma^{\nu}\gamma^{\mu}\gamma_{\mu} \\ &= 2g^{\mu\nu}\gamma_{\mu} - (4-2\epsilon)\gamma^{\nu} \\ \rightsquigarrow \gamma^{\mu}\gamma^{\nu}\gamma_{\mu} &= -2(1-\epsilon)\gamma^{\nu}, \end{aligned} \tag{5.17}$$

and so on.

A difficulty arises in defining the matrix γ_5 in $4 - 2\epsilon$ dimensions. In four dimensions it is defined as

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3. \tag{5.18}$$

It is often fine to use a prescription with an anticommuting γ_5,

$$\{\gamma^{\mu}, \gamma_5\} = 0,$$

in performing the gamma-matrix algebra for QED amplitudes. A discussion of this issue, however, lies beyond the scope of this part. We will return to it in Part V, where γ_5 in d dimensions is at the heart of the axial anomaly.

5.2 The electron propagator at one loop

We now consider the one-loop correction to the electron propagator, which is depicted by the Feynman diagram of Fig. 5.1. For simplicity, we set the mass of the electron to zero,

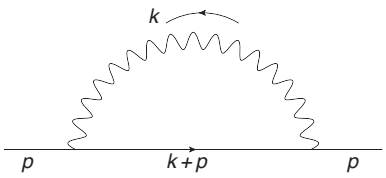


Figure 5.1. The one-loop perturbative correction to the electron propagator.

$m_e = 0$. Following our Feynman rules, this truncated diagram is

$$\begin{aligned} -i\Sigma_{1-loop} &= \int \frac{d^dk}{(2\pi)^d}\left(+ie\mu^{\epsilon}\gamma^{\mu}\right)\frac{i}{\not{k}+\not{p}}\left(+ie\mu^{\epsilon}\gamma^{\nu}\right)\frac{(-i)}{k^2}\left[g_{\mu\nu} - \bar{\xi}\frac{k_{\mu}k_{\nu}}{k^2}\right] \\ &= -e^2\mu^{2\epsilon}\int\frac{d^dk}{(2\pi)^d}\left[\frac{\gamma^{\mu}\not{k}_p\gamma_{\mu}}{k^2k_p^2} - \bar{\xi}\frac{\not{k}\not{k}_p\not{k}}{(k^2)^2k_p^2}\right], \end{aligned} \tag{5.19}$$

where we have introduced the shorthand notation $\bar{\xi} \equiv 1 - \xi$ and $k_p \equiv k + p$. We can simplify the numerator structure with gamma-matrix identities (**exercise**) such as

$$\gamma^\mu \not{k} \gamma_\mu = -2(1-\epsilon)\not{k}, \tag{5.20}$$

etc. We then find two tensor integrals, which can be reduced to scalar integrals with the techniques of Chapter 4 (Eq. 4.43):

$$\int \frac{d^d k}{(2\pi)^d} \frac{k^\mu}{k^2 k_p^2} = -\frac{p^\mu}{2} \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2 k_p^2}, \tag{5.21}$$

$$\int \frac{d^d k}{(2\pi)^d} \frac{k^\mu}{(k^2)^2 k_p^2} = -\frac{p^\mu}{2p^2} \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2 k_p^2} \left[1 + \frac{p^2}{k^2}\right]. \tag{5.22}$$

The two terms on the rhs can be simplified with integration by parts (Section 4.2). We have

$$\begin{aligned} 0 &= \int \frac{d^d k}{(2\pi)^d} \frac{\partial}{\partial k^\mu} \frac{k^\mu}{k^2 k_p^2} \\ \rightsquigarrow \quad & \int \frac{d^d k}{(2\pi)^d} \frac{p^2}{(k^2)^2 k_p^2} = -(1-2\epsilon) \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2) k_p^2}. \end{aligned} \tag{5.23}$$

Therefore,

$$\int \frac{d^d k}{(2\pi)^d} \frac{k^\mu}{(k^2)^2 k_p^2} = -\epsilon \frac{p^\mu}{p^2} \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2 k_p^2}. \tag{5.24}$$

With the tensor reduction at hand, we then easily arrive at

$$-i\Sigma_{1-loop} = e^2 \mu^{2\epsilon} \xi (1-\epsilon) \not{p} \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2 k_p^2}. \tag{5.25}$$

Substituting the expression for the master integral, Eq. 4.14, with the normalisation of the measure of Eq. 3.44, we obtain:

$$\Sigma_{1-loop} = -\frac{e^2}{(4\pi)^2} \left(-\frac{4\pi\mu^2}{p^2}\right)^\epsilon c_\Gamma \frac{(1-\epsilon)}{\epsilon(1-2\epsilon)} \xi \not{p}. \tag{5.26}$$

This expression is divergent as $\epsilon \to 0$. Expanding in ϵ, we obtain:

$$\Sigma_{1-loop} = \Sigma_{V,1-loop}(p^2) \not{p}, \tag{5.27}$$

with

$$\Sigma_{V,1-loop}(p^2) = -\frac{\frac{e^2}{4\pi}}{\pi} (4\pi)^\epsilon \, e^{-\gamma\epsilon} \left[\frac{1}{4\epsilon} + \frac{1}{4} \log\left(-\frac{\mu^2}{p^2}\right) + \frac{1}{4} + O(\epsilon)\right] \xi. \tag{5.28}$$

Notice that the artificial mass μ which we introduced in the Lagrangian to fix the dimensions of the coupling constant in dimensional regularisation combines with the physical scale p^2 in a dimensionless ratio as an argument of a logarithm. To evaluate the logarithm, we should recall that p^2 possesses a small imaginary part $p^2 + i\delta$.

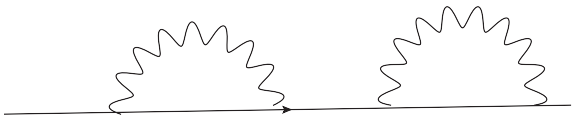


Figure 5.2. An example of a one-particle-reducible graph which contributes to the electron propagator.

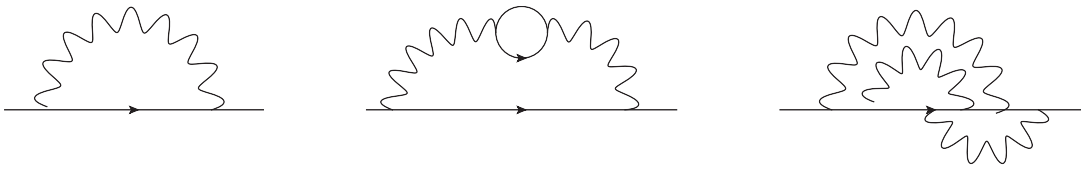


Figure 5.3. A sample of one-particle-irreducible graphs which contribute to the electron propagator.

5.3 The electron propagator at all orders

We can classify the loop corrections of the electron propagator into two classes of Feynman diagrams: one-particle-irreducible (1PI) and one-particle-reducible. You can see an example of a one-particle-reducible graph in Fig. 5.2: by removing one line, the graph separates into two graphs. In contrast, 1PI Feynman diagrams cannot be separated into two graphs by removing any single line. A sample of 1PI graphs is presented in Fig. 5.3.

Let us first examine the sum of all truncated (with the propagators of the external legs removed) 1PI Feynman diagrams, which we conventionally denote by $-i\Sigma(p)$. Obviously, it is not possible to produce an analytic expression for such a sum of an infinite number of graphs. However, we can determine its Lorentz structure as a 4×4 matrix in spin-space. Given that

$$\not{p}^{2n+1} = (p^2)^n \not{p},$$

we can have at most a linear term when expressing Σ as a polynomial in $\not{p}$:

$$\Sigma(p) = A(p^2)\mathbf{1} + \Sigma_V(p^2)\not{p}. \tag{5.29}$$

We can also convince ourselves that the coefficient of the unity $A(p^2)$ must vanish in the limit of a zero electron mass, $m \to 0$. We can isolate this coefficient by taking the trace:

$$A(p^2) = \frac{\mathrm{tr}\left(\Sigma(p)\right)}{4}. \tag{5.30}$$

Examining the truncated graphs of Fig. 5.3, we notice that they have an even number of vertices and an odd number of propagators. Each vertex and each *massless* propagator contributes a single gamma matrix according to the QED Feynman rules. Therefore, the trace of Σ will always be a sum of traces of an odd number of gamma matrices. This is zero. To reflect this, we can therefore cast Σ in the form:

$$\Sigma(p) = m\Sigma_S(p^2)\mathbf{1} + \Sigma_V(p^2)\not{p}. \tag{5.31}$$

Note that the functions Σ_S, Σ_V also have an implicit dependence on the mass of the electron m. In what follows, we will be interested in their UV divergences. For these, the electron

mass can safely be set to zero (after, of course, we have scaled out the mass factor m in the unity coefficient of Eq. 5.31).

Let us now represent the 1PI Feynman diagrams as:

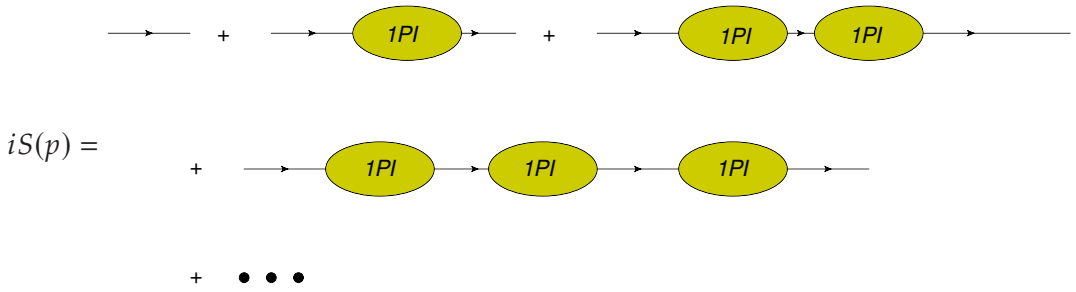


$$= -i\Sigma(p). \tag{5.32}$$

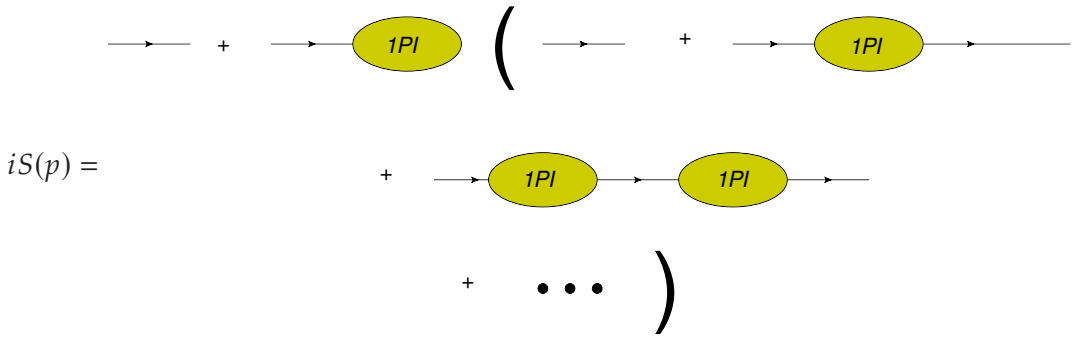
The full propagator

$$iS(p) \equiv \int d^4x e^{-ip\cdot x} \langle\Omega| T\bar{\psi}(x)\psi(0) |\Omega\rangle\,, \tag{5.33}$$

also includes the one-particle-reducible graphs. We can account for all graphs by considering the following series:



We can rearrange the above equation as follows:



In the parentheses, we recognise the original expression for $iS(p)$. Therefore, the electron propagator $iS(p)$ at all orders can be written in the recursive form:

$$iS(p) = iS_0(p) + iS_0(p)(-i\Sigma(p))iS(p), \tag{5.34}$$

where iS_0 is the propagator of the free electron field,

$$iS_0(p) = \frac{i}{\not{p} - m}, \tag{5.35}$$

and $-i\Sigma(p)$ is the sum of all one-particle-irreducible (1PI) graphs at all orders. Multiplying from the right with the inverse of S and from the left with the inverse of S_0, we then obtain

$$\begin{aligned} S^{-1}(p) &= S_0^{-1}(p) - \Sigma(p) \\ &= \not{p} - m - \not{p}\Sigma_V(p^2) - m\Sigma_S(p^2). \end{aligned} \tag{5.36}$$

Inverting the above, we find that

$$iS(p) = \frac{i}{1-\Sigma_V(p^2)} \frac{1}{\not{p} - m\dfrac{1+\Sigma_S(p^2)}{1-\Sigma_V(p^2)}}, \tag{5.37}$$

or, equivalently,

$$iS(p) = \frac{i}{1-\Sigma_V(p^2)} \frac{\not{p} + m f(p^2)}{p^2 - m^2 f^2(p^2)}, \tag{5.38}$$

with

$$f(p^2) = \frac{1+\Sigma_S(p^2)}{1-\Sigma_V(p^2)}. \tag{5.39}$$

Exercise: Show that

$$\Sigma_{S,1-loop}(p^2) = \frac{\frac{e^2}{4\pi}}{\pi} (4\pi)^\epsilon \, e^{-\gamma\epsilon} (3+\xi) \left[\frac{1}{4\epsilon} + O(\epsilon^0) \right]. \tag{5.40}$$

5.3.1 The electron mass

While the pole of the propagator S_0 for a free electron is at $p^2 = m^2$, where m is the electron mass that we inserted in the QED Lagrangian, the location of the pole in Eq. 5.37 is shifted. The physical mass of the electron should actually be determined from the full propagator, as the solution of the equation

$$p^2 - m^2 f^2(p^2) = 0. \tag{5.41}$$

Let us assume that we have solved this equation and found that it is satisfied for $p^2 = m^2_{\text{phys}}$, i.e.

$$m^2_{\text{phys}} - m^2 f^2(m^2_{\text{phys}}) = 0. \tag{5.42}$$

Expressing the mass m in terms of the physical mass and expanding in $p^2 - m^2_{\text{phys}}$, we find that

$$iS(p) = \frac{1}{\left[1-\Sigma_V(m^2_{\text{phys}})\right] \left[1 - m^2 \frac{df^2}{dp^2}(m^2_{\text{phys}})\right]} \frac{i}{\not{p} - m_{\text{phys}}} + \text{continuum}. \tag{5.43}$$

We can compare this result with the Källén–Lehmann formula for the electron propagator,

$$iS(p) = \tilde{Z}_\psi \frac{i}{\not{p} - m_{\text{phys}}} + \text{continuum}, \tag{5.44}$$

from which we obtain

$$\tilde{Z}_\psi = \frac{1}{\left[1-\Sigma_V(m^2_{\text{phys}})\right] \left[1 - m^2 \frac{df^2}{dp^2}(m^2_{\text{phys}})\right]}. \tag{5.45}$$

Exercise: *Calculate m_{phys} and $\tilde{Z}_\psi$ at one loop in perturbation theory. Notice the divergent structure of the results as $\epsilon \to 0$.*

5.4 The photon propagator at one loop

We now consider the one-loop correction to the photon propagator, which is depicted by the Feynman diagram of Fig. 5.4. For simplicity, we set the mass of the electron to zero, $m_e = 0$. Following our Feynman rules, this truncated diagram is

$$\begin{aligned} i\Pi^{\mu\nu}_{1-loop} &= (-1) \int \frac{d^d k}{(2\pi)^d} \text{tr} \left[\frac{i}{\not{k} + \not{p}} \left(+ie\mu^\epsilon \gamma^\mu\right) \frac{i}{\not{k}} \left(+ie\mu^\epsilon \gamma^\nu\right) \right] \\ &= -\left(e\mu^\epsilon\right)^2 \left[\int \frac{d^d k}{(2\pi)^d} \frac{(k_\rho + p_\rho) k_\sigma}{(k+p)^2 k^2} \right] \text{tr}\left[\gamma^\rho \gamma^\mu \gamma^\sigma \gamma^\nu\right]. \end{aligned} \tag{5.46}$$

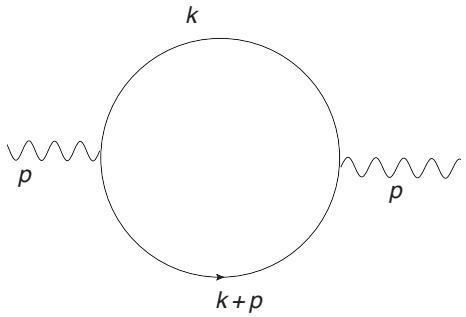


Figure 5.4. The one-loop perturbative correction to the photon propagator.

The trace of gamma matrices is

$$\mathrm{tr}\left[\gamma^\rho\gamma^\mu\gamma^\sigma\gamma^\nu\right] = 4\left[g^{\rho\mu}g^{\sigma\nu} - g^{\rho\sigma}g^{\mu\nu} + g^{\rho\nu}g^{\sigma\mu}\right], \tag{5.47}$$

where the indices μ, ν, ρ, σ are $d-$dimensional with $d = 4 - 2\epsilon$. The tensor loop integral can be computed with the Passarino–Veltman technique of Section 4.4. Using Eqs. 4.43 and 4.53, we find

$$\begin{aligned} I_{\rho\sigma} &\equiv \int \frac{d^dk}{(2\pi)^d}\frac{(k_\rho + p_\rho)k_\sigma}{(k+p)^2k^2} \\ &= -\frac{1}{4(d-1)}\left\{g_{\rho\sigma}p^2 + (d-2)p_\rho p_\sigma\right\}\int \frac{d^dk}{(2\pi)^d}\frac{1}{(k+p)^2k^2}. \end{aligned} \tag{5.48}$$

The expression for the scalar integral has been given in Eq. 4.14. We then find:

$$\Pi^{\mu\nu}_{1-loop} = \Pi_{1-loop}(p^2)\left[p^2g^{\mu\nu} - p^\mu p^\nu\right], \tag{5.49}$$

with

$$\Pi_{1-loop}(p^2) = -\frac{\frac{e^2}{4\pi}}{\pi}\left(-\frac{4\pi\mu^2}{p^2}\right)^\epsilon \frac{c_\Gamma}{\epsilon}\frac{1-\epsilon}{(1-2\epsilon)(3-2\epsilon)}. \tag{5.50}$$

We find that the expression has a $1/\epsilon$ pole and is therefore divergent in exactly $d = 4$ dimensions. Performing an expansion in ϵ, we find:

$$\Pi_{1-loop}(p^2) = -\frac{\frac{e^2}{4\pi}}{\pi}(4\pi)^\epsilon\, e^{-\gamma\epsilon}\left[\frac{1}{3\epsilon} + \frac{1}{3}\log\left(-\frac{\mu^2}{p^2}\right) + \frac{5}{9} + O(\epsilon)\right]. \tag{5.51}$$

Finally, we see that the one-loop correction $\Pi^{\mu\nu}$ is proportional to the tensor:

$$T^{\mu\nu} \equiv g^{\mu\nu} - \frac{p^\mu p^\nu}{p^2}. \tag{5.52}$$

This is a simpler structure than the one that we find in the tree-level Feynman rule for the photon propagator,

$$\sim\!\sim = \frac{i}{p^2}\left[-g_{\mu\nu} + (1-\xi)\frac{p_\mu p_\nu}{p^2}\right] = -\frac{i}{p^2}\left[T^{\mu\nu} + \xi L^{\mu\nu}\right], \tag{5.53}$$

where we find a second tensor:

$$L^{\mu\nu} \equiv \frac{p^\mu p^\nu}{p^2}, \tag{5.54}$$

multiplying the gauge-fixing parameter ξ. The tensor $T^{\mu\nu}$ is "transverse" to the momentum:

$$T^{\mu\nu}p_\mu = T^{\mu\nu}p_\nu = 0. \tag{5.55}$$

The tensor $L^{\mu\nu}$ is "parallel" to the momentum:

$$L^{\mu\nu}p_\mu = p^\nu. \tag{5.56}$$

The one-loop diagram corrects only the part of the propagator which is transverse, leaving the longitudinal part, which carries the gauge parameter, unchanged. As we shall see soon, this observation is more general and holds at all loop orders.

The tensors $T^{\mu\nu}, L^{\mu\nu}$ are projectors. Indeed, we find:

$$T^{\mu\nu}T_\nu^{\ \rho} = T^{\mu\rho}, \quad L^{\mu\nu}L_\nu^{\ \rho} = L^{\mu\rho}, \quad T^{\mu\nu}L_\nu^{\ \rho} = 0. \tag{5.57}$$

We can write any rank-two tensor which is made exclusively out of the metric and a single momentum p^μ as:

$$X^{\mu\nu} = X_T(p^2)T^{\mu\nu} + X_L(p^2)L^{\mu\nu}. \tag{5.58}$$

The inverse can be written down easily:

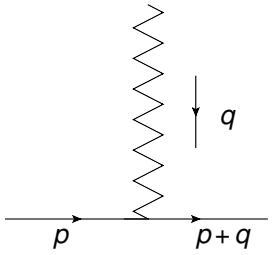
$$\left(X^{-1}\right)^{\mu\nu} = \frac{1}{X_T(p^2)}T^{\mu\nu} + \frac{1}{X_L(p^2)}L^{\mu\nu}. \tag{5.59}$$

Indeed,

$$\begin{aligned} X^{\mu\nu}\left(X^{-1}\right)_\nu^{\ \rho} &= \left(X_T(p^2)T^{\mu\nu} + X_L(p^2)L^{\mu\nu}\right)\left(\frac{1}{X_T(p^2)}T_\nu^{\ \rho} + \frac{1}{X_L(p^2)}L_\nu^{\ \rho}\right) \\ &= T^{\mu\rho} + L^{\mu\rho} = g^{\mu\rho}. \end{aligned} \tag{5.60}$$

5.5 Ward identity*

Consider the tree-level (i.e. zero-loop) electron-electron-photon Green's function:

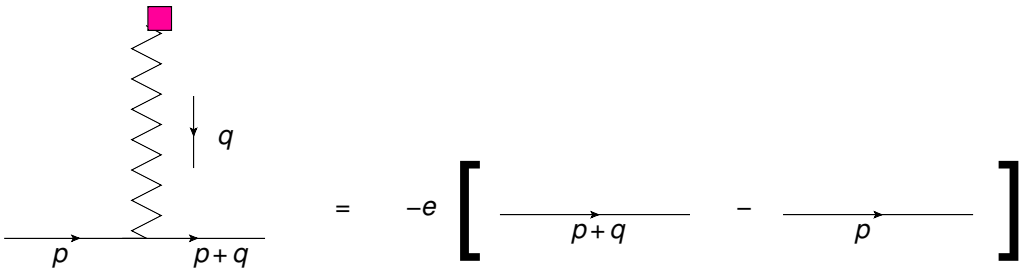


$$\equiv ie\epsilon_\mu V_0^\mu = \epsilon_\mu \frac{i}{\not p + \not q - m}\left(+ie\gamma^\mu\right)\frac{i}{\not p - m} \tag{5.61}$$

where we have truncated the photon external line, but left the electron lines untruncated. Now, let us replace the polarisation vector with the momentum of the photon, $\epsilon_\mu \to q_\mu$. We find

$$\begin{aligned} ieq_\mu V_0^\mu &= \frac{i}{\not p + \not q - m}\left(+ie\not q\right)\frac{i}{\not p - m} \\ &= -ie\left[\frac{1}{\not p + \not q - m}\left(\left[\not p + \not q - m\right] - \left[\not p - m\right]\right)\frac{1}{\not p - m}\right] \\ &= ie\left[\frac{1}{\not p + \not q - m} - \frac{1}{\not p - m}\right] \\ \rightsquigarrow \quad & ieq_\mu V_0^\mu = e\left[iS_0(p+q) - iS_0(p)\right]. \end{aligned} \tag{5.62}$$

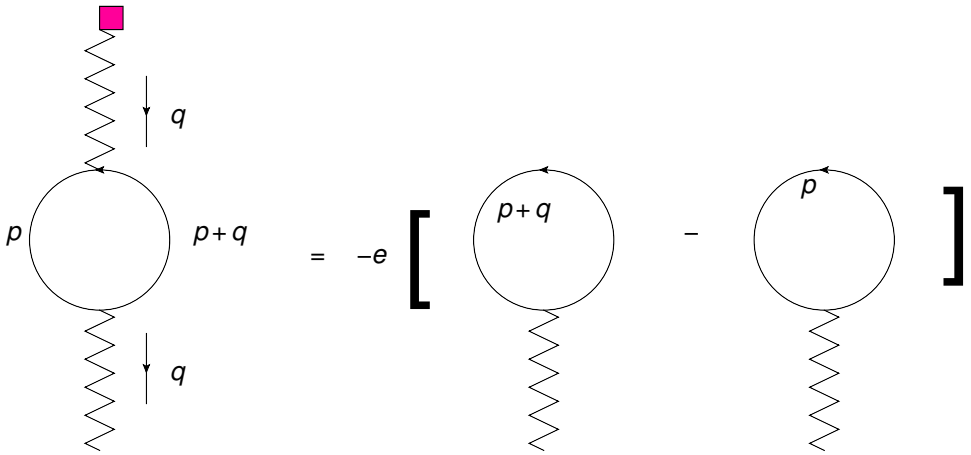
Graphically, we can represent the above equation as:



(5.63)

where the filled square at the start of the incoming photon line represents the operation of substituting the polarisation vector with the momentum vector. This identity has important consequences at all orders in perturbation theory.

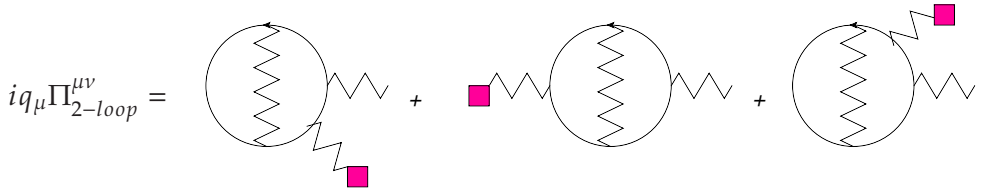
Consider the example of the previous section, the one-loop diagram for the photon propagator $\Pi^{\mu\nu}_{1-loop}$, and contract it with the momentum of the photon, $q_\mu \Pi^{\mu\nu}_{1-loop}$. We have found, after an explicit calculation, that $q_\mu \Pi^{\mu\nu}_{1-loop} = 0$. We can now elucidate this result by means of Eq. 5.63. Indeed, we can join together the ends of the electron lines on both sides of Eq. 5.63 forming one-loop diagrams. In addition, we can attach a second photon to these loops. The resulting identity is then:



(5.64)

Notice that the diagrams on the rhs are identical, except that the momenta $p, p+q$ circulating in their loops differ by a shift. However, the loop momentum p is an integration variable and we are allowed (at least in dimensional regularisation) to make a change of variables $p \longrightarrow p+q$ in the first of the two terms on the rhs. With this transformation, the two integrals on the rhs are shown to be identical and cancel. This is an alternative proof that the one-loop correction to the photon propagator is transverse, $q_\mu \Pi^{\mu\nu}_{1-loop} = 0$.

The generalisation of this identity to higher loop orders is straightforward. The two-loop 1PI diagrams which contribute to the photon propagator $q_\mu \Pi^{\mu\nu}_{2-loop}$ are:

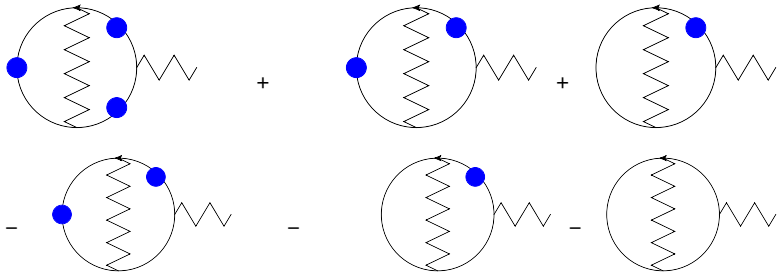


(5.65)

We can now apply the identity of Eq. 5.63 for the photon lines which are contracted with

q_μ. The result is proportional to the following diagrams:

$$iq_\mu \Pi^{\mu\nu}_{2-loop} \propto$$



(5.66)

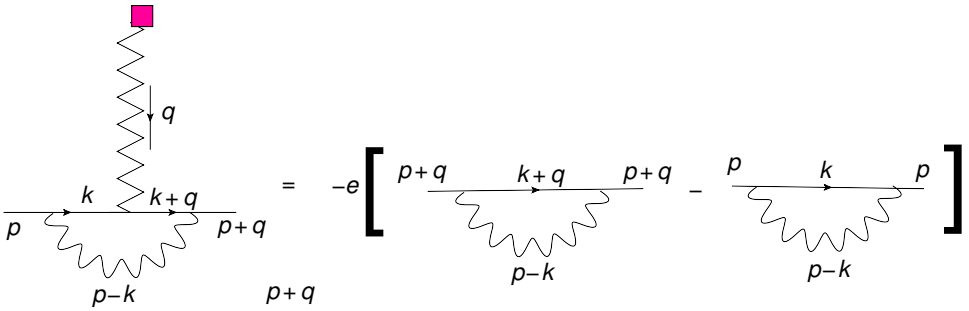
The blue dot on a line signifies that the momentum of this propagator is shifted by q with respect to the analogous propagator in the diagram without any dots. As you can see, four of these terms cancel in pairs without any further manipulation. The first and the last term differ by a shift of both the loop momenta by q^μ. Therefore, they also cancel, as can be shown with an appropriate change of variables. We have therefore shown without performing the loop integrations that

$$iq_\mu \Pi^{\mu\nu}_{2-loop} = 0. \tag{5.67}$$

Exercise: It is not hard to convince yourselves that this result holds at three and higher loop orders as well:

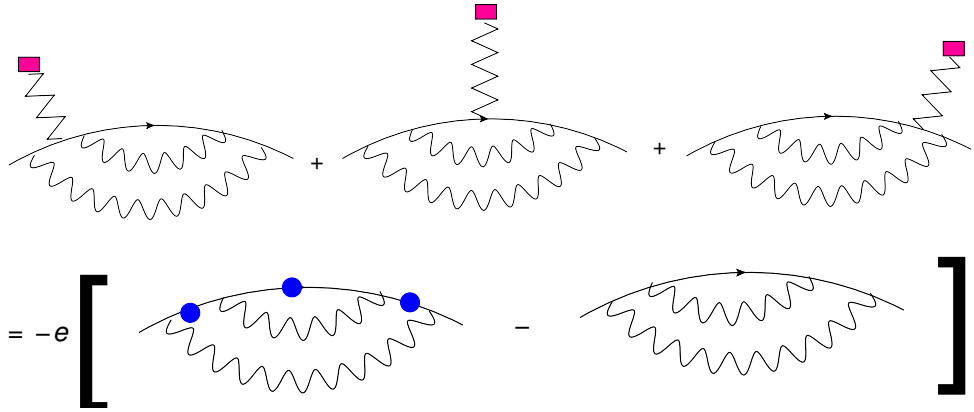
$$q_\mu \Pi^{\mu\nu} = 0. \tag{5.68}$$

Let us now look at the one-loop correction to the three-point function (electron-electron-photon vertex), substituting the photon polarisation with the momentum $\epsilon^\mu \to q^\mu$. Applying the identity of Eq. 5.63, we find that



(5.69)

We observe that by replacing the polarisation with the momentum, the 3-point vertex becomes the difference of propagator diagrams. After cancellations of the type we observed in Eq. 5.66, we find that a similar result holds for all two-loop vertex diagrams.
Exercise: Prove the following:



(where we recall that the dot signifies a propagator momentum shifted by q). This result holds in general. We can write the so-called "Ward identity",

$$q_\mu \left(ie\Gamma^\mu\right) = ie\left[S^{-1}(p+q) - S^{-1}(p)\right], \tag{5.70}$$

where by $ie\epsilon_\mu\Gamma^\mu$ we denote the truncated (one-particle-irreducible) 3-point electron-electron-photon vertex function.

5.6 The photon propagator at all orders

We are now in a position to determine the functional form of the full photon propagator at all orders,

$$-iD_{\mu\nu}(p) \equiv \int d^4x e^{-ip\cdot x} \langle\Omega| T A^\mu(x) A^\nu(0) |\Omega\rangle . \tag{5.71}$$

This contains both one-particle-reducible and one-particle-irreducible Feynman diagrams. Following faithfully the analysis which led to Eq. 5.34, we find in the same way a recursion formula for the photon propagator:

$$-iD_{\mu\nu}(p) = -iD^0_{\mu\nu}(p) + (-iD^0_{\mu\rho}(p))(i\Pi^{\rho\sigma})(-iD_{\sigma\nu}(p)). \tag{5.72}$$

In the above, the propagator of a free photon is given by Eq. 2.32 and $\Pi^{\mu\nu}$ is the sum of all 1PI graphs at all orders in perturbation theory. We can multiply Eq. 5.72 with $\left(D^{-1}\right)^{\lambda\mu}(D_0^{-1})^{\nu\lambda}$ and recast it in the form:

$$\left(D^{-1}\right)_{\mu\nu} = \left(D^0\right)^{-1}_{\mu\nu} - \Pi_{\mu\nu}. \tag{5.73}$$

As we have seen in the previous section, $\Pi^{\mu\nu}$ is a transverse tensor:

$$\Pi^{\mu\nu} = \Pi(p^2)(g^{\mu\nu}p^2 - p^\mu p^\nu) = \Pi(p^2)p^2 T^{\mu\nu}. \tag{5.74}$$

Substituting Eq. 2.32 and Eq. 5.74 into Eq. 5.73 and inverting one more time, we find that the photon propagator at all orders takes the form:

$$-iD^{\mu\nu}(p) = -\frac{i}{p^2}\left[\frac{T^{\mu\nu}}{1-\Pi(p^2)} + \xi L^{\mu\nu}\right]. \tag{5.75}$$

Notice that there is a pole at $p^2 = 0$. Therefore, the photon remains massless when we include perturbative corrections.

6

Renormalisation of QED

We have found in the previous chapter that the one-loop corrections to the electron and photon propagators are divergent; they contain poles $\frac{1}{\epsilon}$ in dimensional regularisation. In this chapter we will see that these infinities can be absorbed into a redefinition of the fields and of the parameters of the QED Lagrangian, as anticipated in Chapter 3.

Let us, for simplicity, work with QED in the case of a massless electron. The QED Lagrangian in the Lorenz gauge is given in this case by

$$\mathcal{L} = i\bar{\psi}\partial\!\!\!/\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2 + e\mu^\epsilon\bar{\psi}A\!\!\!/\psi. \tag{6.1}$$

This Lagrangian is built out of the requirement of gauge symmetry and the requirement of not introducing any mass scales (other than the mass of the electron, which we set to zero).

Without spoiling our symmetry principle, we are allowed to redefine multiplicatively the fields as well as the parameters e, ξ of the Lagrangian:

$$\psi = Z_\psi^{\frac{1}{2}}\tilde{\psi}, \tag{6.2}$$

$$A^\mu = Z_A^{\frac{1}{2}}\tilde{A}^\mu, \tag{6.3}$$

$$e = Z_\alpha^{\frac{1}{2}}\tilde{e}, \tag{6.4}$$

$$\xi = Z_\xi\tilde{\xi}. \tag{6.5}$$

Such rescalings only recalibrate the scales in which we read our results. They do not change the physics, and they respect all the symmetries of the Lagrangian. There is, however, one requirement. We do perturbation theory around the free theory, whose spectrum of free particles we know, and for that the Lagrangian must reduce to the correctly normalised free Lagrangian when the interaction is switched off. This does not force the Z factors to be one. It only requires that they become one when the coupling vanishes, $Z = 1 + O(\alpha)$. The higher-order terms are otherwise unrestricted, and we will exploit this freedom by allowing them to contain $\frac{1}{\epsilon}$ poles, chosen such that they cancel the divergences of Green's functions. With these redefinitions, the Lagrangian takes the form

$$\mathcal{L} = Z_\psi i\bar{\tilde{\psi}}\partial\!\!\!/\tilde{\psi} - \frac{Z_A}{4}\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu} - \frac{Z_A}{2Z_\xi\tilde{\xi}}\left(\partial_\mu\tilde{A}^\mu\right)^2 + Z_\alpha^{\frac{1}{2}}Z_A^{\frac{1}{2}}Z_\psi\tilde{e}\mu^\epsilon\bar{\tilde{\psi}}\tilde{A}\!\!\!/\tilde{\psi}. \tag{6.6}$$

Consider now the photon two-point correlation function:

$$\langle\Omega|\,TA^\mu(x)A^\nu(0)\,|\Omega\rangle = Z_A\,\langle\Omega|\,T\tilde{A}^\mu(x)\tilde{A}^\nu(0)\,|\Omega\rangle\,. \tag{6.7}$$

Taking the Fourier transform, we obtain

$$D^{\mu\nu}(p) = Z_A \tilde{D}^{\mu\nu}(p), \tag{6.8}$$

where $\tilde{D}^{\mu\nu}$ is the propagator for the renormalised field $\tilde{A}^\mu$. Substituting the result of Eq. 5.75, we find

$$\tilde{D}^{\mu\nu} = \frac{1}{p^2}\left[\frac{T^{\mu\nu}}{Z_A(1-\Pi(p^2))} + \frac{Z_\xi}{Z_A}\tilde{\xi}L^{\mu\nu}\right]. \tag{6.9}$$

The factor $1-\Pi(p^2)$ contains infinities in the form of $1/\epsilon$ poles, as we have observed at the one-loop level (Eq. 5.51). For generic values of Z_A, Z_ξ, the expansion in ϵ of the expression above would therefore be divergent. Is there an exception to this?

Notice that the $1/\epsilon$ pole in the expression of Eq. 5.51 has a coefficient which is independent of p^2 and consists of constants only. This allows us to cancel the infinity in $Z_A(1-\Pi(p^2))$ with a judicious choice of Z_A. Indeed, let us set

$$Z_\alpha = 1 + O(\alpha), \tag{6.10}$$

and

$$Z_A = 1 - \frac{\alpha(\mu_R)}{\pi}\left(\frac{1}{3\epsilon}\right) + O(\alpha^2), \tag{6.11}$$

where we have defined the renormalised coupling constant $\alpha(\mu_R)$ via

$$\frac{e^2}{4\pi}e^{-\gamma\epsilon}(4\pi\mu^2)^\epsilon = Z_\alpha\frac{\tilde{e}^2}{4\pi}e^{-\gamma\epsilon}(4\pi\mu^2)^\epsilon \equiv Z_\alpha(\alpha(\mu_R))\alpha(\mu_R)\mu_R^{2\epsilon}. \tag{6.12}$$

The mass parameter μ_R is an arbitrary "renormalisation scale" which in effect replaces the scale μ of dimensional regularisation. With these definitions, we find

$$\begin{aligned}
&Z_A(1-\Pi(p^2)) = Z_A(1-\Pi_{1-loop}(p^2)+O(\alpha^2)) \\
&= \left(1 - \frac{\alpha(\mu_R)}{\pi}\left(\frac{1}{3\epsilon}\right) + O(\alpha^2)\right) \\
&\times \left(1 + \frac{\alpha(\mu_R)}{\pi}\left[\frac{1}{3\epsilon} + \frac{5}{9} + \frac{1}{3}\log\left(-\frac{\mu_R^2}{p^2}\right)\right] + O(\alpha^2)\right) \\
&= 1 + \frac{\alpha(\mu_R)}{\pi}\left[\frac{5}{9} + \frac{1}{3}\log\left(-\frac{\mu_R^2}{p^2}\right)\right] + O(\alpha^2),
\end{aligned} \tag{6.13}$$

which is finite. Therefore, there is a photon field redefinition $A^\mu \to \tilde{A}^\mu$, with $A^\mu = Z_A^{\frac{1}{2}}\tilde{A}^\mu$, which renders the transverse part of the $\tilde{A}^\mu$ photon propagator (Eq. 6.9) finite. In fact, given that loop corrections at all orders are transverse and do not change the coefficient of $L^{\mu\nu}$ in Eq. 6.9, we can guarantee that also the longitudinal part is finite by setting

$$Z_\xi = Z_A. \tag{6.14}$$

We can repeat the same process for the electron propagator:

$$\langle\Omega|\, T\bar{\psi}(x)\psi(0)\,|\Omega\rangle = Z_\psi \langle\Omega|\, T\bar{\tilde{\psi}}(x)\tilde{\psi}(0)\,|\Omega\rangle\,, \tag{6.15}$$

which, by taking the Fourier transform, yields

$$S(p) = Z_\psi \tilde{S}(p). \tag{6.16}$$

As for the renormalised photon field, we will require that the renormalised electron field $\tilde{\psi}$ yields a propagator which is free of $1/\epsilon$ divergences. Substituting Eq. 5.37, we find

$$\tilde{S}(p) = \frac{1}{Z_\psi\left(1-\Sigma_V(p^2)\right)} \frac{1}{\not{p} - \tilde{m}\dfrac{Z_m Z_\psi\left(1+\Sigma_S(p^2)\right)}{Z_\psi\left(1-\Sigma_V(p^2)\right)}}. \tag{6.17}$$

We can fix the factors Z_ψ, Z_m by requiring that the following two quantities are finite:

$$Z_\psi\left(1-\Sigma_V(p^2)\right) = \text{finite}, \quad Z_m Z_\psi\left(1+\Sigma_S(p^2)\right) = \text{finite}. \tag{6.18}$$

Remark. Here we have restored the electron mass, $m = Z_m\tilde{m}$, which was set to zero in the Lagrangian at the beginning of this chapter; a mass term $-Z_\psi Z_m \tilde{m}\bar{\tilde{\psi}}\tilde{\psi}$ is then to be added to Eq. 6.6. With Σ_V and Σ_S of Eqs. 5.28 and 5.40, the poles of the choice below indeed cancel: $Z_\psi(1-\Sigma_V) = 1 + O(\epsilon^0)$ and $Z_m Z_\psi(1+\Sigma_S) = 1 + \frac{\alpha}{4\pi\epsilon}(-3-\xi+3+\xi) + \ldots$; Z_m is gauge independent, as it must be. We can achieve this by choosing, for example,

$$Z_\psi = 1 - \frac{\alpha(\mu_R)}{\pi}\frac{\xi}{4\epsilon} + O(\alpha^2), \tag{6.19}$$

and

$$Z_m = 1 - \frac{\alpha(\mu_R)}{\pi}\frac{3}{4\epsilon} + O(\alpha^2). \tag{6.20}$$

Finally, the renormalisation constant Z_α serves to cancel the UV divergences in the photon-fermion-fermion vertex. In fact, writing the Ward identity of Eq. 5.70 in terms of renormalised quantities, we have

$$q_\mu \Gamma^\mu = Z_\psi^{-1}\left[\tilde{S}^{-1}(p+q) - \tilde{S}^{-1}(p)\right]. \tag{6.21}$$

On the rhs, the divergent part is Z_ψ^{-1} (which is a constant independent of p^μ, q^μ). Therefore, the combination

$$Z_\psi \Gamma^\mu \sim \text{ finite}\,. \tag{6.22}$$

The physical matrix element for an electron-photon interaction is

$$\tilde{Z}_\psi \tilde{Z}_A^{\frac{1}{2}} e\Gamma^\mu = \tilde{Z}_A^{\frac{1}{2}} Z_\alpha^{\frac{1}{2}} \tilde{e}\left(\tilde{Z}_\psi \Gamma^\mu\right), \tag{6.23}$$

where we have included the appropriate factors $\tilde{Z}$ of the LSZ reduction formula of Eq. 12.81 of Part I. The factors $\tilde{Z}$ and Z do not need to be the same. However, comparing the general Källén–Lehmann form of the propagator, in which the $\tilde{Z}$ appear, and the conditions we set for determining the corresponding field renormalisation Z, we arrive at the conclusion that $\tilde{Z}$ must have the same divergences as Z (**exercise**). The matrix element of Eq. 6.23 must be finite in order to make physical sense. Given that the combination $\left(\tilde{Z}_\psi \Gamma^\mu\right)$ is finite, we can achieve this by choosing

$$Z_\alpha = Z_A^{-1}. \tag{6.24}$$

We now have to inspect one-loop matrix elements with four or more external particles for infinities, as for example in the process $e^+e^- \to \gamma\gamma$. Notice that we have already determined all Z factors in the renormalised QED Lagrangian of Eq. 6.6. If we find any infinite matrix elements beyond the ones we have already used to fix the Z constants, we will not be able to absorb these infinities into a renormalisation without adding new terms to the Lagrangian. However, in QED, the infinities appear only in the propagators and the

electron-photon vertex. It also turns out that the Z factors suffice to absorb the infinities at all loop orders and not just the one-loop case that we have considered here. A proof of the above statements will be the subject of QFT II (Part III).

A comment is in order about the choice of the Z constants in Eqs. 6.11, 6.14, 6.19, 6.20 and 6.24, and the definition of the coupling constant in Eq. 6.12. Clearly, these are not unique. We can change any Z by a finite constant,

$$Z \to Z' = Z + \text{finite}, \tag{6.25}$$

achieving the same goal of absorbing the infinities from the propagators and the electron-photon vertex. The choice of finite pieces for the Z factors defines a "renormalisation scheme". Our choice corresponds to the "modified minimal subtraction" ($\overline{\text{MS}}$) renormalisation scheme. The scheme dependence is not an ambiguity in the physical predictions of the theory. Renormalised Green's functions, and therefore the values of renormalised parameters such as $\alpha(\mu_R)$, depend on the scheme. However, we determine the renormalised parameters by comparing some of our results with measurements. Once our predictions are expressed in terms of measured quantities, they are the same in every scheme, up to higher orders in perturbation theory than the ones we have computed.

6.1 Running of the QED coupling constant and the electron mass*

Consider Eq. 6.12,

$$\frac{e^2}{4\pi} e^{-\gamma\epsilon} (4\pi\mu^2)^\epsilon = Z_\alpha(\alpha(\mu_R))\alpha(\mu_R)\mu_R^{2\epsilon}. \tag{6.26}$$

The general structure of the Z_α factor through two loops is

$$Z_\alpha(\alpha(\mu_R)) = 1 - \frac{\beta_0}{\epsilon}\frac{\alpha(\mu_R)}{\pi} + \left[\frac{\beta_0^2}{\epsilon^2} - \frac{1}{2}\frac{\beta_1}{\epsilon}\right]\left(\frac{\alpha(\mu_R)}{\pi}\right)^2 + O(\alpha^3). \tag{6.27}$$

The lhs of the above equation does not depend on the scale μ_R. Differentiating both sides of the equation with respect to μ_R^2, we must have

$$0 = \frac{d}{d\mu_R^2}\left(Z_\alpha(\alpha(\mu_R))\alpha(\mu_R)\mu_R^{2\epsilon}\right). \tag{6.28}$$

This leads to the equation

$$\frac{d\log\alpha}{d\log\mu_R^2} = -\epsilon - \beta(\alpha(\mu_R)), \tag{6.29}$$

where the so-called "beta" function is given by

$$\beta(\alpha) = \beta_0\frac{\alpha}{\pi} + \beta_1\left(\frac{\alpha}{\pi}\right)^2 + \beta_2\left(\frac{\alpha}{\pi}\right)^3 + \cdots. \tag{6.30}$$

Remark. For QED, $Z_\alpha = Z_A^{-1} = 1 + \frac{\alpha}{3\pi\epsilon} + \ldots$ (Eqs. 6.11, 6.24), so $\beta_0 = -\frac{1}{3}$; with one massless Dirac fermion the two-loop coefficient is $\beta_1 = -\frac{1}{4}$. In four dimensions $\frac{d\alpha}{d\log\mu_R^2} = \frac{\alpha^2}{3\pi} + \frac{\alpha^3}{4\pi^2} + \ldots$: the QED coupling grows with the scale (screening), e.g. $\alpha^{-1} \approx 137$ at $q^2 \to 0$ and ≈ 128 at M_Z (see Ref. [2], Sec. 7.5 and Ch. 12, and the Particle Data Group review "Electroweak model and constraints on new physics").

7

Higgs boson production at NLO in QCD

We close this part with a complete calculation at next-to-leading order (NLO) in perturbation theory, for a process which is measured at the Large Hadron Collider (LHC): the production of a Higgs boson in the fusion of two gluons. All the tools which we have developed so far appear in this calculation: dimensional regularisation and its $d = 4 - 2\epsilon$ poles (chapter 3), the one-loop bubble integral and its relatives (chapter 4), and ultraviolet renormalisation (chapter 6). In addition, we will meet for the first time, in a realistic setting, the *infrared* divergences which the remark after eq. (3.15) warned us about. We will see that they are not removed by renormalisation. Some of them cancel when we add real radiation, and the rest are absorbed into the definition of the parton densities of the proton. This is the general pattern of every NLO QCD calculation, and it is worth seeing it once in full detail, with nothing swept under the carpet.

The process has a further pedagogical advantage. Gluons do not couple to the Higgs boson at tree level; the leading order is already a one-loop amplitude. However, in the limit of a heavy top quark the top loop shrinks to a point, and the NLO calculation becomes a one-loop calculation in an effective theory. All the integrals that we need are then of the type that we have already computed.

7.1 Hadron colliders and the factorisation theorem

At LEP and SLC we have learned that the Standard Model describes the electroweak interactions with a remarkable precision. Precision measurements pointed to a light Higgs boson, which was then discovered at the LHC in 2012. Proton colliders such as the Tevatron and the LHC probe directly the physics of electroweak symmetry breaking: the Higgs boson, pairs of electroweak gauge bosons, top quarks, and possible new particles. The colliding objects, however, are protons, i.e. bound states of quarks and gluons, and the theoretical tool which connects the Lagrangian of the Standard Model to what is measured is perturbative Quantum Chromodynamics (QCD). Let us first understand why perturbation theory can be used at all for a process with protons in the initial state.

When a heavy particle of mass M, or a jet of large transverse momentum, is produced, the interaction happens in a very short time (figure 7.1),

$$\tau_{\text{hard}} \sim \frac{1}{M}. \tag{7.1}$$

The interactions which keep the quarks and gluons bound inside the proton are much slower,

$$\tau_{\text{proton}} \sim \frac{1}{M_p} \gg \tau_{\text{hard}}, \tag{7.2}$$

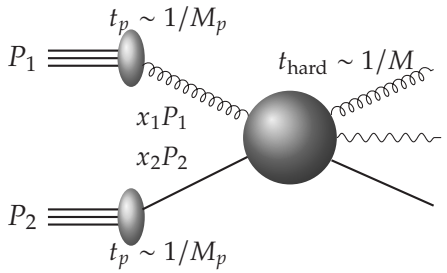


Figure 7.1. A hard scattering in a proton-proton collision. Partons with momenta x_1P_1 and x_2P_2 take part in a short-distance process at a scale $M \gg M_p$.

and they cannot change the content of the proton during the hard scattering. The proton then acts as a beam of quarks and gluons (partons), with momenta $p_1 = x_1P_1$ and $p_2 = x_2P_2$. The momentum distributions $f_i(x)$ of the partons $i = g, u, \bar{u}, d, \ldots$ inside the proton are not computable in perturbation theory, but they do not depend on the hard process. They are universal and can be measured once and for all. The cross-section is then

$$\boxed{\sigma = \sum_{ij} \int_0^1 dx_1\, dx_2\; f_i(x_1, \mu_F)\, f_j(x_2, \mu_F)\; \hat{\sigma}_{ij}\big(\hat{s} = x_1x_2S, \mu_F, \mu_R\big)\,,} \tag{7.3}$$

where $S = (P_1 + P_2)^2$ and $\hat{\sigma}_{ij}$ is the cross-section for the scattering of the partons i and j, the *partonic cross-section*. This is the *factorisation theorem*. The scales μ_F and μ_R will be explained in the course of this chapter. The partonic cross-section is a short-distance quantity. At large scales the strong coupling is small and we can compute it in perturbation theory,

$$\hat{\sigma}_{ij} = \alpha_s^N \left(\text{LO} + \alpha_s\, \text{NLO} + \alpha_s^2\, \text{NNLO} + \ldots\right). \tag{7.4}$$

Does this picture survive beyond the leading order? Perturbation theory contains all types of interactions. A loop integral $\int d^4k/(k^2 \cdots)$ receives contributions from infrared ($k \sim 0$), hard ($k \sim M$) and ultraviolet ($k \to \infty$) momenta. The answer, which is the content of the factorisation proofs of Collins, Soper and Sterman, is *yes*:

- the ultraviolet contributions *renormalise* the parameters of the Lagrangian, e.g. the strong coupling α_s;
- some infrared contributions *cancel* in physically sensible observables;
- the remaining infrared contributions *change the parton distribution functions*.

We will verify each of these statements explicitly in our example.

From the many processes at the LHC we choose the production of a Higgs boson. Its dominant production mechanism is gluon fusion, $gg \to h$, through a loop of top quarks. This is a simple process, with a single particle in the final state, and at the same time a process of great importance: it was the main discovery channel of the Higgs boson, and it is today the channel with the largest rate for the study of its properties.

7.2 A minimum of QCD

QCD is the $SU(3)$ gauge theory of quarks and gluons. Its Lagrangian is eq. (3.34) of Part III,

$$\mathcal{L}_{\text{QCD}} = \sum_{\text{flavours}} \bar{\psi}\,(i\not{D} - m)\,\psi - \frac{1}{4} G^a_{\mu\nu} G^{a\,\mu\nu}, \tag{7.5}$$

with

$$D_\mu = \partial_\mu - ig\,T^a A^a_\mu, \qquad G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\,f^{abc} A^b_\mu A^c_\nu, \tag{7.6}$$

plus gauge-fixing and ghost terms. We will need very little of the general theory of non-abelian gauge theories which is developed in Part III. The gluon A^a_μ, $a = 1, \ldots, N^2-1$, is the analogue of the photon, with $N_A \equiv N^2 - 1 = 8$ colours. The quarks carry $N = 3$ colours. The generators T^a of the fundamental representation and the structure constants f^{abc} satisfy the relations

$$\mathrm{Tr}\left(T^a T^b\right) = T_R\,\delta^{ab}, \qquad \sum_a T^a_{ik} T^a_{kj} = C_F\,\delta_{ij}, \qquad \sum_{c,d} f^{acd} f^{bcd} = C_A\,\delta^{ab}, \tag{7.7}$$

with

$$T_R = \frac{1}{2}, \qquad C_F = \frac{N^2-1}{2N} = \frac{4}{3}, \qquad C_A = N = 3. \tag{7.8}$$

These "colour factors" replace the trivial charge factors of QED. We will keep them general where this helps to recognise the origin of a term. The number of light (massless) quark flavours is $n_f = 5$.

The Feynman rules which we need are, in Feynman gauge and with all momenta incoming:

- gluon propagator: $-\dfrac{i\,g_{\mu\nu}\,\delta^{ab}}{k^2 + i\delta}$;
- quark-gluon vertex: $ig\,\mu^\epsilon\,\gamma^\mu T^a$ (the same sign convention as in QED, eq. (5.10));
- three-gluon vertex, for gluons (a,μ,k_1), (b,ν,k_2), (c,ρ,k_3):

$$V^{abc}_{\mu\nu\rho}(k_1,k_2,k_3) = g\,\mu^\epsilon\,f^{abc}\left[g_{\mu\nu}(k_1-k_2)_\rho + g_{\nu\rho}(k_2-k_3)_\mu + g_{\rho\mu}(k_3-k_1)_\nu\right]; \tag{7.9}$$

- four-gluon vertex, for gluons (a,μ), (b,ν), (c,ρ), (d,σ):

$$\begin{aligned} -ig^2\mu^{2\epsilon}\Big[& f^{abe}f^{cde}\left(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}\right) + f^{ace}f^{bde}\left(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g_{\nu\rho}\right) \\ & + f^{ade}f^{bce}\left(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma}\right)\Big]. \end{aligned} \tag{7.10}$$

As in QED, in d dimensions the coupling acquires the mass dimension ϵ and we write it as $g\mu^\epsilon$. Ghosts will play no role in what follows: they do not couple to the Higgs boson, and the only ghost loops that could appear are self-energy insertions on massless external legs, which are scaleless and vanish (eq. (3.50)).

The strong coupling $\alpha_s = g^2/(4\pi)$ depends on the renormalisation scale μ. At one loop,

$$\mu^2\frac{d\alpha_s}{d\mu^2} = -\beta_0\,\frac{\alpha_s^2}{\pi} + O(\alpha_s^3), \qquad \beta_0 = \frac{11\,C_A - 4\,T_R\,n_f}{12} = \frac{33-2n_f}{12}, \tag{7.11}$$

so that

$$\alpha_s(\mu) = \frac{\alpha_s(\mu_0)}{1 + \beta_0\,\frac{\alpha_s(\mu_0)}{\pi}\ln\frac{\mu^2}{\mu_0^2}} = \alpha_s(\mu_0)\left[1 - \beta_0\,\frac{\alpha_s(\mu_0)}{\pi}\ln\frac{\mu^2}{\mu_0^2} + O(\alpha_s^2)\right]. \tag{7.12}$$

Since $\beta_0 > 0$, the coupling decreases at short distances (asymptotic freedom); this is what makes the partonic cross-section computable. At $\mu = m_h = 125$ GeV we have $\alpha_s \simeq 0.113$, i.e. $\alpha_s/\pi \simeq 0.036$. In this chapter the natural expansion parameter is α_s/π.

In the $\overline{\text{MS}}$ scheme the bare coupling, which appears in the Feynman rules, is related to the renormalised coupling $\alpha_s(\mu)$ by

$$\alpha_s^{\text{bare}} = \alpha_s(\mu)\,\mu^{2\epsilon}\,(4\pi)^{-\epsilon}e^{\gamma\epsilon}\left[1 - \frac{\alpha_s(\mu)}{\pi}\frac{\beta_0}{\epsilon} + O(\alpha_s^2)\right], \tag{7.13}$$

where γ is the Euler constant of eq. (3.38). The factor $(4\pi)^{-\epsilon}e^{\gamma\epsilon}$ removes the $\ln(4\pi)$ and γ constants which accompany every $1/\epsilon$ pole of a loop integral. Note that the β_0 of the counterterm is the same β_0 that drives the running in eq. (7.11): the μ-independence of the bare coupling implies one from the other. This is exactly the structure that we have found for QED in eq. (6.12) and in the section on the running coupling of chapter 6. There $\beta_0 = -\frac{1}{3}$ is negative, and the QED coupling grows at short distances.

7.3 Leading order: $gg \to h$ through a top-quark loop

The Higgs boson couples to quarks proportionally to their mass, $\mathcal{L} \supset -\frac{m_q}{v}\,h\,\bar\psi_q\psi_q$, with $v \simeq 246$ GeV. It does not couple to gluons at tree level. The leading-order amplitude for

$$g(p_1,a,\epsilon_1) + g(p_2,b,\epsilon_2) \to h(p_1+p_2), \qquad p_1^2 = p_2^2 = 0, \qquad s \equiv (p_1+p_2)^2 = 2p_1\cdot p_2 = m_h^2, \tag{7.14}$$

is therefore given by the quark loops of figure 7.2. Light quarks have a negligible Yukawa coupling and we keep only the top quark.

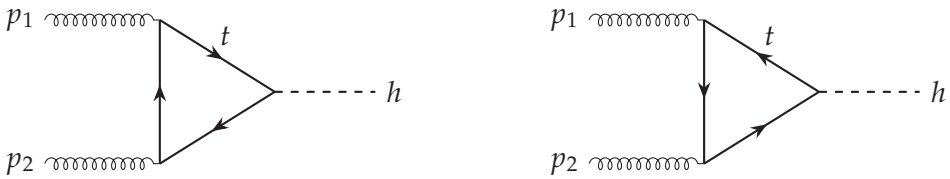


Figure 7.2. The leading-order diagrams for $gg \to h$: a top-quark loop with the two orientations of the fermion line.

Before computing anything, let us ask what the answer can be. The amplitude has the form $\mathcal{M} = \mathcal{M}^{\mu\nu}\epsilon_{1\mu}\epsilon_{2\nu}$, and the colour factor $\text{Tr}(T^aT^b) = T_R\,\delta^{ab}$. The tensor $\mathcal{M}^{\mu\nu}$ can only be built out of $g^{\mu\nu}$ and the two momenta. Terms proportional to p_1^μ or p_2^ν vanish when contracted with the physical polarisations ($p_1\cdot\epsilon_1 = p_2\cdot\epsilon_2 = 0$), so that $\mathcal{M}^{\mu\nu} \to A\,g^{\mu\nu} + B\,p_2^\mu p_1^\nu$. Gauge invariance requires that the amplitude vanishes for $\epsilon_1 \to p_1$, which fixes $B = -A/(p_1\cdot p_2)$. Therefore

$$\mathcal{M}^{\mu\nu} \propto \delta^{ab}\left(g^{\mu\nu}\,p_1\cdot p_2 - p_1^\nu p_2^\mu\right). \tag{7.15}$$

There is a single form factor, which we can extract with the projector

$$P_{\mu\nu} = \frac{1}{d-2}\left(g_{\mu\nu} - \frac{p_{1\mu}p_{2\nu} + p_{2\mu}p_{1\nu}}{p_1\cdot p_2}\right), \qquad P_{\mu\nu}\left(g^{\mu\nu}\,p_1\cdot p_2 - p_1^\nu p_2^\mu\right) = p_1\cdot p_2. \tag{7.16}$$

The projector annihilates all the terms proportional to p_1^μ or p_2^ν, so we never need to compute them.

The calculation is then a standard one-loop calculation. We contract the trace of the fermion loop with $P_{\mu\nu}$, combine the three top-quark propagators with Feynman parameters (eq. (3.63)), shift the loop momentum and integrate it with the formulae of chapter 3.

Individual terms have ultraviolet poles in ϵ, but they cancel in the sum; the amplitude is finite, as it must be, since there is no hgg coupling in the Lagrangian that could absorb a divergence. We find

$$\mathcal{M}_{gg\to h} = -\frac{\alpha_s}{2\pi v}\,\mathcal{G}(\tau)\,\delta^{ab}\left(p_1\cdot p_2\;\epsilon_1\cdot\epsilon_2 - p_1\cdot\epsilon_2\;p_2\cdot\epsilon_1\right), \qquad \tau \equiv \frac{4m_t^2}{m_h^2}, \tag{7.17}$$

with the Feynman-parameter representation

$$\mathcal{G}(\tau) = 2\int_0^1 dx \int_0^{1-x} dy\;\frac{1-4xy}{1-\frac{4xy}{\tau}-i\delta}. \tag{7.18}$$

The integrals are elementary and give

$$\boxed{\mathcal{G}(\tau) = \tau\left[1+(1-\tau)\,f(\tau)\right], \qquad f(\tau) = \begin{cases} \arcsin^2\frac{1}{\sqrt{\tau}}, & \tau \geq 1, \\ -\frac{1}{4}\left[\ln\frac{1+\sqrt{1-\tau}}{1-\sqrt{1-\tau}} - i\pi\right]^2, & \tau < 1. \end{cases}} \tag{7.19}$$

For $\tau < 1$, i.e. $m_h > 2m_t$, the top quarks in the loop can be on shell and the form factor develops an imaginary part, whose sign is fixed by the $i\delta$ prescription. For the physical Higgs boson, $\tau = 4m_t^2/m_h^2 \simeq 7.6$.

The limit of a heavy top quark is of particular interest. Expanding $\arcsin^2 x = x^2 + \frac{x^4}{3} + \frac{8x^6}{45} + \ldots$ we find

$$\mathcal{G}(\tau) = \frac{2}{3} + \frac{7}{45\,\tau} + O\left(\frac{1}{\tau^2}\right) \qquad (\tau \to \infty), \tag{7.20}$$

which also follows immediately from eq. (7.18) at $\tau = \infty$, $2\iint(1-4xy) = 2\left(\frac{1}{2}-\frac{1}{6}\right) = \frac{2}{3}$. The amplitude does not vanish when the top quark becomes infinitely heavy: the factor m_t of the Yukawa coupling compensates the $1/m_t$ suppression of the loop. Heavy quarks do not decouple from the Higgs boson. This also means that $gg \to h$ counts the number of heavy quarks which get their mass from the Higgs mechanism. For the physical value $\tau \simeq 7.6$ we have $\mathcal{G} \simeq 0.688$: the limiting value 2/3 is accurate to 3% in the amplitude, and to 7% in the cross-section.

Let us now compute the cross-section. Summing over the polarisations with $-g_{\mu\nu}$ (allowed here, since the tensor (7.15) is transverse) and over colours,

$$\sum|\mathcal{M}|^2 = \left|\frac{\alpha_s\mathcal{G}}{2\pi v}\right|^2 N_A\,(d-2)\,(p_1\cdot p_2)^2. \tag{7.21}$$

In d dimensions a gluon has $d-2 = 2(1-\epsilon)$ polarisations, and we average over $4(1-\epsilon)^2 N_A^2$ initial states. With the one-particle phase space $\int \frac{d^d p_h}{(2\pi)^{d-1}}\delta^+(p_h^2-m_h^2)(2\pi)^d\delta^d(p_1+p_2-p_h) = 2\pi\,\delta(s-m_h^2)$ and the flux factor $1/(2s)$ we obtain

$$\hat{\sigma}^{\mathrm{LO}}_{gg\to h} = \sigma_0\,\delta(1-z), \qquad z \equiv \frac{m_h^2}{s}, \tag{7.22}$$

with

$$\sigma_0 = \frac{G_F\,\alpha_s^2}{128\sqrt{2}\,\pi}\,|\mathcal{G}(\tau)|^2 \xrightarrow{m_t\to\infty} \frac{G_F\,\alpha_s^2}{288\sqrt{2}\,\pi} = \frac{\pi}{576\,v^2}\left(\frac{\alpha_s}{\pi}\right)^2, \tag{7.23}$$

where $G_F/\sqrt{2} = 1/(2v^2)$. In d dimensions the average over polarisations gives an extra factor $1/(1-\epsilon)$; we write $\sigma_0^{(\epsilon)} = \sigma_0/(1-\epsilon)$ for the d-dimensional Born normalisation. Folding with the gluon densities, eq. (7.3), the δ-function removes one integration:

$$\sigma^{\text{LO}}_{pp\to h} = \sigma_0 \int dx_1 dx_2\, g(x_1) g(x_2)\, \delta\!\left(1 - \frac{\tau_h}{x_1 x_2}\right) = \sigma_0\, \tau_h \int_{\tau_h}^1 \frac{dx}{x}\, g(x)\, g\!\left(\frac{\tau_h}{x}\right), \qquad \tau_h = \frac{m_h^2}{S}. \tag{7.24}$$

Remark.. At NLO in the full theory we would have to compute two-loop diagrams with massive top quarks. This has been done (Spira, Djouadi, Graudenz and Zerwas, 1995), but it is beyond the scope of these lectures. For the QCD effects which we want to understand, the top mass is not important. We will therefore work in the limit $m_h^2 \ll 4m_t^2$, where the calculation simplifies dramatically.

7.4 The heavy-top limit: an effective Lagrangian

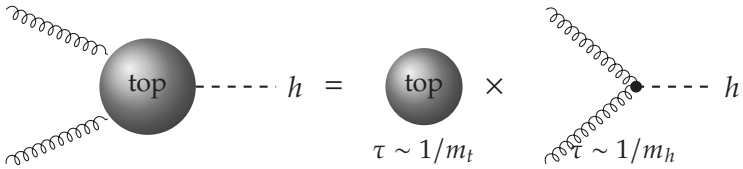


Figure 7.3. For $m_t \gg m_h$ the top loop is a short-distance process: it factorises into a coefficient and a local hgg interaction.

If $m_t \gg m_h$, the interactions inside the top loop happen very fast, within a time $\sim 1/m_t$. They are insensitive to the variations of the external gluon fields, which happen on the much longer time scale $\sim 1/m_h$ (figure 7.3). This is the same separation of scales which gave us the factorisation theorem, now applied inside the hard process. The top loop acts as a point-like vertex, and its effect can be described by a local effective Lagrangian, built from the gluon field strength and the Higgs field. The lowest-dimension gauge-invariant operator is $h\, G^a_{\mu\nu} G^{a\,\mu\nu}$, and we write (Wilczek; Shifman, Vainshtein, Voloshin, Zakharov; Chetyrkin, Kniehl, Steinhauser)

$$\boxed{\mathcal{L}_{\text{eff}} = -\frac{1}{4}\, C\, \frac{h}{v}\, G^a_{\mu\nu} G^{a\,\mu\nu}, \qquad C = -\frac{\alpha_s}{3\pi}\left[1 + \frac{11}{4}\frac{\alpha_s}{\pi} + O(\alpha_s^2)\right].} \tag{7.25}$$

Equivalently, $\mathcal{L}_{\text{eff}} = +\frac{\alpha_s}{12\pi v}\left(1 + \frac{11}{4}\frac{\alpha_s}{\pi}\right) h\, G^a_{\mu\nu} G^{a\,\mu\nu}$. The coefficient C is called a *Wilson coefficient*. It is obtained by *matching*: we require that the effective theory reproduces the amplitudes of the full theory for $m_t \to \infty$. At leading order, comparing with eqs. (7.17) and (7.20), $C = -\frac{\alpha_s}{2\pi} \cdot \frac{2}{3} = -\frac{\alpha_s}{3\pi}$, as we will verify in a moment with the Feynman rules. The correction $\frac{11}{4}\frac{\alpha_s}{\pi}$ requires two-loop diagrams with top quarks at zero external momenta (Dawson; Djouadi, Spira, Zerwas). We will take it from the literature. For a general gauge group it reads $\frac{5C_A - 3C_F}{4}\frac{\alpha_s}{\pi}$.

There is a beautiful way to understand the leading coefficient without computing any diagram: the *low-energy theorem*. A heavy quark contributes to the running of α_s above its mass, and not below it. At scales $\mu \ll m_t$ the effect of the top quark on the gluons is a finite renormalisation of the gluon kinetic term proportional to $\alpha_s \ln m_t^2$. The Higgs field enters through $m_t \to m_t(1 + h/v)$, i.e. $\ln m_t^2 \to \ln m_t^2 + 2h/v$, and the derivative of the kinetic

term with respect to $\ln m_t$ is precisely the operator $h\,GG$. You are asked to work this out in the exercises at the end of the chapter.

The effective Lagrangian (7.25) contains interactions of the Higgs boson with two, three and four gluons, since $G^a_{\mu\nu}$ contains terms quadratic in the gluon field. The three- and four-gluon terms of $G^a G^a$ are, up to the overall factor, the interaction terms of the Yang–Mills Lagrangian. The Feynman rules are therefore simple (figure 7.4). For the two-gluon vertex we expand $-\frac{C}{4v} h \cdot 2\left(\partial_\mu A^a_\nu \partial^\mu A^{a\,\nu} - \partial_\mu A^a_\nu \partial^\nu A^{a\,\mu}\right)$, replace $\partial \to -ip$ for every incoming gluon momentum and count the two ways of attaching the gluons to the fields. With all gluon momenta incoming:

$$h\, g^a_\mu(p_1)\, g^b_\nu(p_2): \qquad i\,\frac{C}{v}\,\delta^{ab}\left(g^{\mu\nu}\, p_1\cdot p_2 - p_1^\nu p_2^\mu\right), \tag{7.26}$$

$$h\, g^a_\mu(p_1)\, g^b_\nu(p_2)\, g^c_\rho(p_3): \qquad \frac{C}{v}\, V^{abc}_{\mu\nu\rho}(p_1,p_2,p_3), \tag{7.27}$$

$$h\, g^a_\mu\, g^b_\nu\, g^c_\rho\, g^d_\sigma: \qquad \frac{C}{v}\times\left[\text{four-gluon vertex, eq. (7.10)}\right], \tag{7.28}$$

with $V^{abc}_{\mu\nu\rho}$ the three-gluon vertex of eq. (7.9). Note that in the $hggg$ vertex the momenta p_1, p_2, p_3 are those of the gluons only; they do not sum to zero, since the Higgs boson carries momentum away.

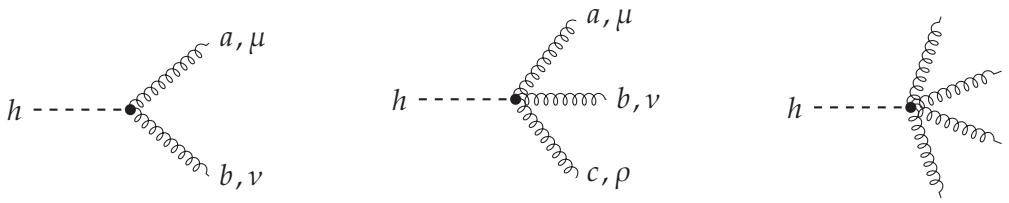


Figure 7.4. The interactions of the effective Lagrangian (7.25): the effective vertex (dot) couples the Higgs boson to two, three and four gluons. The last two are C/v times the three- and four-gluon vertices of QCD.

With the rule (7.26) the leading-order amplitude in the effective theory is

$$\mathcal{M}^{(0)} = \frac{C}{v}\,\delta^{ab}\left(p_1\cdot p_2\;\epsilon_1\cdot\epsilon_2 - p_1\cdot\epsilon_2\; p_2\cdot\epsilon_1\right), \tag{7.29}$$

which agrees with eq. (7.17) for $\mathcal{G} \to \frac{2}{3}$. Repeating the steps which led to eq. (7.23),

$$\sigma_0 = \frac{\pi\, C^2}{8 N_A\, v^2} = \frac{\pi\, C^2}{64\, v^2} = \frac{\pi}{576\, v^2}\left(\frac{\alpha_s}{\pi}\right)^2\left[1 + \frac{11}{2}\frac{\alpha_s}{\pi} + \ldots\right]. \tag{7.30}$$

A last remark before we start the NLO calculation. In practice one uses the effective theory to compute the *ratio* of the NLO to the LO cross-section and multiplies it with the exact LO cross-section (7.23), including the full $\mathcal{G}(\tau)$. Comparisons with the exact two-loop calculation show that this approximation is very good, at the level of a few per cent, for $m_h < 2m_t$: the QCD corrections are dominated by soft and collinear gluons, which cannot resolve the top loop.

7.5 The virtual corrections

At NLO the effective theory gives one-loop corrections to $gg \to h$. Let us first enumerate the diagrams. With the vertices of figure 7.4 and of QCD we can draw: (a) a triangle, in

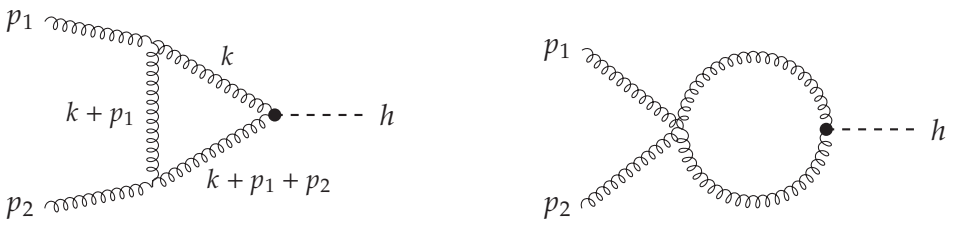


Figure 7.5. The one-loop diagrams of the effective theory which do not vanish in dimensional regularisation. Diagrams with a loop on an external gluon, or with a loop attached to a single external gluon, are scaleless.

which the two external gluons emit gluons which meet at the hgg vertex; (b) a bubble in which the two external gluons meet at a four-gluon vertex, and two gluons propagate from there to the hgg vertex; (c) bubbles in which one external gluon enters the $hggg$ vertex, while the loop is attached to the other external gluon at a three-gluon vertex; (d) a tadpole at the $hgggg$ vertex; (e) self-energy insertions on the external legs, with gluon, ghost or quark loops. The loop integrals of (c) depend only on $p_1^2 = 0$ or $p_2^2 = 0$, those of (d) and (e) on no external scale at all. They are scaleless and vanish, eq. (3.50). We are left with the two diagrams of figure 7.5. The second one has a symmetry factor $\frac{1}{2}$, for the exchange of the two identical gluon propagators.

The amplitude has again the structure of eq. (7.15), and we project it with $P_{\mu\nu}$ of eq. (7.16). The colour factors are the same for both diagrams, $f^{ace}f^{bce} = C_A\delta^{ab}$. For the triangle we have, symbolically,

$$\mathcal{M}^{(1)}_{\text{bare}} = \frac{C}{v}\,\frac{g^2\mu^{2\epsilon}C_A}{(4\pi)^{d/2}}\,\delta^{ab}\,\epsilon_{1\mu}\epsilon_{2\nu}\int\frac{d^dk}{i\pi^{d/2}}\,\frac{N^{\mu\nu}\left(k,p_1,p_2\right)}{k^2\,(k+p_1)^2\,(k+p_1+p_2)^2} + (\text{bubble}), \tag{7.31}$$

where we have used $\int\frac{d^dk}{(2\pi)^d} = \frac{i}{(4\pi)^{d/2}}\int\frac{d^dk}{i\pi^{d/2}}$ to bring the loop integration to the normalisation of chapters 3 and 4. The numerator $N^{\mu\nu}$ is the product of two three-gluon vertices and of the hgg vertex. It is a polynomial of degree four in the loop momentum. After the contraction with $P_{\mu\nu}$, only the scalar products k^2, $k\cdot p_1$ and $k\cdot p_2$ remain. This is a short, if tedious, exercise in Lorentz algebra, best done with a computer-algebra program (FORM, or the `FeynCalc` package of Mathematica).

We now introduce three Feynman parameters, with x_1 for k^2, x_2 for $(k+p_1+p_2)^2$ and x_3 for $(k+p_1)^2$. Following section 4.1, the denominator becomes $\left[\ell^2 + x_1x_2\,s\right]^3$ after a shift $k = \ell - x_3p_1 - x_2(p_1+p_2)$: the only invariant which appears is $(q_1-q_2)^2 = s$ (in the notation of eq. (4.9), with $q_1 = 0$, $q_2 = p_1+p_2$, $q_3 = p_1$), multiplied by x_1x_2. The numerator becomes a polynomial in ℓ^2, $\ell\cdot p_{1,2}$ and the x_i. Terms odd in ℓ vanish, and the even ones are reduced by symmetric integration, e.g. $\ell^\mu\ell^\nu \to g^{\mu\nu}\ell^2/d$. Integrating over ℓ with

$$\int\frac{d^d\ell}{i\pi^{d/2}}\,\frac{(\ell^2)^r}{(\ell^2-\Delta)^n} = (-1)^{n+r}\,\frac{\Gamma\left(r+\frac{d}{2}\right)\Gamma\left(n-r-\frac{d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)\Gamma(n)}\,\Delta^{\frac{d}{2}+r-n}, \tag{7.32}$$

with $\Delta = -x_1x_2\,s - i\delta$ (which generalises eq. (3.46); the $r=1$ case is an easy exercise) we obtain

$$\mathcal{M}^{(1)}_{\text{bare}} = \mathcal{M}^{(0)}\,\frac{\alpha_s^{\text{bare}}}{\pi}\,(4\pi)^\epsilon\,(-s-i\delta)^{-\epsilon}\int_0^1 dx_1dx_2dx_3\;\delta(1-x_1-x_2-x_3)\,\frac{f(x_1,x_2,x_3;\epsilon)}{(x_1x_2)^{1+\epsilon}}, \tag{7.33}$$

with f a polynomial in the Feynman parameters, whose coefficients are rational functions of ϵ (they contain, e.g., the ultraviolet pole of the rank-four tensor integral). The singular factor $(x_1 x_2)^{-1-\epsilon}$ is the origin of the infrared divergences: the region $x_1 \to 0$ and $x_2 \to 0$ corresponds to loop momenta for which the two gluons attached to the Higgs vertex are both on shell, i.e. soft or collinear to the incoming gluons.

How do we extract the poles from such an integral in general? We map the simplex to the unit square,

$$x_1 = \lambda_1, \qquad x_2 = \lambda_2(1-\lambda_1), \qquad x_3 = (1-\lambda_1)(1-\lambda_2), \tag{7.34}$$

with Jacobian $(1-\lambda_1)$, so that

$$\int_0^1 d\lambda_1 d\lambda_2\, (1-\lambda_1)^{-\epsilon}\, \frac{f(\lambda_1, \lambda_2(1-\lambda_1), (1-\lambda_1)(1-\lambda_2))}{(\lambda_1\lambda_2)^{1+\epsilon}}. \tag{7.35}$$

The singularities are now *factorised*: they sit at $\lambda_1 = 0$ and at $\lambda_2 = 0$, each in one variable. We can then use, in the sense of distributions,

$$\boxed{\lambda^{-1-\epsilon} = -\frac{\delta(\lambda)}{\epsilon} + \sum_{n=0}^{\infty} \frac{(-\epsilon)^n}{n!} \left[\frac{\ln^n \lambda}{\lambda}\right]_+ ,} \tag{7.36}$$

where the plus-distributions are defined by their action on a test function $h(\lambda)$,

$$\int_0^1 d\lambda \left[\frac{\ln^n \lambda}{\lambda}\right]_+ h(\lambda) = \int_0^1 d\lambda\, \frac{\ln^n \lambda}{\lambda}\, [h(\lambda) - h(0)]\,. \tag{7.37}$$

Eq. (7.36) follows from writing $\int_0^1 d\lambda\, \lambda^{-1-\epsilon} h(\lambda) = \int_0^1 d\lambda\, \lambda^{-1-\epsilon}[h(\lambda) - h(0)] - \frac{h(0)}{\epsilon}$ and expanding the first integral, which is now convergent, in ϵ. Two such expansions produce at most a $1/\epsilon^2$ pole; this is the double pole of one-loop QCD amplitudes. With this method one can compute numerically the ϵ expansion of very complicated loop integrals, for which no analytic result is available.

In our case, however, we can do better. Since $\Delta = -x_1 x_2 s$ is a monomial, all the Feynman parameter integrals are of Dirichlet type,

$$\int_0^1 dx_1 dx_2 dx_3\, \delta(1 - x_1 - x_2 - x_3)\, x_1^{a_1-1} x_2^{a_2-1} x_3^{a_3-1} = \frac{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)}{\Gamma(a_1+a_2+a_3)}, \tag{7.38}$$

the three-variable generalisation of the Beta function, and the amplitude can be obtained in closed form, for any d. The four-gluon bubble gives integrals of the same type with two Feynman parameters. Adding the two diagrams, all the Γ-functions combine and the result can be expressed through the massless bubble $I[1,1]$ of chapter 4, eq. (4.13), evaluated at $p^2 = s$:

$$\boxed{\mathcal{M}^{(1)}_{\text{bare}} = \mathcal{M}^{(0)}\, \frac{g^2 \mu^{2\epsilon}\, C_A}{(4\pi)^{d/2}} \left[-\frac{2\left(1 - 3\epsilon + 2\epsilon^2 + \epsilon^3\right)}{\epsilon\,(1-\epsilon)}\right] I[1,1](s),} \tag{7.39}$$

where

$$I[1,1](s) = \frac{c_\Gamma}{\epsilon(1-2\epsilon)}\, (-s - i\delta)^{-\epsilon}\,, \tag{7.40}$$

with

$$c_\Gamma \equiv \frac{\Gamma(1+\epsilon)\Gamma(1-\epsilon)^2}{\Gamma(1-2\epsilon)}. \tag{7.41}$$

The triangle integral itself is proportional to the bubble,

$$\int \frac{d^dk}{i\pi^{d/2}} \frac{1}{k^2(k+p_1)^2(k+p_1+p_2)^2} = -\frac{c_\Gamma}{\epsilon^2}(-s-i\delta)^{-1-\epsilon} = \frac{2(d-3)}{(4-d)\,s} I[1,1](s), \tag{7.42}$$

which explains why only one "master integral" appears. (This is the kind of relation which we derived with integration by parts in section 4.2.)

Writing out the bubble, eq. (7.39) becomes

$$\mathcal{M}^{(1)}_{\text{bare}} = \mathcal{M}^{(0)} \frac{g^2\mu^{2\epsilon}\,C_A}{(4\pi)^{d/2}}\, c_\Gamma\,(-s-i\delta)^{-\epsilon}\left[-\frac{2}{\epsilon^2} - \frac{2\epsilon}{(1-\epsilon)(1-2\epsilon)}\right]. \tag{7.43}$$

This is a remarkably simple result: the one-loop correction has a double pole, but no single pole and no finite term in the bracket.

7.5.1 Ultraviolet renormalisation

The result of eq. (7.43) has the form

$$\mathcal{M}^{(1)}_{\text{bare}} \propto \frac{\mathcal{A}_2}{\epsilon^2} + \frac{\mathcal{A}_1}{\epsilon} + \text{finite} + \mathcal{O}(\epsilon), \tag{7.44}$$

and we must renormalise. We express the bare coupling through the $\overline{\text{MS}}$ coupling with eq. (7.13). In the one-loop term this only changes the normalisation, $\frac{g^2\mu^{2\epsilon}}{(4\pi)^{d/2}} \to \frac{\alpha_s(\mu)}{4\pi} e^{\gamma\epsilon}\mu^{2\epsilon}$; the counterterm β_0/ϵ is of higher order there. The Wilson coefficient C, however, is itself proportional to the coupling. Expressing $C \propto \alpha_s^{\text{bare}}$ through $\alpha_s(\mu)$ multiplies the leading-order amplitude by the factor $\left[1 - \frac{\alpha_s}{\pi}\frac{\beta_0}{\epsilon}\right]$, and this is the ultraviolet counterterm of our process. In the language of the effective theory, $G^a_{\mu\nu}G^{a\,\mu\nu}$ is a composite operator which needs its own renormalisation constant, $Z = 1 - \frac{\alpha_s}{\pi}\frac{\beta_0}{\epsilon} + \ldots$, while the combination $\alpha_s\, G^a_{\mu\nu}G^{a\,\mu\nu}$ is not renormalised at this order. The two bookkeepings are equivalent; we must only make sure that we apply the factor once.

The renormalised amplitude is then

$$\mathcal{M} = \mathcal{M}^{(0)}\left[1 + \frac{\alpha_s(\mu)}{\pi}\left(M^{(1)} + \frac{11}{4}\right) + \mathcal{O}(\alpha_s^2)\right], \tag{7.45}$$

where $\mathcal{M}^{(0)}$ contains C at leading order, the $\frac{11}{4}$ is the correction to the Wilson coefficient in eq. (7.25), and

$$M^{(1)} = \frac{C_A}{4} e^{\gamma\epsilon} c_\Gamma \left(\frac{\mu^2}{-s-i\delta}\right)^\epsilon \left[-\frac{2}{\epsilon^2} - \frac{2\epsilon}{(1-\epsilon)(1-2\epsilon)}\right] - \frac{\beta_0}{\epsilon}. \tag{7.46}$$

With the expansion (3.38) of the Γ-function, $e^{\gamma\epsilon}c_\Gamma = 1 - \frac{\pi^2}{12}\epsilon^2 + \mathcal{O}(\epsilon^3)$, and we find

$$\boxed{M^{(1)} = \left(\frac{\mu^2}{-s-i\delta}\right)^\epsilon \left[-\frac{C_A}{2\epsilon^2} + \frac{C_A\pi^2}{24} + \mathcal{O}(\epsilon)\right] - \frac{\beta_0}{\epsilon}.} \tag{7.47}$$

For the physical process $s = m_h^2 > 0$ and $(-s-i\delta)^{-\epsilon} = s^{-\epsilon}e^{i\pi\epsilon}$. The phase produces an imaginary part, $-i\pi C_A/(2\epsilon)$, which drops out of the interference with the (real) Born

amplitude, and an extra $\frac{C_A\pi^2}{4}$ in the real part. The virtual correction to the cross-section is $\hat{\sigma}^V = \sigma_0^{(\epsilon)}\delta(1-z)\,2\mathrm{Re}\left[\frac{\alpha_s}{\pi}\left(M^{(1)}+\frac{11}{4}\right)\right]$:

$$\boxed{\hat{\sigma}^V_{gg\to h} = \sigma_0^{(\epsilon)}\,\frac{\alpha_s}{\pi}\,\delta(1-z)\left\{\left(\frac{\mu^2}{m_h^2}\right)^{\epsilon}\left[-\frac{C_A}{\epsilon^2}+\frac{7\pi^2}{12}C_A\right]-\frac{2\beta_0}{\epsilon}+\frac{11}{2}+O(\epsilon)\right\}.} \tag{7.48}$$

The divergences have not disappeared after ultraviolet renormalisation! We are left with *infrared* singularities, and the virtual cross-section by itself is meaningless. We need to understand where they come from, and what we should add to cancel them.

7.6 The structure of infrared divergences

Ultraviolet divergences are *universal*: they redefine the fields, the couplings and the masses of the Lagrangian in exactly the same way for all processes. Infrared singularities are also universal, and their form is known in advance. They originate from loop momenta for which internal massless particles are soft or collinear to an external leg; in this limit the loop does not resolve the details of the hard process, and the singular part depends only on the colour charges and the momenta of the external partons (Giele, Glover; Kunszt, Soper; Kunszt, Signer, Trócsányi; Catani, Seymour). Without doing any calculation we know that

$$M^{(1)}\,\mathcal{M}^{(0)} = \frac{1}{2}\,\mathbf{I}^{(1)}(\epsilon)\,\mathcal{M}^{(0)} + \mathcal{M}^{(1)}_{\text{finite}}. \tag{7.49}$$

The operator $\mathbf{I}^{(1)}$ was given by Catani. In the normalisation in which amplitudes are expanded in $\alpha_s/(2\pi)$ (hence the factor $\frac{1}{2}$ above, since we expand in α_s/π), it reads

$$\mathbf{I}^{(1)}(\epsilon) = \frac{1}{2}\,\frac{e^{\gamma\epsilon}}{\Gamma(1-\epsilon)}\sum_i\left(\frac{\mathbf{T}_i^2}{\epsilon^2}+\frac{\gamma_i}{\epsilon}\right)\sum_{j\neq i}\frac{\mathbf{T}_i\cdot\mathbf{T}_j}{\mathbf{T}_i^2}\left(\frac{\mu^2}{-s_{ij}-i\delta}\right)^{\epsilon}, \tag{7.50}$$

where the sums run over the external partons, $s_{ij} = 2p_i\cdot p_j$ with all momenta counted as outgoing (so that $s_{ij}>0$ when i,j are both incoming or both outgoing), and

$$\mathbf{T}_q^2 = C_F, \qquad \mathbf{T}_g^2 = C_A, \qquad \gamma_q = \frac{3}{2}C_F, \qquad \gamma_g = \frac{11}{6}C_A - \frac{2}{3}T_R n_f = 2\beta_0. \tag{7.51}$$

The colour operator $\mathbf{T}_i$ acts on the colour index of parton i; it changes the colour of the leading-order amplitude. Colour conservation, $\sum_i \mathbf{T}_i\mathcal{M}^{(0)} = 0$, gives for our two gluons $\mathbf{T}_1\cdot\mathbf{T}_2 = -\mathbf{T}_1^2 = -C_A$, and therefore

$$\frac{1}{2}\,\mathbf{I}^{(1)}(\epsilon) = -\frac{e^{\gamma\epsilon}}{\Gamma(1-\epsilon)}\left(\frac{C_A}{2\epsilon^2}+\frac{\beta_0}{\epsilon}\right)\left(\frac{\mu^2}{-s-i\delta}\right)^{\epsilon} = -\frac{C_A}{2\epsilon^2}-\frac{\beta_0}{\epsilon}+O\left(\frac{1}{\epsilon}\ln\frac{\mu^2}{-s}\right). \tag{7.52}$$

These are exactly the poles of eq. (7.47). The double pole comes from gluons which are simultaneously soft and collinear. The single pole γ_g/ϵ is collinear; note that it contains n_f, although our bare amplitude did not contain any quark loop. The n_f dependence enters through the ultraviolet counterterm: the scaleless quark loops on the external gluons contain $\frac{1}{\epsilon_{\text{UV}}}-\frac{1}{\epsilon_{\text{IR}}}$, and after renormalisation only the infrared part remains.

Universality holds also at higher loops (Catani; Sterman, Tejeda-Yeomans; and later, to all orders, Aybat, Dixon, Sterman; Becher, Neubert; Gardi, Magnea). At two loops, in the $\alpha_s/(2\pi)$ normalisation,

$$\mathcal{M}^{(2)} = \mathbf{I}^{(1)}\mathcal{M}^{(1)} + \mathbf{I}^{(2)}\mathcal{M}^{(0)} + \text{finite}, \tag{7.53}$$

with

$$\mathbf{I}^{(2)}(\epsilon) = -\frac{1}{2}\mathbf{I}^{(1)}(\epsilon)\left(\mathbf{I}^{(1)}(\epsilon) + \frac{2b_0}{\epsilon}\right) + \frac{e^{\gamma\epsilon}\,\Gamma(1-2\epsilon)}{\Gamma(1-\epsilon)}\left(\frac{b_0}{\epsilon} + K\right)\mathbf{I}^{(1)}(2\epsilon) + \mathbf{H}^{(2)}(\epsilon), \tag{7.54}$$

where $b_0 = 2\beta_0$, $K = \left(\frac{67}{18} - \frac{\pi^2}{6}\right)C_A - \frac{10}{9}T_R n_f$, and $\mathbf{H}^{(2)}$ contains only a single pole. The general picture is that an amplitude factorises into jet functions for every external leg, a soft function, and a finite hard function,

$$\mathcal{M}_{\text{all orders}} = \left(\prod_{\text{legs}} J^{\text{leg}}(\alpha_s, \epsilon)\right)\mathbf{S}(\alpha_s, \epsilon)\, H(\alpha_s), \tag{7.55}$$

where all the infrared poles are contained in J and $\mathbf{S}$. Such results are extremely useful in practice: they are a strong check of every loop calculation.

Loop divergences are universal and factorisable. But how do they cancel? The answer is given by the Kinoshita–Lee–Nauenberg theorem: processes which "look the same" in the infrared limit must be added together. A final state h cannot be distinguished from a final state h plus a gluon of zero energy, or h plus a gluon which is exactly collinear to an incoming gluon. We must therefore add the real emission process $gg \to h + g$.

7.7 Real emission: $gg \to h + g$

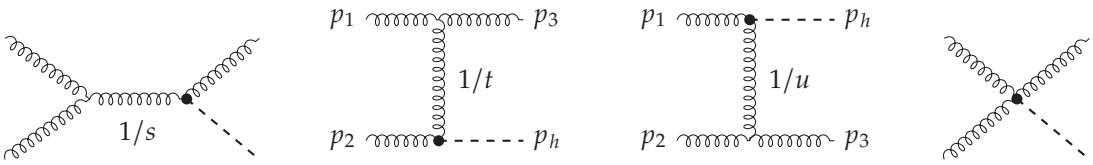


Figure 7.6. Tree diagrams for $g(p_1)g(p_2) \to h(p_h) + g(p_3)$ in the effective theory: s-, t- and u-channel gluon exchange and the $hggg$ contact vertex.

The tree-level amplitude for

$$g(p_1) + g(p_2) \to h(p_h) + g(p_3), \tag{7.56}$$

with

$$s = (p_1 + p_2)^2, \qquad t = (p_1 - p_3)^2, \qquad u = (p_2 - p_3)^2, \qquad s + t + u = m_h^2, \tag{7.57}$$

is given by the four diagrams of figure 7.6. Each of the first three contains a three-gluon vertex, a gluon propagator and the hgg vertex; the fourth is the $hggg$ contact interaction. All four are proportional to the same colour factor f^{abc} (for the gluons a, b, c), so the colour sum gives $f^{abc}f^{abc} = C_A N_A$. A good check of the relative signs of the four diagrams is gauge invariance: the amplitude must vanish when any of the polarisation vectors is replaced by the momentum of its gluon, provided the other two gluons are physical. Each diagram alone fails this test; their sum passes it.

To sum over the polarisations in d dimensions we cannot use $-g_{\mu\nu}$ for all three gluons without adding ghosts in the final state. Instead we use the sum over the $d-2$ physical polarisations,

$$\sum_{\text{pol}} \epsilon^{\mu}(p)\epsilon^{*\nu}(p) = -g^{\mu\nu} + \frac{p^{\mu}n^{\nu} + n^{\mu}p^{\nu}}{p \cdot n}, \tag{7.58}$$

with an auxiliary light-like vector n, $n \cdot p \neq 0$ (for example $n = p_2$ for gluon 1, and $n = p_1$ for gluons 2 and 3). After some algebra (again best done with a computer) we find

$$\boxed{\sum |\mathcal{M}_{gg\to hg}|^2 = g^2\mu^{2\epsilon}\left(\frac{C}{v}\right)^2 \frac{C_A\, N_A}{s\,t\,u}\Big[(1-2\epsilon)\left(m_h^8 + s^4 + t^4 + u^4\right) + \frac{\epsilon}{2}\left(m_h^4 + s^2 + t^2 + u^2\right)^2\Big].} \tag{7.59}$$

In four dimensions this is the classic result $\propto (m_h^8 + s^4 + t^4 + u^4)/(stu)$; note that $stu > 0$ in the physical region, where $t, u < 0$.

7.7.1 Where is the matrix element singular?

Let us work in the partonic centre-of-mass frame (figure 7.7). The incoming gluons have energies $\sqrt{s}/2$, and the emitted gluon has energy and scattering angle

$$E_3 = \frac{\sqrt{s}}{2}(1-z), \qquad \theta \equiv \angle(\vec{p}_1, \vec{p}_3), \qquad z = \frac{m_h^2}{s}. \tag{7.60}$$

The invariants $s_{13} \equiv 2p_1 \cdot p_3 = -t$ and $s_{23} \equiv 2p_2 \cdot p_3 = -u$ are

$$s_{13} = 2E_1E_3(1-\cos\theta) = \frac{s}{2}(1-z)(1-\cos\theta), \qquad s_{23} = \frac{s}{2}(1-z)(1+\cos\theta). \tag{7.61}$$

The matrix element (7.59) is singular when $t \to 0$ or $u \to 0$, i.e.

$$s_{13} = 0 \;\Leftrightarrow\; E_3 = 0 \text{ or } \theta = 0, \qquad s_{23} = 0 \;\Leftrightarrow\; E_3 = 0 \text{ or } \theta = \pi. \tag{7.62}$$

These are the *soft* ($E_3 \to 0$, i.e. $z \to 1$) and the *collinear* ($\theta \to 0, \pi$) singularities. The double singularity, soft *and* collinear, is the counterpart of the $1/\epsilon^2$ of the virtual amplitude. The s-channel pole is harmless: $s \geq m_h^2$.

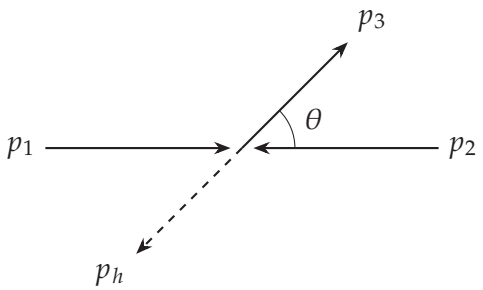


Figure 7.7. Kinematics of $gg \to hg$ in the partonic centre-of-mass frame. The angle θ is measured from the direction of $\vec{p}_1$.

7.7.2 Phase space in d dimensions

The singularities are integrable in $d < 4$ dimensions, and we must perform the phase-space integration in d dimensions, exactly as we did for the loop momentum. For any observable O we need

$$\int d\Phi_2 \, \frac{F(s_{13}, s_{23})\, O(s_{13}, s_{23})}{s_{13}\, s_{23}}, \tag{7.63}$$

where

$$d\Phi_2 = \frac{d^d p_h}{(2\pi)^{d-1}}\,\delta^+(p_h^2 - m_h^2)\,\frac{d^d p_3}{(2\pi)^{d-1}}\,\delta^+(p_3^2)\,(2\pi)^d\delta^d(p_1 + p_2 - p_h - p_3), \tag{7.64}$$

and F is a regular function. We use the δ^d to integrate p_h, write

$$d^{d-1}\vec{p}_3 = |\vec{p}_3|^{d-2} d|\vec{p}_3|\,\sin^{d-3}\theta\,d\theta\,d\Omega_{d-2}, \tag{7.65}$$

and use the remaining δ-function to fix $|\vec{p}_3| = E_3$. The angular integration over the $(d-3)$-sphere transverse to the beam gives $\Omega_{d-2} = 2\pi^{(d-2)/2}/\Gamma\left(\frac{d-2}{2}\right)$, from eq. (3.26). The result is

$$\int d\Phi_2 = \frac{1}{8\pi}\,\frac{1}{\Gamma(1-\epsilon)}\left(\frac{4\pi}{s}\right)^{\epsilon}(1-z)^{1-2\epsilon}\int_0^\pi d\theta\,\frac{\sin^{1-2\epsilon}\theta}{2^{1-2\epsilon}}\;\ldots \tag{7.66}$$

The natural variable is $\cos\theta = 1 - 2\lambda$, for which

$$s_{13} = s(1-z)\lambda, \qquad s_{23} = s(1-z)(1-\lambda), \tag{7.67}$$

and

$$\boxed{\int d\Phi_2 = \frac{1}{8\pi}\,\frac{1}{\Gamma(1-\epsilon)}\left(\frac{4\pi}{s}\right)^{\epsilon}(1-z)^{1-2\epsilon}\int_0^1 d\lambda\;[\lambda(1-\lambda)]^{-\epsilon}\,.} \tag{7.68}$$

Our integral has therefore the structure

$$\int d\Phi_2\,\frac{F\,O}{s_{13}s_{23}} \propto (1-z)^{-1-2\epsilon}\int_0^1 d\lambda\,\left[\lambda^{-1-\epsilon}(1-\lambda)^{-\epsilon} + (1-\lambda)^{-1-\epsilon}\lambda^{-\epsilon}\right]F\,O, \tag{7.69}$$

where we have used $\frac{1}{\lambda(1-\lambda)} = \frac{1}{\lambda} + \frac{1}{1-\lambda}$. The singularities are again factorised, in λ (collinear) and in $1-z$ (soft), and we can apply the expansion eq. (7.36) to $\lambda^{-1-\epsilon}$, $(1-\lambda)^{-1-\epsilon}$ and to

$$(1-z)^{-1-2\epsilon} = -\frac{\delta(1-z)}{2\epsilon} + \left[\frac{1}{1-z}\right]_+ - 2\epsilon\left[\frac{\ln(1-z)}{1-z}\right]_+ + O(\epsilon^2). \tag{7.70}$$

We learn an important lesson: a pole in ϵ develops only from the points $\lambda = 0$, $\lambda = 1$ or $z = 1$ *exactly*. Everywhere else the integrand is finite and can be computed in four dimensions.

7.8 Infrared-safe observables and the cancellation of soft poles

The function $O(s_{13}, s_{23})$ in the phase-space integral defines our measurement. For a cross-section with cuts it is 0 or 1: it decides whether an "event" $gg \to hg$ is included in our measurement or not. We hope that by *adding* the cross-section for $gg \to hg$ to the cross-section for $gg \to h$, which behaves as $\left(-\frac{C_A}{\epsilon^2} + \ldots\right)\delta(1-z)$, we obtain a finite result. For this, the observable must include ($O = 1$) the phase-space points where the real matrix element develops poles, in exactly the same way as it includes the corresponding virtual configuration. In other words, we should measure only *infrared-safe* quantities,

$$O\left(gg \to h + \text{soft or collinear gluon}\right) = O(gg \to h). \tag{7.71}$$

The total cross-section, $O = 1$, is the simplest example, and we compute it now. Other examples are the rapidity distribution of the Higgs boson, or the cross-section with a veto

on jets above some transverse momentum. The number of gluons in the final state, on the other hand, is not infrared safe: a soft gluon cannot be counted.

We insert eqs. (7.59) and (7.68) into $\hat{\sigma}^R = \frac{1}{2s}\frac{1}{4(1-\epsilon)^2 N_A^2}\int d\Phi_2 \sum |\mathcal{M}|^2$, express the coupling through $\alpha_s(\mu)$ with eq. (7.13), and divide by the d-dimensional Born normalisation $\sigma_0^{(\epsilon)}$. With

$$t = -s(1-z)\lambda, \qquad u = -s(1-z)(1-\lambda), \qquad m_h^2 = z\,s, \qquad stu = s^3(1-z)^2\lambda(1-\lambda), \tag{7.72}$$

we find

$$\hat{\sigma}^R_{gg\to hg} = \sigma_0^{(\epsilon)}\,\frac{\alpha_s}{\pi}\,\frac{C_A}{2}\,\frac{e^{\gamma\epsilon}}{(1-\epsilon)\Gamma(1-\epsilon)}\left(\frac{\mu^2}{m_h^2}\right)^{\epsilon} z^{\epsilon}\,(1-z)^{-1-2\epsilon} \times \int_0^1 d\lambda\;[\lambda(1-\lambda)]^{-1-\epsilon}\;f(z,\lambda;\epsilon), \tag{7.73}$$

with

$$f(z,\lambda;\epsilon) = (1-2\epsilon)\left[1+z^4+(1-z)^4\left(\lambda^4+(1-\lambda)^4\right)\right] + \frac{\epsilon}{2}\left[1+z^2+(1-z)^2\left(\lambda^2+(1-\lambda)^2\right)\right]^2. \tag{7.74}$$

The λ-integrals are Beta functions, $\int_0^1 d\lambda\,\lambda^{n-1-\epsilon}(1-\lambda)^{m-1-\epsilon} = \frac{\Gamma(n-\epsilon)\Gamma(m-\epsilon)}{\Gamma(n+m-2\epsilon)}$. Two cases are instructive.

- *Collinear limit.* The endpoints $\lambda \to 0, 1$ give the pole

$$\int_0^1 d\lambda\;[\lambda(1-\lambda)]^{-1-\epsilon}\,f(z,\lambda;\epsilon) = -\frac{2}{\epsilon}\,f(z,0;0) + \mathcal{O}(\epsilon^0) = -\frac{4}{\epsilon}\left(1-z+z^2\right)^2 + \mathcal{O}(\epsilon^0). \tag{7.75}$$

- *Soft limit.* At $z = 1$ the function $f = 2(1-\epsilon)$ does not depend on λ, and

$$\int_0^1 d\lambda\;[\lambda(1-\lambda)]^{-1-\epsilon}\,f(1,\lambda;\epsilon) = -\frac{4(1-\epsilon)}{\epsilon}\,\frac{\Gamma(1-\epsilon)^2}{\Gamma(1-2\epsilon)}. \tag{7.76}$$

Expanding $(1-z)^{-1-2\epsilon}$ with eq. (7.70) and collecting terms, we obtain for $\mu = m_h$

$$\hat{\sigma}^R_{gg\to hg} = \sigma_0^{(\epsilon)}\frac{\alpha_s}{\pi}\left\{C_A\left[\frac{1}{\epsilon^2}-\frac{\pi^2}{4}\right]\delta(1-z) - \frac{2}{\epsilon}\,C_A\left(\left[\frac{1}{1-z}\right]_+ - z(2-z+z^2)\right) + 4C_A\left[\frac{\ln(1-z)}{1-z}\right]_+ - 4C_A\,z(2-z+z^2)\ln(1-z) - 2C_A\frac{(1-z+z^2)^2}{1-z}\ln z - \frac{11}{6}C_A(1-z)^3\right\} + \mathcal{O}(\epsilon). \tag{7.77}$$

The $\delta(1-z)$ term comes from the soft region: the soft limit of the integrand, multiplied by $-\delta(1-z)/(2\epsilon)$, times the collinear $1/\epsilon$. The double pole $+C_A/\epsilon^2$ cancels exactly the double pole of the virtual correction (7.48). This is the Bloch–Nordsieck cancellation of soft divergences, as in QED.

Adding the virtual and the real contributions we find

$$\hat{\sigma}^V + \hat{\sigma}^R = \text{finite} - \frac{2}{\epsilon}\frac{\alpha_s}{\pi}\sigma_0^{(\epsilon)}\, z\, P_{gg}(z), \tag{7.78}$$

where we have combined the remaining poles, $-\frac{2\beta_0}{\epsilon}\delta(1-z)$ from the ultraviolet counterterm and the collinear pole of eq. (7.77), into the universal *gluon splitting function* (Altarelli, Parisi)

$$\boxed{P_{gg}(z) = C_A\left(\left[\frac{1}{1-z}\right]_+ + \frac{1}{z} - 2 + z(1-z)\right) + \beta_0\,\delta(1-z),} \tag{7.79}$$

or

$$z\,P_{gg}(z) = C_A\left(\left[\frac{1}{1-z}\right]_+ - z(2-z+z^2)\right) + \beta_0\,\delta(1-z). \tag{7.80}$$

(We used $z\left[\frac{1}{1-z}\right]_+ = \left[\frac{1}{1-z}\right]_+ - 1$.) The splitting function is written here in the normalisation of an expansion in α_s/π; it is half of the more common one, defined with $\alpha_s/(2\pi)$.

The soft poles have cancelled. A collinear pole, however, remains. It is proportional to the leading-order cross-section and to a universal function: it does not know anything about the Higgs boson. It describes gluons which emit collinear radiation *before* the hard scattering, and the appearance of z in eq. (7.78) is easy to understand: after the collinear emission the gluon which enters the hard process carries a fraction z of the incoming momentum, and the Born cross-section is evaluated at the reduced energy.

7.9 Collinear factorisation and the parton densities

Is the remaining pole a disaster? No. Collinear radiation from the initial state before the hard scattering changes the momentum distribution of the gluons in the proton. This is a long-distance effect, of exactly the type that we cannot compute in perturbation theory and that we have agreed to include in the parton densities. The pole tells us that the parton densities which appear in the factorisation theorem are themselves divergent, "bare", quantities, and must be renormalised, in the same way as we renormalise the coupling. The renormalised parton densities are

$$f_i(x,\mu_F) = \sum_j \int_x^1 \frac{dy}{y}\left[\delta_{ij}\,\delta(1-y) - \frac{\alpha_s(\mu_F)}{\pi}\frac{P_{ij}(y)}{\epsilon} + O(\alpha_s^2)\right] f_j^{\text{bare}}\left(\frac{x}{y}\right), \tag{7.81}$$

where, as for the coupling, the counterterm is a pure pole in the $\overline{\text{MS}}$ scheme, and μ_F is the *factorisation scale*. Inverting this relation to $O(\alpha_s)$ and inserting it into the factorisation formula (7.3), written with bare densities and bare partonic cross-sections, we obtain the finite partonic cross-sections

$$\hat{\sigma}_{ij}(z) = \hat{\sigma}_{ij}^{\text{bare}}(z) + \frac{\alpha_s}{\pi}\frac{1}{\epsilon}\sum_k \int_z^1 dy\left[P_{ki}(y)\,\hat{\sigma}_{kj}^{(0)}\left(\frac{z}{y}\right) + P_{kj}(y)\,\hat{\sigma}_{ik}^{(0)}\left(\frac{z}{y}\right)\right] + O(\alpha_s^2). \tag{7.82}$$

(Check the measure dy: the change of variables $x_1 \to x_1/y$ in eq. (7.3) cancels the $1/y$ of eq. (7.81).) For $gg \to h$ only $\hat{\sigma}_{gg}^{(0)} = \sigma_0\delta(1-z)$ is non-zero at leading order, and the counterterm is

$$2\,\frac{\alpha_s}{\pi}\frac{1}{\epsilon}\int_z^1 dy\, P_{gg}(y)\,\sigma_0^{(\epsilon)}\,\delta\left(1-\frac{z}{y}\right) = +\frac{2}{\epsilon}\frac{\alpha_s}{\pi}\sigma_0^{(\epsilon)}\, z\, P_{gg}(z), \tag{7.83}$$

which cancels exactly the pole of eq. (7.78). This is not a coincidence. The universal structure of the collinear singularities, which we saw in the virtual amplitude through γ_g of Catani's formula and in the real emission through the splitting function, is what makes the redefinition of the parton densities process independent: the same $f_i(x, \mu_F)$, measured in deep-inelastic scattering, can be used for Higgs production at the LHC.

The dependence of the renormalised densities on μ_F follows from the μ_F-independence of the bare ones. It is the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equation,

$$\mu_F^2 \frac{\partial f_i(x,\mu_F)}{\partial \mu_F^2} = \frac{\alpha_s}{\pi} \sum_j \int_x^1 \frac{dy}{y} P_{ij}(y)\, f_j\!\left(\frac{x}{y}, \mu_F\right), \tag{7.84}$$

the analogue for parton densities of the running of the coupling, eq. (7.11). Note that the $\beta_0 \delta(1-y)$ term of P_{gg} guarantees momentum conservation, $\sum_i \int_0^1 dx\, x\, f_i(x) = 1$, once all splitting functions are included.

7.10 The NLO cross-section

Collecting the virtual correction (7.48), the real emission (7.77) and the collinear counterterm, all poles cancel and we can set $\epsilon = 0$. The $\delta(1-z)$ terms add up to

$$\left[\frac{7\pi^2}{12} C_A + \frac{11}{2}\right] + \left[-\frac{\pi^2}{4} C_A\right] = \frac{11}{2} + \frac{\pi^2}{3} C_A = \frac{11}{2} + \pi^2. \tag{7.85}$$

For general scales we use two facts. First, $\sigma_0 \propto \alpha_s^2$, and if we express $\alpha_s(m_h)$ through $\alpha_s(\mu_R)$ with eq. (7.12), we generate the term $+2\beta_0 \ln \frac{\mu_R^2}{m_h^2}\, \delta(1-z)$ at order α_s. Second, the parton densities at μ_F differ from those at m_h by eq. (7.84), and the partonic cross-section must compensate this difference. The final result for the gluon-gluon channel is (Dawson; Djouadi, Spira, Zerwas)

$$\begin{aligned}\hat{\sigma}_{gg\to h+X}(z) = \frac{\pi}{576\, v^2} \left(\frac{\alpha_s(\mu_R)}{\pi}\right)^2 \Bigg[\delta(1-z) + \frac{\alpha_s(\mu_R)}{\pi} \Bigg\{ \left(\frac{11}{2} + \pi^2\right) \delta(1-z) \\ + 12 \left[\frac{\ln(1-z)}{1-z}\right]_+ - 12z\left(2 - z + z^2\right) \ln(1-z) - 6\, \frac{\left(1 - z + z^2\right)^2}{1-z} \ln z \\ - \frac{11}{2}(1-z)^3 + 2\beta_0 \ln \frac{\mu_R^2}{m_h^2}\, \delta(1-z) - 2\, z\, P_{gg}(z) \ln \frac{\mu_F^2}{m_h^2} \Bigg\} \Bigg],\end{aligned} \tag{7.86}$$

with $z = m_h^2/\hat{s}$ and P_{gg} of eq. (7.79) for $C_A = 3$. The scale-independent terms reproduce the $C_A = 3$ value of eq. (7.77): $4C_A = 12$, $2C_A = 6$ and $\frac{11}{6} C_A = \frac{11}{2}$. The $\frac{11}{2}$ in the $\delta(1-z)$ term, on the other hand, is the Wilson coefficient. For $\mu_F = \mu_R$ the two $\beta_0\, \delta(1-z)$ terms cancel.

At the same order new partonic channels open, through the diagrams of figure 7.6 with two gluons replaced by a quark line: $qg \to hq$ (and $\bar{q}g \to h\bar{q}$) and $q\bar{q} \to hg$. The quark-gluon channel has a collinear singularity, $q \to q + g^*$, which is removed by the counterterm with $P_{gq}(z) = \frac{C_F}{2} \frac{1+(1-z)^2}{z}$ in eq. (7.82). The quark-antiquark channel is finite.

The results are

$$\hat{\sigma}_{qg\to h+X}(z) = \sigma_0 \frac{\alpha_s}{\pi}\left\{\frac{2}{3}\left[1+(1-z)^2\right]\left[2\ln(1-z)-\ln z\right]-1+2z-\frac{z^2}{3} - z\,P_{gq}(z)\ln\frac{\mu_F^2}{m_h^2}\right\}, \tag{7.87}$$

$$\hat{\sigma}_{q\bar{q}\to h+X}(z) = \sigma_0 \frac{\alpha_s}{\pi}\frac{32}{27}(1-z)^3. \tag{7.88}$$

They are much smaller than the gluon-gluon contribution, because of the small quark densities at the relevant values of x, and because they lack the large soft terms.

7.10.1 Factorisation and renormalisation scales

The final result depends on two unphysical scales. The renormalisation scale μ_R enters through the strong coupling, the factorisation scale μ_F through the parton densities. Their choice is "arbitrary": the exact hadronic cross-section does not depend on them, and indeed the explicit logarithms in eq. (7.86) are such that

$$\frac{d\sigma}{d\ln\mu_R^2} = O(\alpha_s^4), \qquad \frac{d\sigma}{d\ln\mu_F^2} = O(\alpha_s^4), \tag{7.89}$$

while the LO cross-section varies at $O(\alpha_s^3)$. A truncated perturbative series, however, does depend on the scales, and the size of this dependence is commonly used to estimate the missing higher orders. The natural choice is $\mu_R \sim \mu_F \sim m_h$, the only hard scale of the process, and one varies them by a factor of two around it.

Let us estimate the scale dependence at LO. From $\sigma_{\text{LO}} \propto \alpha_s^2(\mu_R)$ and eq. (7.12),

$$\frac{\delta\sigma_{\text{LO}}}{\sigma_{\text{LO}}} \simeq -2\beta_0\frac{\alpha_s}{\pi}\ln\frac{\mu_R^2}{m_h^2} \simeq \mp 0.19 \qquad \text{for} \quad \mu_R = 2m_h,\ m_h/2, \tag{7.90}$$

with $\beta_0 = \frac{23}{12}$ and $\alpha_s/\pi \simeq 0.036$. A leading-order prediction for $gg \to h$ is therefore uncertain by about ±20% from μ_R alone.

7.11 The numerical picture

How large is the correction? Already the $\delta(1-z)$ term is large: $\frac{\alpha_s}{\pi}\left(\frac{11}{2}+\pi^2\right) \simeq 0.036\times 15.4 \simeq 0.55$. The π^2 comes from the analytic continuation $(-s)^{-\epsilon} \to s^{-\epsilon}e^{i\pi\epsilon}$ of the virtual amplitude, and the $\frac{11}{2}$ from the Wilson coefficient. On top of this, the real emission away from threshold gives a further large positive contribution, mainly through the terms $-12z(2-z+z^2)\ln(1-z)$ and $-6\frac{(1-z+z^2)^2}{1-z}\ln z$, which are positive for all $0<z<1$. (In the form (7.86), with the gluon luminosity at LHC energies, the plus-distribution $12\left[\frac{\ln(1-z)}{1-z}\right]_+$ by itself contributes with a moderate negative sign; how the correction is split between "soft" and "hard" terms depends on how the flux factor $1/z$ is distributed, and is not a physical statement.) Folding with the parton densities, the ratio of the NLO to the LO cross-section at the LHC is

$$K = \frac{\sigma_{\text{NLO}}}{\sigma_{\text{LO}}} \simeq 1.7-2, \tag{7.91}$$

a correction of order 100%! The two bands in figure 7.8 do not even overlap: the LO scale variation badly underestimates the size of the missing higher orders. The scale dependence

itself is only mildly reduced at NLO (the NLO band is about as wide as the LO one); a genuine reduction requires the NNLO and N^{3}LO corrections discussed below.

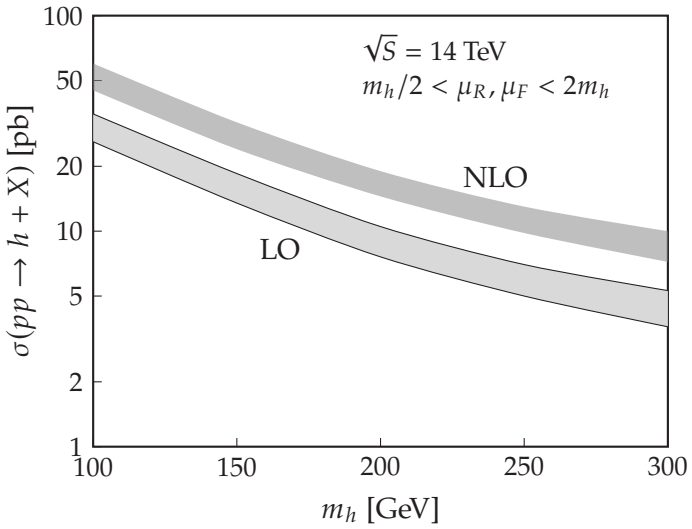


Figure 7.8. Sketch of the LO (light) and NLO (dark) cross-sections for $pp \to h + X$ in gluon fusion at $\sqrt{S} = 14$ TeV, as a function of the Higgs mass; the bands show the variation of μ_R and μ_F between $m_h/2$ and $2m_h$. The band edges are schematic (MRST2001 parton densities, heavy-top limit normalised to the exact LO); the sketch is meant to show the size of the K-factor and of the bands, not precise values.

A correction of this size raises the question of whether perturbation theory converges at all for this process. The answer came with the next-to-next-to-leading order (NNLO) calculations (Harlander, Kilgore; Anastasiou, Melnikov; Ravindran, Smith, van Neerven, 2002–2003), which found a further increase of about 15–25% and a reduced scale dependence, and with the N^{3}LO calculation (Anastasiou, Duhr, Dulat, Herzog, Mistlberger, 2015), which found a correction of a few per cent and a residual scale uncertainty of a few per cent. The series converges, slowly. The NLO calculation of this chapter is the first step of this programme, and contains all its conceptual ingredients.

7.12 Summary

- The factorisation theorem and perturbation theory are the framework for quantitative predictions at hadron colliders.
- Observables must be *infrared safe*.
- Infrared singularities are *universal*. Ultraviolet singularities renormalise α_s (and masses). *Soft* singularities *cancel* between virtual and real corrections. *Collinear* singularities of the initial state are factorised *into* the parton densities.
- Cross-sections at fixed order depend on "arbitrary" renormalisation and factorisation *scales*; the dependence is reduced at higher orders.
- QCD effects are large in our example: $K \simeq 2$ for Higgs production in gluon fusion, and not only there.

Exercises

Exercise.. (*Low-energy theorem.*) Rescale the gluon field, $A \to A/g$, so that the Yang–Mills Lagrangian reads $-\frac{1}{4g^2} G^a_{\mu\nu} G^{a\,\mu\nu}$. At one loop, a heavy quark of mass m changes the effective coupling below its mass according to $\frac{1}{g^2_{\text{low}}(\mu)} = \frac{1}{g^2_{\text{high}}(\mu)} \left[1 + \frac{\alpha_s}{6\pi} \ln \frac{\mu^2}{m^2}\right]$, where the two couplings agree at $\mu = m$. Explain why the coefficient is $\frac{\alpha_s}{\pi}$ times $\frac{4T_R}{12} = \frac{1}{6}$, the contribution of one quark flavour to β_0 in eq. (7.11). (Hint: integrate eq. (7.11) for $1/\alpha_s$ with n_f and $n_f + 1$ flavours.) Then replace $m \to m_t(1 + h/v)$ and show that the term linear in h is the effective Lagrangian (7.25) with $C = -\frac{\alpha_s}{3\pi}$.

Exercise.. (*The one-mass triangle.*) Compute the scalar triangle of eq. (7.42) with Feynman parameters, and show that it equals $\frac{2(d-3)}{(4-d)s} I[1,1](s)$. Derive the same relation with integration by parts, using the identity $0 = \int d^d k \, \frac{\partial}{\partial k^\mu} \frac{(k+p_1)^\mu}{k^2 (k+p_1)^2 (k+p_1+p_2)^2}$ and the IBP relation for the bubble, eq. (4.19). (You should find $(d-4)\,T = I[2,1](s) + I[1,2](s)$, where T is the triangle.)

Exercise.. (*Two-body phase space.*) Derive eq. (7.68). Check it in $d = 4$ against the familiar result $\int d\Phi_2 = \frac{1-z}{8\pi}$ for one massless and one massive particle. Then verify the soft limit used in section 7.8: show that $-\frac{\delta(1-z)}{2\epsilon}$ times the λ-integral at $z = 1$, together with the prefactors of eq. (7.73), gives the $\delta(1-z)$ term $C_A \left[\frac{1}{\epsilon^2} - \frac{\pi^2}{4}\right]$ of eq. (7.77).

Exercise.. (*Scale independence and the $q\bar{q}$ channel.*) (a) Using the DGLAP equation (7.84) and eq. (7.12), show that the hadronic cross-section obtained from eq. (7.86) is independent of μ_R and of μ_F up to terms of order α_s^4 (keep only the gluon density). (b) The amplitude for $q(p_1)\bar{q}(p_2) \to h\, g(p_3)$ is given by a single diagram with an s-channel gluon. Show that $\sum |\mathcal{M}|^2 = g^2 \mu^{2\epsilon} \left(\frac{C}{v}\right)^2 C_F N \, \frac{t^2+u^2-\epsilon(t+u)^2}{s}$, and that integrating over the phase space in four dimensions gives eq. (7.88).

References..

- H. M. Georgi, S. L. Glashow, M. E. Machacek, D. V. Nanopoulos, *Higgs bosons from two-gluon annihilation in proton-proton collisions*, Phys. Rev. Lett. **40** (1978) 692.
- F. Wilczek, *Decays of heavy vector mesons into Higgs particles*, Phys. Rev. Lett. **39** (1977) 1304; M. A. Shifman, A. I. Vainshtein, M. B. Voloshin, V. I. Zakharov, *Low-energy theorems for Higgs boson couplings to photons*, Sov. J. Nucl. Phys. **30** (1979) 711.
- S. Dawson, *Radiative corrections to Higgs boson production*, Nucl. Phys. B**359** (1991) 283.
- A. Djouadi, M. Spira, P. M. Zerwas, *Production of Higgs bosons in proton colliders: QCD corrections*, Phys. Lett. B**264** (1991) 440.
- M. Spira, A. Djouadi, D. Graudenz, P. M. Zerwas, *Higgs boson production at the LHC*, Nucl. Phys. B**453** (1995) 17, arXiv:hep-ph/9504378.
- M. Spira, *QCD effects in Higgs physics*, Fortsch. Phys. **46** (1998) 203, arXiv:hep-ph/9705337.
- K. G. Chetyrkin, B. A. Kniehl, M. Steinhauser, *Decoupling relations to $O(\alpha_s^3)$ and their connection to low-energy theorems*, Nucl. Phys. B**510** (1998) 61, arXiv:hep-ph/9708255.

- G. Altarelli, G. Parisi, *Asymptotic freedom in parton language*, Nucl. Phys. B**126** (1977) 298.
- J. C. Collins, D. E. Soper, G. Sterman, *Factorization of hard processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1, arXiv:hep-ph/0409313.
- T. Kinoshita, J. Math. Phys. **3** (1962) 650; T. D. Lee, M. Nauenberg, Phys. Rev. **133** (1964) B1549.
- S. Catani, M. H. Seymour, *A general algorithm for calculating jet cross sections in NLO QCD*, Nucl. Phys. B**485** (1997) 291, arXiv:hep-ph/9605323.
- S. Catani, *The singular behaviour of QCD amplitudes at two-loop order*, Phys. Lett. B**427** (1998) 161, arXiv:hep-ph/9802439.
- G. Sterman, M. E. Tejeda-Yeomans, *Multi-loop amplitudes and resummation*, Phys. Lett. B**552** (2003) 48, arXiv:hep-ph/0210130.
- R. V. Harlander, W. B. Kilgore, *Next-to-next-to-leading order Higgs production at hadron colliders*, Phys. Rev. Lett. **88** (2002) 201801, arXiv:hep-ph/0201206; C. Anastasiou, K. Melnikov, *Higgs boson production at hadron colliders in NNLO QCD*, Nucl. Phys. B**646** (2002) 220, arXiv:hep-ph/0207004; V. Ravindran, J. Smith, W. L. van Neerven, *NNLO corrections to the total cross section for Higgs boson production in hadron hadron collisions*, Nucl. Phys. B**665** (2003) 325, arXiv:hep-ph/0302135.
- C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, B. Mistlberger, *Higgs boson gluon-fusion production in QCD at three loops*, Phys. Rev. Lett. **114** (2015) 212001, arXiv:1503.06056.

PART III

Quantum Field Theory II

Preface to Part III

The subject of this part is modern applications of quantum field theory, with emphasis on the quantisation of non-abelian gauge theories. The following topics are discussed:

- Path-integral quantisation of bosonic and fermionic fields, and perturbation theory with path integrals.
- Classical gauge transformations and non-abelian gauge theories.
- Quantisation of non-abelian gauge theories: the Faddeev–Popov method and BRST symmetry.
- The quantum effective action and the effective potential.
- Classical symmetries of the effective action: Slavnov–Taylor identities and the Zinn-Justin equation.
- Physical interpretation of the effective action.
- Spontaneous symmetry breaking, the Goldstone theorem, and spontaneous symmetry breaking in theories with local gauge invariance.
- Power counting and ultraviolet divergences in field theories; renormalisable Lagrangians.
- Renormalisation and symmetries of non-abelian gauge theories.

Equations of Part I are cited as Eq. (I.n.m). The spontaneous breaking of local gauge symmetries (Chapter 8) is taken up again, and developed into the electroweak theory, in Part IV.

1

Path integral quantisation in Quantum Mechanics

We have carried out a quantisation programme for simple field theories in the course of QFT I by means of "canonical quantisation" (imposing commutation and anti-commutation relations on fields). Here we will quantise gauge-invariant field theories with a different method, using a formalism based on path integrals. It is imperative to develop this formalism. While the canonical formalism can be successfully applied to Quantum Electrodynamics, it becomes very cumbersome for the gauge field theories which describe the unified strong and electroweak interactions. *Remark.* Non-abelian gauge theories can, in principle, be quantised canonically, e.g. in Coulomb or axial gauge (Christ–Lee) or covariantly with BRST charges and an indefinite-metric state space (Kugo–Ojima, Prog. Theor. Phys. Suppl. 66 (1979) 1). The path integral is preferred because it makes the Faddeev–Popov ghosts and the covariant Feynman rules come out most economically (Weinberg QFT II, ch. 15). As a warm-up we revisit quantum mechanics, formulating quantisation using path integrals.

Conventions

Part III uses the conventions of Parts I and II; we collect here those that enter repeatedly.

- Natural units $\hbar = c = 1$ (except in this chapter, where $\hbar$ is kept), metric $g_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$, Dirac matrices as in Part I.
- Mode expansions: annihilation operators multiply $e^{-ik\cdot x}$, creation operators $e^{+ik\cdot x}$, with the invariant measure $d^3k/[(2\pi)^3 2\omega_k]$ (Part I). Feynman propagators carry $k^2 - m^2 + i\epsilon$; e.g. $\langle 0|T\phi(x)\phi(y)|0\rangle = \int \frac{d^4k}{(2\pi)^4} \frac{i\,e^{-ik\cdot(x-y)}}{k^2-m^2+i\epsilon}$.
- Generating functionals: $Z[J] = \int \mathcal{D}\phi\, e^{iS[\phi]+i\int d^4x\, J\phi}$, $Z = e^{iW}$, so that $\frac{\delta W}{\delta J} = \langle\phi\rangle_J$ and $\frac{\delta^2 W}{\delta J(x)\delta J(y)} = i\langle T\phi(x)\phi(y)\rangle_{\mathrm{c}}$. Effective action $\Gamma[\langle\phi\rangle] = W[J] - \int d^4x\, J\langle\phi\rangle$, $\frac{\delta\Gamma}{\delta\langle\phi\rangle} = -J$; $\frac{1}{i}\frac{\delta^2\Gamma}{\delta\phi\delta\phi}$ is the inverse of the full propagator and, for $n \geq 3$, $i\,\frac{\delta^n\Gamma}{\delta\phi^n}$ is the sum of 1PI graphs (Ch. 6). For constant fields $V_{\mathrm{eff}} = -\Gamma/\int d^4x$; the vacuum is the minimum of V_{eff} (a maximum of Γ), and $V''_{\mathrm{eff}} = -i\,D^{-1}(0) = m^2/Z$ for $D(p) = iZ/(p^2 - m^2)$ (Ch. 8). Grassmann derivatives are left derivatives unless stated otherwise.
- Abelian gauge theory (Part I, Part II): $D_\mu = \partial_\mu - ieA_\mu$, so that $\mathcal{L}_{\mathrm{int}} = +e\bar\psi\gamma^\mu A_\mu\psi$ and the QED vertex is $+ie\gamma^\mu$; gauge transformations $\psi \to e^{ie\theta}\psi$, $A_\mu \to A_\mu + \partial_\mu\theta$ (Ch. 3). Chapters 5 and 7 use instead the abelian limit of the non-abelian convention below, $\psi \to (1 - i\Theta)\psi$, $A_\mu \to A_\mu - \frac{1}{q}\partial_\mu\Theta$ (i.e. $\theta = -\Theta/q$).

- Non-abelian gauge theory (Ch. 3): $[T^a, T^b] = if^{abc}T^c$, $\mathrm{Tr}(T^aT^b) = T_R\delta^{ab}$; $D_\mu = \partial_\mu - igA^a_\mu T^a$, $G_{\mu\nu} = \frac{i}{g}[D_\mu, D_\nu]$, $G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu$; adjoint $D^{ab}_\mu = \partial_\mu\delta^{ab} - gf^{abc}A^c_\mu$. Gauge transformations $U = e^{-i\theta^a T^a}$: $\psi \to (1 - i\theta^a T^a)\psi$, $A^a_\mu \to A^a_\mu - \frac{1}{g}D^{ab}_\mu\theta^b$.
- Gauge fixing and ghosts (Ch. 4): $\mathcal{L}_{\mathrm{gf}} = -\frac{1}{2\xi}(\partial_\mu A^{a\mu})^2$, $\mathcal{L}_{\mathrm{FP}} = -\bar\eta^a\partial^\mu D^{ab}_\mu\eta^b \simeq (\partial^\mu\bar\eta^a)D^{ab}_\mu\eta^b$; propagators $\frac{-i\delta^{ab}}{k^2+i\epsilon}\left[g^{\mu\nu} - (1-\xi)\frac{k^\mu k^\nu}{k^2}\right]$ (gluon) and $\frac{i\delta^{ab}}{k^2+i\epsilon}$ (ghost).
- BRST (Ch. 5, used also in Ch. 7 and in the proof of renormalisability): $\delta_\theta F = \theta\, sF$ with $\delta_\theta\psi = -i\theta\eta^a T^a\psi$, $\delta_\theta A^a_\mu = -\frac{\theta}{g}D^{ab}_\mu\eta^b$, $\delta_\theta\eta^a = \frac{\theta}{2}f^{abc}\eta^b\eta^c$, $\delta_\theta\bar\eta^a = -\frac{\theta}{g}w^a$, $\delta_\theta w^a = 0$; $\mathcal{L}_{\mathrm{gf}} + \mathcal{L}_{\mathrm{FP}} = w^a\mathcal{G}^a + \frac{\xi}{2}w^aw^a + \bar\eta^a\Delta^a = -g\,s\left[\bar\eta^a(\mathcal{G}^a + \frac{\xi}{2}w^a)\right]$; the BRST charge acts as $\delta_\theta F = i\theta[Q, F]_\mp$. Sources for the BRST variations enter as $S[\phi, K] = S[\phi] + \int d^4x\, B_iK_i$ with $\delta_\theta\phi_i = \theta B_i$, and the Zinn-Justin equation reads $(\Gamma, \Gamma) = 0$ (Ch. 7).
- R_ξ gauges (Ch. 8): for scalars $\psi = \langle\psi\rangle + \psi'$ in a real representation, $u^a = iT^a\langle\psi\rangle$, $\mathcal{L}_{\mathrm{gf}} = -\frac{1}{2\xi}\left(\partial_\mu A^{a\mu} + \xi g\, u^a\cdot\psi'\right)^2$; Goldstone bosons and ghosts get masses squared ξM^2 and the massive vector propagator is $\frac{-i}{k^2-m^2+i\epsilon}\left[g^{\mu\nu} - (1-\xi)\frac{k^\mu k^\nu}{k^2-\xi m^2}\right]$.

1.1 The propagator

We consider a quantum mechanical state $|\psi\rangle$ which satisfies the Schrödinger equation:

$$i\hbar\partial_t|\psi\rangle = \hat H|\psi\rangle\,. \tag{1.1}$$

The solution of this equation

$$|\psi(t_2)\rangle = e^{-\frac{i}{\hbar}\hat H(t_2-t_1)}|\psi(t_1)\rangle\,, \tag{1.2}$$

determines the evolution of this state from an initial moment t_1 to a later moment t_2. The wave function $\psi(x_2, t_2) \equiv \langle x_2|\,\psi(t_2)\rangle$ is then

$$\psi(x_2, t_2) = \langle x_2|\,e^{-\frac{i}{\hbar}\hat H(t_2-t_1)}|\psi(t_1)\rangle\,. \tag{1.3}$$

We now insert a unit operator

$$1 = \int d^3x_1\,|x_1\rangle\langle x_1|\,, \tag{1.4}$$

obtaining

$$\begin{aligned}\psi(x_2, t_2) &= \int d^3x_1\,\langle x_2|\,e^{-\frac{i}{\hbar}\hat H(t_2-t_1)}|x_1\rangle\,\langle x_1|\,\psi(t_1)\rangle \\ &= \int d^3x_1\,\langle x_2|\,e^{-\frac{i}{\hbar}\hat H(t_2-t_1)}|x_1\rangle\,\psi(x_1, t_1)\,.\end{aligned} \tag{1.5}$$

In other words, if we know the wave function at one time, we can determine it fully at a later time by integrating,

$$\psi(x_2, t_2) = \int d^3x_1K(x_2, x_1; t_2 - t_1)\psi(x_1, t_1)\,, \tag{1.6}$$

with a kernel

$$K(x_2, x_1; t_2 - t_1) = \langle x_2 | \, e^{-\frac{i}{\hbar}\hat{H}(t_2 - t_1)} \, | x_1 \rangle \tag{1.7}$$

which depends on the Hamiltonian of the system and the elapsed time $t_2 - t_1$. This integration kernel is called the **"propagator"**. For $t_2 = t_1$, the propagator is a delta function

$$K(x_2, x_1; t_1 - t_1) = \langle x_2 | \, x_1 \rangle = \delta^3(x_2 - x_1). \tag{1.8}$$

Exercise: *Prove that the propagator $K(x, x'; t - t')$ satisfies the Schrödinger equation in the variables x, t for times $t > t'$.*

Exercise: *Prove that*

$$\int_{-\infty}^{\infty} dx e^{-ax^2} = \sqrt{\frac{\pi}{a}}.$$

Exercise: *Compute the propagator $K(x, x'; t - t')$ for*

- *a free particle,*
- *the simple harmonic oscillator.*

The propagator is the amplitude for a particle measured at position x_1 at time t_1 to propagate to a new position x_2 at time $t_2 > t_1$. We can verify this easily. Consider a particle measured at a position x_1. After time $t_2 - t_1$ it will be evolved to a new state

$$e^{-\frac{i}{\hbar}\hat{H}(t_2 - t_1)} \, | x_1 \rangle \, . \tag{1.9}$$

The probability amplitude to be measured at a position x_2 is obtained by taking the inner product of the evolved state and a position ket $|x_2\rangle$,

$$\langle x_2 | \, e^{-\frac{i}{\hbar}\hat{H}(t_2 - t_1)} \, | x_1 \rangle = K(x_2, x_1; t_2 - t_1). \tag{1.10}$$

We can attempt to compute the propagator for a transition which takes a very small time $\delta t \to 0$:

$$\begin{aligned} K(x, x'; \delta t) &= \langle x | \, e^{-\frac{i}{\hbar}\hat{H}\delta t} \, | x' \rangle \\ &= \langle x | \, 1 - \frac{i}{\hbar}\hat{H}\delta t \, | x' \rangle + O\left(\delta t^2\right) \\ &= \langle x | \left(1 - \frac{i}{\hbar}\hat{H}\delta t\right) \left(\int d^3p \, |p\rangle \langle p| \right) | x' \rangle + O\left(\delta t^2\right) \\ &= \int d^3p \left\{ \langle x | \, p \rangle \langle p | \, x' \rangle - \frac{i}{\hbar}\delta t \, \langle x | \hat{H} \, | p \rangle \langle p | \, x' \rangle \right\} + O\left(\delta t^2\right). \end{aligned} \tag{1.11}$$

We now specialise to Hamiltonian operators of the form $\hat{H} = f_1(\hat{p}) + f_2(\hat{x})$. Then

$$\langle x | \hat{H} \, | p \rangle = H \langle x | \, p \rangle \, , \tag{1.12}$$

where H is not an operator anymore but the classical Hamiltonian, i.e. a real-valued function of position and momentum. Recall that the position and momentum states are related via a Fourier transform,

$$\langle x | \, p \rangle = \frac{e^{\frac{i}{\hbar}px}}{\sqrt{2\pi\hbar}}. \tag{1.13}$$

For simplicity, let us consider only one dimension; the three-dimensional case is a faithful repetition of the same steps. Then we find for the propagator at small time intervals:

$$\begin{aligned} K(x,x';\delta t) &= \int \frac{dp}{2\pi\hbar} e^{\frac{i}{\hbar}p(x-x')}\left(1-\frac{i}{\hbar}H\delta t\right)+O\left(\delta t^2\right) \\ &= \int \frac{dp}{2\pi\hbar}\exp\left(\frac{i}{\hbar}\left\{p\left(x-x'\right)-H\delta t\right\}\right)+O\left(\delta t^2\right). \end{aligned} \tag{1.14}$$

An interesting form for the propagator for small time intervals arises when the Hamiltonian is of the form

$$H=\frac{p^2}{2m}+V(x). \tag{1.15}$$

Then,

$$\begin{aligned} K(x,x';\delta t) &= \int \frac{dp}{2\pi\hbar}\exp\left(\frac{i}{\hbar}\left\{p\left(x-x'\right)-\frac{p^2}{2m}\delta t-V(x)\delta t\right\}\right)+O\left(\delta t^2\right) \\ &= \int \frac{dp}{2\pi\hbar}\exp\left(\frac{i}{\hbar}\left\{-\frac{\delta t}{2m}\left(p-m\frac{x-x'}{\delta t}\right)^2+\delta t\frac{1}{2}m\left(\frac{x-x'}{\delta t}\right)^2-V(x)\delta t\right\}\right)+O\left(\delta t^2\right) \\ &= \int \frac{d\tilde{p}}{2\pi\hbar}\exp\left(\frac{i\delta t}{\hbar}\left\{-\frac{\tilde{p}^2}{2m}+\frac{1}{2}m\left(\frac{x-x'}{\delta t}\right)^2-V(x)\right\}\right)+O\left(\delta t^2\right), \end{aligned} \tag{1.16}$$

where in the final step we performed a trivial change of integration variable. Our final result reads:

$$K(x,x';\delta t)\simeq\frac{1}{N(\delta t)}\exp\left(\frac{i\delta t}{\hbar}\left\{\frac{1}{2}m\left(\frac{x-x'}{\delta t}\right)^2-V(x)\right\}\right), \tag{1.17}$$

with

$$\frac{1}{N(\delta t)}=\int\frac{dp}{2\pi\hbar}\exp\left(\frac{-i\delta t p^2}{2m\hbar}\right)=\left(\frac{m}{2\pi i\hbar\delta t}\right)^{1/2}. \tag{1.18}$$

1.2 The path integral

We consider now the transition from an initial position (x_i,t_i) to a final position (x_f,t_f), which has a probability amplitude given by the propagator:

$$K(x_f,x_i;t_f-t_i).$$

We can take a "snapshot" at an intermediate time t_1 during this transition:

$$t_i<t_1<t_f.$$

If taking the "snapshot" is a measurement of the position x_1 of the particle at the moment t_1, then the amplitude for the full transition will be:

$$K(x_f,x_i;t_f-t_i)=K(x_f,x_1;t_f-t_1)K(x_1,x_i;t_1-t_i). \tag{1.19}$$

If taking the "snapshot" is only a thought experiment and we do not actually determine the position x_1 of the particle at t_1 with a real measurement, we should integrate over all

probability amplitudes for the particle to have performed this transition via any point. We then have:

$$K(x_f, x_i; t_f - t_i) = \int_{-\infty}^{\infty} dx_1 K(x_f, x_1; t_f - t_1) K(x_1, x_i; t_1 - t_i). \tag{1.20}$$

We may decide to take several "snapshots" during the transition from x_i to x_f, at times $t_i < t_1 < t_2 < \ldots < t_n < t_f$. Using again the superposition principle we must write:

$$K(x_f, x_i; t_f - t_i) = \int_{-\infty}^{\infty} dx_1 \ldots dx_n K(x_f, x_n; t_f - t_n) K(x_n, x_{n-1}; t_n - t_{n-1}) \times \\ \ldots K(x_2, x_1; t_2 - t_1) K(x_1, x_i; t_1 - t_i). \tag{1.21}$$

For simplicity, we now take all intermediate time steps equal, $t_{k+1} - t_k = \delta t = \frac{t_f - t_i}{n+1}$, and let them become infinitesimal. Then we obtain

$$K(x_f, x_i; t_f - t_i) = \lim_{n\to\infty} \int_{-\infty}^{\infty} dx_1 \ldots dx_n K(x_f, x_n; \delta t) K(x_n, x_{n-1}; \delta t) \times \\ \ldots K(x_2, x_1; \delta t) K(x_1, x_i; \delta t). \tag{1.22}$$

We have discretised and taken the limit $n \to \infty$, which is the defining procedure of an integration. However, this new integral is rather unusual. Think of all the paths which connect the initial and final points x_i and x_f. The points which belong to these paths are accounted for by the limit $n \to \infty$ of infinitesimal transitions which we have taken in Eq. (1.22). Therefore, the right-hand side of this equation is an integral over all paths that a particle may take in going from $x_i \to x_f$.

We can insert the expressions for the propagator at small time intervals of Eq. (1.14) or Eq. (1.17) into Eq. (1.22). Notice that in the limit $\delta t \to 0$,

$$\frac{x_n - x_{n-1}}{\delta t} \to \dot{x_n}.$$

Then Eq. (1.22) becomes:

$$K(x_f, x_i; t_f - t_i) = \lim_{n\to\infty} \frac{1}{N(\delta t)^{n+1}} \int_{-\infty}^{\infty} dx_1 \ldots dx_n \\ \exp\left(\frac{i\delta t}{\hbar}\left[\frac{m}{2}\dot{x_1}^2 - V(x_1)\right]\right) \\ \exp\left(\frac{i\delta t}{\hbar}\left[\frac{m}{2}\dot{x_2}^2 - V(x_2)\right]\right) \\ \ldots \\ \exp\left(\frac{i\delta t}{\hbar}\left[\frac{m}{2}\dot{x_f}^2 - V(x_f)\right]\right). \tag{1.23}$$

Equivalently,

$$K(x_f, x_i; t_f - t_i) = \lim_{n\to\infty} \frac{1}{N(\delta t)^{n+1}} \int_{-\infty}^{\infty} dx_1 \ldots dx_n \\ \exp\left(\frac{i}{\hbar}\left\{L(x_1(t_1))\delta t + L(x_2(t_2))\delta t + \ldots L(x_f(t_f))\delta t\right\}\right), \tag{1.24}$$

where

$$L(x) = \frac{m}{2}\dot{x}^2 - V(x),$$

is the Lagrangian of the system. We now have a more concrete understanding of the above integral. This is an integration over “all paths” connecting the fixed points x_i and x_f, which we write symbolically as

$$K(x_f, x_i; t_f - t_i) = \frac{1}{\mathcal{N}} \int \mathcal{D}x \exp\left(\frac{i}{\hbar} \int_{t_i}^{t_f} dt L[x(t)]\right), \tag{1.25}$$

and the sum (integral) in the exponential is the classical action as it is evaluated on each of the paths. Even shorter, we can write:

$$K(x_f, x_i; t_f - t_i) = \frac{1}{\mathcal{N}} \int \mathcal{D}x \exp\left(\frac{i}{\hbar} S[x]\right). \tag{1.26}$$

Exercise: *Consider a Lagrangian of the form*

$$L = \frac{1}{2} f(x)\dot{x}^2 + g(x)\dot{x} - V(x).$$

1. *Compute the Hamiltonian.*
2. *Compute the propagator for a small transition.*
3. *Write the path-integral expression for the propagator for a finite time interval. Notice that the measure of the path integration is modified:*

$$\mathcal{D}x \to \mathcal{D}x\, f(x)^{\frac{1}{2}}.$$

1.3 An “adventurous” transition

We look now at a more eventful transition amplitude. We first prepare a particle at an initial position x_i at a time t_i and let it evolve for some time $t - t_i$ according to a Hamiltonian $\hat{H}$:

$$e^{-\frac{i}{\hbar}\hat{H}(t-t_i)} |x_i\rangle .$$

At the time t, something abrupt occurs (e.g. an interaction with another particle which was originally far away) and modifies the particle state. We will see later how we can describe interactions of particles using path integrals; now, let us consider an “easy” modification of the state where the state is “mixed up” in a simple way, acting on it with the position operator:

$$\hat{x} e^{-\frac{i}{\hbar}\hat{H}(t-t_i)} |x_i\rangle .$$

Then we allow the particle to evolve undistracted for a time $t_f - t$,

$$e^{-\frac{i}{\hbar}\hat{H}(t_f-t)} \hat{x} e^{-\frac{i}{\hbar}\hat{H}(t-t_i)} |x_i\rangle ,$$

and then we place a detector at x_f:

$$\left\langle x_f \right| e^{-\frac{i}{\hbar}\hat{H}(t_f-t)} \hat{x} e^{-\frac{i}{\hbar}\hat{H}(t-t_i)} |x_i\rangle .$$

We can compute this matrix element as a path integral. Before we proceed, we should use some language which is more convenient to describe “eventful” transitions. We can write the same transition amplitude as:

$$\begin{aligned} & \left\{\left\langle x_f \right| e^{-\frac{i}{\hbar}\hat{H}t_f}\right\} \left\{e^{\frac{i}{\hbar}\hat{H}t} \hat{x} e^{-\frac{i}{\hbar}\hat{H}t}\right\} \left\{e^{\frac{i}{\hbar}\hat{H}t_i} |x_i\rangle\right\} \\ = \; & \left\langle x_f, t_f \right| \hat{x}(t) |x_i, t_i\rangle . \end{aligned} \tag{1.27}$$

We have defined states

$$\left|\psi, t\right\rangle \equiv e^{\frac{i}{\hbar}\hat{H}t}\left|\psi\right\rangle, \tag{1.28}$$

which refer to a fixed moment t only and do not evolve:

$$\begin{aligned} & i\hbar\frac{\partial}{\partial t}\left|\psi, t\right\rangle = i\hbar\frac{\partial}{\partial t}\left(e^{\frac{i}{\hbar}\hat{H}t}\left|\psi\right\rangle\right) \\ = \quad & e^{\frac{i}{\hbar}\hat{H}t}\left(i\hbar\frac{\partial}{\partial t}\left|\psi\right\rangle\right) + i\hbar\left(\frac{\partial}{\partial t}e^{\frac{i}{\hbar}\hat{H}t}\right)\left|\psi\right\rangle \\ = \quad & e^{\frac{i}{\hbar}\hat{H}t}\left(\hat{H}-\hat{H}\right)\left|\psi\right\rangle = 0. \end{aligned} \tag{1.29}$$

We have also defined operators which do change with time,

$$\hat{O}(t) = e^{\frac{i}{\hbar}\hat{H}t}\hat{O}e^{-\frac{i}{\hbar}\hat{H}t}. \tag{1.30}$$

We recognise here the “Heisenberg picture” of evolution. For us, it is convenient to assign a time label to a state which denotes a particle or a collection of particles at the beginning of an experiment or at the end of it. However, one could have equally well chosen to work in the probably more familiar “Schrödinger picture”.

Let us now compute this “eventful” transition:

$$\left\langle x_f, t_f\right|\hat{x}(t_j)\left|x_i, t_i\right\rangle \qquad \text{with } t_i < t_j < t_f,$$

following the method we used for the simple transition $\left\langle x_f, t_f\right|\left. x_i, t_i\right\rangle = \left\langle x_f\right| e^{-\frac{i}{\hbar}\hat{H}\left(t_f - t_i\right)}\left|x_i\right\rangle$, and subdividing the transition into small time intervals. We will find a similar path integral for the new case as well:

$$\begin{aligned} \left\langle x_f, t_f\right|\hat{x}(t_j)\left|x_i, t_i\right\rangle \quad = \quad & \int dx_1 \ldots dx_{j-1}dx_j \ldots dx_n \\ & \times\left\langle x_f, t_f\right|\left. x_n, t_n\right\rangle \ldots \left\langle x_{j+1}, t_{j+1}\right|\left. x_j, t_j\right\rangle \\ & \times\left\langle x_j, t_j\right|\hat{x}(t_j)\left|x_{j-1}, t_{j-1}\right\rangle \\ & \times\left\langle x_{j-1}, t_{j-1}\right|\left. x_{j-2}, t_{j-2}\right\rangle \ldots \left\langle x_1, t_1\right|\left. x_i, t_i\right\rangle. \end{aligned} \tag{1.31}$$

The subdivision of time is carefully chosen. We have

$$\begin{aligned} \hat{x}(t_j)\left|x_j, t_j\right\rangle \quad &= \quad \left(e^{\frac{i}{\hbar}\hat{H}t_j}\hat{x}e^{-\frac{i}{\hbar}\hat{H}t_j}\right)\left(e^{\frac{i}{\hbar}\hat{H}t_j}\left|x_j\right\rangle\right) \\ &= \quad e^{\frac{i}{\hbar}\hat{H}t_j}\left(\hat{x}\left|x_j\right\rangle\right) \\ &= \quad x_j e^{\frac{i}{\hbar}\hat{H}t_j}\left|x_j\right\rangle \\ &= \quad x_j\left|x_j, t_j\right\rangle \\ \rightsquigarrow \left\langle x_j, t_j\right|\hat{x}(t_j)\left|x_{j-1}, t_{j-1}\right\rangle \quad &= \quad x_j\left\langle x_j, t_j\right|\left. x_{j-1}, t_{j-1}\right\rangle. \end{aligned} \tag{1.32}$$

We then obtain for the transition amplitude the same succession of propagators as for the simple transition multiplied with an additional factor x_j:

$$\begin{aligned} & \left\langle x_f, t_f\right|\hat{x}(t_j)\left|x_i, t_i\right\rangle = \int dx_1 \ldots dx_j x_j \ldots dx_n \\ & \times\left\langle x_f, t_f\right|\left. x_n, t_n\right\rangle \ldots \left\langle x_{j+1}, t_{j+1}\right|\left. x_j, t_j\right\rangle\left\langle x_j, t_j\right|\left. x_{j-1}, t_{j-1}\right\rangle \ldots \left\langle x_1, t_1\right|\left. x_i, t_i\right\rangle. \end{aligned} \tag{1.33}$$

Introducing, as before, the explicit form for the propagator during a small time transition

$$\langle x_b, t_{n+1}|\, x_a, t_n\rangle \approx \left(\frac{m}{2\pi i\hbar\,(t_{n+1}-t_n)}\right)^{\frac{1}{2}} \exp\left[\frac{i}{\hbar}\left\{\frac{m}{2}\left(\frac{x_b-x_a}{t_{n+1}-t_n}\right)^2 - V(x_a)\right\}(t_{n+1}-t_n)\right]$$

and taking equal time intervals, we obtain the path integral

$$\begin{aligned}\left\langle x_f, t_f\right|\hat{x}(t_j)\left|x_i, t_i\right\rangle &= \lim_{n\to\infty}\frac{1}{N(\delta t)^{n+1}}\int_{-\infty}^{\infty} dx_1\ldots dx_j x_j\ldots dx_n \\ \times\exp\left(\frac{i}{\hbar}\sum_{r=1}^{n+1}L(x_r,\dot{x}_r)\delta t\right) &\qquad \text{with } \delta t = \frac{t_f-t_i}{n+1}. \end{aligned} \tag{1.34}$$

In compact notation we can write

$$\left\langle x_f, t_f\right|\hat{x}(\tau)\left|x_i, t_i\right\rangle = \int \mathcal{D}x x(\tau) e^{\frac{i}{\hbar}S[x]}. \tag{1.35}$$

Exercise:

1. *Prove that*
$$\left\langle x_f, t_f\right|\hat{x}(\tau)\left|x_i, t_i\right\rangle = \int_{-\infty}^{\infty} dx\, x\left\langle x_f, t_f\right|\left. x,\tau\right\rangle\left\langle x,\tau\right|\left. x_i, t_i\right\rangle.$$
2. *Evaluate the above integral explicitly for a free particle.*
3. *Observe the dependence of the result on the intermediate time τ.*
4. *You could have considered the simpler matrix element $\left\langle x_f, t_f\right|\left. x_i, t_i\right\rangle$ and written down an analogous integral. Observe how the intermediate time τ drops out from the final expression, when nothing special occurs then!*

We now elaborate further on the form of the path integral that we have just found. Recall the method for computing exponential integrals of the form

$$I_n = \int_{-\infty}^{\infty} dx x^n e^{-ax^2} \qquad \text{with } n = 1, 2, \ldots, \tag{1.36}$$

from the result of the integral

$$I_0 \equiv \int_{-\infty}^{\infty} dx e^{-ax^2} = \sqrt{\frac{\pi}{a}}. \tag{1.37}$$

We can compute I_n with very simple differentiations, after we add a "source" term to the exponent of the integrand:

$$\begin{aligned} I_J &= \int_{-\infty}^{\infty} dx e^{-ax^2+Jx} = \int_{-\infty}^{\infty} dx e^{-a\left(x-\frac{J}{2a}\right)^2+\frac{J^2}{4a}} \\ &= \int_{-\infty}^{\infty} d\tilde{x} e^{-a\tilde{x}^2+\frac{J^2}{4a}} = \sqrt{\frac{\pi}{a}} e^{\frac{J^2}{4a}}. \end{aligned} \tag{1.38}$$

To compute I_1 it is sufficient to differentiate the last expression with respect to the source and substitute $J = 0$:

$$\left.\frac{dI_J}{dJ}\right|_{J=0} = \int_{-\infty}^{\infty} dx \left.\frac{de^{-ax^2+Jx}}{dJ}\right|_{J=0} = \int_{-\infty}^{\infty} dx x e^{-ax^2}.$$

Similarly,

$$\left.\frac{d^n I_J}{dJ^n}\right|_{J=0} = \int_{-\infty}^{\infty} dx x^n e^{-ax^2}.$$

Adding a source term to the exponent does not increase the difficulty of the computation and it allows us to calculate all integrals where the integrand is multiplied with a polynomial in the integration variable. This is very suggestive, and we will do the same trick for path integrals such as the one that we found in Eq. (1.34). We then add a linear term (source) in the Lagrangian; this is only a computational trick and eventually we will compute all interesting physical quantities as in the above example with the source term set to zero. The simple transition from a state $|x_i, t_i\rangle$ to a state $\left|x_f, t_f\right\rangle$ in the presence of the source has a probability amplitude:

$$\left\langle x_f, t_f\right| x_i, t_i\rangle_J = \lim_{n\to\infty} \frac{1}{N(\delta t)^{n+1}} \int dx_1 \dots dx_n e^{\frac{i}{\hbar}\sum_k dt_k (L(x_k,\dot{x}_k)+J_k x_k)}. \tag{1.39}$$

We can then compute

$$\left\langle x_f, t_f\right| \hat{x}(t_l) \left|x_i, t_i\right\rangle = \frac{\hbar}{i\,\delta t}\frac{\partial}{\partial J_l} \left\langle x_f, t_f\right| x_i, t_i\rangle_J \Big|_{J=0}. \tag{1.40}$$

Let us now differentiate twice with respect to the source:

$$\left(\frac{\hbar}{i\,\delta t}\right)^2 \frac{\partial^2}{\partial J_l \partial J_q} \left\langle x_f, t_f\right| x_i, t_i\rangle_J \Big|_{J=0} = \lim_{n\to\infty} \frac{1}{N(\delta t)^{n+1}} \int dx_1 \dots dx_l x_l \dots dx_q x_q \dots dx_n \times e^{\frac{i}{\hbar}\sum_k dt_k L(x_k,\dot{x}_k)}. \tag{1.41}$$

We can recognise the right-hand side as the expectation value for the product of two position operators $\hat{x}(t_l)\hat{x}(t_q)$ if t_q is earlier than t_l or $\hat{x}(t_q)\hat{x}(t_l)$ otherwise. We then write

$$\left\langle x_f, t_f\right| T\left(\hat{x}(t_l)\hat{x}(t_q)\right) \left|x_i, t_i\right\rangle = \left(\frac{\hbar}{i\,\delta t}\right)^2 \frac{\partial^2}{\partial J_l \partial J_q} \left\langle x_f, t_f\right| x_i, t_i\rangle_J \Big|_{J=0}, \tag{1.42}$$

where we have introduced the notation $T(\hat{O}(t_1)\hat{O}(t_3)\hat{O}(t_2)\dots\hat{O}(t_n))$ to remind us that we should put the operators in the correct time order once the sequence of the moments t_i is known. For example, if $t_1 < t_2 < t_3 < \dots < t_n$ we should write

$$T(\hat{O}(t_1)\hat{O}(t_3)\hat{O}(t_2)\dots\hat{O}(t_n)) = \hat{O}(t_n)\dots\hat{O}(t_3)\hat{O}(t_2)\hat{O}(t_1).$$

Exercise: *Compute* $\left\langle x_f, t_f\right| T\left(\hat{x}(t_l)\hat{x}(t_q)\right) \left|x_i, t_i\right\rangle$ *for a free particle.*

1.4 Functional differentiation

It is cumbersome to work with path integrals by writing explicitly the limit of infinitely fine discretised paths. We introduced earlier a more compact notation,

$$\left\langle x_f, t_f\right| x_i, t_i\rangle_J = \frac{1}{\mathcal{N}} \int Dx e^{\frac{i}{\hbar}\int_{t_i}^{t_f} dt[L(x(t),\dot{x}(t))+J(t)x(t)]}. \tag{1.43}$$

We can write neatly expressions for the expectation values of operators,

$$\left\langle x_f, t_f\right| \hat{x}(t_1) \left|x_i, t_i\right\rangle = \frac{1}{\mathcal{N}} \int Dx\; x(t_1)\; e^{\frac{i}{\hbar}\int_{t_i}^{t_f} dt L(x(t),\dot{x}(t))}$$

or

$$\left\langle x_f, t_f \right| T\left(\hat{x}(t_1)\hat{x}(t_2)\right) \left| x_i, t_i \right\rangle = \frac{1}{\mathcal{N}} \int Dx\; x(t_1)\; x(t_2)\; e^{\frac{i}{\hbar}\int_{t_i}^{t_f} dt L(x(t),\dot{x}(t))}$$

by defining a functional derivative. We consider an integral $F[f]$ over a function $f(t)$ (for example, along a path). We define a functional derivative by slightly changing the function $f(y)$:

$$\frac{\delta F[f(y)]}{\delta f(t)} = \lim_{\epsilon \to 0} \frac{F\left[f(y) + \epsilon\delta(y-t)\right] - F\left[f(y)\right]}{\epsilon}. \tag{1.44}$$

For example, consider the derivative of the action integral in the presence of a source with respect to the source:

$$\begin{aligned} \frac{\delta \int dy \left\{L\left(x(y), \dot{x}(y)\right) + J(y)x(y)\right\}}{\delta J(t)} &= \int dy \delta(t-y)x(y) \\ &= x(t). \end{aligned} \tag{1.45}$$

In practice, we only need the chain rule and the rule

$$\frac{\delta f(x)}{\delta f(y)} = \delta(x-y).$$

The expectation values of time-ordered operators can then be written as

$$\left\langle x_f, t_f \right| T\left(\hat{x}(t_1)\ldots\hat{x}(t_n)\right) \left| x_i, t_i \right\rangle = \left(\frac{\hbar}{i}\right)^n \frac{\delta^n}{\delta J(t_1)\ldots\delta J(t_n)} \left\langle x_f, t_f \right| x_i, t_i \rangle_J \Big|_{J=0}. \tag{1.46}$$

1.5 Vacuum to vacuum transitions

Field theory allows us to compute transitions between states with different particle content. Interestingly, we can build particle states acting with creation (or field) operators on the ground state which contains no particles (vacuum). All transition amplitudes can be described as expectation values of operators in the vacuum:

$$\left\langle 0, t_f \right| T(\ldots) \left| 0, t_i \right\rangle.$$

We can compute expectation values of operators in the vacuum with path integrals. First, we try a direct approach:

$$\left\langle 0, t_f \right| T(\ldots) \left| 0, t_i \right\rangle = \int dx dx' \left\langle 0, t_f \right| x, t\rangle \left\langle x, t \right| T(\ldots) \left| x', t' \right\rangle \left\langle x', t' \right| 0, t_i \rangle. \tag{1.47}$$

This formula is complicated. It requires that we know the wave function of the vacuum and that we are able to convolute it with the result for a path integral. There is a simpler way, with fewer integrations.

Consider a Hamiltonian $\hat{H}$ with eigenstates $|n\rangle$,

$$\hat{H}\left|n\right\rangle = E_n \left|n\right\rangle \qquad \text{(in the Schrödinger picture)}$$

and a general state

$$\left|\psi\right\rangle = \sum_n c_n \left|n\right\rangle.$$

Taking a "Heisenberg photograph" of the state in the very past ($-t \to -\infty$), we find:

$$\left|\psi, -t\right\rangle = e^{i\hat{H}(-t)} \sum_n c_n \left|n\right\rangle . \tag{1.48}$$

Now, we play a mathematical trick; we give a very small imaginary part to the Hamiltonian

$$\hat{H} \to \hat{H}(1 - i\epsilon), \qquad (\epsilon \to 0^+),$$

which has the same energy eigenstates as the physical Hamiltonian in the limit. The general state in the very past with the modified Hamiltonian is

$$\begin{aligned}
\left|\psi, -t\right\rangle &= e^{i\hat{H}(-t)(1-i\epsilon)} \sum_n c_n \left|n\right\rangle \\
&= \sum_n c_n e^{-(\epsilon+i)E_n t} \left|n\right\rangle \\
&= e^{-(\epsilon+i)E_0 t} \left[c_0 \left|0\right\rangle + \sum_{n=1}^{\infty} e^{-(\epsilon+i)(E_n - E_0)t} c_n \left|n\right\rangle \right] \\
&= e^{-\epsilon E_0 t} \left[c_0 \left|0, -t\right\rangle + \sum_{n=1}^{\infty} e^{-\epsilon(E_n - E_0)t} c_n \left|n, -t\right\rangle \right] .
\end{aligned} \tag{1.49}$$

Recall that the states in Eq. (1.49) are in the Heisenberg picture, and thus independent of time. The time argument in the ket is only a label. For a very long time t in the past the exponential $e^{-\epsilon(E_n - E_0)t}$ vanishes; it vanishes faster for larger energy eigenvalues. We can then approximate,

$$\left|\psi, -t\right\rangle \approx e^{-\epsilon E_0 t} c_0 \left|0, -t\right\rangle . \tag{1.50}$$

This is very convenient. An arbitrary Heisenberg state in the very past with the slightly complex Hamiltonian is, essentially, the vacuum state in the very past. Higher energy eigenstates do not contribute to the superposition since the small imaginary part forces them to decay as we take the time $-t$ further back in the past.

Similarly, for a general Heisenberg state $\left\langle\psi', t\right|$ labelled in the far future ($t \to \infty$), we have

$$\begin{aligned}
\left\langle\psi', +t\right| &= \sum_n c_n'^* \left\langle n\right| e^{-iHt(1-i\epsilon)} \\
&\approx c_0'^* \left\langle 0\right| e^{-iE_0 t(1-i\epsilon)} \\
&\approx c_0'^* e^{-\epsilon E_0 t} \left\langle 0, t\right| , \qquad t \to +\infty .
\end{aligned} \tag{1.51}$$

For amplitudes of vacuum-to-vacuum transitions over very long times we can therefore write

$$\left\langle 0, t\right| T(\ldots) \left|0, -t\right\rangle = \frac{e^{2\epsilon E_0 t}}{c_0 c_0'^*} \left\langle\psi', t\right| T(\ldots) \left|\psi, -t\right\rangle , \quad t \to \infty, \tag{1.52}$$

as long as we set $H \to H(1 - i\epsilon)$ in the right-hand side.

We can fix the normalisation by dividing by the simplest amplitude of the kind that corresponds to a vacuum-to-vacuum transition without an "interaction" occurring at an intermediate time. We obtain

$$\frac{\left\langle 0, t\right| T(\ldots) \left|0, -t\right\rangle}{\left\langle 0, t\right|\left. 0, -t\right\rangle} = \frac{\left\langle\psi', t\right| T(\ldots) \left|\psi, -t\right\rangle}{\left\langle\psi', t\right|\left. \psi, -t\right\rangle}, \quad t \to \infty. \tag{1.53}$$

We can choose as states in the right-hand side of Eq. (1.53) two position eigenstates in the Heisenberg picture, one labelled in the very far past and the other in the very far future:

$$\frac{\langle 0,t|\,T(\ldots)\,|0,-t\rangle}{\langle 0,t|\,0,-t\rangle} \;=\; \frac{\langle x_2,t|\,T(\ldots)\,|x_1,-t\rangle}{\langle x_2,t|\,x_1,-t\rangle}, \quad t\to\infty. \tag{1.54}$$

Then we can express vacuum-to-vacuum transitions, from the far past ($t\to-\infty$) to the far future ($t\to+\infty$), as a ratio of path integrals,

$$\begin{aligned}&\frac{\langle 0,\infty|\,T(\hat{x}(t_1)\ldots\hat{x}(t_r))\,|0,-\infty\rangle}{\langle 0,\infty|\,0,-\infty\rangle}\\ &\quad=\lim_{\epsilon\to 0^+}\frac{\int\mathcal{D}x\,x(t_1)\ldots x(t_r)\,e^{\frac{i}{\hbar}\int_{-\infty}^{+\infty}dtL[x(t),\epsilon]}}{\int\mathcal{D}x\,e^{\frac{i}{\hbar}\int_{-\infty}^{+\infty}dtL[x(t),\epsilon]}}.\end{aligned} \tag{1.55}$$

The action integral in the exponent of the path integrals in Eq. (1.55) is the classical action, with a small modification (as indicated by the ϵ in the argument of the Lagrangian) which is inherited from the $\hat{H}\to\hat{H}(1-i\epsilon)$ deformation that we employed in order to "decay" excited states.

1.6 The simple harmonic oscillator

An instructive example is the simple harmonic oscillator, with a classical Hamiltonian

$$H=\frac{p^2}{2m}+\frac{1}{2}m\omega^2x^2. \tag{1.56}$$

The Hamiltonian of Eq. (1.56) is positive definite. For vacuum-to-vacuum transitions over large times, we will need to modify it, $H\to H\cdot(1-i\epsilon)$, giving it a small imaginary part. We can achieve this in an easy way, by giving a small negative imaginary part to the square of the frequency,

$$\omega^2\to\omega^2-i\epsilon,\quad \epsilon\to 0^+. \tag{1.57}$$

The slightly deformed Hamiltonian reads,

$$H=\frac{p^2}{2m}+\frac{1}{2}m(\omega^2-i\epsilon)x^2. \tag{1.58}$$

The Lagrangian now acquires an infinitesimal positive imaginary part,

$$\begin{aligned}L &= p\dot{x}-H\\ &= \frac{1}{2}m\dot{x}^2-\frac{1}{2}m\omega^2(1-i\epsilon)x^2.\end{aligned} \tag{1.59}$$

We will now compute a so-called "generating functional integral",

$$\begin{aligned}Z[J] &= \int\mathcal{D}x\,e^{i\int dt(L+J(t)x(t))}\\ &= \int\mathcal{D}x\,e^{i\int dt\left[\frac{1}{2}m\dot{x}^2-\frac{1}{2}m\omega^2(1-i\epsilon)x^2+J(t)x(t)\right]},\end{aligned} \tag{1.60}$$

which also includes a source term in the action. Functional differentiation with respect to the source J will allow us to compute a large variety of vacuum-to-vacuum transitions including interactions or perturbations.

Things become easier if we perform a Fourier transformation of all quantities which depend on time:

$$x(t) = \int_{-\infty}^{\infty} \frac{dE}{2\pi} e^{-iEt} \tilde{x}(E), \tag{1.61}$$

$$J(t) = \int_{-\infty}^{\infty} \frac{dE}{2\pi} e^{-iEt} \tilde{J}(E). \tag{1.62}$$

The action integral in the exponent of the generating functional becomes

$$\begin{aligned}
S[x] &= \int dt \left[\frac{1}{2}m\dot{x}^2 - \frac{1}{2}m(1-i\epsilon)\omega^2 x^2 + J(t)x(t)\right] \\
&= \frac{1}{4\pi^2}\int dE dE' dt \frac{1}{2} e^{-it(E+E')} \times \left\{\left[-mEE' - m\omega^2(1-i\epsilon)\right]\tilde{x}(E)\tilde{x}(E') \right. \\
&\qquad \left. + \tilde{J}(E)\tilde{x}(E') + \tilde{x}(E)\tilde{J}(E')\right\} \\
&= \int \frac{dE}{4\pi}\left\{m\left[E^2 - \omega^2 + i\epsilon\omega^2\right]\tilde{x}(E)\tilde{x}(-E) \right. \\
&\qquad \left. + \tilde{J}(E)\tilde{x}(-E) + \tilde{x}(E)\tilde{J}(-E)\right\} \\
&= \int \frac{dE}{4\pi}\left\{m\left[E^2 - \omega^2 + i\epsilon\right]\tilde{x}(E)\tilde{x}(-E) + \tilde{J}(E)\tilde{x}(-E) + \tilde{x}(E)\tilde{J}(-E)\right\} \\
&= \int \frac{dE}{4\pi}\left\{\left(\tilde{x}(E) + \frac{\tilde{J}(E)}{m\left[E^2 - \omega^2 + i\epsilon\right]}\right) m\left[E^2 - \omega^2 + i\epsilon\right] \right. \\
&\qquad \times \left(\tilde{x}(-E) + \frac{\tilde{J}(-E)}{m\left[E^2 - \omega^2 + i\epsilon\right]}\right) \\
&\qquad \left. - \frac{\tilde{J}(E)\tilde{J}(-E)}{m\left[E^2 - \omega^2 + i\epsilon\right]}\right\},
\end{aligned} \tag{1.63}$$

where in the third step we have simplified $\omega^2\epsilon \to \epsilon$, since only the sign of the infinitesimal imaginary part matters. We can also define

$$\tilde{y}(E) = \tilde{x}(E) + \frac{\tilde{J}(E)}{m\left[E^2 - \omega^2 + i\epsilon\right]}. \tag{1.64}$$

The second term of the right-hand side is, up to its sign, the Fourier transform of the solution of the Euler–Lagrange equations, i.e. of the classical path. We can verify it easily. The classical equation of motion for the harmonic oscillator with a source term and a small negative imaginary part for the frequency is

$$m\left[\frac{d^2 x_{cl}}{dt^2} + \left(\omega^2 - i\epsilon\right)x_{cl}\right] = J(t), \tag{1.65}$$

and taking the Fourier transform

$$x_{cl} = \int_{-\infty}^{\infty} \frac{dE}{2\pi} e^{-iEt} \tilde{x}_{cl}(E),$$

we obtain

$$\begin{aligned}
& m\left[E^2 - \omega^2 + i\epsilon\right]\tilde{x}_{cl}(E) = -\tilde{J}(E) \\
\rightsquigarrow \qquad & \tilde{x}_{cl}(E) = -\frac{\tilde{J}(E)}{m\left[E^2 - \omega^2 + i\epsilon\right]}.
\end{aligned} \tag{1.66}$$

Then the path integral over paths $x(t)$, under a shift

$$x(t) = x_{cl}(t) + y(t), \tag{1.67}$$

becomes

$$Z[J] = e^{-\frac{1}{2}\int \frac{dE}{2\pi} \frac{\tilde{J}(E)i\tilde{J}(-E)}{m\left[E^2-\omega^2+i\epsilon\right]}} \int \mathcal{D}y e^{\frac{i}{2}\int \frac{dE}{2\pi}\tilde{y}(E)m\left[E^2-\omega^2+i\epsilon\right]\tilde{y}(-E)}. \tag{1.68}$$

The dependence of the path integral on the source terms becomes a simple pre-factor. It is easy to see that the new path integral corresponds to the harmonic oscillator Hamiltonian with no sources. We can therefore write the above expression as,

$$\left\langle x_f, +t\right| x_i, -t\rangle_J = e^{\frac{i}{2}\int \frac{dE}{2\pi} \frac{\tilde{J}(E)(-1)\tilde{J}(-E)}{m\left[E^2-\omega^2+i\epsilon\right]}} \left\langle x_f, +t\right| x_i, -t\rangle, \tag{1.69}$$

where $t \to \infty$. Performing the inverse Fourier transformations, we have:

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{dE}{2\pi} \frac{-\tilde{J}(E)\tilde{J}(-E)}{m\left[E^2-\omega^2+i\epsilon\right]} \\ = & \int_{-\infty}^{\infty} \frac{dE}{2\pi} \frac{-\left(\int dt\, e^{iEt}J(t)\right)\left(\int dt' e^{-iEt'}J(t')\right)}{m\left[E^2-\omega^2+i\epsilon\right]} \\ = & \int dt dt' J(t)G(t-t')J(t'), \end{aligned} \tag{1.70}$$

with

$$G(t-t') = -\frac{1}{m}\int_{-\infty}^{+\infty} \frac{dE}{2\pi} \frac{e^{i(t-t')E}}{E^2-\omega^2+i\epsilon}. \tag{1.71}$$

We then write:

$$\left\langle x_f, +t\right| x_i, -t\rangle_J = e^{\frac{i}{2}\int dt dt' J(t)G(t-t')J(t')} \left\langle x_f, +t\right| x_i, -t\rangle. \tag{1.72}$$

We can now differentiate this expression with respect to the sources as many times as we need. We then obtain:

$$\begin{aligned} \left\langle x_f, +t\right| T\hat{x}(t_n)\ldots\hat{x}(t_1)\left|x_i, -t\right\rangle = & \left(\frac{1}{i^n}\frac{\delta^n}{\delta J(t_1)\ldots\delta J(t_n)} e^{\frac{i}{2}\int dt dt' J(t)G(t-t')J(t')}\right)\Bigg|_{J=0} \\ & \times \left\langle x_f, +t\right| x_i, -t\rangle. \end{aligned} \tag{1.73}$$

Taking the limit $\epsilon \to 0^+$ selects the ground state on both sides of the equation as asymptotic states (the normalisation factors are also the same on both sides of the equation). We therefore have for the vacuum expectation value of time-ordered position operators:

$$\frac{\langle 0, t| T\hat{x}(t_n)\ldots\hat{x}(t_1)|0, -t\rangle}{\langle 0, t|\, 0, -t\rangle} = \left(\frac{1}{i^n}\frac{\delta^n}{\delta J(t_1)\ldots\delta J(t_n)} e^{\frac{i}{2}\int dt dt' J(t)G(t-t')J(t')}\right)\Bigg|_{J=0}. \tag{1.74}$$

Exercise:

1. *Show that*

$$\langle 0, t| T\hat{x}(t_n)\ldots\hat{x}(t_1)|0, -t\rangle = 0,$$

for n odd.

2. *Compute* $\langle 0,t|\,T\hat{x}(t_2)\hat{x}(t_1)\,|0,-t\rangle$ *and* $\langle 0,t|\,T\hat{x}(t_4)\ldots\hat{x}(t_1)\,|0,-t\rangle$ *in terms of* $G(t-t')$.

3. *Find a general expression for* $\langle 0,t|\,T\hat{x}(t_{2n})\ldots\hat{x}(t_1)\,|0,-t\rangle$ *in terms of* $G(t-t')$.

We now compute the propagator $G(t_1-t_2)$ explicitly:

$$\begin{aligned} G(t_1-t_2) &= -\frac{1}{m}\int\frac{dE}{2\pi}\frac{e^{-i(t_2-t_1)E}}{E^2-\omega^2+i\epsilon} \\ &= -\frac{1}{m}\int\frac{dE}{2\pi}\frac{e^{-i(t_2-t_1)E}}{(E-\omega+i\epsilon)\,(E+\omega-i\epsilon)} \\ &= -\frac{1}{m}\int\frac{dE}{2\pi}\frac{e^{-i(t_2-t_1)E}}{2\omega}\left[\frac{1}{E-\omega+i\epsilon}-\frac{1}{E+\omega-i\epsilon}\right]. \end{aligned} \tag{1.75}$$

We can compute this integral using the residue theorem. If $t_2 > t_1$ we can close the contour of integration in the lower half of the complex plane, where the exponential vanishes at infinity. If $t_1 > t_2$ we must instead close the contour in the upper half-plane. For both cases, we obtain:

$$G(t_1-t_2) = \frac{i}{2m\omega}e^{-i\omega|t_2-t_1|}. \tag{1.76}$$

2

Path integrals and scalar fields

In this chapter, we will write down path integrals for quantum field theory. Earlier, in Quantum Mechanics, we found that the canonical quantisation condition and the law of time-evolution,

$$\left[\hat{x}(t), \hat{p}(t)\right] = i, \qquad i\partial_t \left|\psi(t)\right\rangle = \hat{H} \left|\psi(t)\right\rangle \tag{2.1}$$

give rise to a path integral formulation for quantum mechanical correlation functions,

$$\begin{aligned} &\frac{\langle 0, +\infty | \, T\left(\hat{x}(t_1) \dots \hat{x}(t_n)\right) | 0, -\infty \rangle}{\langle 0, +\infty | \, 0, -\infty \rangle} \\ &\qquad = \lim_{\epsilon \to 0^+} \frac{\int \mathcal{D}x \, x(t_1) \dots x(t_n) \, e^{i \int dt L[x(t), \epsilon]}}{\int \mathcal{D}x \, e^{i \int dt L[x(t), \epsilon]}}. \end{aligned} \tag{2.2}$$

In quantum field theory, the canonical quantisation formalism postulates commutation relations, which for scalar fields read

$$\left[\phi(\vec{x}, t), \pi(\vec{y}, t)\right] = i\delta^3(\vec{x} - \vec{y}). \tag{2.3}$$

Comparing the commutation relations in QM and QFT, we observe that in passing from the former to the latter, the role of the time coordinate is played by the set of all four space-time coordinates x^μ and the role of the position $x(t)$ in a path is played by the value $\phi(x^\mu)$ of the field.

To formulate path integrals in QFT, we will heuristically try out a QM to QFT dictionary, which reads

$$\begin{aligned} t &\to x^\mu \,, \\ x(t) &\to \phi(x^\mu) \,, \\ \mathcal{D}x &\to \mathcal{D}\phi \,. \end{aligned} \tag{2.4}$$

We will not attempt to derive a path integral for correlation functions in QFT directly, starting from the canonical quantisation conditions as we did in QM.[1] Instead, we will write down correlation functions in quantum field theory in terms of path integrals applying the dictionary of Eqs. (2.4) to Eq. (2.2). We will test the validity of this heuristic approach by verifying that the QFT correlation functions (or, equivalently, the Feynman rules of perturbation theory) which we will derive in this way are the same as the ones we obtain with the canonical quantisation formalism of Part I.

[1] For scalar fields this is possible.

We will start with a field theory of a single scalar field $\phi(x)$. Motivated by our results for path integrals in quantum mechanics, we postulate a generating functional for the field,

$$Z[J] = \mathcal{N} \int \mathcal{D}\phi e^{i\int d^4x\left(\mathcal{L}+J(x)\phi(x)+i\epsilon\phi^2\right)}. \tag{2.5}$$

In Eq. (2.5), the Lagrangian has already been expressed as the space integral of a Lagrangian density,

$$L = \int d^3\vec{x}\mathcal{L}.$$

The role of the $i\epsilon\phi^2$, with $\epsilon \to 0^+$, will be, as in our QM treatment, to dissipate the contributions of excited states in expectation values of operators, yielding expectation values only in the ground state (the "vacuum" state). We will postulate that correlation functions in QFT are given by

$$\begin{aligned} G(x_1,\ldots,x_n) &\equiv \frac{\langle\Omega,+\infty|\,T\left(\hat{\phi}(x_1)\ldots\hat{\phi}(x_n)\right)|\Omega,-\infty\rangle}{\langle\Omega,+\infty|\,\Omega,-\infty\rangle} \\ &= \lim_{\epsilon\to 0^+}\frac{(-i)^n}{Z[0]}\left.\frac{\delta^n Z[J]}{\delta J(x_1)\ldots\delta J(x_n)}\right|_{J=0}. \end{aligned} \tag{2.6}$$

2.1 Real Klein–Gordon field

Let us now try to compute some correlation functions using the path integral of Eq. (2.5) and Eq. (2.6) in the simplest possible case of the real Klein–Gordon field. The Lagrangian density for a free real scalar field is

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2\right). \tag{2.7}$$

The action integral in the exponent of Eq. (2.5) is

$$\begin{aligned} S &= \int d^4x\left\{\frac{1}{2}\left(\partial_\mu\phi\partial^\mu\phi - m^2\phi^2\right) + i\epsilon\phi^2 + J\phi\right\} \\ &= \int d^4x\left\{\frac{1}{2}\left[\partial_\mu\left(\phi\partial^\mu\phi\right) - \phi\partial^2\phi - m^2\phi^2\right] + i\epsilon\phi^2 + J\phi\right\} \\ &= -\int d^4x\left\{\frac{1}{2}\phi\left[\partial^2 + m^2 - i\epsilon\right]\phi - J\phi\right\}. \end{aligned} \tag{2.8}$$

In the above, we performed an integration by parts, where we could discard a surface term. The path integral for this action

$$Z[J] = \mathcal{N}\int \mathcal{D}\phi e^{iS},$$

is a straightforward generalisation in four dimensions of the path-integral of the simple harmonic oscillator.

We can then compute this path integral by repeating the steps of our computation for the harmonic oscillator in QM. An important step was to change the path integral integration variable from a generic path $x(t)$ to its difference $y(t) = x(t) - x_{cl}(t)$ from the

solution of the classical equations of motion. We will do the same here, for the QFT path integral. We shift

$$\phi(x^\mu) \longrightarrow \phi(x^\mu) + \phi_{classical}(x^\mu).$$

We then find

$$\begin{aligned} Z[J] &= e^{-\frac{1}{2}\int d^4x d^4y J(x)\Delta_F(x-y)J(y)} \times \\ &\quad \mathcal{N}\int \mathcal{D}\phi e^{-\frac{i}{2}\int d^4x \phi\left(\partial^2+m^2-i\epsilon\right)\phi}\,, \end{aligned} \tag{2.9}$$

where we recognise the Feynman propagator of a free scalar field,

$$\Delta_F(x-y) = i\int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot(x-y)}}{k^2-m^2+i\epsilon}. \tag{2.10}$$

Exercise:

a) Write the Euler–Lagrange equations for the free real scalar field.

b) Evaluate the generating path integral for the free real scalar field working in Fourier space and following the analogous derivation of the simple harmonic oscillator.

c) You can also obtain the result of Eq. (2.9) directly, without resorting to a Fourier transform. Try it out!

d) Find the Fourier representation of the propagator.

e) Integrate the Fourier representation of the propagator over the energy using the Cauchy theorem. Pay attention to the conditions on the time variable in order to be able to use a closed contour of integration.

As we can easily compute, Eq. (2.6) gives for the one-, two-, three- and four-point correlation functions the following results:

$$G(x_1) = 0\,, \tag{2.11}$$

$$G(x_1,x_2) = \Delta_F(x_1-x_2)\,, \tag{2.12}$$

$$G(x_1,x_2,x_3) = 0\,, \tag{2.13}$$

$$\begin{aligned} G(x_1,x_2,x_3,x_4) &= \Delta_F(x_1-x_2)\Delta_F(x_3-x_4) \\ &\quad +\Delta_F(x_1-x_3)\Delta_F(x_2-x_4)+\Delta_F(x_1-x_4)\Delta_F(x_2-x_3)\,. \end{aligned} \tag{2.14}$$

The above results are in agreement with what we obtain in the canonical formalism, by applying Wick's theorem.

For interacting theories, the path integral will serve as a very convenient formalism which enables us to develop a method for deriving Feynman rules and performing perturbative expansions. Also for interacting theories, we will find that the path integral formalism for scalar fields reproduces the results of the canonical formalism.

There is a convenient graphical representation which facilitates computations of correlation functions, such as in deriving the correlation functions of Eqs. (2.11)–(2.14). In a graphical notation à la Feynman, we represent the propagator of Eq. (2.10) with a line,

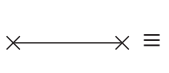
$$\overline{x \qquad y} \equiv \Delta_F(x-y). \tag{2.15}$$

The integral over one endpoint of a propagator times a source evaluated at this endpoint is represented graphically by



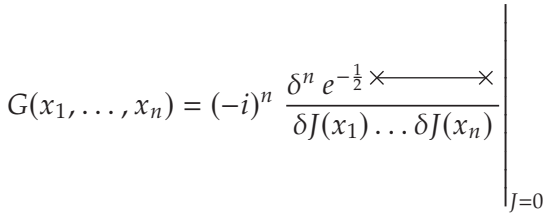
$$\equiv \int d^4x J(x)\Delta_F(x-y). \tag{2.16}$$

Finally, the integral over both endpoints of a propagator multiplied with sources at each endpoint is represented with



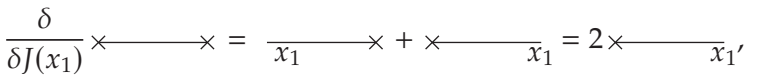
$$\equiv \int d^4x d^4y J(x)\Delta_F(x-y)J(y). \tag{2.17}$$

With this notation, Eq. (2.6) takes the form,



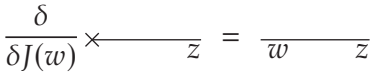
$$G(x_1,\ldots,x_n) = (-i)^n \left.\frac{\delta^n\, e^{-\frac{1}{2}}}{\delta J(x_1)\ldots\delta J(x_n)}\right|_{J=0} . \tag{2.18}$$

The above differentiations are straightforward with our graphical mnemonic. For example, we have



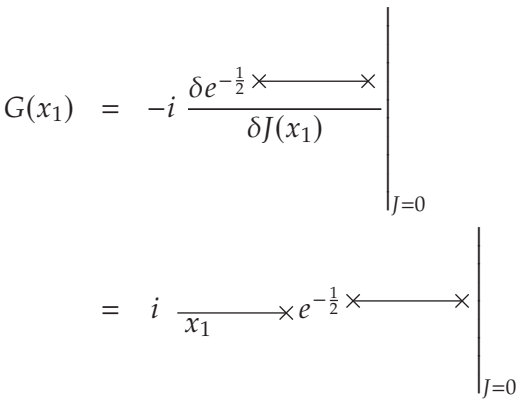
$$\frac{\delta}{\delta J(x_1)} \quad = \quad + \quad , \tag{2.19}$$

and



$$\frac{\delta}{\delta J(w)} \quad = \quad . \tag{2.20}$$

For the one-point correlation function, we find



$$\begin{aligned} G(x_1) &= -i \left.\frac{\delta e^{-\frac{1}{2}}}{\delta J(x_1)}\right|_{J=0} \\ &= i \quad e^{-\frac{1}{2}} \Big|_{J=0} \\ &= 0. \end{aligned} \tag{2.21}$$

For the two-point correlation function we have,

$$G(x_1,x_2) = (-i)^2 \left.\frac{\delta^2 e^{-\frac{1}{2}\times\longrightarrow\times}}{\delta J(x_1)\delta J(x_2)}\right|_{J=0}$$

$$= \frac{\delta}{\delta J(x_2)} \left(\overline{x_1}\times e^{-\frac{1}{2}\times\longrightarrow\times} \right)\Bigg|_{J=0}$$

$$= \left(\overline{x_1\quad x_2} - \overline{x_1}\times\quad\overline{\times\, x_2} \right) e^{-\frac{1}{2}\times\longrightarrow\times}\Bigg|_{J=0}$$

$$= \overline{x_1\quad x_2}, \tag{2.22}$$

and so on.

2.2 Functional integration and determinants of differential operators

So far, we have avoided carrying out any path integration, as all dependence on the sources $J(x)$ was a "constant" with respect to the integration over paths after a shift of the integration variable by the classical path. The path integral factor was eliminated by the normalisation in the denominator of the right-hand side of Eq. (2.6). Nevertheless, it will be useful to gain some insight into the meaning of the path integral itself. In what follows, we will see that the path integral has a meaning as a "determinant" of the differential operator which appears in the classical Euler–Lagrange equations. For this purpose, we will develop some additional simple mathematical tools.

We start with the familiar integral

$$\int_{-\infty}^{\infty} dx e^{-ax^2} = \sqrt{\frac{\pi}{a}}, \qquad a > 0. \tag{2.23}$$

For n variables, we then obtain

$$\int_{-\infty}^{\infty} dx_1 \dots dx_n e^{-\sum_{i=1}^n a_i x_i^2} = \frac{\pi^{\frac{n}{2}}}{(a_1 \dots a_n)^{\frac{1}{2}}}, \qquad a_i > 0. \tag{2.24}$$

We can define

$$A = \mathrm{diag}\,(a_1, a_2, \dots, a_n)\,, \tag{2.25}$$

$$x^T = (x_1, \dots, x_n). \tag{2.26}$$

We then rewrite the above integral as

$$\int_{-\infty}^{\infty} \left(\prod_i \frac{dx_i}{\sqrt{\pi}} \right) e^{-x^T A x} = \frac{1}{(\det A)^{\frac{1}{2}}}. \tag{2.27}$$

Let us now perform a transformation

$$x_i = R_{ij} y_j, \quad \text{or} \quad x = Ry. \tag{2.28}$$

The integral can be written as

$$\int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} \frac{dy_i}{\sqrt{\pi}} \right) (\det R)\, e^{-y^T (R^T A R) y} = \frac{1}{(\det A)^{\frac{1}{2}}}. \tag{2.29}$$

We define the matrix

$$B = R^T A R. \tag{2.30}$$

We can easily verify that B is symmetric:

$$B^T = (R^T A R)^T = R^T A^T \left(R^T\right)^T = R^T A R = B.$$

The determinant of B is then

$$\det B = \det(R^T A R) = (\det A)(\det R)^2.$$

We then find

$$\int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} \frac{dy_i}{\sqrt{2\pi}} \right) e^{-\frac{1}{2} y^T B y} = \frac{1}{(\det B)^{\frac{1}{2}}}, \tag{2.31}$$

where B is any real, positive definite (positive eigenvalues), symmetric matrix.

Let us finally take the limit $n \to \infty$. The exponent on the left-hand side is a double sum over the two indices of the $n \times n$ matrix B. In the limit, we can write:

$$\int \mathcal{D}\phi e^{-\frac{1}{2} \int \phi(x) \hat{A}(x) \phi(x)} = \frac{1}{\sqrt{\det \hat{A}}}, \tag{2.32}$$

which we can view as a definition for the determinant of a continuous differential operator.

We can also define the determinant of a Hermitian operator via path integration over complex fields. Consider

$$\begin{aligned} \frac{\pi}{a} &= \int_{-\infty}^{\infty} dx dy e^{-a\left(x^2+y^2\right)} \\ &= \int_{-\infty}^{\infty} dx dy e^{-a(x-iy)(x+iy)}. \end{aligned} \tag{2.33}$$

We define

$$x = \frac{z+z^*}{2}, \qquad y = \frac{z-z^*}{2i}.$$

We obtain

$$\int \frac{dz^* dz}{2\pi i} e^{-azz^*} = \frac{1}{a}. \tag{2.34}$$

Following the same steps as before, we can write the determinant of a Hermitian operator as:

$$\int \mathcal{D}\phi \mathcal{D}\phi^* e^{-\int d^4x \phi^*(x) \hat{A}(x) \phi(x)} = \frac{1}{\det \hat{A}}. \tag{2.35}$$

2.3 Path integrals and interacting fields

For a free scalar field, the computation of correlation functions with the path integral formalism was straightforward. The only integration required was for the evaluation of the Feynman propagator,

$$\Delta_F(x_1 - x_2) \equiv \overline{x_1 \qquad x_2}.$$

Using the Feynman propagator as a kernel of a generating functional, we could obtain Green's functions by taking functional derivatives with respect to sources, without carrying out explicit functional integrations.

In general, for field theories with interactions, it is not possible to compute correlation functions exactly in four spacetime dimensions. However, the path integral formalism can be very useful, as it makes perturbative expansions rather straightforward; one can compute a few[2] terms in the series with advanced but established computational methods. The terms of the perturbative expansion will be represented graphically, in terms of Feynman diagrams. In this section, we will see how they arise with the path integral approach. The results will be the same as with the canonical formalism and we will verify that this is so in some cases.

Let us consider a real scalar field which can interact with itself,

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_I, \tag{2.36}$$

where

$$\mathcal{L}_0 = -\frac{1}{2}\phi(x)\left(\partial^2 + m^2 - i\epsilon\right)\phi(x), \tag{2.37}$$

and

$$\mathcal{L}_I = -\frac{\lambda}{4!}\phi(x)^4. \tag{2.38}$$

According to our previous conclusions, we can obtain the correlation functions of this theory by taking functional derivatives of the following path integral,

$$Z[J] = \mathcal{N}\int \mathcal{D}\phi e^{i\int d^4x\left\{\mathcal{L}_0 + \phi(x)J(x) + \mathcal{L}_I\right\}}. \tag{2.39}$$

We consider the case of a small coupling constant $\lambda \ll 1$ and assume this condition to suffice for a perturbative expansion around $\lambda = 0$. We separate the action into a "free" and an "interacting" part,

$$S = S_0 + S_I, \tag{2.40}$$

where the "free" part includes the sources as well as the quadratic terms of the Lagrangian in the field ϕ,

$$S_0[J,\phi] = \int d^4x\left\{\mathcal{L}_0 + \phi(x)J(x)\right\}, \tag{2.41}$$

and the "interacting" part is

$$S_I[\phi] = \int d^4x\,\mathcal{L}_I. \tag{2.42}$$

We were able to compute correlation functions exactly for $\lambda = 0$. We will attempt to build solutions for small λ as perturbations of the known solutions at $\lambda = 0$, which we hope will

[2] Typically three.

furnish the leading contributions to the correlation functions of the interacting theory. A Taylor series expansion of the path integral in λ gives

$$\begin{aligned} Z[J] &= \mathcal{N}\int \mathcal{D}\phi e^{iS_0[J,\phi]+iS_I[\phi]} = \mathcal{N}\int \mathcal{D}\phi \sum_{n=0}^{\infty}\frac{(iS_I[\phi])^n}{n!}e^{iS_0[J,\phi]} \\ &= \mathcal{N}\int \mathcal{D}\phi \sum_{n=0}^{\infty}\frac{\left[i\int d^4y\,\left(\frac{-\lambda}{4!}\phi(y)^4\right)\right]^n}{n!}e^{iS_0[J,\phi]} \\ &= \mathcal{N}\sum_{n=0}^{\infty}\frac{\left[i\int d^4y\left(\frac{-\lambda}{4!}\left(\frac{\delta}{i\delta J(y)}\right)^4\right)\right]^n}{n!}\int \mathcal{D}\phi e^{iS_0[J,\phi]}. \end{aligned} \tag{2.43}$$

In the last step, we used the fact that we can generate powers of ϕ in the integrand of the path integral by acting with functional derivatives. In a short notation, we "sum back" the series into an exponential of functional derivatives. The result for the generating functional is compactly written in the form,

$$Z[J] = \mathcal{N}\, e^{i\int d^4y\mathcal{L}_I\left(\frac{1}{i}\frac{\delta}{\delta J(y)}\right)}\int \mathcal{D}\phi e^{iS_0[J,\phi]}. \tag{2.44}$$

Let us stress that the exponential of functional derivatives is only meant as a short notation. The precise meaning of it is defined through the Taylor series in the last line of Eq. (2.43). On the right-hand side of Eq. (2.44), we find the path integral of the free field theory,

$$\begin{aligned} Z_0[J] &\equiv \mathcal{N}_0\int \mathcal{D}\phi e^{iS_0[J,\phi]} \\ &= \mathcal{N}_0\int \mathcal{D}\phi e^{iS_0[0,\phi]}\, e^{-\frac{1}{2}\times\!\!\longrightarrow\!\!\times}. \end{aligned} \tag{2.45}$$

Substituting Eq. (2.45) into Eq. (2.44) and defining an overall normalisation,

$$\mathcal{N}' = \mathcal{N}\int \mathcal{D}\phi e^{iS_0[0,\phi]},$$

we obtain

$$Z[J] = \mathcal{N}'\, e^{i\int d^4y\mathcal{L}_I\left(\frac{1}{i}\frac{\delta}{\delta J(y)}\right)}\, e^{-\frac{1}{2}\times\!\!\longrightarrow\!\!\times}. \tag{2.46}$$

We fix the normalisation of the generating functional by requiring that the amplitude for a vacuum-to-vacuum transition over infinitely long times is unity,

$$Z[0] = 1.$$

We then obtain,

$$Z[J] = \frac{e^{i\int d^4y\mathcal{L}_I\left(\frac{1}{i}\frac{\delta}{\delta J(y)}\right)}\, e^{-\frac{1}{2}\times\!\!\longrightarrow\!\!\times}}{e^{i\int d^4y\mathcal{L}_I\left(\frac{1}{i}\frac{\delta}{\delta J(y)}\right)}\, e^{-\frac{1}{2}\times\!\!\longrightarrow\!\!\times}}\Bigg|_{J=0}. \tag{2.47}$$

We will compute explicitly the generating functional $Z[J]$ of Eq. (2.47) through order $O(\lambda^2)$. Let us consider the numerator of the right-hand side of Eq. (2.47),

$$\text{Num}[J] = e^{i\int d^4y \mathcal{L}_I\left(\frac{1}{i}\frac{\delta}{\delta J(y)}\right)} e^{-\frac{1}{2}} \times\!\!\longleftarrow\!\!\times$$

$$= \left(1 - \frac{i\lambda}{4!}\int d^4z \frac{\delta^4}{\delta J(z)^4}\right) e^{-\frac{1}{2}} \times\!\!\longleftarrow\!\!\times + O(\lambda^2). \tag{2.48}$$

After carrying out the four functional differentiations, we obtain

$$\text{Num}[J] = e^{-\frac{1}{2}} \times\!\!\longleftarrow\!\!\times \left(1 - i\frac{\lambda}{4!}\Bigg[\;\times\!\!\times\; - 6 \times\!\!\longleftarrow\!\!\times$$

$$+ 3 \quad \text{z} \quad \Bigg]\Bigg) + O(\lambda^2). \tag{2.49}$$

The denominator of $Z[J]$ in Eq. (2.47) is given by

$$\text{Den} = \text{Num[J]}|_{J=0} = 1 - i\frac{\lambda}{4!}\Bigg[3 \quad \text{z} \quad \Bigg] + O(\lambda^2). \tag{2.50}$$

We can then compute the generating functional, dividing our results for numerator and denominator and expanding consistently in λ. We find

$$Z[J] = \frac{\text{Num}[J]}{\text{Den}}$$

$$= e^{-\frac{1}{2}} \times\!\!\longleftarrow\!\!\times \left(1 - i\frac{\lambda}{4!}\Bigg[\;\times\!\!\times\; - 6 \times\!\!\longleftarrow\!\!\times \Bigg]\right)$$

$$+ O(\lambda^2). \tag{2.51}$$

Notice the effect of dividing by the denominator, Den, which guarantees the normalisation $Z[0] = 1$, on our result of Eq. (2.51). After expanding in λ, the denominator cancels all vacuum graphs which also appear in the numerator Num$[J]$. All graphs which remain in the final result are connected to a source.

We can now compute Green's functions perturbatively, through order $O(\lambda)$, using the master formula

$$G(x_1, \ldots, x_n) = (-i)^n \left.\frac{\delta^n Z[J]}{\delta J(x_1)\ldots\delta J(x_n)}\right|_{J=0}, \tag{2.52}$$

and substituting the perturbative expansion of Eq. (2.51). We find,

$$G(x_1) = 0\,, \tag{2.53}$$

$$G(x_1, x_2) = \quad - \frac{i\lambda}{2} \quad + O(\lambda^2)\,, \tag{2.54}$$

$$G(x_1, x_2, x_3) = 0\,, \tag{2.55}$$

$$G(x_1, x_2, x_3, x_4) \quad = \quad + \quad + \quad$$

$$-i\frac{\lambda}{2}\Bigg(\quad + \quad + \quad$$

$$\quad + \quad + \quad \Bigg)$$

$$-i\lambda \quad + O(\lambda^2)\,, \tag{2.56}$$

and so on. We observe the same Feynman diagram expressions as one obtains with perturbation theory in the canonical quantisation formalism.

Notice that Eq. (2.52) generates *all* Feynman diagrams for the four-point Green's function on the right-hand side of Eq. (2.56). These comprise diagrams which do not connect all four external points x_1, x_2, x_3, and x_4, along with the one diagram (the last graph on the right-hand side of Eq. (2.56)) which is fully connected. Indeed, we can recast Eq. (2.52) as

$$G(x_1, x_2, x_3, x_4) = -i\lambda \quad + G(x_1, x_2)G(x_3, x_4) + G(x_1, x_3)G(x_2, x_4) + G(x_1, x_4)G(x_2, x_3) + O(\lambda^2)\,. \tag{2.57}$$

As we know from scattering theory and the LSZ formula, the disconnected graphs, which belong to products of lower-point Green's functions, do not contribute to the scattering amplitudes. *Remark.* More precisely, after LSZ amputation the products $G(x_1, x_2)G(x_3, x_4)$ give the forward, non-scattering part (the "1" in $S = 1 + iT$, momentum-conserving delta functions for each particle separately); they do not contribute to the transition amplitude T for $2 \to 2$ scattering (Peskin–Schroeder sec. 4.6).

2.3.1 Generating functional of connected Feynman diagrams

There is a simple and general method to generate directly fully connected Feynman diagrams, by differentiating the logarithm of the path integral $\log Z[J]$, rather than differentiating $Z[J]$. Let us define a functional $W[J]$ via

$$Z[J] = e^{iW[J]}. \tag{2.58}$$

For the one-point function, we find

$$G(x_1) = (-i)\frac{\delta Z[J]}{\delta J(x_1)}\bigg|_{J=0} = \frac{\delta W[J]}{\delta J(x_1)}\bigg|_{J=0}. \tag{2.59}$$

For the two-point correlation function, we obtain

$$\begin{aligned} G(x_1,x_2) &= (-i)^2 \frac{\delta^2 Z[J]}{\delta J(x_1)\delta J(x_2)}\bigg|_{J=0} \\ &= -i\frac{\delta^2 W[J]}{\delta J(x_1)\delta J(x_2)}\bigg|_{J=0} + G(x_1)\,G(x_2)\,. \end{aligned} \tag{2.60}$$

Rearranging, we find that acting with two derivatives on the logarithm $W[J]$ of the path integral $Z[J]$ has the effect of removing the product of one-point correlation functions, leaving only the fully connected diagrams,

$$\begin{aligned} -i\frac{\delta^2 W[J]}{\delta J(x_1)\delta J(x_2)}\bigg|_{J=0} &= G(x_1,x_2) - G(x_1)G(x_2) \\ &\equiv G_{\text{conn.}}(x_1,x_2). \end{aligned} \tag{2.61}$$

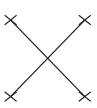
Similarly, for the three-point function we find

$$\begin{aligned} &(-i)^2\frac{\delta^3 W[J]}{\delta J(x_1)\delta J(x_2)\delta J(x_3)}\bigg|_{J=0} = G_{\text{conn.}}(x_1,x_2,x_3) \\ &\equiv G(x_1,x_2,x_3) - G_{\text{conn.}}(x_1,x_2)\,G(x_3) - G_{\text{conn.}}(x_2,x_3)\,G(x_1) \\ &-G_{\text{conn.}}(x_1,x_3)\,G(x_2) - G(x_1)\,G(x_2)\,G(x_3)\,, \end{aligned} \tag{2.62}$$

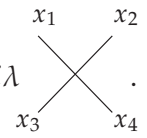
and so on. The general result reads,

$$G_{\text{conn.}}(x_1,x_2,\ldots,x_n) = (-i)^{n-1}\frac{\delta^n W[J]}{\delta J(x_1)\delta J(x_2)\ldots\delta J(x_n)}\bigg|_{J=0}. \tag{2.63}$$

In the case of the ϕ^4 scalar theory, we have

$W[J] = \frac{i}{2}$ $+\frac{\lambda}{4}$ $-\frac{\lambda}{4!}$  $+\mathcal{O}(\lambda^2).$ (2.64)

Applying Eq. (2.63), we find indeed only the connected graph for the four-point function,

$$\begin{aligned} G_{\text{conn.}}(x_1,x_2,x_3,x_4) &= (-i)^3\frac{\delta^4 W[J]}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)}\bigg|_{J=0} \\ &= -i\lambda \end{aligned}$$ 

. (2.65)

2.4 Fermionic path integrals

We have discussed the path integral for scalar fields which are bosonic. To describe fermions, we need to formulate path integration over anti-commuting functions.

We will introduce a new kind of numbers, the *Grassmann numbers*, which are defined to obey

$$\{c_i, c_j\} = c_i c_j + c_j c_i = 0. \tag{2.66}$$

A Grassmann number c is also defined to commute with a regular number x,

$$x\, c = c\, x\,. \tag{2.67}$$

From the definition of Eq. (2.66), we have

$$c_i^2 = 0. \tag{2.68}$$

Therefore, any function constructed out of Grassmann variables is at most linear in any of these variables.

Exercise:*Show that the product of two Grassmann variables $a = c_1 c_2$ commutes with a Grassmann variable c_3,*

$$[a, c_3] = 0.$$

Exercise: *Show that linear combinations of Grassmann variables are also Grassmann variables.*

We will now develop our calculus for functions of Grassmann variables, defining derivatives and integration. This will lead us to a definition of a path integration over Grassmann fields, which we will postulate to describe fermions.

We start by defining a derivative similarly to derivatives of commuting numbers,

$$\frac{\partial c_i}{\partial c_j} = \delta_{ij}. \tag{2.69}$$

We will also assume that we can apply a chain rule for differentiating functions of Grassmann variables.

Our definition of Grassmann differentiation, as it stands now, is a bit naive. When it is used to differentiate functions of multiple Grassmann variables it leads to ambiguities. To obtain unambiguous results, we must introduce a further rule.

Let us consider a concrete example of a function of two Grassmann variables c_1, c_2. The general form of the function is at most linear in each variable, and it is given by

$$f(c_1, c_2) = a_0 + a_1 c_1 + a_2 c_2 + a_{12} c_1 c_2, \tag{2.70}$$

where the coefficients a_0, a_1, a_2, a_{12} are regular commuting numbers. Ambiguities arise because a reordering of Grassmann variables in a product may give rise to minus signs. For example, we can decide to commute the c_2 variable to the left of the c_1 Grassmann variable in the last term of the right-hand side of Eq. (2.70), obtaining an equivalent expression:

$$f(c_1, c_2) = a_0 + a_1 c_1 + a_2 c_2 - a_{12} c_2 c_1. \tag{2.71}$$

A naive application of Eq. (2.69) could yield two different results,

$$\frac{\partial f}{\partial c_2} = a_2 + a_{12} c_1, \quad \text{or} \quad \frac{\partial f}{\partial c_2} = a_2 - a_{12} c_1.$$

We will need to fix this ambiguity before we take derivatives.

We will define a "left" derivative with respect to c_2 which acts in the following way,

$$\begin{aligned}\frac{\partial^L f}{\partial c_2} &= \frac{\partial^L}{\partial c_2}(a_0 + a_1 c_1 + a_2 c_2 + a_{12} c_1 c_2) \\ &= \frac{\partial^L}{\partial c_2}(a_0 + a_1 c_1 + a_2 c_2 - a_{12} c_2 c_1) \\ &= a_2 - a_{12} c_1. \end{aligned} \tag{2.72}$$

The meaning of "left" (as the superscript denotes) is that we must bring the variable which we differentiate to the left of any other Grassmann variable in a product, anticommuting by means of Eq. (2.66), before we eliminate the variable by differentiating with the rule of Eq. (2.69).

Analogously, we could have defined a "right" derivative, where we always anti-commute a Grassmann variable to the right of any other variable before we eliminate it. In our example, we have:

$$\frac{\partial^R f}{\partial c_2} = a_2 + a_{12} c_1. \tag{2.73}$$

Throughout these notes, we will always use left derivatives, unless it is explicitly stated otherwise. From now on, we will drop the superscript L on derivatives of Grassmann variables and we will assume implicitly that the derivative is a left one.

Exercise:*Prove that*

$$\left\{ c_i, \frac{\partial}{\partial c_j} \right\} = \delta_{ij}, \tag{2.74}$$

and

$$\left\{ \frac{\partial}{\partial c_i}, \frac{\partial}{\partial c_j} \right\} = 0. \tag{2.75}$$

Our next step is to define an integration over Grassmann variables c_i. As functions of Grassmann variables are linear, we only need to know two integrals,

$$\int dc\ 1 = ?$$

and

$$\int dc\ c = ?$$

We will set

$$\int dc\ 1 = 0. \tag{2.76}$$

In addition, we will set

$$\int dc\ c = x,$$

where x is a usual commuting number of a fixed value. As the only other possible integral, Eq. (2.76), vanishes, the actual value of x can be absorbed into a redefinition of the integration measure. It is convenient to take

$$\int dc\ c = 1. \tag{2.77}$$

For integration, as well as for differentiation, we will need to fix sign ambiguities when integrating over functions of multiple Grassmann variables. Our rule will be to order

products of Grassmann variables so that the Grassmann variable that we integrate over appears on the left. For example, in integrating a product of two Grassmann variables, we have

$$\int dc_1 c_2 c_1 = -\int dc_1 c_1 c_2 = -c_2.$$

We can now reveal a striking fact about our definitions of differentiation and integration on Grassmann variables: differentiation and integration are identical operations,

$$\begin{aligned}\int dc_2 f &= \int dc_2 \left(a_0 + a_1 c_1 + a_2 c_2 + a_{12} c_1 c_2\right) \\ &= \int dc_2 \left(a_0 + a_1 c_1 + a_2 c_2 - a_{12} c_2 c_1\right) \\ &= a_2 - a_{12} c_1 \\ &= \frac{\partial f}{\partial c_2}. \end{aligned} \tag{2.78}$$

Multiple integrals over Grassmann variables can be used to express determinants. This property will turn out to be very useful in defining a path integral for non-abelian gauge theories. Consider a general integral

$$\int dc_1 dc_2 \ldots dc_n f(c_1, c_2, \ldots, c_n),$$

and perform a linear transformation

$$c_i = M_{ij} b_j. \tag{2.79}$$

We shall have

$$\int dc_1 dc_2 \ldots dc_n f(c_i) = (\text{Jacobian}) \int db_1 db_2 \ldots db_n f\left(M_{ij} b_j\right). \tag{2.80}$$

What is the Jacobian? In the case of a double integral, we have:

$$\begin{aligned}\int dc_1 dc_2 c_1 c_2 &= (\text{Jacobian}) \int db_1 db_2 \left(M_{11} b_1 + M_{12} b_2\right)\left(M_{21} b_1 + M_{22} b_2\right) \\ &= (\text{Jacobian}) \int db_1 db_2 \left(M_{11} M_{22} b_1 b_2 + M_{12} M_{21} b_2 b_1\right) \\ &= (\text{Jacobian}) \int db_1 db_2 b_1 b_2 \left(M_{11} M_{22} - M_{12} M_{21}\right) \\ &= (\text{Jacobian}) \det(M) \int db_1 db_2 b_1 b_2 \\ &\rightsquigarrow \text{Jacobian} = \frac{1}{\det(M)}. \end{aligned} \tag{2.81}$$

Exercise: *Prove that the same result holds for multiple integrals of arbitrary dimensions.*
If we recall that Grassmann integration is in reality differentiation, it is not surprising that the Jacobian of the transformation on Grassmann variables is the inverse of what emerges in integrations of normal commuting variables.
Exercise: *Consider the complex linear combinations*

$$y = \frac{c_1 + i c_2}{\sqrt{2}}, \quad \bar{y} = \frac{c_1 - i c_2}{\sqrt{2}},$$

of two real Grassmann variables c_1, c_2. Show that:

$$\{y, \bar{y}\} = 0,$$
$$\int dc_1 dc_2 f(c_1, c_2) = -i \int dy d\bar{y} f\left(\frac{y+\bar{y}}{\sqrt{2}}, \frac{y-\bar{y}}{\sqrt{2}i}\right).$$

After we have defined integration over Grassmann variables, we can now study multiple integrals over exponentials of Grassmann variables. This will be an intermediate step in formulating a path integral for fermionic fields, as an analogous step emerged in the formulation of a path integral for scalar bosonic fields. Let us start with two independent vectors of Grassmann variables $x^T = (x_1, x_2)$ and $y^T = (y_1, y_2)$. We have

$$x^T y = x_1 y_1 + x_2 y_2, \tag{2.82}$$

and

$$\begin{aligned} \left(x^T y\right)^2 &= (x_1 y_1 + x_2 y_2)(x_1 y_1 + x_2 y_2) \\ &= x_1 y_1 x_2 y_2 + x_2 y_2 x_1 y_1 \\ &= 2 x_1 y_1 x_2 y_2. \end{aligned} \tag{2.83}$$

We can easily see that

$$\left(x^T y\right)^n = 0, \qquad n > 2. \tag{2.84}$$

Then, for the exponential, we have

$$\begin{aligned} e^{-x^T y} &= 1 - x^T y + \frac{1}{2}(x^T y)^2 - \frac{1}{3!}(x^T y)^3 + \ldots \\ &= 1 - x_1 y_1 - x_2 y_2 + x_1 y_1 x_2 y_2. \end{aligned} \tag{2.85}$$

Therefore, using this expansion, we have

$$\begin{aligned} \int dx_1 dy_1 dx_2 dy_2 e^{-x^T y} &= \int dx_1 dy_1 dx_2 dy_2 \left[1 - \left(x_1 y_1 + x_2 y_2\right) + x_1 y_1 x_2 y_2\right] \\ &= 1. \end{aligned} \tag{2.86}$$

We can now perform linear transformations on both x and y,

$$x \to M x',$$
$$y \to N y'.$$

We find

$$\begin{aligned} 1 &= \int dx dy e^{-x^T y} = \det(M^T)^{-1} \det(N)^{-1} \int dx dy e^{-x^T M^T N y} \\ &= \det(M^T N)^{-1} \int dx dy e^{-x^T M^T N y}. \end{aligned} \tag{2.87}$$

Defining $A = M^T N$, we obtain

$$\int dx dy e^{-x^T A y} = \det(A). \tag{2.88}$$

Exercise: *Prove the above for* $x^T = (x_1, x_2, x_3)$ *and* $y^T = (y_1, y_2, y_3)$*. Generalise to arbitrary dimensionality.*
Recall that for normal commuting variables we have

$$\int dx dx^\dagger e^{-x^\dagger A x} \sim \frac{1}{\det A}. \tag{2.89}$$

Let us now define a path integral over Grassmann variables by considering an infinite number of them and taking the continuum limit. This path integral will give correlation functions for the fermionic field ψ by taking functional derivatives with respect to sources, similarly to what we have seen in the scalar field case. It will differ however from the Green's functions of the bosonic field, in that the order of differentiations will matter. Different orderings will yield minus signs, as expected from Fermi–Dirac statistics. Let us consider the Lagrangian of a free Dirac fermion,

$$\mathcal{L} = \bar{\psi}\left(i\partial\!\!\!/ - m\right)\psi. \tag{2.90}$$

We can write a generating functional,

$$Z_0\left[a, \bar{a}\right] = \mathcal{N} \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{i\int d^4x \left[\mathcal{L} + \bar{a}\psi + \bar{\psi}a\right]}. \tag{2.91}$$

In this path integral the integration variables $\psi, \bar{\psi}$ and the sources $a, \bar{a}$ are all independent Grassmann functions. The constant $\mathcal{N}$ is fixed as usual by requiring that

$$Z_0\left[0, 0\right] = \langle 0|\, 0\rangle = 1.$$

Everything works in an analogous way as in the case of the path integral for the scalar field. For a theory without interactions, we can "complete the square" and compute the generating functional explicitly if we shift the fields $\psi, \bar{\psi}$ by a function corresponding to their classical value at which the action is stationary. We need the inverse of the Dirac wave operator,

$$(i\partial\!\!\!/ - m\mathbf{1})S(x-y) = i\,\delta^4\left(x-y\right)\mathbf{1}. \tag{2.92}$$

As for the scalar field and the harmonic oscillator, we can write a Fourier representation,

$$S(x-y) = i\int \frac{d^4k}{(2\pi)^4} \frac{k\!\!\!/ + m}{k^2 - m^2 + i\epsilon} e^{-ik\cdot(x-y)}. \tag{2.93}$$

Exercise:*Write the Euler–Lagrange equations for the action with sources* $a, \bar{a}$*. Solve these equations in Fourier space.* The values of $\psi, \bar{\psi}$ at which the action is stationary are now given by

$$\psi_{cl} = i\int d^4y S(x-y)a(y),$$

$$\bar{\psi}_{cl}(x) = i\int d^4y\, \bar{a}(y)S(y-x),$$

and the change of variables

$$\psi = \psi_{cl} + \eta, \qquad \bar{\psi} = \bar{\psi}_{cl} + \bar{\eta},$$

yields

$$Z_0[a, \bar{a}] = e^{-\int d^4x d^4y \bar{a}(x)S(x-y)a(y)}. \tag{2.94}$$

We obtain expectation values of time-ordered products of field operators from the generating functional via

$$\langle 0,+\infty|\, T\ldots\psi_i(x)\ldots\bar{\psi}_j(y)\ldots|0,-\infty\rangle = \ldots\frac{1}{i}\frac{\delta}{\delta\bar{a}_i(x)}\ldots i\frac{\delta}{\delta a_j(y)}\ldots Z_0[a,\bar{a}], \tag{2.95}$$

where the indices i, j are spinor indices. **Exercise:***Calculate the expectation values*

- $\langle 0,+\infty|\, T\psi_i(x_1)\bar{\psi}_j(x_2)\,|0,-\infty\rangle$
- $\langle 0,+\infty|\, T\psi_i(x_1)\bar{\psi}_j(x_2)\psi_k(x_3)\bar{\psi}_l(x_4)\,|0,-\infty\rangle$

and compare them with your results from canonical quantisation. We can also develop a perturbative expansion by faithfully repeating the steps we performed in the ϕ^4 scalar field theory. It is not hard to convince ourselves that a completely analogous formula should be valid here for the generating functional when an interaction term $\mathcal{L}_I$ is present in the Lagrangian:

$$Z[a,\bar{a}] = \frac{e^{i\int d^4z\,\mathcal{L}_I\left(i\frac{\delta}{\delta a(z)},\frac{1}{i}\frac{\delta}{\delta\bar{a}(z)}\right)}Z_0[a,\bar{a}]}{e^{i\int d^4z\,\mathcal{L}_I\left(i\frac{\delta}{\delta a(z)},\frac{1}{i}\frac{\delta}{\delta\bar{a}(z)}\right)}Z_0[a,\bar{a}]\Big|_{a=\bar{a}=0}}. \tag{2.96}$$

Despite the similarities in the appearance of the formulae, there are also very important differences which are encoded in the Grassmann algebra of the functions which we integrate over. It is very important to remember that functional derivatives anticommute, generalising the result that we found for the derivatives of discrete Grassmann variables:

$$\left\{\frac{\delta}{\delta a(x)},\frac{\delta}{\delta a(y)}\right\} = \left\{\frac{\delta}{\delta\bar{a}(x)},\frac{\delta}{\delta\bar{a}(y)}\right\} = \left\{\frac{\delta}{\delta a(x)},\frac{\delta}{\delta\bar{a}(y)}\right\} = 0. \tag{2.97}$$

We should also remember that these derivatives are, by convention, left derivatives, and that the order in which the sources appear in the integrand matters.
Exercise: *Consider a theory with a fermion, a real scalar and a Yukawa interaction $\lambda\bar{\psi}\psi\phi$. Calculate the fermion and scalar propagators through $O(\lambda^2)$. Derive the Feynman diagrams contributing to the scattering of four fermions. Derive the Feynman diagrams contributing to the scattering of two fermions and two scalars.*[3]

[3] This exercise is very important for checking your understanding of the material covered so far. Please spend as much time as needed until you get it right.

3

Non-abelian gauge theories

We will now discuss quantum field theories with a gauge symmetry. Realistic theories such as QED, QCD and the full Standard Model are all symmetric under local gauge transformations.

3.1 Gauge invariance and QED

We shall start with the familiar case of Quantum Electrodynamics. We will see how the QED Lagrangian emerges from extending the Dirac Lagrangian of free electrons, by requiring that the "global" symmetry of $U(1)$ transformations found in the Dirac Lagrangian becomes "local".

The Dirac Lagrangian for a fermion field ψ is

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(x)\,(i\partial\!\!\!/ - m)\,\psi(x). \tag{3.1}$$

This is clearly invariant under a "global" transformation,

$$\psi(x) \to \psi'(x) = \exp(i\,g\,\theta)\,\psi(x), \tag{3.2}$$

where the parameter θ is constant $\left(\frac{\partial\theta}{\partial x} = 0\right)$. Under a local gauge transformation,

$$U(x) = \exp\left(i\,g\,\theta(x)\right), \tag{3.3}$$

the free Dirac Lagrangian is not invariant:

$$\bar{\psi}\partial\!\!\!/\psi \to \bar{\psi}\partial\!\!\!/\psi + \bar{\psi}e^{-ig\theta}\left[\partial\!\!\!/ e^{ig\theta}\right]\psi. \tag{3.4}$$

We will now modify the Dirac Lagrangian in order to make it invariant under *local* gauge transformations with $\theta = \theta(x)$. To achieve this goal, we must introduce another field A^μ, which we will recognise as the photon field.

The term which causes the Dirac Lagrangian not to be invariant under *local* transformations is the derivative term,

$$\partial\!\!\!/\psi \to \partial\!\!\!/\psi' = e^{i\,g\,\theta}\left[\partial\!\!\!/ + ig\,(\partial\!\!\!/\theta)\right]\psi.$$

We will substitute the simple derivative ∂_μ with a covariant derivative D_μ which transforms more conveniently. Specifically, we look for a covariant derivative which, under a *local* gauge transformation, transforms as:

$$D\!\!\!/\psi \to U(x)D\!\!\!/\psi. \tag{3.5}$$

To obtain a covariant derivative, we add to the usual derivative a new function (field):

$$D_\mu = \partial_\mu - igA_\mu(x). \tag{3.6}$$

The new field $A_\mu(x)$ must be a vector field, as it must have the same Lorentz transformation as the derivative ∂_μ which is a vector. The field A_μ will also have its own gauge transformation,

$$A_\mu \to A'_\mu,$$

which we will define in such a manner that we obtain the covariant derivative transformation of Eq. (3.5). We need:

$$\begin{aligned}
& D_\mu\psi(x) \to D'_\mu\psi' = U(x)D_\mu\psi \\
\rightsquigarrow\ & \left(\partial_\mu - igA'_\mu\right)\left(U(x)\psi\right) = U(x)\left(\partial_\mu - igA_\mu\right)\psi \\
\rightsquigarrow\ & U(x)\partial_\mu\psi + \left[\partial_\mu U(x)\right]\psi - igA'_\mu U(x)\psi = U(x)\partial_\mu\psi - igA_\mu U(x)\psi \\
\rightsquigarrow\ & A'_\mu = A_\mu - \frac{i}{g}U^{-1}(x)\partial_\mu U(x).
\end{aligned} \tag{3.7}$$

The covariant derivative transforms as:

$$\begin{aligned}
& D_\mu \to D'_\mu = \partial_\mu - igA'_\mu \\
= \ & \partial_\mu - ig\left(A_\mu - \frac{i}{g}U^{-1}\partial_\mu U\right) \\
= \ & \partial_\mu - igA_\mu - U^{-1}\left(\partial_\mu U\right) \\
= \ & \partial_\mu - igA_\mu + U\left(\partial_\mu U^{-1}\right) \\
= \ & U(x)\left(\partial_\mu - igA_\mu\right)U^{-1}(x).
\end{aligned} \tag{3.8}$$

Therefore:

$$D_\mu \to D'_\mu = U(x)D_\mu U^{-1}(x). \tag{3.9}$$

We can now replace the free Lagrangian of the spin-1/2 field with a new Lagrangian which is also gauge invariant:

$$\begin{aligned}
& \mathcal{L}' = \bar\psi\left[i\not{D} - m\right]\psi \\
& \to \bar\psi U^{-1}U\left[i\not{D} - m\right]U^{-1}U\psi \\
& = \bar\psi\left[i\not{D} - m\right]\psi.
\end{aligned}$$

If A_μ is a physical field, we need to introduce a kinetic term in the Lagrangian for it. We will insist on constructing a fully gauge-invariant Lagrangian. For this purpose, we can use the covariant derivative as a building block. Consider the gauge transformation of the product of two covariant derivatives:

$$\begin{aligned}
& D_\mu D_\nu \to D'_\mu D'_\nu = UD_\mu U^{-1}UD_\nu U^{-1} \\
& = UD_\mu D_\nu U^{-1}.
\end{aligned}$$

This is not a gauge-invariant object. Now let us look at the commutator:

$$\begin{aligned}
& \left[D_\mu, D_\nu\right] \to \left[D'_\mu, D'_\nu\right] \\
& = U\left[D_\mu, D_\nu\right]U^{-1}.
\end{aligned} \tag{3.10}$$

This is gauge invariant. To convince ourselves, we write the commutator explicitly:

$$\begin{aligned} &\left[D_\mu, D_\nu\right] = \left(\partial_\mu - igA_\mu\right)\left(\partial_\nu - igA_\nu\right) - \left[\mu \leftrightarrow \nu\right] \\ = \;& \partial_\mu\partial_\nu - ig\left(\partial_\mu A_\nu\right) - igA_\nu\partial_\mu - igA_\mu\partial_\nu + (ig)^2 A_\mu A_\nu - \left[\mu \leftrightarrow \nu\right] \\ = \;& -ig\left[\partial_\mu A_\nu - \partial_\nu A_\mu\right]. \end{aligned} \tag{3.11}$$

Inserting Eq. (3.11) into Eq. (3.10), we find that the commutator of covariant derivatives (in the abelian $U(1)$ case) is gauge invariant. We have also found that it is proportional to the field strength tensor of the gauge (photon) field:

$$F_{\mu\nu} = \frac{i}{g}\left[D_\mu, D_\nu\right] = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{3.12}$$

We now have invariant terms for a Lagrangian with an "electron" and a "photon" field. The Lagrangian for QED reads

$$\mathcal{L}_{\text{QED}} = \bar{\psi}\left(i\not{D} - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{3.13}$$

Exercise: *Find the Noether current and conserved charge due to the invariance under the $U(1)$ gauge transformation of the QED Lagrangian.*
Exercise: *Think of at least five more operators that one can add to the QED Lagrangian without spoiling gauge invariance. There is an infinite number of them. Which of them do not have mass dimension four? As we shall see, operators of higher dimension spoil renormalisability.*

3.2 Non-abelian (global) $SU(N)$ transformations

Lagrangians for theories with more than one field may be symmetric under more complicated transformations than a $U(1)$ symmetry. In particular, a symmetry group which is relevant for quantum chromodynamics and the theory of weak interactions is the $SU(N)$ symmetry group. Let us consider, as an example, a collection of N scalar fields with the Lagrangian

$$\mathcal{L} = \left(\partial_\mu\phi_i\right)\left(\partial^\mu\phi_i^*\right) - m^2\left(\phi_i\phi_i^*\right) - \frac{\lambda}{4}\left(\phi_i\phi_i^*\right)^2, \qquad i = 1\ldots N, \tag{3.14}$$

where we have used Einstein's summation convention for repeated indices. The Lagrangian is symmetric under $SU(N)$ transformations of the ϕ_i fields,

$$\phi_i \to \phi_i' = U_{ij}\,\phi_j\,, \quad U^\dagger = U^{-1}\,,\ \det U = 1\,. \tag{3.15}$$

Indeed,

$$\begin{aligned} &\phi_i^*\phi_i \to \phi_i'^*\phi_i' = U_{ij}^*\phi_j^* U_{ik}\phi_k \\ &= U_{ji}^\dagger U_{ik}\phi_j^*\phi_k = \delta_{jk}\phi_j^*\phi_k = \phi_i^*\phi_i\,. \end{aligned} \tag{3.16}$$

We can learn a lot by studying small $SU(N)$ transformations. Since they form a group, large transformations can be obtained by repeating (infinitely) many small ones. We write:

$$U_{ij} = \delta_{ij} - i\theta^a T_{ij}^a + O(\theta^2), \tag{3.17}$$

where we choose θ^a to be real parameters.

The $N \times N$ matrices T^a are the generators of $SU(N)$. There are $N^2 - 1$ of them: An arbitrary $N \times N$ complex matrix has $2N^2$ real elements. For a unitary matrix $U^\dagger = U^{-1}$, only N^2 elements are independent. The condition $\det U = 1$ ("special") adds one more constraint, leaving $N^2 - 1$ independent elements. Remember the dimensionality of the indices defining the generators,

$$T^a_{i,j}: \qquad a = 1 \ldots (N^2 - 1) \textbf{ and } i, j = 1 \ldots N\,,$$

as they will be needed in various situations.
Exercise: *$SO(N)$ group: $N \times N$ matrices with*

$$R_{ij} R_{kl} \delta_{jl} = \delta_{ik}.$$

Find the number of generators.
Exercise: *The symplectic group $Sp(2N)$ can be defined as $2N \times 2N$ matrices S with*

$$S_{ij} S_{kl} \eta_{jl} = \eta_{ik},$$

where

$$\eta_{ij} = -\eta_{ji} \qquad \textit{and} \qquad \eta^2 = -\mathbf{1}.$$

Find the number of generators.

The $SU(N)$ generators are **Hermitian**,

$$\begin{aligned} & U^\dagger U = 1 \\ \rightsquigarrow \quad & \left(1 + i(T^a)^\dagger \theta^a\right)\left(1 - iT^b\theta^b\right) = \mathbf{1} + \mathcal{O}(\theta^2) \\ \rightsquigarrow \quad & T^{a\dagger} = T^a, \end{aligned} \tag{3.18}$$

and **traceless**:

$$\begin{aligned} & \det U = 1 \rightsquigarrow \log(\det U) = 0 \\ \rightsquigarrow \quad & \mathrm{Tr}\left(\log U\right) = 0 \\ \rightsquigarrow \quad & \mathrm{Tr} \log\left(1 - i\theta^a T^a\right) = 0 \\ \rightsquigarrow \quad & \mathrm{Tr}\left(i\theta^a T^a\right) = 0 \\ \rightsquigarrow \quad & \mathrm{Tr}\left(T^a\right) = 0. \end{aligned} \tag{3.19}$$

We can choose a normalisation condition for the $SU(N)$ generators. By convention, we choose

$$\mathrm{Tr}\left(T^a T^b\right) \equiv T^a_{ij} T^b_{ji} = T_R\, \delta_{ab}, \quad T_R = \frac{1}{2}\,. \tag{3.20}$$

A very basic property of the generators is that they satisfy a Lie algebra,

$$\left[T^a, T^b\right] = i f^{abc} T^c, \tag{3.21}$$

where f^{abc} are the structure constants of the algebra.
Exercise: *Prove Eq. (3.21) by considering a transformation $U'^{-1} U^{-1} U' U$, with U, U' independent $SU(N)$ transformations.*
From Eq. (3.21), we can derive

$$\begin{aligned} & \left[T^a, T^b\right] = i f^{abd} T^d \rightsquigarrow \left[T^a, T^b\right] T^c = i f^{abd} T^d T^c \\ \rightsquigarrow \quad & f^{abc} = -2i\mathrm{Tr}\left(\left[T^a, T^b\right] T^c\right). \end{aligned} \tag{3.22}$$

Exercise: *Prove that the structure constants are fully antisymmetric and real.*

3.3 Local non-abelian gauge symmetries

We now consider N fields ϕ_i (scalar or spinor). We are interested in Lagrangians which are invariant under a local $SU(N)$ transformation:

$$\phi_i(x) \to \phi_i'(x) = U_{ij}(x)\phi_j(x). \tag{3.23}$$

As in QED, the building block for the construction of a gauge-invariant Lagrangian will be a covariant derivative:

$$D_\mu = \partial_\mu - igA_\mu, \tag{3.24}$$

such that

$$D_\mu \to D_\mu' = UD_\mu U^\dagger, \qquad \text{with } U^\dagger = U^{-1}. \tag{3.25}$$

For a scalar field ϕ we have

$$(D_\mu'\phi')^\dagger(D'^\mu\phi') = (D_\mu\phi)^\dagger(D^\mu\phi).$$

Similarly, the kinetic term with the same covariant derivative for a fermion field is invariant. There are many similarities with QED; however, there are also many important differences. Let us start by pointing out that the gauge field A_μ is an $N \times N$ matrix.

We can easily find the transformation for the gauge field:

$$\begin{aligned}
& D_\mu \to D_\mu' = U(x)D_\mu U^\dagger(x) \\
\rightsquigarrow\ & \left(\partial_\mu - igA_\mu'\right) = U\left(\partial_\mu - igA_\mu\right)U^\dagger \\
\rightsquigarrow\ & \partial_\mu - igA_\mu' = \partial_\mu + U\left(\partial_\mu U^\dagger\right) - UigA_\mu U^\dagger \\
\rightsquigarrow\ & A_\mu' = U(x)A_\mu U^\dagger(x) + \frac{i}{g}U(x)\left(\partial_\mu U^\dagger(x)\right).
\end{aligned} \tag{3.26}$$

This formula is analogous to the gauge transformation of the photon in QED. However, here the gauge field A_μ and the gauge transformation U are complex $N \times N$ matrices rather than complex numbers (A_μ Hermitian and traceless, U unitary).

We now compute the corresponding commutator:

$$\begin{aligned}
& \left[D_\mu, D_\nu\right] = \left(\partial_\mu - igA_\mu\right)\left(\partial_\nu - igA_\nu\right) - \left[\mu \leftrightarrow \nu\right] \\
= \ & \partial_\mu\partial_\nu - ig\left(\partial_\mu A_\nu\right) - igA_\nu\partial_\mu - igA_\mu\partial_\nu + (ig)^2A_\mu A_\nu - \left[\mu \leftrightarrow \nu\right] \\
= \ & -ig\left\{\partial_\mu A_\nu - \partial_\nu A_\mu - ig\left[A_\mu, A_\nu\right]\right\}.
\end{aligned} \tag{3.27}$$

The commutator term was absent in the case of QED.

The gauge field strength,

$$G_{\mu\nu} \equiv \frac{i}{g}\left[D_\mu, D_\nu\right] = \partial_\mu A_\nu - \partial_\nu A_\mu - ig\left[A_\mu, A_\nu\right], \tag{3.28}$$

is no longer gauge invariant:

$$G_{\mu\nu} \to G_{\mu\nu}' = U(x)G_{\mu\nu}U^\dagger(x).$$

However, the trace

$$\begin{aligned}
& \mathrm{Tr}(G_{\mu\nu}G^{\mu\nu}) \\
\to\ & \mathrm{Tr}(UG_{\mu\nu}U^\dagger UG^{\mu\nu}U^\dagger) \\
=\ & \mathrm{Tr}(U^\dagger UG_{\mu\nu}U^\dagger UG^{\mu\nu}) \\
=\ & \mathrm{Tr}(G_{\mu\nu}G^{\mu\nu})
\end{aligned}$$

is gauge invariant.

We can expand the gauge field in the basis of generators:

$$A_\mu = A_\mu^a T^a. \tag{3.29}$$

Equivalently,

$$A_\mu^a = \frac{\mathrm{Tr}(A_\mu T^a)}{T_R}. \tag{3.30}$$

We also have

$$G_{\mu\nu} = G_{\mu\nu}^a T^a, \tag{3.31}$$

with

$$G_{\mu\nu}^a = \frac{\mathrm{Tr}(G_{\mu\nu} T^a)}{T_R}. \tag{3.32}$$

Explicitly,

$$\begin{aligned} G_{\mu\nu}^a &= \frac{1}{T_R}\mathrm{Tr}(G_{\mu\nu}T^a) \\ &= \frac{1}{T_R}\mathrm{Tr}\left(\partial_\mu A_\nu T^a - \partial_\nu A_\mu T^a - ig\left[A_\mu, A_\nu\right]T^a\right) \\ &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c. \end{aligned} \tag{3.33}$$

A prominent example of a theory with a non-abelian local gauge symmetry is QCD, which is invariant under $SU(3)$ gauge transformations:

$$\mathcal{L}_{\mathrm{QCD}} = \bar{\psi}\left(i\not{D} - m\mathbf{1}\right)\psi - \frac{1}{4}G^{c\mu\nu}G_{\mu\nu}^c. \tag{3.34}$$

Exercise:*Expand all terms in the QCD Lagrangian using the explicit expressions in terms of the gauge field for the covariant derivative and the gauge field strength. Sketch the interactions (the precise Feynman rules will be derived in the next chapter).*

Exercise:*Write a Lagrangian for a scalar field which is invariant under $SU(N)$ gauge transformations. This case appears in supersymmetric theories for the scalar partners of quarks. Sketch the interactions.*

3.3.1 Adjoint representation

By expanding the commutators we can easily prove the Jacobi identity,

$$\left[\left[T^a, T^b\right], T^c\right] + \left[\left[T^b, T^c\right], T^a\right] + \left[\left[T^c, T^a\right], T^b\right] = 0. \tag{3.35}$$

From this we derive a relation for the structure constants:

$$\begin{aligned} &\rightsquigarrow \quad \left[f^{abd}T^d, T^c\right] + \left[f^{bcd}T^d, T^a\right] + \left[f^{cad}T^d, T^b\right] = 0 \\ &\rightsquigarrow \quad f^{abd}f^{dce} + f^{bcd}f^{dae} + f^{cad}f^{dbe} = 0. \end{aligned} \tag{3.36}$$

We define the matrices

$$\tilde{T}_{ac}^b = i f^{abc}. \tag{3.37}$$

Then the above relation can be written as

$$\left[\tilde{T}^b, \tilde{T}^c\right] = i f^{bcd}\tilde{T}^d. \tag{3.38}$$

Therefore, the matrices $\tilde{T}$ furnish a representation of the same Lie algebra. This is called the adjoint representation.

We can now consider fields ψ_a which transform in the adjoint representation. An example emerges in supersymmetric extensions of QCD, where the gluino, the supersymmetric partner of the gluon, transforms in the adjoint. Specifically, the transformation is

$$\psi_a \to \psi'_a = {U_A}^{ab}\psi_b \quad \text{with} \quad U_A^{ab} = e^{-i\theta^c \tilde{T}^c_{ab}} \quad \text{and} \quad a,b,c = 1\ldots(N^2-1). \tag{3.39}$$

Or, for small transformations,

$$\begin{aligned} & \psi_a \to \psi'_a = \psi_a - i\theta^b \tilde{T}^b_{ac}\psi_c \\ \rightsquigarrow \quad & \psi'_a = \psi_a - i\theta^b \left(if^{abc}\psi_c\right) \\ \rightsquigarrow \quad & \psi'_a = \psi_a + f^{abc}\theta^b\psi_c. \end{aligned} \tag{3.40}$$

We can find a covariant derivative for the transformations of the adjoint representation in complete analogy with the "fundamental" representation:

$$\begin{aligned} D_\mu\psi_a &= \partial_\mu\psi_a - igA^b_\mu \tilde{T}^b_{ac}\psi_c \\ &= \partial_\mu\psi_a + gf^{abc}A^b_\mu\psi_c. \end{aligned} \tag{3.41}$$

Now consider a general member of the representation, $\psi = \psi_a T^a$. The covariant derivative acts on it as

$$\begin{aligned} D_\mu\psi &= \left(\partial_\mu\psi_a + gf^{abc}A^b_\mu\psi_c\right)T^a \\ &= \partial_\mu\psi - ig\left[T^b, T^c\right]\psi_c A^b_\mu, \end{aligned} \tag{3.42}$$

or equivalently

$$D_\mu\psi = \partial_\mu\psi - ig\left[A_\mu, \psi\right], \quad \psi \equiv \psi_a T^a, \quad A_\mu \equiv A^a_\mu T^a. \tag{3.43}$$

Exercise: *Take an element $\xi = \xi^a T^a$ of the Lie algebra, transforming in the adjoint representation, and a field χ transforming in the fundamental representation. Prove the Leibniz rule for the covariant derivative:*

$$D_\mu(\xi\chi) = \left(D_\mu\xi\right)\chi + \xi\left(D_\mu\chi\right). \tag{3.44}$$

3.3.2 Euler–Lagrange equations and conserved currents*

We consider the variation of the gauge field

$$A_\mu \to A_\mu + \delta A_\mu$$

and

$$\partial_\nu A_\mu \to \partial_\nu A_\mu + \delta\left(\partial_\nu A_\mu\right).$$

The corresponding variation of the gauge field strength is

$$\begin{aligned} & \delta G_{\mu\nu} = \delta\left(\partial_\mu A_\nu - \partial_\nu A_\mu - ig\left[A_\mu, A_\nu\right]\right) \\ & = \partial_\mu\delta A_\nu - \partial_\nu\delta A_\mu - ig\left[\delta A_\mu, A_\nu\right] - ig\left[A_\mu, \delta A_\nu\right] \\ \rightsquigarrow \quad & \delta G_{\mu\nu} = D_\mu\left(\delta A_\nu\right) - D_\nu\left(\delta A_\mu\right). \end{aligned} \tag{3.45}$$

We can now look at the variation of the gauge-field term in the QCD action:

$$\begin{aligned}
& \delta \int d^4x \mathrm{Tr}\left(G_{\mu\nu}G^{\mu\nu}\right) \\
= \; & 2\int d^4x \mathrm{Tr}\left(G_{\mu\nu}\delta G^{\mu\nu}\right) \\
= \; & 2\int d^4x \mathrm{Tr}\left(G^{\mu\nu}\left[D_\mu\left(\delta A_\nu\right) - D_\nu\left(\delta A_\mu\right)\right]\right) \\
= \; & 4\int d^4x \mathrm{Tr}\left(G^{\mu\nu}D_\mu\left(\delta A_\nu\right)\right) \qquad \text{(antisymmetry)} \\
= \; & 4\int d^4x \left\{\partial_\mu\left[\mathrm{Tr}\left(G^{\mu\nu}\left(\delta A_\nu\right)\right)\right] - \mathrm{Tr}\left[D_\mu\left(G^{\mu\nu}\right)\delta A_\nu\right]\right\}. \qquad (3.46)
\end{aligned}$$

The first trace is gauge invariant, so that its covariant derivative is the ordinary derivative and the first term is a surface term, which we can drop. *Remark.* Explicitly, $\mathrm{Tr}(XD_\mu Y) = \partial_\mu \mathrm{Tr}(XY) - \mathrm{Tr}((D_\mu X)Y)$ for adjoint X, Y, and the total derivative $\partial_\mu(\ldots)$ integrates to zero for variations δA_ν vanishing at infinity. So we have

$$\delta \int d^4x \mathrm{Tr}\left[\frac{-1}{2}G_{\mu\nu}G^{\mu\nu}\right] = 2\int d^4x \mathrm{Tr}\left[D_\mu\left(G^{\mu\nu}\right)\delta A_\nu\right]. \qquad (3.47)$$

The fermionic term in the Lagrangian varies as:

$$\begin{aligned}
& \delta \int d^4x \bar{\psi}\left(i\not{D} - m\right)\psi = \delta \int d^4x \bar{\psi}\left(i\not{\partial} + g\not{A}\right)\psi \\
= \; & g\int d^4x \bar{\psi}\delta\not{A}\psi = g\int d^4x \bar{\psi}\gamma^\mu T^a \psi \delta A^a_\mu \\
= \; & 2g\int d^4x \left(\bar{\psi}_i \gamma^\mu T^a_{ij}\psi_j\right)\mathrm{Tr}\left[\delta A_\mu T^a\right] \\
= \; & -2\mathrm{Tr}\left[J_\mu \delta A^\mu\right], \qquad (3.48)
\end{aligned}$$

with

$$J_\mu = J^a_\mu T^a, \qquad (3.49)$$

and

$$J^a_\mu = -g\bar{\psi}\gamma_\mu T^a \psi. \qquad (3.50)$$

Combining the variation of both fermionic and gauge boson terms in the Lagrangian of Eq. (3.34), we derive the Euler–Lagrange equation

$$D_\mu G^{\mu\nu} = J^\nu. \qquad (3.51)$$

We can act with a second covariant derivative on the above equation:

$$D_\nu D_\mu G^{\mu\nu} = D_\nu J^\nu. \qquad (3.52)$$

The left-hand side is

$$\begin{aligned}
& D_\nu D_\mu G^{\mu\nu} = D_\nu\left(\partial_\mu G^{\mu\nu} - ig\left[A_\mu, G^{\mu\nu}\right]\right) \\
& = \partial_\nu\partial_\mu G^{\mu\nu} \\
& \quad -ig\left[\partial_\nu A_\mu, G^{\mu\nu}\right] \\
& \quad -ig\left[A_\mu, \partial_\nu G^{\mu\nu}\right] - ig\left[A_\nu, \partial_\mu G^{\mu\nu}\right] \\
& \quad -g^2\left[A_\nu, \left[A_\mu, G^{\mu\nu}\right]\right]. \qquad (3.53)
\end{aligned}$$

Using that $G^{\mu\nu} = -G^{\nu\mu}$ and

$$[A,[B,C]] + [B,[C,A]] + [C,[A,B]] = 0$$

one can prove that

$$D_\nu D_\mu G^{\mu\nu} = \frac{ig}{2}\left[G_{\mu\nu}, G^{\mu\nu}\right] = 0. \tag{3.54}$$

Therefore:

$$D_\mu J^\mu = 0. \tag{3.55}$$

The fermionic current is thus no longer conserved, as it was in QED; it is, rather, covariantly conserved!

Exercise: *In an abelian gauge theory, consider the dual tensor*

$$\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}.$$

Show that

$$F^{\mu\nu}\tilde{F}_{\mu\nu} = \partial_\mu K^\mu, \tag{3.56}$$

with

$$K^\mu = \epsilon^{\mu\nu\rho\sigma}A_\nu F_{\rho\sigma}.$$

Exercise: *In a non-abelian gauge theory, consider the dual tensor*

$$\tilde{G}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}G_{\rho\sigma}.$$

Show that

$$\mathrm{Tr}\left(G^{\mu\nu}\tilde{G}_{\mu\nu}\right) = \partial_\mu K^\mu, \tag{3.57}$$

with

$$K^\mu = \epsilon^{\mu\nu\rho\sigma}\mathrm{Tr}\left[A_\nu G_{\rho\sigma} + \frac{2ig}{3}A_\nu A_\rho A_\sigma\right].$$

4

Quantisation of non-abelian gauge theories

We now have sufficient tools to study correlation functions with the path integral formalism in a non-abelian gauge theory. The classical Yang–Mills Lagrangian is

$$\mathcal{L}_{\rm YM} = -\frac{1}{4} G^{a\mu\nu} G^a_{\mu\nu}, \tag{4.1}$$

with

$$G^a_{\mu\nu} \;=\; \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu. \tag{4.2}$$

Proceeding in analogy with the path integral quantisation of the scalar field theory, we may write the following path integral for the Yang–Mills theory,

$$Z[J^\mu_a] \;=\; \mathcal{N} \int \mathcal{D} A^a_\mu e^{i \int d^4x \left[\mathcal{L}_{\rm YM}(A^\mu_a) + J^\mu_a A_{a\mu}\right]}. \tag{4.3}$$

We would like to develop a perturbation theory programme. Let us recall the steps of the scalar $-\lambda \frac{\phi^4}{4!}$ case. We could try to define the analogous steps here, namely:

- find the propagator of the free field by inverting the differential operator in the quadratic part of the Lagrangian. This would give us an expression for the path integral when all interactions are switched off ($g = 0$):

$$Z_0[J^\mu_a] = \mathcal{N}' e^{\frac{i}{2} \int d^4x d^4y J^\mu_a(x) \Delta_{\mu\nu ab}(x-y) J^\nu_b(y)}. \tag{4.4}$$

- derive the perturbative expansion from

$$Z[J] \sim e^{i \int d^4z \mathcal{L}_{int}\left(\frac{1}{i}\frac{\delta}{\delta J^\mu_a(z)}\right)} Z_0[J^\mu_a]. \tag{4.5}$$

We will see that the first of the two steps is problematic in a naive treatment.

The free-field action ($g = 0$) is

$$\begin{aligned}
S_{free} \;&=\; -\frac{1}{4} \int d^4x \left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu\right) \left(\partial^\mu A^{\nu a} - \partial^\nu A^{a\mu}\right) \\
&=\; -\frac{1}{2} \int d^4x \left[\left(\partial_\mu A^a_\nu\right) \left(\partial^\mu A^{\nu a}\right) - \left(\partial_\nu A^a_\mu\right) \left(\partial^\mu A^{\nu a}\right)\right] \\
&=\; -\frac{1}{2} \int d^4x \left[\partial_\mu \left(A^a_\nu \partial^\mu A^{\nu a}\right) - A^a_\nu \partial_\mu \partial^\mu A^{\nu a} - \partial_\nu \left(A^a_\mu \partial^\mu A^{\nu a}\right) + A^a_\mu \partial_\nu \partial^\mu A^{\nu a}\right] \\
&=\; -\frac{1}{2} \int d^4x A^{a\mu} \delta^{ab} \left[-\partial^2 g_{\mu\nu} + \partial_\mu \partial_\nu\right] A^{\nu b}.
\end{aligned} \tag{4.6}$$

We now need to find the inverse of the operator

$$\left[-\partial^2 g_{\mu\nu} + \partial_\mu \partial_\nu\right] \delta^{ab}.$$

However, it turns out that there is none! This operator has zero eigenvalues, and its determinant is zero. In particular, we obtain zero when it acts on any function that can be written as a total derivative:

$$\left[-\partial^2 g_{\mu\nu} + \partial_\mu \partial_\nu\right] \partial^\nu \Lambda(x) = -\partial^2 \partial_\mu \Lambda(x) + \partial^2 \partial_\mu \Lambda(x) = 0. \tag{4.7}$$

Our naive attempt to establish a perturbation expansion in g using a path integral formalism has failed at the first step. However, there is a property of the theory, gauge invariance, which we have not yet used and we can exploit it to remove the zero-modes of the operator in the free part of the Lagrangian.

Let us start by defining a δ-functional, in analogy to a δ-function, which we will need shortly. The integral over a δ-function is

$$\int df \delta(f) = 1.$$

We can change variables, $f = f(w)$, and we obtain

$$\int dw \frac{\partial f}{\partial w} \delta(f(w)) = 1.$$

The multidimensional generalisation of this equation is:

$$1 = \int dw_1 \dots dw_n \det\left(\frac{\partial f_i}{\partial w_j}\right) \delta(f_1(w_1, \dots, w_n)) \dots \delta(f_n(w_1, \dots, w_n)).$$

We can take the limit $n \to \infty$ which yields a functional integral over $w(x)$, where x is the continuous variable corresponding to the index $i = 1 \dots n$. We define the infinite product of delta functions as a delta functional. We write:

$$\begin{aligned} & \int \mathcal{D}w \det\left(\frac{\delta f(x)}{\delta w(y)}\right) \delta\left[f(w)\right] \\ \equiv & \lim_{n\to\infty} \int dw_1 \dots dw_n \det\left(\frac{\partial f_i}{\partial w_j}\right) \delta(f_1(w_1, \dots, w_n)) \dots \delta(f_n(w_1, \dots, w_n)) \\ = & \ 1. \end{aligned} \tag{4.8}$$

Notice the emergence of a functional determinant, due to changing variables in the measure of a functional integral.

We now return to the gauge theory; the action is of course invariant under gauge transformations

$$A_\mu \to A'_\mu = U(x) A_\mu U^\dagger(x) + \frac{i}{g} U(x) \left(\partial_\mu U^\dagger(x)\right), \tag{4.9}$$

with $A_\mu \equiv A^a_\mu T^a$. The transformation matrices U are determined by as many independent parameters as the generators of the Lie group,

$$U(x) = e^{-i\theta^a(x)T^a} = 1 - i\theta^a(x)T^a + \mathcal{O}(\theta^2). \tag{4.10}$$

The integration $\int \mathcal{D}A^a_\mu$ in the path integral formalism does not discriminate among the fields which are connected via a gauge transformation.

Consider all the fields $A_\mu^{a(independent)}$ which cannot be connected via a gauge transformation. We never know them explicitly, but we can impose that they satisfy gauge-fixing conditions of the form

$$F(A_\mu^a) = 0,$$

which remove the superfluous degrees of freedom. For example, a very common choice is the Lorenz gauge-fixing condition,

$$\partial_\mu A^{a\mu} = 0.$$

Of course, we will need to apply this gauge condition an infinite number of times, corresponding to all possible values of the gauge-transformation parameters $\theta^a(x)$ at each x which are sampled in the path integral.

Let us now go back to the path integral for the gauge theory without sources:

$$Z = \mathcal{N} \int \mathcal{D}A_\mu^a e^{i\int d^4x[\mathcal{L}_{\rm YM}(A_a^\mu)]}. \tag{4.11}$$

This integrates over all fields including the ones related by gauge transformations. In other words, had we thought of all possible gauge fixings, it integrates over all these possibilities. We can write:

$$\begin{aligned} Z &= \mathcal{N} \int \mathcal{D}A_\mu^a e^{i\int d^4x[\mathcal{L}_{\rm YM}(A_a^\mu)]} \times \mathbf{1} \\ &= \mathcal{N} \int \mathcal{D}A_\mu^a e^{i\int d^4x[\mathcal{L}_{\rm YM}(A_a^\mu)]} \times \int \mathcal{D}F^a \delta[F^a(A^{a\mu})], \end{aligned} \tag{4.12}$$

where

$$F^a(A^{\mu a}) = 0 \rightsquigarrow A^{a\mu} = A^{a\mu}(\theta^b).$$

Different gauge fixings F^a correspond to different group parameters, so we may change variables $F^a \to \theta^a$ in the second functional integration:

$$\begin{aligned} Z &= \mathcal{N} \int \mathcal{D}A_\mu^a e^{i\int d^4x[\mathcal{L}_{\rm YM}(A_a^\mu)]} \int \mathcal{D}\theta^b \det\left(\frac{\delta F^a(A^{a\mu})}{\delta\theta^b}\right) \delta[F^a(A^{a\mu}(\theta^b))] \\ &= \mathcal{N} \int \mathcal{D}\theta^b \int \mathcal{D}A_\mu^a e^{i\int d^4x[\mathcal{L}_{\rm YM}(A_a^\mu)]} \det\left(\frac{\delta F^a(A_\mu^a)}{\delta\theta^b}\right)\Bigg|_{F^a(A^{a\mu}(\theta^b))=0} \delta[F^a(A^{a\mu}(\theta^b))] \end{aligned} \tag{4.13}$$

Now there is a crucial observation to be made. No term in the inner functional integral depends on θ^b. Let us justify this statement. We have

$$\begin{aligned} 1 &= \int \mathcal{D}F^a \delta\left[F^a(A^{a\mu})\right] \\ \rightsquigarrow \frac{1}{\det\left(\frac{\delta F^a(A^{a\mu})}{\delta\theta^b}\right)\Big|_{F^a(A^{a\mu}(\theta^b))=0}} &= \int \mathcal{D}\theta^b \delta(F^a(A^{a\mu}(\theta^b))). \end{aligned} \tag{4.14}$$

This determinant is therefore a gauge-invariant functional of A_μ^a (not a number: in the non-abelian case it depends on A_μ^a), as it does not depend on gauge parameters which are integrated over on the right-hand side. *Remark.* The argument assumes that the condition $F^a(A^\theta) = 0$ has exactly one solution θ on each gauge orbit. This holds in perturbation theory (small fields), but for large non-abelian fields Lorenz and Coulomb gauges have

multiple solutions (Gribov copies; V. N. Gribov, Nucl. Phys. B139 (1978) 1; I. M. Singer, Commun. Math. Phys. 60 (1978) 7), so the Faddeev–Popov procedure is a perturbative statement. The exponential of the Yang–Mills action is of course gauge invariant:

$$e^{i\int d^4x \mathcal{L}_{\rm YM}(A_\mu)} = e^{i\int d^4x \mathcal{L}_{\rm YM}(A^a_\mu(\theta^b))}.$$

Finally, the gauge-boson fields are also members of the Lie algebra and only get reshuffled by gauge transformations. The measure $\mathcal{D}A^a_\mu$ can be evaluated in any gauge:

$$\mathcal{D}A^a_\mu = \mathcal{D}A^a_\mu(\theta^b). \tag{4.15}$$

Exercise: *Consider a gauge transformation $A^a_\mu \to A^{a\prime}_\mu$. Prove that $\mathcal{D}A^a_\mu = \mathcal{D}A^{a\prime}_\mu$. You only need to consider an infinitesimal gauge transformation.*
We then compute all terms in the inner path integral in the specific gauge chosen by the delta functional:

$$\begin{aligned} Z &= \mathcal{N}\int \mathcal{D}\theta^b \int \mathcal{D}A^a_\mu e^{i\int d^4x[\mathcal{L}_{\rm YM}(A^\mu_a)]} \det\left(\frac{\delta F^a(A^a_\mu)}{\delta\theta^b}\right)\Bigg|_{F^a(A^{a\mu}(\theta^b))=0} \delta[F^a(A^{a\mu}(\theta^b))] \\ &= \mathcal{N}\int \mathcal{D}\theta^b \int \mathcal{D}A^a_\mu(\theta^b) e^{i\int d^4x[\mathcal{L}_{\rm YM}(A^\mu_a(\theta^b))]} \det\left(\frac{\delta F^a(A^a_\mu(\theta^b))}{\delta\theta^b}\right) \delta[F^a(A^{a\mu}(\theta^b))] \\ &= \mathcal{N}\left(\int \mathcal{D}\theta^b\right) \int \mathcal{D}A^a_\mu e^{i\int d^4x[\mathcal{L}_{\rm YM}(A^\mu_a)]} \det\left(\frac{\delta F^a(A^a_\mu)}{\delta\theta^b}\right) \delta[F^a(A^{a\mu})]. \end{aligned} \tag{4.16}$$

In the last line we used the fact that the gauge-fixed field variable $A^a_\mu(\theta^b)$ is a dummy integration variable. The path integration over all possible gauge transformations (corresponding to all possible gauge fixings) is an overall normalisation factor. This is an infinite factor, an integration over infinitely many Lie algebra parameters. However, this is not a problem if we want to compute physical quantities, since the overall normalisation will cancel. We therefore end up with the gauge-fixed path integral

$$Z = \mathcal{N}'\left(\int \mathcal{D}\theta^b\right) \int \mathcal{D}A^a_\mu e^{i\int d^4x[\mathcal{L}_{\rm YM}(A^\mu_a)]} \det\left(\frac{\delta F^a(A^a_\mu)}{\delta\theta^b}\right) \delta[F^a(A^{a\mu})]. \tag{4.17}$$

It is very important to notice that (up to an irrelevant infinite normalisation) this path integral does not depend on the gauge-fixing condition $F(A^{a\mu})$ that we may choose!

The new path integral in Eq. (4.17) integrates over fields which cannot be related via gauge transformations; it should therefore be suitable for deriving Green's functions for fields which are physically distinct. However, the new expression has two features, the gauge-fixing delta-functional and the determinant, which were absent in our formulation of perturbation theory. It is therefore unclear at first sight how to establish a calculable perturbative expansion with the mathematics that we know. Two very clever tricks will come to our rescue.

Without loss of generality we write

$$F^a(A^{a\mu}) = \mathcal{G}^a(A^{a\mu}) - w^a(x). \tag{4.18}$$

We are allowed to multiply the path integral by an overall constant without any physical consequences. We then multiply by the factor

$$C = \int \mathcal{D}w^a e^{-i\int d^4x \frac{w^a(x)^2}{2\xi}}. \tag{4.19}$$

We can do this because Z in Eq. (4.17) does not depend on $w(x)$. We have

$$\begin{aligned} Z &\sim \left(\int \mathcal{D}w^a e^{-i\int d^4x \frac{w^a(x)w^a(x)}{2\xi}}\right)\int \mathcal{D}A^a_\mu e^{i\int d^4x[\mathcal{L}_{\rm YM}(A^\mu_a)]}\det\left(\frac{\delta\mathcal{G}^a(A^a_\mu)}{\delta\theta^b}\right) \\ &\quad\times\delta[\mathcal{G}^a(A^{a\mu})-w^a(x)] \\ &= \int \mathcal{D}A^a_\mu e^{i\int d^4x[\mathcal{L}_{\rm YM}(A^\mu_a)]}\det\left(\frac{\delta\mathcal{G}^a(A^a_\mu)}{\delta\theta^b}\right) \\ &\quad\times\int \mathcal{D}w^a e^{-i\int d^4x\frac{w(x)^2}{2\xi}}\delta[\mathcal{G}^a(A^{a\mu})-w(x)] \\ &= \int \mathcal{D}A^a_\mu e^{i\int d^4x\left[\mathcal{L}_{\rm YM}(A^\mu_a)-\frac{1}{2\xi}(\mathcal{G}^a(A^{a\mu}))^2\right]}\det\left(\frac{\delta\mathcal{G}^a(A^a_\mu)}{\delta\theta^b}\right). \end{aligned} \tag{4.20}$$

With this trick, we remove the delta functional from the integrand and modify the exponent of the path-integral. This also yields a well-defined propagator for the gauge boson field. If we choose for example the gauge-fixing condition

$$\mathcal{G}^a(A^{a\mu}) = \partial_\mu A^{a\mu},$$

the free part ($g = 0$) of the action in the exponent becomes:

$$S_{free} = -\frac{1}{2}\int d^4x A^{a\mu}\delta^{ab}\left[-\partial^2 g_{\mu\nu}+\left(1-\frac{1}{\xi}\right)\partial_\mu\partial_\nu\right]A^{b\nu}. \tag{4.21}$$

Now, the new differential operator has an inverse:

$$\Delta^{ab}_{\mu\nu} = \delta^{ab}\int\frac{d^4k}{(2\pi)^4}\frac{e^{-ikx}}{k^2+i\epsilon}\left[g_{\mu\nu}-(1-\xi)\frac{k_\mu k_\nu}{k^2+i\epsilon}\right]. \tag{4.22}$$

Exercise: *Find the inverse of the operator*

$$\left[\left(-\partial^2-M^2\right)g_{\mu\nu}+\left(1-\frac{1}{\xi}\right)\partial_\mu\partial_\nu\right].$$

This is the case for a massive gauge boson such as the W or the Z.
Exercise: *Find the gauge boson propagator in an axial gauge,*

$$G(A) = n_\mu A^{a\mu},$$

where n is a light-like vector, $n^2 = 0$.

We now need to deal with the determinant in the integrand of the path integral. Here we will use a result that we found for integrals over Grassmann variables. We proved earlier that:

$$\int dx_1\ldots dx_n dy_1\ldots dy_n e^{-x^TAy} = \det(A), \tag{4.23}$$

where A is an $n\times n$ matrix and x, y are *Grassmann* variables. We can take the limit $n\to\infty$ and express a functional determinant as a fermionic path integral over two independent Grassmann functions x and y:

$$\det(-igA) = \int \mathcal{D}y\mathcal{D}x e^{i\int d^4x_1 d^4x_2\, y(x_1)(gA(x_1-x_2))x(x_2)}. \tag{4.24}$$

The Faddeev–Popov idea was to introduce two new fields with the "wrong" spin-statistics (Grassmann variables in the path integral), a ghost and an anti-ghost, and write the determinant as a functional integral over an exponential. We write

$$\begin{aligned} Z &\sim \int \mathcal{D}A^a_\mu e^{i\int d^4x\left[\mathcal{L}_{\rm YM}(A^\mu_a)-\frac{1}{2\xi}(\mathcal{G}^a(A^{a\mu}))^2\right]} \det\left(-ig\frac{\delta\mathcal{G}^a(A^a_\mu)}{\delta\theta^b}\right) \\ &\sim \int \mathcal{D}A^a_\mu\mathcal{D}\bar\eta^a\mathcal{D}\eta^a e^{i\int d^4x\left[\mathcal{L}_{\rm YM}(A^\mu_a)-\frac{1}{2\xi}(\mathcal{G}^a(A^{a\mu}))^2\right]}\times \\ &\quad\times e^{i\int d^4x_1d^4x_2\bar\eta^a(x_1)\left(g\frac{\delta\mathcal{G}^a(A^a_\mu(\theta^a))}{\delta\theta^b}\right)\eta^b(x_2)}. \end{aligned} \tag{4.25}$$

We need not worry about computing precisely the overall normalisation of the path integral. This will drop out when we require that vacuum-to-vacuum transitions have a unit amplitude. The factor $-ig$ is convenient, as it will later allow us to combine all terms under a common exponential.

Let us consider a local gauge transformation

$$U(x) = e^{-i\theta^aT^a},$$

under which a gauge field transforms as

$$A_\mu(\theta) = U(x)A_\mu U^\dagger(x) + \frac{i}{g}U(x)\left(\partial_\mu U^\dagger(x)\right), \tag{4.26}$$

where, as usual, $A_\mu = A^a_\mu T^a$. For a small transformation, we obtain

$$A^{\mu,a}(\theta) = A^{\mu,a} - \frac{1}{g}\left[\partial^\mu\delta^{ab} - gf^{abc}A^{\mu,c}\right]\theta^b. \tag{4.27}$$

We recognise in the above expression the covariant derivative in the adjoint representation:

$$D^{ab}_\mu \equiv \partial_\mu\delta^{ab} - gf^{abc}A^c_\mu, \tag{4.28}$$

so we can write:

$$A^{\mu,a}(\theta) = A^{\mu,a} - \frac{1}{g}D^{\mu,ab}\theta^b. \tag{4.29}$$

We then have

$$\frac{\delta A^{\mu,a}(\theta(x))}{\delta\theta^b(y)} = -\frac{1}{g}\frac{\delta\left(D^{\mu,ac}(x)\theta^c(x)\right)}{\delta\theta^b(y)} = -\frac{1}{g}D^{\mu,ab}(x)\delta(x-y). \tag{4.30}$$

We can now compute the functional derivative

$$\begin{aligned} g\frac{\delta G(A^{\mu,a}(\theta(x)))}{\delta\theta^b(y)} &= g\int d^4z\frac{\delta G(A^{\mu,a}(\theta(x)))}{\delta(A^{\nu,c}(\theta(z)))}\frac{\delta(A^{\nu,c}(\theta(z)))}{\delta\theta^b(y)} \\ &= -\int d^4z\frac{\delta G(A^{\mu,a}(x))}{\delta(A^{\nu,c}(z))}D^{\nu,cb}(z)\delta(z-y) \\ &= -\frac{\delta G(A^{\mu,a}(x))}{\delta(A^{\nu,c}(y))}D^{\nu,cb}(y). \end{aligned} \tag{4.31}$$

We can then write the following expression for the path integral:

$$\begin{aligned} Z &\sim \int \mathcal{D}A^a_\mu\mathcal{D}\bar\eta^a\mathcal{D}\eta^a e^{i\int d^4x\left[\mathcal{L}_{\rm YM}(A^\mu_a)-\frac{1}{2\xi}(\mathcal{G}^a(A^{a\mu}))^2\right]}\times \\ &\quad\times e^{i\int d^4x_1d^4x_2\bar\eta^a(x_1)\left(-\frac{\delta\mathcal{G}^a(A^a_\mu(x_1))}{\delta A^c_\nu(x_2)}D^{cb}_\nu(x_2)\right)\eta^b(x_2)}. \end{aligned} \tag{4.32}$$

This is valid for any gauge-fixing condition $\mathcal{G}^a(A^a_\mu)$.

It will be instructive and practical (this is what we need to do when we compute elements of the S-matrix) to choose a gauge. A common choice is the Lorenz gauge:

$$\mathcal{G}^a(A^a_\mu) = \partial_\mu A^{\mu,a}, \tag{4.33}$$

for which

$$\frac{\delta \partial_\mu A^{\mu,a}(x_1)}{\delta A^{\nu,c}(x_2)} = \partial_\mu g^\mu_\nu \delta^{ac} \delta(x_1 - x_2) = \partial_\nu \delta^{ac} \delta(x_1 - x_2). \tag{4.34}$$

Therefore, in the Lorenz gauge, the path integral is (up to a normalisation):

$$Z = \int \mathcal{D}A^a_\mu \mathcal{D}\bar{\eta}^a \mathcal{D}\eta^a e^{i\int d^4x \left[-\frac{1}{4}G^a_{\mu\nu}G^{a,\mu\nu} - \frac{1}{2\xi}\left(\partial_\mu A^{a\mu}\right)^2 - \bar{\eta}^a(x)\partial_\mu D^{\mu;ab}\eta^b(x)\right]}. \tag{4.35}$$

After an integration by parts, we obtain

$$Z = \int \mathcal{D}A^a_\mu \mathcal{D}\bar{\eta}^a \mathcal{D}\eta^a e^{i\int d^4x \left[-\frac{1}{4}G^a_{\mu\nu}G^{a,\mu\nu} - \frac{1}{2\xi}\left(\partial_\mu A^{a\mu}\right)^2 + \left(\partial_\mu \bar{\eta}^a\right)D^{\mu;ab}\eta^b\right]}. \tag{4.36}$$

In this last expression, we have exponentiated all terms which arose from gauge-fixing. The argument of the exponential is a new action, modified by a gauge-fixing term and contributions from the ghost fields.

Notice that the form of the quadratic term in the ghost fields is the same as for a complex *scalar* field. However, the variables $\eta, \bar{\eta}$ in the path integral are anti-commuting Grassmann variables. Therefore, the ghost field is a scalar field with the "wrong" spin-statistics.

We also observe that for an abelian gauge theory $f^{abc} = 0$ we have $D^{ab}_\mu = \delta^{ab}\partial_\mu$ and there is no interaction term for the ghost field and the gauge boson. In this case, the ghost field is a free field and we can integrate out its contribution to the path integral (changing the irrelevant overall normalisation). This is why we never needed to introduce it in QED.

Exercise: *Find the expression of the path integral for an $SU(N)$ Yang–Mills theory in an axial gauge,*

$$G(A) = n_\mu A^{a\mu},$$

where n is a light-like vector, $n^2 = 0$.

4.1 Perturbative QCD

After gauge fixing with the Faddeev–Popov method, we can formulate a path integral for QCD, where the action is

$$S = \int d^4x \mathcal{L},$$

with

$$\mathcal{L} = \mathcal{L}_{\text{YANG–MILLS}} + \mathcal{L}_{\text{FERMION}} + \mathcal{L}_{\text{GAUGE–FIXING}} + \mathcal{L}_{\text{FADDEEV–POPOV}}.$$

The classical Yang–Mills Lagrangian is

$$\mathcal{L}_{\text{YANG–MILLS}} = -\frac{1}{4}G^a_{\mu\nu}G^{a,\mu\nu},$$

and

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu.$$

The gauge-fixing and Faddeev–Popov terms in the Lorenz gauge are

$$\mathcal{L}_{\text{GAUGE–FIXING}} = -\frac{1}{2\xi}\left(\partial^\mu A^a_\mu\right)^2,$$

$$\mathcal{L}_{\text{FADDEEV–POPOV}} = \left(\partial^\mu \bar{\eta}^a\right) D^{ab}_\mu \eta^b,$$

and the fermion term is

$$\mathcal{L}_{\text{FERMION}} = \bar{\psi}^i\left(i\gamma^\mu D^{ij}_\mu - m\delta^{ij}\right)\psi^j.$$

The covariant derivatives in the adjoint and fundamental representations are

$$D^{ab}_\mu = \partial_\mu \delta^{ab} - g f^{abc} A^c_\mu$$

and

$$D^{ij}_\mu = \partial_\mu \delta^{ij} - i g T^c_{ij} A^c_\mu,$$

respectively.

We would like to compute Green's functions using perturbation theory. If we switch off the coupling, $g = 0$, then we are left with terms which are quadratic in the fields, and we can compute the corresponding path integral for the free action. We define

$$\mathcal{L} = \mathcal{L}_{free} + \mathcal{L}_{interacting}, \tag{4.37}$$

with

$$\mathcal{L}_{free} = \mathcal{L}|_{g=0}.$$

Explicitly,

$$\begin{aligned}\mathcal{L}_{free} &= -\frac{1}{4}\left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu\right)\left(\partial^\mu A^{a,\nu} - \partial^\nu A^{a,\mu}\right) - \frac{1}{2\xi}\left(\partial^\mu A^a_\mu\right)\left(\partial^\nu A^a_\nu\right)\\ &\quad + \left(\partial_\mu \bar{\eta}^a\right)\left(\partial^\mu \eta^a\right)\\ &\quad + \bar{\psi}^i\left(i\gamma^\mu \partial_\mu - m\right)\psi^i.\end{aligned} \tag{4.38}$$

It is convenient to use integration by parts and cast the free Lagrangian in a "standard form":

$$-\frac{1}{2}\left(\text{Field}_A\right)\hat{O}\left(\text{Field}_A\right) + \partial_\mu\left(\ldots\right)$$

or, for terms with independent fields (appearing separately in the measure of the path integral),

$$-\left(\text{Field}_A\right)\hat{O}\left(\text{Field}_B\right) + \partial_\mu\left(\ldots\right).$$

We have

$$\begin{aligned}\mathcal{L}_{free} &= -\frac{1}{2}A^a_\mu K^{ab,\mu\nu} A^b_\nu\\ &\quad -\bar{\eta}^a K^{ab}\eta^b\\ &\quad -\bar{\psi}^i \Lambda^{ij}\psi^j\\ &\quad +\partial_\mu\left(\ldots\right),\end{aligned} \tag{4.39}$$

with ($\partial^2 \equiv \partial_\mu\partial^\mu$)

$$K^{ab,\mu\nu} = \delta^{ab}\left[-g^{\mu\nu}\partial^2 + \left(1 - \frac{1}{\xi}\right)\partial^\mu\partial^\nu\right], \tag{4.40}$$

$$K^{ab} = \delta^{ab}\partial^2, \tag{4.41}$$

$$\Lambda^{ij} = \delta^{ij}\left(m - i\not{\partial}\right). \tag{4.42}$$

An essential step in order to compute the generating functional for the free path integral is to find the inverse of the above operators, i.e. objects which diagonalise them in all indices. We define:

$$\begin{aligned} K^{ac}_{\mu\rho}(x)D^{cb,\rho\nu}(x-y) &= \delta^{ab}g^{\nu}_{\mu}\delta^{(4)}(x-y), \\ K^{ac}(x)D^{cb}(x-y) &= \delta^{ab}\delta^{(4)}(x-y), \\ \Lambda^{ik}(x)S^{kj}(x-y) &= \delta^{ij}\delta^{(4)}(x-y). \end{aligned} \tag{4.43}$$

Exercise: *Solve these equations. The solutions are given next in the text and it is easy to verify whether they are correct or not by inserting them above. However, it will be instructive to think how to find them if nobody told you the answer!*

We find

$$D^{ab}_{\mu\nu}(x) = \delta^{ab}\int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot x}}{k^2+i\epsilon}\left[g_{\mu\nu} - (1-\xi)\frac{k_\mu k_\nu}{k^2+i\epsilon}\right], \tag{4.44}$$

$$D^{ab}(x) = \delta^{ab}\int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot x}}{k^2+i\epsilon}(-1), \tag{4.45}$$

$$S^{ij}(x) = \delta^{ij}\int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot x}}{k^2-m^2+i\epsilon}(-1)\left(\not{k}+m\right). \tag{4.46}$$

The generating functional for the free Lagrangian is

$$\begin{aligned} Z_0\left[J_\psi, J_{\bar\psi}, J_\eta, J_{\bar\eta}, J_A\right] = \int \mathcal{D}\psi\mathcal{D}\bar\psi\mathcal{D}\eta\mathcal{D}\bar\eta\mathcal{D}A_\mu \exp\Bigg(i\int d^4x \\ \Bigg[-\frac{1}{2}A^a_\mu K^{ab,\mu\nu}A^b_\nu - \bar\eta^a K^{ab}\eta^b - \bar\psi^i\Lambda^{ij}\psi^j \\ + J^{a,\mu}_A A^a_\mu + J^i_{\bar\psi}\psi^i + \bar\psi^i J^i_\psi + J^a_{\bar\eta}\eta^a + \bar\eta^a J^a_\eta\Bigg]\Bigg), \end{aligned} \tag{4.47}$$

where we have included independent sources for all fermion, anti-fermion, ghost, anti-ghost, and gauge fields. We should keep in mind that only the source $J^{a,\mu}_A$ for the gauge field is a bosonic (commuting) variable; all other sources are Grassmann variables. We can now "complete squares" by shifting the fields as

$$A^{a,\mu}(x) \to A^{a,\mu}(x) + \int d^4y D^{ab}_{\mu\nu}(x-y)J^{b,\nu}_A(y) \tag{4.48}$$

$$\eta^a(x) \to \eta^a(x) + \int d^4y D^{ab}(x-y)J^b_\eta(y) \tag{4.49}$$

$$\bar\eta^a(x) \to \bar\eta^a(x) + \int d^4y J^b_{\bar\eta}(y)D^{ba}(x-y) \tag{4.50}$$

$$\psi^i(x) \to \psi^i(x) + \int d^4y S^{ij}(x-y)J^j_\psi(y) \tag{4.51}$$

$$\bar\psi^i(x) \to \bar\psi^i(x) + \int d^4y J^j_{\bar\psi}(y)S^{ji}(x-y). \tag{4.52}$$

We then obtain (up to an undetermined overall constant factor)

$$Z_0\left[J_\psi, J_{\bar\psi}, J_\eta, J_{\bar\eta}, J_A\right] = \mathcal{N}\exp\left(i\int d^4x d^4y\left[\frac{1}{2}J_A^{a,\mu}(x)D_{\mu\nu}^{ab}(x-y)J_A^{b,\nu}(y)\right.\right.$$
$$\left.\left.+J_{\bar\eta}^a(x)D^{ab}(x-y)J_\eta^b(y)+J_{\bar\psi}^i(x)S^{ij}(x-y)J_\psi^j(y)\right]\right). \tag{4.53}$$

We will now deal with the interaction Lagrangian; this is defined as

$$\mathcal{L}_{interaction} = \mathcal{L} - \mathcal{L}_{free}. \tag{4.54}$$

We find

$$\begin{aligned}\mathcal{L}_{interaction} &= g\bar\psi^i T_{ij}^a \not{A}^a \psi^j \\ &\quad -g\left(\partial_\mu\bar\eta^a\right) f^{abc}A^{c,\mu}\eta^b \\ &\quad -\frac{g}{2}f^{abc}A^{b,\mu}A^{c,\nu}\left(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a\right) \\ &\quad -\frac{g^2}{4}f^{abc}f^{ade}A_\mu^b A_\nu^c A^{d,\mu}A^{e,\nu}.\end{aligned} \tag{4.55}$$

The generating functional for the full theory can be obtained perturbatively, expanding in g,

$$\begin{aligned}&Z\left[J_\psi, J_{\bar\psi}, J_\eta, J_{\bar\eta}, J_A\right] \\ =\ &\mathcal{N}\exp\left\{i\int d^4z\,\mathcal{L}_{interaction}\left(-i\frac{\delta}{\delta J_A}, i\frac{\delta}{\delta J_\psi}, -i\frac{\delta}{\delta J_{\bar\psi}}, i\frac{\delta}{\delta J_\eta}, -i\frac{\delta}{\delta J_{\bar\eta}}\right)\right\} \\ &Z_0\left[J_\psi, J_{\bar\psi}, J_\eta, J_{\bar\eta}, J_A\right],\end{aligned} \tag{4.56}$$

where we must replace in the interaction Lagrangian the field variables with functional derivatives with respect to their corresponding sources.

Exercise:

- *Compute* $Z\left[J_\psi, J_{\bar\psi}, J_\eta, J_{\bar\eta}, J_A\right]$ *through* $O(g^2)$ *in the expansion around* $g = 0$.
- *Compute the generating functional of connected diagrams*

$$W\left[J_\psi, J_{\bar\psi}, J_\eta, J_{\bar\eta}, J_A\right] = -i\log Z\left[J_\psi, J_{\bar\psi}, J_\eta, J_{\bar\eta}, J_A\right]$$

  through the same order.
- *Compute the gauge-boson propagator through the same order,*

$$\langle 0, +\infty | T A^{a,\mu}(x)A^{b,\nu}(0) | 0, -\infty\rangle$$

- *Transform this expression to momentum space.*

5

BRST symmetry

We recall here the path integral for a non-abelian gauge theory with a fermion field,

$$Z \sim \int \mathcal{D}A^a_\mu \mathcal{D}\bar{\psi}^i \mathcal{D}\psi^i \mathcal{D}\bar{\eta}^a \mathcal{D}\eta^a e^{i\int d^4x \left[\mathcal{L}_{\rm YM}+\mathcal{L}_{\rm fermion}-\frac{1}{2\xi}(\mathcal{G}^a(A^{\mu a}))^2+\bar{\eta}^a\Delta^a\right]}, \tag{5.1}$$

where we have defined

$$\Delta^a(x) = -\int d^4y \left(\frac{\delta \mathcal{G}^a(A^a_\mu(x))}{\delta A^c_\nu(y)} D^{cb}_\nu(y)\right)\eta^b(y). \tag{5.2}$$

It is convenient to linearise the action in the gauge-fixing term. If we introduce (yet) another (bosonic) field, w^a, we can rewrite

$$\begin{aligned}\int \mathcal{D}w^a e^{i\int d^4x\left(\frac{\xi}{2}w^a w^a + w^a\mathcal{G}^a\right)} &= \int \mathcal{D}w^a e^{i\frac{\xi}{2}\int d^4x\left[\left(w^a+\frac{\mathcal{G}^a}{\xi}\right)^2-\frac{\mathcal{G}^a\mathcal{G}^a}{\xi^2}\right]} \\ &= \mathcal{N}_x e^{i\int d^4x \frac{-\mathcal{G}^a\mathcal{G}^a}{2\xi}}.\end{aligned} \tag{5.3}$$

We will denote by

$$\mathcal{L}_{cl} = \mathcal{L}_{\rm YM} + \mathcal{L}_{\rm fermion}, \tag{5.4}$$

the classical Lagrangian, which satisfies local gauge invariance. The path integral, up to a normalisation, is equal to

$$Z = \int \mathcal{D}A^a_\mu \mathcal{D}\bar{\psi}^i \mathcal{D}\psi^i \mathcal{D}\bar{\eta}^a \mathcal{D}\eta^a \mathcal{D}w^a e^{i\int d^4x\left[\mathcal{L}_{cl}+\bar{\eta}^a\Delta^a + w^a\mathcal{G}^a + \frac{\xi}{2}w^a w^a\right]}. \tag{5.5}$$

If we want to go back to the original form of Z we simply need to integrate out the field w^a. The exponent is not gauge invariant, except for the term with the classical Lagrangian $\mathcal{L}_{cl}$.

However, there is a closely related symmetry, called Becchi–Rouet–Stora–Tyutin or, shorter, BRST symmetry, which leaves invariant not only the classical Lagrangian but also the sum of the gauge-fixing and Faddeev–Popov terms. For the gauge boson A^a_μ and the fermion ψ, the fields which enter the classical Lagrangian $\mathcal{L}_{cl}$, the BRST transformations are *gauge transformations* of a special form: the gauge parameter is made out of the ghost field and a Grassmann variable. Specifically, the fermion transforms as:

$$\delta_\theta \psi = -iT^a\left(\theta\eta^a\right)\psi. \tag{5.6}$$

This is equivalent to a classical gauge transformation with the replacement

$$\theta^a(x) \to \theta\eta^a.$$

where both θ and the ghost field η^a are Grassmann variables. Of course, their product is a bosonic variable, as expected for a classical gauge transformation. We take the parameter θ to be global,

$$\partial_\mu \theta = 0.$$

For the gauge boson, we require that the BRST transformation is also a gauge transformation with gauge parameter $\theta\eta^a$:

$$\delta_\theta A^a_\mu = -\frac{\theta}{g} D^{ab}_\mu \eta^b. \tag{5.7}$$

Similarly, the anti-fermion field transforms as:

$$\delta_\theta \bar{\psi} = \bar{\psi} i T^a \left(\theta \eta^a\right).$$

Notice that the Grassmann variable entering here is η, so that we perform the same gauge transformation on all classical fields $\psi, \bar{\psi}, A^a_\mu$.

Before we present the transformations for the remaining fields, we state a characteristic property of the transformation: **Two successive BRST transformations on an arbitrary function of fields give zero (nilpotent transformation)**:

$$\delta_{\theta_2}\delta_{\theta_1} F(A, \psi, \bar{\psi}, \eta, \bar{\eta}) = 0. \tag{5.8}$$

If we insist on this property, we obtain:

$$\begin{aligned}
0 &= \delta_{\theta_2}\left(\delta_{\theta_1}\psi\right) \\
&= \delta_{\theta_2}\left(-iT^a\theta_1\eta^a\psi\right) \\
&= -iT^a\theta_1\left[(\delta_{\theta_2}\eta^a)\psi + \eta^a(\delta_{\theta_2}\psi)\right] \\
&= -iT^a\theta_1\left[(\delta_{\theta_2}\eta^a)\psi + \eta^a\left(-iT^b\theta_2\eta^b\psi\right)\right] \\
&= -iT^a\theta_1\left[(\delta_{\theta_2}\eta^a) - i\eta^a\theta_2\eta^b T^b\right]\psi.
\end{aligned} \tag{5.9}$$

Equivalently,

$$\begin{aligned}
& T^c\delta_{\theta_2}\eta^c = iT^bT^c\eta^b\theta_2\eta^c \\
\rightsquigarrow\ & \mathrm{tr}(T^aT^c)\delta_{\theta_2}\eta^c = -i\theta_2\mathrm{tr}(T^aT^bT^c)\eta^b\eta^c \qquad \text{(beware of the Grassmann signs!)} \\
\rightsquigarrow\ & \frac{\delta^{ac}}{2}\delta_{\theta_2}\eta^c = -i\theta_2\mathrm{tr}\left(T^aT^bT^c\right)\frac{\eta^b\eta^c - \eta^c\eta^b}{2} \\
\rightsquigarrow\ & \delta_{\theta_2}\eta^a = -i\theta_2\mathrm{tr}\left(T^a\left[T^b, T^c\right]\right)\eta^b\eta^c \\
\rightsquigarrow\ & \delta_{\theta_2}\eta^a = \theta_2 f^{bcd}\mathrm{tr}(T^aT^d)\eta^b\eta^c.
\end{aligned} \tag{5.10}$$

Therefore we require the ghost field to transform as:

$$\delta_\theta \eta^a = \frac{\theta}{2} f^{abc}\eta^b\eta^c. \tag{5.11}$$

Two successive transformations on the gauge field produce,

$$\begin{aligned}
\delta_{\theta_2}\delta_{\theta_1}A^a_\mu &= -\frac{\theta_1}{g}\delta_{\theta_2}\left[\partial_\mu\eta^a - gf^{abc}\eta^bA^c_\mu\right] \\
&= -\frac{\theta_1}{g}\left[\partial_\mu\left(\delta_{\theta_2}\eta^a\right) - gf^{abc}\left(\delta_{\theta_2}\eta^b\right)A^c_\mu - gf^{abc}\eta^b\left(\delta_{\theta_2}A^c_\mu\right)\right] \\
&= -\frac{\theta_1}{g}\left[D^{ab}_\mu\left(\delta_{\theta_2}\eta^b\right) + f^{abc}\eta^b\theta_2 D^{cd}_\mu\eta^d\right] = \ldots = 0.
\end{aligned} \tag{5.12}$$

Exercise: *Fill in the dots: prove the above statement using the anticommutation of Grassmann variables and the Jacobi identity for the structure constants.*

For the remaining two independent fields in the action of the path integral, we have no unambiguous guidance to construct their BRST transformations. We will make two very simple choices. We perform no transformation on the auxiliary bosonic field,

$$\delta_\theta w^a = 0.$$

For the antighost we require that under a BRST transformation it gets shifted by the auxiliary field,

$$\delta_\theta \bar{\eta}^a = -\frac{1}{g}\theta w^a.$$

Since w^a is itself BRST invariant, nilpotency on the antighost is then automatic. This choice, as we will see, guarantees BRST invariance of the quantum action. Indeed,

$$\delta_{\theta_1}\delta_{\theta_2}\bar{\eta}^a = \delta_{\theta_1}\delta_{\theta_2} w^a = 0.$$

Let us now make the notation more compact. We consider any field F from the set $\{A_\mu, \psi, \bar{\psi}, \eta, \bar{\eta}, w\}$ and introduce the shorthand notation

$$\delta_\theta F \equiv \theta(sF).$$

We have pulled out an explicit factor of the BRST transformation parameter, and sF denotes the remainder of the expression for the transformed field F. For example, we write

$$\delta_\theta \psi = -iT^a \theta \eta^a \psi \rightsquigarrow s\psi = -iT^a \eta^a \psi.$$

Notice that if the field F is a bosonic field then sF is Grassmannian and vice versa. If we perform two consecutive BRST transformations, we have:

$$\delta_{\theta_1}\delta_{\theta_2} F = \theta_2\theta_1 s^2 F.$$

Nilpotency means that:

$$s^2 F = 0. \tag{5.13}$$

Nilpotency is a property valid for a product of two fields as well. Performing a BRST transformation on the product of two fields we have:

$$\begin{aligned} \delta_{\theta_1}(F_1F_2) &= (\delta_{\theta_1}F_1)F_2 + F_1(\delta_{\theta_1}F_2) \\ &= \theta_1(sF_1)F_2 + F_1\theta_1(sF_2) \\ &= \theta_1\left[(sF_1)F_2 \pm F_1(sF_2)\right], \end{aligned} \tag{5.14}$$

where the minus sign arises if F_1 is a Grassmann variable. We have used here that the fields F and sF always have opposite statistics. If we perform a double BRST transformation on the product F_1F_2 we then find,

$$\begin{aligned} \delta_{\theta_2}\delta_{\theta_1}(F_1F_2) &= \theta_1\delta_{\theta_2}\left[(sF_1)F_2 \pm F_1(sF_2)\right] \\ &= \theta_1\left[(sF_1)\theta_2(sF_2) \pm \theta_2(sF_1)(sF_2)\right] \\ &= \theta_1\theta_2\left[\mp(sF_1)(sF_2) \pm (sF_1)(sF_2)\right] \\ &= 0. \end{aligned} \tag{5.15}$$

We can continue in the same spirit. We find that:

$$\delta_{\theta_2}\delta_{\theta_1}(F_1F_2\ldots F_n) = 0. \tag{5.16}$$

In fact every functional of these fields satisfies,

$$\delta_{\theta_2}\delta_{\theta_1}G[F_1, F_2, \ldots, F_n] = 0.$$

We will return to this shortly.

We now investigate the effect of a BRST transformation on the gauge-fixing term $\mathcal{G}^a$ in the Lagrangian. $\mathcal{G}^a$ is a functional of the gauge field and we use the chain rule:

$$\begin{aligned}\delta_\theta \mathcal{G}^a(A^a_\mu(x)) &= \int d^4y \frac{\delta \mathcal{G}^a(x)}{\delta A^b_\mu(y)} \delta_\theta A^b_\mu(y) \\ &= -\frac{1}{g}\theta \int d^4y \frac{\delta \mathcal{G}^a(x)}{\delta A^b_\mu(y)} D^{bc}_\mu \eta^c(y) \\ &= \frac{\theta}{g}\Delta^a. \end{aligned} \tag{5.17}$$

Let us now consider the variation:

$$\begin{aligned}\delta_\theta \left[\bar\eta^a \left(\mathcal{G}^a + \frac{\xi}{2} w^a\right)\right] &= (\delta_\theta \bar\eta^a)\left(\mathcal{G}^a + \frac{\xi}{2} w^a\right) + \bar\eta^a(\delta_\theta \mathcal{G}^a) \\ &= -\frac{\theta}{g}\left(\bar\eta^a \Delta^a + w^a \mathcal{G}^a + \frac{\xi}{2} w^a w^a\right). \end{aligned} \tag{5.18}$$

In other words, the non-classical part of the Lagrangian is itself the BRST variation of a functional. Owing to nilpotency, such terms are invariant under a BRST transformation:

$$\delta_\theta \left[\mathcal{L}_{cl} + \bar\eta^a \Delta^a + w^a \mathcal{G}^a + \frac{\xi}{2} w^a w^a\right] = (\delta_\theta \mathcal{L}_{cl}) - g\delta_\theta \left(s\left[\bar\eta^a \left(\mathcal{G}^a + \frac{\xi}{2} w^a\right)\right]\right) = 0. \tag{5.19}$$

We recall that the classical part of the Lagrangian is BRST invariant owing to its gauge invariance.

Exercise: *Prove that the Jacobian of a BRST transformation is unity. This completes the proof that the full path integral is BRST invariant.*

The BRST symmetry allows us to characterise the physical asymptotic states of the S-matrix. Let us compute the S-matrix element

$$\langle \alpha |\, \beta \rangle\,,$$

in two different gauges, $\mathcal{G}_1$ and $\mathcal{G}_2$, where the two gauge-fixing conditions differ slightly:

$$\mathcal{G}_2 = \mathcal{G}_1 + \tilde\delta \mathcal{G}\,.$$

We require that the initial $|\beta\rangle$ and final $\langle\alpha|$ states are physical and thus the same in both gauges:

$$\langle\alpha|_{\mathcal{G}_1} = \langle\alpha|_{\mathcal{G}_2}\,, \qquad |\beta\rangle_{\mathcal{G}_1} = |\beta\rangle_{\mathcal{G}_2}\,.$$

The matrix element is computed through the path integral Z, which has an apparent but fake dependence on the gauge. A difference due to the different gauge-fixing conditions shows up through the change of the action in the exponent of the path integral. We have

$$\tilde\delta Z = Z_{\mathcal{G}_2} - Z_{\mathcal{G}_1} = \int DA \ldots e^{i\int d^4x \left[\mathcal{L}_{cl} - g s\, K|_{\mathcal{G}_2}\right]} - \int DA \ldots e^{i\int d^4x \left[\mathcal{L}_{cl} - g s\, K|_{\mathcal{G}_1}\right]}.$$

Since the two gauge-fixing conditions differ only infinitesimally, we have, schematically,

$$\tilde{\delta}Z = Z_{\mathcal{G}_1}(-ig)s\tilde{\delta}K.$$

Demanding that this difference has no physical consequences leads to a selection criterion for physical states. Promoting $\tilde{\delta}K$ into an operator, we would like it to yield zero when inserted in a matrix element for a transition between physical states:

$$0 = \langle \text{out}| \, s\tilde{\delta}K \, |\text{in}\rangle \, .$$

We can construct an operator Q which is the generator of the BRST transformations for canonical fields,

$$\delta_\theta(Field) = i\,[\theta Q, Field] = i\theta\,[Q, Field]_\mp \, , \tag{5.20}$$

where the subscript denotes a commutator for a bosonic field and an anticommutator for a Grassmann field, with $[A, B]_\pm \equiv AB \pm BA$. Using the BRST generator we write

$$\langle \text{out}| \, s\tilde{\delta}K \, |\text{in}\rangle = i \, \langle \text{out}| \left[Q, \tilde{\delta}K\right]_\mp |\text{in}\rangle = i \, \langle \text{out}| \, Q\tilde{\delta}K \, |\text{in}\rangle \mp i \, \langle \text{out}| \, \tilde{\delta}KQ \, |\text{in}\rangle \, .$$

The matrix element is the same in all gauges if the above vanishes. For this to happen, we must impose a condition on the states,

$$\langle \text{out}| \, Q = Q \, |\text{in}\rangle = 0. \tag{5.21}$$

We have arrived at a selection criterion for asymptotic (in and out) physical states. Physical in or out states must necessarily be annihilated by the generator of BRST transformations,

$$Q \left|\text{phys}\right\rangle = 0 \, . \tag{5.22}$$

However, not all "in" or "out" states which are annihilated by Q need be considered as physical. Specifically, there is a class of states, called "BRST exact", which are of the form

$$|\text{exact}\rangle = Q \, |\chi\rangle \ , \tag{5.23}$$

where $|\chi\rangle$ on the right-hand side can be any other state. These have no physical relevance, owing to the nilpotency of the BRST transformations. We can write

$$0 = s(s \, Field) = [Q, [Q, Field]_\mp]_\pm = \left[Q^2, Field\right]_- \, .$$

For the above to be satisfied for every field we need

$$Q^2 = 0.$$

Clearly, "BRST exact" states are annihilated by the generator Q,

$$Q \, |\text{exact}\rangle = Q^2 \, |\chi\rangle = 0 \, |\chi\rangle = 0. \tag{5.24}$$

However, BRST exact states do not contribute to physical scattering matrix elements, for the following reason. Assume a truly physical state $\left|\text{phys}\right\rangle$ which is not an exact state. The matrix element for the scattering into an exact state is zero. Indeed,

$$\left\langle \text{phys}\right| \text{exact}\rangle = \left\langle \text{phys}\right| Q \, |\chi\rangle = 0. \tag{5.25}$$

Similarly, we find

$$\langle \text{exact}' | \, \text{exact} \rangle = \langle \chi' | \, Q^2 \, | \chi \rangle = 0. \tag{5.26}$$

In summary, consistency of canonical quantisation and the path integral requires that we select as physical in or out states the ones that are annihilated by the generator Q of BRST transformations, modulo BRST exact states, which do not contribute to physical matrix elements. *Remark.* The step $\langle \text{phys} | \, \text{exact} \rangle = \langle \text{phys} | \, Q \, | \chi \rangle = 0$ uses that Q is hermitian, $Q^\dagger = Q$, so that $\langle \text{phys} | \, Q = (Q \, | \text{phys} \rangle)^\dagger = 0$. The precise statement is that the physical Hilbert space is the BRST cohomology $\operatorname{Ker} Q / \operatorname{Im} Q$; that it has a positive-definite norm and that the S-matrix is unitary on it is the Kugo–Ojima quartet mechanism (T. Kugo, I. Ojima, Prog. Theor. Phys. Suppl. 66 (1979) 1; Weinberg, QFT II, Sec. 15.7).

5.1 Application to the free electromagnetic field

If we turn off the coupling constant $g = 0$, non-abelian gauge theories reduce to (copies of) free electrodynamics. In and out states of the S-matrix are constructed from the requirement that asymptotic states at times very far in the future and in the past are states of the interaction-free Lagrangian. It is therefore very useful to study the BRST symmetry and its consequences due to Eq. (5.22) on free photon states.

The generating functional for electrodynamics in the Lorenz gauge takes the form

$$Z = \int \mathcal{D}A_\mu \mathcal{D}\bar{\eta} \mathcal{D}\eta \mathcal{D}w \exp\left\{ i \int d^4x \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{\xi}{2} w^2 + w \partial_\mu A^\mu - \bar{\eta} \partial^2 \eta \right] \right\}. \tag{5.27}$$

The BRST transformations take a simple enough form even if we integrate out the auxiliary w field and work with a Lagrangian which is non-linear in the gauge-fixing condition,

$$Z = \int \mathcal{D}A_\mu \mathcal{D}\bar{\eta} \mathcal{D}\eta \exp\left\{ i \int d^4x \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{\left(\partial_\mu A^\mu\right)^2}{2\xi} - \bar{\eta} \partial^2 \eta \right] \right\}. \tag{5.28}$$

The classical Euler–Lagrange equation of motion for the auxiliary field w is

$$w = -\frac{1}{\xi} \partial_\mu A^\mu. \tag{5.29}$$

Notice that we can obtain the path integral with the field integrated out, Eq. (5.28), by substituting the equation of motion (5.29) into the path integral of Eq. (5.27).

The BRST transformations (with the w field integrated out) are

$$\delta_\theta \eta = 0, \quad \delta_\theta \bar{\eta} = \frac{\theta}{g\xi} \partial_\mu A^\mu, \quad \delta_\theta A^\mu = -\frac{\theta}{g} \partial^\mu \eta. \tag{5.30}$$

Indeed, we can verify that

$$\begin{aligned} &\delta_\theta \left[\frac{\left(\partial_\mu A^\mu\right)^2}{2\xi} + \bar{\eta} \partial^2 \eta \right] \\ &= \frac{\partial_\nu A^\nu}{\xi} \partial_\mu \delta_\theta A^\mu + \left(\delta_\theta \bar{\eta}\right) \partial^2 \eta \\ &= -\theta \frac{\partial_\mu A^\mu}{g\xi} \partial^2 \eta + \theta \frac{\partial_\mu A^\mu}{g\xi} \partial^2 \eta = 0 \,. \end{aligned} \tag{5.31}$$

We now write the quantum fields as superpositions of plane waves,

$$A^\mu(x) = \int \frac{d^3k}{(2\pi)^3 2\omega_k} \sum_\lambda \left[\epsilon^\mu_\lambda(k) a_\lambda e^{-ik\cdot x} + \epsilon^\mu_\lambda(k)^* a^*_\lambda e^{ik\cdot x}\right], \tag{5.32}$$

$$\bar\eta(x) = \int \frac{d^3k}{(2\pi)^3 2\omega_k} \left[b e^{-ik\cdot x} + b^* e^{ik\cdot x}\right], \tag{5.33}$$

$$\eta(x) = \int \frac{d^3k}{(2\pi)^3 2\omega_k} \left[c e^{-ik\cdot x} + c^* e^{ik\cdot x}\right], \tag{5.34}$$

where the operators b, b^* and c, c^* are not necessarily Hermitian conjugates. Following the canonical quantisation programme of Part I, the operators a^*_λ are creation operators of photon states with polarisation λ,

$$a^*_\lambda |0\rangle = |\gamma_\lambda\rangle \ . \tag{5.35}$$

Similarly, the b^* and c^* operators generate antighost and ghost states in the Fock space, respectively,

$$b^* |0\rangle = \left|\bar\eta\right\rangle \ , \ \ c^* |0\rangle = \left|\eta\right\rangle \ . \tag{5.36}$$

We choose the polarisation vectors ϵ^μ_λ to satisfy the normalisation

$$\epsilon_\lambda \cdot \epsilon^*_{\lambda'} = g^{\lambda\lambda'}. \tag{5.37}$$

Specifically, we take the zeroth polarisation to be "scalar", pointing in the time direction,

$$\epsilon^\mu_0(k) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{5.38}$$

the third polarisation to be "longitudinal", pointing in the direction of motion of the photon,

$$\epsilon^\mu_3(k) = \frac{1}{\left|\vec{k}\right|} \begin{pmatrix} 0 \\ k^1 \\ k^2 \\ k^3 \end{pmatrix}, \tag{5.39}$$

and the other two polarisation vectors $\epsilon^\mu_{1,2}(k)$ to be transverse to the time and longitudinal directions.

Substituting the plane-wave expansion of the fields into the BRST transformations of Eq. (5.30), we obtain (**exercise**) (absorbing the factor $1/g$ into θ):

$$[Q, a_\lambda] = +\frac{k\cdot\epsilon_\lambda}{\epsilon_\lambda\cdot\epsilon^*_\lambda}\, c, \tag{5.40}$$

$$\left[Q, a^*_\lambda\right] = -\frac{k\cdot\epsilon_\lambda}{\epsilon_\lambda\cdot\epsilon^*_\lambda}\, c^*, \tag{5.41}$$

$$\{Q, b\} = -\frac{1}{\xi}\sum_\lambda (\epsilon_\lambda\cdot k)\, a_\lambda, \tag{5.42}$$

$$\{Q, b^*\} = \frac{1}{\xi}\sum_\lambda (\epsilon_\lambda\cdot k)\, a^*_\lambda, \tag{5.43}$$

$$\{Q, c\} = 0, \tag{5.44}$$

$$\{Q, c^*\} = 0. \tag{5.45}$$

Let us act with both sides of the three equations above which contain creation operators on the vacuum, and use that $Q|0\rangle = 0$. We obtain

$$Q|\gamma_\lambda\rangle = -\frac{k\cdot\epsilon_\lambda}{\epsilon_\lambda\cdot\epsilon_\lambda^*}|\eta\rangle\,, \tag{5.46}$$

$$Q|\bar{\eta}\rangle = \frac{1}{\xi}\sum_\lambda(\epsilon_\lambda\cdot k)|\gamma_\lambda\rangle\,, \tag{5.47}$$

$$Q|\eta\rangle = 0\,. \tag{5.48}$$

For a transverse photon ($\lambda = 1, 2$), we have from Eq. (5.46) that

$$Q|\gamma_1\rangle = Q|\gamma_2\rangle = 0. \tag{5.49}$$

These states are annihilated by the BRST generator and we consider them to be physical. For a scalar or a longitudinal photon, Eq. (5.46) gives

$$Q|\gamma_0\rangle = Q|\gamma_3\rangle = -k^0|\eta\rangle \neq 0. \tag{5.50}$$

Therefore, purely longitudinal or purely scalar photon states are not annihilated by the BRST generator and they are unphysical.

From Eq. (5.50), we see that there is a linear combination of a scalar and a longitudinal photon which annihilates the BRST operator,

$$Q[|\gamma_0\rangle - |\gamma_3\rangle] = 0. \tag{5.51}$$

However, this state turns out to be "BRST exact" and it does not contribute to the S-matrix. Indeed, writing explicitly the surviving terms in the sum on the right-hand side of Eq. (5.47), we find

$$\frac{k^0}{\xi}\Big[|\gamma_0\rangle - |\gamma_3\rangle\Big] = Q|\bar{\eta}\rangle\,. \tag{5.52}$$

Reading Eq. (5.50) as

$$|\eta\rangle = -\frac{1}{k^0}Q|\gamma_0\rangle\,, \tag{5.53}$$

we find that in or out states with a ghost are BRST exact and do not contribute to the S-matrix. *Remark.* The antighost state is excluded for a different reason: by Eq. (5.47) it is not annihilated by Q. The four states $|\gamma_0\rangle + |\gamma_3\rangle$ (not closed), $|\gamma_0\rangle - |\gamma_3\rangle$ (exact), $|\eta\rangle$ (exact) and $|\bar{\eta}\rangle$ (not closed) form a Kugo–Ojima quartet, which drops out of the cohomology $\mathrm{Ker}\,Q/\mathrm{Im}\,Q$; only the two transverse photons remain.

Finally, the Lagrangian has a symmetry which preserves the *ghost number*. This symmetry dictates that ghosts cannot appear alone: they are always accompanied by antighosts. Ghost-number conservation alone, however, does not remove ghosts from the asymptotic states (a ghost–antighost pair has ghost number zero). The decisive argument is the one above: ghosts are BRST exact and antighosts are not BRST closed, so neither belongs to the physical subspace $\mathrm{Ker}\,Q/\mathrm{Im}\,Q$ (Kugo–Ojima; Weinberg, QFT II, Sec. 15.7).

Suggested further reading:

"A BRST primer"
D. Nemeschansky, C. R. Preitschopf and M. Weinstein.
10.1016/0003-4916(88)90233-3
Annals Phys. **183**, 226 (1988).

6

Quantum effective action and the effective potential

We have started to collect essential tools for the renormalisation of gauge field theories by proving the existence of the BRST symmetry. In the following chapters we will convince ourselves that gauge theories, such as QCD, are renormalisable. It turns out that for renormalisation we need to worry only about one-particle-irreducible (1PI) Feynman diagrams. If these are rendered finite, then the full S-matrix elements will possess no other ultraviolet singularities. In this chapter, we will introduce a new functional, the quantum effective action, which generates only 1PI graphs. The quantum effective action is also a very important tool to define the ground state of the quantum field theory, and to study symmetry breaking via quantum effects.

We have worked with the generating functional

$$Z[J] = \int \mathcal{D}\phi e^{i\left(S[\phi]+\int d^4x J(x)\phi(x)\right)},$$

from which Green's functions are obtained as

$$\langle 0|\, T\phi(x_1)\dots\phi(x_n)\,|0\rangle = \frac{1}{i^n}\frac{1}{Z[J]}\frac{\delta^n Z[J]}{\delta J(x_1)\dots\delta J(x_n)}\bigg|_{J=0}.$$

We have seen that if we require only connected graphs to be generated, which are the ones relevant for computing S-matrix elements, we should use the generating functional $W[J]$, with

$$Z[J] = e^{iW[J]} \rightsquigarrow W[J] = -i\log Z[J].$$

Then,

$$\langle 0|\, T\phi(x_1)\dots\phi(x_n)\,|0\rangle_{\text{conn}} = \frac{1}{i^{n-1}}\frac{\delta^n W[J]}{\delta J(x_1)\dots\delta J(x_n)}\bigg|_{J=0}.$$

An important Green's function is the one-point function

$$\langle 0|\,\phi\,|0\rangle\,(x) = \frac{\delta W[J]}{\delta J(x)}\bigg|_{J=0},$$

which we typically find to vanish for physical fields that can create asymptotic states. However, there are situations in which a field (external field) fills the vacuum, $\langle 0|\,\phi\,|0\rangle\,(x) \neq 0$. Such a constant vacuum value cannot be in the final or initial state of a scattering process, but it can be a background in which the scattering of other fields takes place. For example, the Higgs field may have such a role. *Remark.* A field with $\langle\phi\rangle \neq 0$ can still create particles:

its fluctuation $\phi - \langle\phi\rangle$ has vanishing vev and interpolates for asymptotic states. This is how the Higgs boson, the excitation of the Higgs field around its vev $v \simeq 246$ GeV, was discovered at the LHC in 2012 ($m_H \simeq 125$ GeV). It is the constant vev itself, not the field, that is a background. The vacuum expectation value of a field also need not vanish when the corresponding source in the generating functional is not set to zero. In the presence of sources, we have

$$\langle\phi\rangle_J \equiv \langle 0|\,\phi\,|0\rangle_J(x) = \frac{\delta W[J]}{\delta J(x)}.$$

The above equation defines a relation between the source $J(x)$ and the vacuum expectation value of the corresponding field. We can then trade source functions in the path integral and in functional derivatives for the corresponding vevs (vacuum expectation values),

$$\langle\phi\rangle_J = \text{function}(J), \qquad J = \text{function}^{-1}(\langle\phi\rangle_J).$$

Regarding the relation between $W[J]$ and $\langle\phi\rangle_J$ as a differential equation, we write its solution as

$$W[J] = \int d^4x\,\langle\phi\rangle_J(x)J(x) + \Gamma\left[\langle\phi\rangle_J\right], \tag{6.1}$$

where the last term is a constant of integration and does not depend on the source $J(x)$. This constant of integration is the quantum effective action,

$$\Gamma\left[\langle\phi\rangle_J\right] = W[J] - \int d^4x J(x)\,\langle\phi\rangle_J(x), \tag{6.2}$$

and it is a functional of field vacuum expectation values with very interesting properties. From the above we find the simple equation

$$\frac{\delta\Gamma\left[\langle\phi\rangle\right]}{\delta\,\langle\phi\rangle_J(x)} = -J(x). \tag{6.3}$$

Recall the role of the classical action $S[\phi]$. The equations of motion for the classical field are found by requiring that the action is stationary,

$$\left.\frac{\delta S[\phi]}{\delta\phi}\right|_{\phi=\phi_{\text{classical}}} = 0.$$

At the quantum level, fields are promoted to operators. The analogue of the classical fields in quantum field theory is the expectation value of the field operator in the state of the system, usually the ground state. In the absence of sources, the quantum effective action yields the equations of motion for the average values of quantum fields:

$$\frac{\delta\Gamma\left[\langle\phi\rangle\right]}{\delta\,\langle\phi\rangle_0(x)} = 0. \tag{6.4}$$

6.1 The quantum effective action as a generating functional

We may take a further step and use the quantum effective action to generate new Green's functions. We use it in the exponent of a path integral:

$$e^{iW_\Gamma[g,J]} = \int \mathcal{D}\langle\phi\rangle\, e^{\frac{i}{\hbar}\{\Gamma[\langle\phi\rangle]+\int d^4x J(x)\langle\phi\rangle(x)\}}. \tag{6.5}$$

We can establish perturbation theory using this path integral. It is possible to derive propagators by identifying the quadratic terms in the action and inverting the corresponding operator. We will not do this explicitly here; we are rather interested in counting powers of the arbitrary constant g.

Every propagator in a graph (since it is produced by inversion) contributes a single power of g. Vertices are derived from the non-quadratic terms in the Lagrangian without any inversion. Thus, each vertex contributes a factor $1/g$ to a Feynman graph. For a graph with N_I propagators and N_V vertices the overall power of the coupling is:

$$g^{N_I - N_V}.$$

All connected graphs generated by W_Γ [1] *with N_I propagators and N_V vertices have $L = 1 + N_I - N_V$ loops.* It only takes trying out a few examples to convince ourselves of the above statement. Alternatively, assign N_I unconstrained momenta, one for each propagator. Each vertex will provide one constraint, of which one combination is an overall delta function stating that the sum of momenta of incoming and outgoing particles is zero. $N_I - N_V + 1$ momenta are left unconstrained and they are thus loop momenta. The power of g for a graph is therefore determined exclusively from the number of loops that it possesses:

$$g^{L-1}.$$

We can then perform an expansion:

$$W_\Gamma[g, J] = \sum_{L=0}^{\infty} g^{L-1} W_\Gamma^{(L)}[J]. \tag{6.6}$$

Of course, we are still interested in the case $g = 1$. What the above expression tells us,

$$W_\Gamma[1, J] = \sum_{L=0}^{\infty} W_\Gamma^{(L)}[J], \tag{6.7}$$

is that the generating functional $W_\Gamma[1, J]$ can be decomposed as a sum of independent generating functionals $W_\Gamma^{(L)}[J]$ corresponding to different loop orders. The functionals $W_\Gamma^{(L)}$ are independent in the sense that shifts in the measure do not mix them; symmetries of the full action should therefore be symmetries of each one of the loop contributions separately.

Let us explore the limit of a very small parameter g. We can expand the exponent in the path integral around the value:

$$\left\langle \phi \right\rangle = \left\langle \phi \right\rangle_J + \eta, \qquad \text{with} \qquad \frac{\delta \Gamma[\left\langle \phi \right\rangle]}{\delta \left\langle \phi \right\rangle_J (x)} = -J(x).$$

We have for the exponent of the path integral:

$$\begin{aligned} \Gamma[\left\langle \phi \right\rangle] + \int d^4x J \left\langle \phi \right\rangle &= \Gamma[\left\langle \phi \right\rangle_J] + \int d^4x J \left\langle \phi \right\rangle_J \\ &\quad + \int d^4x \eta(x) \left[\frac{\delta \Gamma[\left\langle \phi \right\rangle]}{\delta \left\langle \phi \right\rangle_J} + J \right] \\ &\quad + \frac{1}{2} \int d^4x d^4y \eta(x) \frac{\delta^2 \Gamma[\left\langle \phi \right\rangle]}{\delta \left\langle \phi \right\rangle_J (x) \delta \left\langle \phi \right\rangle_J (y)} \eta(y) + \dots . \end{aligned} \tag{6.8}$$

[1] They are connected because W_Γ is the logarithm of a path integral.

The linear term in η vanishes. We therefore have

$$\Gamma[\langle\phi\rangle] + \int d^4x J \langle\phi\rangle = \left\{\Gamma[\langle\phi\rangle_J] + \int d^4x J \langle\phi\rangle_J\right\} + O(\eta^2) = W[J] + O(\eta^2). \tag{6.9}$$

We then have

$$e^{\frac{i}{g}\sum_{L=0}^{\infty} g^L W_\Gamma^{(L)}[J]} = e^{\frac{i}{g}W[J]} \int \mathcal{D}\eta e^{\frac{i}{g}\int O(\eta^2)}. \tag{6.10}$$

The path integral over η can be computed perturbatively. The factor $\frac{1}{g}$ can be eliminated by redefining $\eta = \eta' g^{1/2}$. Then we are left with a "canonically" normalised quadratic part. The important observation is that this perturbative expansion starts at order $O(g^0)$ at the earliest. After taking the logarithm of both sides of the above equation, and by comparing the $\frac{1}{g}$ coefficients, we find that

$$W[J] = W_\Gamma^{(0)}[J]. \tag{6.11}$$

In other words, the full generating functional of connected graphs $W[J]$ can be obtained from a generating functional in which we have replaced the classical action with the quantum effective action:

$$-i \log \int \mathcal{D}\phi e^{i\left[S[\phi] + \int d^4x J(x)\phi(x)\right]} = -i \log \int_{\text{TREE}} \mathcal{D}\phi e^{i\left[\Gamma[\phi] + \int d^4x J(x)\phi(x)\right]}, \tag{6.12}$$

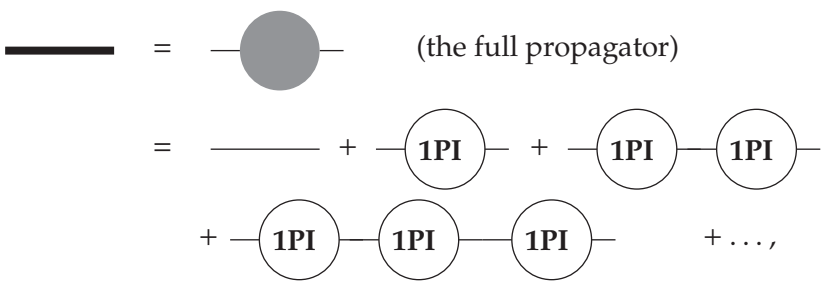
keeping only the tree contributions (denoted by the subscript on the integral sign).

This is a remarkable result; it states that it is possible to reorganise the perturbative expansion, which gives rise to both tree and loop graphs, into a new expansion where only tree graphs appear. Of course, $W[J]$ and $W_\Gamma^{(0)}[J]$ are equal. The corresponding perturbative expansions are therefore equivalent; the apparent lack of loops in the expansion from the path integral with the effective action, $W_\Gamma^{(0)}[J]$, should be explained by a rewriting of the usual expansion from $W[J]$ with modified vertices and propagators. These new exotic vertices and propagators should be "dressed" to account for all loop effects that we have encountered in the path integral of the classical action.

The above result is of great importance for renormalisation. "Trees" do not have any ultraviolet divergences. Therefore, we only need to establish a renormalisation procedure which renders finite the "dressed" propagators and vertices of the quantum effective action.

Let us take the "tree-only" statement seriously, and write down all possible graphs that we might have for two-, three-, and four-point functions. This will be sufficient to establish a pattern for the Green's functions derived from the effective action. Actually, we can only have a very small number of tree graphs for a small number of external legs.

We can figure out the propagators and vertices of the tree diagrams of Fig. 6.1 by comparing with the usual Feynman diagrams which we obtain by expanding $W[J]$. The two-point function in Fig. 6.1 must be equivalent to the full propagator, computed at all orders in perturbation theory from $W[J]$:



(6.13)

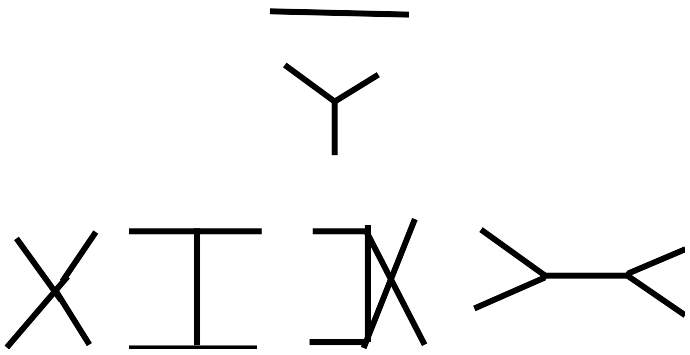
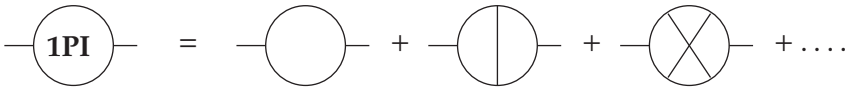


Figure 6.1. Possible tree graphs for two-, three-, and four-point functions. This very small number of connected graphs, which arises from the perturbative expansion of the path integral $W_\Gamma[J]$ with the effective action $\Gamma[\langle\phi\rangle]$, should contain in the propagators and vertices all loop effects found in the usual path integral $W[J]$.

where we sum all possible Feynman diagrams with two external legs. We can write the sum of all graphs contributing to the full propagator as a geometric series of one-particle-irreducible (1PI) two-point loop Feynman diagrams,



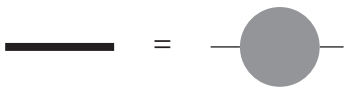
(6.14)

One-particle irreducible diagrams are those which cannot be separated into two diagrams after we cut one of the propagators. Knowledge of the 1PI propagator graphs is sufficient to determine the full propagator. Let us work, as an example, with a scalar field theory. We denote, in Fourier space,



$$= \frac{i}{p^2 - m^2}\left(-i\Sigma(p^2)\right)\frac{i}{p^2 - m^2}.$$

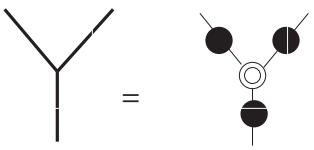
From Eq. (6.13) we find



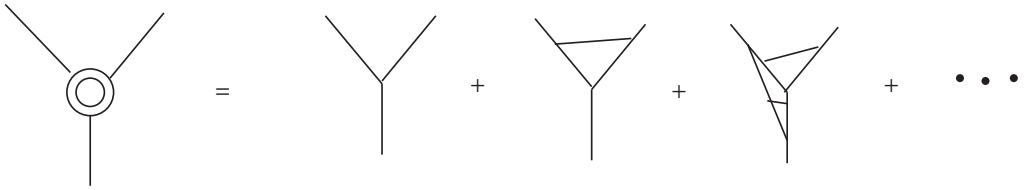
$$= \frac{i}{p^2 - m^2}\sum_{n=0}^{\infty}\left[\frac{\Sigma(p^2)}{p^2 - m^2}\right]^n = \frac{i}{p^2 - m^2 - \Sigma(p^2)}.$$

Indeed, the full propagator is then computed using only 1PI graphs.

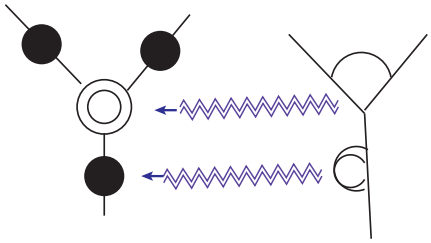
We proceed to compare the three-point Green's functions of Fig. 6.1 with the result of the full perturbative expansion from $W[J]$. We know that the propagators connected to the triple vertex are full propagators:



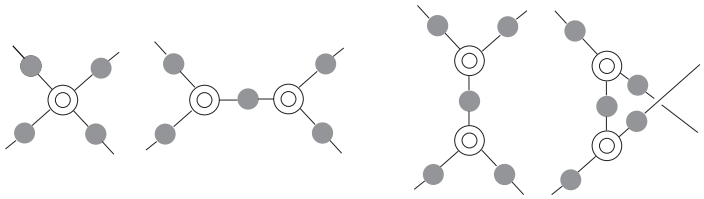
The triple vertex must then be the sum of all 1PI three-point functions:



Three-point graphs which are one-particle reducible always contribute a two-point subgraph to the full propagators of the external legs and a 1PI three-point subgraph to the vertex:



We can now convince ourselves easily that the four-point vertex in Fig. 6.1 contains all one-particle irreducible four-point functions:



The same of course holds for higher multiplicities. Our conclusion is that we can always rearrange the sum of graphs in the perturbative expansion, derived via $W[J]$ and containing both loops and trees, into an equivalent "tree-level graphs only" expansion where the propagator is the full "two-point" function and the vertices are all one-particle irreducible graphs with the same number of external legs as in the vertex.

The important statement is that $W[J] = W_{\Gamma}^{(0)}[J]$ and that all Green's functions can be obtained automatically from the tree-level expansion of a generating functional with the effective action replacing the classical action. Let us verify that the two-, three- and four-point functions are derivable from the effective action $\Gamma[\langle\phi\rangle]$.

We first introduce the short notation

$$\langle\phi\rangle_J(x) \equiv \phi_x, \quad J(x) \equiv J_x.$$

We start from the equation,

$$\frac{\delta\Gamma}{\delta\phi_x} = -J_x.$$

Differentiating with respect to the source, we obtain

$$\begin{aligned}
\delta(x-y) &= -\frac{\delta}{\delta J_y}\frac{\delta\Gamma}{\delta\phi_x} \\
&= -\int d^4z \frac{\delta\phi_z}{\delta J_y}\frac{\delta^2\Gamma}{\delta\phi_z\delta\phi_x} \\
&= \int d^4z \left[\frac{\delta^2\Gamma}{i\delta\phi_x\delta\phi_z}\right]\left[\frac{\delta^2 W}{i\delta J_z\delta J_y}\right].
\end{aligned} \tag{6.15}$$

From the above we see that

$$\frac{1}{i}\frac{\delta^2\Gamma}{\delta\phi_x\delta\phi_y}$$

is the inverse of the full two-point function

$$\Delta(x-y) \equiv \langle 0|\, T\phi(x)\phi(y)\,|0\rangle_{\text{conn}} = \frac{1}{i}\frac{\delta^2 W}{\delta J_x \delta J_y}.$$

Before we compute the three-point function we need two tricks:

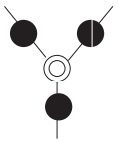
- Chain rule:

$$\begin{aligned}\frac{\delta}{\delta J_x} &= \int d^4z \frac{\delta\phi_z}{\delta J_x}\frac{\delta}{\delta\phi_z} \\ &= \int d^4z \frac{\delta^2 W[J]}{\delta J_x \delta J_z}\frac{\delta}{\delta\phi_z} = i\int d^4z \Delta(x-z)\frac{\delta}{\delta\phi_z}. \end{aligned} \tag{6.16}$$

- Differentiation of an inverse matrix:

$$\begin{aligned} & 1 = MM^{-1} \\ \rightsquigarrow\ & 0 = \frac{\partial\left(MM^{-1}\right)}{\partial\lambda} = \frac{\partial M}{\partial\lambda}M^{-1} + M\frac{\partial M^{-1}}{\partial\lambda} \\ \rightsquigarrow\ & \frac{\partial M^{-1}}{\partial\lambda} = -M^{-1}\frac{\partial M}{\partial\lambda}M^{-1}. \end{aligned} \tag{6.17}$$

We have:



$$\begin{aligned} &= \frac{1}{i^2}\frac{\delta^3 W[J]}{\delta J_{x_1}\delta J_{x_2}\delta J_{x_3}} \\ &= \int d^4y_1 \Delta(x_1-y_1)\frac{\delta}{\delta\phi_{y_1}}\left[\frac{\delta^2 W}{i\delta J_{x_2}\delta J_{x_3}}\right] \\ &= \int d^4y_1 \Delta(x_1-y_1)\frac{\delta}{\delta\phi_{y_1}}\left[\frac{\delta^2\Gamma}{i\delta\phi_{x_2}\delta\phi_{x_3}}\right]^{-1} \\ &= -\int d^4y_1 d^4y_2 d^4y_3 \Delta(x_1-y_1)\left[\frac{\delta^2\Gamma}{i\delta\phi_{x_2}\delta\phi_{y_2}}\right]^{-1}\frac{\delta^3\Gamma}{i\delta\phi_{y_1}\delta\phi_{y_2}\delta\phi_{y_3}}\left[\frac{\delta^2\Gamma}{i\delta\phi_{y_3}\delta\phi_{x_3}}\right]^{-1} \\ &= \int d^4y_1 d^4y_2 d^4y_3 \Delta(x_1-y_1)\Delta(x_2-y_2)\Delta(x_3-y_3)\, i\,\frac{\delta^3\Gamma}{\delta\phi_{y_1}\delta\phi_{y_2}\delta\phi_{y_3}}. \end{aligned} \tag{6.18}$$

Now we may compare the graphs on the left- and right-hand sides of this equation. We have explicitly found that the full three-point function is the convolution of propagators, one for each external leg, and the third derivative of the effective action. From our earlier discussion we know that after we factor out full propagators for the external legs, the remainder is the sum of one-particle irreducible three-point diagrams.

Exercise: *Prove that*

$$i\frac{\delta^4\Gamma}{\delta\phi_{y_1}\delta\phi_{y_2}\delta\phi_{y_3}\delta\phi_{y_4}}$$

is the sum of 1PI four-point functions.

In summary, we can deduce from the quantum effective action all physical predictions of a quantum field theory:

- The second derivative of $\Gamma[\langle\phi\rangle]$ is the inverse propagator. The zeros of the inverse propagator yield the mass values of the physical particles in the theory.
- Higher derivatives of the effective action are 1PI Green's functions. Connecting them with full propagators to form trees, we can derive all connected amplitudes which are required for S-matrix element computations.

Additionally, solving the equation

$$\frac{\delta\Gamma}{\delta\langle\phi\rangle} = 0$$

yields the values of vevs where the effective action is stationary; the true ground state is the solution at which the effective potential V_{eff} defined below is minimal (for constant vevs $\Gamma = -\int d^4x\, V_{\text{eff}}$, so Γ is maximal there). This will serve to define the true ground state of the theory. The location of the minimum will also reveal whether any symmetries of the Lagrangian are broken spontaneously.

6.2 The effective potential

We have just observed that by differentiating the effective action functional with respect to the field vevs, we generate one-particle-irreducible Feynman diagrams. All functional derivatives of $\Gamma[\langle\phi\rangle]$ are therefore represented in terms of Feynman diagrams; if we could compute all these diagrams we could compute the full effective action by summing up all the terms of a Taylor series expansion.

Specifically, we can expand

$$\begin{aligned}\Gamma\left[\langle\phi\rangle\right] &= \sum_{n=1}^{\infty}\frac{1}{n!}\int d^4x_1\ldots d^4x_n\,\frac{\delta^n\Gamma\left[\langle\phi\rangle\right]}{\delta\langle\phi(x_1)\rangle\ldots\delta\langle\phi(x_n)\rangle}\langle\phi(x_1)\rangle\ldots\langle\phi(x_n)\rangle \\ &= \sum_{n=1}^{\infty}\frac{-i}{n!}\int d^4x_1\ldots d^4x_n\,\Gamma^{(n)}(x_1,\ldots,x_n)\langle\phi(x_1)\rangle\ldots\langle\phi(x_n)\rangle\,, \qquad (6.19)\end{aligned}$$

where

$$\Gamma^{(n)}(x_1,\ldots,x_n) \equiv i\frac{\delta^n\Gamma\left[\langle\phi\rangle\right]}{\delta\langle\phi(x_1)\rangle\ldots\delta\langle\phi(x_n)\rangle}, \qquad (6.20)$$

are one-particle-irreducible Green's functions (in coordinate space).

We consider the case where the ground state (vacuum) is translation invariant, i.e. a constant in space-time. There are situations where this is not true (solitons, domain walls, inhomogeneous phases); however, the case of a space-time-independent vacuum is interesting and common. We then have

$$\langle\phi(x)\rangle = \text{constant} \equiv \phi. \qquad (6.21)$$

Remark. Instantons are not an example of a non-translation-invariant vacuum: they are finite-action Euclidean saddle points describing tunnelling, and the resulting θ-vacua are translation invariant. The Green's functions simplify if we use a Fourier transformation

(momentum space),

$$\begin{aligned}
&\int d^4x_1 \dots d^4x_n \Gamma^{(n)}(x_1, \dots, x_n) \left\langle \phi(x_1) \right\rangle \dots \left\langle \phi(x_n) \right\rangle \\
= \;& \phi^n \int d^4x_1 \dots d^4x_n \Gamma^{(n)}(x_1, \dots, x_n) \\
= \;& \phi^n \int d^4x_1 \dots d^4x_n \\
& \int \frac{d^4k_1}{(2\pi)^4} \dots \frac{d^4k_n}{(2\pi)^4} e^{-ik_1x_1} \dots e^{-ik_nx_n} (2\pi)^4 \delta^{(4)}(k_1 + k_2 + \dots + k_n) \\
& \times \tilde{\Gamma}^{(n)}(k_1, \dots, k_n) \\
= \;& \phi^n \int d^4k_1 \dots d^4k_n \delta^{(4)}(k_1) \dots \delta^{(4)}(k_n) (2\pi)^4 \delta^{(4)}(k_1 + k_2 + \dots + k_n) \\
& \times \tilde{\Gamma}^{(n)}(k_1, \dots, k_n) \\
= \;& \phi^n (2\pi)^4 \tilde{\Gamma}^{(n)}(0, 0, \dots, 0) \delta^{(4)}(0).
\end{aligned} \tag{6.22}$$

Notice that we have explicitly shown the delta-function which imposes momentum conservation. The multiple integrations over space-time x_i are simple because of the assumption of an x-independent vev ϕ. The factor $(2\pi)^4\delta^{(4)}(0)$ is the space-time volume,

$$(2\pi)^4 \delta^{(4)}(0) = \int d^4x e^{-i0\cdot x} = \left(\int d^4x \right).$$

We then have for the effective action

$$\Gamma[\phi] = \left(\int d^4x \right) \sum_{n=1}^{\infty} (-i) \frac{\phi^n}{n!} \tilde{\Gamma}^{(n)}(0, 0, \dots, 0). \tag{6.23}$$

The effective potential is defined from the effective action, factoring out the space-time volume,

$$V_{\text{eff}}(\phi) \equiv -\frac{\Gamma[\phi]}{\left(\int d^4x \right)}. \tag{6.24}$$

We obtain

$$V_{\text{eff}}(\phi) = i \sum_{n=1}^{\infty} \frac{\phi^n}{n!} \tilde{\Gamma}^{(n)}(0, 0, \dots, 0). \tag{6.25}$$

Therefore, the recipe to compute the effective potential is:

- Compute all 1PI Green's functions with increasing number of external legs in momentum space, setting all external momenta to zero.
- For each external leg include a power of the classical vev of the corresponding field.
- Sum up the series without forgetting to include the $i/n!$ from the Taylor series expansion.

Let us consider the Lagrangian of a real scalar field with a quartic interaction,

$$\mathcal{L} = \frac{1}{2} \left(\partial_\mu \phi \right)^2 - \frac{m^2}{2} \phi^2 - \frac{\lambda}{4!} \phi^4. \tag{6.26}$$

A computation of the effective potential including all orders in perturbation theory is impossible. We can, however, compute the effective potential easily in the tree and one-loop approximations.

We observe that the only two **1PI** Green's functions that we can write in the tree approximation are the two-point (amputated propagator) and four-point (vertex) functions. From Eq. (6.25) we find

$$V_{\text{eff}}^{\text{tree}} = +\frac{m^2}{2}\phi^2 + \frac{\lambda}{4!}\phi^4. \tag{6.27}$$

The effective potential at tree level is the same as the potential of the classical Lagrangian.

The one-loop computation of the effective potential is left as an exercise. For reference, the one-loop result for this model is $V_{\text{eff}} = \frac{m^2}{2}\phi^2 + \frac{\lambda}{4!}\phi^4 - \frac{i}{2}\int \frac{d^4k}{(2\pi)^4} \log \frac{k^2 - m^2 - \lambda\phi^2/2 + i0}{k^2 - m^2 + i0}$ (up to a ϕ-independent constant), which after renormalisation becomes $\frac{m^2}{2}\phi^2 + \frac{\lambda}{4!}\phi^4 + \frac{(m^2+\lambda\phi^2/2)^2}{64\pi^2}\left(\log\frac{m^2+\lambda\phi^2/2}{\mu^2} - \frac{3}{2}\right)$ in $\overline{\text{MS}}$ (Coleman–Weinberg, Phys. Rev. D7 (1973) 1888; Peskin–Schroeder Sec. 11.4). Beyond perturbation theory the exact V_{eff} is convex; a non-convex perturbative V_{eff} signals that the homogeneous-state assumption fails between the minima.

7

Symmetries of the path integral and the effective action

Our guiding principle in constructing realistic theories for particle interactions is invariance of the classical action under certain symmetries (e.g. BRST symmetry for Yang–Mills theories). Symmetries of the classical action S are not automatically symmetries of the effective action Γ. However, the effective action Γ satisfies very general equations (Slavnov–Taylor identities) due to these classical symmetry constraints.

7.1 Slavnov–Taylor identities

Consider a theory of interacting fields ϕ_i with arbitrary (bosonic or fermionic) statistics. We assume that this theory is symmetric under an infinitesimal symmetry transformation:

$$\phi_i \to \phi_i{}' = \phi_i + \epsilon F_i(x, \phi_i),$$

where ϵ is a small parameter and F_i is usually an ordinary function of the fields ϕ_i and their derivatives. Then, we require that both the action and the path-integral measure of the fields to be invariant under this transformation:

$$\begin{aligned} S[\phi_i + \epsilon F_i(x, \phi_i)] &= S[\phi_i] \\ \mathcal{D}\left(\phi_i + \epsilon F_i(x, \phi_i)\right) &= \mathcal{D}\phi_i. \end{aligned}$$

After transforming the fields, the generating functional is

$$\begin{aligned} Z[J_i] &= \int \mathcal{D}\phi_i' e^{iS[\phi_i']+i\int d^4x \phi_i' J_i} \\ &= \int \mathcal{D}\left(\phi_i + \epsilon F_i(x, \phi_i)\right) e^{iS\left[\phi_i+\epsilon F_i(x,\phi_i)\right]+i\int d^4x\left(\phi_i+\epsilon F_i(x,\phi_i)\right)J_i} \\ &= \int \mathcal{D}\phi_i e^{iS[\phi_i]+i\int d^4x\left(\phi_i+\epsilon F_i(x,\phi_i)\right)J_i}. \end{aligned}$$

We can now expand in the small parameter ϵ,

$$\begin{aligned} Z[J_i] &= \int \mathcal{D}\phi_i e^{iS[\phi_i]+i\int d^4x \phi_i J_i}\left(1 + i\epsilon \int d^4y F_i(y, \phi_i) J_i(y) + \mathcal{O}(\epsilon^2)\right) \\ &= Z[J_i] + i\epsilon \int \mathcal{D}\phi_i \left(\int d^4y F_i(y, \phi_i) J_i(y)\right) e^{iS[\phi_i]+i\int d^4x \phi_i J_i} \\ &\rightsquigarrow \int \mathcal{D}\phi_i \left(\int d^4y F_i(y, \phi_i) J_i(y)\right) e^{iS[\phi_i]+i\int d^4x \phi_i J_i} = 0, \end{aligned} \tag{7.1}$$

or, dividing by the path integral,

$$\int d^4y \left[\frac{\int \mathcal{D}\phi_i F_i(y,\phi_i) e^{iS[\phi_i]+i\int d^4x \phi_i J_i}}{Z[J_i]} \right] J_i(y) = 0. \tag{7.2}$$

In the square brackets we recognise the average of the transformation over all field configurations,

$$\left\langle F_i(y,\phi_i) \right\rangle_J \equiv \frac{\int \mathcal{D}\phi_i F_i(y,\phi_i) e^{iS[\phi_i]+i\int d^4x \phi_i J_i}}{Z[J_i]}. \tag{7.3}$$

We then find the identity

$$\int d^4y \left\langle F_i(y,\phi_i) \right\rangle_J J_i(y) = 0, \tag{7.4}$$

concluding that if there exists an infinitesimal symmetry transformation of the classical action, then there is a constraint on the "**average**" value of the transformation. Eq. (7.4) depends on arbitrary sources, and by differentiating multiple times with respect to the sources, we can obtain an infinite number of identities. These are called Slavnov–Taylor identities; we shall consider an example soon.

7.2 Slavnov–Taylor identities in QED

We now consider an example of Slavnov–Taylor identities in QED. The classical Lagrangian is, in the Lorenz gauge,

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} + \bar{\psi}\left(i\not{D} - m\right)\psi - \frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2. \tag{7.5}$$

The corresponding path integral is

$$Z\left[J^\mu, \bar{\rho}, \rho\right] = \int \mathcal{D}A_\mu \mathcal{D}\bar{\psi}\mathcal{D}\psi e^{i\int d^4x \left[\mathcal{L} + A^\mu J_\mu + \bar{\psi}\rho + \bar{\rho}\psi\right]}. \tag{7.6}$$

An infinitesimal local gauge transformation is (with $D_\mu = \partial_\mu - iqA_\mu$):

$$\begin{aligned} A_\mu &\rightarrow A_\mu - \frac{1}{q}\partial_\mu \Theta(x), \\ \psi &\rightarrow (1 - i\Theta(x))\psi, \\ \bar{\psi} &\rightarrow (1 + i\Theta(x))\bar{\psi}. \end{aligned}$$

The path-integral measure is invariant under the gauge transformation. In the exponent of the path integral we can identify a part which is invariant under these transformations, while the remainder, which includes the gauge-fixing term and the source terms, is not invariant:

$$\mathcal{L}_{\text{non-inv}} = -\frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2 + A^\mu J_\mu + \bar{\psi}\rho + \bar{\rho}\psi. \tag{7.7}$$

By performing a gauge transformation on the path integral we can derive, as before, the Slavnov–Taylor identity

$$\int d^4x \left\langle \delta \mathcal{L}_{\text{non-inv}} \right\rangle = 0. \tag{7.8}$$

We can work out the change in the non-invariant part of the Lagrangian. The gauge-fixing term transforms as

$$\begin{aligned} -\frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2 &\to -\frac{1}{2\xi}\left[\partial_\mu\left(A^\mu - \frac{1}{q}\partial^\mu\Theta\right)\right]^2 \\ &= -\frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2 + \frac{1}{q\xi}\partial_\mu A^\mu \partial^2\Theta(x) + \mathcal{O}\left(\Theta^2\right) \\ \rightsquigarrow \delta\left(-\frac{1}{2\xi}\left(\partial_\mu A^\mu\right)^2\right) &= \frac{1}{q\xi}\partial_\mu A^\mu\partial^2\Theta(x). \end{aligned} \tag{7.9}$$

Adding the variation of the source terms, we obtain

$$\delta\mathcal{L}_{\text{non-inv}} = \frac{1}{q\xi}\partial_\mu A^\mu\partial^2\Theta(x) - i\Theta(x)\bar\rho\psi + i\Theta(x)\bar\psi\rho - \frac{1}{q}J^\mu\partial_\mu\Theta(x). \tag{7.10}$$

The Slavnov–Taylor identity is

$$\begin{aligned} &\int d^4x\left\langle\frac{1}{q\xi}\partial_\mu A^\mu\partial^2\Theta(x) - i\Theta(x)\bar\rho\psi + i\Theta(x)\bar\psi\rho - \frac{1}{q}J^\mu\partial_\mu\Theta(x)\right\rangle = 0 \\ \rightsquigarrow &\int d^4x\left[\frac{1}{q\xi}\partial_\mu\left\langle A^\mu\right\rangle\partial^2\Theta(x) - i\Theta(x)\bar\rho\left\langle\psi\right\rangle + i\Theta(x)\left\langle\bar\psi\right\rangle\rho - \frac{1}{q}J^\mu\partial_\mu\Theta(x)\right] = 0 \\ \rightsquigarrow &\int d^4x\Theta(x)\left[\frac{1}{\xi}\partial_\mu\partial^2\left\langle A^\mu\right\rangle - iq\bar\rho\left\langle\psi\right\rangle + iq\left\langle\bar\psi\right\rangle\rho + \partial_\mu J^\mu\right] = 0, \end{aligned} \tag{7.11}$$

where we have used integration by parts. The above must be valid for arbitrary $\Theta(x)$; therefore the integrand in the square brackets must vanish identically:

$$\frac{1}{\xi}\partial_\mu\partial^2\left\langle A^\mu\right\rangle - iq\bar\rho\left\langle\psi\right\rangle + iq\left\langle\bar\psi\right\rangle\rho + \partial_\mu J^\mu = 0. \tag{7.12}$$

Replacing vacuum expectation values by functional derivatives of the generating functional for connected graphs, $W = -i\ln Z$, we write

$$\frac{1}{\xi}\partial_\mu\partial^2\frac{\delta W}{\delta J_\mu} - iq\bar\rho\frac{\delta W}{\delta\bar\rho} - iq\frac{\delta W}{\delta\rho}\rho + \partial_\mu J^\mu = 0, \tag{7.13}$$

where the functional derivatives are left derivatives for the fermionic sources.

Eq. (7.13) provides constraints for Green's functions in QED at all orders in perturbation theory. We find the simplest example by differentiating this equation with respect to a photon source and then setting all sources to zero,

$$\begin{aligned} &\frac{\delta}{\delta J_\nu(y)}\left(\frac{1}{\xi}\partial_\mu\partial^2\frac{\delta W}{\delta J_\mu} - iq\bar\rho\frac{\delta W}{\delta\bar\rho} - iq\frac{\delta W}{\delta\rho}\rho + \partial_\mu J^\mu\right)\Bigg|_{J,\rho,\bar\rho=0} = 0, \\ \rightsquigarrow &\frac{1}{\xi}\partial_\mu\partial^2\left.\frac{\delta^2 W}{\delta J_\mu(x)\delta J_\nu(y)}\right|_{J,\rho,\bar\rho=0} + \partial^\nu\delta(x-y) = 0 \\ \rightsquigarrow &\frac{1}{\xi}\partial_\mu\partial^2\left\langle 0\right|T A^\mu(x)A^\nu(y)\left|0\right\rangle = i\partial^\nu\delta(x-y). \end{aligned} \tag{7.14}$$

We now write the Fourier representations

$$\delta(x-y) = \int\frac{d^4k}{(2\pi)^4}e^{-ik(x-y)},$$

and

$$\langle 0|\, TA^{\mu}(x)A^{\nu}(y)\,|0\rangle = \int \frac{d^4k}{(2\pi)^4} e^{-ik(x-y)} D^{\mu\nu}(k).$$

Substituting into Eq. (7.14) we find that the photon propagator in momentum space must satisfy

$$k_\mu D^{\mu\nu}(k) = -i\xi \frac{k^\nu}{k^2}. \tag{7.15}$$

We can write, in complete generality,

$$D_{\mu\nu} = A(k^2) g_{\mu\nu} + B(k^2) \frac{k_\mu k_\nu}{k^2}. \tag{7.16}$$

Substituting into Eq. (7.15), we find

$$A(k^2) + B(k^2) = -i\frac{\xi}{k^2}. \tag{7.17}$$

Thus, the photon propagator in momentum space has the form

$$D^{\mu\nu}(k) = A(k^2)\left(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\right) - i\frac{\xi}{k^2}\frac{k^\mu k^\nu}{k^2}. \tag{7.18}$$

This result is valid at all orders in perturbation theory.

Notice that the term which depends on the gauge-fixing parameter is fully known. We can compare this with the result at leading order in perturbation theory,

$$D^{\mu\nu}(k) = \frac{-i}{k^2}\left(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\right) - i\frac{\xi}{k^2}\frac{k^\mu k^\nu}{k^2} + O(q^2). \tag{7.19}$$

We can see that the gauge-fixing contribution is fully accounted for in the leading-order result, and therefore it is not modified at higher orders in perturbation theory. Higher-order corrections modify only the function $A(k^2)$. For this reason, the gauge-fixing parameter ξ does not receive any renormalisation. *Remark.* More precisely, the gauge-fixing term needs no counterterm of its own. In terms of bare quantities the longitudinal part is $-i\xi_0 k^\mu k^\nu/k^4$ for the bare field $A_0 = Z_3^{1/2} A$, so the renormalised gauge parameter is $\xi = \xi_0/Z_3$, i.e. $Z_\xi = Z_3$, the standard result of QED renormalisation. The statement "ξ is not renormalised" refers to this.

A second important observation is that the part of the propagator which does not depend on ξ,

$$D_T^{\mu\nu} = A(k^2)\left(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\right),$$

is transverse to the photon momentum. Indeed, we easily find that

$$D_T^{\mu\nu}(k) k_\mu = 0.$$

7.3 Symmetry constraints on the effective action

The general Slavnov–Taylor identity of Eq. (7.4) tells us that the effective action possesses a symmetry. Substituting

$$J_i(y) = -\frac{\delta\Gamma}{\delta\left\langle\phi_i(y)\right\rangle_J},$$

we obtain

$$\int d^4y \left\langle F_i(y,\phi_i)\right\rangle_J \frac{\delta\Gamma}{\delta\left\langle\phi_i(y)\right\rangle_J} = 0. \tag{7.20}$$

Equivalently,

$$\begin{aligned}
&\frac{\Gamma\left[\left\langle\phi_i\right\rangle_J\right]}{\epsilon} + \int d^4y \left\langle F_i(y,\phi_i)\right\rangle_J \frac{\delta\Gamma}{\delta\left\langle\phi_i(y)\right\rangle_J} = \frac{\Gamma\left[\left\langle\phi_i\right\rangle_J\right]}{\epsilon} \\
\rightsquigarrow\ &\Gamma\left[\left\langle\phi_i\right\rangle_J\right] + \epsilon\int d^4y \left\langle F_i(y,\phi_i)\right\rangle_J \frac{\delta\Gamma}{\delta\left\langle\phi_i(y)\right\rangle_J} = \Gamma\left[\left\langle\phi_i\right\rangle_J\right] \\
\rightsquigarrow\ &\Gamma\left[\left\langle\phi_i\right\rangle_J + \epsilon\left\langle F_i(y,\phi_i)\right\rangle_J\right] = \Gamma\left[\left\langle\phi_i\right\rangle_J\right] + O\left(\epsilon^2\right).
\end{aligned} \tag{7.21}$$

Therefore, the effective action is symmetric under the transformation

$$\left\langle\phi_i\right\rangle \to \left\langle\phi_i\right\rangle' = \left\langle\phi_i\right\rangle + \epsilon\left\langle F_i\left(\phi_i\right)\right\rangle. \tag{7.22}$$

We recall that the classical action is symmetric under the transformation

$$\phi_i \to \phi_i' = \phi_i + \epsilon F_i\left(\phi_i\right). \tag{7.23}$$

Are these two transformations the same? In other words, is $F_i = \langle F_i\rangle$? In general they are not! The symmetries of the classical action are in general not symmetries of the quantum effective action. Consider, as an example, a classical action symmetric under the field transformation

$$\phi(x) \to \phi(x) + \epsilon\phi^2(x).$$

The quantum effective action should be symmetric under the transformation

$$\left\langle\phi(x)\right\rangle \to \left\langle\phi(x)\right\rangle + \epsilon\left\langle\phi^2(x)\right\rangle.$$

Given that

$$\left\langle\phi(x)\right\rangle^2 \neq \left\langle\phi^2(x)\right\rangle,$$

the two transformations are different.

Nevertheless, we can identify many symmetries in classical actions for realistic field theories which are linear:

$$F_i[\phi_i,x] = c_i(x) + \int d^4y T^{ij}(x,y)\phi_j(y). \tag{7.24}$$

The equivalent symmetry transformation for the effective action is

$$\begin{aligned}
\left\langle F_i[\phi_i,x]\right\rangle &= \left\langle c_i(x) + \int d^4y T^{ij}(x,y)\phi_j(y)\right\rangle \\
&= c_i(x) + \int d^4y T^{ij}(x,y)\left\langle\phi_j(y)\right\rangle \\
&= F_i[\left\langle\phi_i\right\rangle,x],
\end{aligned} \tag{7.25}$$

and it is identical to the classical transformation. It is useful to remember that **linear symmetry transformations of the classical action are automatically symmetry transformations of the effective action.**

7.4 Constraints on the effective action from BRST: the Zinn-Justin equation

The BRST transformations are not linear; therefore they are only a symmetry of the classical action [1] and not of the effective action. Nevertheless, the effective action is constrained by the BRST symmetry of the classical Lagrangian, Eq. (7.4). For these nilpotent transformations Eq. (7.4) takes a very special form, the so-called Zinn-Justin equation.

We start with a classical action $S[\phi_i]$ of fields ϕ_i which is symmetric under the BRST transformation

$$\delta_\theta \phi_i = \theta B_i. \tag{7.26}$$

Since the transformation is nilpotent, we also have

$$\begin{aligned} & \delta_{\theta'} \delta_\theta \phi_i = 0 \\ \rightsquigarrow \quad & \delta_{\theta'} B_i = 0. \end{aligned} \tag{7.27}$$

We realise that, because of Eq. (7.27), there is a more general action which has the same symmetry as the original $S[\phi_i]$. It is easy to verify that the action

$$S[\phi_i, K_i] = S[\phi_i] + \int d^4x B_i(x) K_i(x), \tag{7.28}$$

is indeed symmetric under the same transformation.

The functions K_i are arbitrary (sources). We recall, however, that the functions B_i have the opposite statistics to the corresponding field ϕ_i. Since the product $B_i K_i$ must be bosonic (as the action S is), we deduce that the source K_i and the field ϕ_i also have opposite statistics.

We can write the generating functional W for connected graphs as

$$e^{iW[J_i,K_i]} = \int \mathcal{D}\phi_i e^{iS[\phi_i]+i\int d^4x B_i K_i + i\int d^4x \phi_i J_i}. \tag{7.29}$$

The fields ϕ_i may be bosonic or fermionic, therefore the order in which we write the terms in the exponent is important. Conventionally, we have placed the sources J_i, K_i on the right.

The vacuum expectation value $\langle \phi_i \rangle$ is given by

$$\begin{aligned} \langle \phi_i(y) \rangle &= \frac{\int \mathcal{D}\phi_i \, \phi_i(y) \, e^{iS[\phi_i]+i\int d^4x B_i K_i + i\int d^4x \phi_i J_i}}{\int \mathcal{D}\phi_i e^{iS[\phi_i]+i\int d^4x B_i K_i + i\int d^4x \phi_i J_i}} \\ &= \frac{\delta_R W[J_i, K_i]}{\delta J_i(y)}. \end{aligned} \tag{7.30}$$

This is an implicit relation among $J_i, K_i, \langle \phi_i \rangle$, and we will consider $K_i, \langle \phi_i \rangle$ as independent variables, and the sources J_i as being expressed in terms of these two variables,

$$J_i = J_i(\langle \phi_i \rangle, K_i).$$

[1] From now on, by "classical action" for a gauge theory we mean the action obtained after gauge-fixing using the Faddeev–Popov method.

The exact form of $J_i(\langle\phi_i\rangle, K_i)$ can be found if we evaluate explicitly the effective action

$$\Gamma\left[\langle\phi_i\rangle, K_i\right] = W\left[J_i(\langle\phi_i\rangle, K_i), K_i\right] - \int d^4x \langle\phi_i\rangle J_i(\langle\phi_i\rangle, K_i), \tag{7.31}$$

and take a left derivative,

$$\frac{\delta_L\Gamma\left[\langle\phi_i\rangle, K_i\right]}{\delta\langle\phi_i(x)\rangle} = -J_i(\langle\phi_i\rangle, K_i)(x). \tag{7.32}$$

We now compute the derivative of the effective action with respect to the sources K_i:

$$\begin{aligned}
\frac{\delta_R\Gamma\left[\langle\phi_i\rangle, K_i\right]}{\delta K_i(x)} &= \frac{\delta_R}{\delta K_i(x)}\left(W\left[J_i(\langle\phi_i\rangle, K_i), K_i\right] - \int d^4x \langle\phi_i\rangle J_i(\langle\phi_i\rangle, K_i)\right) \\
&= \left.\frac{\delta_R W[J_i, K_i]}{\delta K_i(x)}\right|_{J_i = J_i(\langle\phi_i\rangle, K_i)} \\
&\quad + \int d^4y \left(\left.\frac{\delta_R W[J_i, K_i]}{\delta J_m(y)}\right|_{J_m = J_m(\langle\phi_i\rangle, K_i)}\right)\left(\frac{\delta_R J_m(\langle\phi_i\rangle, K_i)(y)}{\delta K_i(x)}\right) \\
&\quad - \int d^4y \langle\phi_m\rangle(y)\frac{\delta_R J_m(\langle\phi_i\rangle, K_i)(y)}{\delta K_i(x)} \\
&= \left.\frac{\delta_R W[J_i, K_i]}{\delta K_i(x)}\right|_{J_i = J_i(\langle\phi_i\rangle, K_i)} \\
&\quad + \int d^4y \langle\phi_m\rangle(y)\frac{\delta_R J_m(\langle\phi_i\rangle, K_i)(y)}{\delta K_i(x)} \\
&\quad - \int d^4y \langle\phi_m\rangle(y)\frac{\delta_R J_m(\langle\phi_i\rangle, K_i)(y)}{\delta K_i(x)} \\
&= \left.\frac{\delta_R W[J_i, K_i]}{\delta K_i(x)}\right|_{J_i = J_i(\langle\phi_i\rangle, K_i)} \\
&= -i\,\frac{\delta_R}{\delta K_i(x)} \ln\left(\int \mathcal{D}\phi_i e^{iS[\phi_i] + i\int d^4x \phi_i J_i + i\int d^4x B_i K_i}\right)\Bigg|_{J_i = \text{fixed}} \\
&= \langle B_i\rangle .
\end{aligned} \tag{7.33}$$

From the general Slavnov–Taylor identity we have

$$\begin{aligned}
&\int d^4x \langle B_i\rangle J_i = 0 \\
\rightsquigarrow\quad &\int d^4x \langle B_i\rangle \frac{\delta_L\Gamma}{\delta\langle\phi_i\rangle} = 0 \\
\rightsquigarrow\quad &\int d^4x \frac{\delta_R\Gamma}{\delta K_i}\frac{\delta_L\Gamma}{\delta\langle\phi_i\rangle} = 0.
\end{aligned} \tag{7.34}$$

This is a constraint which depends only on the effective action $\Gamma[\langle\phi_i\rangle, K_i]$ (Zinn-Justin equation). It is a very useful form to study the consequences of symmetry for the effective action, especially in connection with renormalisation proofs and the study of anomalies.

For later use, we define the product

$$(F, G) \equiv \int d^4x \left(\frac{\delta_R F}{\delta K_i}\frac{\delta_L G}{\delta\langle\phi_i\rangle} - \frac{\delta_R F}{\delta\langle\phi_i\rangle}\frac{\delta_L G}{\delta K_i}\right), \tag{7.35}$$

for functionals $F[\langle\phi_i\rangle, K_i]$, $G[\langle\phi_i\rangle, K_i]$ of the functions $\langle\phi_i\rangle, K_i$. Recall that $\langle\phi_i\rangle$ and K_i have opposite statistics. Then, the Zinn-Justin equation can be written as

$$(\Gamma, \Gamma) = 0. \tag{7.36}$$

8

Spontaneous symmetry breaking

Symmetry transformations that leave the effective action invariant need not be symmetries of the physical states and of the vacuum state. Such symmetries are "spontaneously broken". Spontaneous symmetry breaking is associated with a degeneracy of the ground state (vacuum). Let us assume that an effective action is symmetric under

$$\langle\phi\rangle \to -\langle\phi\rangle \rightsquigarrow \Gamma[\langle\phi\rangle] = \Gamma[-\langle\phi\rangle],$$

a linear symmetry which, as we have seen in the previous chapter, is inherited unchanged from the classical action. Assume also that the physical vacuum expectation value of the field, determined by

$$\langle\phi\rangle_a \quad : \quad \frac{\delta\Gamma}{\delta\langle\phi\rangle_a} = 0, \tag{8.1}$$

is non-vanishing, $\langle\phi\rangle_a \neq 0$. In other words, there exists a state $|v_a\rangle$ with

$$\langle\phi\rangle_a \equiv \frac{\langle v_a|\,\hat{\phi}(x)\,|v_a\rangle}{\langle v_a|\,v_a\rangle} \neq 0, \tag{8.2}$$

for which $V_{\text{eff}}(\langle\phi\rangle_a)$ is a minimum, i.e. $\Gamma[\langle\phi\rangle_a] \equiv \Gamma_a$ is a maximum. Then there must be a second value of the field vev, in a different state, which gives the same value for the effective action:

$$\exists\,|b\rangle \quad : \quad \langle\phi\rangle_b = -\langle\phi\rangle_a \quad \text{with} \quad \Gamma[\langle\phi\rangle_a] = \Gamma[\langle\phi\rangle_b] = \text{maximum}.$$

So, while the transformation $\phi \to -\phi$ preserves the action and the effective action, it does not preserve the states and transforms one state into another:

$$|v_a\rangle \to |v_b\rangle\,.$$

The symmetry is broken as long as the system is in one of the degenerate states.

8.1 Goldstone theorem

Consider a symmetry transformation

$$\phi_n(x) \to \phi'_n(x) = \phi_n(x) + i\epsilon \sum_m t_{nm}\phi_m(x), \tag{8.3}$$

with ϵ a small parameter and t_{nm} generators of the transformation. The transformation leaves the effective action intact:

$$\Gamma\left[\langle\phi_n(x)\rangle\right] = \Gamma\left[\langle\phi'_n(x)\rangle\right]. \tag{8.4}$$

The symmetry of the effective action gives rise to Slavnov–Taylor identities:

$$\int d^4x \frac{\delta\Gamma}{\delta\left\langle\phi_n(x)\right\rangle} t_{nm}\left\langle\phi_m(x)\right\rangle = 0. \tag{8.5}$$

Taking a further derivative with respect to $\left\langle\phi_l(y)\right\rangle$, we have

$$0 = \int d^4x \frac{\delta^2\Gamma}{\delta\left\langle\phi_l(y)\right\rangle \delta\left\langle\phi_n(x)\right\rangle} t_{nm}\left\langle\phi_m(x)\right\rangle + \frac{\delta\Gamma}{\delta\left\langle\phi_n(y)\right\rangle} t_{nl}. \tag{8.6}$$

For physical systems with zero external sources, $J_n(x) = 0$, we have

$$\frac{\delta\Gamma}{\delta\left\langle\phi_l(y)\right\rangle} = 0, \tag{8.7}$$

and we arrive at the equation

$$0 = \int d^4x \frac{\delta^2\Gamma}{\delta\left\langle\phi_l(y)\right\rangle \delta\left\langle\phi_n(x)\right\rangle} t_{nm}\left\langle\phi_m(x)\right\rangle. \tag{8.8}$$

We now make the important assumption that the vacuum state $|\Omega\rangle$ is translation invariant. We then find that the vacuum expectation value of the field is the same at all space-time points. Since we can find the value of the field operator at a space-time point x from the value of the field at the origin with a translation, using the momentum operator as a generator, we can write

$$\begin{aligned}\left\langle\phi(x)\right\rangle &= \langle\Omega|\,\hat{\phi}(x)\,|\Omega\rangle \\ &= \langle\Omega|\,e^{i\hat{P}x}\hat{\phi}(0)e^{-i\hat{P}x}\,|\Omega\rangle \\ &= \langle\Omega|\,\hat{\phi}(0)\,|\Omega\rangle = \text{ constant } \equiv \left\langle\phi\right\rangle.\end{aligned} \tag{8.9}$$

Then the effective action can be written as

$$\Gamma\left[\left\langle\phi_n\right\rangle\right] = -\int d^4x V_{\text{eff}}(\left\langle\phi_n\right\rangle) = -\left(\int d^4x\right) V_{\text{eff}}(\left\langle\phi_n\right\rangle). \tag{8.10}$$

Eq. (8.8) then yields for the effective potential V_{eff} the following constraint:

$$\sum_{nm} t_{nm}\left\langle\phi_m\right\rangle \frac{\partial^2 V_{\text{eff}}}{\partial\left\langle\phi_l\right\rangle \partial\left\langle\phi_n\right\rangle} = 0. \tag{8.11}$$

This is a constraint on the mass spectrum of the theory. To see this, we recall that the second derivative of the effective action is the inverse of a two-point Green's function:

$$
\begin{aligned}
& \frac{\delta W[J]}{\delta J_m(x)} = \left\langle \phi_m(x) \right\rangle \\
\rightsquigarrow \quad & \frac{\delta^2 W[J]}{\delta \left\langle \phi_n(y) \right\rangle \delta J_m(x)} = \delta(x-y)\delta_{nm} \\
& \cdots \\
\rightsquigarrow \quad & \int d^4z \, \langle \Omega | \, T\phi_n(y)\phi_k(z) \, | \Omega \rangle_{\mathrm{c}} \, \frac{\delta^2 \Gamma}{\delta \left\langle \phi_k(z) \right\rangle \delta \left\langle \phi_m(x) \right\rangle} = i \, \delta(x-y)\delta_{nm} \\
\rightsquigarrow \quad & \int d^4z d^4y \, \langle \Omega | \, T\phi_n(y)\phi_k(z) \, | \Omega \rangle_{\mathrm{c}} \, \frac{\delta^2 \Gamma}{\delta \left\langle \phi_k(z) \right\rangle \delta \left\langle \phi_m(x) \right\rangle} = i \, \delta_{nm} \\
\rightsquigarrow \quad & \int d^4z d^4y \, \langle \Omega | \, T\phi_n(y)\phi_k(z) \, | \Omega \rangle_{\mathrm{c}} \, \frac{\partial^2 V_{\mathrm{eff}}}{\partial \left\langle \phi_k \right\rangle \partial \left\langle \phi_m \right\rangle} \delta(z-x) = -i \, \delta_{nm} \\
\rightsquigarrow \quad & \frac{\partial^2 V_{\mathrm{eff}}}{\partial \left\langle \phi_k \right\rangle \partial \left\langle \phi_m \right\rangle} \int d^4y \, \langle \Omega | \, T\phi_n(y)\phi_k(x) \, | \Omega \rangle_{\mathrm{c}} = -i \, \delta_{nm}.
\end{aligned}
\tag{8.12}
$$

Substituting the Fourier transform of the two-point function,

$$
\langle \Omega | \, T\phi_n(y)\phi_k(x) \, | \Omega \rangle_{\mathrm{c}} = \int \frac{d^4p}{(2\pi)^4} D_{nk}(p^2) e^{-ip\cdot(x-y)}, \tag{8.13}
$$

the integration over the y variable yields a delta function setting the momentum to $p^\mu = 0$. We therefore have

$$
D_{nk}(0) \frac{\partial^2 V_{\mathrm{eff}}}{\partial \left\langle \phi_k \right\rangle \partial \left\langle \phi_m \right\rangle} = -i \, \delta_{nm}. \tag{8.14}
$$

or, equivalently,

$$
\frac{\partial^2 V_{\mathrm{eff}}}{\partial \left\langle \phi_n \right\rangle \partial \left\langle \phi_m \right\rangle} = -i \, D^{-1}_{nm}(0). \tag{8.15}
$$

Eq. (8.11) then yields

$$
\sum_{nm} D^{-1}_{ln}(0) t_{nm} \left\langle \phi_m \right\rangle = 0. \tag{8.16}
$$

When is this equation satisfied? Let us write the combination $\sum_m t_{nm} \left\langle \phi_m \right\rangle = \delta \left\langle \phi_n \right\rangle$ as the variation of the vev under the symmetry transformation. Then Eq. (8.16) becomes

$$
\sum_{n} D^{-1}_{ln}(0) \delta \left\langle \phi_n \right\rangle = 0. \tag{8.17}
$$

If the transformation leaves the vacuum state and, thus, the vacuum expectation values of the fields invariant, $\delta \left\langle \phi_n \right\rangle = 0$, then Eq. (8.17) is fulfilled. What if the symmetry is broken and the symmetry transformation of the effective action changes the vacuum, so that there are some $\delta \left\langle \phi_i \right\rangle \neq 0$? Let us rewrite Eq. (8.17) in matrix notation:

$$
\begin{pmatrix} & D^{-1}_{ln}(0) & \end{pmatrix} \begin{pmatrix} \delta \left\langle \phi_n \right\rangle \end{pmatrix} = \mathbf{0} \begin{pmatrix} \delta \left\langle \phi_n \right\rangle \end{pmatrix}. \tag{8.18}
$$

We observe that the matrix $D^{-1}_{nl}(0)$ has zero eigenvalues, as many as the independent vectors $\delta \left\langle \phi_n \right\rangle$ which are non-vanishing. In the simplest case of only one field, the inverse

propagator of the field at zero momentum is proportional to the squared mass of the particle excitation of the field:

$$D(p) = \frac{iZ}{p^2 - m^2} + \text{continuum} \rightsquigarrow D^{-1}(0) \propto m^2.$$

In general, $D^{-1}_{nl}(0)$ is the mass matrix of the theory. Redefining the fields appropriately, $\phi_n = R_{nm}\tilde{\phi}_m$, eliminates the non-diagonal terms, and the diagonal terms, the eigenvalues of the matrix, are (up to the factor i and the wave-function normalisations Z, cf. $-iD^{-1}(0) = m^2/Z$) the squared masses of the physical particle excitations of the fields $\tilde{\phi}_i$.

We have just proven Goldstone's theorem: for each independent non-vanishing variation of the vacuum, $\delta\langle\phi_n\rangle = \sum_m T_{nm}\langle\phi_m\rangle$, there exists a massless particle in the spectrum of the theory. The symmetry generators T_{nm} which change the vev of the fields are called "broken" generators. There is an alternative proof of Goldstone's theorem, due to Weinberg, which we present in Part IV, Chapter 2. This proof also demonstrates that

- The massless states are one-particle states.
- They are also invariant under rotations and correspond to spin-0 particles, the so-called Goldstone bosons.
- The Goldstone bosons have the same "quantum numbers" as the conserved currents corresponding to the broken generators.

Goldstone's theorem seems very powerful and its proof appears to leave no room for exceptions. Nevertheless, we will soon be able to find a loophole: it is possible to have spontaneous symmetry breaking without giving rise to massless particles. We note that our proof requires a manifestly covariant, translation-invariant formulation as well as states with positive norm. These requirements cannot be satisfied simultaneously in quantum theories with local gauge invariance. *Remark.* The precise statement is that a gauge theory cannot be quantised with both manifest Lorentz covariance (locality) and a positive-definite state space: in covariant gauges the Goldstone theorem holds but the massless states have zero norm and decouple from the physical (BRST-invariant) states, while in non-covariant gauges such as the Coulomb or unitary gauge the norm is positive but the proof fails because the gauge-fixed theory is not manifestly covariant (Guralnik, Hagen, Kibble 1964; Weinberg Vol. II, Sec. 21.1). Moreover, by Elitzur's theorem a local gauge symmetry cannot be broken spontaneously; what is "broken" is the global part of the symmetry, or a gauge-fixed remnant of it.

8.2 General broken global symmetries

Let us assume a pattern of spontaneous symmetry breaking

$$G \to H,$$

where G is the symmetry group that leaves the effective action invariant and H a subgroup of G which leaves the vacuum invariant. We also assume that the symmetry is global. In other words, the effective action remains invariant, $\Gamma[\psi_n] = \Gamma[\psi'_n]$, for

$$\psi'_n = \sum_m g_{nm}\psi_m, \quad \frac{\partial g_{nm}}{\partial x^\mu} = 0, \quad g_{nm} \in G, \tag{8.19}$$

and the vacuum remains invariant,

$$\sum_m h_{nm} \left\langle \psi_m \right\rangle = \left\langle \psi_n \right\rangle , \quad \forall\, h_{nm} \in H. \tag{8.20}$$

According to Goldstone's theorem, the mass matrix of the theory has zero eigenvalues for the eigenvectors

$$\sum_m T^a_{nm} \left\langle \psi_m \right\rangle \equiv \delta \left\langle \psi_n \right\rangle , \tag{8.21}$$

where T^a_{nm} is a broken generator.

Which independent linear combinations of the fields in the Lagrangian of the theory correspond to Goldstone fields and which do not? We shall prove that all fields ψ_n (including non-Goldstones) can be obtained from Goldstone-free fields $\tilde{\psi}_n$ by performing a *local* group transformation:

$$\psi_n(x) = \sum_m \gamma_{nm}(x) \tilde{\psi}_m(x). \tag{8.22}$$

We start by observing that Goldstone-free field combinations $\tilde{\psi}_n$ (the "heavy" fields of the theory) must be orthogonal to the vectors of Eq. (8.21), that is,

$$\sum_{nm} \tilde{\psi}_n(x) T^a_{nm} \left\langle \psi_m \right\rangle = 0. \tag{8.23}$$

Without loss of generality, we will assume that the elements $g \in G$ belong to a real and orthogonal representation of the group which is compact. Then the quantity

$$V_\psi(g) = \psi_n g_{nm} \left\langle \psi_m \right\rangle \tag{8.24}$$

is a bounded, continuous, real-valued function on the compact group, and it therefore attains a maximum. Let us now choose $g = \gamma(x)$ such that $V_{\psi(x)}(g)$ is a maximum at every space-time point x. Then, under a small variation of the group parameter,

$$\delta\gamma_{nm} = i \sum_a \epsilon^a \gamma_{nl} T^a_{lm}, \tag{8.25}$$

$V_\psi(g)$ is stationary:

$$0 = \delta V_\psi(g) = i \sum_a \epsilon^a \sum_{nlm} \psi_n(x) \gamma_{nl}(x) T^a_{lm} \left\langle \psi_m \right\rangle . \tag{8.26}$$

Recalling that we have chosen an orthogonal and real representation of the group, we have

$$[\gamma_{nl}] = [\gamma_{ln}]^{-1} . \tag{8.27}$$

Thus,

$$0 = i \sum_a \epsilon^a \sum_{lm} \left[\sum_n \gamma^{-1}_{ln} \psi_n \right] \left(T^a_{lm} \left\langle \psi_m \right\rangle \right) . \tag{8.28}$$

Therefore, the field combinations

$$\tilde{\psi}_l = \sum_n \gamma^{-1}_{ln} \psi_n \tag{8.29}$$

are orthogonal to the vectors

$$\sum_m T^a_{lm} \left\langle \psi_m \right\rangle = \delta \left\langle \psi_l \right\rangle , \tag{8.30}$$

and they are **not** Goldstone bosons.

Let us rewrite the Lagrangian of the theory by making the substitution which we have just found,

$$\psi(x) = \gamma(x)\tilde{\psi}(x), \tag{8.31}$$

rewriting the fields of the theory as explicit non-Goldstones $\tilde{\psi}(x)$ and the remaining Goldstone fields contained in $\gamma(x)$. We recall that the Lagrangian is invariant only under a global gauge transformation, while the above transformation is a local gauge transformation which does **not** leave the Lagrangian invariant. We have

$$\begin{aligned} \mathcal{L}\left[\gamma(x)\tilde{\psi}(x)\right] &= \mathcal{L}\left[\gamma(x_0)\tilde{\psi}(x)\right] \\ &\quad +\text{terms with derivatives of } \gamma(x), \tilde{\psi}(x). \end{aligned} \tag{8.32}$$

Owing to the global gauge invariance of the theory, $\mathcal{L}\left[\gamma(x_0)\tilde{\psi}(x)\right] = \mathcal{L}\left[\tilde{\psi}(x)\right]$, we have

$$\begin{aligned} \mathcal{L}\left[\gamma(x)\tilde{\psi}(x)\right] &= \mathcal{L}\left[\tilde{\psi}(x)\right] \\ &\quad +\text{terms with derivatives of } \gamma(x), \tilde{\psi}(x), \end{aligned} \tag{8.33}$$

where the first term does not have any Goldstone bosons. Goldstone bosons appear only through derivatives. This forbids mass terms

$$m_B^2 B(x)B(x)$$

for them. Also, at low energies, Goldstone interactions vanish. Indeed, the Feynman rules for fields that appear only through derivatives are proportional to the momenta of the particles:

$$\partial_\mu \gamma(x) \to \partial_\mu B(x) \to p_\mu \qquad \text{(in Feynman rules)},$$

and vanish for zero momenta $p^\mu \to 0$.

8.3 Spontaneous symmetry breaking of local gauge symmetries

Let us now assume that our Lagrangian is invariant under a local gauge symmetry. Repeating the reasoning of the previous section and rewriting

$$\psi = \gamma\tilde{\psi},$$

we have

$$\mathcal{L}\left[\gamma(x)\tilde{\psi}(x)\right] = \mathcal{L}\left[\tilde{\psi}(x)\right]. \tag{8.34}$$

In other words, our carefully selected gauge transformation eliminates all Goldstone boson fields from the Lagrangian. We have just found an exception to Goldstone's theorem in theories with local gauge invariance, where the symmetry is spontaneously broken but there are no physical massless Goldstone fields due to the breaking of the symmetry. The rewriting $\psi = \gamma\tilde{\psi}$ is equivalent to choosing the gauge-fixing condition

$$\tilde{\psi} \cdot \left(T^a \left\langle \psi \right\rangle\right) = 0. \tag{8.35}$$

Lagrangians which are locally invariant under a continuous symmetry transformation require gauge bosons in order to form covariant derivatives. Let us look at the quadratic terms in the covariant derivatives.

For simplicity, we consider a gauge group with a single coupling g and real scalar fields ψ_n in a real, orthogonal representation, as in the previous section, with Hermitian (thus imaginary and antisymmetric) generators T^a. With the conventions of the chapter on non-abelian gauge theories, the covariant derivative is

$$D_\mu \psi_n = \partial_\mu \psi_n - igA^a_\mu T^a_{nm} \psi_m, \tag{8.36}$$

and the kinetic term of the scalars is $\frac{1}{2}\left(D_\mu\psi\right)^T \left(D^\mu\psi\right)$. Notice that the rewriting $\psi = \gamma\tilde{\psi}$ is now a gauge transformation, and it acts also on the gauge fields. The Lagrangian is the same function of $\tilde{\psi}$ and of the transformed gauge fields, which we denote again by A^a_μ. We expand around the vacuum,

$$\tilde{\psi} = \left\langle\psi\right\rangle + \psi', \qquad \psi' \cdot \left(T^a \left\langle\psi\right\rangle\right) = 0, \tag{8.37}$$

where ψ' contains only the "heavy" fields. We find:

$$D_\mu\tilde{\psi} = \partial_\mu\psi' - igA^a_\mu T^a \left\langle\psi\right\rangle - igA^a_\mu T^a \psi'. \tag{8.38}$$

The term which is quadratic in the gauge fields and does not contain any scalar field is

$$\begin{aligned} \frac{1}{2}\left(D_\mu\tilde{\psi}\right)^T \left(D^\mu\tilde{\psi}\right) &= \frac{1}{2}\left(-ig\right)^2 A^a_\mu A^{b\mu} \left\langle\psi\right\rangle^T \left(T^a\right)^T T^b \left\langle\psi\right\rangle + \dots \\ &= \frac{1}{2} g^2 \left\langle\psi\right\rangle^T T^a T^b \left\langle\psi\right\rangle A^a_\mu A^{b\mu} + \dots, \end{aligned} \tag{8.39}$$

where we have used $\left(T^a\right)^T = -T^a$. We recall that a free massive vector field has the Lagrangian $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu$. Therefore, the gauge bosons have acquired the mass matrix

$$\left(M^2\right)_{ab} = g^2 \left\langle\psi\right\rangle^T T^a T^b \left\langle\psi\right\rangle. \tag{8.40}$$

For a product of groups with different couplings, as in the Standard Model, we replace $g^2 \to g_a g_b$. The physical masses are the eigenvalues of M^2, and the physical gauge fields are the corresponding orthogonal combinations of the A^a_μ.

Which gauge bosons remain massless? Let us define the real vectors $u^a \equiv iT^a \left\langle\psi\right\rangle$. Then $\left(M^2\right)_{ab} = g^2\, u^a \cdot u^b$, which is a real, symmetric and positive semi-definite matrix. A combination $c_a A^a_\mu$ is massless if and only if $\sum_a c_a T^a \left\langle\psi\right\rangle = 0$, i.e. if $c_a T^a$ is a generator of the unbroken subgroup H. In other words, the number of massive gauge bosons is equal to the number of broken generators, $\dim G - \dim H$, which is exactly the number of Goldstone fields that we have removed from the Lagrangian with the gauge transformation $\gamma(x)$. The counting of degrees of freedom works out: a massless vector boson has two polarisations, while a massive one has three. The Goldstone boson has not disappeared; it has been "eaten" by the gauge boson and it has become its longitudinal polarisation. This is the **Higgs mechanism**.

Exercise: *Prove that M^2 is positive semi-definite and that its null vectors are the generators of H.*

The gauge $\tilde{\psi} \cdot \left(T^a \left\langle\psi\right\rangle\right) = 0$ is known as the **unitary gauge**. Only physical particles appear in it, but the propagator of a massive vector boson,

$$\frac{i\left(-g^{\mu\nu} + k^\mu k^\nu / m^2\right)}{k^2 - m^2 + i\epsilon}, \tag{8.41}$$

does not fall off at large momenta, and power counting wrongly suggests that the theory is not renormalisable. Without fixing the unitary gauge, we can write $\psi = \langle\psi\rangle + \psi'$ with all components in ψ'. The kinetic term then contains a mixing of the gauge bosons with the Goldstone directions, $-gA^a_\mu\,\partial^\mu\left(u^a\cdot\psi'\right)$. 't Hooft's R_ξ **gauges** remove it with the gauge-fixing term

$$\mathcal{L}_{\rm gf} = -\frac{1}{2\xi}\sum_a \left(\partial_\mu A^{a\mu} + \xi\, g\, u^a\cdot\psi'\right)^2, \tag{8.42}$$

whose cross term cancels the mixing up to a total derivative. The Goldstone fields remain in the Lagrangian with masses squared ξM^2 (as do the Faddeev–Popov ghosts), and the gauge-boson propagator becomes

$$\frac{-i}{k^2 - m^2 + i\epsilon}\left[g^{\mu\nu} - (1-\xi)\,\frac{k^\mu k^\nu}{k^2 - \xi m^2}\right], \tag{8.43}$$

which falls off as $1/k^2$. Physical amplitudes do not depend on ξ, as a consequence of the BRST symmetry and the Zinn-Justin equation of Sec. 7.4. The limit $\xi\to\infty$ is the unitary gauge, while for $\xi = 1$ (Feynman–'t Hooft gauge) the massive propagator is the simplest.

The abelian Higgs model

Let us consider the simplest example, a complex scalar field coupled to a $U(1)$ gauge field,

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \left(D_\mu\phi\right)^* D^\mu\phi - \lambda\left(\phi^*\phi - \frac{v^2}{2}\right)^2, \qquad D_\mu = \partial_\mu - igA_\mu, \tag{8.44}$$

which is invariant under $\phi\to e^{i\alpha(x)}\phi$, $A_\mu\to A_\mu + \frac{1}{g}\partial_\mu\alpha(x)$. The potential is minimal for $\left|\langle\phi\rangle\right| = v/\sqrt{2}$. We choose $\alpha(x)$ such that the phase of ϕ is removed (unitary gauge),

$$\phi(x) = \frac{v + h(x)}{\sqrt{2}}, \qquad h(x)\ \text{real}. \tag{8.45}$$

Substituting, we obtain:

$$\begin{aligned}\mathcal{L} \;=\; & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2v^2A_\mu A^\mu + \frac{1}{2}\partial_\mu h\,\partial^\mu h - \frac{1}{2}\left(2\lambda v^2\right)h^2 \\ & + g^2 v\,h\,A_\mu A^\mu + \frac{1}{2}g^2h^2A_\mu A^\mu - \lambda v\,h^3 - \frac{\lambda}{4}h^4.\end{aligned} \tag{8.46}$$

The photon has acquired the mass $m_A = gv$ and the Higgs boson h has the mass $m_h = \sqrt{2\lambda}\,v$. The massless Goldstone boson, the phase of ϕ, is nowhere to be found! Before symmetry breaking we had $2+2=4$ degrees of freedom (a complex scalar and a massless photon); after it we have $1+3=4$ (the real field h and a massive vector boson).

Exercise: *Write $\phi = \left(\psi_1 + i\psi_2\right)/\sqrt{2}$, find the generator T of the $U(1)$ in the real basis $\left(\psi_1,\psi_2\right)$ and verify that Eq. (8.40) gives $m_A^2 = g^2v^2$.*

The electroweak gauge bosons

The most important application is the Standard Model, with gauge group $SU(2)\times U(1)_Y$ and a complex scalar doublet Φ of hypercharge $Y = 1/2$,

$$D_\mu\Phi = \left(\partial_\mu - igW^a_\mu\frac{\tau^a}{2} - ig'\frac{1}{2}B_\mu\right)\Phi, \qquad \langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}. \tag{8.47}$$

Substituting the vacuum expectation value into the kinetic term $\left(D_\mu\Phi\right)^\dagger D^\mu\Phi$, we find the mass terms

$$\frac{g^2v^2}{8}\left[W^1_\mu W^{1\mu}+W^2_\mu W^{2\mu}\right]+\frac{v^2}{8}\left(gW^3_\mu-g'B_\mu\right)\left(gW^{3\mu}-g'B^\mu\right). \tag{8.48}$$

The charged bosons $W^\pm_\mu=\left(W^1_\mu\mp iW^2_\mu\right)/\sqrt{2}$ and the neutral combinations

$$Z_\mu=\frac{gW^3_\mu-g'B_\mu}{\sqrt{g^2+g'^2}},\qquad A_\mu=\frac{g'W^3_\mu+gB_\mu}{\sqrt{g^2+g'^2}}, \tag{8.49}$$

have the masses

$$m_W=\frac{gv}{2},\qquad m_Z=\frac{\sqrt{g^2+g'^2}\,v}{2}=\frac{m_W}{\cos\theta_W},\qquad m_A=0, \tag{8.50}$$

with $\cos\theta_W=g/\sqrt{g^2+g'^2}$ and $e=g\sin\theta_W$. The photon remains massless because the electric charge $Q=\tau^3/2+Y$ annihilates $\langle\Phi\rangle$: $SU(2)\times U(1)_Y\to U(1)_{\rm em}$. Three generators are broken, three Goldstone bosons are eaten by $W^\pm$ and Z, and one of the four real fields in Φ remains as a physical particle, the Higgs boson. It was discovered at the LHC in 2012, with a mass of about 125 GeV. From the Fermi constant, $v=\left(\sqrt{2}G_F\right)^{-1/2}\simeq 246$ GeV, and the measured masses are $m_W\simeq 80.4$ GeV and $m_Z\simeq 91.2$ GeV.

Exercise: *Derive the above mass terms and verify that the matrix of Eq. (8.40), with $g^2\to g_a g_b$, has the eigenvalues m_W^2 (twice), m_Z^2 and 0.*

The relation $m_W=m_Z\cos\theta_W$ is a consequence of the special structure of the Higgs doublet. How well it survives quantum corrections, and what this teaches us about physics beyond the Standard Model, is one of the topics of Part IV, Chapter 9.

9

Renormalisation: counting the degree of ultraviolet divergences

Consider a Lagrangian with fields f and generic interaction operators O_i,

$$\mathcal{L} = \text{kinetic terms} + g_1 O_1 + g_2 O_2 + \ldots + g_N O_N. \tag{9.1}$$

Each operator O_i is a product of fields and/or their derivatives. In QED, for example, we have one such interaction term,

$$\bar{\psi} \not{A} \psi.$$

In QCD more operators emerge, for example

$$f^{abc} \partial_\mu A^a_\nu A^{\mu,b} A^{\nu,c}, \qquad f^{abe} f^{cde} A^a_\mu A^b_\nu A^{\mu,c} A^{\nu,d}, \ldots$$

We would like to keep this discussion as general as possible. At the end, we will be able to make statements on whether we can remove ultraviolet infinities from arbitrary Lagrangians via renormalisation. Most of our arguments will be derived using simple dimensional analysis.

We consider a generic one-particle irreducible Feynman diagram in perturbation theory. We will first find a simple formula to test whether it has the most obvious of all possible divergences, the so-called superficial ultraviolet divergence. If the diagram has L loops, a superficial divergence corresponds to an infinity of the diagram in the limit

$$|k_1| = |k_2| = \ldots = |k_L| = \kappa \to \infty,$$

where k_i are the loop momenta. A Feynman diagram might have divergences in other limits, where only some momenta or linear combinations of them are taken to infinity while the remaining independent momenta remain fixed. A Feynman diagram in the superficial ultraviolet limit behaves as

$$\int^\infty d\kappa \kappa^{D-1}, \tag{9.2}$$

where D is an integer, called the *superficial degree of divergence*.

- If $D > 0$ the Feynman diagram has a power-like divergence,
- if $D = 0$ it diverges logarithmically,
- if $D < 0$ it is convergent *(only in the superficial limit, since it might have other divergences)*.

We can compute D (or an upper bound of it) for any Feynman diagram on general grounds. We assume that our 1PI Feynman diagram has

- I_f internal propagators for each of the fields f,
- E_f external legs for each of the fields f and
- N_i vertices corresponding to the term $g_i O_i$ in the Lagrangian.

Recall, as examples, the Feynman rules for propagators in gauge theories and how they behave in the limit of infinite momentum:

- a photon propagator,
$$\sim \frac{-g_{\mu\nu} + \frac{k_\mu k_\nu}{k^2}}{k^2} \sim \kappa^{-2};$$
- a fermion propagator,
$$\sim \frac{\not{k} + m}{k^2 - m^2} \sim \kappa^{-1};$$
- for a scalar,
$$\sim \frac{1}{k^2 - m^2} \sim \kappa^{-2}.$$

For each of the internal propagators of the field f in the Feynman diagram there is a contribution to the superficial divergence,

$$\Delta_f \sim k^{-2+2s_f},$$

where $s_f = 0$ for a boson and $s_f = \frac{1}{2}$ for a fermion. The total contribution to the asymptotic limit from propagators is then

$$\kappa^{\sum_f 2I_f(s_f-1)}. \tag{9.3}$$

The contribution from vertices is easy to find if we know how many loop momenta appear in the corresponding Feynman rules. For a vertex due to an operator O_i this number is equal to the number of space-time derivatives d_i which can be found in the expression for O_i. Recall that a Feynman rule for a vertex is essentially the Fourier transform of the expression of its operator and therefore momenta arise only from derivatives. The total contribution from vertices to the superficial ultraviolet limit is

$$\kappa^{\sum_i N_i d_i}. \tag{9.4}$$

Finally, due to the integration measure d^4k_i for each loop, the total contribution from the loop momenta to the superficial UV limit is

$$\kappa^{4L}, \tag{9.5}$$

where L is the number of loops in the graph. L is known if we are given the number of internal propagators I_f and the number of vertices N_i in the graph. The number of momenta carried by internal propagators is $\sum_f I_f$. The vertices provide $\sum_i N_i$ constraints, of which one is not for loop momenta but for the external momenta. Therefore, the number of loop momenta is

$$L = \sum_f I_f - \sum_i N_i + 1.$$

Putting together the contributions from the loop integration measures, vertices, and internal propagators, we find that the asymptotic behaviour at infinity has a superficial degree of divergence

$$D = \sum_f 2I_f(1+s_f) - \sum_i N_i(4-d_i) + 4. \tag{9.6}$$

We can express the number of internal propagators in terms of the number of external legs. Let us assume that we have N_{if} particles f in the vertex corresponding to the operator O_i in the Lagrangian. The total number of legs of the particle f attached to all the vertices of the graph is

$$\sum_i N_i N_{if}.$$

Of these, E_f are external and the remaining ones are internal. Every propagator of f has two ends, so the number of internal legs is

$$2I_f.$$

We then have the identity

$$2I_f + E_f = \sum_i N_i N_{if}. \tag{9.7}$$

We can therefore write the degree of divergence as

$$D = 4 - \sum_f E_f(1+s_f) - \sum_i N_i \left[4 - d_i - \sum_f N_{if}(1+s_f)\right]. \tag{9.8}$$

Notice that the square bracket in the last expression depends only on the functional form of the operator O_i. If this operator is multiplied by a coupling constant g_i in the Lagrangian, i.e. $\mathcal{L} = g_i O_i + \ldots$, we can prove that this square bracket is exactly the mass dimension of the coupling g_i:

$$[g_i] = 4 - d_i - \sum_f N_{if}(1+s_f). \tag{9.9}$$

Indeed, the mass dimension of each term in the action must be zero. We then have

$$\left[d^4x\right] + [g_i] + [O_i] = 0,$$

where the operator O_i has d_i derivatives and N_{if} fields f. Thus,

$$-4 + [g_i] + d_i - \sum_f N_{if}[f] = 0,$$

and $[f] = 1 + s_f$ is the mass dimension of the field. Indeed,

$$\begin{aligned} &\langle 0|\, T f(x_1) f(x_2)\, |0\rangle \sim \int_{k\to\infty} d^4k k^{-2+2s_f} e^{-ikx} \\ \rightsquigarrow\quad & 2[f] = 4 - 2 + 2s_f \rightsquigarrow [f] = 1 + s_f. \end{aligned} \tag{9.10}$$

In conclusion, we can write a very suggestive expression for the superficial degree of divergence,

$$D = 4 - \sum_f E_f(1+s_f) - \sum_i N_i[g_i]. \tag{9.11}$$

If the Lagrangian does not contain any couplings with negative mass dimension, $[g_i] \geq 0$, we find a superficial ultraviolet divergence, $D \geq 0$, only in Feynman diagrams with a small number of external legs,

$$D \geq 0 \rightsquigarrow \sum_f E_f(1 + s_f) \leq 4.$$

In particular, superficial divergences do not appear in *(one-particle-irreducible)* Feynman diagrams with five external legs or more.

Examples of theories where superficial divergences may appear in only a limited number of Green's functions are QED and QCD. All interaction operators have dimension four and their coefficients are dimensionless. Superficial ultraviolet divergences are limited to 1PI Green's functions such as $\langle 0|\, T\bar{\psi}(x_1)\psi(x_2)\,|0\rangle$, $\langle 0|\, TA^a_\mu(x_1)A^b_\nu(x_2)\,|0\rangle$, $\langle 0|\, T\bar{\psi}(x_1)A^b_\nu(x_2)\psi(x_3)\,|0\rangle \ldots$. On the contrary, $\langle 0|\, T\bar{\psi}(x_1)\psi(x_2)\bar{\psi}(x_3)\psi(x_4)\,|0\rangle_{1PI}$ is (superficially) finite.

Theories with $[g_i] \geq 0$ are called **renormalisable**. As we shall see, these superficial divergences in a finite number of Green's functions can be removed by adding a finite number of extra terms in the original Lagrangian (counterterms).

If the Lagrangian contains a coupling with negative mass dimension, $[g_j] < 0$, then from Eq. (9.11) we see that **all** Green's functions, at some loop order, will develop a superficial divergence. It is therefore impossible to cancel the infinities by adding a finite number of counterterms. Such theories are called non-renormalisable. *Remark.* Two qualifications. (i) The counting assumes that every boson propagator falls like κ^{-2}; a massive vector in unitary gauge, $(-g_{\mu\nu} + k_\mu k_\nu/M^2)/(k^2 - M^2) \sim \kappa^0$, violates this, so for spontaneously broken gauge theories one must count in a renormalisable (R_ξ) gauge. (ii) "Non-renormalisable" does not mean "useless": with infinitely many operators allowed and an expansion in E/Λ, a theory with $[g_j] < 0$ is renormalised order by order in the EFT expansion with a finite number of counterterms at each order (Weinberg, QFT I, sec. 12.3; Weinberg, Physica A 96 (1979) 327).

9.1 Subdivergences

We should stress that the counting of the superficial degree of divergence is not sufficient to prove that a Feynman diagram is finite. Consider the two example graphs of Fig. 9.1. Both two-loop graphs have the same superficial degree of divergence $D = -2$. One could

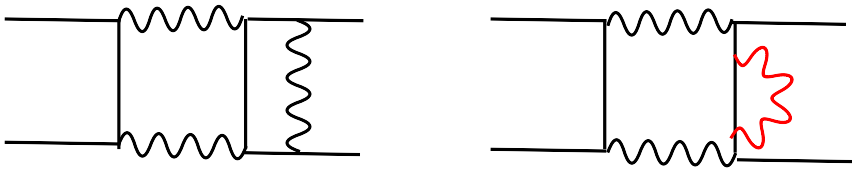


Figure 9.1. Both two-loop graphs have the same superficial degree of divergence $D = -2$. However, the Feynman diagram on the right has a self-energy one-particle-irreducible subgraph which has a superficial degree of divergence $D = 1$. A sufficient condition for a graph to be UV finite is that the graph and all its subgraphs have $D < 0$.

naively conclude that both two-loop graphs are likely to be finite. We know, though, that this is not the case. Let us compute the superficial degree of divergence for all one-loop subgraphs that we can spot in the two diagrams. For the left diagram, we find that all subgraphs have a negative superficial degree of divergence. It also turns out with an explicit calculation (beyond the scope of these lectures) that the diagram is indeed UV

finite. However, the two-loop Feynman diagram on the right has a one-loop self-energy subgraph; this has a superficial degree of divergence $D = 1$. The self-energy is $1/(d-4)$ divergent, where d is the space-time dimensionality. Such a divergence remains even after we embed the one-loop self-energy as a subgraph inside a two-loop graph (there is no mechanism to cancel it). Against our naive counting for the global superficial degree of divergence, the two-loop diagram on the right is divergent. The lesson from the above examples is that for a diagram to be UV finite it is sufficient that the superficial degrees of divergence for the full graph and all of its subgraphs are negative. *Remark.* This is Weinberg's theorem (S. Weinberg, Phys. Rev. 118 (1960) 838): $D < 0$ for the graph and for all its 1PI subgraphs guarantees convergence. The converse is not a theorem: $D \geq 0$ for a (sub)graph usually signals a divergence, but symmetries can make it finite, e.g. the QED light-by-light 1PI four-photon function has $D = 0$ yet is finite by gauge invariance. In the right-hand diagram the divergence is real (the fermion self-energy, $D = 1$, is log divergent after the chiral-symmetry suppression).

9.2 Cancellation of superficial divergences with counterterms

We have derived a criterion to decide whether a Green's function will develop the most "obvious" type of divergence (superficial) in the limit where the magnitudes of all loop momenta tend simultaneously to infinity. We also found that for renormalisable theories this type of divergence appears in only a finite number of Green's functions with a small number of external legs.

Infinities are not acceptable for physical theories. A way out of this problem is to recognise that the Lagrangian that we started with has a certain degree of arbitrariness. The guiding principle for constructing a Lagrangian is to respect a set of symmetries (e.g. BRST symmetry). However, this is not a tight enough constraint to fix, for example, the actual values of independent mass and coupling parameters. It may be possible to redefine the parameters and fields of the Lagrangian or even add more operators to it without destroying the symmetries of the Lagrangian. How can we fix the fields and parameters of the Lagrangian, choosing among their various possible redefinitions? In renormalisable theories, we (partially) fix this arbitrariness so that all Green's functions calculated with the redefined ("renormalised") fields, couplings and masses are finite.

We have seen that for "renormalisable theories" the "disease" of infinities spreads only to a few Green's functions. Redefining the fields and parameters of the Lagrangian ($\psi = Z\psi_R = \psi_R + \delta Z \psi_R, \ldots$) gives rise to a few new terms (counterterms) with coefficients engineered to cancel exactly the UV infinities which emerge order by order in perturbation theory. But, is it possible mathematically that we can cancel the infinities from loop diagrams with counterterms? For this method to work, it is essential that diagrams with counterterms at a loop order have the same kinematic dependence as the UV infinities of loop diagrams without counterterms at higher orders. At the first two orders in perturbation theory, this statement means that tree diagrams with counterterms must have the same kinematic dependence as the infinities of one-loop diagrams without counterterms.

If, for example, the $1/\epsilon$ terms of a Green's function at one-loop order (where $d = 4 - 2\epsilon$ in dimensional regularisation) are logarithms of external momenta,

$$\frac{\ln(p^2)}{\epsilon},$$

such a contribution cannot be cancelled by the tree-level contribution of a counterterm. Recalling the Feynman rules for vertices which enter tree-level calculations in all theories

that we have examined so far, we find no such logarithms in their expressions. Feynman rules always yield simple polynomial expressions for the vertices of tree diagrams. For such tree-level expressions made out of counterterms to cancel the infinities of one-loop diagrams the latter have also to be constants or simple polynomials of momenta. A necessary condition for the counterterm programme to be successful is that one-loop infinities are "local", i.e. they appear as simple polynomials in the external momenta as the usual Feynman rules do.

Let us look at the functional form of the superficial infinities at one-loop order in perturbation theory, and convince ourselves that indeed this is exactly what happens in practice. Take, as an example, the one-loop correction to the four-point function in the $-\frac{\lambda}{4!}\phi^4$ theory. The superficial degree of divergence is $D = 4 - 4 \times 1 = 0$ and the graph is indeed divergent. In dimensional regularisation, the corresponding Feynman-parameter integral yields

$$\begin{aligned}\langle 0|\, T\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\,|0\rangle\big|_{\text{1-loop}} &\sim \lambda^2 \int d^dk \frac{1}{(k^2-m^2)\left[(k+p)^2-m^2\right]} \\ &\sim \lambda^2\Gamma(\epsilon)\int_0^1 dx_1 dx_2 \frac{\delta(1-x_1-x_2)}{(m^2-x_1x_2s)^\epsilon} \\ &\sim \frac{\lambda^2}{\epsilon} - \lambda^2\gamma - \lambda^2\int_0^1 dx \ln\left(m^2-x(1-x)s\right), \end{aligned} \tag{9.12}$$

with $s = (p_1+p_2)^2 = (p_3+p_4)^2$.

Loop integrals, in general, contain logarithms or integrals of logarithms (polylogarithms) whose arguments are kinematic invariants formed from external momenta. Our example result is not an exception and we indeed find logarithmic contributions in the finite part. We cannot fail to observe, however, that the divergent part is very simple: it is just a constant. We can then modify the interaction terms in the Lagrangian,

$$\frac{-\lambda}{4!}\phi^4 \to \frac{-\lambda}{4!}\phi^4 + \frac{\sharp\lambda^2}{\epsilon}\phi^4,$$

and adjust the coefficient $\sharp$ so that it cancels exactly the divergent part of this one-loop integral.

The divergent parts of one-loop integrals are simple polynomials of the external momenta, as in the above example. If we use Feynman parameters, any one-loop integral may be written as

$$I_{\text{1-loop}} \sim \Gamma\left(N-\frac{d}{2}\right)\int_0^1 \frac{dx_1\ldots dx_n\,\delta(1-x_1-\ldots x_n)}{\left(m_1^2x_1+\ldots+m_n^2x_n-\sum s_{i\ldots j}x_ix_j-i\delta\right)^{N-\frac{d}{2}}}, \tag{9.13}$$

where N is an integer (equal to the number of propagators) and $d = 4-2\epsilon$ the dimension. The denominator contains a sum over masses and kinematic invariants of the external momenta. Divergences may arise from two sources: the Gamma function $\Gamma(N-2+\epsilon)$ and the denominator of the integrand. The argument of the Gamma function $N - d/2 = -D/2$ is proportional to the superficial degree of divergence D of the integral. The denominator of the Feynman integral does not have any ultraviolet divergences. It could become divergent when masses or invariants become zero, but it is finite when all the propagators are massive. If this is not the case, and there are massless particles propagating in a loop, singularities from the denominator are of infrared nature, connected to the small or zero values of $|k|$ rather than the UV $|k| \to \infty$ limit. Infrared singularities can also be regulated

by attributing a small mass to massless particles and/or considering them to be slightly off-shell. We shall not worry here about infrared one-loop singularities and focus on the ultraviolet divergences which can be found, at one-loop order, in the Gamma-function prefactor of the Feynman-parameter integral,

$$\Gamma(N - d/2) = \Gamma(-1 + \epsilon),\ \Gamma(\epsilon)$$

for $N = 1, 2$. Using the identity

$$\Gamma(x) = \frac{\Gamma(1 + x)}{x},$$

and

$$\Gamma(1 + \epsilon) = 1 - \gamma\epsilon + O(\epsilon^2),$$

and the fact that for $N = 1, 2$ (responsible for the UV divergences) the denominator of Eq. (9.13) turns into a numerator ($N - d/2 \leq 0$ in four dimensions), we can see that the coefficient of the $1/\epsilon$ pole can only be a polynomial in the external momenta. A loop diagram with superficial degree of divergence D,

$$\int_{|k|\to\infty} |k|^{D-1}, \tag{9.14}$$

has mass dimension D. Therefore, the polynomial can only be of degree D in the external momenta. Each term in this polynomial multiplying $1/\epsilon$ must be cancelled by a separate counterterm operator with a different number of derivatives. Naturally, the number of derivatives must satisfy

$$d_i \leq D. \tag{9.15}$$

Exercise: *Prove that for the cancellation of UV divergences we need at most as many counterterms in the Lagrangian as there are divergent Green's functions.*

9.3 Nested and overlapping divergences

It can be proven that we only need to worry about removing superficial divergences from loop integrals. Nested and overlapping singularities are "automatically" removed with this procedure as well. We refer to the original literature for this topic:

> K. Hepp, "Proof of the Bogolyubov-Parasiuk theorem on renormalization," Commun. Math. Phys. **2** (1966) 301.

Remark. More precisely (BPHZ), subdivergences are removed by the counterterms already fixed at lower orders, inserted into the diagram (Bogoliubov's R-operation); after that, only the overall (superficial) divergence of each 1PI graph remains and is local. Besides Hepp, see N. N. Bogoliubov and O. S. Parasiuk, Acta Math. 97 (1957) 227; W. Zimmermann, Commun. Math. Phys. 15 (1969) 208 (forest formula).

10

Proof of renormalisability for non-abelian gauge theories

Consider a theory with action $S[\phi]$ which is invariant under BRST transformations of the fields ϕ_i: $\delta_\theta \phi_i = \theta B_i$. We can add to the classical action source terms which, owing to nilpotency, preserve the invariance under BRST transformations,

$$S[\phi_i, K_i] = S[\phi_i] + \int d^4x B_i K_i. \tag{10.1}$$

We now split the action into two terms,

$$S[\phi_i, K_i] = S_R[\phi_i, K_i] + S_\infty[\phi_i, K_i]. \tag{10.2}$$

The first term is the action with the fields, masses and coupling constants set to their renormalised values. The second term contains the counterterms. S and S_R have the same functional form. Therefore, they possess the same set of symmetries. It follows that S_∞ must also possess the same set of symmetries.

The effective action can be cast as an expansion in loops,

$$\Gamma[\phi_i, K_i] = \sum_{L=0}^{\infty} \Gamma_L[\phi_i, K_i]. \tag{10.3}$$

We recall that all terms in the expansion are separately symmetric and that we can perform independent shifts in the measure of the path integral for each one of them.

The Slavnov–Taylor identities for the BRST symmetry transformations result in the Zinn-Justin equation, Eq. (7.34), which is written in short notation as

$$(\Gamma, \Gamma) = 0. \tag{10.4}$$

Inserting the loop expansion of the effective action, we obtain

$$\begin{aligned} 0 \;=\; & (\Gamma_0, \Gamma_0) \\ & +(\Gamma_0, \Gamma_1) + (\Gamma_1, \Gamma_0) \\ & +(\Gamma_0, \Gamma_2) + (\Gamma_1, \Gamma_1) + (\Gamma_2, \Gamma_0) \\ & + \dots . \end{aligned} \tag{10.5}$$

Every line in the above expression must be separately zero, since it corresponds to a different order in the loop expansion (equivalently, the $\hbar$ expansion). For the N-th term of the expansion we have

$$\sum_{L=0}^{N} (\Gamma_L, \Gamma_{N-L}) = 0. \tag{10.6}$$

At each loop order we find UV infinities. We decompose the L-loop effective action into a finite and a divergent part:

$$\Gamma_L = \Gamma_{L,\text{fin}} + \Gamma_{L,\infty}. \tag{10.7}$$

At zeroth order we only find tree graphs and there are no infinities. In addition, the tree-level effective action is equal to the classical action. We therefore have

$$\Gamma_{0,\text{fin}} = S_R, \quad \Gamma_{0,\infty} = 0. \tag{10.8}$$

We will prove by induction that we can remove all infinities from the effective action, rendering all $\Gamma_{L,\infty} = 0$, with the counterterms in S_∞. Let us assume that we have achieved this for all loops up to $N-1$,

$$\Gamma_{L,\infty} = 0, \quad L = 1 \dots N-1. \tag{10.9}$$

Then, taking the infinite part of Eq. (10.6), we obtain

$$\big(\Gamma_{0,\text{fin}}, \Gamma_{N,\infty}\big) + \big(\Gamma_{N,\infty}, \Gamma_{0,\text{fin}}\big) = 0. \tag{10.10}$$

or, equivalently,

$$(S_R, \Gamma_{N,\infty}) = 0. \tag{10.11}$$

As we have discussed in the previous chapter, we expect the infinities of momentum-space Green's functions in $\Gamma_{N,\infty}$ to have a simple polynomial dependence on the momenta, given that all divergences at the previous loop orders are cancelled. We now make two observations:

- As we have shown earlier, the infinities of $\Gamma_{N,\infty}$ arise in Green's functions with a small number of external legs. As we have assumed that the infinities of all previous loop orders have been cancelled, at the N-th loop order we cannot have any subdivergences. Thus, the N-th loop order divergences correspond to the superficial limit where all loop momenta are taken to infinity. For the superficial divergences we have derived that they should originate from local field operators (products of fields and their derivatives as well as sources K_i) in $\Gamma_{N,\infty}$ whose dimension is less than or equal to four.

- $\Gamma_{N,\infty}$ has all the linear symmetries of S_R. These are:

  - Lorentz transformations
  - Global gauge transformations
  - Antighost translations
  - Ghost phase transformations ($\rightsquigarrow$ ghost number conservation)

  The last two symmetries are apparent by inspecting the ghost terms of the Lagrangian,

Conventions. This chapter uses the conventions of the BRST chapter: $D_\mu = \partial_\mu - igA^a_\mu T^a$, $D^{ab}_\mu = \partial_\mu \delta^{ab} - g f^{abc} A^c_\mu$, and, in Lorenz gauge, $\delta_\theta \psi = -i\theta \eta^a T^a \psi$, $\delta_\theta A^a_\mu = -\frac{\theta}{g} D^{ab}_\mu \eta^b$, $\delta_\theta \eta^a = \frac{\theta}{2} f^{abc} \eta^b \eta^c$, $\delta_\theta \bar\eta^a = -\frac{\theta}{g} w^a$, $\delta_\theta w^a = 0$, with $\mathcal{L}_{\text{gf+gh}} = w^a \partial_\mu A^{a\mu} + \frac{\xi}{2} w^a w^a + (\partial^\mu \bar\eta^a) D^{ab}_\mu \eta^b = -g\, s\big[\bar\eta^a (\partial_\mu A^{a\mu} + \frac{\xi}{2} w^a)\big]$. The proof follows Weinberg, QFT II, Sec. 17.2, whose conventions absorb g into the generators and differ in signs; the dictionary is $t^a_{\text{W}} = gT^a$, $f^{abc}_{\text{W}} = g f^{abc}$, $\eta^a_{\text{W}} = -\eta^a/g$, $\bar\eta^a_{\text{W}} = g\bar\eta^a$,

$\omega^a_{\rm W} = w^a$, which maps his $\delta_\theta\psi = i\theta\eta^a_{\rm W}t^a_{\rm W}\psi$, $\delta_\theta A^a_\mu = \theta(\partial_\mu\eta^a_{\rm W} + f^{abc}_{\rm W}A^b_\mu\eta^c_{\rm W})$, $\delta_\theta\eta^a_{\rm W} = -\frac{\theta}{2}f^{abc}_{\rm W}\eta^b_{\rm W}\eta^c_{\rm W}$, $\delta_\theta\bar\eta^a_{\rm W} = -\theta\omega^a_{\rm W}$ and ghost term $-(\partial^\mu\bar\eta^a_{\rm W})D^{ab}_{{\rm W}\mu}\eta^b_{\rm W}$ onto ours.

$$\mathcal{L}_{\rm FP} = \left(\partial^\mu\bar\eta^a\right)D^{ab}_\mu\eta^b \qquad \text{(Lorenz gauge)}. \tag{10.12}$$

The antighost field enters the Lagrangian only through its derivative, and thus the Lagrangian is invariant if we shift the field globally by a constant. In addition, the Lagrangian is invariant under a phase transformation of the ghost and antighost fields:

$$\eta^a(x) \to e^{i(+1)\rho}\eta^a(x), \quad \bar\eta^a(x) \to e^{i(-1)\rho}\bar\eta^a(x),$$
$$A^{a\mu} \to e^{i(+0)\rho}A^{a\mu}, \quad \psi \to e^{i(+0)\rho}\psi.$$

The conserved charge of this symmetry is called the ghost number. The ghost numbers of the fields $\phi_i = \left\{A^{a\mu}, \psi, \eta^a, \bar\eta^a\right\}$ are $\gamma_i = \{0, 0, +1, -1\}$ respectively. The above phase transformations leave the action $S[\phi_i]$ invariant. For the extended action $S[\phi_i, K_i]$ to be invariant, we need to assign ghost numbers to the sources K_i as well. From the BRST transformations we see that if a field ϕ_i has a ghost number γ_i, the variation under the transformation B_i of the field has a ghost number $\gamma_i + 1$. The term $\int d^4x B_iK_i$ ought to remain invariant under the ghost phase transformation. We must therefore assign ghost numbers $-\gamma_i - 1$ for the sources K_i. Specifically, the ghost numbers for $K_A, K_\psi, K_{\bar\eta}, K_\eta$ are $-1, -1, 0, -2$ respectively.

Lemma: $\Gamma_{N,\infty}$ is linear in the sources K_i.
Proof: To prove this we shall use dimensional analysis and symmetries. First we determine the mass dimension of the sources K_i. For a field ϕ_i with dimension d_i the operators B_i have dimension $d_i + 1$, as can be seen from the expressions of the BRST transformations. The term $\int d^4x B_iK_i$ ought to be dimensionless. Therefore we conclude that the sources K_i have dimension $3 - d_i$. Hence the dimension of the BRST sources corresponding to scalar and vector fields, $K_A, K_\eta, K_{\bar\eta}$, is 2, while the dimension of the BRST source corresponding to fermion fields, K_ψ, is 3/2.

Since the operators of $\Gamma_{N,\infty}$ have dimension at most four, we can have operators with at most two sources K_i:

- $K_{\rm scalar/vector}K_{\rm scalar/vector}$,
- $K_{\rm fermion}K_{\rm fermion}$,
- $K_{\rm fermion}K_{\rm fermion}$field, with [field] ≤ 1.

All quadratic terms in the sources K_i have a non-zero ghost number and they are therefore excluded, with the exception of

$$K_{\bar\eta^a}K_{\bar\eta^a}$$

operator, which has zero ghost number. We can exclude this operator for a different reason. The BRST symmetry transformation of the classical action for an antighost is linear and not quadratic,

$$\delta_\theta\bar\eta^a = -\frac{\theta}{g}w^a. \tag{10.13}$$

Therefore,

$$\frac{\delta_L\Gamma[\left\langle\phi_i\right\rangle, K_i]}{\delta K_{\bar\eta^a}} = -\frac{1}{g}\left\langle w^a\right\rangle \rightsquigarrow \frac{\delta_L\Gamma[\phi_i, K_i]}{\delta K_{\bar\eta^a}} = -\frac{1}{g}w^a \;\rightsquigarrow\; \frac{\delta_L\Gamma_{N\geq 1}}{\delta K_{\bar\eta^a}} = 0, \tag{10.14}$$

where in the last step we used that the transformation is linear so that, in that case, the transformation of the "average" is equal to the "average" of the transformation. The above differential equation tells us that Γ is at most linear in the source $K_{\bar{\eta}^a}$. We have just shown that $\Gamma_{N,\infty}$ is at most linear in all sources K_i. We write

$$\Gamma_{N,\infty}[\phi_i, K_i] = \Gamma_{N,\infty}[\phi_i, 0] + \int d^4x \tilde{B}_i K_i. \tag{10.15}$$

Recall that the classical action is also linear:

$$S_R[\phi_i, K_i] = S_R[\phi_i] + \int d^4x B_i K_i. \tag{10.16}$$

Substituting in $(S_R, \Gamma_{N,\infty}) = 0$ we obtain two equations for the zeroth- and first-order terms in K_i, namely

$$\int d^4x \left[B_i \frac{\delta_L \Gamma_{N,\infty}}{\delta \phi_i} + \tilde{B}_i \frac{\delta_L S_R}{\delta \phi_i} \right] = 0, \tag{10.17}$$

$$\int d^4x \left[B_i \frac{\delta_L \tilde{B}_j}{\delta \phi_i} + \tilde{B}_i \frac{\delta_L B_j}{\delta \phi_i} \right] = 0. \tag{10.18}$$

We now define

$$\Gamma^{(\epsilon)} \equiv S_R + \epsilon \Gamma_{N,\infty}, \tag{10.19}$$

$$B_i^\epsilon \equiv B_i + \epsilon \tilde{B}_i, \tag{10.20}$$

where ϵ is a very small parameter. With Eqs. (10.17) and (10.18), we can prove that under the field transformation

$$\phi_i \to \phi_i + \theta B_i^\epsilon, \tag{10.21}$$

- $\Gamma^{(\epsilon)}$ is invariant;
- the transformation is nilpotent (up to $O(\epsilon)$).

We leave the proof of the above statements to the reader as an **exercise**.

From the above equations we infer that the dimensionality of $\tilde{B}_i$ is at most the dimensionality of B_i. From Eq. (10.18), we also infer that B_i, $\tilde{B}_i$ and thus B_i^ϵ all have the same ghost number (**exercise**). With these constraints, the allowed form of the transformations of Eq. (10.21) is

$$\psi \to \psi - i \left(\theta \eta^a\right) \mathcal{T}^a \psi, \tag{10.22}$$

$$A^{\mu,a} \to A^{\mu a} - \theta \left[\frac{1}{g} B^{ab} \partial^\mu \eta^b + D^{abc} A^{\mu,b} \eta^c \right], \tag{10.23}$$

$$\eta^a \to \eta^a + \frac{1}{2} \theta E^{abc} \eta^b \eta^c, \tag{10.24}$$

with $E^{abc} = -E^{acb}$ due to the ghost field η^b being a Grassmann variable.

We can place more constraints on the coefficients of Eqs. (10.22) by exploiting the nilpotency of the transformations:

- From $\delta_{\theta_1}\delta_{\theta_2}\eta^a = 0$, we find that

$$E^{abc}E^{bde} + E^{abe}E^{bcd} + E^{abd}E^{bec} = 0, \tag{10.25}$$

which reveals that E^{abc} must be a structure constant of some Lie algebra. It would not be a surprise if this Lie algebra is the same as that of the non-abelian gauge group of the classical action S_R. Indeed, if we set the small parameter ϵ to zero, then $\Gamma^{(\epsilon)}\big|_{\epsilon=0} = S_R$. The structure constants E^{abc} must therefore be proportional to the structure constants of the non-abelian gauge group of the classical action:

$$E^{abc} = \lambda f^{abc}. \tag{10.26}$$

- The nilpotency of the transformation of the gauge field $A^{\mu,a}$ yields two constraints, namely

$$D^{abc}D^{bde} - D^{abe}D^{bdc} = E^{bec}D^{adb} = \lambda f^{bec}D^{adb}, \tag{10.27}$$

$$B^{ab}E^{bcd} = D^{abd}B^{bc}. \tag{10.28}$$

Eq. (10.27) tells us that the matrices $\tilde{t}^a_{bc} = -iD^{bca}$ satisfy the commutation relation of generators in some representation of the non-abelian gauge group:

$$\left[\tilde{t}^c, \tilde{t}^e\right] = i\lambda f^{ceb}\tilde{t}^b.$$

The only representation of the Lie group with the dimensionality of D^{abc} is the adjoint representation. Therefore, the solution of Eq. (10.27) is

$$D^{abc} = \lambda f^{abc}. \tag{10.29}$$

Remark. The statement "the only representation with the dimensionality of D^{abc} is the adjoint" is not true as such (e.g. for $SU(2)$ a 3-dimensional representation may also be $\mathbf{2}\oplus\mathbf{1}$ or three singlets), and the $\epsilon \to 0$ argument for $E^{abc} \propto f^{abc}$ is heuristic. The clean argument (Weinberg, QFT II, sec. 17.2) uses the linear global gauge symmetry: $\Gamma_{N,\infty}$ and hence $\tilde{B}_i$ are covariant, so E^{abc}, D^{abc}, B^{ab} must be invariant tensors; for a simple group the only antisymmetric one is f^{abc}, $D^{abc} = \alpha f^{abc} + \beta d^{abc}$ with $\beta = 0$ forced by Eq. (10.27), and $B^{ab} \propto \delta^{ab}$ by Schur's lemma. For a product group (e.g. $SU(3)\times SU(2)\times U(1)$) there is one λ and one N per factor. Eq. (10.28) reveals that the matrix B_{ab} commutes with the structure constants which can be chosen to be totally antisymmetric. The only possible solution is therefore a diagonal matrix (**exercise**). We can verify this easily in the special case of an $SU(2)$ group where the structure constants are the totally antisymmetric Levi-Civita symbol. Eq. (10.28) takes the form

$$B^{ab}\epsilon_{bcd} = B^{bc}\epsilon_{abd},$$

which, as examples, for $(a,c,d) = (1,2,3)$ yields $B^{11} = B^{22}$ and for $(a,c,d) = (1,2,2)$ yields $B^{32} = 0$. Similarly, one finds all diagonal terms to be equal and the non-diagonal terms to vanish. We write

$$B^{ab} = N\lambda\delta_{ab}. \tag{10.30}$$

- Nilpotency of the fermion field transformation yields for the matrices $\mathcal{T}^a$ that they also satisfy the Lie algebra of the non-abelian group of the classical action,

$$\left[\mathcal{T}^b, \mathcal{T}^c\right] = iE^{abc}\mathcal{T}^a = i\lambda f^{abc}\mathcal{T}^a. \tag{10.31}$$

Therefore, as suggested from the $\epsilon = 0$ limit, we have

$$\mathcal{T}^a = \lambda T^a, \tag{10.32}$$

where T^a are the generators of the representation for the fermions in S_R.

We have therefore found that $\Gamma^{(\epsilon)}$ is symmetric under the same BRST symmetry transformation as S_R, up to some rescalings. Explicitly, the BRST symmetry transformations of $\Gamma^{(\epsilon)}$ take the form

$$\begin{aligned}
\psi &\to \psi - i\left(\lambda\theta\eta^a\right)T^a\psi, && (10.33)\\
A^{\mu,a} &\to A^{\mu a} - \lambda\theta\left[\frac{N}{g}\partial^\mu\eta^a + f^{abc}A^{\mu,b}\eta^c\right], && (10.34)\\
\eta^a &\to \eta^a + \frac{1}{2}\lambda\theta f^{abc}\eta^b\eta^c, && (10.35)\\
\bar\eta^a &\to \bar\eta^a - \frac{\theta}{g}w^a, && (10.36)\\
w^a &\to w^a. && (10.37)
\end{aligned}$$

The last two transformations are linear symmetry transformations of the classical action and they are automatically symmetry transformations of the effective action and of $\Gamma^{(\epsilon)}$ as well.

Recall that we expect $\Gamma^{(\epsilon)}$ to be made out of local operators. We write

$$\Gamma^{(\epsilon)} = \int d^4x\mathcal{L}^{(\epsilon)}. \tag{10.38}$$

The dimension of the operators is bounded by the power-counting arguments of the previous chapter. In addition, $\mathcal{L}^{(\epsilon)}$ should consist of a combination of operators that respect the BRST symmetry which we have just discovered, Eqs. (10.33). Finally, $\mathcal{L}^{(\epsilon)}$ is constrained further to respect all the linear symmetries of the classical Lagrangian,

$$\mathcal{L} = \mathcal{L}_{\text{fermion}} - \frac{1}{4}G^{a\mu\nu}G^a_{\mu\nu} + \left(\partial_\mu\bar\eta^a\right)\left(\partial^\mu\eta^a\right) + gf^{abc}\left(\partial_\mu\bar\eta^a\right)A^{b,\mu}\eta^c + w^a\partial_\mu A^{a,\mu} + \frac{\xi}{2}w^aw^a.$$

These linear symmetries are:

- Lorentz invariance.
- Global gauge invariance. Explicitly, the global symmetry transformations are

$$\begin{aligned}
\delta\psi &= -i\epsilon^aT^a\psi,\ \delta A^a_\mu = f^{abc}\epsilon^bA^c_\mu,\\
\delta\eta^a &= f^{abc}\epsilon^b\eta^c,\ \delta\bar\eta^a = f^{abc}\epsilon^b\bar\eta^c,\ \delta w^a = f^{abc}\epsilon^bw^c.
\end{aligned}$$

- Antighost translation invariance: $\bar\eta^a \to \bar\eta^a + c$.
- Ghost-number conservation.

Can we write a Lagrangian density $\mathcal{L}^{(\epsilon)}$ with operators in addition to those of the classical $\mathcal{L}$, still satisfying the list of constraints that we have found above? If such operators exist, then we can establish some simple rules for them. To preserve ghost number, the ghost and antighost fields must appear in pairs or not appear at all in such novel operators. Because of antighost translation invariance, the antighost must always be differentiated. We therefore conclude that the antighost should appear in the form

$$\left(\partial_\mu\bar\eta^a\right). \tag{10.39}$$

Let us recall the dimensions of the fields,

$$\left[A^{a,\mu}\right] = \left[\eta^a\right] = \left[\bar\eta^a\right] = 1, \quad \left[w^a\right] = 2.$$

The combination of fields in Eq. (10.39), together with the ghost η^b that must accompany it, has dimension three. Operators must have dimension at most four; thus they can include at most one such combination of ghost fields. Altogether, we can have the following operators:

- ghost operators,
$$\left(\partial_\mu \bar{\eta}^a\right)\left(\partial^\mu \eta^b\right),\ \left(\partial_\mu \bar{\eta}^a\right) A^{b,\mu}\eta^c, \tag{10.40}$$
- auxiliary-field operators,
$$w^a\left(\partial_\mu A^{b,\mu}\right),\ w^a A^c_\mu A^{b,\mu}, \tag{10.41}$$
- and operators which contain only fermion and gauge-boson fields. We denote their sum by $\mathcal{L}_{\psi A}$.

Therefore, the most general Lagrangian density $\mathcal{L}^{(\epsilon)}$ is

$$\begin{aligned}\mathcal{L}^{(\epsilon)} &= \mathcal{L}_{\psi A} + \frac{\xi'}{2} w^a w^a + C w^a \left(\partial_\mu A^{a,\mu}\right) \\ &\quad - e_{abc} w^a A^b_\mu A^{c,\mu} + Z_\eta \left(\partial_\mu \bar{\eta}^a\right)\left(\partial^\mu \eta^a\right) + d_{abc}\left(\partial_\mu \bar{\eta}^a\right) A^{b,\mu}\eta^c,\end{aligned} \tag{10.42}$$

where $\xi', C, Z_\eta, d_{abc}, e_{abc}$ are unknown constants, with e_{abc} symmetric in the last two indices, $e_{abc} = e_{acb}$; they are, however, constrained by global gauge invariance.

We now use that $\mathcal{L}^{(\epsilon)}$ is invariant under the BRST transformations of Eqs. (10.33). We recall that for fermion and gauge-boson fields, the BRST transformation has the same functional form as a classical local gauge transformation with a local gauge parameter made out of a Grassmann constant and the ghost field. Thus, the $\mathcal{L}_{\psi A}$ part of the Lagrangian has to be not only globally gauge invariant but also locally gauge invariant, with gauge parameter

$$\epsilon^a \to \lambda\theta\eta^a,$$

and with a gauge coupling $g \to g/N$ in the covariant derivative and in the field strength. BRST invariance of the ghost and auxiliary field part of $\mathcal{L}^{(\epsilon)}$ determines the constants. Specifically, we find (**exercise**)

$$C = \frac{Z_\eta}{\lambda N}, \tag{10.43}$$

$$d_{abc} = \frac{g Z_\eta}{N} f^{abc}, \tag{10.44}$$

$$e_{abc} = 0. \tag{10.45}$$

Summarising the effect of all constraints, we can cast $\mathcal{L}^{(\epsilon)}$ in the form

$$\begin{aligned}\mathcal{L}^{(\epsilon)} &= -\frac{Z_A}{4}\tilde{G}^{a\mu\nu}\tilde{G}^a_{\mu\nu} + i Z_\psi \bar{\psi}\gamma^\mu\left[\partial_\mu - i\frac{g}{N}T^a A^a_\mu\right]\psi \\ &\quad + \frac{\xi'}{2} w^a w^a + \left(\frac{Z_\eta}{N\lambda}\right) w^a \partial_\mu A^{a\mu} + Z_\eta\left(\partial_\mu \bar{\eta}^a\right)\left(\partial^\mu \eta^a\right) \\ &\quad + Z_\eta \frac{g}{N} f^{abc}\left(\partial_\mu \bar{\eta}^a\right) A^{b,\mu}\eta^c,\end{aligned} \tag{10.46}$$

where the field strength tensor $\tilde{G}^{a\mu\nu}$ is evaluated as in $G^{a\mu\nu}$ with the replacement $g \to g/N$. For $\lambda = N = Z_A = Z_\psi = Z_\eta = 1$, $\xi' = \xi$ this is the classical Lagrangian $\mathcal{L}$ above.

This is a Lagrangian which is very similar to the classical Lagrangian $\mathcal{L}$, differing only in multiplicative constants. This tells us that the two Lagrangians describe the same physics, since we are allowed to rescale the definitions of fields, couplings and masses at will. We can exploit this freedom to remove all ultraviolet divergences. With explicit calculations of a few Green's functions at the N-th loop order, we can find how the constants that emerged in $\mathcal{L}^{(\epsilon)}$ (which necessarily contain the infinities of all matrix elements) are related to the original parameters and field definitions of the classical Lagrangian. With this information at hand, and by reverse engineering, we can redefine the fields, fermion masses and coupling constant so that at the N-th loop order we have $\Gamma^{(\epsilon)} = S_R$, which renders $\Gamma_{N,\infty} = 0$.

We have proven that non-abelian gauge theories are renormalisable, in the sense that multiplicative redefinitions of fields and parameters order by order in the loop expansion can remove all ultraviolet infinities from Green's functions. This is one of the biggest successes of quantum field theory, since we now have realistic theories with predictive power for physical (i.e. finite) observables. *Remark.* The proof assumes that the Zinn-Justin equation $(\Gamma, \Gamma) = 0$ holds for the regularised theory at every order, i.e. that there is a BRST-preserving regulator. Dimensional regularisation provides one for vector-like theories such as QCD; for chiral gauge theories (the electroweak Standard Model) this requires the cancellation of gauge anomalies, otherwise $(\Gamma, \Gamma) \neq 0$ at one loop and renormalisability fails. Fermion masses ($\bar{\psi}\psi$, dimension 3) are also allowed in $\mathcal{L}_{\psi A}$ and are omitted in Eq. (10.46). See Weinberg, QFT II, sec. 17.2 and ch. 22; Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, ch. 21.

PART IV

The Physics of Electroweak Symmetry Breaking

Preface to Part IV

This part is the script of the third course in the sequence, *The Physics of Electroweak Symmetry Breaking*. The gauge theory of the electroweak interactions is extremely successful, yet the mechanism which gives mass to the W and Z bosons raises questions of its own. Why is the electroweak scale so much smaller than the scales at which new physics might appear? Could the Higgs boson be a composite particle, like the pion? How can precision measurements tell us about physics which is too heavy to be produced directly? To address such questions we need a good understanding of spontaneous symmetry breaking and of effective field theories, and this is what the course aims to provide.

We start with a recap of the quantum effective action and of spontaneous symmetry breaking, which were introduced in Part III, and give a second proof of Goldstone's theorem, due to Weinberg, which relies only on the existence of a conserved current and on the spectral properties of the theory. We then derive the soft theorems for the emission of Goldstone bosons, and study the $O(N)$ linear sigma model as an explicit example. Symmetries of nature are often only approximate; we show how a small explicit breaking aligns the vacuum and gives masses to pseudo-Goldstone bosons. The chiral symmetry of QCD and the pions are the prime example: we discuss the Goldberger–Treiman relation and the pion mass, and we construct the effective field theory of pions, the non-linear sigma model.

The second half of the course is devoted to electroweak symmetry breaking proper. We study how non-renormalisable effective theories behave at one loop, with the Fermi theory and the Standard Model in the limit of a very heavy Higgs boson as examples, and how loops point to the scale where new physics must appear. We then develop the formalism of oblique corrections and of the S, T and U parameters, with which precision electroweak data constrain extensions of the Higgs sector. The last chapter introduces little Higgs models, in which the Higgs boson is a pseudo-Goldstone boson of a larger global symmetry and its mass is protected by collective symmetry breaking.

The chapter on little Higgs models was written by Giuliano Panico, who was the assistant of the course. I am grateful to him for this contribution.

1

Recap: the effective action and spontaneous symmetry breaking

Conventions

Throughout this part we use the following conventions, which agree with Parts I–III and, up to the metric, with Weinberg, *The Quantum Theory of Fields*, Vol. II, Ch. 19.

- Metric $g_{\mu\nu} = \mathrm{diag}(+1,-1,-1,-1)$, $p\cdot x = p^0 x^0 - \vec{p}\cdot\vec{x}$.
- Heisenberg picture and translations: $\hat{O}(x) = e^{i\hat{P}\cdot x}\,\hat{O}(0)\,e^{-i\hat{P}\cdot x}$, hence $\langle a|\,\hat{O}(x)\,|b\rangle = e^{i(P_a-P_b)\cdot x}\,\langle a|\,\hat{O}(0)\,|b\rangle$. One-particle states are normalised to $\delta^{(3)}(\vec{p}-\vec{p}\,')$, so that $\langle 0|\,\hat{O}(x)\,|p\rangle \propto e^{-ip\cdot x}/\big((2\pi)^{3/2}\sqrt{2p^0}\big)$.
- Internal symmetries: $\phi_n \to \phi_n + i\epsilon^a\,(T^a)_{nm}\,\phi_m$ with hermitian generators T^a (purely imaginary and antisymmetric for real fields; written t_{nm} in this chapter and in the chapter on spontaneous symmetry breaking of Part III). The conserved charges act as $\big[Q_a,\phi_n\big] = -(T^a)_{nm}\,\phi_m$, with Noether current $J^\mu_a = -i\,\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_n)}\,(T^a)_{nm}\phi_m$. The symbol $\delta\phi$ (or $\delta\,\langle\phi\rangle$) is defined locally in each chapter where it is used.
- Spectral function, with ρ_n real:

$$\sum_N \delta^{(4)}(p-P_N)\,\langle 0|\,J^\mu(0)\,|N\rangle\,\langle N|\,\phi_n(0)\,|0\rangle = i\,(2\pi)^{-3}\,p^\mu\rho_n(p^2)\,\theta(p^0).$$

- Goldstone bosons: $\langle 0|\,J^\mu_a(x)\,|B_b\rangle = iF_{ab}\,p^\mu e^{-ip\cdot x}/\big((2\pi)^{3/2}\sqrt{2p^0}\big)$ and $\langle B_a|\,\phi_n(x)\,|0\rangle = Z_{an}\,e^{+ip\cdot x}/\big((2\pi)^{3/2}\sqrt{2p^0}\big)$, with F, Z real, satisfying the sum rule $i\sum_b F_{ab}Z_{bn} = -\sum_m T^a_{nm}\,\langle\phi_m\rangle$.
- Pions: $F_{ab} = \frac{F_\pi}{2}\delta_{ab}$ with $F_\pi \simeq 184$ MeV ($= 2f_\pi$, $f_\pi \simeq 92$ MeV).

1.1 The effective action

In a quantised field theory, we compute expectation values and physical observables through a "partition functional",

$$Z[J] = e^{iW[J]} = \int \mathcal{D}\phi\; e^{i\left(S[\phi]+\int d^4x\,J(x)\phi(x)\right)}, \tag{1.1}$$

where the path integral extends over all field paths. For simplicity we consider here only one (scalar) field and one source term $J(x)$, but the discussion can be generalised. $Z[J]$ is the sum of all Feynman diagrams, while $W[J]$ is the sum of all connected Feynman diagrams.

Expectation values of field operators in the ground state can be obtained by acting with functional derivatives $\frac{1}{i}\frac{\delta}{\delta J(x_i)}$ on $Z[J]$ and $W[J]$. In the simplest cases,

$$\langle \text{vac}|\, \text{vac}\rangle_J = Z[J], \tag{1.2}$$

$$\langle \text{vac}|\, \hat{\phi}(x_1)\, |\text{vac}\rangle_J = \frac{1}{i}\frac{\delta}{\delta J(x_1)} Z[J], \tag{1.3}$$

and so on. The vacuum expectation value of the field operator is

$$\left\langle \hat{\phi}(x)\right\rangle_J \equiv \frac{\langle \text{vac}|\, \hat{\phi}(x)\, |\text{vac}\rangle_J}{\langle \text{vac}|\, \text{vac}\rangle_J} = \frac{1}{i}\frac{1}{Z[J]}\frac{\delta Z[J]}{\delta J(x)} \quad \rightsquigarrow \quad \boxed{\left\langle \hat{\phi}(x)\right\rangle_J = \frac{\delta W[J]}{\delta J(x)}.} \tag{1.4}$$

We have considered a situation where the system is "fed" with an external source $J(x)$. The more interesting situation with $J(x) = 0$ gives the field vacuum expectation value of a closed system which is left to relax to its ground state.

We can integrate the above differential equation and obtain

$$W[J] = \int d^4x\, \left\langle \phi(x)\right\rangle_J J(x) + \Gamma\left[\left\langle \phi\right\rangle_J\right], \tag{1.5}$$

where $\Gamma[\left\langle \phi\right\rangle_J]$ plays the role of a "constant of integration". It is the quantum effective action (see the chapter on the quantum effective action in Part III for its detailed properties). $\Gamma[\left\langle \phi\right\rangle_J]$ depends on $\left\langle \phi\right\rangle_J$, and only through it does it depend on the source J. Differentiating Eq. (1.5) with respect to the field vev, we obtain

$$\begin{aligned}\frac{\delta\Gamma[\left\langle \phi\right\rangle_J]}{\delta\left\langle \phi\right\rangle_J(x)} &= \frac{\delta W[J]}{\delta\left\langle \phi\right\rangle_J(x)} - \frac{\delta}{\delta\left\langle \phi\right\rangle_J(x)}\int d^4y\, J(y)\left\langle \phi\right\rangle_J(y) \\ &= \int d^4y\, \frac{\delta J(y)}{\delta\left\langle \phi\right\rangle_J(x)}\frac{\delta W[J]}{\delta J(y)} - J(x) - \int d^4y\, \frac{\delta J(y)}{\delta\left\langle \phi\right\rangle_J(x)}\left\langle \phi\right\rangle_J(y).\end{aligned} \tag{1.6}$$

Since $\delta W[J]/\delta J(y) = \left\langle \phi\right\rangle_J(y)$, the first and last terms cancel and

$$\boxed{\frac{\delta\Gamma[\left\langle \phi\right\rangle_J]}{\delta\left\langle \phi\right\rangle_J(x)} = -J(x).} \tag{1.7}$$

Consider the physical case of interest with no external sources (a closed system, the universe). Then

$$\frac{\delta\Gamma[\left\langle \phi\right\rangle]}{\delta\left\langle \phi\right\rangle} = 0, \tag{1.8}$$

and the vacuum expectation value of the field operator corresponds to a stationary point of the effective action.

1.2 Symmetries of the effective action

1.2.1 A discrete symmetry: $\phi \to -\phi$

Assume that the action $S[\phi]$ is symmetric under

$$\phi \leftrightarrow -\phi \ : \qquad S[\phi] = S[-\phi].$$

Then the effective action shares the same symmetry:

$$\Gamma\left[\langle\phi\rangle_J\right] = \Gamma\left[-\langle\phi\rangle_J\right]. \tag{1.9}$$

Proof: The functional

$$\begin{aligned} e^{iW[J]} &= \int \mathcal{D}\phi \; e^{i\left(S[\phi]+\int d^4x\, J(x)\phi(x)\right)} = \int \mathcal{D}\phi \; e^{i\left(S[-\phi]+\int d^4x\,(-J(x))(-\phi(x))\right)} \\ &\overset{\bar\phi=-\phi}{=} \int \mathcal{D}\bar\phi \; e^{i\left(S[\bar\phi]+\int d^4x\,(-J(x))\bar\phi(x)\right)} = e^{iW[-J]}, \end{aligned} \tag{1.10}$$

so that

$$\boxed{W[-J] = W[J].} \tag{1.11}$$

The vacuum expectation value in the presence of a source $J_p(x) = -J(x)$ is

$$\langle\phi\rangle_{J_p}(x) = \frac{\delta W[J_p]}{\delta J_p(x)}$$

and therefore

$$\langle\phi\rangle_{-J}(x) = \frac{\delta W[-J]}{\delta(-J(x))} = -\frac{\delta W[-J]}{\delta J(x)} \overset{(1.11)}{=} -\frac{\delta W[J]}{\delta J(x)} \quad \rightsquigarrow \quad \langle\phi\rangle_{-J}(x) = -\langle\phi\rangle_J(x). \tag{1.12}$$

Then we can easily arrive at

$$\begin{aligned} \Gamma\left[-\langle\phi\rangle_J(x)\right] &= \Gamma\left[\langle\phi\rangle_{-J}(x)\right] = W[-J] - \int d^4x\, \langle\phi\rangle_{-J}(x)\,(-J(x)) \\ &= W[J] - \int d^4x\, \left(-\langle\phi\rangle_J(x)\right)(-J(x)) = W[J] - \int d^4x\, \langle\phi\rangle_J(x) J(x) \\ &= \Gamma\left[\langle\phi\rangle_J(x)\right]. \end{aligned} \tag{1.13}$$

1.2.2 General symmetry transformations

Does the effective action always share the symmetries of the classical action? No! Let us see when exactly this happens. Take a symmetry transformation

$$\phi \to \phi + \epsilon F(x,\phi), \tag{1.14}$$

with F a functional of ϕ, which leaves the integration measure and the action invariant:

$$\mathcal{D}\left(\phi + \epsilon F(x,\phi)\right) = \mathcal{D}\phi, \qquad S\left[\phi + \epsilon F(x,\phi)\right] = S[\phi].$$

Under this transformation the partition functional becomes

$$\begin{aligned} Z[J] &= \int \mathcal{D}(\phi+\epsilon F)\; e^{i\left[S[\phi+\epsilon F]+\int d^4y\,(\phi+\epsilon F)J\right]} \\ &= \int \mathcal{D}\phi \; e^{i\left(S[\phi]+\int d^4y\,\phi(y)J(y)\right)+i\epsilon\int d^4y\,F(y,\phi(y))J(y)}, \end{aligned} \tag{1.15}$$

which after Taylor expansion in ϵ gives

$$Z[J] = Z[J] + i\epsilon\, Z[J] \int d^4y\, \left\langle F(y,\phi(y))\right\rangle_J J(y), \tag{1.16}$$

with

$$\left\langle F(y,\phi(y))\right\rangle_J = \frac{\int \mathcal{D}\phi\; e^{i\left[S[\phi]+\int d^4x\, J\phi\right]} F(y,\phi(y))}{Z[J]} \tag{1.17}$$

the "quantum average" of the functional $F(y,\phi(y))$. The transformation $\phi \to \phi + \epsilon F(x,\phi)$ is thus also a symmetry transformation of $Z[J]$ at the quantum level, $Z[J] \to Z[J]$, only if

$$0 = \int d^4y\, \langle F\rangle_J(y) J(y). \tag{1.18}$$

But, from Eq. (1.7),

$$J(y) = -\frac{\delta\Gamma[\langle\phi\rangle_J]}{\delta\langle\phi\rangle_J(y)},$$

and therefore

$$\boxed{\int d^4y\, \langle F\rangle_J(y) \frac{\delta\Gamma[\langle\phi\rangle_J]}{\delta\langle\phi\rangle_J(y)} = 0} \tag{1.19}$$

(Slavnov–Taylor identities; compare Eq. (7.4) of Part III).

We have just proven that if $\phi \to \phi + \epsilon F$ is a symmetry transformation of the partition function $Z[J]$, then

$$\langle\phi\rangle \to \langle\phi\rangle + \epsilon\langle F\rangle_J$$

is a symmetry transformation of the effective action $\Gamma[\phi]$. The two transformations are not necessarily the same, since F may or may not be $\langle F\rangle$!

1.2.3 Linear transformations

For linear transformations we have

$$F = \langle F\rangle \qquad \text{(exercise)}.$$

Indeed, consider

$$\phi_i(x) \to \phi_i'(x) = \phi_i(x) + \epsilon \underbrace{\left(\int d^4y\, T_{ij}(x,y)\phi_j(y) + s_i(x)\right)}_{F_i}, \tag{1.20}$$

and

$$\langle F_i\rangle = s_i(x) + \int d^4y\, T_{ij}(x,y)\left\langle\phi_j\right\rangle_{J_\phi}(y). \tag{1.21}$$

But

$$\left\langle\phi\right\rangle_{J_\phi} = \phi \qquad \text{(prove it!)},$$

where J_ϕ is the source for which the vacuum expectation value of the field is ϕ. So

$$\langle F\rangle = F.$$

1.2.4 Summary

Let us summarise, adding some further points. For a quantised field theory with a generating functional $Z[J] = e^{iW[J]}$ we can compute the quantum effective action

$$\Gamma[\langle\phi\rangle] = W[J] - \int d^4x\, J(x)\,\langle\phi\rangle\,(x). \tag{1.22}$$

The vacuum expectation value in the presence of a source J is

$$\langle\phi\rangle_J = \frac{\langle 0|\,\phi(x)\,|0\rangle_J}{\langle 0|\,0\rangle_J} = \frac{\delta W[J]}{\delta J(x)}. \tag{1.23}$$

Also,

$$\frac{\delta\Gamma[\langle\phi\rangle]}{\delta\,\langle\phi\rangle_J\,(x)} = -J(x). \tag{1.24}$$

The effective action $\Gamma[\langle\phi\rangle]$ has an extremum for a physical ($J = 0$) vacuum expectation value of the field operator:

$$J = 0 \quad \rightsquigarrow \quad \frac{\delta\Gamma}{\delta\,\langle\phi\rangle} = 0.$$

Assume that $Z[J]|_{J=0}$ is symmetric under a symmetry transformation

$$\phi_i \to \phi_i' = \phi_i + \delta\phi_i = \phi_i + F[\phi_i].$$

Then the effective action is symmetric under

$$\langle\phi_i\rangle \to \langle\phi_i\rangle + \langle F[\phi_i]\rangle\,.$$

For discrete symmetries or linear symmetry transformations we have

$$\langle F[\phi_i]\rangle = F[\langle\phi_i\rangle],$$

and *the symmetries of the action are also symmetries of the effective action*. *Remark.* This presupposes, as assumed above, that the path-integral measure (and the regularisation) is also invariant. If the measure is not invariant the symmetry is anomalous and is not inherited by Γ, even for linear transformations (e.g. the axial $U(1)$ of QCD, or the scale symmetry of massless ϕ^4); see Weinberg, QFT Vol. II, Sec. 16.4 and Ch. 22.

1.3 Spontaneous symmetry breaking

Symmetry transformations which leave the effective action intact may not be symmetries of the physical states and the vacuum state. These symmetries are "*spontaneously broken*". (See also the chapter on spontaneous symmetry breaking in Part III.)

Spontaneous symmetry breaking is associated with a degeneracy of the ground state (vacuum). Assume that an effective action is symmetric under

$$\langle\phi\rangle \to -\langle\phi\rangle\,,$$

a symmetry which is inherited from the classical action (as we have seen in Section 1.2). Assume that the physical field vev is

$$\langle\phi\rangle_a \;:\quad \frac{\delta\Gamma}{\delta\,\langle\phi\rangle_a} = 0, \quad \text{with} \quad \langle\phi\rangle_a \neq 0.$$

In other words, there is a vacuum state $|V_a\rangle$ with

$$\left\langle \phi \right\rangle_a \equiv \frac{\langle V_a | \hat{\phi}(x) | V_a \rangle}{\langle V_a | V_a \rangle} \neq 0, \tag{1.25}$$

for which $V_{\text{eff}}(\left\langle \phi \right\rangle_a)$ is minimum, i.e. $\Gamma[\left\langle \phi \right\rangle_a] = \Gamma_a$ is maximum ($\Gamma = -\int d^4x\, V_{\text{eff}}$ for constant fields). Then there should be a second value of the field vev in a different state which also gives the same value for the effective action:

$$\exists\, |V_b\rangle \;:\quad \left\langle \phi \right\rangle_b = -\left\langle \phi \right\rangle_a \;:\quad \Gamma[\left\langle \phi \right\rangle_b] = \Gamma[\left\langle \phi \right\rangle_a] = \text{maximum}.$$

So, while $\phi \to -\phi$ preserves the effective action,

$$\Gamma[\left\langle \phi \right\rangle] \to \Gamma[-\left\langle \phi \right\rangle] = \Gamma[\left\langle \phi \right\rangle],$$

under the same transformation

$$|V_a\rangle \longrightarrow |V_b\rangle\,,$$

which is a different state. The symmetry is broken, as long as the true vacuum is one of the two states.

1.3.1 Why the system chooses one vacuum

The two states $|V_a\rangle$ and $|V_b\rangle$ are not the only candidates for the vacuum: any linear combination of them is degenerate with them as well. How do we know that the system will choose one of the two states to live in, and does not tunnel from one to the other? By symmetry ($\phi \to -\phi$) we know that

$$\begin{aligned} \langle V_a | H | V_a \rangle &= \langle V_b | H | V_b \rangle = E_{\text{diag}}, \\ \langle V_a | H | V_b \rangle &= \langle V_b | H | V_a \rangle = E_{\text{non-diag}}. \end{aligned} \tag{1.26}$$

In field theory, where the volume is taken to be infinite, we can show that

$$E_{\text{non-diag}} = 0$$

for any Hamiltonian.

To see this, take two local operators A and B and a point $\vec{x}$ outside the light-cone of the origin. By causality,

$$\langle u | \left[B(0), A(\vec{x}) \right] | v \rangle = \langle u | 0 | v \rangle = 0, \tag{1.27}$$

which leads to

$$\sum_w \langle u | A(0) | w \rangle \langle w | B(0) | v \rangle = \sum_w \langle u | B(0) | w \rangle \langle w | A(0) | v \rangle\,. \tag{1.28}$$

Take $\hat{A}$, $\hat{B}$ Hermitian. Then the matrices $(\langle u | A(0) | w \rangle)$ and $(\langle u | B(0) | w \rangle)$ commute,

$$\sum_w (A_{uw} B_{wv} - B_{uw} A_{wv}) = 0, \qquad A_{uw} \equiv \langle u | A(0) | w \rangle\,,$$

for any $\hat{A}$ and $\hat{B}$. All $\langle u | A(0) | v \rangle$ can be simultaneously diagonalised, and so

$$\langle u | A(0) | v \rangle = \delta_{uv}\, a_u, \qquad a_u \in \mathbb{R}, \tag{1.29}$$

with $\langle u | v \rangle = \delta_{uv}$. For $u \neq v$,

$$\langle u | A(0) | v \rangle = 0.$$

Remark. The argument above is a compressed version of Weinberg, QFT Vol. II, Sec. 19.1. There A, B are replaced by their spatial averages $\bar{A} = \lim_{V\to\infty} V^{-1}\int_V d^3x\, A(\vec{x},0)$, whose commutator vanishes by causality as $V \to \infty$, and u, v, w run only over the degenerate vacua ($\bar{A}$ has no matrix elements from a vacuum to non-vacuum states, by translation invariance). The conclusion is that there is a basis of vacua in which all such matrices are diagonal; it holds in that basis, not for an arbitrary choice of degenerate vacua.

Different but degenerate vacua diagonalise any perturbation made out of local operators which break the symmetry. Linear combinations do not diagonalise such perturbations and are therefore unstable.

1.4 Goldstone bosons

Take a symmetry transformation which is linear,

$$\phi_n(x) \to \phi_n(x) + i\epsilon \sum_m t_{nm}\phi_m(x), \tag{1.30}$$

and leaves the action, and thus also the effective action, invariant:

$$\Gamma\left[\left\langle\phi_n\right\rangle + i\epsilon\sum_m t_{nm}\left\langle\phi_m\right\rangle\right] = \Gamma\left[\left\langle\phi_n\right\rangle\right] \quad\rightsquigarrow\quad \int d^4x\,\frac{\delta\Gamma[\phi]}{\delta\phi_n(x)}\,t_{nm}\phi_m(x) = 0 \tag{1.31}$$

(a Slavnov–Taylor identity). Taking a second derivative,

$$\int d^4x\,\frac{\delta^2\Gamma[\phi]}{\delta\phi_l(y)\delta\phi_n(x)}\,t_{nm}\left\langle\phi_m(x)\right\rangle + \frac{\delta\Gamma[\phi]}{\delta\phi_n(y)}\,t_{nl} = 0. \tag{1.32}$$

Consider physical systems where $J = 0$ and

$$\left.\frac{\delta\Gamma}{\delta\phi}\right|_{\phi=\langle\phi\rangle} = 0.$$

Then

$$\int d^4x\,\frac{\delta^2\Gamma[\phi]}{\delta\left\langle\phi_l(y)\right\rangle\,\delta\left\langle\phi_n(x)\right\rangle}\,t_{nm}\left\langle\phi_m(x)\right\rangle = 0. \tag{1.33}$$

For systems with translational invariance,

$$\left\langle\phi(x)\right\rangle = \langle 0|\,e^{i\hat{P}\cdot x}\phi(0)e^{-i\hat{P}\cdot x}\,|0\rangle = \left\langle\phi\right\rangle_0 = \text{constant}, \tag{1.34}$$

so that $\left\langle\phi_l(y)\right\rangle = \left\langle\phi_l\right\rangle$. Also

$$\Gamma[\left\langle\phi\right\rangle] = -\int d^4x\, V(\left\langle\phi\right\rangle),$$

with V the effective potential (Eq. (6.25) of Part III), and we get

$$\boxed{\sum_{m,n}\frac{\partial^2 V(\phi)}{\partial\left\langle\phi_l\right\rangle\,\partial\left\langle\phi_n\right\rangle}\,t_{nm}\left\langle\phi_m\right\rangle = 0.} \tag{1.35}$$

1.4.1 The second derivative of the effective potential and the propagator

Let us start from

$$\begin{aligned}
&\frac{\delta W[J]}{\delta J_m(x)} = \langle \phi_m(x) \rangle \\
\rightsquigarrow \quad &\frac{\delta^2 W[J]}{\delta \langle \phi_n(y) \rangle \, \delta J_m(x)} = \delta(x-y)\, \delta_{nm} \\
\rightsquigarrow \quad &-\delta(x-y)\, \delta_{nm} = \int d^4z \, \frac{\delta^2 W[J]}{\delta J_k(z) \delta J_m(x)} \left(-\frac{\delta J_k(z)}{\delta \langle \phi_n \rangle (y)} \right) \\
&\qquad = i \int d^4z \, \langle 0 | \, T \phi_k(z) \phi_m(x) \, | 0 \rangle_c \, \frac{\delta^2 \Gamma}{\delta \langle \phi_n \rangle (y) \delta \langle \phi_k \rangle (z)}. \qquad (1.36)
\end{aligned}$$

But

$$\langle 0 | \, T \phi_k(z) \phi_n(x) \, | 0 \rangle_c = \int \frac{d^4p}{(2\pi)^4} \, e^{-ip\cdot(x-z)} D_{kn}(p). \qquad (1.37)$$

Inserting this above and integrating over y (since, for constant fields, $\Gamma = -\int d^4x \, V$ implies $\int d^4y \, \delta^2\Gamma / \delta \langle \phi_n \rangle (y) \delta \langle \phi_k \rangle (z) = -\partial^2 V / \partial \langle \phi_n \rangle \, \partial \langle \phi_k \rangle$), we find

$$-\delta_{nl} = -i \int d^4z \int \frac{d^4p}{(2\pi)^4} \, e^{-ip\cdot(x-z)} \, D_{kl}(p) \, \frac{\partial^2 V(\langle \phi \rangle)}{\partial \langle \phi_n \rangle \, \partial \langle \phi_k \rangle} = -i \, \frac{\partial^2 V(\langle \phi \rangle)}{\partial \langle \phi_n \rangle \, \partial \langle \phi_k \rangle} \, D_{kl}(0), \qquad (1.38)$$

$$\rightsquigarrow \quad \boxed{\delta_{nl} = i \sum_k \frac{\partial^2 V}{\partial \langle \phi_n \rangle \, \partial \langle \phi_k \rangle} \, D_{kl}(0)} \quad \rightsquigarrow \quad \boxed{\frac{\partial^2 V}{\partial \langle \phi_n \rangle \, \partial \langle \phi_l \rangle} = -i D^{-1}_{nl}(0).} \qquad (1.39)$$

For a single field with $D(p) = iZ/(p^2 - m^2) + \ldots$ this gives $V'' = m^2/Z$, the (positive) mass squared.

1.4.2 Massless particles

Going back to Eq. (1.35),

$$\boxed{\sum_{m,n} D^{-1}_{ln}(0) \, t_{nm} \, \langle \phi_m \rangle = 0.} \qquad (1.40)$$

When is Eq. (1.40) satisfied? Let us write

$$\delta \langle \phi_n \rangle = i\epsilon \sum_m t_{nm} \langle \phi_m \rangle .$$

Then Eq. (1.40) becomes

$$\boxed{\sum_n D^{-1}_{ln}(0) \, \delta \langle \phi_n \rangle = 0.} \qquad (1.41)$$

If the symmetry leaves the vacuum state unchanged, $\delta \langle \phi_n \rangle = 0$, then Eq. (1.41) is satisfied. If the symmetry is broken,

$$\left(D^{-1}_{lm}(0) \right) \left(\delta \langle \phi_m \rangle \right) = 0,$$

i.e. $\delta \langle \phi_m \rangle$ is an eigenvector of $D^{-1}_{lm}(0)$ with zero eigenvalue. So $D^{-1}_{lm}(p) \to 0$ as $p \to 0$ in the direction $\delta \langle \phi_m \rangle$, and

$$D_{lm}(p) \sim \frac{1}{p^2}$$

has a pole at $p^2 = 0$. There is a massless particle in the spectrum of the theory: the "*Goldstone boson*" (compare Eqs. (8.8)–(8.16) of Part III).

1.4.3 Second proof

A second proof starts from the observation that a continuous symmetry implies a conserved current $J^\mu(x)$. It is developed in the next chapter.

2

Goldstone's theorem from the spectral representation

2.1 Review: effective action and broken symmetries

We briefly recall what we have learned about the effective action and spontaneously broken symmetries (see the chapter on spontaneous symmetry breaking in Part III).

- **Effective action.** The physical vacuum expectation values of the field operators are stationary points of the effective action,

$$\frac{\delta\Gamma}{\delta\langle\phi\rangle} = 0. \tag{2.1}$$

  A symmetry transformation of the fields, $\phi \to \phi + \epsilon F[\phi]$, induces the transformation

$$\langle\phi\rangle \to \langle\phi\rangle + \epsilon\langle F[\phi, x]\rangle \tag{2.2}$$

  of the vacuum expectation values, which keeps the effective action invariant. Remember that, in general, $\langle F[\phi]\rangle \neq F[\langle\phi\rangle]$. Linear continuous symmetry transformations of the action are also symmetry transformations of the effective action.

- **Spontaneously broken symmetries.** The symmetries of the effective action do not leave the vacuum state invariant. There is a degeneracy: $\langle\phi\rangle$ and $L\langle\phi\rangle$, with L a symmetry transformation, correspond to degenerate vacuum states, characterised by different vacuum expectation values of the field operator. These orthonormal vacuum states are stable: once the system is in one of them, a tunnelling through infinite space is required to move to another.

- **Goldstone's theorem.** Spontaneous symmetry breaking implies massless particles.

In Part III we derived Goldstone's theorem from the symmetry of the effective potential (Eq. (8.11)). Here we give a different proof, due to Weinberg, which relies only on the existence of a conserved current and on general properties of the spectrum of states.

2.2 Conserved currents and charges

Consider an action which is invariant under a continuous transformation. From Noether's theorem we find a conserved current

$$J^\mu(x) = J^\mu(\vec{x}, t), \qquad \partial_\mu J^\mu = 0, \tag{2.3}$$

due to the symmetry. This gives a conserved (time independent) charge

$$Q = \int d^3\vec{x}\, J^0(\vec{x}, t). \tag{2.4}$$

Under the symmetry transformation the quantum fields transform as

$$\left[Q, \phi_n\right] = i\delta\phi_n, \tag{2.5}$$

with

$$\epsilon\, \delta\phi_n = +i\epsilon\, T_{nm}\phi_m \tag{2.6}$$

for a continuous symmetry transformation; all of ϕ_n, $\delta\phi_n$, Q are now operators. (The classical field transforms as $\phi \to \phi + \epsilon\, \delta\phi$.)

To study the spontaneous breaking of the global symmetry we start from the vacuum average of the commutator of the current with the field,

$$\left\langle \left[J^\mu(y), \phi_n(x)\right] \right\rangle = \frac{\langle 0|\left[J^\mu(y), \phi_n(x)\right]|0\rangle}{\langle 0|\,0\rangle}. \tag{2.7}$$

We take the vacuum normalised, $\langle 0|\,0\rangle = 1$. Writing out the commutator and inserting a complete set of momentum eigenstates,

$$\hat{P}^\mu |N\rangle = P_N^\mu |N\rangle, \tag{2.8}$$

we obtain

$$\begin{aligned} \langle 0|\left[J^\mu(y), \phi_n(x)\right]|0\rangle &= \langle 0|J^\mu(y)\phi_n(x) - \phi_n(x)J^\mu(y)|0\rangle \\ &= \sum_N \Big\{ \langle 0|J^\mu(y)|N\rangle\langle N|\phi_n(x)|0\rangle - \langle 0|\phi_n(x)|N\rangle\langle N|J^\mu(y)|0\rangle \Big\} \end{aligned} \tag{2.9}$$

2.3 The spectral representation of the commutator

Translation symmetry is represented on operators as

$$\hat{O}(x) = e^{+i\hat{P}\cdot x}\, \hat{O}(0)\, e^{-i\hat{P}\cdot x}. \tag{2.10}$$

Remark. This is the Heisenberg evolution of Part I (Eq. (13.6)) in our metric (signature $+,-,-,-$). Weinberg (QFT Vol. II, Sec. 19.2) writes $e^{-iP\cdot x}\hat{O}(0)e^{iP\cdot x}$, which is the same statement in his mostly-plus metric; all phases below are written in our metric. For two momentum eigenstates $\hat{P}|a\rangle = P_a|a\rangle$ and $\hat{P}|b\rangle = P_b|b\rangle$ we then have

$$\langle a|\hat{O}(x)|b\rangle = e^{+iP_a\cdot x}e^{-iP_b\cdot x}\langle a|\hat{O}(0)|b\rangle = e^{+i(P_a-P_b)\cdot x}\langle a|\hat{O}(0)|b\rangle. \tag{2.11}$$

Thus,

$$\begin{aligned} \langle 0|\left[J^\mu(y), \phi_n(x)\right]|0\rangle &= \sum_N \Big\{ e^{-iP_N\cdot(y-x)}\langle 0|J^\mu(0)|N\rangle\langle N|\phi_n(0)|0\rangle \\ &\qquad - e^{-iP_N\cdot(x-y)}\langle 0|\phi_n(0)|N\rangle\langle N|J^\mu(0)|0\rangle \Big\} \\ &= \int d^4p \sum_N \delta^{(4)}(p - P_N)\Big\{ e^{-ip\cdot(y-x)}\langle 0|J^\mu(0)|N\rangle\langle N|\phi_n(0)|0\rangle \\ &\qquad - e^{-ip\cdot(x-y)}\langle 0|\phi_n(0)|N\rangle\langle N|J^\mu(0)|0\rangle \Big\}. \end{aligned} \tag{2.12}$$

Let us perform the sum over the states N. The result can depend only on the momentum p^μ, and by Lorentz invariance it must take the form

$$\sum_N \delta^{(4)}(p - P_N)\ \langle 0|\, J^\mu(0)\,|N\rangle\,\langle N|\,\phi_n(0)\,|0\rangle = i\,(2\pi)^{-3}\,p^\mu\,\rho_n(p^2)\,\theta(p^0) \tag{2.13}$$

and

$$\sum_N \delta^{(4)}(p - P_N)\ \langle 0|\,\phi_n(0)\,|N\rangle\,\langle N|\, J^\mu(0)\,|0\rangle = -i\,(2\pi)^{-3}\,p^\mu\,\tilde\rho_n(p^2)\,\theta(p^0). \tag{2.14}$$

For hermitian J^μ and ϕ_n, the complex conjugate of Eq. (2.13) is Eq. (2.14), so that $\tilde\rho_n = \rho_n^*$; we shall not need this until the end.

Inserting these into the commutator we find

$$\begin{aligned}
\langle 0|\left[J^\mu(y),\phi_n(x)\right]|0\rangle &= \int \frac{d^4p}{(2\pi)^3}\,\theta(p^0)\left[ip^\mu e^{-ip\cdot(y-x)}\rho_n(p^2) + ip^\mu e^{-ip\cdot(x-y)}\tilde\rho_n(p^2)\right] \\
&= -\frac{\partial}{\partial y_\mu}\int \frac{d^4p}{(2\pi)^3}\,\theta(p^0)\left[\rho_n(p^2)e^{-i(y-x)\cdot p} - \tilde\rho_n(p^2)e^{-i(x-y)\cdot p}\right] \\
&= -\frac{\partial}{\partial y_\mu}\int dm^2\Big\{\rho_n(m^2)\left[\int \frac{d^4p}{(2\pi)^3}\,\delta(p^2-m^2)\,\theta(p^0)\,e^{-i(y-x)\cdot p}\right] \\
&\qquad -\tilde\rho_n(m^2)\left[\int \frac{d^4p}{(2\pi)^3}\,\delta(p^2-m^2)\,\theta(p^0)\,e^{-i(x-y)\cdot p}\right]\Big\}
\end{aligned} \tag{2.15}$$

We define

$$\Delta_+(x-y;m^2) \equiv \int \frac{d^4p}{(2\pi)^3}\,\delta(p^2-m^2)\,\theta(p^0)\,e^{-ip\cdot(x-y)}, \tag{2.16}$$

so that

$$\langle 0|\left[J^\mu(y),\phi_n(x)\right]|0\rangle = -\frac{\partial}{\partial y_\mu}\int dm^2\left[\rho_n(m^2)\,\Delta_+(y-x;m^2) - \tilde\rho_n(m^2)\,\Delta_+(x-y;m^2)\right]. \tag{2.17}$$

Let us consider a space-like distance $(x-y)$. Then

$$\Delta_+(x-y;m^2) = \Delta_+(y-x;m^2). \tag{2.18}$$

Exercise: Prove that for a space-like interval $(x-y)$, $\Delta_+(x-y;m^2) = \Delta_+(y-x;m^2)$.

For space-like distances also the commutator vanishes, $\left[J^\mu(y),\phi_n(x)\right] = 0$, by causality. Thus,

$$\tilde\rho_n(m^2) = +\rho_n(m^2). \tag{2.19}$$

Remark. Combined with $\tilde\rho_n = \rho_n^*$ this shows that ρ_n is real. This is consistent with the sum rule (2.27) below: taking the hermitian conjugate of $\left[Q,\phi_n\right] = -T_{nm}\phi_m$ for hermitian Q, ϕ_n gives $T^*_{nm} = -T_{nm}$, so $i\,T_{nm}\left\langle\phi_m\right\rangle$ is real as well. Using this, for generic x and y we obtain

$$\langle 0|\left[J^\mu(y),\phi_n(x)\right]|0\rangle = -\frac{\partial}{\partial y_\mu}\int dm^2\,\rho_n(m^2)\left[\Delta_+(y-x;m^2) - \Delta_+(x-y;m^2)\right]. \tag{2.20}$$

2.4 The Goldstone theorem

2.4.1 The spectral function is concentrated at zero mass

Taking a second derivative, i.e. the divergence with respect to y, and using current conservation, we have

$$\begin{aligned} 0 = \langle 0| \left[\partial_\mu J^\mu(y), \phi_n(x)\right] |0\rangle &= -\int dm^2\, \rho_n(m^2)\, \partial^2 \left[\Delta_+(y-x;m^2) - \Delta_+(x-y;m^2)\right] \\ &= +\int dm^2\, m^2 \rho_n(m^2) \left[\Delta_+(y-x;m^2) - \Delta_+(x-y;m^2)\right] \end{aligned} \tag{2.21}$$

where we used $(\partial^2 + m^2)\Delta_+ = 0$. This leads to

$$\boxed{m^2 \rho_n(m^2) = 0}. \tag{2.22}$$

Let us now consider equal times, $y^0 = x^0 = t$, and the time component $\mu = 0$:

$$\begin{aligned} &\langle 0| \left[J^0(t,\vec{y}), \phi_n(t,\vec{x})\right] |0\rangle \\ &\qquad = -\int dm^2\, \rho_n(m^2) \frac{\partial}{\partial y_0} \left[\Delta_+(y-x;m^2) - \Delta_+(x-y;m^2)\right]_{y^0=x^0}. \end{aligned} \tag{2.23}$$

Writing Δ_+ as an integral over three-momenta,

$$\Delta_+(x-y;m^2) = \int \frac{d^3\vec{p}}{(2\pi)^3\, 2\sqrt{\vec{p}^2+m^2}}\, e^{-i\sqrt{\vec{p}^2+m^2}\,(x^0-y^0)+i\vec{p}\cdot(\vec{x}-\vec{y})}, \tag{2.24}$$

the time derivative brings down a factor $\pm i\sqrt{\vec{p}^2+m^2}$ which cancels the energy denominator, and at equal times we are left with

$$\langle 0| \left[J^0(t,\vec{y}), \phi_n(t,\vec{x})\right] |0\rangle = i \int dm^2\, \rho_n(m^2)\, \delta^{(3)}(\vec{y}-\vec{x}). \tag{2.25}$$

Integrating over $\vec{y}$ we obtain the commutator with the charge $Q = \int d^3\vec{y}\, J^0(t,\vec{y})$,

$$\langle 0| \left[Q, \phi_n(t,\vec{x})\right] |0\rangle = i \int dm^2\, \rho_n(m^2). \tag{2.26}$$

With Eqs. (2.5)–(2.6), $\left[Q, \phi_n\right] = i\delta\phi_n = -T_{nm}\phi_m$, so that

$$\boxed{\int dm^2\, \rho_n(m^2) = i\, T_{nm} \left\langle \phi_m \right\rangle}. \tag{2.27}$$

If spontaneous symmetry breaking occurs, then there is a generator such that the vacuum expectation value is not invariant,

$$\left\langle \phi_n \right\rangle \to \left\langle \phi_n \right\rangle' = \left\langle \phi_n \right\rangle + i\epsilon\, T_{nm} \left\langle \phi_m \right\rangle \neq \left\langle \phi_n \right\rangle, \qquad \text{i.e.} \qquad T_{nm} \left\langle \phi_m \right\rangle \neq 0. \tag{2.28}$$

At the same time, we have $m^2\rho_n(m^2) = 0$. To reconcile the two, the function $\rho_n(m^2)$ must have a singularity at $m^2 = 0$: the "mass spectrum" of the field must be

$$\boxed{\rho_n(m^2) = \delta(m^2)\, i\, T_{nm} \left\langle \phi_m \right\rangle}. \tag{2.29}$$

Indeed, for this form both $m^2\rho_n(m^2) = 0$ and $\int dm^2 \rho_n(m^2) = i\, T_{nm} \left\langle \phi_m \right\rangle$ are satisfied.

2.4.2 Physical meaning: a massless state

What is the physical meaning of $\rho_n(m^2)$? From its definition, Eq. (2.13), we have

$$i\,(2\pi)^{-3}\,p^\mu \rho_n(p^2)\,\theta(p^0) = \sum_N \langle 0|\,J^\mu(0)\,|N\rangle\,\langle N|\,\phi_n(0)\,|0\rangle\;\delta^{(4)}(p-P_N). \tag{2.30}$$

It is made out of a summation over all states. For $\rho_n(p^2) \propto \delta(p^2)$ there must exist a state $|N\rangle$ with $P_N = p$ such that $P_N^2 = 0$.

> This is Goldstone's theorem, as proven by Weinberg. We have found that the spectrum of the theory contains a massless state for each broken generator, $T_{nm}\left\langle\phi_m\right\rangle \neq 0$.

2.5 Properties of the Goldstone bosons

We can get more out of this derivation. Further conclusions one can draw are the following.

One-particle state.. The state $|N\rangle$ (the Goldstone boson state with $p^2 = 0$), which yields the $\delta(p^2)$, must be a one-particle state. Multi-particle states can yield $p^2 = 0$, but in a continuous manner: they give a continuum, not a δ-function. Take, for example, two massless particles flying away from each other at an angle θ. When $\theta \to 0$ they become parallel and their total invariant mass becomes zero.

Spin zero.. The Goldstone boson state $|N\rangle$ is a scalar, i.e. rotationally invariant. In the matrix element

$$\langle N|\,\phi_n(0)\,|0\rangle \tag{2.31}$$

the transformation of $\langle N|$ under rotations must counter the transformation of $\phi_n(0)\,|0\rangle$. The latter is rotationally invariant, since $|0\rangle$ is rotationally invariant and ϕ_n is evaluated at the centre of our reference frame, $x^\mu = 0$:

$$\phi_n(0)\,|0\rangle \to R(\theta)\,\phi_n(0)\,|0\rangle = u(\theta)\,\phi_n\big(R(\theta)\cdot 0\big)\,|0\rangle = u(\theta)\,\phi_n(0)\,|0\rangle = \phi_n(0)\,|0\rangle\,. \tag{2.32}$$

For $\langle N|\,\phi_n(0)\,|0\rangle$ to be invariant under rotations, $\langle N|$ must have zero spin. Therefore, Goldstone bosons are scalars. By the same argument, we conclude that

$$\langle N|\,\phi_n(0)\,|0\rangle = 0 \tag{2.33}$$

for every state $|N\rangle$ with non-zero spin or helicity.

Quantum numbers.. The matrix element

$$\langle N|\,J^0(0)\,|0\rangle \tag{2.34}$$

must be invariant if $|N\rangle$ is a Goldstone state: $\langle N|$ must cancel the symmetry transformation of the state $J^0(0)\,|0\rangle$. Therefore, $|N\rangle$ must have the same symmetry transformations and the same corresponding "quantum numbers" (charges) as $J^0(0)\,|0\rangle$. In addition, this matrix element vanishes,

$$\langle N|\,J^0(0)\,|0\rangle = 0, \tag{2.35}$$

for all states $\langle N|$ with quantum numbers different from those of $J^\mu(0)\,|0\rangle$.

2.6 Goldstone boson decay constant and field normalisation

Let us now focus on the states $|N\rangle = |B\rangle$ which are Goldstone bosons. The matrix element $\langle 0|\, J^\mu(x)\, |B\rangle$ is of the form

$$\langle 0|\, J^\mu(x)\, |B\rangle = \frac{i\, F\, P_B^\mu\, e^{-iP_B\cdot x}}{(2\pi)^{3/2}\sqrt{2P_B^0}}, \qquad P_B^0 = |\vec{P}_B|, \quad P_B^2 = 0. \tag{2.36}$$

F is a constant: by Lorentz invariance it can only depend on P_B^2, so $F = F(P_B^2) = F(0)$. Similarly, for a one-particle state,

$$\langle B|\, \phi_n(y)\, |0\rangle = \frac{Z_n\, e^{+iP_B\cdot y}}{(2\pi)^{3/2}\sqrt{2P_B^0}}, \qquad Z_n = \text{constant}. \tag{2.37}$$

Then, the contribution of the Goldstone boson to the spectral function is

$$\begin{aligned} i\,(2\pi)^{-3}\rho_n(p^2)\, p^\mu\, \theta(p^0) &= \int d^3\vec{P}_B\; \delta^{(4)}(p - P_B)\; \langle 0|\, J^\mu(0)\, |B\rangle\, \langle B|\, \phi_n(0)\, |0\rangle \\ &= \int \frac{d^3\vec{P}_B}{(2\pi)^3}\; \delta^{(3)}(\vec{p} - \vec{P}_B)\, \frac{\delta(p^0 - |\vec{p}|)}{2P_B^0}\; i\, p^\mu F Z_n \\ &= \theta(p^0)\, \delta(p^2)\, i\,(2\pi)^{-3}\, p^\mu F Z_n, \end{aligned} \tag{2.38}$$

so that

$$\rho_n(p^2) = F Z_n\, \delta(p^2). \tag{2.39}$$

We also have $\rho_n(m^2) = \delta(m^2)\, i\, T_{nm}\, \langle \phi_m \rangle$, Eq. (2.29), which gives

$$\boxed{i\, F \cdot Z_n = -T_{nm}\, \langle \phi_m \rangle}. \tag{2.40}$$

Here F is a measure of symmetry breaking, Z_n is the field wave-function normalisation, and $T_{nm}\, \langle \phi_m \rangle$ is the magnitude of symmetry breaking.

2.6.1 Several broken generators

It may happen that we have more than one "broken" generator, T^a_{nm}, $a = 1, 2, \ldots$. For each broken generator there is a Goldstone boson state $|B_a\rangle$. We define

$$\langle 0|\, J_a^\mu(x)\, |B_b\rangle = \frac{i\, F_{ab}\, P_B^\mu\, e^{-iP_B\cdot x}}{(2\pi)^{3/2}\sqrt{2P_B^0}}, \tag{2.41}$$

$$\langle B_a|\, \phi_n(y)\, |0\rangle = \frac{Z_{an}\, e^{+iP_B\cdot y}}{(2\pi)^{3/2}\sqrt{2P_B^0}}. \tag{2.42}$$

Then

$$\boxed{i\sum_b F_{ab} Z_{bn} = -\sum_m T^a_{nm}\, \langle \phi_m \rangle}. \tag{2.43}$$

3

Amplitudes with Goldstone bosons

3.1 Current matrix elements between momentum eigenstates

Consider the matrix element

$$\left\langle \beta \right| J^\mu(x) \left| \alpha \right\rangle , \tag{3.1}$$

where $|\alpha\rangle$, $\left|\beta\right\rangle$ are momentum eigenstates,

$$\hat{P} \left| \alpha \right\rangle = P_\alpha \left| \alpha \right\rangle , \qquad \hat{P} \left| \beta \right\rangle = P_\beta \left| \beta \right\rangle , \tag{3.2}$$

and J^μ is the Noether current corresponding to a broken generator T_{nm}, i.e. $T_{nm} \left\langle \phi_m \right\rangle \neq 0$. Let us assume that we can use perturbation theory and Feynman diagrams; our conclusions will, however, also be valid for non-perturbative theories. There are three types of Feynman diagrams:

(I) the current couples to an external particle;

(II) the current couples to an internal propagator;

(III) the current couples first to a Goldstone boson, and the Goldstone boson is attached somewhere in the diagram.

These are shown in Fig. 3.1.

The matrix element of the current at the point x is related to the one at the origin by translation,

$$\left\langle \beta \right| J^\mu(x) \left| \alpha \right\rangle = e^{-iq\cdot x} \left\langle \beta \right| J^\mu(0) \left| \alpha \right\rangle , \qquad q^\mu \equiv P_\alpha^\mu - P_\beta^\mu . \tag{3.3}$$

Current conservation then gives

$$0 = \left\langle \beta \right| \partial_\mu J^\mu(x) \left| \alpha \right\rangle = -i\, e^{-iq\cdot x}\, q_\mu \left\langle \beta \right| J^\mu(0) \left| \alpha \right\rangle \quad \Rightarrow \quad \boxed{q_\mu \left\langle \beta \right| J^\mu(0) \left| \alpha \right\rangle = 0} . \tag{3.4}$$

3.2 Diagrams of type III: the Goldstone pole

The coupling of the current to the corresponding Goldstone boson $\left|B, q\right\rangle$ is, from Eq. (2.36),

$$\left\langle 0 \right| J^\mu(0) \left| B, q \right\rangle \propto \frac{i\, F q^\mu e^{-iq\cdot 0}}{(2\pi)^{3/2}\sqrt{2q^0}} . \tag{3.5}$$

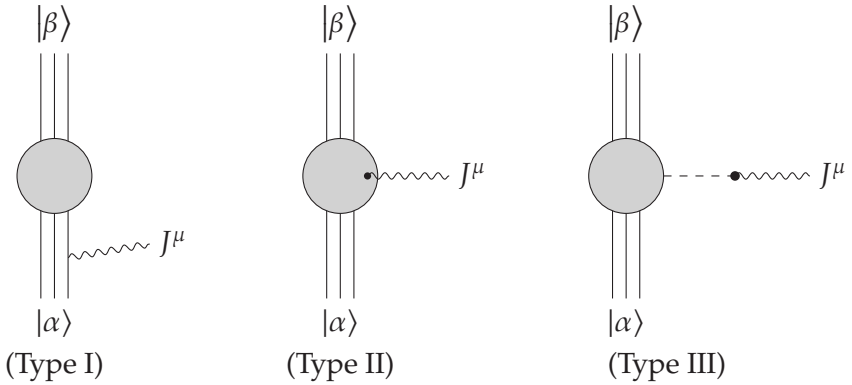


Figure 3.1. The three types of diagrams contributing to $\langle\beta|\, J^\mu\, |\alpha\rangle$. Type I: the current (wavy line) attaches to an external particle. Type II: the current attaches to an internal propagator inside the blob. Type III: the current turns into a Goldstone boson (dashed line), which attaches to the blob.

Also, the intermediate Goldstone propagator is

$$\propto \frac{1}{q^2}. \tag{3.6}$$

We then write

$$\langle\beta|\, J^\mu(0)\, |\alpha\rangle\Big|_{\text{type III}} = \frac{i\,F q^\mu}{q^2}\, M_{\beta\alpha}, \tag{3.7}$$

where

$$M_{\beta\alpha}\,(2\pi)^{-3/2}(2q^0)^{-1/2}\;\delta^{(4)}(P_\alpha - P_\beta - q) \tag{3.8}$$

is the amplitude for the transition $|\alpha\rangle \to |\beta\rangle$ emitting a Goldstone boson $|B, q\rangle$ (and not the current J^μ), see Fig. 3.2. Then

$$q_\mu\, \langle\beta|\, J^\mu(0)\, |\alpha\rangle\Big|_{\text{type III}} = i\,F\,M_{\beta\alpha}. \tag{3.9}$$

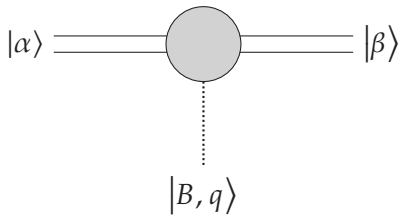


Figure 3.2. The amplitude $M_{\beta\alpha}$: the transition $|\alpha\rangle \to |\beta\rangle$ with the emission of a Goldstone boson $|B, q\rangle$.

3.3 Diagrams of type I: current on an external line

Let us now look at diagrams of type I, Fig. 3.3. The propagator adjacent to the current vertex is

$$\frac{1}{(p+q)^2 - m^2} = \frac{1}{p^2 - m^2 + q^2 + 2p\cdot q}, \tag{3.10}$$

where $p^2 = m^2$ (external particle) and $q^2 = 0$, so that

$$\frac{1}{(p+q)^2 - m^2} = \frac{1}{2p \cdot q}. \tag{3.11}$$

Since q is light-like, we can write

$$q^\mu = |\vec{q}|\,(1, \vec{n}) \equiv |\vec{q}|\, n^\mu, \qquad |\vec{n}| = 1. \tag{3.12}$$

Then

$$\frac{1}{(p+q)^2 - m^2} = \frac{1}{|\vec{q}|} \frac{1}{2p \cdot n}. \tag{3.13}$$

We can then write

$$q_\mu \left\langle \beta \right| J^\mu(0) \left| \alpha \right\rangle \Big|_{\text{type I}} = \left(|\vec{q}|\, n_\mu\right) \frac{1}{|\vec{q}|\,(2p \cdot n)} N^\mu_{\beta\alpha} = \frac{n_\mu N^\mu_{\beta\alpha}}{2p \cdot n}, \tag{3.14}$$

where $N^\mu_{\beta\alpha}$ denotes the remaining part of the diagram.

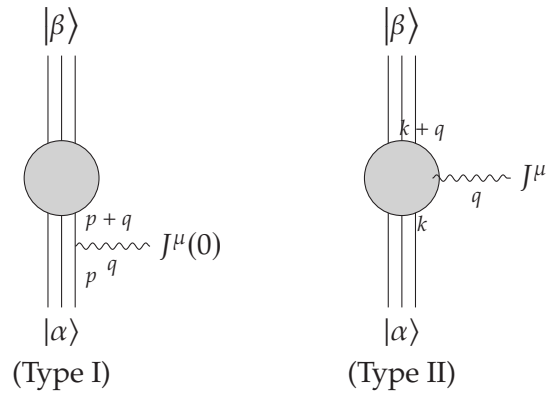


Figure 3.3. Momentum routing for the type I and type II diagrams. In type I the current with momentum q attaches to an external line of momentum p; in type II it attaches to an internal propagator of momentum k.

3.4 Diagrams of type II: current on an internal line

Now we look at the diagrams of type II, where the current attaches to an internal propagator, with momentum, say, k (Fig. 3.3, right). Since the propagator is internal, $k^2 \neq m^2$: it is off-shell. The second propagator attached to the current is

$$\frac{1}{(k+q)^2 - m^2} \tag{3.15}$$

and it remains finite as $q \to 0$. Therefore,

$$\lim_{q \to 0} \left\langle \beta \right| J^\mu(0) \left| \alpha \right\rangle \Big|_{\text{type II}} \neq \infty. \tag{3.16}$$

3.5 The soft Goldstone boson theorem

Now we take the limit $q^\mu \to 0$ of current conservation:

$$\begin{aligned} 0 = \lim_{q\to 0} q_\mu \left\langle \beta \right| J^\mu \left| \alpha \right\rangle &= \lim_{q\to 0} q_\mu \left\langle \beta \right| J^\mu \left| \alpha \right\rangle_{\text{type I}} \\ &\quad \underbrace{+ \lim_{q\to 0} q_\mu \left\langle \beta \right| J^\mu \left| \alpha \right\rangle_{\text{type II}}}_{=0} + \lim_{q\to 0} q_\mu \left\langle \beta \right| J^\mu \left| \alpha \right\rangle_{\text{type III}} \\ &= \frac{n_\mu N^\mu_{\beta\alpha}}{2p\cdot n} + i\,F\,M_{\beta\alpha}, \end{aligned} \tag{3.17}$$

which gives

$$\boxed{M_{\beta\alpha} = \frac{i}{F}\,\frac{n_\mu N^\mu_{\beta\alpha}}{2p\cdot n}}. \tag{3.18}$$

Remark. Eq. (3.18) shows a single external line. In general the type-I contribution is a sum over all external lines of $|\alpha\rangle$ and $\left|\beta\right\rangle$: with $q = P_\alpha - P_\beta$ flowing out through the current, an outgoing line gives $1/(2p\cdot q)$ as above, an incoming line $1/((p-q)^2 - m^2) = -1/(2p\cdot q)$. The limit is finite but depends on the direction n, and for $q \to 0$ the current vertex reduces to the matrix of the broken charge between (degenerate) external one-particle states. If no such pole terms exist, $M_{\beta\alpha} \to 0$ as $q \to 0$: the Adler zero (Weinberg, QFT Vol. II, Ch. 19; Adler 1965).

> The amplitude $M_{\beta\alpha}$ to emit a soft Goldstone boson ($q^\mu \to 0$) can be obtained from an amplitude where the Goldstone boson is replaced by the corresponding current, where the latter is attached *only* to external particles.

4

The $O(N)$ linear sigma model

4.1 An example of a spontaneously broken global symmetry

Consider the Lagrangian

$$\mathcal{L} = \frac{1}{2}\sum_n \left(\partial_\mu \phi_n\right)^2 - \frac{M^2}{2}\sum_n \phi_n^2 - \frac{g}{4}\left(\sum_n \phi_n^2\right)^2 \tag{4.1}$$

of N real scalar fields ϕ_n, $n = 1, \ldots, N$. The Lagrangian is symmetric under rotations

$$\phi_n \to \phi_n' = R_{nm}\phi_m, \qquad R^T R = 1, \tag{4.2}$$

and the symmetry group is $O(N)$.

The vacuum expectation values of the fields can be computed by minimising the effective potential,

$$\left.\frac{\partial V_{eff}}{\partial \phi_l}\right|_{\phi_l = \langle \phi_l \rangle} = 0. \tag{4.3}$$

Ignoring loop effects, the effective potential is equal to the classical potential,

$$V_{eff}(\phi) \approx \frac{M^2}{2}\sum_n \phi_n^2 + \frac{g}{4}\left(\sum_n \phi_n^2\right)^2, \tag{4.4}$$

and its extrema are solutions of

$$\langle \phi_l \rangle \left(M^2 + g\sum_n \langle \phi_n \rangle^2\right) = 0. \tag{4.5}$$

For $M^2, g > 0$, we find a minimum

$$\langle \phi_l \rangle = 0, \tag{4.6}$$

which is invariant under $O(N)$ rotations.

Spontaneous symmetry breaking occurs for $M^2 < 0$ and $g > 0$. The minima are found to satisfy

$$\boxed{\sum_n \langle \phi_n \rangle^2 = -\frac{M^2}{g}} \tag{4.7}$$

and they are degenerate. Only one is chosen (spontaneously), but it could be any of the solutions of the above equation.

4.2 The mass spectrum

Let us look at the spectrum of the theory, and specifically at the masses of the scalar particles. These are given by diagonalising the quadratic terms in the effective potential. The corresponding "mass matrix" is

$$\begin{aligned} D^{-1}_{nm}(0) &\approx \left.\frac{\partial^2 V}{\partial\phi_n \partial\phi_m}\right|_{\phi=\langle\phi\rangle} + (\text{loop corrections}) \\ &= \underbrace{M^2\delta_{nm} + g\,\delta_{nm}\sum_l \langle\phi_l\rangle^2}_{=0 \text{ by Eq. (4.7)}} + 2g\,\langle\phi_n\rangle\langle\phi_m\rangle \end{aligned} \tag{4.8}$$

or

$$M^2_{nm} = 2g\,\langle\phi_n\rangle\langle\phi_m\rangle\,, \tag{4.9}$$

or, in matrix form,

$$M^2_{nm} = 2g \begin{pmatrix} \langle\phi_1\rangle^2 & \langle\phi_1\rangle\langle\phi_2\rangle & \cdots & \langle\phi_1\rangle\langle\phi_N\rangle \\ \langle\phi_2\rangle\langle\phi_1\rangle & \langle\phi_2\rangle^2 & \cdots & \langle\phi_2\rangle\langle\phi_N\rangle \\ \vdots & & \ddots & \vdots \\ \langle\phi_N\rangle\langle\phi_1\rangle & \cdots & \cdots & \langle\phi_N\rangle^2 \end{pmatrix}. \tag{4.10}$$

To diagonalise it and find the mass eigenvalues, we solve

$$\det\left(M^2_{nm} - \mu^2\delta_{nm}\right) = 0 \;\rightsquigarrow\; \left(\mu^2\right)^{N-1}\left(\mu^2 - 2g\sum_i \langle\phi_i\rangle^2\right) = 0. \tag{4.11}$$

We find one massive particle with mass

$$\mu^2 = 2g\sum_i \langle\phi_i\rangle^2 = 2|M|^2 \qquad (\text{from Eq. (4.7)}), \tag{4.12}$$

and $(N-1)$ Goldstone boson particles with zero mass.

4.3 Counting broken generators

The original group $O(N)$ has

$$\frac{1}{2}N(N-1) \tag{4.13}$$

generators. The spontaneous breaking picks out one direction in the space of $\langle\phi_i\rangle$, $i = 1,\ldots,N$. The linear combinations of fields which are orthogonal to this direction can be freely rotated into each other: the vacuum has a remaining $O(N-1)$ symmetry. The number of generators which have been broken[1] is

$$\frac{1}{2}N(N-1) - \frac{1}{2}(N-1)(N-2) = N-1, \tag{4.14}$$

in agreement with the number of massless Goldstone bosons found above. We will get back to this soon.

[1] A broken symmetry generator transforms the vacuum state into a different one. An unbroken generator leaves the vacuum invariant.

4.4 Goldstone boson states and the sum rule $F \cdot Z = v$

Let us now go back to the second (spectral) proof of Goldstone's theorem, where we found that the Goldstone boson states $|B\rangle$ yield

$$\int d^3\vec{p}_B \, \langle 0|\, J^\mu(x)\,|B\rangle \langle B|\, \phi_n(y)\,|0\rangle \; \delta^{(3)}(\vec{p} - \vec{p}_B) = \frac{i}{(2\pi)^3} \rho_n(p^2)\, p^\mu. \tag{4.15}$$

From Lorentz invariance we infer that

$$\langle 0|\, J^\mu(x)\,|B\rangle = \frac{i\, p_B^\mu\, F\, e^{-ip_B\cdot x}}{(2\pi)^{3/2}\sqrt{2p_B^0}} \tag{4.16}$$

and

$$\langle B|\, \phi_n(x)\,|0\rangle = \frac{Z_n\, e^{+ip_B\cdot x}}{(2\pi)^{3/2}\sqrt{2p_B^0}}, \tag{4.17}$$

which yields

$$\rho_n(m^2) = F\, Z_n\, \delta(m^2). \tag{4.18}$$

We had also shown that

$$\left\langle \left[J^0(\vec{y},t), \phi_n(\vec{x},t)\right]\right\rangle = i\,\delta^{(3)}(\vec{x}-\vec{y}) \int \rho_n(m^2)\, dm^2, \tag{4.19}$$

and, integrating over $\vec{y}$,

$$\left\langle \underbrace{\left[Q, \phi_n(\vec{x},t)\right]}_{\delta\phi_n} \right\rangle = i\, F\, Z_n \;\rightsquigarrow\; \left\langle \delta\phi_n \right\rangle = i\, F\, Z_n. \tag{4.20}$$

In the case of multiple broken generators T^a, we have an equal number of currents $J^\mu_{(a)}$ and, therefore, of Goldstone bosons. This is because we can repeat the same proof for each one of the $J^\mu_{(a)}$ currents, and we have

$$\langle B_a|\, \phi_n(x)\,|0\rangle = \frac{Z_{an}\, e^{+ip_B\cdot x}}{(2\pi)^{3/2}\sqrt{2p_B^0}}, \tag{4.21}$$

$$\langle 0|\, J^\mu_a(x)\,|B_b\rangle = \frac{i\, F_{ab}\, p_B^\mu\, e^{-ip_B\cdot x}}{(2\pi)^{3/2}\sqrt{2p_B^0}}. \tag{4.22}$$

This yields the sum rule

$$i \sum_b F_{ab} Z_{bn} = -\sum_m T^a_{nm} \left\langle \phi_m \right\rangle, \tag{4.23}$$

where the variation of the field under the transformation generated by T^a is $i\epsilon\, \delta\phi_n^{(a)} = -i\epsilon \sum_m T^a_{nm}\phi_m$. *Remark.* These are the conventions of the whole part (see the Conventions section of Chapter 1): $\left[Q^a, \phi_n\right] = -T^a_{nm}\phi_m$, so that here $\delta\phi_n \equiv \left[Q, \phi_n\right]$, and Eq. (4.23) is Eq. (2.43) of Chapter 2, with real F and Z (Weinberg, QFT Vol. II, Sec. 19.2).

In our example of the $O(N)$-symmetric Lagrangian, we can choose a vacuum aligned with one of the axes in field space (Fig. 4.1),

$$\left\langle \phi_n \right\rangle = v\, \delta_{n1}. \tag{4.24}$$

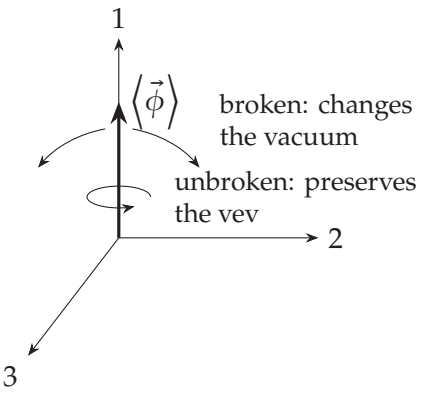


Figure 4.1. The vacuum $\langle\phi_n\rangle = v\,\delta_{n1}$ of the $O(N)$ model (shown for $N = 3$). Rotations which move the vev along the vacuum manifold are broken; rotations around the 1-axis preserve the vev and are unbroken.

Let us take the index "a" to correspond to a broken generator, $a \in \{2, \ldots, N\}$, corresponding to the rotations in the $(1, a)$ plane, which turn the vev from the 1-axis towards the a-axis. Then we have

$$i \sum_b F_{ab} Z_{bn} = -v\,(T^a)_{n1}\,. \tag{4.25}$$

Exercise: We can write the generators $(T^a)_{n1}$ as

$$T^a_{a1} = -T^a_{1a} = -i \quad \text{for } n = a, \qquad T^a_{n1} = -T^a_{1n} = 0 \quad \text{for } n \neq a. \tag{4.26}$$

Prove it for $O(3) \to O(2)$.

Then we have

$$\sum_b F_{ab} Z_{ba} = v, \qquad a = 2, \ldots, N. \tag{4.27}$$

Using the unbroken $O(N-1)$ symmetry one can show that

$$F_{ab} = \delta_{ab} F \qquad \text{(exercise)} \tag{4.28}$$

and $Z_{ab} = \delta_{ab} Z$, and we can choose $Z = 1$ by renormalising the field ϕ. We then have

$$\boxed{F = v.} \tag{4.29}$$

So $F \sim \langle 0|\, J^\mu_a\, |B_b\rangle$ is associated with the strength of spontaneous symmetry breaking.

4.5 Interactions of Goldstone bosons

What is the probability of producing a Goldstone boson in a certain interaction? To compute this, we consider a matrix element

$$\big\langle\beta\big|\, J^\mu(x)\,|\alpha\rangle\,, \tag{4.30}$$

in which the current $J^\mu(x)$ is inserted into a transition from an initial state $|\alpha\rangle$ to a final state $\big|\beta\big\rangle$ (Fig. 4.2, left).

A Goldstone boson has the same quantum numbers as J^μ and it therefore couples to it. If we were to write a Feynman diagram, it would be something like Fig. 4.2 (right),

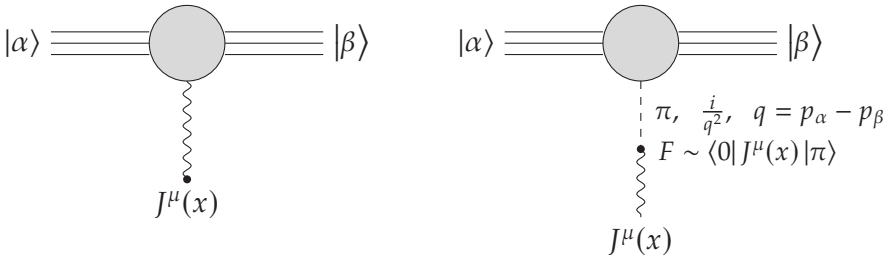


Figure 4.2. Left: insertion of the current $J^\mu(x)$ in the transition $|\alpha\rangle \to |\beta\rangle$. Right: the contribution in which the current creates a Goldstone boson π which is absorbed in the transition.

where the current creates a Goldstone boson π with strength $F \sim \langle 0|\, J^\mu(x)\, |\pi\rangle$, which then propagates, i/q^2, with $q \equiv p_\alpha - p_\beta$, and is absorbed in the transition.

We can trivially predict the dependence of this amplitude on x:

$$\langle\beta|\, J^\mu(x)\, |\alpha\rangle = e^{-i(p_\alpha - p_\beta)\cdot x}\, \langle\beta|\, J^\mu(0)\, |\alpha\rangle\,. \tag{4.31}$$

The matrix element $\langle\beta|\, J^\mu(0)\, |\alpha\rangle$ may involve unknown Feynman rules and receives contributions from more Feynman diagrams than the one drawn above. But, given that the Goldstone boson is a scalar, we can infer

$$\langle\beta|\, J^\mu(0)\, |\alpha\rangle = \frac{iF}{q^2}\, q^\mu\, M_{\beta\alpha} + N^\mu_{\beta\alpha}, \tag{4.32}$$

where $M_{\beta\alpha}$ is the transition amplitude from $|\alpha\rangle \to |\beta\rangle$ emitting a Goldstone boson. The second term collects the diagrams in which the current does not couple to a Goldstone boson propagator, so that it has no Goldstone pole at $q^2 = 0$. (It may have other singularities as $q \to 0$, of the form $1/(p\cdot q)$, as we shall explain.)

Multiplying by q_μ and taking $q^\mu \to 0$, current conservation, $\partial_\mu J^\mu = 0$, gives

$$0 = \langle\beta|\, q\cdot J(0)\, |\alpha\rangle = q\cdot N_{\beta\alpha} + iF\, M_{\beta\alpha} \tag{4.33}$$

$$\rightsquigarrow \qquad \boxed{M_{\beta\alpha} = \frac{i}{F}\, q\cdot N_{\beta\alpha}.} \tag{4.34}$$

$M_{\beta\alpha}$ is an amplitude to produce a Goldstone boson. $N_{\beta\alpha}$ is an amplitude without Goldstone boson vertices attached to the current $J^\mu(0)$.

Let us take the limit of a soft Goldstone boson,

$$q \to 0. \tag{4.35}$$

Then $M_{\beta\alpha}|_{q\to 0}$ is made out of the residues of the poles of $N_{\beta\alpha}$. These arise when the current is inserted on an external line of momentum p, with $p^2 = m^2$ (Fig. 4.3):

$$\frac{i}{(p+q)^2 - m^2} = \frac{i}{p^2 - m^2 + 2p\cdot q + q^2} \overset{p^2=m^2,\, q\to 0}{\longrightarrow} \frac{i}{2p\cdot q}. \tag{4.36}$$

Notice that only insertions in external legs can produce soft singularities.

So, the amplitude $M_{\beta\alpha}$ to emit a soft Goldstone boson in the transition $|\alpha\rangle \to |\beta\rangle$ is obtained from the amplitude for the same transition without the Goldstone boson, by inserting the current $J^\mu(0)$ on its external legs.

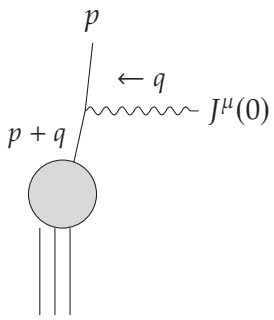


Figure 4.3. Insertion of the current on an external line of momentum p produces the soft singularity $i/(2p \cdot q)$.

The coupling constant associated with the emission of a Goldstone boson is

$$\frac{1}{F}, \tag{4.37}$$

the inverse of the strength of spontaneous symmetry breaking.

5

Approximate symmetries and pseudo-Goldstone bosons

5.1 Exact symmetries and Goldstone's theorem revisited

Consider a theory with an effective potential $V_0(\phi_n)$. Under a global symmetry transformation the effective potential remains invariant:

$$\phi_n \to \phi_n + i\epsilon^a T^a_{nm}\phi_m \;\rightsquigarrow\; V_0(\phi_n) = V_0\left(\phi_n + i\epsilon^a T^a_{nm}\phi_m\right) \;\rightsquigarrow\; \frac{\partial V_0}{\partial \phi_n}\left(T^a_{nm}\phi_m\right) = 0, \tag{5.1}$$

where T^a_{nm} are the generators of the symmetry group (summation over repeated indices is understood). Taking a second derivative, we have

$$\boxed{\frac{\partial^2 V_0}{\partial\phi_l \partial\phi_n}\left(T^a_{nm}\phi_m\right) + \frac{\partial V_0}{\partial\phi_n}T^a_{nl} = 0.} \tag{5.2}$$

For the physical system, when $\phi_m = \langle\phi_m\rangle$ is the physical vacuum expectation value, we have

$$\frac{\partial V_0}{\partial\langle\phi_n\rangle} = 0 \tag{5.3}$$

and thus

$$\boxed{\frac{\partial^2 V_0}{\partial\langle\phi_l\rangle\,\partial\langle\phi_n\rangle}T^a_{nm}\langle\phi_m\rangle = 0.} \tag{5.4}$$

The matrix $\partial^2 V_0/\partial\langle\phi_l\rangle\,\partial\langle\phi_n\rangle$ is the mass matrix of the theory, and the above equation shows that for spontaneously broken generators ($T^a_{nm}\langle\phi_m\rangle \neq 0$) there must be a massless particle in the spectrum of the theory (a Goldstone boson). This is the effective-action proof of Goldstone's theorem of Part III, Eq. (8.11).

5.2 Explicit breaking by a small perturbation

Let us now introduce a very small non-symmetric term in the effective potential,

$$V = V_0 + V_I, \tag{5.5}$$

with

$$\delta V_I \neq 0 \quad \text{under} \quad \phi_n \to \phi_n + i\epsilon^a T^a_{nm}\phi_m, \tag{5.6}$$

while $\delta V_0 = 0$ under the same transformation,

$$\frac{\partial V_0}{\partial \langle \phi_n \rangle} T^a_{nm} \langle \phi_m \rangle = 0. \tag{5.7}$$

The term $V_I(\phi)$ does not respect the symmetry and breaks it explicitly. Since V_I is relatively small, the symmetry, although not exact, is considered to hold approximately. We have studied the spontaneous breaking of exact symmetries earlier. Does spontaneous breaking also occur for an approximate symmetry? In other words, should we expect that the ground state varies by applying a symmetry transformation? This is indeed the case. In contrast to the spontaneous breaking of an exact global symmetry, we will find, however, that the vacuum state which is realised is constrained by the symmetry-breaking term $V_I(\phi)$ in the potential.

5.3 The vacuum alignment condition

Let us split

$$\langle \phi_n \rangle = \langle \phi^0_n \rangle + \langle \phi^I_n \rangle , \tag{5.8}$$

where $\langle \phi^0_n \rangle$ is the vev of the theory without the perturbation (exact symmetry, $V_I = 0$). We have that

$$\frac{\partial V_0}{\partial \langle \phi^0_n \rangle} = 0. \tag{5.9}$$

For the full system the vacuum is found by solving

$$\frac{\partial V}{\partial \langle \phi_n \rangle} = 0, \tag{5.10}$$

and, expanding around the vacuum of the system with the exact symmetry,

$$\underbrace{\frac{\partial V_0}{\partial \langle \phi^0_n \rangle}}_{0^{\text{th}} \text{ order} = 0} + \underbrace{\frac{\partial V_I}{\partial \langle \phi^0_n \rangle} + \frac{\partial^2 V_0}{\partial \langle \phi^0_n \rangle \partial \langle \phi^0_l \rangle} \langle \phi^I_l \rangle}_{1^{\text{st}} \text{ order in the } \langle \phi^I \rangle \text{ expansion}} + \ldots = 0. \tag{5.11}$$

Multiplying by $T^a_{nm} \langle \phi^0_m \rangle$ and summing over n,

$$\underbrace{\frac{\partial^2 V_0}{\partial \langle \phi^0_n \rangle \partial \langle \phi^0_l \rangle} T^a_{nm} \langle \phi^0_m \rangle}_{=0 \text{ (Goldstone's theorem)}} \langle \phi^I_l \rangle + \frac{\partial V_I}{\partial \langle \phi^0_n \rangle} T^a_{nm} \langle \phi^0_m \rangle + \ldots = 0. \tag{5.12}$$

We therefore have

$$\boxed{\sum_{nm} \frac{\partial V_I}{\partial \langle \phi^0_n \rangle} T^a_{nm} \langle \phi^0_m \rangle = 0.} \tag{5.13}$$

This is the "vacuum alignment condition". The symmetry breaking of the vacuum ($T^a_{nm} \langle \phi^0_m \rangle$) is not chosen randomly: it is aligned with the symmetry-breaking term V_I, in the sense that the gradient of V_I is orthogonal to every direction $T^a \langle \phi^0 \rangle$ along the vacuum orbit,

$$\left(\frac{\partial V_I}{\partial \langle \phi^0_n \rangle} \right) \perp \left(T^a_{nm} \langle \phi^0_m \rangle \right) \quad \text{for all } a. \tag{5.14}$$

The vacuum state $|0\rangle$ is forced into an alignment with the symmetry-breaking term V_I.

5.4 Existence of an aligned vacuum

We can prove that the vacuum alignment condition has a "solution", i.e. that there is a vacuum state which satisfies it.

Proof: $V_0(\phi)$ is invariant under group transformations

$$\phi \to L(\theta)\,\phi, \tag{5.15}$$

where θ is the parameter (assume one for simplicity) of the group. If $\phi = \phi^*$ corresponds to a minimum, then also $L(\theta)\phi^*$ corresponds to a minimum:

$$V_0\left(L(\theta)\phi^*\right) = V_0(\phi^*) = \text{minimum}. \tag{5.16}$$

Consider the perturbation

$$V_I\left(L(\theta)\phi^*\right) \tag{5.17}$$

as a function of the group parameters θ^a. For compact groups, $L(\theta)\phi^*$ spans a compact manifold and, if V_I is a continuous function, then $V_I(L(\theta)\phi^*)$ must have a minimum (Fig. 5.1). At the minimum,

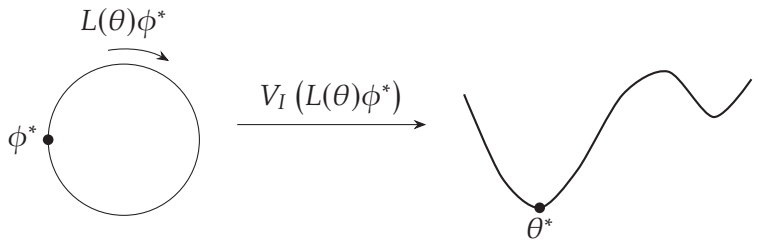


Figure 5.1. The orbit $L(\theta)\phi^*$ of degenerate minima of V_0 (left) is mapped by the perturbation to the function $V_I(L(\theta)\phi^*)$ of the group parameters (right); on a compact orbit it has a minimum at θ^*.

$$0 = \frac{\partial V_I\left(L(\theta)\phi^*\right)}{\partial\theta^a} = \sum_n \left.\frac{\partial V_I(\phi)}{\partial\phi_n}\right|_{\phi=L(\theta)\phi^*} \frac{\partial\left(L(\theta)\phi^*\right)_n}{\partial\theta^a}. \tag{5.18}$$

Recall that

$$L(\theta) = e^{iT^a\theta^a}, \qquad T^a : \text{generators}, \tag{5.19}$$

$$\rightsquigarrow \quad \frac{\partial L_{nm}}{\partial\theta^a} = iT^a_{nk}\left(L(\theta)\right)_{km}. \tag{5.20}$$

Then:

$$\boxed{\left.\frac{\partial V_I}{\partial\phi_n}\right|_{\phi=L(\theta^*)\phi^*} T^a_{nm}\;\phi_m\big|_{\phi_m=(L(\theta^*)\phi^*)_m} = 0.} \tag{5.21}$$

This is the vacuum alignment condition, Eq. (5.13); we have proven that it is satisfied for

$$\left\langle\phi_n\right\rangle = L(\theta^*)_{nm}\left\langle\phi^*_m\right\rangle, \tag{5.22}$$

where θ^* minimises $V_I(L(\theta)\phi^*)$.

5.5 Finding the vacuum

Therefore, in order to find the vacuum which is chosen, among the many, in the spontaneous breaking of an approximate symmetry, we need to:

α) Find the vevs which minimise the "exact-symmetry" potential $V_0(\phi)$. These are degenerate,

$$L(\theta)\left\langle\phi^*\right\rangle \;:\; 0 = \left.\frac{\partial V_0}{\partial\phi}\right|_{\phi=L(\theta)\left\langle\phi^*\right\rangle}, \tag{5.23}$$

and parametrised by the group parameters.

β) Choose the value θ^* of the group parameters which minimises also the symmetry-breaking perturbation,

$$\theta^* :\; V_I\left(L(\theta^*)\left\langle\phi^*\right\rangle\right) \text{ is minimum.} \tag{5.24}$$

The true vacuum state is the one with a field vev

$$\left\langle\phi\right\rangle = L(\theta^*)\left\langle\phi^*\right\rangle. \tag{5.25}$$

The recipe is very intuitive. Among the degenerate vacuum states of the exact-symmetry theory, we select the one which minimises the perturbation of the potential.

5.6 Pseudo-Goldstone bosons

5.6.1 Goldstone fields

For exact symmetries we have found that for each broken generator T^a there is a massless particle state $|B_a\rangle$. These states can be created from field operators acting on the vacuum,

$$\langle B_a|\,\phi_n(x)\,|0\rangle = \frac{Z_{an}\,e^{ip_B\cdot x}}{(2\pi)^{3/2}\sqrt{2p_B^0}}, \tag{5.26}$$

and have the quantum numbers of the conserved currents corresponding to the broken generators,

$$\langle 0|\,J_a^\mu(x)\,|B_b\rangle = \frac{iF_{ab}\,p_B^\mu\,e^{-ip_B\cdot x}}{(2\pi)^{3/2}\sqrt{2p_B^0}}. \tag{5.27}$$

We have proven (Chapter 4, Eq. (4.23)) that

$$\boxed{i\sum_b F_{ab}Z_{bn} = -\sum_m T^a_{nm}\left\langle\phi_m\right\rangle} \qquad (\text{"}F\cdot Z = v\text{"}). \tag{5.28}$$

The Goldstone fields Π_a are linear combinations of the fields of the Lagrangian which generate exactly the Goldstone states:

$$\langle B_a|\,\Pi_b(x)\,|0\rangle = \frac{e^{+ip_B\cdot x}\,\delta_{ab}}{(2\pi)^{3/2}\sqrt{2p_B^0}}. \tag{5.29}$$

We can then rewrite

$$\phi_n(x) = \sum_a C_{na}\Pi_a(x) + \text{non-Goldstone fields} \tag{5.30}$$

$$\rightsquigarrow \quad \langle B_a|\,\phi_n(x)\,|0\rangle = \sum_b C_{nb}\,\langle B_a|\,\Pi_b(x)\,|0\rangle + 0 \quad \Rightarrow \quad Z_{an} = \sum_b C_{nb}\,\delta_{ab} \tag{5.31}$$

$$\Rightarrow \quad \boxed{C_{na} = Z_{an}}. \tag{5.32}$$

We therefore obtain

$$\langle\phi_n\rangle \equiv \langle 0|\,\phi_n(x)\,|0\rangle = \sum_a Z_{an}\,\langle\Pi_a\rangle + \text{vevs of non-Goldstone fields.} \tag{5.33}$$

5.6.2 The mass matrix

To compute the mass matrix of the physical Goldstone bosons, we have

$$M^2_{ab} = \frac{\partial^2 V_{eff}}{\partial\langle\Pi_a\rangle\,\partial\langle\Pi_b\rangle} = \sum_{nm}\frac{\partial\langle\phi_n\rangle}{\partial\langle\Pi_a\rangle}\frac{\partial\langle\phi_m\rangle}{\partial\langle\Pi_b\rangle}\frac{\partial^2 V_{eff}}{\partial\langle\phi_n\rangle\,\partial\langle\phi_m\rangle} \tag{5.34}$$

$$\rightsquigarrow \quad M^2_{ab} = \sum_{nm} Z_{an}Z_{bm}\frac{\partial^2 V_{eff}}{\partial\langle\phi_n\rangle\,\partial\langle\phi_m\rangle}. \tag{5.35}$$

For an exact symmetry which is broken spontaneously, all Goldstone bosons are massless and thus

$$M^2_{ab} = 0 \qquad \text{(exact symmetry)}. \tag{5.36}$$

In the case of an approximate symmetry we write

$$M^2_{ab} = \sum_{nm} Z_{an}Z_{bm}\left.\frac{\partial^2 V_{eff}}{\partial\phi_n\partial\phi_m}\right|_{\phi=\langle\phi^0\rangle+\langle\phi^I\rangle}. \tag{5.37}$$

Expanding around $\langle\phi^I\rangle = 0$, using that $M^2_{ab} = 0$ at leading order, and that, from Eq. (5.28),

$$Z_{an} = i\sum_b F^{-1}_{ab}\,\delta^b\,\langle\phi^0_n\rangle\,, \qquad \delta^b\,\langle\phi_n\rangle = \sum_m T^b_{nm}\,\langle\phi^0_m\rangle\,, \tag{5.38}$$

we obtain

$$M^2_{cd} = -\sum_{ab} F^{-1}_{ca}F^{-1}_{db}\left\{\delta^a\,\langle\phi^0_n\rangle\;\delta^b\,\langle\phi^0_m\rangle\;\frac{\partial^2 V_I(\langle\phi\rangle)}{\partial\langle\phi^0_n\rangle\,\partial\langle\phi^0_m\rangle} + \left(T^a\delta^b\,\langle\phi^0\rangle\right)_n\frac{\partial V_I(\langle\phi\rangle)}{\partial\langle\phi_n\rangle}\right\}. \tag{5.39}$$

Remark. The step to Eq. (5.39) goes as follows. Differentiating the invariance identity Eq. (5.2) once more and contracting with $\delta^b\,\langle\phi^0\rangle$ and $\langle\phi^I\rangle$ gives $V_0'''\,\delta^a\,\langle\phi^0\rangle\;\delta^b\,\langle\phi^0\rangle\;\langle\phi^I\rangle = -(T^a\delta^b\,\langle\phi^0\rangle)_n\,(\partial^2 V_0\;\langle\phi^I\rangle)_n = (T^a\delta^b\,\langle\phi^0\rangle)_n\,\partial V_I/\partial\langle\phi_n\rangle$, using Eq. (5.4) and the first-order vacuum condition $\partial^2 V_0\;\langle\phi^I\rangle = -\partial V_I$. The antisymmetric part in $a \leftrightarrow b$ vanishes by the alignment condition (Weinberg, QFT Vol. II, Sec. 19.3). We can prove that this is positive semi-definite.

5.6.3 Positivity of the mass matrix

To tie up the loose ends of the discussion of the spontaneous breaking of approximate symmetries, we show that:

1. We can find a solution of the vacuum alignment condition (Sec. 5.4).
2. The Goldstone-boson mass matrix is positive semi-definite.

For the proof of the second statement we use that

$$\left\langle \phi^0 \right\rangle = L(\theta^*) \left\langle \phi^* \right\rangle . \tag{5.40}$$

We then obtain that

$$\boxed{M^2_{ab} = \sum_{cd} F^{-1}_{ac} F^{-1}_{bd} \left. \frac{\partial^2 V_I \left(L(\theta) \left\langle \phi^0 \right\rangle \right)}{\partial \theta_c \partial \theta_d} \right|_{\theta=0}} . \tag{5.41}$$

Given that $V_I(L(\theta)\left\langle \phi^* \right\rangle)$ is a minimum of the perturbation at $\theta = \theta^*$, M^2_{ab} is positive semi-definite (positive definite if V_I lifts the vacuum degeneracy in every broken direction). We shall see examples of this.

The Goldstone bosons of a spontaneously broken approximate symmetry are massive. For this reason we call them "pseudo-Goldstone bosons". The pions of QCD are pseudo-Goldstone bosons. The same may be true for the "Higgs boson". The pseudo-Goldstone mechanism is also used to address the hierarchy problem in electroweak theory. *Remark.* The 125 GeV boson discovered at the LHC in 2012 has couplings consistent with the elementary SM Higgs at the 5–10% level; a pseudo-Goldstone (composite) Higgs, e.g. from $SO(5) \to SO(4)$ with $v^2/f^2 \lesssim 0.1$, remains viable but requires some tuning, and no top partners have been found at the LHC so far (see the PDG review "Status of Higgs boson physics").

6

Chiral symmetry of QCD

6.1 The $SU(2) \otimes SU(2)$ symmetry of the light quarks

Consider the part of the QCD Lagrangian which involves the two lightest quarks, u and d, ignoring their masses for the moment:

$$\mathcal{L} = \bar{u}\, i\partial\!\!\!/\, u + \bar{d}\, i\partial\!\!\!/\, d + \ldots = (\bar{u}, \bar{d}) \begin{pmatrix} i\partial\!\!\!/ & 0 \\ 0 & i\partial\!\!\!/ \end{pmatrix} \begin{pmatrix} u \\ d \end{pmatrix} + \ldots \tag{6.1}$$

This Lagrangian has the symmetry

$$\begin{pmatrix} u \\ d \end{pmatrix} \to e^{i\left[\vec{\theta}_V \cdot \vec{\tau} + \gamma_5 \vec{\theta}_A \cdot \vec{\tau}\right]} \begin{pmatrix} u \\ d \end{pmatrix}, \tag{6.2}$$

with

$$\vec{\tau} \cdot \vec{\theta} = \frac{1}{2} \begin{pmatrix} \theta_3 & \theta_1 - i\theta_2 \\ \theta_1 + i\theta_2 & -\theta_3 \end{pmatrix}, \qquad \vec{\tau} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}, \qquad \vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}, \tag{6.3}$$

where $\tau_i = \sigma_i/2$, with σ_i the Pauli matrices.

6.1.1 Proof of the symmetry

Consider first the transformation without γ_5,

$$\begin{pmatrix} u \\ d \end{pmatrix} \to e^{i\vec{\theta}\cdot\vec{\tau}} \begin{pmatrix} u \\ d \end{pmatrix} \quad \Rightarrow \quad (\bar{u}, \bar{d}) \to (\bar{u}, \bar{d})\, e^{-i\vec{\theta}\cdot\vec{\tau}^\dagger}. \tag{6.4}$$

Then

$$(\bar{u}, \bar{d})\, (i\partial\!\!\!/) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ d \end{pmatrix} \to (\bar{u}, \bar{d})\, (i\partial\!\!\!/) \underbrace{e^{-i\vec{\theta}\cdot\vec{\tau}^\dagger}\, \mathbf{1}_{2\times 2}\, e^{i\vec{\theta}\cdot\vec{\tau}}}_{\mathbf{1}} \begin{pmatrix} u \\ d \end{pmatrix}. \tag{6.5}$$

For the transformation with γ_5 we expand the exponential,

$$e^{i\gamma_5 \vec{\theta}_A \cdot \vec{\tau}} = \mathbf{1} + i\gamma_5\, \vec{\theta}_A \cdot \vec{\tau} + i^2 \gamma_5^2 \frac{\left(\vec{\theta}_A \cdot \vec{\tau}\right)^2}{2!} + \ldots \tag{6.6}$$

In the Dirac representation

$$\gamma_5^2 = \begin{pmatrix} 0 & \mathbf{1}_{2\times 2} \\ \mathbf{1}_{2\times 2} & 0 \end{pmatrix} \begin{pmatrix} 0 & \mathbf{1}_{2\times 2} \\ \mathbf{1}_{2\times 2} & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{2\times 2} & 0 \\ 0 & \mathbf{1}_{2\times 2} \end{pmatrix} = \mathbf{1}_{4\times 4}, \tag{6.7}$$

and therefore

$$e^{i\gamma_5\vec{\theta}_A\cdot\vec{\tau}} = \cos\left(\vec{\theta}_A\cdot\vec{\tau}\right) + i\gamma_5\sin\left(\vec{\theta}_A\cdot\vec{\tau}\right). \tag{6.8}$$

Abbreviating $c_\theta \equiv \cos(\vec{\theta}_A\cdot\vec{\tau})$ and $s_\theta \equiv \sin(\vec{\theta}_A\cdot\vec{\tau})$ we have

$$\begin{aligned}
\begin{pmatrix} u \\ d \end{pmatrix} &\to (c_\theta + i\gamma_5 s_\theta)\begin{pmatrix} u \\ d \end{pmatrix} \\
\Rightarrow \left(u^\dagger, d^\dagger\right) &\to \left(u^\dagger, d^\dagger\right)\left(c_\theta^\dagger - i\gamma_5 s_\theta^\dagger\right) \\
\Rightarrow \left(u^\dagger\gamma^0, d^\dagger\gamma^0\right) &\to \left(u^\dagger, d^\dagger\right)\left(c_\theta^\dagger - i\gamma_5 s_\theta^\dagger\right)\gamma^0 = \left(u^\dagger, d^\dagger\right)\gamma^0\left(c_\theta^\dagger + i\gamma_5 s_\theta^\dagger\right) \\
\Rightarrow (\bar{u}, \bar{d}) &\to (\bar{u}, \bar{d})\left(c_\theta^\dagger + i\gamma_5 s_\theta^\dagger\right) \\
\Rightarrow (\bar{u}, \bar{d})\,\gamma^\mu &\to (\bar{u}, \bar{d})\,\gamma^\mu\left(c_\theta^\dagger - i\gamma_5 s_\theta^\dagger\right) = (\bar{u}, \bar{d})\,\gamma^\mu e^{-i\gamma_5\vec{\theta}_A\cdot\vec{\tau}^\dagger},
\end{aligned} \tag{6.9}$$

where we used that γ_5 anticommutes with γ^0 and γ^μ. Since $\vec{\tau}^\dagger = \vec{\tau}$, this shows that

$$(\bar{u}, \bar{d})\,(i\partial\!\!\!/)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u \\ d \end{pmatrix} \tag{6.10}$$

remains invariant.

6.2 Generators of the symmetry and Lie algebra

We have the following generators:

$$\vec{\tau} = (\tau_1, \tau_2, \tau_3) \quad \text{and} \quad \vec{x} = \gamma_5\vec{\tau}. \tag{6.11}$$

They satisfy the algebra

$$\boxed{\left[\tau_i, \tau_j\right] = i\epsilon_{ijk}\tau_k} \qquad \text{(SU(2) subgroup)}, \tag{6.12}$$

$$\left[\tau_i, x_j\right] = \left[\tau_i, \gamma_5\tau_j\right] = i\epsilon_{ijk}\gamma_5\tau_k \quad \Rightarrow \quad \boxed{\left[\tau_i, x_j\right] = i\epsilon_{ijk}x_k}, \tag{6.13}$$

and

$$\left[x_i, x_j\right] = \gamma_5^2\left[\tau_i, \tau_j\right] \quad \Rightarrow \quad \boxed{\left[x_i, x_j\right] = i\epsilon_{ijk}\tau_k}. \tag{6.14}$$

Now consider the linear combinations

$$\vec{\tau}_R = \frac{1}{2}(\mathbf{1} + \gamma_5)\,\vec{\tau} = \frac{\vec{\tau} + \vec{x}}{2}, \qquad \vec{\tau}_L = \frac{1}{2}(\mathbf{1} - \gamma_5)\,\vec{\tau} = \frac{\vec{\tau} - \vec{x}}{2}. \tag{6.15}$$

We have

$$P_R \equiv \frac{1}{2}(1 + \gamma_5), \qquad P_L \equiv \frac{1}{2}(1 - \gamma_5) \tag{6.16}$$

being projectors:

$$P_R^2 = \frac{1}{4}\left(1 + \gamma_5^2 + 2\gamma_5\right) = \frac{1}{2}(1 + \gamma_5) = P_R, \qquad P_L^2 = P_L, \tag{6.17}$$

and

$$P_R P_L = \frac{1 - \gamma_5^2}{4} = 0. \tag{6.18}$$

Thus

$$\left[\tau_{R,i},\tau_{R,j}\right]=P_R^2\left[\tau_i,\tau_j\right]\quad\Rightarrow\quad\boxed{\left[\tau_{R,i},\tau_{R,j}\right]=i\epsilon_{ijk}\tau_{R,k}}\qquad\text{(SU(2) subgroup).}\tag{6.19}$$

Similarly,

$$\left[\tau_{L,i},\tau_{L,j}\right]=i\epsilon_{ijk}\tau_{L,k}\qquad\text{(SU(2) subgroup),}\tag{6.20}$$

and

$$\left[\tau_{R,i},\tau_{L,j}\right]=P_RP_L\left[\tau_i,\tau_j\right]=0,\tag{6.21}$$

i.e. the two subgroups do not talk to each other. Therefore, the symmetry group of

$$(\bar{u},\bar{d})\,i\partial\!\!\!/\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}u\\d\end{pmatrix}\tag{6.22}$$

is $SU(2)\otimes SU(2)$.

6.3 Conserved currents

Let us compress our notation:

$$q\equiv\begin{pmatrix}u\\d\end{pmatrix},\tag{6.23}$$

in which case the Lagrangian becomes

$$\mathcal{L}=\bar{q}\,i\partial\!\!\!/\,q+\ldots\tag{6.24}$$

and it is invariant under

$$q\to e^{i\left[\vec{\theta}_V+\gamma_5\vec{\theta}_A\right]\cdot\vec{\tau}}q,\tag{6.25}$$

or, infinitesimally,

$$q\to q+\delta q,\qquad\delta q=i\left[\vec{\theta}_V+\gamma_5\vec{\theta}_A\right]\cdot\vec{\tau}\,q.\tag{6.26}$$

Also, $q^\dagger\to q^\dagger+\delta q^\dagger$ with

$$\begin{aligned}\delta q^\dagger&=-iq^\dagger\left[\vec{\theta}_V+\gamma_5\vec{\theta}_A\right]\cdot\vec{\tau}^\dagger,\\\delta\bar{q}&=-iq^\dagger\gamma^0\left[\vec{\theta}_V-\gamma_5\vec{\theta}_A\right]\cdot\vec{\tau}^\dagger=-i\bar{q}\left[\vec{\theta}_V-\gamma_5\vec{\theta}_A\right]\cdot\vec{\tau}^\dagger,\\\Rightarrow\quad\delta\bar{q}\,\gamma^\mu&=\bar{q}\gamma^\mu\left(-i\left[\vec{\theta}_V+\gamma_5\vec{\theta}_A\right]\cdot\vec{\tau}^\dagger\right).\end{aligned}\tag{6.27}$$

The Lagrangian changes as

$$\begin{aligned}\delta\mathcal{L}&=\mathcal{L}(q+\delta q)-\mathcal{L}(q)=\left(\bar{q}+\delta\bar{q}\right)i\partial\!\!\!/\left(q+\delta q\right)-\bar{q}\,i\partial\!\!\!/\,q\\&=\delta\bar{q}\,i\partial\!\!\!/\,q+\bar{q}\,i\partial\!\!\!/\,\delta q\\&=\bar{q}\,i\partial\!\!\!/\left[-i\left(\vec{\theta}_V+\gamma_5\vec{\theta}_A\right)\cdot\vec{\tau}^\dagger\right]q+\bar{q}\,i\partial\!\!\!/\left[i\left(\vec{\theta}_V+\gamma_5\vec{\theta}_A\right)\cdot\vec{\tau}\right]q=0,\end{aligned}\tag{6.28}$$

since the τ_i are hermitian, e.g.

$$\tau_2^\dagger=\frac{1}{2}\begin{pmatrix}0&-i\\i&0\end{pmatrix}^\dagger=\tau_2.\tag{6.29}$$

That is, the Lagrangian is invariant (for a symmetry it would suffice that it changes by a total derivative). By Noether's theorem,

$$J^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu q)}\,\delta q \tag{6.30}$$

is a conserved current:

$$J^\mu = \frac{\partial\left(\bar{q}\, i\gamma^\nu \partial_\nu q\right)}{\partial(\partial_\mu q)}\,\delta q \quad \Rightarrow \quad J^\mu = -\,\bar{q}\gamma^\mu\left[\vec{\theta}_V + \gamma_5\vec{\theta}_A\right]\cdot\vec{\tau}\, q. \tag{6.31}$$

We therefore have two conserved currents:

(i) Vector:

$$\vec{T}^\mu = \bar{q}\gamma^\mu\vec{\tau}q, \tag{6.32}$$

(ii) Axial-vector:

$$\vec{X}^\mu = \bar{q}\gamma^\mu\gamma_5\vec{\tau}q. \tag{6.33}$$

The charges are

$$\vec{T} \equiv \int d^3\vec{x}\;\vec{T}^0(x), \qquad \vec{X} \equiv \int d^3\vec{x}\;\vec{X}^0(x). \tag{6.34}$$

Quantisation of the Dirac field gives (setting the QCD coupling to zero)

$$q_i(x) = \int \frac{d^3\vec{p}}{(2\pi)^3\sqrt{2p^0}} \sum_s \left[a_s^i(p)\, u_s(p)\, e^{-ip\cdot x} + b_s^{i\,\dagger}(p)\, v_s(p)\, e^{ip\cdot x}\right], \qquad q_1 \equiv u, \quad q_2 \equiv d, \tag{6.35}$$

where a_s^i and b_s^i are annihilation operators.

Exercise:. Calculate the charges $\vec{T}$ and $\vec{X}$ in terms of creation and annihilation operators.

Exercise:. Prove that

$$\left[T_i, T_j\right] = i\epsilon_{ijk}T_k, \qquad \left[T_i, X_j\right] = i\epsilon_{ijk}X_k, \qquad \left[X_i, X_j\right] = i\epsilon_{ijk}T_k. \tag{6.36}$$

Exercise:. Prove that

$$\left[\vec{T}, q\right] = -\vec{\tau}\, q, \qquad \left[\vec{X}, q\right] = -\vec{x}\, q = -\gamma_5\vec{\tau}\, q, \tag{6.37}$$

i.e. the conserved charges generate the $SU(2)\otimes SU(2)$ transformations of the quark fields.

6.4 Explicit breaking of the symmetry

This $SU(2)\otimes SU(2)$ symmetry is not exact. It is broken explicitly by the quark mass terms, e.g.

$$m_u\bar{u}u. \tag{6.38}$$

Writing

$$u = (P_L + P_R)\,u = (P_L u) + (P_R u) \equiv u_L + u_R, \tag{6.39}$$

we have

$$\bar{u}_L = (P_L u)^\dagger\gamma^0 = u^\dagger P_L^\dagger\gamma^0 = u^\dagger\gamma^0 P_R = \bar{u}P_R, \tag{6.40}$$

and

$$\begin{aligned} m\bar{u}u &= m\left(\bar{u}_L + \bar{u}_R\right)\left(u_L + u_R\right) = m\left(\bar{u}_L u_L + \bar{u}_R u_R + \bar{u}_L u_R + \bar{u}_R u_L\right) \\ &= m\left(\bar{u}_L u_R + \bar{u}_R u_L\right). \end{aligned} \tag{6.41}$$

Under the symmetry transformation

$$q \to \left(1 + i\left(\vec{\theta}_V + \vec{\theta}_A\gamma_5\right)\cdot\vec{\tau}\right) q \tag{6.42}$$

the left-handed fields transform as

$$\begin{aligned} q_L &\to q_L + P_L\, i\left(\vec{\theta}_V(P_L + P_R) + \vec{\theta}_A(P_R - P_L)\right)\cdot\vec{\tau}\, q \\ &= q_L + i\left(\vec{\theta}_V P_L - \vec{\theta}_A P_L\right)\cdot\vec{\tau}\, P_L q \\ \Rightarrow\quad q_L &\to q_L + i\left(\vec{\theta}_V - \vec{\theta}_A\right)\cdot\vec{\tau}_L\, q_L, \end{aligned} \tag{6.43}$$

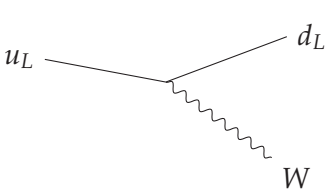
and

$$\boxed{q_L \to e^{i\vec{\theta}_L\cdot\vec{\tau}_L} q_L} \qquad \text{and} \qquad \boxed{q_R \to e^{i\vec{\theta}_R\cdot\vec{\tau}_R} q_R}. \tag{6.44}$$

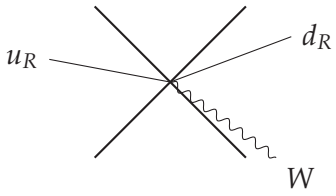
Left and right fields transform independently, with $\vec{\theta}_L = \vec{\theta}_V - \vec{\theta}_A$ and $\vec{\theta}_R = \vec{\theta}_V + \vec{\theta}_A$. The $\bar{u}_L u_R$ term is therefore not invariant. However, the masses of the light quarks are small,

$$m_u, m_d \ll M_{\text{proton}}. \tag{6.45}$$

A second small breaking is due to weak interactions, where left-handed and right-handed fields have different interactions. For example, the vertex



exists,



does not exist.

6.5 Spontaneous breaking and the pions

We have excellent evidence that this $SU(2)_L \otimes SU(2)_R$ is broken spontaneously. If the symmetry were unbroken, the hadron states

$$|h\rangle \quad \text{and} \quad \hat{X}\,|h\rangle \tag{6.46}$$

would be degenerate, i.e. have similar masses. Assume that $|h\rangle$ has parity +1,

$$\hat{\Pi}\,|h\rangle = |h\rangle\,. \tag{6.47}$$

Then

$$\begin{aligned} \hat{\Pi}\hat{X}\,|h\rangle &\propto \hat{\Pi}\int d^3\vec{x}\,\left(\bar{q}\gamma^0\gamma_5 q\right)\hat{\Pi}^{-1}\,\hat{\Pi}\,|h\rangle = \int d^3\vec{x}\,\underbrace{\hat{\Pi}\left(\bar{q}\gamma^0\gamma_5 q\right)\hat{\Pi}^{-1}}_{-\bar{q}\gamma^0\gamma_5 q}\,|h\rangle \\ &= -\hat{X}\,|h\rangle\,. \end{aligned} \tag{6.48}$$

Hence $|h\rangle$ and $\hat{X}\,|h\rangle$ have opposite parity. All other quantum numbers must be the same, e.g. spin and baryon number. No such parity doublets are seen in the spectrum of known hadrons.

Let us assume that

$$SU(2)_L \otimes SU(2)_R \longrightarrow SU(2)_{\text{vector}}. \tag{6.49}$$

Where are the (pseudo-)Goldstone bosons corresponding to the broken generators of $\hat{X}$? They are the pions:

$$\begin{array}{ccc} \pi^0, & \pi^+, & \pi^- \\ \downarrow & & \downarrow \\ \gamma_5\tau^3 & \gamma_5\tau^\pm \equiv \gamma_5\dfrac{\tau_1 \pm i\tau_2}{2} & \end{array} \tag{6.50}$$

with quark content

$$\pi^+\,(u\bar{d}), \qquad \pi^-\,(d\bar{u}), \qquad \pi^0\left(\frac{u\bar{u}-d\bar{d}}{\sqrt{2}}\right). \tag{6.51}$$

The pions have

- negative parity;
- zero spin;
- zero baryon number (baryon number is the common phase transformation of u and d, $q \longrightarrow e^{i\theta_{\text{baryon}}}q$).

6.6 Weak interactions of pions

At low energies, the weak interactions of the light quarks are described by

$$\mathcal{L}_{\text{weak}} = 2\,G_{wk}\left(T_+^\mu - X_+^\mu\right)\left(\sum_{\text{leptons}} \bar{\ell}\gamma_\mu P_L \nu_\ell\right) + \text{h.c.}, \qquad G_{wk} = G_F\cos\theta_C, \tag{6.52}$$

with

$$T_\pm^\mu = \frac{T_1^\mu \pm iT_2^\mu}{\sqrt{2}}, \qquad X_\pm^\mu = \frac{X_1^\mu \pm iX_2^\mu}{\sqrt{2}}. \tag{6.53}$$

This Lagrangian reflects the fact that hadrons can decay into leptons:

$$u \longrightarrow d + e^+ + \nu_e, \qquad u \longrightarrow d + \mu^+ + \nu_\mu, \tag{6.54}$$

for example

$$\pi^+ \longrightarrow \pi^0 + e^+ + \nu_e, \qquad n \longrightarrow p + e^- + \bar{\nu}_e, \qquad {}^{14}\text{O} \longrightarrow {}^{14}\text{N}^* + e^+ + \nu_e. \tag{6.55}$$

6.6.1 Transitions between spin-zero states

For transitions between zero-spin states such as $\pi^+ \longrightarrow \pi^0 + e^+ + \nu_e$ and ${}^{14}\text{O} \longrightarrow {}^{14}\text{N}^* + e^+ + \nu_e$, the amplitude factorises,

$$\mathcal{M} = \left\langle \pi^0\, e^+\nu_e \right| J_H^\mu J_{L\,\mu} \left|\pi^+\right\rangle G_{wk} = \underbrace{\left\langle \pi^0 \right| J_H^\mu \left|\pi^+\right\rangle}_{\text{known (group theory)}} \underbrace{\left\langle e^+\nu_e \right| J_{L\,\mu} \left|0\right\rangle}_{\text{pert. theory}} G_{wk}, \tag{6.56}$$

with the hadronic current $J_H \equiv T$ (vector). Since the hadronic matrix element is fixed by group theory and the leptonic one by perturbation theory, G_{wk} can be extracted from the measured rates!

6.6.2 Pion decay

In pion decay,

$$\pi^+ \to \mu^+ + \nu_\mu, \tag{6.57}$$

the amplitude involves

$$\langle 0|\, X_i^\mu \,|\pi_j\rangle \quad \underbrace{\langle \mu^+ \nu_\mu |\, J_{L\,\mu}\, |0\rangle}_{\text{pert. theory}}. \tag{6.58}$$

But

$$\langle 0|\, X_i^\mu(x) \,|\pi_j\rangle = \frac{+iF_\pi \delta_{ij}\, p_\pi^\mu\, e^{-ip_\pi\cdot x}}{2(2\pi)^{3/2}\sqrt{2p_\pi^0}}. \tag{6.59}$$

From the matrix element,

$$\Gamma(\pi \to \mu + \nu) = \frac{G_{wk}^2 F_\pi^2 m_\mu^2 \left(m_\pi^2 - m_\mu^2\right)^2}{16\pi\, m_\pi^3}. \tag{6.60}$$

Comparing with the measured pion lifetime we can extract F_π:

$$\boxed{F_\pi \simeq 184 \text{ MeV}}. \tag{6.61}$$

6.7 Nucleon matrix elements of the axial current and the Goldberger–Treiman relation

Let us look at

$$\begin{aligned} &\langle \text{proton}|\, X_1^\mu(x) + iX_2^\mu(x)\,|\text{neutron}\rangle \\ &= (2\pi)^{-3} e^{-iq\cdot x}\, \bar{u}_p \left[\gamma^\mu\gamma_5\, f(q^2) + q^\mu\gamma_5\, g(q^2) + iq_\nu\,[\gamma^\mu,\gamma^\nu]\,\gamma_5\, h(q^2)\right] u_n, \qquad q \equiv p_n - p_p \end{aligned} \tag{6.62}$$

If $SU(2)_L \otimes SU(2)_R$ were exact, then

$$q_\mu \langle p|\, X_1^\mu + iX_2^\mu\,|n\rangle = 0. \tag{6.63}$$

Then, with

$$\bar{u}_p \not{p}_p = m_p\, \bar{u}_p, \qquad \not{p}_n u_n = m_n u_n, \tag{6.64}$$

$$\Rightarrow \quad q_\mu \bar{u}_p \gamma^\mu \gamma_5 u_n = -2m_N\, \bar{u}_p \gamma_5 u_n \tag{6.65}$$

$$\Rightarrow \quad \boxed{2m_N f(q^2) = q^2 g(q^2)}. \tag{6.66}$$

As $q \to 0$, if $g(q^2)$ had no singularity, we would need $m_N f(0) = 0$. This is not true: the nucleon is massive and, as we now see, $f(0) \neq 0$.

From neutron decay we can measure $f(0)$:

$$n \to p + e^- + \bar{\nu}_e, \qquad f(0) \simeq 1.2573(28). \tag{6.67}$$

Remark. The value 1.2573(28) is the one quoted in the 1990s (e.g. Weinberg, QFT Vol. II, Sec. 19.4). The modern value, dominated by neutron-decay asymmetry measurements, is $|g_A| = 1.2754(13)$ (PDG 2024). The tensor term h is a "second-class" current and vanishes in the isospin limit. Then $g(q^2)$ must have a pole,

$$g(q^2) \to \frac{2m_N g_A}{q^2}, \qquad g_A \equiv f(0). \tag{6.68}$$

The pion can do the job!

6.7.1 The pion pole

The pion-nucleon Hamiltonian is

$$H_{\pi N} \simeq -2i\, G_{\pi N}\, \vec{\pi}\cdot \bar{N}\gamma_5 \vec{\tau} N. \tag{6.69}$$

The matrix element

$$\langle 0|\, X_i^\mu(x)\, |\pi_j\rangle = \frac{+iF_\pi\, \delta_{ij}\, p_\pi^\mu\, e^{-ip_\pi\cdot x}}{2(2\pi)^{3/2}\sqrt{2p_\pi^0}} \tag{6.70}$$

can be reproduced by replacing

$$\vec{X}^\mu \to -F_\pi \frac{\partial^\mu \vec{\pi}}{2}. \tag{6.71}$$

Then the pion-exchange contribution to the nucleon matrix element is

$$\left\langle p\right|\, X_1^\mu(x) + iX_2^\mu(x)\, |n\rangle \to \frac{e^{-iq\cdot x}}{(2\pi)^3}\left[-2G_{\pi N}\, \bar{u}_p \gamma_5 u_n\right]\left[\frac{i}{q^2 - m_\pi^2}\right]\left[iq^\mu \frac{F_\pi}{2}\right], \tag{6.72}$$

and therefore

$$g(q^2) \to \frac{G_{\pi N} F_\pi}{q^2} \qquad (m_\pi \to 0). \tag{6.73}$$

Comparing with $g(q^2) \to 2m_N g_A/q^2$ gives the Goldberger–Treiman relation,

$$2m_N g_A = G_{\pi N} F_\pi, \tag{6.74}$$

to be compared with the measured values

$$m_N = \frac{m_p + m_n}{2} \simeq 940\ \text{MeV}, \qquad g_A \simeq 1.257, \qquad G_{\pi N} \simeq 13.5. \tag{6.75}$$

Remark. With current inputs, $m_N = 938.9\,\text{MeV}$, $g_A = 1.2754$, $F_\pi = 184.1\,\text{MeV}$ and $G_{\pi N}^2/4\pi \simeq 13.7$ ($G_{\pi N} \simeq 13.1$, from πN scattering and pionic atoms), one finds $2m_N g_A = 2.395\,\text{GeV}$ versus $G_{\pi N} F_\pi = 2.416\,\text{GeV}$: the Goldberger–Treiman discrepancy is about 1–2%, of order $m_\pi^2/(1\,\text{GeV})^2$, as expected from explicit chiral symmetry breaking.

7

Pions: approximate chiral symmetry and effective field theory

QCD exhibits an $SU(2) \otimes SU(2)$ symmetry, and spontaneous symmetry breaking $SU(2) \otimes SU(2) \to SU(2)$ has interesting consequences: pions! A quantitative consequence is the Goldberger–Treiman relation (see Chapter 6).

7.1 An approximate $SU(2) \otimes SU(2)$ symmetry

$SU(2) \otimes SU(2)$ is only an approximate symmetry of QCD, due to the quark mass terms. The symmetry breaking part of the Hamiltonian is

$$\begin{aligned} H_I &= m_u \bar{u}u + m_d \bar{d}d \\ &= (m_u + m_d)\,\frac{\bar{u}u + \bar{d}d}{2} + (m_u - m_d)\,\frac{\bar{u}u - \bar{d}d}{2}. \end{aligned} \tag{7.1}$$

In terms of the doublet

$$q = \begin{pmatrix} u \\ d \end{pmatrix}$$

we have

$$H_I = (m_u + m_d)\left(\bar{q}\,\frac{\mathbf{1}}{2}\,q\right) + (m_u - m_d)\left(\bar{q}\,\tau_3\,q\right), \tag{7.2}$$

where $\tau_a = \sigma_a/2$ are the isospin generators. Indeed,

$$\bar{q}\tau_3 q = \frac{1}{2}\left(\bar{u}, \bar{d}\right)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} u \\ d \end{pmatrix} = \frac{1}{2}\left(\bar{u}u - \bar{d}d\right).$$

Let us define two four-vectors of operators, ϕ_n^+ and ϕ_n^-:

$$\phi_n^+ : \quad \begin{pmatrix} \phi_i^+ \\ \phi_4^+ \end{pmatrix} = \begin{pmatrix} i\bar{q}\gamma_5\tau_i q \\ \frac{1}{2}\bar{q}q \end{pmatrix}, \qquad i = 1,2,3, \tag{7.3}$$

and

$$\phi_n^- : \quad \begin{pmatrix} \phi_i^- \\ \phi_4^- \end{pmatrix} = \begin{pmatrix} \bar{q}\tau_i q \\ -\frac{i}{2}\bar{q}\gamma_5 q \end{pmatrix}, \qquad i = 1,2,3. \tag{7.4}$$

Then

$$H_I = (m_u + m_d)\,\phi_4^+ + (m_u - m_d)\,\phi_3^-. \tag{7.5}$$

The two four-vectors ϕ^+ and ϕ^- transform in the same way under $SU(2)\otimes SU(2)$ (they differ in parity). Denoting by T_i the (isospin) and by X_i the (axial) generators of $SU(2)\otimes SU(2)$, we find

$$\begin{aligned}\left[T_i, \phi_n^\pm\right] &= -\sum_m \left(\tilde{T}_i\right)_{nm} \phi_m^\pm, \\ \left[X_i, \phi_n^\pm\right] &= -\sum_m \left(\tilde{X}_i\right)_{nm} \phi_m^\pm,\end{aligned} \tag{7.6}$$

where $\tilde{T}_i$, $\tilde{X}_i$ turn out to be generators of the fundamental (four-dimensional) representation of $SO(4)$.

Exercise: Show that $\tilde{T}_i$, $\tilde{X}_i$ are generators of the fundamental representation of $SO(4)$.

Exercise: Prove that the $SO(4)$ Lie algebra is equivalent to the $SU(2)\otimes SU(2)$ Lie algebra.

7.2 Vacuum alignment in QCD

The vacuum alignment condition is

$$\sum_{nm} \frac{\partial V_I}{\partial \phi_n} T^\alpha_{nm} \left\langle \phi_m \right\rangle = 0. \tag{7.7}$$

If the symmetry-breaking potential is linear in a set of operators Φ_n,

$$V_I = \sum_n u_n \Phi_n, \tag{7.8}$$

which satisfy

$$[T_\alpha, \Phi_n] = -\sum_m T^\alpha_{nm} \Phi_m, \tag{7.9}$$

with T_α the conserved charges and T^α_{nm} the symmetry generators (in the representation furnished by the Φ_n), then the vacuum alignment condition becomes

$$\boxed{\sum_{nm} u_n T^\alpha_{nm} \left\langle \phi_m \right\rangle = 0} \tag{7.10}$$

or, equivalently,

$$0 = \langle [T^\alpha, H_I] \rangle . \tag{7.11}$$

For the QCD symmetry breaking Hamiltonian of Eq. (7.5) the vacuum alignment condition becomes

$$\begin{aligned}0 &= (m_u + m_d) \left\langle \left[T_i, \phi_4^+\right] \right\rangle + (m_u - m_d) \left\langle \left[T_i, \phi_3^-\right] \right\rangle , \\ 0 &= (m_u + m_d) \left\langle \left[X_i, \phi_4^+\right] \right\rangle + (m_u - m_d) \left\langle \left[X_i, \phi_3^-\right] \right\rangle .\end{aligned} \tag{7.12}$$

For T_3 we get $0 = 0$. For T_1, T_2, X_1, X_2, X_3 we obtain

$$\left\langle \phi_2^- \right\rangle = \left\langle \phi_1^- \right\rangle = \left\langle \phi_1^+ \right\rangle = \left\langle \phi_2^+ \right\rangle = 0 \tag{7.13}$$

and

$$(m_u + m_d) \left\langle \phi_3^+ \right\rangle - (m_u - m_d) \left\langle \phi_4^- \right\rangle = 0. \tag{7.14}$$

We know that QCD preserves parity, i.e. parity is an unbroken symmetry. Therefore, only parity-even operators are allowed to have a non-zero vacuum expectation value. We then have that

$$\left\langle \phi_1^+ \right\rangle = \left\langle \phi_2^+ \right\rangle = \left\langle \phi_3^+ \right\rangle = \left\langle \phi_4^- \right\rangle = 0 \tag{7.15}$$

and the vacuum alignment condition is satisfied. Conversely, once $\left\langle \phi_4^- \right\rangle = 0$, Eq. (7.14) gives $\left\langle \phi_3^+ \right\rangle = 0$, and together with $\left\langle \phi_1^+ \right\rangle = \left\langle \phi_2^+ \right\rangle = 0$ all parity-odd vacuum expectation values vanish, as parity requires.

Therefore, the $\vec{T}$ generators are unbroken. The explicit small breaking of the symmetry in the Lagrangian due to H_I therefore forces us to have a spontaneous symmetry breaking such that only

$$\left\langle \phi_4^+ \right\rangle \neq 0 \qquad \left(\Leftrightarrow \left\langle \frac{\bar{u}u + \bar{d}d}{2} \right\rangle \neq 0 \right) \tag{7.16}$$

is permitted. This is a viable pattern for symmetry breaking, compatible with isospin invariance.

7.3 Mass of the pion

We can now turn to the pseudo-Goldstone boson mass matrix, which can give us information on the pion mass. We have found that (general formula; T^α is a general broken generator)

$$M_{cd}^2 = -\sum_{\alpha\beta} F_{c\alpha}^{-1} F_{d\beta}^{-1} \left[\sum_{nm} \delta^\alpha \left\langle \phi_n \right\rangle \delta^\beta \left\langle \phi_m \right\rangle \frac{\partial^2 V_I(\langle \phi \rangle)}{\partial \left\langle \phi_n \right\rangle \partial \left\langle \phi_m \right\rangle} + \sum_{nm} T_{nm}^\alpha \delta^\beta \left\langle \phi_m \right\rangle \frac{\partial V_I(\langle \phi \rangle)}{\partial \left\langle \phi_n \right\rangle} \right]. \tag{7.17}$$

For

$$V_I = \sum_n u_n \Phi_n, \qquad [T_\alpha, \Phi_n] = -T_{nm}^\alpha \Phi_m$$

(with T_{nm}^α the generators of the group) we have

$$M_{cd}^2 = -\sum_{\alpha\beta} F_{c\alpha}^{-1} F_{d\beta}^{-1} \left\langle \left[T_\alpha, \left[T_\beta, H_I \right] \right] \right\rangle. \tag{7.18}$$

Isospin (rotational) symmetry, due to the unbroken generators $\vec{T}$, gives that

$$F_{ab} = \frac{F_\pi}{2} \delta_{ab}, \tag{7.19}$$

where F_{ab} is defined from the matrix element of the broken currents between the vacuum and a Goldstone boson state,

$$\left\langle 0 \right| J_\alpha^\mu(x) \left| B_b \right\rangle = \frac{i F_{\alpha b} \, p_B^\mu \, e^{-i p_B \cdot x}}{(2\pi)^{3/2} \sqrt{2 p_B^0}}. \tag{7.20}$$

Thus,

$$M_{cd}^2 = -\frac{4}{F_\pi^2} \left\langle \left[T_c, \left[T_d, H_I \right] \right] \right\rangle. \tag{7.21}$$

In our case, with the broken generators X_a, we find

$$\begin{aligned} \left[X_a, \left[X_b, \phi_4^+ \right] \right] &= \delta_{ab} \, \phi_4^+, \\ \left[X_a, \left[X_b, \phi_3^- \right] \right] &= \delta_{b3} \, \phi_a^-. \end{aligned} \tag{7.22}$$

Since $\langle \phi_a^- \rangle = 0$, only the first term contributes. Thus, the pion mass matrix is

$$M_{ab}^2 = \delta_{ab}\, m_\pi^2 \tag{7.23}$$

with

$$\boxed{m_\pi^2 = -\frac{4(m_u + m_d)\,\langle \phi_4^+ \rangle}{F_\pi^2}} \tag{7.24}$$

Remark. The vacuum minimises $\langle H_I \rangle$ on the vacuum orbit, which forces $(m_u + m_d)\,\langle \phi_4^+ \rangle < 0$, so $m_\pi^2 > 0$ (this is the positivity of Chapter 5). With $f_\pi = F_\pi/2 \simeq 92$ MeV and $\langle \bar{u}u \rangle = \langle \bar{d}d \rangle$ this is the Gell-Mann–Oakes–Renner relation $m_\pi^2 f_\pi^2 = -(m_u + m_d)\,\langle \bar{u}u \rangle$. The isospin-breaking $(m_u - m_d)$ term does not contribute at first order, so the $\pi^\pm$–π^0 splitting is dominantly electromagnetic (Dashen); see Weinberg, QFT Vol. II, Sec. 19.3, and Donoghue, Golowich, Holstein, *Dynamics of the Standard Model*. All pions π^+, π^-, π^0 are mass degenerate. The small difference in the mass of the charged and neutral pions is due to the electromagnetic interaction.

7.4 An effective Lagrangian approach for pion interactions

QCD is a non-perturbative theory at low energies. Nevertheless, in our study of pions and their interactions we were able to constrain transition amplitudes and "calculate" basic quantities (such as the mass) using symmetry and powerful theorems of spontaneous symmetry breaking. This helped us to bypass the obstacle of not being able to compute in QCD, due to its non-perturbative behaviour at low energies.

Since symmetry is the basic feature which led to our predictions, could we substitute the QCD Lagrangian (with quark and gluon degrees of freedom) with a different Lagrangian (directly in terms of the "composite" pion fields) which is simpler, perhaps even perturbative, in the sense that it describes the same physics at low energies, having the same symmetries and symmetry breaking patterns as QCD? If symmetry is the crucial ingredient, we should arrive at the same conclusions! Such a theory is an *effective theory*; see Fig. 7.1.

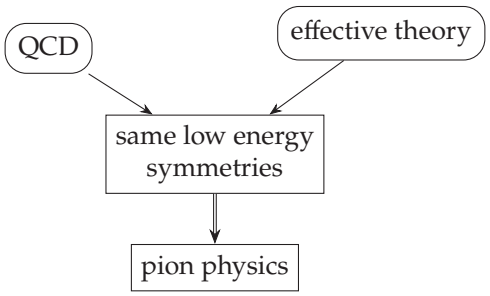


Figure 7.1. QCD and the effective theory share the same low energy symmetries, which determine pion physics.

7.5 The $SO(4)$ sigma model in polar variables

For example, we can consider an $SO(4)$-invariant Lagrangian

$$\mathcal{L} = \frac{1}{2}\,(\partial_\mu \phi_n)\,(\partial^\mu \phi_n) - \frac{\lambda}{4}\,(\phi_n \phi_n - v^2)^2\,, \qquad n = 1,2,3,4, \tag{7.25}$$

where $\vec{\phi} \equiv (\phi_1, \phi_2, \phi_3)$ is an "isovector" pseudoscalar field (corresponding to $\phi_i^+ \equiv i\bar{q}\gamma_5\tau_i q$) and ϕ_4 an "isoscalar" scalar field (corresponding to $\phi_4^+ \equiv \frac{\bar{q}q}{2}$). We can use such a Lagrangian to do perturbation theory in the (small) energies of the Goldstone bosons. *Remark.* The low-energy expansion is in powers of E/M_σ and, in the non-linear theory below, of $E^2/(4\pi F)^2$ (Weinberg's power counting).

We would like to cast this Lagrangian in terms of Goldstone boson fields plus degrees of freedom "orthogonal" to the Goldstone bosons. First, we observe that we can generate all fields $\phi_n(x)$ from one element in the field space,

$$(0, 0, 0, \sigma(x)),$$

using a symmetry rotation. That is, we write the field $\phi_n(x)$ as an $SO(4)$ rotation acting on

$$\phi_n^{\text{spec}}(x) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma(x) \end{pmatrix}, \tag{7.26}$$

$$\phi_n(x) = R_{nm}(x)\,\phi_m^{\text{spec}}(x) = R_{n4}(x)\,\sigma(x), \tag{7.27}$$

where

$$R^T(x)R(x) = \mathbf{1}. \tag{7.28}$$

With this substitution, we have

$$\phi_n\phi_n = R_{n4}R_{n4}\,\sigma(x)^2 = \underbrace{R_{4n}^T R_{n4}}_{1}\,\sigma(x)^2 \;\Rightarrow\; \phi_n\phi_n = \sigma(x)^2, \tag{7.29}$$

i.e. $\sigma(x) = \sqrt{\sum_n \phi_n(x)^2}$. The field $\sigma(x)$ is invariant under all $SO(4)$ transformations (in particular the unbroken ones), and so is its vacuum expectation value. Thus it is not a Goldstone boson.

Also,

$$\begin{aligned} \left(\partial_\mu\phi_n\right)^2 &= \left[(\partial_\mu\sigma)R_{n4} + (\partial_\mu R_{n4})\,\sigma\right]^2 \\ &= (\partial_\mu\sigma)^2 + (\partial_\mu R_{n4})^2\sigma^2(x) + 2\sigma\,\partial_\mu\sigma\,\underbrace{R_{n4}\partial^\mu R_{n4}}_{\frac{1}{2}\partial^\mu R_{n4}^2 = \frac{1}{2}\partial^\mu 1 = 0}, \end{aligned} \tag{7.30}$$

so that

$$\left(\partial_\mu\phi_n\right)^2 = (\partial_\mu\sigma)^2 + (\partial_\mu R_{n4})^2\sigma^2(x). \tag{7.31}$$

The Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\sigma)^2 + \frac{1}{2}\sigma^2\left(\partial_\mu R_{n4}\right)^2 - \frac{\lambda}{4}\left(\sigma^2 - v^2\right)^2. \tag{7.32}$$

This is just a rewriting.

The effective potential is

$$V_{eff} = \frac{\lambda}{4}\left(\langle\sigma\rangle^2 - v^2\right)^2 + O(\hbar), \qquad V'_{eff} = \lambda\left(\langle\sigma\rangle^2 - v^2\right)\langle\sigma\rangle, \tag{7.33}$$

and

$$\langle\sigma\rangle = v \neq 0 \quad \text{for} \quad v^2 > 0. \tag{7.34}$$

The $\sigma(x)$ field gets a vacuum expectation value $\Rightarrow$ spontaneous symmetry breaking. In the equivalent parametrisation $V = \frac{\mu^2}{2}\sigma^2 + \frac{\lambda}{4}\sigma^4$, $V'_{eff} = \langle\sigma\rangle\left(\mu^2 + \lambda\langle\sigma\rangle^2\right)$ and, for $\mu^2 < 0$, $\langle\sigma\rangle = |\mu|/\sqrt{\lambda}$. Instead of ϕ_n we now have as degrees of freedom the shifted field

$$\sigma' = \sigma - \frac{|\mu|}{\sqrt{\lambda}} \qquad (= \sigma - v)$$

and the variables which parametrise the rotation R_{n4}.

Notice that the Goldstone fields in R_{n4} enter the Lagrangian *only as derivatives*. Let us parametrise the rotation R_{nm} as follows. We define

$$\zeta_a = \frac{\phi_a}{\phi_4 + \sigma}, \qquad a = 1, 2, 3, \tag{7.35}$$

and

$$R_{a4} = \frac{2\zeta_a}{1 + \vec{\zeta}^2} = -R_{4a}, \qquad R_{44} = \frac{1 - \vec{\zeta}^2}{1 + \vec{\zeta}^2}, \qquad R_{ab} = \delta_{ab} - \frac{2\zeta_a\zeta_b}{1 + \vec{\zeta}^2}. \tag{7.36}$$

Indeed,

$$R_{a4}^2 + R_{44}^2 = \frac{(2\zeta_a)^2}{(1 + \vec{\zeta}^2)^2} + \frac{(1 - \vec{\zeta}^2)^2}{(1 + \vec{\zeta}^2)^2} = \frac{(1 + \vec{\zeta}^2)^2}{(1 + \vec{\zeta}^2)^2} = 1. \tag{7.37}$$

Exercise: Check that all rows of R are normalised, $\sum_m R_{nm}^2 = 1$, for $n = 1, 2, 3, 4$.

The Lagrangian becomes

$$\boxed{\mathcal{L} = \frac{1}{2}(\partial_\mu\sigma)(\partial^\mu\sigma) - \frac{\lambda}{4}\left(\sigma^2 - v^2\right)^2 + 2\sigma^2\,\vec{D}_\mu \cdot \vec{D}^\mu} \tag{7.38}$$

with

$$\vec{D}_\mu \equiv \frac{\partial_\mu\vec{\zeta}}{1 + \vec{\zeta}^2}. \tag{7.39}$$

Notice the absence of a mass term for the $\vec{\zeta}$ fields in the Lagrangian: the $\vec{\zeta}$ are massless degrees of freedom (no term $\propto \vec{\zeta}^2$ can be created). They are zero mass particles with derivative interactions. Expanding in

$$\frac{1}{1 + \vec{\zeta}^2} = 1 - \vec{\zeta}^2 + (\vec{\zeta}^2)^2 - \ldots$$

we see that all interactions of $\vec{\zeta}$ particles vanish as the momenta go to zero, i.e. they vanish at low energies.

7.6 Symmetries of the Lagrangian

The Lagrangian is still $SO(4)$ invariant! Under isospin transformations (generated by $\vec{T}$, unbroken), $\vec{\zeta}$ transforms as an "isovector",

$$\delta\vec{\zeta} = \vec{\theta} \times \vec{\zeta}, \tag{7.40}$$

while σ does not transform ("isoscalar"),

$$\delta\sigma = 0. \tag{7.41}$$

$\vec{D}_\mu$ is then also an isovector,

$$\delta\vec{D}_\mu = \delta\frac{\partial_\mu\vec{\zeta}}{1+\vec{\zeta}^{\,2}} = \frac{\partial_\mu\delta\vec{\zeta}}{1+\vec{\zeta}^{\,2}} = \frac{\partial_\mu\left(\vec{\theta}\times\vec{\zeta}\right)}{1+\vec{\zeta}^{\,2}} = \vec{\theta}\times\vec{D}_\mu, \qquad \delta D^i_\mu = \epsilon_{ijk}\,\theta^j D^k_\mu, \tag{7.42}$$

and

$$\delta\left(\vec{D}_\mu\cdot\vec{D}^\mu\right) = 2\,\epsilon_{ijk}\left(\theta^j D^k_\mu\right)D^{\mu\,i} = 0. \tag{7.43}$$

This was to be expected. We must work a bit further for the transformations of the broken generators (rotating from the 4 direction to the 1,2,3 directions; Fig. 7.2). In terms of the original fields ϕ_n, a symmetry transformation with parameters $(\epsilon_1,\epsilon_2,\epsilon_3)\equiv\vec{\epsilon}$ changes the fields as

$$\delta\vec{\phi} = 2\vec{\epsilon}\,\phi_4, \qquad \delta\phi_4 = -2\vec{\epsilon}\cdot\vec{\phi}. \tag{7.44}$$

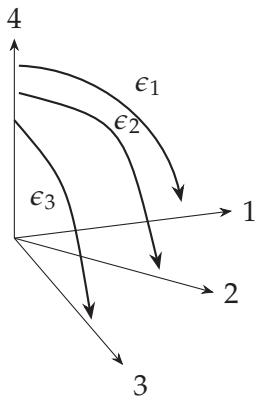


Figure 7.2. The broken generators rotate the direction 4 of the vacuum into the directions $1,2,3$.

We deduce that (exercise)

$$\delta\vec{\zeta} = \vec{\epsilon}\left(1-\vec{\zeta}^{\,2}\right) + 2\vec{\zeta}\left(\vec{\epsilon}\cdot\vec{\zeta}\right), \qquad \delta\sigma = 0. \tag{7.45}$$

This is a *non-linear* symmetry transformation! Nevertheless, the Lagrangian is invariant because $\vec{D}_\mu$ transforms linearly (although with a field-dependent transformation): indeed,

$$\begin{aligned}\delta\vec{\zeta}^{\,2} &= 2\zeta_i\,\delta\zeta_i = 2\zeta_i\left(\epsilon_i(1-\vec{\zeta}^{\,2}) + 2\zeta_i(\vec{\epsilon}\cdot\vec{\zeta})\right)\\ &= 2(\vec{\zeta}\cdot\vec{\epsilon})(1-\vec{\zeta}^{\,2}) + 4\vec{\zeta}^{\,2}(\vec{\epsilon}\cdot\vec{\zeta}) = 2\,\vec{\zeta}\cdot\vec{\epsilon}\,\left(1+\vec{\zeta}^{\,2}\right)\end{aligned} \tag{7.46}$$

and

$$\begin{aligned}\delta\left(\frac{\partial_\mu\zeta_i}{1+\vec{\zeta}^{\,2}}\right) &= \frac{\partial_\mu(\delta\zeta_i)}{1+\vec{\zeta}^{\,2}} - \frac{\partial_\mu\zeta_i\;\delta\vec{\zeta}^{\,2}}{(1+\vec{\zeta}^{\,2})^2}\\ &= \frac{1}{1+\vec{\zeta}^{\,2}}\left[\partial_\mu(\delta\zeta_i) - \frac{(\partial_\mu\zeta_i)\,\delta\vec{\zeta}^{\,2}}{1+\vec{\zeta}^{\,2}}\right]\\ &= \frac{1}{1+\vec{\zeta}^{\,2}}\left[\partial_\mu\left(\epsilon_i(1-\vec{\zeta}^{\,2}) + 2\zeta_i(\vec{\epsilon}\cdot\vec{\zeta})\right) - 2(\partial_\mu\zeta_i)\,\vec{\zeta}\cdot\vec{\epsilon}\right]\end{aligned} \tag{7.47}$$

and, after some algebra (exercise),

$$\boxed{\delta\vec{D}_\mu = 2\left(\vec{\epsilon}\times\vec{\zeta}\right)\times\vec{D}_\mu} \tag{7.48}$$

This is a non-linear realisation of the $SU(2)\otimes SU(2)$ group! Its $SU(2)$ isospin subgroup is realised linearly.

7.7 The non-linear sigma model

If we drop the σ field, we still have an $SO(4)$ invariant Lagrangian. We can do that assuming that the $\sigma(x)$ mass

$$m_\sigma^2 = 2\lambda v^2 \to \infty \tag{7.49}$$

as v stays finite, i.e. $\lambda\to\infty$. Then $\sigma = v$ is frozen, and the potential term $\lambda(\sigma^2-v^2)^2 \sim \infty\cdot 0$ merely enforces this constraint and drops out of the Lagrangian. So, the Lagrangian

$$\mathcal{L} = \frac{F^2}{2}\,\vec{D}_\mu\cdot\vec{D}^\mu = \frac{F^2}{2}\,\frac{(\partial_\mu\vec{\zeta})\cdot(\partial^\mu\vec{\zeta})}{(1+\vec{\zeta}^{\,2})^2} \tag{7.50}$$

is $SO(4)$ invariant. (Comparing with Eq. (7.38) at $\sigma = v$, $F^2/2 = 2v^2$, i.e. $F = 2v$.) We can canonically normalise, so that

$$\vec{\zeta} = \frac{\vec{\pi}}{F}. \tag{7.51}$$

Then

$$\boxed{\mathcal{L} = \frac{1}{2}\,\frac{(\partial_\mu\vec{\pi})\cdot(\partial^\mu\vec{\pi})}{\left[1+\frac{\vec{\pi}^{\,2}}{F^2}\right]^2}} = \frac{1}{2}(\partial_\mu\vec{\pi})^2 - \frac{\vec{\pi}^{\,2}(\partial_\mu\vec{\pi})^2}{F^2} + \dots \tag{7.52}$$

This is the *non-linear σ-model* for

$$SU(2)_L\times SU(2)_R \to SU(2)_V.$$

7.8 Linear versus non-linear representation

We could have kept $\phi_4(x) = \sigma(x)$ and $\phi_{1,2,3}(x) = \vec{\pi}(x)$ as our fields, with $\vec{\pi}$ the Goldstone bosons (linear representation). But then it would have been difficult to eliminate the σ-field and get a theory of pions alone. Indeed,

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2}(\partial_\mu\sigma)^2 + \frac{1}{2}(\partial_\mu\vec{\pi})^2 - \frac{\lambda}{4}\left(\vec{\pi}^{\,2}+\sigma^2-v^2\right)^2 \\
&\overset{\sigma\to\sigma+v}{=} \frac{1}{2}(\partial_\mu\sigma)^2 + \frac{1}{2}(\partial_\mu\vec{\pi})^2 - \frac{\lambda}{4}\left(\vec{\pi}^{\,2}+\sigma(\sigma+2v)\right)^2 \\
&= \frac{1}{2}(\partial_\mu\sigma)^2 + \frac{1}{2}(\partial_\mu\vec{\pi})^2 - \lambda v^2\sigma^2 - \frac{\lambda}{4}\vec{\pi}^{\,4} - \frac{\lambda}{2}\vec{\pi}^{\,2}\sigma^2 - \lambda v\,\vec{\pi}^{\,2}\sigma + \dots
\end{aligned} \tag{7.53}$$

The σ mass term is $\frac{1}{2}M_\sigma^2\sigma^2$, and the $\pi\pi\sigma$ coupling is $\propto\lambda v$. The exchange of a σ between two pairs of pions (Fig. 7.3) gives

$$\sim\frac{\lambda^2v^2}{M_\sigma^2}\to\frac{\lambda^2v^2}{\lambda v^2} = \lambda\to\infty, \tag{7.54}$$

which does not decouple in the limit $M_\sigma \to \infty$. *Remark.* The growth $\sim \lambda$ is an artefact of the linear field basis. At zero momentum the σ-exchange diagrams cancel exactly against the contact term $-\frac{\lambda}{4}\vec{\pi}^4$, as required by Goldstone's theorem; e.g. the s-channel amplitude is $-2\lambda\, s/(s - M_\sigma^2) \to s/v^2$ for $M_\sigma \to \infty$, the leading-order result of the non-linear model. The physics agrees; only the cancellation is hidden (Weinberg, QFT Vol. II, Sec. 19.5; Donoghue, Golowich, Holstein).

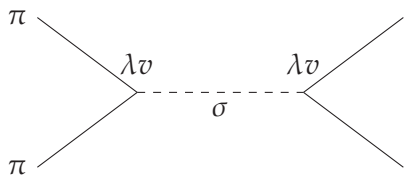


Figure 7.3. σ exchange between pions in the linear representation.

With our (non-linear) Lagrangian,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \sigma)^2 - \frac{\lambda}{4}\left(\sigma^2 - v^2\right)^2 + 2\sigma^2\, \vec{D}_\mu \cdot \vec{D}^\mu,$$

shifting $\sigma \to \sigma + v$ gives couplings of σ to the pions only through $\vec{D}_\mu$, i.e. through derivatives of the pion fields. Diagrams with a σ exchanged between $\vec{D}_\mu$ vertices (Fig. 7.4) behave as

$$\sim \frac{1}{M_\sigma^2} \to 0.$$

We can take the $M_\sigma \to \infty$ limit at the diagram level, ignoring all diagrams originating from σ terms in the Lagrangian.

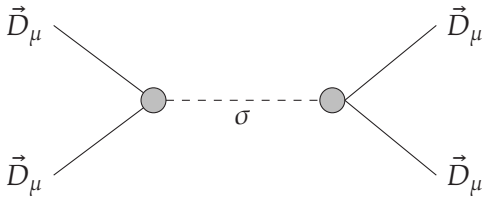


Figure 7.4. σ exchange in the non-linear representation: the vertices involve $\vec{D}_\mu$ only, and the diagram is suppressed as $1/M_\sigma^2$.

7.9 Alternative formulation: the U matrix

Consider the 2×2 matrix

$$\Phi = h_4 \mathbf{1} + i h_i \sigma^i \tag{7.55}$$

constrained by

$$\Phi^\dagger \Phi = \left(h_4 \mathbf{1} - i h_i \sigma^i\right)\left(h_4 \mathbf{1} + i h_j \sigma^j\right) = \left(h_4^2 + \vec{h}^{\,2}\right)\mathbf{1} = v^2\, \mathbf{1} \quad \text{(constant)}. \tag{7.56}$$

Then

$$\Phi = U \cdot v, \qquad U \in SU(2). \tag{7.57}$$

We can parametrise

$$U = e^{i\pi_a \sigma^a / v}. \tag{7.58}$$

This transforms as

$$U \to g_L\, U\, g_R^\dagger \tag{7.59}$$

under an $SU(2)_L \times SU(2)_R$ symmetry transformation. Take

$$\boxed{\mathcal{L} = \frac{v^2}{4}\,\mathrm{tr}\left(\partial_\mu U^\dagger \partial^\mu U\right) = \frac{F^2}{16}\,\mathrm{tr}\left(\partial_\mu U^\dagger \partial^\mu U\right), \qquad v \equiv \frac{F}{2}.} \tag{7.60}$$

Under global $SU(2) \times SU(2)$ transformations this stays invariant:

$$\mathcal{L} \to \frac{F^2}{16}\,\mathrm{tr}\left(\partial_\mu \left(g_R U^\dagger g_L^\dagger\right) \partial^\mu \left(g_L U g_R^\dagger\right)\right) = \frac{F^2}{16}\,\mathrm{tr}\left(\partial_\mu U^\dagger \partial^\mu U\, g_R^\dagger g_R\right) = \mathcal{L}. \tag{7.61}$$

7.10 Goldstone bosons in general

The construction of this chapter, separating the fields into Goldstone-free fields (σ) and a space-time dependent group rotation ($R_{n4}(x)$, or $U(x)$) which contains the Goldstone bosons, works for any pattern of spontaneous breaking $G \to H$ of a global symmetry. This was shown in Part III (section "General broken global symmetries", around Eq. (8.21)): writing $\psi(x) = \gamma(x)\tilde{\psi}(x)$ with $\tilde{\psi}$ orthogonal to the Goldstone directions $T^\alpha \langle\psi\rangle$, the Goldstone bosons appear in the Lagrangian only through derivatives of $\gamma(x)$. This forbids mass terms $\sim m_B^2 B(x)B(x)$ for them, and predicts that at low energies Goldstone interactions vanish, since $\partial_\mu B \to p_B^\mu$ in the Feynman rules. If the symmetry is local, the same rewriting leaves the Lagrangian invariant and the Goldstone fields disappear completely from the spectrum (see the section "Spontaneous symmetry breaking of local gauge symmetries" of Part III).

8

Non-renormalisable effective theories at one loop

In this chapter we follow the lectures of M. K. Gaillard, *Effective nonrenormalizable theories at one loop* (Cargèse Summer School on Particle Physics, 1987; preprint LBL-24114). We study how loop corrections behave in a non-renormalisable effective theory, and compare them with the exact calculation in the underlying renormalisable theory, using two examples drawn from the Standard Model: the Fermi theory of the weak interactions and the non-linear σ-model obtained from the Higgs sector in the limit of a very heavy Higgs boson. References of the form (G1.3) point to the corresponding equations of Gaillard's lectures.

8.1 Effective non-renormalisable theories in physics

8.1.1 Infrared limits of the Standard Model

Two familiar effective theories describe the low-energy limit of an underlying renormalisable theory.

- The Fermi theory of β-decay describes the weak charged-current interactions in the limit of small momentum transfer,

$$|q^2| \ll m_W^2 = \frac{\sqrt{2}\, g^2}{8 G_F}. \tag{8.1}$$

- The $SU(2)_L \otimes SU(2)_R$ chiral invariant σ-model describes pion dynamics at low energies. In this case, however, we cannot find a parameter of QCD which yields this model in a certain limit (in the way that $m_W \to \infty$ yields the Fermi theory from the electroweak theory).

Take the Fermi theory at tree level,

$$\mathcal{L}_{\text{tree}} = 2\sqrt{2}\, G_F\, (\bar{\psi}_L \gamma^\mu \psi_L)(\bar{\psi}_L \gamma_\mu \psi_L), \tag{8.2}$$

and start computing loop corrections to the four-fermion function (Fig. 8.1). We estimate the one-loop diagram by dimensional analysis, cutting off the loop momentum at $|p| = \Lambda$: each vertex carries a factor $2\sqrt{2}G_F$, each fermion propagator counts as $i/\not{p} \sim 1/\Lambda$, and the loop integration gives

$$\int \frac{d^4 q}{(2\pi)^4} \sim \frac{\Lambda^4}{16\pi^2}$$

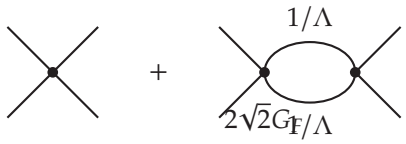


Figure 8.1. Tree-level four-fermion interaction of the Fermi theory and its one-loop correction. In the power counting each fermion propagator counts as $1/\Lambda$ and the loop integral as $\Lambda^4/(16\pi^2)$.

(there is a factor $(4\pi)^{-2}$ for each loop integration).

This gives (G1.3)

$$\mathcal{L}_{\text{eff}}^{\text{1-loop}} \sim \frac{1}{16\pi^2}\Lambda^2 \left(2\sqrt{2}G_F\right)^2 (\bar{\psi}_L\gamma\psi_L)^2 \sim \frac{G_F^2}{2\pi^2}\Lambda^2(\bar{\psi}_L\gamma\psi_L)^2. \tag{8.3}$$

The one-loop correction is quadratically divergent. In the Standard Model we know that the Fermi theory is valid only for $|q^2| \ll m_W^2$. We therefore identify the cut-off with the scale at which the Fermi theory ceases to be valid,

$$\Lambda^2 \approx m_W^2 = \frac{\sqrt{2}\,g^2}{8G_F},$$

and obtain (G1.4)

$$\mathcal{L}_{\text{eff}} = g^2\,\frac{\sqrt{2}\,G_F}{16\pi^2}\,(\bar{\psi}_L\gamma\psi_L)^2 = \frac{\alpha}{8\pi}\,\mathcal{L}_{\text{tree}}, \qquad \alpha \equiv \frac{g^2}{4\pi}, \tag{8.4}$$

where α is the weak "fine structure" constant.

Let us now do the "exact" calculation in the Standard Model. The same four-fermion function now arises from the finite box diagram with two W bosons exchanged (Fig. 8.2). Each W-fermion vertex carries a factor $g/\sqrt{2}$, each W propagator a factor $1/\Lambda^2$ and each fermion propagator a factor $1/\Lambda$, while the loop integration gives Λ^4. The integral is dominated by loop momenta of order m_W, and we find (G1.5)

$$\mathcal{L}_{\text{eff}}^{\text{1-loop}} \sim \frac{1}{16\pi^2}\left(\frac{g}{\sqrt{2}}\right)^4 \frac{1}{m_W^2}(\bar{\psi}_L\gamma\psi_L)^2 \sim \frac{\alpha}{8\pi}\,\mathcal{L}_{\text{tree}}, \tag{8.5}$$

the same estimate as in the effective theory.

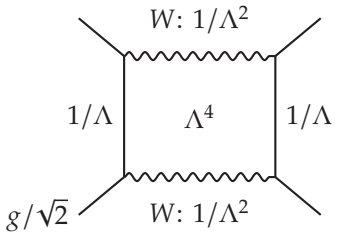


Figure 8.2. The finite box diagram of the Standard Model which corresponds to the one-loop correction of Fig. 8.1, with the power counting of its ingredients.

Notice that, given that the one-loop correction is proportional to $\mathcal{L}_{\text{tree}}$, it can be absorbed into a redefinition of the Fermi constant. But let us look at more complicated processes,

such as the eight-fermion function of Fig. 8.3 (G1.6):

$$\mathcal{L}_{\text{eff}} \sim \frac{1}{16\pi^2}\left(2\sqrt{2}G_F\right)^4 \ln\left(\frac{\Lambda^2}{\mu^2}\right)(\bar{\psi}_L\gamma\psi_L)^4. \tag{8.6}$$

Here four fermion propagators ($1/\Lambda$ each) and the loop integration (Λ^4) combine to a logarithmic divergence, and μ is a fermion mass or an external momentum, $|q_{\text{ext}}|$. This is a new operator, not present in the tree-level Lagrangian (8.2). Let us do the same in the Standard Model. The corresponding diagram has four W propagators ($1/m_W^2$ each) and four fermion propagators, and is finite:

$$\frac{\Lambda^4}{(\Lambda^2)^4\,\Lambda^4} \sim \frac{1}{m_W^8},$$

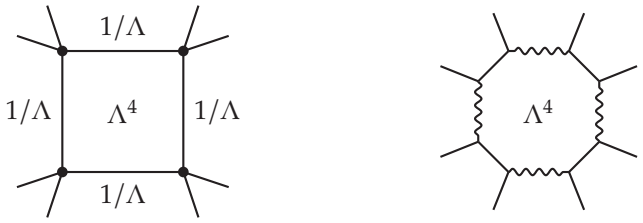
so that (G1.7)

$$\mathcal{L}_{\text{eff}} \sim \frac{1}{16\pi^2}\left(\frac{g}{\sqrt{2}}\right)^8 \frac{1}{m_W^8}\ln\left(\frac{m_W^2}{\mu^2}\right)(\bar{\psi}_L\gamma\psi_L)^4. \tag{8.7}$$

Comparing the two, using (8.1), we deduce that

$$\Lambda = O(m_W).$$

New physics, the W boson, enters to damp the divergences of the effective theory.



Fermi theory Standard Model

Figure 8.3. Eight-fermion function at one loop: logarithmically divergent in the Fermi theory (left); finite in the Standard Model (right), where each four-fermion vertex is resolved into W exchange, so that the loop contains four W and four fermion propagators.

8.2 The large Higgs mass limit of the Standard Model

Neglecting gauge interactions, the Standard Model Lagrangian reduces to that of the Higgs sector (G1.8),

$$\mathcal{L}_{SM} \xrightarrow{g\to 0} \mathcal{L}_H = \partial_\mu\varphi^\dagger\,\partial^\mu\varphi - \lambda\left(|\varphi|^2 - \frac{v^2}{2}\right)^2. \tag{8.8}$$

This is invariant under $SO(4)$ (or $SU(2)\times SU(2)$) linear transformations among the components of

$$\varphi = \begin{pmatrix}\varphi^+\\ \varphi^0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} i\pi_1 + \pi_2\\ \sigma - i\pi_3\end{pmatrix}. \tag{8.9}$$

Explicitly, $\mathcal{L}_H$ is the linear σ-model (G1.10),

$$\mathcal{L}_H = \frac{1}{2}(\partial_\mu\sigma)(\partial^\mu\sigma) + \frac{1}{2}\partial_\mu\vec{\pi}\cdot\partial^\mu\vec{\pi} - \frac{1}{4}\lambda\left(\sigma^2 + \vec{\pi}^2 - v^2\right)^2. \tag{8.10}$$

To separate the Goldstone degrees of freedom explicitly, we can write

$$\varphi = \frac{1}{\sqrt{2}}\, e^{i\vec{\theta}\cdot\vec{\tau}/v} \begin{pmatrix} 0 \\ \rho \end{pmatrix} \quad \rightsquigarrow \quad \mathcal{L}_H = \mathcal{L}_{KE}(\theta,\rho) - \frac{\lambda}{4}\left(\rho^2 - v^2\right)^2, \tag{8.11}$$

where $\mathcal{L}_{KE}$ denotes the kinetic terms. In this form the Goldstone modes θ_i appear only through derivative couplings.

The two formulations give, of course, identical results for S-matrix elements. The expansion is about the vacuum with

$$v = \langle\rho\rangle = \langle\sigma\rangle = \sqrt{2}\,\langle|\varphi|\rangle, \tag{8.12}$$

and

$$\rho = \sigma + O\left(|\varphi| - \frac{v}{\sqrt{2}}\right)^2, \qquad \theta_i = \pi_i + O\left(|\varphi| - \frac{v}{\sqrt{2}}\right)^2. \tag{8.13}$$

Let us now switch on the gauge interactions,

$$\partial_\mu\varphi \to D_\mu\varphi = (\partial_\mu + iA_\mu)\varphi, \qquad A_\mu = \frac{g}{2}\, T_\alpha A^\alpha_\mu, \tag{8.14}$$

where T^α are the generators of the $SU(2)_L \times U(1)$ group and A^α_μ the gauge fields. The gauged Lagrangian is invariant under

$$\varphi' = U(x)\,\varphi, \qquad A'_\mu = UA_\mu U^{-1} + i(\partial_\mu U)U^{-1}, \qquad \mathcal{L}(\varphi', A'_\mu) = \mathcal{L}(\varphi, A_\mu). \tag{8.15}$$

We can choose the *unitary gauge*,

$$U = e^{-i\vec{\theta}\cdot\vec{\tau}/v}, \qquad \varphi' = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \rho \end{pmatrix}. \tag{8.16}$$

The Lagrangian in this gauge is

$$\mathcal{L}_U = \mathcal{L}(A,\rho), \tag{8.17}$$

and the physical Higgs boson is $H = \rho - v$. This gauge is good for exhibiting the physical fields, but not so good for loop calculations and dimensional analysis. These are best done in a renormalisable gauge, where φ is represented linearly and the unphysical scalars π_i appear in the Lagrangian,

$$\mathcal{L}_R = \mathcal{L}(A,\sigma,\vec{\pi}), \qquad H \equiv \sigma - v. \tag{8.18}$$

In either case we find

$$m_H^2 = 2\lambda v^2. \tag{8.19}$$

The vacuum expectation value can be determined from experiment. Expanding the covariant derivative around the vacuum we find (G1.19)

$$(D_\mu\varphi)^\dagger D^\mu\varphi = \frac{g^2v^2}{4}\left(W^+_\mu W^{-\mu} + \frac{1}{2\cos^2\theta_W} Z_\mu Z^\mu\right) + m_W W^+_\mu \partial^\mu\pi^- + \cdots, \tag{8.20}$$

and therefore

$$v^2 = \frac{4m_W^2}{g^2} = \left(\sqrt{2}\,G_F\right)^{-1} \approx \left(\tfrac{1}{4}\,\text{TeV}\right)^2. \tag{8.21}$$

Since v is fixed by experiment, the limit $m_H \to \infty$ corresponds, by (8.19), to $\lambda \to \infty$: the scalar self-interactions become strong.

For the potential energy term

$$\lambda \left(\sigma^2 + \vec{\pi}^2 - v^2\right)^2$$

to remain finite, the only possibility is

$$\sigma = \sqrt{v^2 - \vec{\pi}^2}, \qquad \text{or equivalently} \qquad \rho^2 = v^2. \tag{8.22}$$

The variable ρ (or σ), and with it the physical Higgs boson, can be eliminated from the Lagrangian. The constraint

$$\rho^2 = \vec{\pi}^2 + \sigma^2 = v^2$$

is invariant under $SO(4)$ or $SU(2) \times SU(2)$ transformations.

Upon imposing it, the linear transformations (G1.23)

$$\delta\pi_i = \epsilon_{ijk}\,\alpha_j\,\pi_k + \beta_i\,\sigma, \qquad \delta\sigma = -\beta_i\,\pi_i, \tag{8.23}$$

become (G1.24)

$$\delta\pi_i = \epsilon_{ijk}\,\alpha_j\,\pi_k + \beta_i\left(v^2 - \vec{\pi}^2\right)^{1/2}, \tag{8.24}$$

which is non-linear. The Lagrangian becomes (G1.25), (G1.26)

$$\mathcal{L} = \frac{1}{2}\,g_{ij}(\pi)\,\partial_\mu\pi^i\,\partial^\mu\pi^j, \qquad g_{ij} = \delta_{ij} + \frac{\pi_i\pi_j}{v^2 - \vec{\pi}^2} \quad \text{(scalar metric)}. \tag{8.25}$$

This defines a non-renormalisable effective theory: the non-linear σ-model.

8.3 The gauged non-linear σ-model

Let us define the curvature tensor of the scalar metric (G1.59),

$$R^i{}_{jkl} = \frac{1}{v^2}\left(\delta^i_l\,g_{jk} - \delta^i_k\,g_{jl}\right), \tag{8.26}$$

and the Ricci tensor (G1.60),

$$R_{ij} = \frac{(1-N)}{v^2}\,g_{ij}, \tag{8.27}$$

with N the number of real scalars π_i. The Ricci tensor is proportional to the metric because (8.25) is the only two-derivative term invariant under the $SU(2) \times SU(2)$ transformations.

The one-loop effective action can be computed (by the background-field method; see Sect. 1.3 of Gaillard's lectures), and we obtain (G1.61)

$$\mathcal{L}_{\text{eff}} = \frac{1}{2}\,g_{ij}\,\partial^\mu\pi^i\,\partial_\mu\pi^j\left(1 - \frac{(N-1)}{16\pi^2}\frac{\Lambda^2}{v^2}\right) + \frac{1}{64\pi^2}\,\mathrm{Tr}\left(R^2 + \frac{1}{3}G_{\mu\nu}G^{\mu\nu}\right)\ln\left(\frac{\Lambda^2}{?}\right) + \cdots \tag{8.28}$$

(the question mark is Gaillard's, see below). The first term renormalises the pion fields and the vev,

$$\pi_R = Z\pi, \qquad v_R = Zv, \qquad Z^2 = 1 - \frac{(N-1)}{16\pi^2}\frac{\Lambda^2}{v^2}. \tag{8.29}$$

The second, logarithmically divergent, term involves couplings which are not present at tree level. The argument of the logarithm must be dimensionless, but in the massless σ-model there is no scale parameter with which to scale the cut-off. The situation is similar to QCD, where a Λ_{QCD} emerges after a resummation of infrared divergent

contributions. *Remark.* The analogy should not be taken literally. $\Lambda_{\rm QCD}$ arises by dimensional transmutation, from the renormalisation-group running of α_s; in the massless σ-model the point (Gaillard, Sect. 1.4) is rather that the derivative expansion of the one-loop action is increasingly infrared divergent and has to be resummed, so that the physical momenta ($\partial^2 \to -s, -t, -u$) set the scale in the logarithm, as in (8.31) and in chiral perturbation theory (Gasser–Leutwyler). Since R^2 and G^2 are at least quartic in the scalar fields, the only dimensionful quantity that can appear in the argument of the logarithm is the derivative operator; then (G1.63)

$$\mathcal{L}_{\rm eff} \to \frac{1}{2}\, g_{ij}\, \partial^\mu \pi^i\, \partial_\mu \pi^j\, Z^2 + \frac{1}{64\pi^2} \mathrm{Tr}\left[R \left(\ln \frac{\Lambda^2}{\partial^2} + a \right) R + \frac{1}{3} G_{\mu\nu} \left(\ln \frac{\Lambda^2}{\partial^2} + a' \right) G^{\mu\nu} \right] + O\left(\frac{\partial}{\Lambda} \right), \tag{8.30}$$

where a and a' are unknown constants of order one.

We can go on and compute scattering amplitudes. For $N = 3$ and $a = a' = 0$ one finds for the elastic $\pi^+\pi^-$ scattering amplitude at one loop (G1.64)

$$\mathcal{M}(\pi^+\pi^- \to \pi^+\pi^-) = -\frac{u}{v^2} + \frac{1}{64\pi^2 v^4} \left[3s^2 \ln\left(\frac{\Lambda^2}{-s}\right) + 3t^2 \ln\left(\frac{\Lambda^2}{-t}\right) + 2u^2 \ln\left(\frac{\Lambda^2}{-u}\right) + \cdots \right], \tag{8.31}$$

where the dots stand for similar terms; in Gaillard's paper they read $-\frac{1}{3}t^2 \ln(\Lambda^2/-s) - \frac{1}{3}s^2 \ln(\Lambda^2/-t) + \frac{1}{3}u^2[\ln(\Lambda^2/-s) + \ln(\Lambda^2/-t)]$. Here s, t, u are the usual Mandelstam variables; $s > 0$, $t, u < 0$, and $\ln(\Lambda^2/-s) = \ln(\Lambda^2/s) + i\pi$ contains the absorptive part due to on-shell rescattering.

Recall that the matrix elements for Goldstone boson scattering can be obtained by replacing them with the currents that they are coupled to. Based on this we can derive the "equivalence theorem": S-matrix elements for (W_L, Z_L) scattering can be obtained, up to corrections of $O(m_W/E_W)$ and $O(m_Z/E_Z)$, by replacing $W_L^\pm, Z_L$ with $\pi^\pm, \pi^0$.

Intuitively, a transformation

$$W_\mu \to W'_\mu + \partial_\mu \pi$$

introduces a longitudinal component ($\partial_\mu \pi$) into the gauge field. So we anticipate that, e.g., $W_L W_L \to W_L W_L$ grows with energy, as the amplitude (8.31) does.

How can the divergences of the theory be damped? Let us do the same computation in the Standard Model at $m_H \to \infty$ (forgetting gauge interactions for a moment). We obtain the same result as before, but with Λ being now the renormalisation scale.

Including gauge interactions, the Standard Model effective action at one loop, at $m_H \to \infty$, is (G1.76)

$$\mathcal{L}_{\rm eff}^{\text{1-loop}} = -\frac{\Lambda^2}{16\pi^2} |D_\mu \varphi|^2 + \frac{\ln \Lambda^2}{16\pi^2} \left(a F_{\mu\nu} F^{\mu\nu} + b |D_\mu \varphi|^2 + c \left(\varphi^\dagger D_\mu \varphi \right)^2 + \cdots \right) + \text{finite terms}. \tag{8.32}$$

The c-term is the one which breaks the custodial symmetry, as we now show. In the unitary gauge (G1.78)

$$\left(\varphi^\dagger D_\mu \varphi \right)^2 \Big|_{\pi=0} = -\frac{g^2 v^4}{16 \cos^2 \theta_W} Z_\mu Z^\mu. \tag{8.33}$$

There is thus a divergent term which breaks the $SO(4)$ invariance but preserves $SU(2)\times U(1)$, and which contributes a shift to the Z-boson mass but not to the W mass. It therefore contributes to the parameter (G1.79)

$$\rho - 1 \equiv \frac{m_W^2}{m_Z^2 \cos^2 \theta_W} - 1 = \frac{-3g^2}{64\pi^2} \tan^2 \theta_W \ln\left(\frac{\Lambda^2}{m_W^2}\right) + \text{finite}. \tag{8.34}$$

ρ is well measured and very close to one. This gives a bound $\Lambda \lesssim 3$ TeV. *Remark.* This is the bound of Gaillard's 1987 lectures (with $|\rho - 1| < 0.004$). Today's electroweak fits (αT known to about 10^{-3}) constrain this logarithm much more strongly; before 2012 this was the origin of the indirect upper bound $m_H \lesssim 200$ GeV, and the Higgs boson was found at $m_H \simeq 125$ GeV, so the $m_H \to \infty$ limit is now only a theoretical laboratory (cf. PDG review "Electroweak model and constraints on new physics"). If we put $\Lambda^2 = m_H^2$, we get the leading-logarithm term of the Standard Model calculation.

Experimentally, custodial symmetry cannot be broken significantly; see Chapter 9. The Standard Model with a Higgs boson passes this test, $|\rho - 1|$ being small, but other models must satisfy this as well.

9

Oblique corrections and precision tests

In this chapter we follow M. E. Peskin and T. Takeuchi, *Estimation of oblique electroweak corrections*, Phys. Rev. D **46** (1992) 381.

9.1 Motivation: the ρ parameter and custodial symmetry

We have great experimental evidence that the $SU(2)_L \otimes U(1)_Y$ gauge theory correctly describes the electroweak interactions. However, the Higgs sector, which gives the W and Z bosons their masses, is a mystery. There are many hypotheses and extensions of the Standard Model, and until 2012 none of them, including the minimal SM theory, had been verified by a discovery of the particle or particles which they predict. *Remark.* In July 2012 ATLAS and CMS discovered a Higgs boson, now with $m_h = 125.2\,\text{GeV}$ (PDG 2024); its measured couplings agree with the minimal SM at the 10% level or better. The minimal SM Higgs sector is therefore no longer hypothetical, although its dynamics (the potential, naturalness) remains untested, and precision observables still constrain extended or composite Higgs sectors.

We can test indirectly the Higgs sector with precision measurements of the parameters of the weak interaction. General models of the Higgs sector allow large deviations from the minimal SM. However, as we have seen, the minimal SM has a very successful prediction. At tree level,

$$\rho = \frac{m_W^2}{m_Z^2 \cos^2\theta_W} = 1, \tag{9.1}$$

and it is satisfied to a better than 1% precision. *Remark.* Today the constraint is much stronger: the PDG 2024 global fit gives, for the new-physics part of the ρ parameter, $\rho_0 = 1.00031 \pm 0.00019$ (J. Erler and A. Freitas, PDG review "Electroweak model and constraints on new physics", 2024), i.e. agreement with 1 at the level of a few parts in 10^4. This restricts the possibilities for alternative Higgs sectors. Nonetheless, a large class of models which respect an approximate "custodial" $SU(2)_L \otimes SU(2)_R$ symmetry (approximate, because it becomes exact only when the Yukawa couplings and the gauge coupling of $U(1)_Y$ are taken to zero) predict the same ρ at tree level. Custodial symmetry requires that $\phi^\dagger$ and ϕ are treated on the same footing and can be interchanged. For example:

operator	custodial symmetry
$(D_\mu\phi)^\dagger(D^\mu\phi)$	✓
$(\phi^\dagger\phi)$	✓
$\left\lvert\phi^\dagger D_\mu\phi\right\rvert^2$	×

We also find large deviations from the minimal Standard Model if we extend the $SU(2) \times U(1)$ gauge group. This would require a mixing of the Z boson with new Z' bosons. We shall restrict our search for new physics to models with a custodial symmetry,

$$\rho = \frac{m_W^2}{m_Z^2 \cos^2 \theta_W} = 1 \quad \text{at tree level,} \tag{9.2}$$

and where $SU(2)_L \otimes U(1)_Y$ is the correct group for electroweak interactions up to scales quite a lot above the electroweak scale m_W.

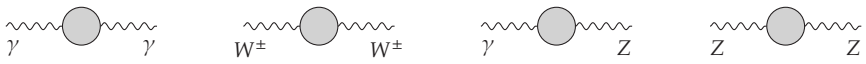
For Higgs sectors with weak interactions, we can employ perturbation theory in order to compute and analyse the impact of radiative corrections to measured parameters. We shall see that we can also analyse the impact of strong sectors, despite not being able to use a perturbative approach, and despite the fact that Goldstone bosons do not arise perturbatively either.

9.2 Formalism of oblique corrections

We consider weak-interaction processes with only light fermions as external particles. For example,

$$e^+e^- \to e^+e^-,\ \mu^+\mu^-,\ b\bar{b}, \qquad q\bar{q} \to e^+e^-,\ \mu^+\mu^-,\ e\bar{\nu}, \qquad \text{etc.} \tag{9.3}$$

These processes involve the propagation of electroweak gauge bosons, with vacuum-polarisation insertions



We find three types of contributions to four-fermion interactions, shown in Fig. 9.1: bubbles (self-energy insertions on the gauge-boson propagator, between bosons V_1 and V_2), triangles (vertex corrections) and boxes.

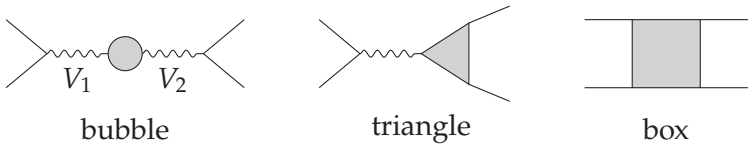


Figure 9.1. The three types of corrections to four-fermion processes with light external fermions.

We are interested in contributions due to new physics. Given that the external particles are light, the relative contributions of the three components scale as

$$\text{(triangle)} \sim \text{(bubble)} \times \frac{1}{M_{NP}^2}, \qquad \text{(box)} \sim \text{(bubble)} \times \frac{1}{M_{NP}^4}, \tag{9.4}$$

where M_{NP} is the mass scale of the new physics. *Remark.* As written, these relations are not dimensionally consistent; they are a mnemonic. The precise argument of Peskin and Takeuchi is that the new particles are assumed to couple to the light external fermions only through the gauge bosons (or with Yukawa strength $\propto m_f/v$), so that new-physics vertex and box corrections are suppressed by powers of m_f^2/m_W^2 or q^2/M_{NP}^2, whereas the vacuum polarisations $\Pi_{XY}(0)$, $\Pi'_{XY}(0)$ need not decouple (e.g. for chiral fermions getting mass from the Higgs). The $Zb\bar{b}$ vertex is the standard exception, for new physics coupling to the third generation.

The dominant new-physics corrections thus arise from self-energy corrections to the propagators. These corrections are called "oblique" for reasons that we shall explain shortly.

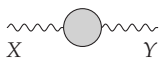
9.2.1 Notation

The electroweak currents are J^μ_Q (electromagnetic) and J^μ_3, $J^\mu_\pm$ (weak isospin), and

$$\mathcal{L} \ni \frac{e}{\sqrt{2}\,s}\left(W^+_\mu J^\mu_+ + W^-_\mu J^\mu_-\right) + \frac{e}{sc}\,Z_\mu\left(J^\mu_3 - s^2 J^\mu_Q\right) + eA_\mu J^\mu_Q, \tag{9.5}$$

with $s \equiv \sin\theta_W$, $c \equiv \cos\theta_W$.

In a four-fermion process, the propagator



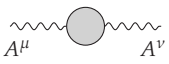
$$= G_{XY}\,g^{\mu\nu} + \left(p^\mu p^\nu \text{ terms}\right) \tag{9.6}$$

contributes to the amplitude only through its $g^{\mu\nu}$ coefficient. Indeed, the $p^\mu p^\nu$ terms are contracted with the light-fermion lines; for an incoming pair with (massless) momenta p_A, p_B, $p = p_A + p_B$, the fermion trace gives

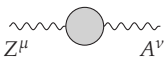
$$(p_A + p_B)_\mu\,(p_A + p_B)_\nu\ \mathrm{tr}\left(\not{p}_A\gamma^\mu\ \not{p}_B\gamma^\nu\right) \to 0, \tag{9.7}$$

since $\not{p}_A\ \not{p}_A = p_A^2 = 0$ and $\not{p}_B\ \not{p}_B = p_B^2 = 0$.

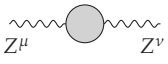
We have four propagators (when including higher orders):



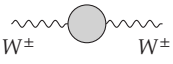
$$= G_{AA}\,g^{\mu\nu} + \ldots,$$



$$= G_{ZA}\,g^{\mu\nu} + \ldots,$$



$$= G_{ZZ}\,g^{\mu\nu} + \ldots,$$



$$= G_{WW}\,g^{\mu\nu} + \ldots. \tag{9.8}$$

We can compute the matrix elements of the charged- and neutral-current interactions:

$$\begin{aligned}\mathcal{M}_{NC} = {} & e^2 QQ'\,G_{AA} + \frac{e^2}{sc}\,G_{ZA}\left[Q\left(I'_3 - s^2Q'\right) + \left(I_3 - s^2Q\right)Q'\right] \\ & + \frac{e^2}{s^2c^2}\left(I_3 - s^2Q\right)\left(I'_3 - s^2Q'\right)G_{ZZ},\end{aligned} \tag{9.9}$$

$$\mathcal{M}_{CC} = \frac{e^2}{2s^2}\,I_+I_-\,G_{WW}. \tag{9.10}$$

(I_3, Q) and (I'_3, Q') are the weak isospin and the electric charge of the two fermion lines, and $I_\pm$ are weak-isospin raising and lowering matrices.

At leading order in perturbation theory

$$G^{(LO)}_{ZA} = 0 \equiv D_{ZA} \tag{9.11}$$

(there is no tree-level γ–Z transition, but at one loop, e.g. through an e^+e^- loop, the γ–Z mixing is non-zero), and

$$D_{BB} = G^{(LO)}_{BB} = \frac{1}{q^2 - m^2_{0B}}, \qquad B = A, Z, W^\pm, \tag{9.12}$$

with

$$m^2_{0A} = 0, \qquad m^2_{0Z} = \frac{e^2}{s^2c^2}\frac{v^2}{4}, \qquad m^2_{0W} = \frac{e^2}{s^2}\frac{v^2}{4} \quad \left(m^2_{0W} = m^2_{0Z}\,c^2\right). \tag{9.13}$$

As we discussed, the leading modifications due to new physics arise from vacuum-polarisation effects. These change the values of G_{XY} but do not change the form of the $\mathcal{M}_{NC}$ and $\mathcal{M}_{CC}$ amplitudes. This is the reason that they are called "oblique", unlike triangle and box corrections which change also the form of the interactions.

9.2.2 Vacuum-polarisation amplitudes

Let us define the one-particle-irreducible (1PI) vacuum-polarisation amplitudes

$$\int d^4y\, e^{-iq\cdot y}\, \left\langle J_X^\mu(y) J_Y^\nu(0)\right\rangle = i g^{\mu\nu}\, \Pi_{XY}(q^2) + \left(q^\mu q^\nu \text{ terms}\right), \tag{9.14}$$

with

$$(XY) \equiv (11),\ (22),\ (33),\ (3Q),\ (QQ). \tag{9.15}$$

We also define

$$\Pi_{XY}(q^2) = \Pi_{XY}(0) + q^2\, \Pi'_{XY}(q^2). \tag{9.16}$$

We can then express the 1PI self-energies of the gauge bosons as



$$= ie^2\, \Pi_{QQ}\, g^{\mu\nu} + \ldots, \tag{9.17}$$



$$= i\frac{e^2}{cs}\left(\Pi_{3Q} - s^2\Pi_{QQ}\right) g^{\mu\nu} + \ldots, \tag{9.18}$$



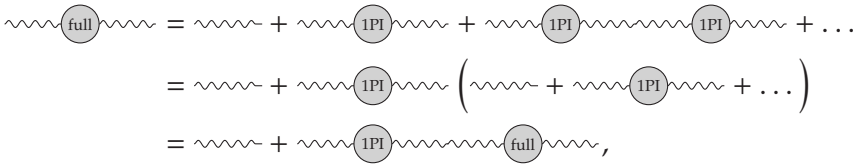
$$= i\frac{e^2}{c^2s^2}\left(\Pi_{33} - 2s^2\Pi_{3Q} + s^4\Pi_{QQ}\right) g^{\mu\nu} + \ldots, \tag{9.19}$$



$$= i\frac{e^2}{s^2}\, \Pi_{11}\, g^{\mu\nu} + \ldots, \tag{9.20}$$

and we denote the corresponding coefficients of $ig^{\mu\nu}$ by Π_{AA}, Π_{ZA}, Π_{ZZ} and Π_{WW} respectively.

Recall that we can perform a Dyson resummation of 1PI graphs:



(9.21)

i.e. (with an implicit sum over the intermediate bosons E, F)

$$\underbrace{G_{CD}}_{\text{full}} = \underbrace{D_{CD}}_{\text{tree}} + \underbrace{D_{CE}}_{\text{tree}}\ \underbrace{\Pi_{EF}}_{\text{1PI}}\ \underbrace{G_{FD}}_{\text{full}}. \tag{9.22}$$

After Dyson resummation, the NC and CC amplitudes become

$$\mathcal{M}_{NC} = \frac{e^2QQ'}{q^2 - \Pi_{AA}} + \frac{e^2}{s^2c^2}\, \frac{\left\{I_3 - \left[s^2 - sc\,\frac{\Pi_{ZA}}{q^2-\Pi_{AA}}\right] Q\right\}\left\{I'_3 - \left[s^2 - sc\,\frac{\Pi_{ZA}}{q^2-\Pi_{AA}}\right] Q'\right\}}{q^2 - m_{0Z}^2 - \Pi_{ZZ} - \frac{(\Pi_{ZA})^2}{q^2-\Pi_{AA}}}, \tag{9.23}$$

$$\mathcal{M}_{CC} = \frac{e^2}{2s^2}\, \frac{I_+I_-}{q^2 - m_{0W}^2 - \Pi_{WW}}. \tag{9.24}$$

9.3 Starred couplings

Let us define the "effective" electric charge and the "effective" weak mixing angle:

$$e_*^2(q^2) \equiv \frac{e^2}{1 - e^2\Pi'_{QQ}(q^2)} = e^2\left(1 + e^2\Pi'_{QQ}(q^2)\right), \tag{9.25}$$

$$s_*^2(q^2) = s^2 - sc\,\frac{\Pi_{ZA}(q^2)}{q^2 - \Pi_{AA}(q^2)} = s^2 - e^2\left(\Pi'_{3Q}(q^2) - s^2\Pi'_{QQ}(q^2)\right). \tag{9.26}$$

It is fine to linearise, since we are dealing with typically small corrections. With these redefinitions of the couplings we cast the $\mathcal{M}_{NC}$ and $\mathcal{M}_{CC}$ amplitudes as

$$\mathcal{M}_{NC} = e_*^2\, Q\, \frac{1}{q^2}\, Q' + \frac{e^2}{s^2c^2}\left(I_3 - s_*^2 Q\right)\left(I_3' - s_*^2 Q'\right) \times \frac{1}{q^2 - \frac{e^2}{s^2c^2}\left[\frac{v^2}{4} + \left(\Pi_{33} - 2s^2\Pi_{3Q} + s^4\Pi_{QQ}\right)\right]}, \tag{9.27}$$

$$\mathcal{M}_{CC} = \frac{e^2}{2s^2}\, I_+ \frac{1}{q^2 - \frac{e^2}{s^2}\left[\frac{v^2}{4} + \Pi_{11}\right]}\, I_-, \tag{9.28}$$

which maintain the same form as the Born amplitudes, and the radiative corrections are accounted for by redefining s_θ^2, e^2 and m_{0Z}^2.

9.3.1 Renormalisation

These redefinitions involve infinite quantities and we must now renormalise. The W and Z masses are the poles of their respective propagators:

$$m_Z^2 = \frac{e^2}{s^2c^2}\frac{v^2}{4} + \frac{e^2}{s^2c^2}\left(\Pi_{33} - 2s^2\Pi_{3Q} + s^4\Pi_{QQ}\right)\Big|_{q^2=m_Z^2}, \tag{9.29}$$

$$m_W^2 = \frac{e^2}{s^2}\frac{v^2}{4} + \frac{e^2}{s^2}\,\Pi_{11}(m_W^2). \tag{9.30}$$

The wave-function renormalisation constants Z_Z, Z_W are the residues of the poles in the Z and W propagators:

$$Z_Z^{-1} = 1 - \frac{e^2}{s^2c^2}\frac{d}{dq^2}\left(\Pi_{33} - 2s^2\Pi_{3Q} + s^4\Pi_{QQ}\right)\Big|_{q^2=m_Z^2}, \tag{9.31}$$

$$Z_W^{-1} = 1 - \frac{e^2}{s^2}\frac{d}{dq^2}\Pi_{11}\Big|_{q^2=m_W^2}. \tag{9.32}$$

This completes our renormalisation procedure, which expresses the amplitudes in terms of the physical pole masses.

However, we are allowed to use other finite quantities to express our amplitudes. We can use a running-mass scheme and define

$$\frac{1}{q^2 - \frac{e^2}{s^2}\left[\frac{v^2}{4} + \Pi_{11}\right]} = \frac{Z_W}{q^2 - M_{W*}^2(q^2)}, \tag{9.33}$$

$$\frac{1}{q^2 - \frac{e^2}{s^2c^2}\left[\frac{v^2}{4} + \left(\Pi_{33}(q^2) - 2s^2\Pi_{3Q}(q^2) + s^4\Pi_{QQ}(q^2)\right)\right]} = \frac{Z_Z}{q^2 - M_{Z*}^2(q^2)}. \tag{9.34}$$

Obviously, at $q^2 = m_W^2$, m_Z^2 we have

$$M_{Z*}^2(m_Z^2) = m_Z^2, \quad \frac{d}{dq^2}M_{Z*}^2\Big|_{q^2=m_Z^2} = 0, \qquad M_{W*}^2(m_W^2) = m_W^2, \quad \frac{d}{dq^2}M_{W*}^2\Big|_{q^2=m_W^2} = 0. \tag{9.35}$$

Finally, we define

$$\frac{e_*^2}{s_*^2c_*^2}\,Z_{Z*} \equiv \frac{e^2}{s^2c^2}\,Z_Z, \qquad \frac{e_*^2}{s_*^2}\,Z_{W*} \equiv \frac{e^2}{s^2}\,Z_W, \qquad c_*^2 = 1 - s_*^2. \tag{9.36}$$

Explicitly,

$$\begin{aligned} Z_{Z*} &= Z_Z \left[1 - e^2\Pi'_{QQ}(q^2) - \frac{e^2(c^2-s^2)}{s^2c^2}\left(\Pi'_{3Q}(q^2) - s^2\Pi'_{QQ}(q^2)\right)\right] \\ &= 1 + \frac{e^2}{s^2c^2}\frac{d}{dq^2}\left(\Pi_{33} - 2s^2\Pi_{3Q} + s^4\Pi_{QQ}\right)\Big|_{q^2=m_Z^2} - \frac{e^2s^2}{c^2}\Pi'_{QQ}(q^2) - \frac{e^2(c^2-s^2)}{s^2c^2}\Pi'_{3Q}(q^2), \end{aligned} \tag{9.37}$$

$$\begin{aligned} Z_{W*} &= Z_W \left[1 - e^2\Pi'_{QQ}(q^2) - \frac{e^2}{s^2}\left(\Pi'_{3Q}(q^2) - s^2\Pi'_{QQ}(q^2)\right)\right] \\ &= 1 + \frac{e^2}{s^2}\frac{d}{dq^2}\Pi_{11}\Big|_{q^2=m_W^2} - \frac{e^2}{s^2}\Pi'_{3Q}(q^2). \end{aligned} \tag{9.38}$$

Finally,

$$\mathcal{M}_{NC} = e_*^2\, Q\, \frac{1}{q^2}\, Q' + \frac{e_*^2}{s_*^2c_*^2}\left(I_3 - s_*^2Q\right)\frac{Z_{Z*}}{q^2 - M_{Z*}^2}\left(I'_3 - s_*^2Q'\right), \tag{9.39}$$

$$\mathcal{M}_{CC} = \frac{e_*^2}{2s_*^2}\, I_+ \frac{Z_{W*}}{q^2 - M_{W*}^2}\, I_-. \tag{9.40}$$

All starred quantities are finite but depend on the energy (q^2). We find that the oblique corrections are fully accounted for if in the tree-level expressions we substitute bare quantities with starred quantities: the photon exchange with couplings Qe_*, $Q'e_*$ and propagator $1/q^2$, plus the Z exchange with couplings $\frac{e_*}{s_*c_*}(I_3 - s_*^2Q)$, $\frac{e_*}{s_*c_*}(I'_3 - s_*^2Q')$ and propagator $Z_{Z*}/(q^2 - M_{Z*}^2)$, reproduce

$$(\text{tree } \gamma) + (\text{tree } Z) + (\gamma\gamma \text{ bubble}) + (\gamma Z \text{ bubble}) + (ZZ \text{ bubble}), \tag{9.41}$$

and the W exchange with couplings $\frac{e_*}{\sqrt{2}s_*}I_+$, $\frac{e_*}{\sqrt{2}s_*}I_-$ and propagator $Z_{W*}/(q^2 - M_{W*}^2)$ reproduces the tree W exchange plus the WW bubble.

9.3.2 Examples of observables

Oblique corrections due to new physics can affect many measurements. For example, the width of the Z boson is

$$\Gamma_Z = Z_{Z*}\frac{\alpha_*\, m_Z}{6s_*^2c_*^2}\sum_f \left[\left(I_{3f} - s_*^2Q_f\right)^2 + \left(s_*^2Q_f\right)^2\right] N_f\Big|_{q^2=m_Z^2}, \tag{9.42}$$

$$N_f = N_{\text{col}}\left[1 + \frac{\alpha_s}{\pi} + \ldots\right] \text{ (quarks)}, \qquad N_f = 1 \text{ (leptons)}, \tag{9.43}$$

and the left-right asymmetry is

$$A_{LR}(q^2) = \frac{2\left[1 - 4s_*^2(q^2)\right]}{1 + \left[1 - 4s_*^2(q^2)\right]^2}, \tag{9.44}$$

and so on. Note that in the SM oblique (vacuum-polarisation) corrections are not always dominant. This is especially manifest for observables where real photon radiation must be accounted for. However, new-physics contributions to the oblique corrections are dominant, with non-oblique corrections suppressed by powers of $1/M_{NP}$. We should always check if our new models introduce big oblique corrections which may cause disagreements with measurements such as those of Γ_Z, Γ_W, A_{LR}, etc.

9.3.3 Low energies and the ρ parameter

At low energies, $q^2 = 0$, we obtain

$$\mathcal{M}_{NC}^{(Z)} = -\frac{\left[I_3 - s_*^2(0)Q\right]\left[I_3' - s_*^2(0)Q'\right]}{\frac{v^2}{4} + \Pi_{33}(0)}, \tag{9.45}$$

$$\mathcal{M}_{CC} = -\frac{1}{2} I_+ \frac{1}{\frac{v^2}{4} + \Pi_{11}(0)} I_-. \tag{9.46}$$

The same amplitudes can be obtained by using Fermi's effective theory,

$$\mathcal{L}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} \left\{ J_\mu^+ J^{-\mu} + \rho_*(0) \left[J_3^\mu - s_*^2(0) J_Q^\mu \right]^2 \right\}. \tag{9.47}$$

This yields

$$\frac{1}{4\sqrt{2}\, G_F} = \frac{v^2}{4} + \Pi_{11}(0), \tag{9.48}$$

$$\frac{1}{\rho_*(0)} = \frac{\frac{v^2}{4} + \Pi_{33}(0)}{\frac{v^2}{4} + \Pi_{11}(0)} = 1 - 4\sqrt{2}\, G_F \left(\Pi_{11}(0) - \Pi_{33}(0)\right). \tag{9.49}$$

9.3.4 Summary

- Oblique electroweak corrections may be computed from leading-order matrix elements by substituting $s \to s_*$, $m_Z \to M_{Z*}$, $m_W \to M_{W*}$, etc.
- The dependence of the starred quantities on the vacuum-polarisation graphs is relatively simple. It is easy to see what the effect of new physics is on measured quantities.
- New physics may affect measurements at the lowest energies.

9.4 The S, T, U parameters

$\mathcal{M}_{CC}$ and $\mathcal{M}_{NC}$ depend on the starred quantities, and these, even at lowest order, depend on three parameters,

$$s_\theta, \quad v, \quad e. \tag{9.50}$$

We need at least three measurements to constrain them:

$$\alpha^{-1} = 137.035989\ldots, \qquad G_F = 1.16637 \times 10^{-5}\,\text{GeV}^{-2}, \qquad m_Z = 91.17\,\text{GeV}. \tag{9.51}$$

Remark. Current values (CODATA 2022, PDG 2024): $\alpha^{-1} = 137.035999177(21)$, $G_F = 1.1663788(6) \times 10^{-5}\,\text{GeV}^{-2}$, $m_Z = 91.1880(20)\,\text{GeV}$.

All starred quantities for new physics with a scale $q^2 \sim m_Z^2 \ll M_{NP}^2$ can be expressed linearly through the S, T and U parameters:

$$\alpha S = 4e^2 \left[\Pi_{33}'(0) - \Pi_{3Q}'(0)\right], \tag{9.52}$$

$$\alpha T = \frac{e^2}{s^2 c^2 m_Z^2} \left(\Pi_{11}(0) - \Pi_{33}(0)\right), \tag{9.53}$$

$$\alpha U = 4e^2 \left[\Pi_{11}'(0) - \Pi_{33}'(0)\right]. \tag{9.54}$$

This is because for $q^2 \to 0$

$$\Pi_{QQ}(q^2) = q^2\,\Pi'_{QQ}(0), \qquad \Pi_{3Q}(q^2) = q^2\,\Pi'_{3Q}(0), \tag{9.55}$$

$$\Pi_{33}(q^2) = \Pi_{33}(0) + q^2\,\Pi'_{33}(0), \qquad \Pi_{11}(q^2) = \Pi_{11}(0) + q^2\,\Pi'_{11}(0). \tag{9.56}$$

We get, for example,

$$\rho_*(0) = 1 + \alpha T, \tag{9.57}$$

$$Z_{Z*}(q^2) - 1 = \frac{\alpha}{4s^2c^2}\,S, \tag{9.58}$$

$$Z_{W*}(q^2) - 1 = \frac{\alpha}{4s^2}\,(S+U)\,. \tag{9.59}$$

In the above, we can assume that $\Pi_{11} \simeq \Pi_{33}$ due to custodial symmetry. So $U \sim 0$. *Remark.* The argument needs refinement: exact custodial symmetry would give $\Pi_{11} = \Pi_{33}$ at all q^2, hence $T = 0$ as well, yet T is the parameter most sensitive to custodial breaking. The reason U is small is that in an EFT expansion it first arises from a dimension-8 operator, so it is suppressed relative to T by an extra m_Z^2/M_{NP}^2; compare the heavy doublet below, where $U/T \sim m_Z^2/m_N^2$. This is why fits are often quoted with $U = 0$ (PDG 2024).

Suppose that the SM prediction for the W mass is known very precisely. Then

$$m_W^2 = m_W^2(\text{ref}) + \frac{\alpha c^2}{c^2 - s^2}\,m_Z^2\left[-\frac{1}{2}S + c^2 T + \frac{c^2 - s^2}{4s^2}\,U\right], \tag{9.60}$$

where the left-hand side is the W mass with new physics and $m_W^2(\text{ref})$ is the precise SM calculation; and so on.

The data constrain the parameters as follows:

- $s_*^2 \leftarrow Z$-pole asymmetries;
- $S \leftarrow m_Z$; $T \leftarrow \Gamma_Z$, α_s; $m_t \leftarrow$ direct measurements (CDF and D0 at the Tevatron);
- $U \leftarrow m_W$.

Remark. The top mass is now dominated by LHC measurements, $m_t = 172.57 \pm 0.29\,\text{GeV}$ (PDG 2024), and the Higgs mass $m_h = 125.2\,\text{GeV}$ is also a fit input; before 2012 m_h was itself a fit parameter.

9.5 Examples

9.5.1 A heavy fermion doublet

Let us consider a new heavy fermion doublet (N, E) with the usual couplings to $SU(2)_L \times U(1)_Y$ and masses m_N, m_E. For $m_N, m_E \gg m_Z$,

$$S = \frac{1}{6\pi}\left[1 - Y\ln\frac{m_N^2}{m_E^2}\right] \quad (Q = I_3 + Y/2), \tag{9.61}$$

$$T = \frac{1}{16\pi s^2c^2m_Z^2}\left[m_N^2 + m_E^2 - \frac{2m_N^2m_E^2}{m_N^2 - m_E^2}\ln\frac{m_N^2}{m_E^2}\right], \tag{9.62}$$

$$U = \frac{1}{6\pi}\left[-\frac{5m_N^4 - 22m_N^2m_E^2 + 5m_E^4}{3\left(m_N^2 - m_E^2\right)^2} + \frac{m_N^6 - 3m_N^4m_E^2 - 3m_N^2m_E^4 + m_E^6}{\left(m_N^2 - m_E^2\right)^3}\ln\frac{m_N^2}{m_E^2}\right]. \tag{9.63}$$

For $\Delta m = |m_N - m_E| \ll m_N, m_E$:

$$S = \frac{1}{6\pi} \quad \left(\text{for quarks} \to \frac{N_c}{6\pi}\right), \qquad T = \frac{1}{12\pi s^2 c^2}\frac{(\Delta m)^2}{m_Z^2}, \qquad U = \frac{2}{15\pi}\frac{(\Delta m)^2}{m_N^2}. \tag{9.64}$$

U is suppressed (relative to T) by m_Z^2/m_N^2, i.e. by the new-physics scale.

The global fit to the data gives (PDG 2010) for the new-physics contributions

$$S = 0.01 \pm 0.10, \qquad T = 0.03 \pm 0.11, \qquad U = 0.06 \pm 0.10. \tag{9.65}$$

Remark. The PDG 2024 fit (reference SM with the measured m_h and m_t) gives $S = -0.04 \pm 0.10$, $T = 0.01 \pm 0.12$, $U = -0.01 \pm 0.09$, and with $U = 0$: $S = -0.05 \pm 0.07$, $T = 0.00 \pm 0.06$ (strongly correlated). Including the CDF m_W measurement would shift T to ≈ 0.07. A full degenerate fourth family ($S \approx 2/(3\pi) \approx 0.21$) is thus strongly disfavoured. The allowed region in the S–T plane is sketched in Fig. 9.2.

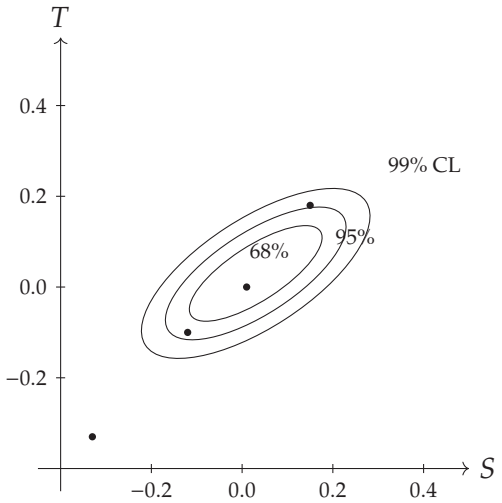


Figure 9.2. Sketch of the experimentally allowed region in the S–T plane (68%, 95% and 99% CL contours, schematic).

9.5.2 The Higgs boson

Let us now look at the contribution of a Higgs boson, through the diagrams of Fig. 9.3. One finds

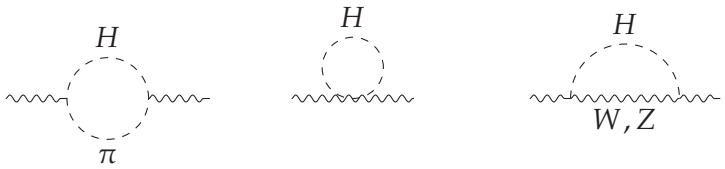


Figure 9.3. Higgs-boson contributions to the W and Z vacuum polarisation (π: Goldstone boson).

$$S \simeq \frac{1}{12\pi}\ln\frac{m_H^2}{m_{H,\mathrm{ref}}^2}, \qquad T \simeq -\frac{3}{16\pi c^2}\ln\frac{m_H^2}{m_{H,\mathrm{ref}}^2}, \qquad U \simeq 0. \tag{9.66}$$

Remark. These are leading-logarithmic results for $m_H \gg m_Z$, relative to a reference mass. Before 2012 they were used to bound m_H indirectly (the fits preferred $m_H \lesssim 160\,\mathrm{GeV}$);

since the discovery, $m_h = 125.2\,\text{GeV}$ fixes the SM reference point, and a heavy Higgs-like state in an extended sector would show up as $S > 0$, $T < 0$ in the way shown here.

10

Little Higgs models

This chapter was written by Giuliano Panico.

We follow M. Schmaltz and D. Tucker-Smith, *Little Higgs review*, hep-ph/0502182, and M. Schmaltz, *The simplest little Higgs*, hep-ph/0407143.

The basic idea of little Higgs models is that the Higgs doublet arises as a pseudo-Goldstone boson in the breaking of a larger, continuous *global* symmetry (not a gauge symmetry). Recall:

- a *Goldstone boson* is exactly massless and has only derivative couplings; it is associated with the breaking of an exact symmetry;
- a *pseudo-Goldstone boson* is associated with the breaking of an approximate symmetry; it behaves similarly to a Goldstone boson, with (approximately) only derivative couplings.

Remark. More precisely, a pseudo-Goldstone boson is associated with a symmetry that is only approximate: it acquires a mass and non-derivative interactions (a potential), both proportional to the small explicit breaking, and only the symmetric part of its interactions is purely derivative. In little Higgs models this radiatively generated potential is the Higgs potential (cf. the pion mass, Chapter 5).

10.1 Goldstone bosons

10.1.1 The $U(1)$ Goldstone boson

Consider

$$\mathcal{L} = (\partial_\mu \phi)^* (\partial^\mu \phi) - V(\phi^* \phi), \tag{10.1}$$

invariant under $U(1)$: $\phi \to e^{i\alpha}\phi$. If ϕ has a vev f, we write

$$\phi(x) = \frac{1}{\sqrt{2}} \big(f + r(x)\big)\, e^{i\theta(x)/f}, \tag{10.2}$$

with $r(x)$ the radial mode and $\theta(x)$ the Goldstone-boson mode. Under the $U(1)$ the Goldstone field shifts,

$$\theta \to \theta + \alpha f. \tag{10.3}$$

After integrating out $r(x)$, $\mathcal{L}_{\text{eff}}$ contains no mass term for $\theta(x)$, and only derivative couplings.

10.1.2 $SU(N) \to SU(N-1)$

Consider global symmetry breaking $SU(N) \to SU(N-1)$. There is one Goldstone boson per broken generator,

$$(N^2-1) - \left[(N-1)^2 - 1\right] = 2N-1, \tag{10.4}$$

The breaking is due to the vev of a field ϕ in the fundamental representation of $SU(N)$. We parametrise

$$\phi = \exp\left\{\frac{i}{f}\begin{pmatrix} -\frac{\pi^0}{\sqrt{2}(N-1)}\mathbf{1}_{N-1} & \begin{matrix}\pi_1\\ \vdots\\ \pi_{N-1}\end{matrix} \\ \begin{matrix}\pi_1^* & \cdots & \pi_{N-1}^*\end{matrix} & \pi^0/\sqrt{2}\end{pmatrix}\right\}\begin{pmatrix}0\\ \vdots\\ 0\\ f\end{pmatrix} = e^{i\Pi/f}\,\phi_0. \tag{10.5}$$

The Goldstone bosons are π^0 (real) and $\vec{\pi} = (\pi_1, \ldots, \pi_{N-1})$ (complex).

Transformation properties of the Goldstone bosons:

a) Under the unbroken symmetries $SU(N-1)$,

$$U_{N-1} = \begin{pmatrix}\hat{U}_{N-1} & 0\\ 0 & 1\end{pmatrix}, \tag{10.6}$$

we have

$$\phi \to U_{N-1}\phi = \left(U_{N-1}e^{i\Pi/f}U_{N-1}^\dagger\right)U_{N-1}\phi_0 = \exp\left(\frac{i}{f}\,U_{N-1}\Pi U_{N-1}^\dagger\right)\phi_0, \tag{10.7}$$

since $U_{N-1}\phi_0 = \phi_0$. Therefore

$$\Pi \to U_{N-1}\,\Pi\,U_{N-1}^\dagger : \tag{10.8}$$

$\pi^0 \to \pi^0$ is a singlet, and

$$\begin{pmatrix}0 & \vec{\pi}\\ \vec{\pi}^\dagger & 0\end{pmatrix} \to U_{N-1}\begin{pmatrix}0 & \vec{\pi}\\ \vec{\pi}^\dagger & 0\end{pmatrix}U_{N-1}^\dagger = \begin{pmatrix}0 & \hat{U}_{N-1}\vec{\pi}\\ \vec{\pi}^\dagger\hat{U}_{N-1}^\dagger & 0\end{pmatrix}, \tag{10.9}$$

such that $\vec{\pi} \to \hat{U}_{N-1}\vec{\pi}$ transforms in the fundamental representation of $SU(N-1)$.

b) Under the broken symmetries (the coset $SU(N)/SU(N-1)$),

$$\begin{aligned}\phi \to U e^{i\Pi/f}\phi_0 &= \exp\left\{i\begin{pmatrix}0 & \vec{\alpha}\\ \vec{\alpha}^\dagger & 0\end{pmatrix}\right\}\exp\left\{\frac{i}{f}\begin{pmatrix}0 & \vec{\pi}\\ \vec{\pi}^\dagger & 0\end{pmatrix}\right\}\phi_0\\ &= \exp\left\{\frac{i}{f}\begin{pmatrix}0 & \vec{\pi}'\\ \vec{\pi}'^\dagger & 0\end{pmatrix}\right\}\underbrace{U_{N-1}(\vec{\alpha},\vec{\pi})\,\phi_0}_{=\phi_0} = \exp\left\{\frac{i}{f}\begin{pmatrix}0 & \vec{\pi}'\\ \vec{\pi}'^\dagger & 0\end{pmatrix}\right\}\phi_0,\end{aligned} \tag{10.10}$$

with $\vec{\pi}' = \vec{\pi}'(\vec{\pi},\vec{\alpha})$; to linear order $\vec{\pi}' = \vec{\pi} + \vec{\alpha}$ (with $\vec{\alpha}$ measured in units of f). Again, $\vec{\pi}$ is massless, with only derivative interactions.

10.1.3 Effective Lagrangian

The effective Lagrangian is the most general $SU(N)$-invariant function of $\phi = e^{i\Pi/f}\phi_0$, expanded in ∂_μ/Λ:

$$\mathcal{L} = \text{const} + |\partial_\mu\phi|^2 + \mathcal{O}\left(\frac{1}{\Lambda}\right). \tag{10.11}$$

Without derivatives it can contain only $SU(N)$ singlets:

$$\mathbf{1} \in N \otimes \bar{N} \;\to\; \phi^\dagger \phi = f^2, \tag{10.12}$$

$$\mathbf{1} \in \underbrace{N \otimes N \otimes \cdots \otimes N}_{N \text{ times}} \;\to\; \epsilon^{i_1 \cdots i_N} \phi_{i_1} \cdots \phi_{i_N}, \tag{10.13}$$

both of which are independent of the Goldstone fields (the second in fact vanishes for a single ϕ).

10.2 Introducing gauge interactions

Consider the simplest case, $SU(3) \to SU(2)$: $\dim SU(3) = 8$, $\dim SU(2) = 3$, hence five Goldstone bosons,

$$\Pi = \begin{pmatrix} -\eta/2 & 0 & \\ 0 & -\eta/2 & h \\ & h^\dagger & \eta \end{pmatrix}, \tag{10.14}$$

with h an $SU(2)$ doublet and η an $SU(2)$ singlet. There is still a global $SU(2)$. Now:

- coupling h to $SU(2)$ gauge bosons breaks $SU(3)$ (explicitly);
- gauging $SU(3)$ instead, h and η are absorbed into the longitudinal degrees of freedom of the $SU(3)/SU(2)$ gauge bosons.

The way out: gauge $SU(3)$ and introduce two copies of ϕ triplets, with aligned vevs $f \simeq O(\text{TeV})$,

$$\phi_1 = e^{i\Pi_1/f} \begin{pmatrix} 0 \\ 0 \\ f \end{pmatrix}, \qquad \phi_2 = e^{i\Pi_2/f} \begin{pmatrix} 0 \\ 0 \\ f \end{pmatrix}, \tag{10.15}$$

with kinetic Lagrangian

$$\mathcal{L} = |D_\mu \phi_1|^2 + |D_\mu \phi_2|^2. \tag{10.16}$$

Only one linear combination of Π_1 and Π_2 is absorbed into the $SU(3)/SU(2)$ gauge bosons. A convenient parametrisation is

$$\begin{aligned} \phi_1 &= \underbrace{\exp\left(\frac{i}{f}\begin{pmatrix} & & k \\ & & \\ k^\dagger & & \end{pmatrix}\right)}_{\text{GB of } SU(3)/SU(2)} \underbrace{\exp\left(\frac{i}{f}\begin{pmatrix} & & h \\ & & \\ h^\dagger & & \end{pmatrix}\right)}_{\text{pseudo-GB}} \begin{pmatrix} 0 \\ 0 \\ f \end{pmatrix}, \\ \phi_2 &= \exp\left(\frac{i}{f}\begin{pmatrix} & & k \\ & & \\ k^\dagger & & \end{pmatrix}\right) \exp\left(-\frac{i}{f}\begin{pmatrix} & & h \\ & & \\ h^\dagger & & \end{pmatrix}\right) \begin{pmatrix} 0 \\ 0 \\ f \end{pmatrix}. \end{aligned} \tag{10.17}$$

The first factor, common to ϕ_1 and ϕ_2, contains the Goldstone bosons k eaten by the $SU(3)/SU(2)$ gauge bosons; the second contains the pseudo-Goldstone boson h, which enters ϕ_1 and ϕ_2 with opposite signs.

10.2.1 Radiative Higgs potential

The Higgs mass and the Higgs potential are generated from radiative corrections (gauge loops). The quadratically divergent one-loop diagrams with ϕ_1 external legs give (with a cut-off Λ)

$$\frac{g^2}{16\pi^2} \Lambda^2 \phi_1^\dagger \phi_1 = \frac{g^2}{16\pi^2} \Lambda^2 f^2, \tag{10.18}$$

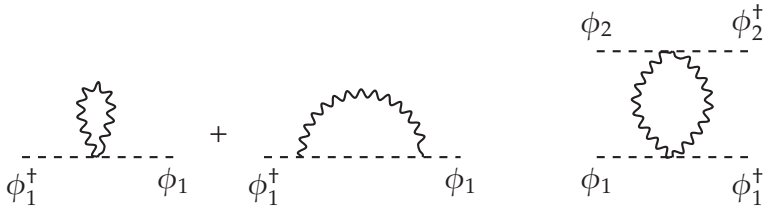


Figure 10.1. Gauge-boson loops in the two-triplet model. Left: quadratically divergent contributions with two external ϕ_1 legs, proportional to $\phi_1^\dagger\phi_1 = f^2$. Right: the logarithmically divergent diagram involving both triplets (two seagull vertices joined by two gauge propagators), proportional to $|\phi_1^\dagger\phi_2|^2$.

which does not involve h, and only corrects the vacuum energy; the same holds for ϕ_2.

The diagram which involves both triplets (Fig. 10.1, right) is only logarithmically divergent,

$$\frac{g^4}{16\pi^2}\log\left(\frac{\Lambda^2}{f^2}\right)\left|\phi_1^\dagger\phi_2\right|^2. \tag{10.19}$$

Using

$$\begin{aligned}\phi_1^\dagger\phi_2 &= \begin{pmatrix}0 & f\end{pmatrix}\exp\left\{-\frac{2i}{f}\begin{pmatrix} & h\\ h^\dagger & \end{pmatrix}\right\}\begin{pmatrix}0\\ f\end{pmatrix}\\ &= \left(f^2\mathbf{1} - 2if\begin{pmatrix} & h\\ h^\dagger & \end{pmatrix} - 2\begin{pmatrix}hh^\dagger & \\ & h^\dagger h\end{pmatrix} + \cdots\right)_{33}\\ &= f^2 - 2h^\dagger h + \cdots,\end{aligned} \tag{10.20}$$

where the common factor containing k has cancelled, we see that the logarithmically divergent term generates a mass term for h,

$$m_h^2 \sim \frac{g^4}{16\pi^2}\log\left(\frac{\Lambda^2}{f^2}\right)f^2 \sim \frac{g^2}{16\pi^2}\log\left(\frac{\Lambda^2}{f^2}\right)m_{W'}^2 \ll f^2. \tag{10.21}$$

- The Higgs mass is free of quadratic one-loop divergences.
- Mass scales:
  - f: vev of $SU(3)/SU(2)$, O(TeV);
  - $f/4\pi$: electroweak scale;
  - $\Lambda \sim 4\pi f$: cut-off scale.

10.3 The simplest little Higgs model

10.3.1 Gauge boson masses

Consider the breaking $SU(3)_w \otimes U(1)_x \to SU(2)_L \otimes U(1)_Y$, with the kinetic terms

$$\left|\left(\partial_\mu + igA_\mu^a T^a - \frac{i}{3}g_x A_\mu^x\right)\phi_i\right|^2 \xrightarrow{\text{vev}} \text{tr}\left[\left(gA_\mu^a T^a - \frac{1}{3}g_x A_\mu^x\right)^2 \phi_i\phi_i^\dagger\right], \tag{10.22}$$

and

$$\left\langle\phi_1\phi_1^\dagger + \phi_2\phi_2^\dagger\right\rangle = \begin{pmatrix}\left\langle hh^\dagger\right\rangle & 0\\ 0 & f^2\end{pmatrix}, \tag{10.23}$$

with the electroweak-symmetry-breaking vev

$$\langle h \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{10.24}$$

Remark. Here $\phi_{1,2}$ have vevs $f_{1,2}$ and $f^2 \equiv f_1^2 + f_2^2$ (Schmaltz); with the equal vevs f of the previous section, f^2 in this section stands for $2f^2$. One obtains:

$$SU(2): \qquad m_{W^\pm}^2 = \frac{g^2 v^2}{4} \quad \text{as in the SM}, \tag{10.25}$$

$$SU(3)/SU(2): \qquad m_{W'^\pm}^2 = m_{W'^0}^2 = \frac{g^2}{2} f^2. \tag{10.26}$$

Neutral gauge bosons: there are two diagonal $SU(3)$ generators, T_3 and T_8. Then

$$\begin{aligned} SU(2): \qquad & W_\mu^3 = A_\mu^3, \\ & B_\mu = \frac{-g_x A_\mu^8 + \sqrt{3}\, g B_\mu^x}{\sqrt{3g^2 + g_x^2}} \qquad (U(1)_Y \text{ gauge boson}), \\ & Z'_\mu = \frac{\sqrt{3}\, g A_\mu^8 + g_x B_\mu^x}{\sqrt{3g^2 + g_x^2}}, \end{aligned} \tag{10.27}$$

and the $U(1)_Y$ coupling is

$$g' = \frac{g_x}{\sqrt{1 + g_x^2/(3g^2)}}. \tag{10.28}$$

After $SU(2)_L \otimes U(1)_Y \to U(1)_Q$:

$$m_\gamma^2 = 0, \tag{10.29}$$

$$m_Z^2 = \frac{g^2 v^2}{4}\left(1 + \tan^2\theta_W\right) \qquad \text{as in the SM}, \tag{10.30}$$

$$m_{Z'}^2 = g^2 f^2 \frac{2}{3 - \tan^2\theta_W}. \tag{10.31}$$

The resulting spectrum is sketched in Fig. 10.2.

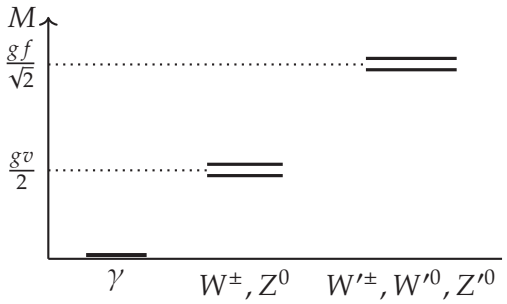


Figure 10.2. Gauge-boson spectrum of the simplest little Higgs model (schematic).

10.3.2 Matter content and Yukawa couplings

Matter is embedded in $(SU(3)_c, SU(3)_w)_{U(1)_x}$ multiplets:

	Quarks	Leptons
left-handed	$\Psi_Q = (3,3)_{1/3}$	$\Psi_L = (1,3)_{-1/3}$
right-handed	$d^c = (\bar{3},1)_{1/3}$	$e^c = (1,1)_1$
	$2\times u^c = (\bar{3},1)_{-2/3}$	$\nu^c = (1,1)_0$

Symmetry breaking proceeds through the vevs of two complex scalar triplet fields,

$$\phi_1,\ \phi_2 = (1,3)_{-1/3}. \tag{10.32}$$

The Yukawa terms in this model (which give the fermion masses) are

$$\begin{aligned}
&\lambda_1^u\, u_1^c\, \phi_1^\dagger \Psi_Q && u\text{-type quarks}\\
&+\lambda_2^u\, u_2^c\, \phi_2^\dagger \Psi_Q && \text{heavy partners of } u\text{-type quarks}\\
&+\lambda^d\, d^c\, \frac{\phi_1\phi_2\Psi_Q}{\Lambda} && d\text{-type quarks}\\
&+\lambda^\nu\, \nu^c\, \phi_1^\dagger \Psi_L && \text{neutrinos}\\
&+\lambda^e\, e^c\, \frac{\phi_1\phi_2\Psi_L}{\Lambda} && \text{charged leptons}
\end{aligned} \tag{10.33}$$

with

$$\phi_1^\dagger \Psi_Q \to 3\otimes\bar{3} \ni \text{singlet}, \tag{10.34}$$

$$\phi_1\phi_2\Psi_Q \to 3\otimes 3\otimes 3 \ni \text{singlet} \quad (\epsilon_{ijk}\phi_1^i\phi_2^j\Psi_Q^k). \tag{10.35}$$

As before, $\dim SU(3) = 8$ and $\dim SU(2) = 3$, so there are five broken generators. *Remark.* The labels "u-type quarks" and "heavy partners" are schematic: the two λ^u terms together give a heavy top partner T of mass $\sqrt{\lambda_1^2 f_1^2 + \lambda_2^2 f_2^2}$, and the orthogonal combination of u_1^c, u_2^c gets the SM top Yukawa, $\lambda_t \propto \lambda_1\lambda_2/\sqrt{\lambda_1^2+\lambda_2^2}$ for $f_1 = f_2$. Each term alone preserves an $SU(3)$ under which h is an exact Goldstone boson, so the top loop contributes to m_h^2 only logarithmically (collective breaking). With two triplets the global symmetry is $[SU(3)\times U(1)]^2 \to [SU(2)\times U(1)]^2$, giving 10 Goldstone bosons: the five broken generators of the gauged $SU(3)$ eat five of them, and h (four real fields) and η remain (Schmaltz, hep-ph/0407143).

PART V

Advanced Field Theory

Preface to Part V

In the previous parts we have treated quantum field theories mostly in perturbation theory, as theories of particles which scatter off each other. In this part we will learn that a quantum field theory contains much more. Symmetries of the classical Lagrangian may be destroyed by quantum fluctuations, the vacuum may have a non-trivial topological structure, and the non-linear field equations may have solutions which look nothing like small waves around the vacuum. These phenomena are not curiosities. They explain the decay of the neutral pion into two photons and the absence of a light ninth pseudoscalar meson, they constrain the fermion content of the Standard Model, and they play an essential role in the history of the early universe.

Chapter 1 asks what an integral in d dimensions really is, faces the problem of γ_5 in dimensional regularisation, and computes the axial anomaly, first in two and then in four space-time dimensions. Chapter 2 applies these results to the chiral symmetry of the strong interactions: the decay $\pi^0 \to \gamma\gamma$, the $U(1)_A$ problem, the linear sigma model, chiral perturbation theory, the θ vacuum of QCD, the axion and the instantons. Chapter 3 is devoted to topological solitons (kinks, vortices and magnetic monopoles) and to the physics of the hot early universe: phase transitions at finite temperature, the anomalous violation of baryon and lepton number, sphalerons and baryogenesis. Finally, Chapter 4 turns to the expanding universe, inflation, and the quantum theory of a scalar field in an expanding background.

We assume familiarity with the material of Parts I–III; for chiral symmetry and effective field theories we also build on Part IV. Our conventions are those of the previous parts, and they are summarised at the beginning of Chapter 1. Exercises are interspersed in the text; they are an integral part of it.

This part grew out of the course *Advanced Field Theory*, which Thomas Gehrmann (University of Zurich) and I taught jointly at ETH Zurich in the spring semester of 2010. I thank Thomas for a very enjoyable collaboration, and Peter Reiter, who first typeset the lecture notes of that course.

1

Dimensional regularisation and the axial anomaly

Conventions

Throughout this part we use the conventions of Parts I–IV. For reference:

- Metric $g_{\mu\nu} = \mathrm{diag}(+1,-1,-1,-1)$; $\epsilon^{0123} = -\epsilon_{0123} = +1$; $\hbar = c = 1$ and Heaviside-Lorentz units, $\alpha = e^2/(4\pi)$.
- $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ (Eq. 7.47), so that $\mathrm{tr}\,(\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma_5) = -4i\,\epsilon^{\mu\nu\rho\sigma}$; chiral projections $\psi_{R,L} = \frac{1}{2}(1 \pm \gamma_5)\psi$.
- QED (Eq. 1.18 of Part II): $D_\mu = \partial_\mu - ieA_\mu$, hence $\mathcal{L}_{\mathrm{int}} = +e\bar{\psi}\not{A}\psi$ and the vertex $+ie\gamma^\mu$; the quanta of ψ carry the electric charge $-e$. Field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, with $F_{0i} = E^i$, $F_{ij} = -\epsilon_{ijk}B^k$, so that $\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} = -8\,\vec{E}\cdot\vec{B}$.
- Non-abelian gauge fields as in Part III (Eq. 3.26): $D_\mu = \partial_\mu - igA^a_\mu T^a$, $G_{\mu\nu} = \frac{i}{g}\left[D_\mu, D_\nu\right] = \partial_\mu A_\nu - \partial_\nu A_\mu - ig\left[A_\mu, A_\nu\right]$, i.e. $G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$, $\mathrm{tr}(T^aT^b) = \frac{1}{2}\delta^{ab}$ in the fundamental representation; dual field $\tilde{G}^{a\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}G^a_{\rho\sigma}$.
- With these conventions the axial anomaly of a Dirac fermion in a representation R (Eq. 1.66) reads $\partial_\mu j^{\mu 5} = -\frac{g^2 T(R)}{8\pi^2}\,G^a_{\mu\nu}\tilde{G}^{a\mu\nu}$, $\mathrm{tr}(T^aT^b) = T(R)\delta^{ab}$; for QED, $\partial_\mu j^{\mu 5} = \frac{e^2}{2\pi^2}\vec{E}\cdot\vec{B}$. The same sign is used in the chapters on chiral symmetry ($\pi^0 \to \gamma\gamma$, $U(1)_A$, θ term) and on $B+L$ violation.
- Chiral symmetry as in Part IV (Section "Conventions" of Chapter 1, and Chapters 6 and 7 of Part IV): $q_{L,R} \to g_{L,R}\,q_{L,R}$, the Goldstone matrix $U = \exp\left(i\vec{\sigma}\cdot\vec{\pi}/f\right)$ transforms as $U \to g_L U g_R^\dagger$. We use the pion decay constant $f_\pi \simeq 92$ MeV, $\langle 0|\,\bar{q}\gamma^\mu\gamma_5\frac{\sigma^a}{2}q\,\left|\pi^b(p)\right\rangle = -ip^\mu f_\pi\delta^{ab}e^{-ip\cdot x}$ (relativistic normalisation); in Part IV the constant $F_\pi = 2f_\pi \simeq 184$ MeV is used.

Symmetries of a classical Lagrangian are not automatically symmetries of the quantum theory. To compute loop amplitudes we must first regularise them, and the regulator may not respect all the symmetries of the Lagrangian. In most cases we can choose a regulator which does, or we can restore the symmetry with suitable counterterms. However, there are symmetries for which this is impossible! The most famous example is the axial (chiral) symmetry of massless fermions coupled to a gauge field. The divergence of the axial current, which vanishes for the classical equations of motion, acquires a non-zero value

from quantum fluctuations. This is the **axial anomaly**, or **Adler-Bell-Jackiw anomaly**. It is not an academic curiosity. It is responsible for the decay $\pi^0 \to \gamma\gamma$, it explains why there is no light isosinglet pseudoscalar meson in QCD, and it puts strong constraints on the fermion content of the Standard Model.

In this chapter we will first recall why we need a regulator and what dimensional regularisation really is. The integrals themselves have been treated in detail in Chapters 3 and 4 of Part II, and we will not repeat them here. We will rather concentrate on the one object of the Dirac algebra which does not like to live in d dimensions, the matrix γ_5. Then we will review the chiral symmetry of massless QED, and we will compute the axial anomaly, first in two space-time dimensions, where everything can be done in a few lines, and then in four dimensions.

1.1 Ultraviolet and infrared divergences

Loop integrals in field theory can diverge in two very different regions of the loop momentum. They can diverge for very large momenta, $k^2 \to \infty$ (**ultraviolet** divergences), and, in theories with massless particles, for small momenta or for momenta which become parallel to an external massless momentum (**infrared** divergences). Let us look at one example of each kind.

A simple ultraviolet divergent integral is the one-loop bubble,

$$\int \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2-m^2)\left[(k-q)^2-m^2\right]} \overset{k\to\infty}{\longrightarrow} \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^4}. \tag{1.1}$$

Counting powers, $d^4k \sim |k|^3 d|k|$, we find that it diverges logarithmically at large $|k|$. We have learned in Chapter 6 of Part II how to absorb such divergences into a redefinition of the parameters of the Lagrangian.

The infrared divergences are more subtle. Consider the vertex integral of a massless theory with two light-like external momenta,

$$\int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2(k+p_1)^2(k-p_2)^2}, \qquad p_1^2 = p_2^2 = 0. \tag{1.2}$$

There is no mass in the propagators to protect the integral from small denominators. We can identify two dangerous regions:

- The **soft** region, where all components of the loop momentum become small, $k^\mu \to \lambda k^\mu$, $\lambda \to 0$. Then $k^2 \sim \lambda^2$, $(k+p_1)^2 = k^2 + 2k\cdot p_1 \sim \lambda$ and $(k-p_2)^2 = k^2 - 2k\cdot p_2 \sim \lambda$, while $d^4k \sim \lambda^4$. The integral behaves as $\int d\lambda/\lambda$, a logarithmic divergence.
- The **collinear** region, where k becomes parallel to p_1 (or to p_2).

To study the collinear region it is useful to introduce light-cone coordinates. We work in a frame where p_1 and p_2 are back-to-back along the z-axis, $p_1^\mu = (E,0,0,E)$ and $p_2^\mu = (E,0,0,-E)$. We introduce the light-like vectors

$$n^\mu = \frac{1}{\sqrt{2}}(1,0,0,1), \qquad \bar{n}^\mu = \frac{1}{\sqrt{2}}(1,0,0,-1), \qquad n^2 = \bar{n}^2 = 0, \quad n\cdot\bar{n} = 1, \tag{1.3}$$

and decompose any four-vector as

$$k^\mu = k^+ n^\mu + k^- \bar{n}^\mu + k_\perp^\mu, \qquad k^\pm = \frac{k^0 \pm k^3}{\sqrt{2}}, \qquad k_\perp^\mu = (0,k^1,k^2,0). \tag{1.4}$$

This is known as the Sudakov (or light-cone) parametrisation. We have

$$k^2 = (k^0 - k^3)(k^0 + k^3) - \vec{k}_\perp^2 = 2k^+k^- - \vec{k}_\perp^2, \qquad d^4k = dk^+\, dk^-\, d^2\vec{k}_\perp, \tag{1.5}$$

since the Jacobian of the transformation $(k^0, k^3) \to (k^+, k^-)$ is one. With $p_1 = \sqrt{2}E\, n$ and $p_2 = \sqrt{2}E\, \bar{n}$ the propagators read

$$k^2 = 2k^+k^- - \vec{k}_\perp^2, \quad (k + p_1)^2 = k^2 + 2\sqrt{2}E\, k^-, \quad (k - p_2)^2 = k^2 - 2\sqrt{2}E\, k^+. \tag{1.6}$$

In the region where k is collinear to p_1 we can scale $k^+ \sim E$, $k_\perp \sim \lambda E$, $k^- \sim \lambda^2 E$. Then $k^2 \sim \lambda^2$ and $(k + p_1)^2 \sim \lambda^2$, while $(k - p_2)^2 \sim 1$. The measure scales as $dk^+dk^-d^2k_\perp \sim \lambda^4$, and we find again a logarithmic divergence, $\int d\lambda/\lambda$. The soft and the collinear regions overlap, and in dimensional regularisation the integral of Eq. 1.2 has a double pole $1/\epsilon^2$.

Exercise: *Using the Feynman parametrisation and the results of Chapter 3 of Part II, show that in $d = 4 - 2\epsilon$ dimensions*

$$\int \frac{d^dk}{i\pi^{d/2}} \frac{1}{k^2(k + p_1)^2(k - p_2)^2} = -\frac{c_\Gamma}{\epsilon^2} (-s - i\delta)^{-1-\epsilon}, \qquad s = (p_1 + p_2)^2,$$

with $c_\Gamma = \Gamma(1 + \epsilon)\Gamma(1 - \epsilon)^2/\Gamma(1 - 2\epsilon)$. Which region of the loop momentum is responsible for the double pole?

1.2 What is an integral in d dimensions?

In Chapter 3 of Part II we have computed loop integrals in $d = 4 - 2\epsilon$ dimensions by using spherical coordinates and by continuing the result, a function of d, to complex values of d. This is a very practical recipe. But what do we really mean by an integral over a space with a non-integer number of dimensions? Let us take a step back and ask which properties we want such an operation to have.

We consider a (Euclidean, after Wick rotation) vector p and a functional $f \to \int d^dp\, f(p)$. For it to deserve the name "integral" we ask:

- **Linearity:** for any complex numbers a, b,

$$\int d^dp \left[a f(p) + b g(p)\right] = a \int d^dp\, f(p) + b \int d^dp\, g(p). \tag{1.7}$$

- **Translation invariance:** for any vector q,

$$\int d^dp\, f(p + q) = \int d^dp\, f(p). \tag{1.8}$$

- **Scaling:** for any number s,

$$\int d^dp\, f(sp) = s^{-d} \int d^dp\, f(p). \tag{1.9}$$

We also require that the result is covariant under rotations. Linearity holds for any integral. Translation and rotation invariance are the basic properties of a Euclidean space, and it is the scaling property which carries the information about the number of dimensions.

It can be shown that these properties determine the integral uniquely, up to an overall normalisation, which we fix by

$$\int d^d p\, e^{-p^2} = \pi^{d/2}. \tag{1.10}$$

For integer d we recover the usual integral; for non-integer d the answer is the analytic continuation of the formulae of Chapter 3 of Part II.

Notice that these few axioms already contain non-trivial results. Consider a scaleless integral, $I = \int d^d p\, (p^2)^{-a}$. The scaling property gives

$$s^{-d} I = \int d^d p\, (s^2 p^2)^{-a} = s^{-2a} I \quad \rightsquigarrow \quad \left(s^{-d} - s^{-2a}\right) I = 0$$

for every s, and therefore $I = 0$ unless $d = 2a$. This is the statement that scaleless integrals vanish in dimensional regularisation (Eq. 3.50 of Part II). It will be very important for us that the shift of the integration variable, Eq. 1.8, is always allowed in dimensional regularisation, even for integrals which diverge in four dimensions.

1.3 The Dirac algebra in d dimensions

Dimensional regularisation changes the number of components of the loop momentum. It also changes the number of values that a Lorentz index can take. We must then define the Dirac matrices with d-dimensional indices. We keep the Clifford algebra and the metric,

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}\mathbf{1}, \qquad g^\mu_{\ \mu} = d, \tag{1.11}$$

together with $\gamma^{0\dagger} = \gamma^0$ and $\gamma^{i\dagger} = -\gamma^i$ for the space-like directions. We do not need to construct the matrices explicitly (for non-integer d they would have to be infinite dimensional). All that enters a Feynman diagram are traces, and these we can compute from Eq. 1.11, from the cyclicity of the trace, and from one normalisation,

$$\mathrm{tr}\,\mathbf{1} = f(d), \qquad f(4) = 4. \tag{1.12}$$

The difference $f(d) - f(4)$ multiplies only divergent integrals and can be absorbed in the choice of renormalisation scheme. The usual choice, which we also make here, is $f(d) = 4$ for all d. In this way the fermions keep their four-dimensional number of spin degrees of freedom, while the gauge boson acquires $d - 2$ physical polarisations. We have already used these rules in Chapter 5 of Part II,

$$\gamma^\mu\gamma_\mu = d, \qquad \gamma^\mu\gamma^\nu\gamma_\mu = -(d-2)\gamma^\nu, \qquad \ldots$$

and we noted there that a difficulty arises with γ_5. It is time to face it.

In four dimensions (see Eq. 7.47)

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = -\frac{i}{4!}\,\epsilon_{\mu\nu\rho\sigma}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma, \qquad \epsilon^{0123} = -\epsilon_{0123} = +1, \tag{1.13}$$

and it has two properties:

$$\{\gamma_5, \gamma^\mu\} = 0, \qquad \mathrm{tr}\,(\gamma_5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma) = -4i\,\epsilon^{\mu\nu\rho\sigma}. \tag{1.14}$$

The second property is what makes the axial current $\bar\psi\gamma^\mu\gamma_5\psi$ different from a vector current in a Feynman diagram. Can we keep both properties in d dimensions? No!

Exercise: *Assume that γ_5 anticommutes with all d-dimensional γ^μ. Insert $\gamma_\alpha\gamma^\alpha = d$ in the traces $\mathrm{tr}(\gamma_5\gamma^\mu\gamma^\nu)$ and $\mathrm{tr}(\gamma_5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma)$, use the cyclicity of the trace and the d-dimensional contraction identities to show that*

$$(d-2)(d-4)\,\mathrm{tr}\,(\gamma_5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma) = 0.$$

Therefore, an anticommuting γ_5 gives a vanishing trace with four γ matrices for $d \neq 4$, and, by continuity, also at $d = 4$. A regularisation with a fully anticommuting γ_5 loses the ϵ-tensor altogether. For many QED and QCD calculations in which γ_5 enters only through open fermion lines this is harmless, and an anticommuting γ_5 is a perfectly good practical prescription. For a closed fermion loop with an odd number of γ_5, such as the triangle that we will compute below, it is inconsistent.

A consistent definition was given by 't Hooft and Veltman, and it was worked out systematically by Breitenlohner and Maison. We simply keep the **four-dimensional** definition, Eq. 1.13, with $\epsilon_{\mu\nu\rho\sigma}$ a purely four-dimensional object. To make this precise, we split the d-dimensional metric into a four-dimensional and a $(d-4)$-dimensional part,

$$g^{\mu\nu} = \bar{g}^{\mu\nu} + \hat{g}^{\mu\nu}, \qquad \bar{g}^\mu_{\ \mu} = 4, \qquad \hat{g}^\mu_{\ \mu} = d-4, \qquad \bar{g}^{\mu\rho}\hat{g}_\rho^{\ \nu} = 0, \tag{1.15}$$

where $\hat{g}^\mu_{\ \rho}\hat{g}^{\rho\nu} = \hat{g}^{\mu\nu}$ is a projector onto the extra dimensions. Accordingly, every vector and every Dirac matrix splits as

$$p^\mu = \bar{p}^\mu + \hat{p}^\mu, \qquad \gamma^\mu = \bar{\gamma}^\mu + \hat{\gamma}^\mu, \qquad \hat{p}^\mu = \hat{g}^\mu_{\ \nu}p^\nu, \quad \hat{\gamma}^\mu = \hat{g}^\mu_{\ \nu}\gamma^\nu, \tag{1.16}$$

with

$$\{\bar{\gamma}^\mu, \bar{\gamma}^\nu\} = 2\bar{g}^{\mu\nu}, \qquad \{\hat{\gamma}^\mu, \hat{\gamma}^\nu\} = 2\hat{g}^{\mu\nu}, \qquad \{\bar{\gamma}^\mu, \hat{\gamma}^\nu\} = 0. \tag{1.17}$$

The ϵ-tensor lives in four dimensions, $\epsilon_{\mu\nu\rho\sigma}\hat{g}^{\sigma\alpha} = 0$, and the product of two ϵ-tensors is defined as the four-dimensional determinant

$$\epsilon^{\mu_1\mu_2\mu_3\mu_4}\epsilon^{\nu_1\nu_2\nu_3\nu_4} = -\sum_{\pi\in S_4}\mathrm{sgn}(\pi)\prod_{i=1}^{4}\bar{g}^{\mu_i\nu_{\pi(i)}} = -\det\left[\bar{g}^{\mu_i\nu_j}\right]. \tag{1.18}$$

(Check the sign with $\mu_i = \nu_i = (0,1,2,3)$.) With γ_5 defined by Eq. 1.13, it is now easy to show that

$$\{\gamma_5, \bar{\gamma}^\mu\} = 0, \qquad [\gamma_5, \hat{\gamma}^\mu] = 0, \qquad \gamma_5^2 = \mathbf{1}, \qquad \gamma_5^\dagger = \gamma_5. \tag{1.19}$$

Indeed, $\hat{\gamma}^\mu$ anticommutes with each of the four matrices $\gamma^0, \gamma^1, \gamma^2, \gamma^3$ in the product, and therefore commutes with γ_5. The trace formula of Eq. 1.14 holds unchanged, with the ϵ-tensor projecting all indices onto four dimensions.

The price that we pay is that γ_5 is not invariant under d-dimensional Lorentz transformations, but only under those of the first four dimensions. This is a nuisance in practical calculations, since Ward identities of the axial current are violated by terms proportional to $\hat{g}^{\mu\nu}$ and must sometimes be restored by finite counterterms. But the prescription is algebraically consistent, and, as we will see, it gives the correct axial anomaly.

1.4 Chiral symmetry of massless QED

Let us recall the QED Lagrangian (Eq. 1.18 of Part II),

$$\mathcal{L} = \bar{\psi}\,(i\not{D} - m)\,\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \qquad D_\mu = \partial_\mu - ieA_\mu. \tag{1.20}$$

We decompose the fermion field into its chiral components,

$$\psi = \psi_L + \psi_R, \qquad \psi_R = \frac{1+\gamma_5}{2}\psi, \qquad \psi_L = \frac{1-\gamma_5}{2}\psi. \tag{1.21}$$

Since $\gamma^0\gamma^\mu$ commutes with γ_5, the kinetic and the gauge interaction terms do not mix the two chiralities,

$$\bar\psi\, i\not{D}\,\psi = \bar\psi_L\, i\not{D}\,\psi_L + \bar\psi_R\, i\not{D}\,\psi_R, \tag{1.22}$$

while the mass term mixes them,

$$m\bar\psi\psi = m\left(\bar\psi_L\psi_R + \bar\psi_R\psi_L\right). \tag{1.23}$$

Exercise: *Prove the last two equations, using* $\bar\psi_R = \bar\psi\frac{1-\gamma_5}{2}$ *and* $\bar\psi_L = \bar\psi\frac{1+\gamma_5}{2}$.
For $m = 0$, left- and right-handed fermions are two independent species which talk to each other only through the photon. The Lagrangian is then invariant under two independent phase rotations,

$$\psi_L \to e^{i\alpha_L}\psi_L, \qquad \psi_R \to e^{i\alpha_R}\psi_R, \tag{1.24}$$

which we can also write as a vector and an axial transformation,

$$\psi \to e^{i\alpha}\psi \quad \text{(vector)}, \qquad \psi \to e^{i\beta\gamma_5}\psi, \quad \bar\psi \to \bar\psi e^{i\beta\gamma_5} \quad \text{(axial)}. \tag{1.25}$$

The corresponding Noether currents (Eq. 3.62) are the vector and the axial vector currents,

$$j^\mu = \bar\psi\gamma^\mu\psi = \bar\psi_R\gamma^\mu\psi_R + \bar\psi_L\gamma^\mu\psi_L, \qquad j^{\mu 5} = \bar\psi\gamma^\mu\gamma_5\psi = \bar\psi_R\gamma^\mu\psi_R - \bar\psi_L\gamma^\mu\psi_L. \tag{1.26}$$

Using the equations of motion we find

$$\partial_\mu j^\mu = 0, \qquad \partial_\mu j^{\mu 5} = 2im\,\bar\psi\gamma_5\psi. \tag{1.27}$$

Exercise: *Derive Eq. 1.27 from the Dirac equation* $(i\not{D} - m)\,\psi = 0$ *and its conjugate. Notice that the gauge field drops out.*
The vector current is conserved for any mass. Its charge $N = \int d^3x\, j^0$ is the fermion number (particles minus antiparticles), and $dN/dt = 0$. For massless fermions, the axial current is conserved as well, and so is $N_5 = N_R - N_L$. Together with $N = N_R + N_L$, this means that the numbers of right- and left-handed fermions would be separately conserved. In what follows we will show that this conclusion, which is correct in the classical theory, is false in the quantum theory. The chiral symmetries of QCD, where the same decomposition is made for the light quarks, are discussed in Chapter 6 of Part IV, and in Chapter 2.

1.5 The axial anomaly in two dimensions

Before we enter the technicalities of four dimensions, let us find the anomaly in a model where the calculation takes a few lines: massless QED in one space and one time dimension,

$$\mathcal{L} = \bar\psi\, i\not{D}\,\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \qquad \mu,\nu = 0,1, \qquad D_\mu = \partial_\mu - ieA_\mu. \tag{1.28}$$

In two dimensions the Clifford algebra $\{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu}$ is realised by 2×2 matrices. We choose

$$\gamma^0 = \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \gamma^1 = i\sigma^1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \qquad \gamma_5 \equiv \gamma^0\gamma^1 = \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.29}$$

The matrix γ_5 anticommutes with γ^0 and γ^1, it is Hermitian, and $\gamma_5^2 = 1$. We label the two components of the spinor field by their γ_5 eigenvalue, $\psi = (\psi_+, \psi_-)^T$. With $\bar\psi = \psi^\dagger\gamma^0$, $(\gamma^0)^2 = 1$ and $\gamma^0\gamma^1 = \sigma^3$, the Lagrangian of the fermions becomes

$$\mathcal{L}_\psi = \psi^\dagger\, i\left(D_0 + \sigma^3 D_1\right)\psi = \psi_+^\dagger\, i\,(D_0 + D_1)\,\psi_+ + \psi_-^\dagger\, i\,(D_0 - D_1)\,\psi_-. \tag{1.30}$$

The free field equations, $(\partial_0 + \partial_1)\psi_+ = 0$ and $(\partial_0 - \partial_1)\psi_- = 0$, are solved by $\psi_+ = f(x^1 - t)$ and $\psi_- = g(x^1 + t)$. The quanta of ψ_+ move to the right with the speed of light, and the quanta of ψ_- to the left. In two dimensions "right-handed" means "right-moving". The two species do not mix in the Lagrangian, and the classical theory conserves the vector current $j^\mu = \bar\psi\gamma^\mu\psi$ and the axial current $j^{\mu 5} = \bar\psi\gamma^\mu\gamma_5\psi$, i.e. the numbers N_R and N_L of right- and left-movers separately. Notice that $j^{05} = \psi^\dagger\gamma_5\psi = \psi_+^\dagger\psi_+ - \psi_-^\dagger\psi_-$.

Two dimensions have a special feature which makes our life easy. With the antisymmetric symbol $\epsilon^{\mu\nu}$, $\epsilon^{01} = -\epsilon^{10} = +1$, we can verify directly that

$$\gamma^\mu\gamma_5 = -\epsilon^{\mu\nu}\gamma_\nu \qquad \rightsquigarrow \qquad j^{\mu 5} = -\epsilon^{\mu\nu} j_\nu. \tag{1.31}$$

For example, $\gamma^0\gamma_5 = \gamma^0\gamma^0\gamma^1 = \gamma^1 = -\epsilon^{01}\gamma_1$. The axial current is not an independent object; it is the "dual" of the vector current. Therefore, if we know how the vector current responds to an external electromagnetic field, we also know how the axial current responds.

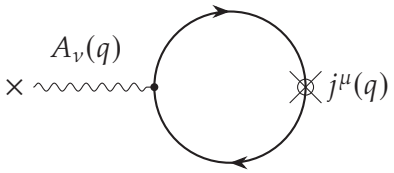


Figure 1.1. The expectation value of the current j^μ induced by a weak background field A_ν at one loop. The diagram is the vacuum polarisation with one photon leg replaced by the current.

Let us then compute the current induced by a weak background field. The coupling of the fermions to the field is $\mathcal{L}_{\rm int} = +e\, j^\mu A_\mu$. If we integrate out the fermions, we obtain an effective action $W[A]$ for the background field. Its quadratic term is given by the vacuum polarisation of Fig. 1.1,

$$W[A] = \frac{1}{2}\int \frac{d^2q}{(2\pi)^2}\, A_\mu(-q)\,\Pi^{\mu\nu}(q)\,A_\nu(q) + \ldots, \qquad \Pi^{\mu\nu}(q) = \left(q^2 g^{\mu\nu} - q^\mu q^\nu\right)\Pi(q^2), \tag{1.32}$$

where $i\Pi^{\mu\nu}$ is the one-loop photon self-energy in the conventions of Eq. 5.51 of Part II. Since A_μ couples to $+ej^\mu$, the induced current is

$$\left\langle j^\mu(x)\right\rangle = +\frac{1}{e}\frac{\delta W}{\delta A_\mu(x)} \qquad \rightsquigarrow \qquad \left\langle j^\mu(q)\right\rangle = +\frac{1}{e}\left(q^2 g^{\mu\nu} - q^\mu q^\nu\right)\Pi(q^2)\,A_\nu(q), \tag{1.33}$$

where $A_\mu(x) = \int \frac{d^2q}{(2\pi)^2} e^{-iq\cdot x} A_\mu(q)$, and similarly for the current. In Chapter 5 of Part II we have computed $\Pi(q^2)$ for massless fermions in d dimensions with $\mathrm{tr}\,\mathbf{1} = 4$. For general d and $\mathrm{tr}\,\mathbf{1}$ the same calculation gives

$$\Pi(q^2) = -\frac{\mathrm{tr}\,\mathbf{1}}{4}\,\frac{8e^2}{(4\pi)^{d/2}}\,\Gamma\left(2 - \frac{d}{2}\right)\int_0^1 dx\,\frac{x(1-x)}{\left[-x(1-x)q^2\right]^{2-\frac{d}{2}}}. \tag{1.34}$$

In $d = 4 - 2\epsilon$ it reproduces the pole $-e^2/(12\pi^2\epsilon)$ of Eq. 5.51 of Part II. In two dimensions, with $\mathrm{tr}\,\mathbf{1} = 2$, the integral is finite, and we find a remarkably simple result,

$$\Pi(q^2) = -\frac{1}{2}\frac{8e^2}{4\pi}\int_0^1 dx\,\frac{x(1-x)}{-x(1-x)q^2} = \frac{e^2}{\pi q^2}. \tag{1.35}$$

Exercise: *Derive Eq. 1.34. Then show that the photon propagator of the two-dimensional theory, $-ig_{\mu\nu}/\left[q^2(1-\Pi(q^2))\right]$ (in Feynman gauge, transverse part), has a pole at $q^2 = e^2/\pi$: the photon of massless two-dimensional QED is massive! (This is the Schwinger model.)*

Substituting Eq. 1.35, the induced vector current reads

$$\left\langle j^\mu(q)\right\rangle = +\frac{e}{\pi}\left(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right)A_\nu(q). \tag{1.36}$$

It is conserved, $q_\mu\left\langle j^\mu(q)\right\rangle = 0$, as it must be: we have computed it with a gauge invariant regulator. Notice that it contains a pole at $q^2 = 0$; in a massless theory, a non-local response is not forbidden. As a check of the sign, consider a static potential $A_0 = \phi(x^1)$. With $D_\mu = \partial_\mu - ieA_\mu$ the quanta of ψ carry the charge $-e$ (for $e > 0$, ψ is the electron field), and the induced charge density is $-e\left\langle j^0\right\rangle = -(e^2/\pi)\phi$: charges are pushed away from the region where their potential energy is high, as they should be.

Using Eq. 1.31, the axial current follows at once,

$$\left\langle j^{\mu 5}(q)\right\rangle = -\epsilon^{\mu\nu}\left\langle j_\nu(q)\right\rangle = -\frac{e}{\pi}\,\epsilon^{\mu\nu}\left(A_\nu(q) - \frac{q_\nu\, q^\lambda A_\lambda(q)}{q^2}\right). \tag{1.37}$$

If the axial current were conserved, its contraction with q_μ would vanish. Instead, the second term drops by antisymmetry of $\epsilon^{\mu\nu}$ and we find

$$q_\mu\left\langle j^{\mu 5}(q)\right\rangle = -\frac{e}{\pi}\,\epsilon^{\mu\nu}q_\mu A_\nu(q). \tag{1.38}$$

In position space, $\partial_\mu \to -iq_\mu$, this reads

$$\partial_\mu j^{\mu 5} = -\frac{e}{\pi}\,\epsilon^{\mu\nu}\partial_\mu A_\nu = -\frac{e}{2\pi}\,\epsilon^{\mu\nu}F_{\mu\nu} = -\frac{e}{\pi}\,F_{01}. \tag{1.39}$$

This is the axial anomaly in two dimensions. The axial current is not conserved in the presence of an electric field, $F_{01} = \partial_0 A_1 - \partial_1 A_0 = E$. The symmetry of the classical Lagrangian, under which left- and right-movers are counted separately, has been broken by the quantum fluctuations of the fermions.

We could have kept the axial current conserved instead, with a different (non gauge invariant) treatment of the vacuum polarisation. Then the vector current would not have been conserved. What we cannot do is to keep both. Since the vector current is coupled to the gauge field, its conservation is necessary for the consistency of the gauge theory, and it is the axial current which must give way.

1.5.1 Where do the right-movers go?

Our result has a striking global consequence. Let us integrate Eq. 1.39 over space-time. For a free theory the left-hand side would give

$$\int d^2x\,\partial_\mu j^{\mu 5} = \int dt\,\frac{d}{dt}\,(N_R - N_L) = \Delta\,(N_R - N_L)\,, \tag{1.40}$$

the change of the difference of right- and left-movers between $t = -\infty$ and $t = +\infty$. The right-hand side is a total derivative,

$$\epsilon^{\mu\nu} F_{\mu\nu} = 2\partial_\mu \left(\epsilon^{\mu\nu} A_\nu\right), \tag{1.41}$$

and one might think that its integral vanishes. This is true for fields which fall off fast enough at infinity, but not for a constant electric field acting for a finite time. Then the anomaly equation predicts

$$\Delta\left(N_R - N_L\right) = -\frac{e}{2\pi} \int d^2x\, \epsilon^{\mu\nu} F_{\mu\nu}. \tag{1.42}$$

Fermions are created or destroyed! How can this happen in a theory in which each species is conserved by the Hamiltonian?

Let us look at the fermions in a box of length L with periodic boundary conditions, in a background field $A_0 = 0$, $A^1 = A(t)$ (so $A_1 = -A$), which is constant in space and varies very slowly in time. Notice that such a constant A^1 cannot be removed by a gauge transformation which respects the periodic boundary conditions (it would require e^{+ieAx}, which is not periodic for general A). The electric field is $E = F_{01} = -\dot{A}$. From Eq. 1.30, the field equations are

$$i\partial_0 \psi_\pm = \mp i D_1 \psi_\pm = \mp i \left(\partial_1 + ieA\right) \psi_\pm. \tag{1.43}$$

Equivalently, the Hamiltonian is

$$H = \int_0^L dx \left\{\psi_+^\dagger \left[-i\left(\partial_1 + ieA\right)\right] \psi_+ - \psi_-^\dagger \left[-i\left(\partial_1 + ieA\right)\right] \psi_-\right\}. \tag{1.44}$$

It is diagonalised by plane waves $e^{ik_n x}$, with $k_n = 2\pi n/L$, $n \in \mathbb{Z}$, which satisfy the boundary conditions. The single-particle energies are

$$\psi_+ : \quad E_n = +\left(k_n + eA\right), \qquad \psi_- : \quad E_n = -\left(k_n + eA\right). \tag{1.45}$$

Each species has an infinite tower of equally spaced levels, $2\pi/L$ apart. The ground state is the Dirac sea, with all negative energy levels filled; an empty negative energy level is an antiparticle.

Now we change A adiabatically by

$$\Delta A = -\frac{2\pi}{eL}. \tag{1.46}$$

At the end the spectrum is the same as at the beginning: every level of ψ_+ has moved down by one spacing to the position of its neighbour, and every level of ψ_- has moved up by one spacing (Fig. 1.2). An adiabatic change does not change the occupation of the levels. Therefore, one empty level of ψ_+ has moved below zero energy, and one filled level of ψ_- has moved above zero. Compared with the vacuum, we have lost one right-mover (we have created a right-moving antiparticle) and we have gained one left-mover:

$$\Delta\left(N_R - N_L\right) = -1 - 1 = -2. \tag{1.47}$$

Let us compare with the anomaly equation, Eq. 1.42:

$$-\frac{e}{2\pi} \int d^2x\, \epsilon^{\mu\nu} F_{\mu\nu} = -\frac{e}{\pi} \int dt \int_0^L dx\, \left(-\dot{A}\right) = +\frac{e}{\pi} L\, \Delta A = -2. \tag{1.48}$$

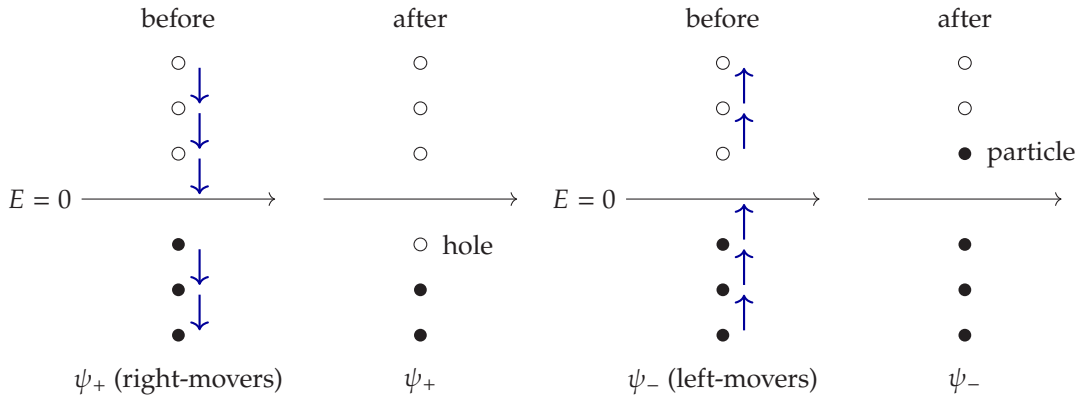


Figure 1.2. Spectral flow in two-dimensional QED. Under the adiabatic change $\Delta A = -2\pi/(eL)$ the levels of ψ_+ move down and those of ψ_- move up by one spacing, keeping their occupation (full circles: occupied levels). The vacuum loses one right-mover and gains one left-mover.

It actually works! The quantum-mechanical counting of levels reproduces exactly the anomaly which we found from the vacuum polarisation diagram.

Notice where the fermions come from. The levels of ψ_+ flow down, and those of ψ_- flow up, from and into the bottom of the Dirac sea at $E \to -\infty$. In counting N_R and N_L we have assumed, implicitly, that nothing happens at infinitely negative energies; we have imposed a gauge invariant cut-off on the deep sea. Even in this simple example we cannot avoid the question of ultraviolet regularisation. A gauge invariant treatment of the infinitely many states of the sea is compatible with the conservation of $N_R + N_L$, but not with the separate conservation of N_R and N_L.

1.6 The axial anomaly in four dimensions

We now return to four-dimensional massless QED. Is the axial current conserved? The two-dimensional example suggests a test. The anomaly equation of two dimensions, Eq. 1.39, is linear in the field strength. In four dimensions the analogous pseudoscalar built from field strengths is quadratic, $\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta}$, and a non-zero divergence of $j^{\mu 5}$ would show up in the matrix element of $\partial_\mu j^{\mu 5}$ to create two photons. Let us then compute

$$\langle p, \nu; k, \lambda | \, j^{\mu 5}(0) \, |0\rangle = M^{\mu\nu\lambda}(p,k)\, \epsilon^*_\nu(p)\, \epsilon^*_\lambda(k), \tag{1.49}$$

where p, k are the momenta of the two photons ($p^2 = k^2 = 0$) and $q = p + k$ is the momentum injected by the current. Since $j^{\mu 5}(x) = e^{iP\cdot x} j^{\mu 5}(0) e^{-iP\cdot x}$, the matrix element of the divergence is

$$\langle p, k | \, \partial_\mu j^{\mu 5}(0) \, |0\rangle = i q_\mu M^{\mu\nu\lambda}(p,k)\, \epsilon^*_\nu(p)\, \epsilon^*_\lambda(k). \tag{1.50}$$

At lowest order we expand the time-ordered exponential of $i\int d^4x\, \mathcal{L}_{\text{int}}$, with $\mathcal{L}_{\text{int}} = +e\bar\psi \not{A}\psi$, to second order and contract the fermion fields into a closed loop. The loop gives a factor (-1), the two vertices a factor $(i)^2(+e)^2$, and the two assignments of the photons to the two vertices give the two diagrams of Fig. 1.3. With $S(l) = i\not{l}/l^2$ for the massless fermion propagator, we find

$$M^{\mu\nu\lambda} = e^2 \int \frac{d^d l}{(2\pi)^d} \operatorname{tr}\left[\gamma^\mu\gamma_5\, S(l)\, \gamma^\nu\, S(l+p)\, \gamma^\lambda\, S(l+q)\right] + \left(p, \nu \leftrightarrow k, \lambda\right). \tag{1.51}$$

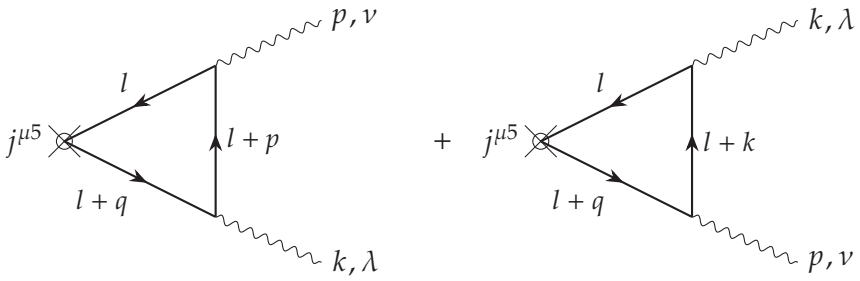


Figure 1.3. The two triangle diagrams for the matrix element of the axial current between the vacuum and a two-photon state. The arrows show the fermion-number flow; the loop momenta flow along the arrows. The current injects the momentum $q = p + k$.

Each diagram is linearly divergent by power counting. We have written it in d dimensions, but let us first argue as if we were in four dimensions and as if all manipulations were allowed.

We contract with q_μ and use, as in the proof of a Ward identity, the denominators of the propagators adjacent to the vertex. Writing $\not{q} = (\not{l} + \not{q}) - \not{l}$ and anticommuting γ_5 with $\not{l}$,

$$\not{q}\,\gamma_5 = \left(\not{l} + \not{q}\right)\gamma_5 + \gamma_5\,\not{l} \qquad \text{(four dimensions)}. \tag{1.52}$$

In the trace, $\not{l} + \not{q}$ meets $S(l+q)$ on its left (by cyclicity) and $\not{l}$ meets $S(l)$ on its right, and $S(l)\not{l} = \not{l}S(l) = i$. We obtain, for the first diagram,

$$\begin{aligned} iq_\mu M_1^{\mu\nu\lambda} &= ie^2 \int \frac{d^d l}{(2\pi)^d} \left\{ i\,\mathrm{tr}\left[\gamma_5\, S(l)\gamma^\nu S(l+p)\gamma^\lambda\right] + i\,\mathrm{tr}\left[\gamma_5\gamma^\nu S(l+p)\gamma^\lambda S(l+q)\right]\right\} \\ &= -e^2 \int \frac{d^d l}{(2\pi)^d} \left\{ -\mathrm{tr}\left[\gamma_5\gamma^\lambda S(l)\gamma^\nu S(l+p)\right] + \mathrm{tr}\left[\gamma_5\gamma^\nu S(l)\gamma^\lambda S(l+k)\right]\right\}, \end{aligned} \tag{1.53}$$

where in the first term we have moved γ^λ to the front with cyclicity and one anticommutation with γ_5, and in the second term we have **shifted** the loop momentum, $l \to l - p$. The second diagram gives the same expression with $(p,\nu) \leftrightarrow (k,\lambda)$, which is exactly the negative of the first. The divergence vanishes. In fact, each term vanishes separately: it is a two-point integral which depends on one momentum only, say p, and the trace with γ_5 can only give $\epsilon^{\lambda\alpha\nu\beta} p_\alpha p_\beta = 0$.

Is the argument correct? The shift of the integration variable is innocent for a convergent integral, but here it is applied to a linearly divergent one. With a momentum cut-off, or with Pauli-Villars regulators, the shift leaves behind a finite surface term which does not vanish. We should therefore redo the calculation with a regulator in which the shift is legal: dimensional regularisation (Eq. 1.8). But then we must also use a consistent γ_5, and this is where the anomaly will hide.

1.6.1 The triangle in dimensional regularisation

We use the 't Hooft-Veltman γ_5 of Section 1.3. The external momenta p, k, q and the external indices μ, ν, λ are four-dimensional. The loop momentum has components in all d dimensions,

$$l = \bar{l} + \hat{l}. \tag{1.54}$$

Since γ_5 commutes with $\hat{\not{l}}$ (Eq. 1.19),

$$-\not{l}\gamma_5 = -\left(\bar{\not{l}} + \hat{\not{l}}\right)\gamma_5 = \gamma_5\bar{\not{l}} - \gamma_5\hat{\not{l}} = \gamma_5\not{l} - 2\gamma_5\hat{\not{l}},$$

and Eq. 1.52 is modified to

$$\not{q}\,\gamma_5 = \left(\not{l}+\not{q}\right)\gamma_5 + \gamma_5\,\not{l} - 2\gamma_5\,\hat{\not{l}}. \tag{1.55}$$

The first two terms are treated exactly as before; now the shift of the loop momentum is justified, and their contribution cancels. The third term is new. It gives

$$\begin{aligned} iq_\mu M_1^{\mu\nu\lambda} &= ie^2 \int \frac{d^d l}{(2\pi)^d}\,\mathrm{tr}\left[-2\gamma_5\hat{\not{l}}S(l)\gamma^\nu S(l+p)\gamma^\lambda S(l+q)\right] \\ &= -2e^2 \int \frac{d^d l}{(2\pi)^d}\,\frac{\mathrm{tr}\left[\gamma_5\,\hat{\not{l}}\not{l}\gamma^\nu\left(\not{l}+\not{p}\right)\gamma^\lambda\left(\not{l}+\not{q}\right)\right]}{l^2(l+p)^2(l+q)^2}, \end{aligned} \tag{1.56}$$

where we have used $i\cdot i^3 = 1$. At first sight this is a term of order $(d-4)$, since $\hat{l}$ has only $d-4$ components. However, the integral is divergent, and the divergence can compensate the factor $(d-4)$. Let us see how.

We combine the denominators with Feynman parameters,

$$\frac{1}{l^2(l+p)^2(l+q)^2} = 2\int_0^1 dx_1 dx_2 dx_3\,\delta(1-x_1-x_2-x_3)\,\frac{1}{\left[(l+P)^2-\Delta\right]^3}, \tag{1.57}$$

with $P = x_2 p + x_3 q$ and $\Delta = -x_1 x_3 q^2$, and we shift $l \to l - P$. Since P is four-dimensional, $\hat{l}$ is not affected by the shift. The numerator becomes

$$\mathrm{tr}\left[\gamma_5\,\hat{\not{l}}(\not{l}+\not{a})\,\gamma^\nu\,(\not{l}+\not{b})\,\gamma^\lambda\,(\not{l}+\not{c})\right], \qquad a = -P, \quad b = p-P, \quad c = q-P.$$

Which terms survive? The trace with γ_5 needs four four-dimensional γ matrices: γ^ν, γ^λ and two of $\not{a}, \not{b}, \not{c}$. The remaining factor of the loop momentum must pair with $\hat{l}$, since under symmetric integration

$$\hat{l}^\alpha l^\beta \to \frac{l^2}{d}\,\hat{g}^{\alpha\beta}. \tag{1.58}$$

(Terms with four powers of the loop momentum leave only γ^ν, γ^λ as four-dimensional matrices and vanish in the trace with γ_5.) We then need $\hat{\gamma}_\alpha X \hat{\gamma}^\alpha$, where X is a product of n four-dimensional matrices. Anticommuting $\hat{\gamma}^\alpha$ through X and using $\hat{\gamma}_\alpha\hat{\gamma}^\alpha = \hat{g}^\alpha_\alpha = d-4$,

$$\hat{\gamma}_\alpha X \hat{\gamma}^\alpha = (-1)^n (d-4) X. \tag{1.59}$$

The loop momentum l can sit in the first, second or third slot, with $n = 0, 2, 4$ respectively. In all three cases the sign is positive, and the numerator becomes

$$(d-4)\frac{l^2}{d}\,T, \qquad T = \mathrm{tr}\left[\gamma_5\gamma^\nu\not{b}\gamma^\lambda\not{c}\right] + \mathrm{tr}\left[\gamma_5\not{a}\gamma^\nu\gamma^\lambda\not{c}\right] + \mathrm{tr}\left[\gamma_5\not{a}\gamma^\nu\not{b}\gamma^\lambda\right]. \tag{1.60}$$

Exercise: *Using Eq. 1.14, show that T does not depend on P, and that*

$$T = \mathrm{tr}\left[\gamma_5\gamma^\nu\not{p}\gamma^\lambda\not{k}\right] = -4i\,\epsilon^{\nu\alpha\lambda\beta}p_\alpha k_\beta.$$

The remaining loop integral is the standard one (Chapter 3 of Part II),

$$\int \frac{d^d l}{(2\pi)^d}\,\frac{l^2}{(l^2-\Delta)^3} = \frac{i}{(4\pi)^{d/2}}\,\frac{d}{2}\,\frac{\Gamma\left(2-\frac{d}{2}\right)}{\Gamma(3)}\,\Delta^{\frac{d}{2}-2}. \tag{1.61}$$

Multiplied by $(d-4)/d$, with $d = 4-2\epsilon$ and $\Gamma(\epsilon) = 1/\epsilon + \ldots$, it gives

$$\frac{d-4}{d}\int \frac{d^d l}{(2\pi)^d}\frac{l^2}{(l^2-\Delta)^3} = \frac{i}{(4\pi)^{d/2}}\frac{(d-4)}{4}\Gamma(\epsilon)\,\Delta^{-\epsilon} \xrightarrow{\epsilon\to 0} -\frac{i}{32\pi^2}. \tag{1.62}$$

This is the heart of the matter. A factor $(d-4)$, which we would naively drop, multiplies a $1/(d-4)$ ultraviolet pole and leaves a finite remainder, independent of Δ. The integral over the Feynman parameters is then trivial, $2\int dx_1 dx_2 dx_3\,\delta(1-\sum x_i) = 1$, and

$$iq_\mu M_1^{\mu\nu\lambda} = -2e^2\left(-\frac{i}{32\pi^2}\right)T = \frac{ie^2}{16\pi^2}(-4i)\,\epsilon^{\nu\alpha\lambda\beta}p_\alpha k_\beta = \frac{e^2}{4\pi^2}\,\epsilon^{\nu\alpha\lambda\beta}p_\alpha k_\beta. \tag{1.63}$$

This is symmetric under $(p,\nu)\leftrightarrow(k,\lambda)$, since $\epsilon^{\lambda\beta\nu\alpha} = \epsilon^{\nu\alpha\lambda\beta}$. The second diagram therefore gives an equal contribution, and

$$\langle p,k|\,\partial_\mu j^{\mu 5}(0)\,|0\rangle = \frac{e^2}{2\pi^2}\,\epsilon^{\alpha\nu\beta\lambda}p_\alpha\epsilon_\nu^*(p)\,k_\beta\epsilon_\lambda^*(k). \tag{1.64}$$

Which operator has this matrix element? For two outgoing photons, $\partial_\alpha A_\beta \to ip_\alpha\epsilon_\beta^*(p)$, and the two photons can be assigned to the two field strengths in two ways:

$$\langle p,k|\,\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta}(0)\,|0\rangle = 2\cdot 4\,\epsilon^{\alpha\beta\gamma\delta}\left(ip_\alpha\epsilon_\beta^*(p)\right)\left(ik_\gamma\epsilon_\delta^*(k)\right) = -8\,\epsilon^{\alpha\beta\gamma\delta}p_\alpha\epsilon_\beta^*(p)\,k_\gamma\epsilon_\delta^*(k). \tag{1.65}$$

Comparing with Eq. 1.64, we have found the **Adler-Bell-Jackiw anomaly**,

$$\partial_\mu j^{\mu 5} = -\frac{e^2}{16\pi^2}\,\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta} = \frac{e^2}{2\pi^2}\,\vec{E}\cdot\vec{B}, \qquad \left(\epsilon^{0123} = +1\right). \tag{1.66}$$

In the last step we have used $F_{0i} = E^i$ and $F_{ij} = -\epsilon_{ijk}B^k$, which give $\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta} = -8\vec{E}\cdot\vec{B}$. (With the opposite convention $\epsilon^{0123} = -1$, the sign in front of the ϵFF form changes, but the physical statement in terms of $\vec{E}\cdot\vec{B}$ does not.) It can be shown that Eq. 1.66 is exact: there are no further corrections from higher orders of perturbation theory (Adler-Bardeen theorem).

The right-hand side is the four-dimensional analogue of the two-dimensional anomaly. It is again a total derivative,

$$\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta} = 4\,\partial_\alpha\left(\epsilon^{\alpha\beta\gamma\delta}A_\beta\partial_\gamma A_\delta\right),$$

and again there are field configurations, parallel electric and magnetic fields, for which its integral does not vanish. Notice also the role of the regulator. In dimensional regularisation the vector current is automatically conserved (gauge invariance is respected), and the anomaly appears in the axial current. As in two dimensions, one could move the anomaly into the vector current by adding a finite local counterterm, but one cannot remove it from both.

Exercise: *We can understand Eq. 1.66 as we understood the two-dimensional anomaly, by counting levels.*

- *Integrate Eq. 1.66 to show that* $\Delta N_R - \Delta N_L = \frac{e^2}{2\pi^2}\int d^4x\,\vec{E}\cdot\vec{B}$.
- *In the chiral representation, show that for* $A_0 = 0$ *the Hamiltonian of massless fermions is* $H = \int d^3x\left[\psi_R^\dagger\left(-i\vec{\sigma}\cdot\vec{D}\right)\psi_R - \psi_L^\dagger\left(-i\vec{\sigma}\cdot\vec{D}\right)\psi_L\right]$, *with* $\vec{D} = \vec{\nabla} + ie\vec{A}$.

- *Take a constant magnetic field B along the z-axis, $\vec{A} = (0, Bx^1, A)$, with A constant in space and slowly varying in time. Show that the spectrum of ψ_R consists of Landau levels, and that, for $eB < 0$ (i.e. for a positive product of B and of the charge $-e$ of the quanta), the lowest Landau level is a one-dimensional fermion moving along $+\vec{B}$ with energy $E = k_3 + eA$ and degeneracy $|eB|L^2/(2\pi)$ in a box of side L.*

- *Change A adiabatically by $\Delta A = -2\pi/(eL)$ and use the two-dimensional result to show that the vacuum loses $|eB|L^2/(2\pi)$ right-handed fermions and gains as many left-handed ones. Check that this agrees with the first item.*

Exercise: *Repeat the calculation for massive QED, where $\partial_\mu j^{\mu 5} = 2im\bar{\psi}\gamma_5\psi$ classically. Define the axial current by point splitting,*

$$j^{\mu 5}(x) = \lim_{\delta\to 0} \bar{\psi}\left(x + \tfrac{\delta}{2}\right)\gamma^\mu\gamma_5 \exp\left[+ie\int_{x-\delta/2}^{x+\delta/2} dz\cdot A(z)\right]\psi\left(x - \tfrac{\delta}{2}\right),$$

which is gauge invariant, and take the limit $\delta \to 0$ symmetrically, $\delta^\mu/\delta^2 \to 0$, $\delta^\mu\delta^\nu/\delta^2 \to g^{\mu\nu}/4$. Show that the anomaly of Eq. 1.66 appears as an additional term, independent of m.

We will see in the next chapter how the anomaly controls the decay $\pi^0 \to \gamma\gamma$ and the absence of a ninth light pseudoscalar meson in QCD.

2

Chiral symmetry of the strong interactions

In Chapter 1 we have found that a symmetry of a classical Lagrangian need not survive quantisation: the axial current of a massless fermion coupled to a gauge field has an anomalous divergence. We now want to see what this means for the strong interactions. The u and d quarks are very light compared to the typical hadronic scale of about 1 GeV, and QCD with massless u and d quarks has a large global symmetry, the **chiral symmetry**. Some of it is realised in the spectrum of hadrons as isospin and baryon number. Some of it is spontaneously broken, and the pions are the Goldstone bosons. And one part of it, the axial $U(1)$, is not a symmetry at all, because of the anomaly.

We have met the chiral symmetry of QCD already in Part IV, in Chapters 6 and 7, where we used it to derive the Goldberger-Treiman relation and the pion mass, and where we constructed the non-linear sigma model. We will not repeat these derivations. In this chapter we will first collect the symmetries of QCD systematically and ask which of them are anomalous. This will explain why there is no light ninth pseudoscalar meson, and it will give us the decay rate of the neutral pion into two photons. We will then return to the linear sigma model, now with nucleons, and learn that very different-looking Lagrangians give the same physics. We will then set up chiral perturbation theory, the systematic low-energy effective theory of QCD, including its couplings to photons and W bosons and its first quantum corrections. Finally, we will see that the anomaly of the axial $U(1)$ has a topological meaning: the vacuum of QCD carries an angle θ, which leads to the strong CP problem, to the axion, and to the instantons.

2.1 The chiral symmetries of QCD

Let us keep only the two lightest quarks. With

$$Q = \begin{pmatrix} u \\ d \end{pmatrix}, \qquad Q_L = \frac{1-\gamma_5}{2}\, Q, \qquad Q_R = \frac{1+\gamma_5}{2}\, Q, \tag{2.1}$$

the quark part of the QCD Lagrangian reads

$$\begin{aligned} \mathcal{L} &= \bar{Q}\, i\not{D}\, Q - \bar{Q}\, m\, Q \\ &= \bar{Q}_L\, i\not{D}\, Q_L + \bar{Q}_R\, i\not{D}\, Q_R - \left(\bar{Q}_L\, m\, Q_R + \bar{Q}_R\, m\, Q_L\right), \qquad m = \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix}, \end{aligned} \tag{2.2}$$

where D_μ is the colour covariant derivative. The gluon couples to the vector current and does not mix the two chiralities. Only the mass term connects Q_L with Q_R. If we drop it,

the Lagrangian is invariant under two independent unitary rotations,

$$Q_L \to U_L\, Q_L, \qquad Q_R \to U_R\, Q_R, \qquad U_L, U_R \in U(2). \tag{2.3}$$

This is a larger group than isospin, which is the special case $U_L = U_R$. Separating the phases from the $SU(2)$ parts, the symmetry group of massless two-flavour QCD is

$$SU(2)_L \times SU(2)_R \times U(1)_L \times U(1)_R \;\simeq\; SU(2)_L \times SU(2)_R \times U(1)_V \times U(1)_A, \tag{2.4}$$

where $U(1)_V$ ($U_L = U_R = e^{i\alpha}$) is baryon number and $U(1)_A$ ($U_L = U_R^{-1} = e^{-i\beta}$) is the axial phase rotation. In Chapter 6 of Part IV we have shown explicitly that $SU(2)_L \times SU(2)_R$ is a symmetry, and we have constructed its Noether currents and its Lie algebra. With $\tau^a = \sigma^a/2$, as there, the currents of all eight generators are

$$j^\mu = \bar{Q}\gamma^\mu Q, \qquad j^{\mu a} = \bar{Q}\gamma^\mu \tau^a Q, \qquad j^{\mu 5} = \bar{Q}\gamma^\mu\gamma_5 Q, \qquad j^{\mu 5a} = \bar{Q}\gamma^\mu\gamma_5\tau^a Q. \tag{2.5}$$

The left- and right-handed currents $\bar{Q}_{L,R}\gamma^\mu\tau^a Q_{L,R}$ are the half-difference and half-sum of the isospin current $j^{\mu a}$ and of the axial isospin current $j^{\mu 5a}$.

The vector symmetries are manifest in nature. Baryon number is conserved, and isospin multiplets such as (p, n) or (π^+, π^0, π^-) are nearly degenerate. What about the axial ones? A hadron $|h\rangle$ and its axial partner $X^a|h\rangle$, with X^a the axial charge, would have opposite parity and the same mass. As we have seen in Section 6.5 of Part IV, there are no such parity doublets in the hadron spectrum. The resolution, proposed by Nambu and Jona-Lasinio in 1960, is that the axial symmetries are good symmetries of the Lagrangian which are spontaneously broken by the vacuum. This simple idea gives a surprisingly detailed description of the strong interactions at low energies. We have, however, been cavalier in one respect. All the currents of Eq. 2.5 are conserved classically. We must still ask whether the axial ones survive quantisation, given what we have learned about anomalies. We will see in Section 2.1.3 that the answer is yes for the isotriplet currents $j^{\mu 5a}$, and no for the isosinglet $j^{\mu 5}$.

2.1.1 Why should chiral symmetry break?

Why should the vacuum of QCD break the axial symmetries in the first place? We know a situation in which something similar happens. In a superconductor a weak attraction between electrons is enough to fill the ground state with a condensate of Cooper pairs. In QCD the force between a quark and an antiquark is strongly attractive, and for massless quarks it costs almost no energy to create an extra quark-antiquark pair. We therefore expect the ground state to be filled with a condensate of $q\bar{q}$ pairs. Each pair must have zero total momentum and zero angular momentum. A massless quark with momentum $\vec{p}$ and an antiquark with momentum $-\vec{p}$ have zero total spin only if their helicities are equal. The antiparticle of a right-handed field has opposite helicity, so the pair consists of a left-handed quark and the antiparticle of a right-handed quark (or vice versa), see Fig. 2.1. Such a pair carries net chiral charge.

The vacuum is then characterised by a non-vanishing expectation value of the scalar density,

$$\langle 0|\,\bar{Q}Q\,|0\rangle = \langle 0|\,\bar{Q}_L Q_R + \bar{Q}_R Q_L\,|0\rangle \neq 0. \tag{2.6}$$

This operator is invariant under $U_L = U_R$ but not under $U_L \neq U_R$. The condensate therefore breaks $SU(2)_L \times SU(2)_R$ to the isospin subgroup $SU(2)_V$, and three axial generators are broken (the fourth, $U(1)_A$, will be dealt with by the anomaly). It also mixes the two helicities, exactly like a mass term. A quark moving through this vacuum behaves as if it had a sizeable mass, of a few hundred MeV, even if its mass in the Lagrangian is zero.

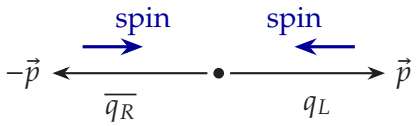


Figure 2.1. A quark-antiquark pair in the QCD vacuum: zero momentum and zero angular momentum require equal helicities, i.e. a left-handed quark paired with the antiparticle of a right-handed quark.

This has a nice consequence for isospin. In the chiral limit the u and d quarks acquire large and equal effective masses from the vacuum. As long as the Lagrangian masses m_u and m_d are small compared to these, u and d behave inside hadrons as if they were degenerate, whatever the ratio m_u/m_d is. Isospin is then an "accident" of the smallness of the quark masses, and not a symmetry of the fundamental Lagrangian. The same argument explains the approximate $SU(3)$ flavour symmetry, since m_s is also smaller than the strong-interaction scale. The quark masses themselves show no trace of flavour symmetry: $m_u : m_d : m_s \approx 1 : 2 : 40$.

2.1.2 The pion decay constant and the pion mass

Three broken generators mean three Goldstone bosons, with the quantum numbers of the axial charges: spin zero, negative parity, isospin one. These are the pions. Their coupling to the broken currents is parametrised by

$$\langle 0|\, j^{\mu 5a}(x)\,|\pi^b(p)\rangle = -ip^\mu f_\pi\, \delta^{ab}\, e^{-ip\cdot x}, \tag{2.7}$$

with relativistically normalised states, $\langle \pi^a(p)|\, \pi^b(p')\rangle = 2p^0(2\pi)^3\delta^{ab}\delta^3(\vec{p}-\vec{p}\,')$. The constant f_π has the dimension of a mass. Comparing with Eq. 6.59 of Part IV, where the states are normalised to $\delta^3(\vec{p}-\vec{p}\,')$, we have $f_\pi = F_\pi/2$; the overall phase, $-i$ here and $+i$ there, is a convention for the sign of the pion field. The phase $-i$ is the one which follows from the U matrix of Section 7.9 of Part IV, $U = \exp(i\vec{\sigma}\cdot\vec{\pi}/f) \to g_L U g_R^\dagger$, which we use in this chapter (see Section 2.3.2). Contracting Eq. 2.7 with p_μ we find

$$\langle 0|\, \partial_\mu j^{\mu 5a}(x)\,|\pi^b(p)\rangle = -p^2 f_\pi\, \delta^{ab}\, e^{-ip\cdot x}. \tag{2.8}$$

If the axial current is conserved, the left-hand side vanishes, and the pion must be massless, $p^2 = 0$, as Goldstone's theorem demands.

Let us now switch on the quark masses. Using the equations of motion

$$i\not{D}\, Q = m\, Q, \qquad -i\,(D_\mu \bar{Q})\,\gamma^\mu = \bar{Q}\, m, \tag{2.9}$$

we have

$$\begin{aligned} \partial_\mu j^{\mu 5a} &= (\partial_\mu \bar{Q})\,\gamma^\mu\gamma_5\tau^a Q + \bar{Q}\gamma^\mu\gamma_5\tau^a\partial_\mu Q \\ &= i\bar{Q}\, m\,\gamma_5\tau^a Q - \bar{Q}\gamma_5\tau^a\, \not{\partial} Q = i\bar{Q}\, m\,\gamma_5\tau^a Q + i\bar{Q}\gamma_5\tau^a m\, Q \\ &= i\bar{Q}\,\{m,\tau^a\}\,\gamma_5 Q \end{aligned} \tag{2.10}$$

(the gluon field drops out, since the current is a colour singlet). For $a = 1, 2$ we have $\{m,\tau^a\} = (m_u + m_d)\tau^a$, and for $a = 3$ the anticommutator contains in addition an isosinglet piece $\frac{1}{2}(m_u - m_d)\mathbf{1}$, which cannot create a pion in the isospin limit. Therefore

$$-m_\pi^2 f_\pi\, \delta^{ab} = (m_u + m_d)\,\langle 0|\, P^a(0)\,|\pi^b(p)\rangle\,, \qquad P^a \equiv \bar{Q}\, i\gamma_5\tau^a Q. \tag{2.11}$$

The pion mass squared is linear in the quark masses. This is the characteristic behaviour of a pseudo-Goldstone boson, which we have derived in a different way in Section 7.3 of Part IV.

What is the matrix element of the pseudoscalar density? To leading order in the quark masses we can compute it in the chiral limit, where the axial charges $X^a = \int d^3\vec{x}\, j^{05a}$ are conserved. Consider

$$\langle 0| \left[X^b, P^a(0)\right] |0\rangle \tag{2.12}$$

and insert a complete set of states between the two operators. Only one-pion states contribute, since X^b acting on the vacuum creates a zero-momentum pion. From Eq. 2.7,

$$\langle 0| X^b \left|\pi^c(p)\right\rangle = -ip^0 f_\pi\, \delta^{bc}(2\pi)^3\delta^3(\vec{p}), \tag{2.13}$$

and with $\langle 0| P^a(0) \left|\pi^b(p)\right\rangle = G_\pi\, \delta^{ab}$, G_π real, we find

$$\begin{aligned}\langle 0| \left[X^b, P^a(0)\right] |0\rangle &= \int \frac{d^3\vec{p}}{(2\pi)^3 2p^0}\Big[\langle 0| X^b \left|\pi^c(p)\right\rangle \left\langle \pi^c(p)\right| P^a(0) |0\rangle \\ &\qquad - \langle 0| P^a(0) \left|\pi^c(p)\right\rangle \left\langle \pi^c(p)\right| X^b |0\rangle \Big] \\ &= -\frac{i}{2} f_\pi G_\pi \delta^{ab} - \frac{i}{2} f_\pi G_\pi \delta^{ab} = -i f_\pi G_\pi \delta^{ab}. \end{aligned} \tag{2.14}$$

On the other hand, the commutator can be computed from the canonical anticommutation relations of the quark fields. We have $\left[X^b, Q\right] = -\gamma_5\tau^b Q$ and $\left[X^b, \bar{Q}\right] = -\bar{Q}\gamma_5\tau^b$, and therefore

$$\left[X^b, P^a\right] = -\bar{Q}\gamma_5\tau^b\, i\gamma_5\tau^a Q - \bar{Q}\, i\gamma_5\tau^a\gamma_5\tau^b Q = -i\bar{Q}\left\{\tau^a, \tau^b\right\} Q = -\frac{i}{2}\delta^{ab}\, \bar{Q}Q. \tag{2.15}$$

Comparing the two results, $G_\pi = \left\langle \bar{Q}Q\right\rangle / (2f_\pi)$. Inserting into Eq. 2.11 and using $\langle \bar{u}u\rangle = \left\langle \bar{d}d\right\rangle = \frac{1}{2}\left\langle \bar{Q}Q\right\rangle$, we obtain the **Gell-Mann-Oakes-Renner relation**,

$$m_\pi^2 f_\pi^2 = -(m_u + m_d)\, \langle \bar{u}u\rangle + O(m_q^2). \tag{2.16}$$

The pion is light because the quarks are light, and its mass measures the product of the quark masses with the condensate. Notice that the condensate must be negative for $m_\pi^2 > 0$ (with positive quark masses). The Lagrangian masses are tiny on hadronic scales: with $m_u + m_d \approx 7$ MeV ($\overline{\text{MS}}$ at 2 GeV), $m_\pi \approx 135$ MeV and $f_\pi \approx 92$ MeV, we find $\langle \bar{u}u\rangle \approx -(280\text{ MeV})^3$. The quark masses are a small perturbation of the strong interactions.

Exercise: *Prove the commutators* $\left[X^b, Q\right] = -\gamma_5\tau^b Q$ *and* $\left[X^b, \bar{Q}\right] = -\bar{Q}\gamma_5\tau^b$ *from the equal-time anticommutator* $\left\{Q_\alpha(t,\vec{x}), Q_\beta^\dagger(t,\vec{y})\right\} = \delta_{\alpha\beta}\delta^3(\vec{x}-\vec{y})$. *Compare with the exercise in Section 6.3 of Part IV.*

How do we measure f_π? The charged pion decays weakly, $\pi^+ \to \mu^+\nu_\mu$, and at low energies the weak interaction couples the leptons to the quark current $\bar{u}\gamma^\mu(1-\gamma_5)d$. The axial part of this current is $\bar{u}\gamma^\mu\gamma_5 d = j^{\mu 51} + ij^{\mu 52}$, whose matrix element between the vacuum and a pion is fixed by Eq. 2.7. The vector part has no such matrix element, by parity. We have done this computation in Section 6.6 of Part IV; let us repeat it as an exercise in the present conventions.

Exercise: *Consider the semileptonic Lagrangian* $\mathcal{L} = -\frac{4G_F}{\sqrt{2}} V_{ud}\left(\bar{\nu}_L\gamma^\mu\mu_L\right)\left(\bar{d}_L\gamma_\mu u_L\right) + \text{h.c.}$, *with* $\psi_L = \frac{1}{2}(1-\gamma_5)\psi$.

- *Show that* $\bar{u}_L\gamma^\mu d_L = \frac{1}{2}\left(j^{\mu 1} + ij^{\mu 2} - j^{\mu 51} - ij^{\mu 52}\right)$.

- *With* $|\pi^+\rangle = \frac{1}{\sqrt{2}}\left(|\pi^1\rangle + i\,|\pi^2\rangle\right)$ *and Eq. 2.7, show that the amplitude for* $\pi^+(p) \to \mu^+(k)\nu_\mu(q)$ *is, up to a phase,* $\mathcal{M} = G_F V_{ud} f_\pi\, \bar{u}(q)\, \not{p}\,(1-\gamma_5)\, v(k)$.
- *Show that*

$$\Gamma(\pi^+ \to \mu^+\nu_\mu) = \frac{G_F^2 |V_{ud}|^2 f_\pi^2}{4\pi}\, m_\mu^2\, m_\pi \left(1 - \frac{m_\mu^2}{m_\pi^2}\right)^2 . \tag{2.17}$$

  Why does the rate vanish for $m_\mu \to 0$*? Compute the ratio* $\Gamma(\pi^+ \to e^+\nu_e)/\Gamma(\pi^+ \to \mu^+\nu_\mu)$ *and compare with experiment.*
- *From the measured lifetime,* $\tau_{\pi^+} = 2.60 \times 10^{-8}\ s$*, determine* f_π.

You should find $f_\pi \approx 93$ MeV, and a ratio of about 1.3×10^{-4}: the decay into an electron, although it has much more phase space, is suppressed by helicity. This is a beautiful confirmation of the $V - A$ structure of the weak current.

2.1.3 Anomalies of the chiral currents

Up to now we have treated the chiral currents classically. In Chapter 1 we found that the axial current of a massless Dirac fermion of charge e, coupled to a photon, has the anomalous divergence of Eq. 1.66,

$$\partial_\mu j^{\mu 5} = -\frac{e^2}{16\pi^2}\, \epsilon^{\alpha\beta\mu\nu} F_{\alpha\beta} F_{\mu\nu} = \frac{e^2}{2\pi^2}\, \vec{E}\cdot\vec{B}, \qquad \left(\epsilon^{0123} = +1\right).$$

The anomaly comes from the triangle diagrams with the axial current at one vertex and the two photons at the others. In QCD the quarks couple to gluons and to photons, and the chiral currents carry flavour matrices. What changes? The triangle diagrams are the same as in QED, see Fig. 2.2, but each vertex now carries a matrix: $\gamma^\mu\gamma_5\tau^a$ for the current, $g\gamma^\alpha t^c$ for a gluon, $eQ\gamma^\alpha$ for a photon, with t^c the colour generators and $Q = \mathrm{diag}(\frac{2}{3}, -\frac{1}{3})$ the quark charge matrix. The anomaly is therefore the abelian one multiplied by the trace of these matrices over colour and flavour. There is one more point. The axial current is gauge invariant, so its divergence must be gauge invariant as well. For gluons, the abelian $F^c_{\alpha\beta}$ in the triangle must be completed into the full non-abelian field strength, $F^c_{\alpha\beta} = \partial_\alpha A^c_\beta - \partial_\beta A^c_\alpha + g f^{cde} A^d_\alpha A^e_\beta$. The terms quadratic in A come from the box diagrams with three gluons, which we do not need to compute.

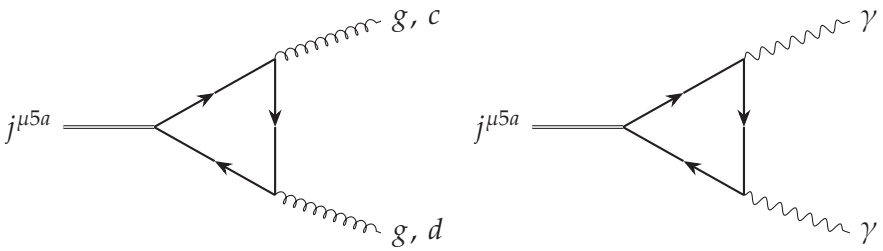


Figure 2.2. Triangle diagrams (plus the crossed ones) giving the anomaly of a flavour axial current: coupling to two gluons (left) and to two photons (right).

For two gluons we obtain

$$\partial_\mu j^{\mu 5a} = -\frac{g^2}{16\pi^2}\, \epsilon^{\alpha\beta\mu\nu} F^c_{\alpha\beta} F^d_{\mu\nu}\ \mathrm{tr}\left[\tau^a t^c t^d\right], \tag{2.18}$$

where the trace runs over colour and flavour. Colour and flavour matrices act on different indices, so the trace factorises,

$$\mathrm{tr}\left[\tau^a t^c t^d\right] = \mathrm{tr}\left[\tau^a\right]\mathrm{tr}\left[t^c t^d\right] = 0. \tag{2.19}$$

The isotriplet axial currents have no QCD anomaly! Their conservation in the chiral limit, and everything we have derived from it (Goldstone pions, f_π, Goldberger-Treiman, the pion mass), is safe. For the isosinglet current $j^{\mu 5}$ the flavour matrix τ^a is replaced by the unit matrix, and with $\mathrm{tr}[t^c t^d] = \frac{1}{2}\delta^{cd}$,

$$\partial_\mu j^{\mu 5} = -\frac{g^2 n_f}{32\pi^2}\,\epsilon^{\alpha\beta\mu\nu} F^c_{\alpha\beta} F^c_{\mu\nu}, \tag{2.20}$$

with $n_f = 2$ the number of light flavours. The $U(1)_A$ symmetry is not a symmetry of quantum QCD. There is therefore no fourth Goldstone boson, and there is no reason for an isosinglet pseudoscalar meson to be light. Indeed, the lightest candidate, the η', has $m_{\eta'} = 958\ \mathrm{MeV} \gg m_\pi$.

Notice that there is a subtlety which we have swept under the carpet. The right-hand side of Eq. 2.20 is a total derivative, as in the two-dimensional example of Chapter 1, $\epsilon^{\alpha\beta\mu\nu} F^c_{\alpha\beta} F^c_{\mu\nu} = \partial_\mu K^\mu$. A total derivative cannot change the action in perturbation theory, and one may wonder whether a modified, conserved axial charge can be constructed from $j^{\mu 5}$ and K^μ. It cannot, because there are gauge field configurations for which $\int d^4x\, \epsilon^{\alpha\beta\mu\nu} F^c_{\alpha\beta} F^c_{\mu\nu}$ is non-zero. These topologically non-trivial configurations, the instantons, will be studied in Sections 2.4 and 2.6.

2.1.4 The decay of the neutral pion into two photons

The isotriplet currents do not feel the gluons, but they do feel the photons. From the right diagram of Fig. 2.2,

$$\partial_\mu j^{\mu 5a} = -\frac{e^2}{16\pi^2}\,\epsilon^{\alpha\beta\mu\nu} F_{\alpha\beta} F_{\mu\nu}\ \mathrm{tr}\left[\tau^a Q^2\right]. \tag{2.21}$$

The colour trace gives a factor $N_c = 3$. The flavour trace is non-zero only for $a = 3$,

$$\mathrm{tr}\left[\tau^a Q^2\right] = N_c\,\delta^{a3}\,\frac{1}{2}\left[\left(\frac{2}{3}\right)^2 - \left(-\frac{1}{3}\right)^2\right] = \frac{N_c}{6}\,\delta^{a3} = \frac{1}{2}\,\delta^{a3}, \tag{2.22}$$

and therefore

$$\partial_\mu j^{\mu 53} = -\frac{e^2}{32\pi^2}\,\epsilon^{\alpha\beta\mu\nu} F_{\alpha\beta} F_{\mu\nu}. \tag{2.23}$$

The current $j^{\mu 53}$ annihilates a π^0. The anomaly must therefore have something to do with the decay $\pi^0 \to \gamma\gamma$. We will now show that it determines the decay amplitude completely, in the limit of massless u and d quarks.

Consider the matrix element of the current between the vacuum and two photons with momenta p, k and polarisations $\epsilon(p), \epsilon(k)$,

$$\left\langle \gamma(p)\gamma(k)\right| j^{\mu 53}(x)\left|0\right\rangle = e^{iq\cdot x}\,\epsilon^*_\nu(p)\epsilon^*_\lambda(k)\,\mathcal{M}^{\mu\nu\lambda}(p,k), \qquad q = p + k. \tag{2.24}$$

In QED we have computed this object from the triangle diagram. Here we cannot compute it, since the quarks are strongly interacting. Instead, let us find out how much of it is fixed by symmetry. The tensor $\mathcal{M}^{\mu\nu\lambda}$ is a pseudotensor, built from p, k and the ϵ-tensor. It must be symmetric under $(p,\nu) \leftrightarrow (k,\lambda)$, and it must satisfy the QED Ward

identities $p_\nu \mathcal{M}^{\mu\nu\lambda} = k_\lambda \mathcal{M}^{\mu\nu\lambda} = 0$. For on-shell photons, $p^2 = k^2 = 0$, and dropping terms proportional to p^ν or k^λ which vanish against the physical polarisations, these conditions leave three structures:

$$\begin{aligned}\mathcal{M}^{\mu\nu\lambda} &= q^\mu \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta \, M_1 + \left(\epsilon^{\mu\nu\alpha\beta} k^\lambda - \epsilon^{\mu\lambda\alpha\beta} p^\nu\right) k_\alpha p_\beta \, M_2 \\ &\quad + \left[\left(\epsilon^{\mu\nu\alpha\beta} p^\lambda - \epsilon^{\mu\lambda\alpha\beta} k^\nu\right) k_\alpha p_\beta - \epsilon^{\mu\nu\lambda\sigma} (p-k)_\sigma \, p\cdot k\right] M_3, \end{aligned} \tag{2.25}$$

with form factors $M_i(q^2)$.

Exercise: *Verify that each of the three structures satisfies the Ward identities on shell.*

Let us now take the divergence. With Eq. 2.24, ∂_μ becomes iq_μ. In the M_2 structure and in the first part of the M_3 structure, $q_\mu = (p+k)_\mu$ is contracted with an ϵ-tensor which already carries p and k, and gives zero. In the last term, $\epsilon^{\mu\nu\lambda\sigma} q_\mu (p-k)_\sigma = -2\epsilon^{\mu\nu\lambda\sigma} p_\mu k_\sigma = -2\epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta$, and with $q^2 = 2p\cdot k$ we find

$$iq_\mu \mathcal{M}^{\mu\nu\lambda} = iq^2 \, \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta \, (M_1 + M_3) \, . \tag{2.26}$$

The divergence is proportional to q^2, and it seems to vanish at $q^2 \to 0$. However, the anomaly equation tells us something else. Taking the matrix element of the right-hand side of Eq. 2.23, where each $F_{\alpha\beta}$ creates one of the photons ($\partial_\alpha \to ip_\alpha$ or ik_α), we find

$$iq_\mu \mathcal{M}^{\mu\nu\lambda} = -\frac{e^2}{4\pi^2} \, \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta , \tag{2.27}$$

which is a non-zero constant. The two equations are compatible only if one of the form factors has a pole at $q^2 = 0$. Is there a massless particle which can produce it? Yes, the pion! The current creates a π^0, which then decays into the two photons, Fig. 2.3.

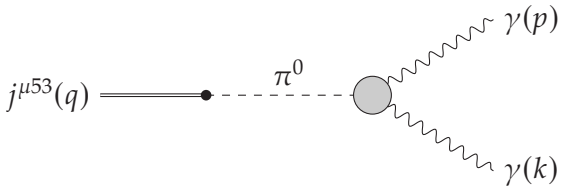


Figure 2.3. The pion pole in the matrix element of the axial current between the vacuum and two photons.

Let us parametrise the decay amplitude by a constant A,

$$i\mathcal{M}(\pi^0 \to \gamma\gamma) = iA \, \epsilon_\nu^*(p) \epsilon_\lambda^*(k) \, \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta , \tag{2.28}$$

which is the only Lorentz and parity invariant form. The diagram of Fig. 2.3 is the product of the amplitude for the current to create the pion, $iq^\mu f_\pi$ from Eq. 2.7, of the pion propagator i/q^2 and of the decay amplitude:

$$\mathcal{M}^{\mu\nu\lambda} \supset \left(iq^\mu f_\pi\right) \frac{i}{q^2} \left(iA \, \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta\right) \quad \rightsquigarrow \quad M_1 = \frac{-i f_\pi A}{q^2} + \left(\text{regular at } q^2 = 0\right) . \tag{2.29}$$

Inserting into Eq. 2.26 at $q^2 \to 0$ gives $iq_\mu \mathcal{M}^{\mu\nu\lambda} = f_\pi A \, \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta$, and the anomaly equation, Eq. 2.27, fixes

$$A = -\frac{e^2}{4\pi^2 f_\pi} = -\frac{\alpha}{\pi f_\pi} . \tag{2.30}$$

This is a remarkable result. The decay amplitude of a hadron, a bound state of strongly interacting quarks, is given exactly (in the chiral limit) by a one-loop coefficient and by f_π, which we have measured in pion decay. All the strong dynamics has dropped out.

It is now easy to compute the decay rate. We have computed the amplitude for a massless pion, but we use the physical kinematics, $q^2 = 2p\cdot k = m_\pi^2$. Summing over polarisations, $\sum_{\rm pol}\left|\epsilon_\nu^*\epsilon_\lambda^*\,\epsilon^{\nu\lambda\alpha\beta}p_\alpha k_\beta\right|^2 = 2(p\cdot k)^2$, and including a factor $\frac{1}{2}$ for two identical photons in the final state, we obtain

$$\begin{aligned}\Gamma(\pi^0\to\gamma\gamma) &= \frac{1}{2m_\pi}\,\frac{1}{2}\,\frac{1}{8\pi}\sum_{\rm pol}\left|\mathcal{M}(\pi^0\to\gamma\gamma)\right|^2 = \frac{1}{32\pi m_\pi}\,A^2\,2(p\cdot k)^2 = \frac{A^2m_\pi^3}{64\pi}\\ &= \frac{\alpha^2 m_\pi^3}{64\pi^3 f_\pi^2}.\end{aligned} \tag{2.31}$$

With $f_\pi = 92.1$ MeV and $m_{\pi^0} = 135.0$ MeV this gives $\Gamma = 7.78$ eV, to be compared with the measured 7.80 ± 0.12 eV. The agreement is impressive, and it is a direct measurement of the coefficient of the anomaly. Notice that the amplitude is proportional to N_c; with $N_c = 1$ the rate would be nine times smaller. Historically, this was one of the first pieces of evidence for three colours.

Exercise: *Verify the polarisation sum*

$$\sum_{\rm pol}\left|\epsilon_\nu^*\epsilon_\lambda^*\,\epsilon^{\nu\lambda\alpha\beta}p_\alpha k_\beta\right|^2 = 2(p\cdot k)^2,$$

using $\sum_{\rm pol}\epsilon_\nu\epsilon_{\nu'}^* \to -g_{\nu\nu'}$.

2.2 The linear sigma model with nucleons

It is instructive to see all the ingredients of spontaneously broken chiral symmetry at work in a model which we can solve in perturbation theory. The simplest such model is the linear sigma model of Gell-Mann and Lévy, with pions and nucleons. We have studied its pion sector in Part IV (Chapter 4 and Section 7.5). Here we add the nucleons and compare two different ways of writing the same theory. The model is renormalisable, and it has the right symmetries. It is not QCD, however: it has too few parameters and fixes, for example, $g_A = 1$. We should regard it as a laboratory, not as the low-energy theory of the strong interactions.

The Lagrangian is

$$\begin{aligned}\mathcal{L}_\sigma &= \frac{1}{2}\left(\partial_\mu\sigma\partial^\mu\sigma + \partial_\mu\vec\pi\cdot\partial^\mu\vec\pi\right) - \frac{\lambda}{4}\left(\sigma^2+\vec\pi^{\,2}-v^2\right)^2\\ &\quad + \bar\psi\, i\partial\!\!\!/\,\psi - g\,\bar\psi\left(\sigma + i\gamma_5\,\vec\tau\cdot\vec\pi\right)\psi, \qquad \psi = \begin{pmatrix} p\\ n\end{pmatrix},\end{aligned} \tag{2.32}$$

where $\vec\tau$ are the Pauli matrices acting on the nucleon isospin. To exhibit the symmetry, we collect the four bosons into a 2×2 matrix,

$$\Sigma = \sigma\,\mathbf{1} + i\,\vec\tau\cdot\vec\pi, \qquad \frac{1}{2}\mathrm{tr}\left(\Sigma^\dagger\Sigma\right) = \sigma^2+\vec\pi^{\,2}. \tag{2.33}$$

Then

$$\begin{aligned}\mathcal{L}_\sigma &= \frac{1}{4}\mathrm{tr}\left(\partial_\mu\Sigma\,\partial^\mu\Sigma^\dagger\right) - \frac{\lambda}{16}\left[\mathrm{tr}\left(\Sigma^\dagger\Sigma\right) - 2v^2\right]^2\\ &\quad + \bar\psi_L\, i\partial\!\!\!/\,\psi_L + \bar\psi_R\, i\partial\!\!\!/\,\psi_R - g\left(\bar\psi_L\,\Sigma\,\psi_R + \bar\psi_R\,\Sigma^\dagger\psi_L\right),\end{aligned} \tag{2.34}$$

which is invariant under

$$\psi_L \to g_L\,\psi_L, \qquad \psi_R \to g_R\,\psi_R, \qquad \Sigma \to g_L\,\Sigma\,g_R^\dagger, \qquad (g_L, g_R) \in SU(2)_L \times SU(2)_R. \tag{2.35}$$

The nucleon fields transform exactly like the quark doublet Q. A nucleon mass term $\bar\psi_L\psi_R$ is forbidden by the symmetry; the nucleon can only get its mass from Σ.

Exercise: *Verify Eq. 2.34, in particular that* $\bar\psi_L\Sigma\psi_R + \bar\psi_R\Sigma^\dagger\psi_L = \bar\psi(\sigma + i\gamma_5\vec\sigma\cdot\vec\pi)\psi$.

2.2.1 Linear representation

For $v^2 > 0$ the potential is minimised by $\sigma^2 + \vec\pi^{\,2} = v^2$, and we choose the vacuum $\langle\sigma\rangle = v$, $\langle\vec\pi\rangle = 0$. This vacuum is invariant under $g_L = g_R$, i.e. under isospin, and breaks the axial transformations. Writing $\sigma = v + \tilde\sigma$, we obtain

$$\begin{aligned}\mathcal{L}_\sigma &= \frac{1}{2}\left(\partial_\mu\tilde\sigma\right)^2 - \lambda v^2\,\tilde\sigma^2 + \frac{1}{2}\left(\partial_\mu\vec\pi\right)^2 - \lambda v\,\tilde\sigma\left(\tilde\sigma^2 + \vec\pi^{\,2}\right) - \frac{\lambda}{4}\left(\tilde\sigma^2 + \vec\pi^{\,2}\right)^2 \\ &\quad + \bar\psi\left(i\partial\!\!\!/ - gv\right)\psi - g\,\bar\psi\left(\tilde\sigma + i\gamma_5\,\vec\sigma\cdot\vec\pi\right)\psi.\end{aligned} \tag{2.36}$$

We read off the spectrum:

$$m_N = gv, \qquad m_\sigma^2 = 2\lambda v^2, \qquad m_\pi = 0. \tag{2.37}$$

The nucleon has acquired its mass from the spontaneous breaking, and the three pions are the massless Goldstone bosons of the three broken axial generators. The pion couples to the nucleon through the pseudoscalar Yukawa coupling $-ig\,\bar\psi\gamma_5\vec\sigma\cdot\vec\pi\,\psi$, so that $g_{\pi NN} = g = m_N/v$.

Something looks strange, however. The pions have non-derivative couplings, $\lambda v\,\tilde\sigma\vec\pi^{\,2}$ and $\lambda(\vec\pi^{\,2})^2$. In Section 4.5 of Part IV we have shown that the amplitudes for the emission of a Goldstone boson vanish when its momentum goes to zero. How can this be? Let us compute the simplest process, $\pi^+\pi^0 \to \pi^+\pi^0$, at tree level. The diagrams are the contact term and the exchange of a $\tilde\sigma$ in the t-channel, Fig. 2.4.

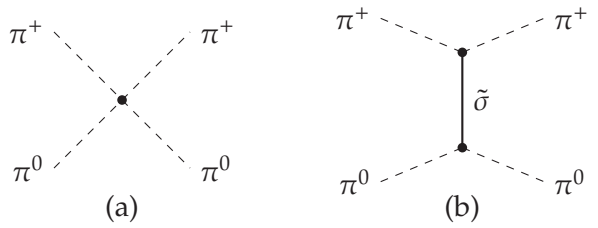


Figure 2.4. Tree-level contributions to $\pi^+\pi^0 \to \pi^+\pi^0$: (a) contact interaction, (b) exchange of the massive scalar.

With $\pi^\pm = (\pi_1 \mp i\pi_2)/\sqrt{2}$ we have $\vec\pi^{\,2} = 2\pi^+\pi^- + \pi_3^2$. The contact term contains $-\lambda\,\pi^+\pi^-\pi_3^2$, which gives the vertex $-2i\lambda$. The cubic terms contain $-\lambda v\,\tilde\sigma\pi_3^2$ and $-2\lambda v\,\tilde\sigma\pi^+\pi^-$, which give the vertices $-2i\lambda v$ each. With $t = (p_+ - p'_+)^2$ the momentum transfer between the two charged pions, we find

$$\begin{aligned}i\mathcal{M} &= -2i\lambda + (-2i\lambda v)^2\,\frac{i}{t - m_\sigma^2} = -2i\lambda\left[1 + \frac{2\lambda v^2}{t - 2\lambda v^2}\right] = -2i\lambda\,\frac{t}{t - m_\sigma^2} \\ &= i\,\frac{t}{v^2} + O\left(\frac{t^2}{v^2 m_\sigma^2}\right).\end{aligned} \tag{2.38}$$

The two diagrams are separately large, of order λ, but they cancel at $t = 0$. What remains vanishes as the momenta go to zero, as the Goldstone theorem requires, and at low energies it does not depend on λ at all. This is the cancellation which we have discussed qualitatively in Section 7.8 of Part IV; here we see it at work explicitly.

2.2.2 Non-linear representation

The cancellation in Eq. 2.38 is automatic, but it is hidden. We would like to choose field variables in which the pions have only derivative couplings, so that the low-energy behaviour is manifest diagram by diagram. As in Section 7.9 of Part IV, we write the matrix Σ in polar form,

$$\Sigma(x) = (v + S(x))\, U(x), \qquad U(x) = \exp\left(\frac{i\,\vec{\sigma}\cdot\vec{\phi}(x)}{v}\right), \tag{2.39}$$

with S a real field and $U \in SU(2)$. To first order $S = \tilde{\sigma}$ and $\vec{\phi} = \vec{\pi}$, but the relation between the two sets of fields is non-linear. Under the chiral group S is invariant and $U \to g_L U g_R^\dagger$, as in Part IV. What do we do with the nucleons? Let us define the square root of U,

$$u(x) = \exp\left(\frac{i\,\vec{\sigma}\cdot\vec{\phi}(x)}{2v}\right), \qquad u^2 = U, \tag{2.40}$$

and new nucleon fields

$$N_L = u^\dagger \psi_L, \qquad N_R = u\,\psi_R. \tag{2.41}$$

Then the Yukawa term becomes

$$\bar{\psi}_L\,\Sigma\,\psi_R = (v+S)\,\bar{N}_L\,u^\dagger\,uu\,u^\dagger N_R = (v+S)\,\bar{N}_L N_R, \tag{2.42}$$

and the pions have disappeared from it! They reappear in the kinetic term of the nucleons, where the derivative acts on u:

$$\begin{aligned}\bar{\psi}_L\,i\partial\!\!\!/\,\psi_L + \bar{\psi}_R\,i\partial\!\!\!/\,\psi_R &= \bar{N}_L\,i\gamma^\mu\left(\partial_\mu + u^\dagger\partial_\mu u\right)N_L + \bar{N}_R\,i\gamma^\mu\left(\partial_\mu + u\,\partial_\mu u^\dagger\right)N_R \\ &= \bar{N}\,i\gamma^\mu\left(\partial_\mu + \Gamma_\mu\right)N + \bar{N}\,i\gamma^\mu\gamma_5\,\Delta_\mu\,N,\end{aligned} \tag{2.43}$$

with

$$\Gamma_\mu = \frac{1}{2}\left(u^\dagger\partial_\mu u + u\,\partial_\mu u^\dagger\right), \qquad \Delta_\mu = \frac{1}{2}\left(u\,\partial_\mu u^\dagger - u^\dagger\partial_\mu u\right). \tag{2.44}$$

The full Lagrangian in the new variables reads

$$\begin{aligned}\mathcal{L}_\sigma &= \frac{1}{2}\left(\partial_\mu S\right)^2 - \lambda v^2 S^2 - \lambda v S^3 - \frac{\lambda}{4}S^4 + \frac{(v+S)^2}{4}\,\mathrm{tr}\left(\partial_\mu U\,\partial^\mu U^\dagger\right) \\ &\quad + \bar{N}\,i\gamma^\mu\left(\partial_\mu + \Gamma_\mu\right)N + \bar{N}\,i\gamma^\mu\gamma_5\,\Delta_\mu\,N - g\,(v+S)\,\bar{N}N.\end{aligned} \tag{2.45}$$

All pion couplings now carry derivatives. Expanding to first order in the pion field, $u^\dagger\partial_\mu u = -u\,\partial_\mu u^\dagger = i\vec{\sigma}\cdot\partial_\mu\vec{\phi}/(2v) + \ldots$, so that $\Gamma_\mu = O(\phi^2)$ and

$$\bar{N}\,i\gamma^\mu\gamma_5\,\Delta_\mu\,N = +\frac{1}{2v}\,\bar{N}\gamma^\mu\gamma_5\,\vec{\sigma}\cdot\partial_\mu\vec{\phi}\;N + O(\phi^3). \tag{2.46}$$

This is a pseudovector pion-nucleon coupling. Comparing with the matrix element of the axial current, the coefficient defines the axial coupling of the nucleon: the general form

is $+\frac{g_A}{2f_\pi}\bar{N}\gamma^\mu\gamma_5\vec{\sigma}\cdot\partial_\mu\vec{\phi}\,N$, and the sigma model gives $g_A = 1$, $f_\pi = v$. Between nucleons on the mass shell, $\bar{u}(p')\gamma^\mu\gamma_5 u(p)\,q_\mu = -2m_N\,\bar{u}(p')\gamma_5 u(p)$ for $q = p - p'$, and the pseudovector coupling is equivalent (up to the sign convention of the pion field) to a pseudoscalar one with

$$g_{\pi NN} = \frac{g_A\, m_N}{f_\pi} = \frac{gv}{v} = g, \tag{2.47}$$

in agreement with the linear representation. This is the Goldberger-Treiman relation of Section 6.7 of Part IV, with $g_A = 1$. In nature $g_A = 1.27$, and the relation $g_{\pi NN} = g_A m_N/f_\pi \approx 13.0$ is satisfied at the level of two per cent. The linear sigma model gets the structure right but not the number.

Exercise: *Show that under a chiral transformation $u \to g_L\, u\, h^\dagger(x) = h(x)\, u\, g_R^\dagger$, which defines a field-dependent $SU(2)$ matrix $h(x)$. Show that $N_{L,R} \to h(x)N_{L,R}$, that $\Gamma_\mu \to h\Gamma_\mu h^\dagger + h\partial_\mu h^\dagger$ and that $\Delta_\mu \to h\Delta_\mu h^\dagger$. Conclude that Eq. 2.45 is chirally invariant, and that the nucleon now transforms non-linearly.*

2.2.3 Representation independence

Let us now recompute $\pi^+\pi^0 \to \pi^+\pi^0$ with the non-linear Lagrangian, Eq. 2.45. Expanding the kinetic term of U to fourth order in the pion fields, we find

$$\frac{v^2}{4}\,\mathrm{tr}\left(\partial_\mu U\,\partial^\mu U^\dagger\right) = \frac{1}{2}\left(\partial_\mu\vec{\phi}\right)^2 + \frac{1}{6v^2}\left[\left(\vec{\phi}\cdot\partial_\mu\vec{\phi}\right)^2 - \vec{\phi}^{\,2}\left(\partial_\mu\vec{\phi}\right)^2\right] + O(\phi^6). \tag{2.48}$$

The exchange of an S is now suppressed: S couples to two pions through $\frac{S}{v}\left(\partial_\mu\vec{\phi}\right)^2$, so diagram (b) of Fig. 2.4 is of order $p^4/(v^2m_\sigma^2)$. To order p^2 only the contact term of Eq. 2.48 contributes. It gives

$$\mathcal{M}(\pi^+\pi^0 \to \pi^+\pi^0) = \frac{t}{v^2} + O(p^4), \tag{2.49}$$

the same as Eq. 2.38. In fact, the result is more general. For $\pi^a\pi^b \to \pi^c\pi^d$ the contact term of Eq. 2.48 gives

$$\mathcal{M}(\pi^a\pi^b \to \pi^c\pi^d) = \frac{s}{v^2}\,\delta^{ab}\delta^{cd} + \frac{t}{v^2}\,\delta^{ac}\delta^{bd} + \frac{u}{v^2}\,\delta^{ad}\delta^{bc}, \tag{2.50}$$

which is **Weinberg's low-energy theorem** for pion-pion scattering in the chiral limit. With $v \to f_\pi$ it is a prediction for QCD.

Exercise: *Derive Eq. 2.48 and Eq. 2.50. Then repeat the computation in the linear representation, including the $\tilde{\sigma}$ exchange in the s, t and u channels, and show that it agrees with Eq. 2.50 up to corrections of order p^2/m_σ^2. Finally, use the parametrisation $U = \left(\sqrt{v^2 - \vec{\phi}^{\,2}} + i\vec{\sigma}\cdot\vec{\phi}\right)/v$ of Section 7.5 of Part IV, whose quartic term is $\frac{1}{2v^2}(\vec{\phi}\cdot\partial_\mu\vec{\phi})^2$, and show that it gives the same amplitude.*

The two representations of the sigma model look very different. They have different interactions and even different Feynman diagrams, and yet they give the same scattering amplitude. This is not an accident. If two fields are related by a non-linear transformation $\phi = \chi\,F(\chi)$ with $F(0) = 1$, then the S-matrix computed with $\mathcal{L}(\phi)$ is the same as the one computed with $\mathcal{L}(\chi F(\chi))$. This **representation independence** (Haag 1958; Chisholm, and Kamefuchi, O'Raifeartaigh and Salam 1961) is easy to understand physically. The condition $F(0) = 1$ means that both fields create the same one-particle states from the vacuum, with the same normalisation. The two fields have the same single-particle poles

and residues, and the S-matrix is read off from exactly these, via the LSZ reduction formula. An experiment cannot know whether a theorist has called the particle ϕ or χ. Off-shell Green's functions, on the other hand, do depend on the choice of fields, and they are not observable.

Representation independence is a very useful freedom. We can choose the variables in which the physics is most transparent. For Goldstone bosons, these are the non-linear variables: the derivative couplings make the low-energy behaviour manifest diagram by diagram, while the linear representation produces it only through cancellations between large terms. The non-linear form also lets us drop the heavy fields S (and, for pion-pion scattering, N) altogether at low energies, which is the starting point of chiral perturbation theory.

2.3 Chiral perturbation theory

We now have all the ingredients to build the effective theory of QCD at low energies in a systematic way. The guiding principle is the one of Section 7.4 of Part IV: the effective theory must have the same symmetries and the same pattern of symmetry breaking as QCD. Beyond that, we want three more things: a systematic expansion in powers of momenta and quark masses, a way to couple the pions to photons and W bosons, and a way to include the explicit breaking by the quark masses. This programme is called **chiral perturbation theory**.

Consider QCD with N_f massless quarks, $q = (u, d, \ldots)$, where $N_f = 2$ or 3:

$$\mathcal{L}^0_{\text{QCD}} = \bar{q}_L \, i\not{D} \, q_L + \bar{q}_R \, i\not{D} \, q_R - \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu}, \qquad q_{R,L} = \frac{1 \pm \gamma_5}{2} \, q. \tag{2.51}$$

Its global symmetry is $SU(N_f)_L \times SU(N_f)_R \times U(1)_V \times U(1)_A$. As in Section 2.1.3, $U(1)_A$ is broken by the anomaly, and $U(1)_V$ is baryon number, which is conserved and plays no role for the mesons. The **chiral group** $G = SU(N_f)_L \times SU(N_f)_R$ is spontaneously broken to $H = SU(N_f)_V$, and there are $N_f^2 - 1$ Goldstone bosons, which live in the coset space G/H. As in Part IV, we collect them in a unitary matrix

$$U(\phi) \to g_L \, U(\phi) \, g_R^\dagger, \qquad (g_L, g_R) \in G. \tag{2.52}$$

There are many ways to parametrise U in terms of fields, which correspond to different coordinates on G/H. By representation independence they all give the same physics. We use the exponential one,

$$U(\phi) = \exp\left(\frac{i \, \lambda^a \phi^a}{f} \right), \tag{2.53}$$

where λ^a are the Pauli matrices for $N_f = 2$ and the Gell-Mann matrices for $N_f = 3$, normalised as $\text{tr}(\lambda^a \lambda^b) = 2\delta^{ab}$, and f is a constant of dimension mass which we will identify shortly. In terms of the physical mesons,

$$\frac{\vec{\sigma} \cdot \vec{\phi}}{\sqrt{2}} = \begin{pmatrix} \frac{\pi^0}{\sqrt{2}} & \pi^+ \\ \pi^- & -\frac{\pi^0}{\sqrt{2}} \end{pmatrix}, \qquad \frac{\lambda^a \phi^a}{\sqrt{2}} = \begin{pmatrix} \frac{\pi^0}{\sqrt{2}} + \frac{\eta_8}{\sqrt{6}} & \pi^+ & K^+ \\ \pi^- & -\frac{\pi^0}{\sqrt{2}} + \frac{\eta_8}{\sqrt{6}} & K^0 \\ K^- & \bar{K}^0 & -\frac{2\eta_8}{\sqrt{6}} \end{pmatrix}. \tag{2.54}$$

For $N_f = 3$ we have an octet of pseudo-Goldstone bosons: the pions, the kaons and the η. The η' is missing, as it should be.

2.3.1 External fields

The real world is not chirally symmetric. The quarks have masses, and the weak interactions couple differently to left- and right-handed quarks. How can we include these effects without losing the power of the symmetry? The trick, due to Gasser and Leutwyler, is to couple the quarks to external classical fields, and to let these fields transform in such a way that the extended Lagrangian is chirally invariant. Consider

$$\mathcal{L} = \mathcal{L}^0_{\rm QCD} + \bar{q}\gamma^\mu \left(v_\mu + a_\mu \gamma_5\right) q - \bar{q}\left(s - ip\,\gamma_5\right) q, \tag{2.55}$$

where v_μ, a_μ, s and p are hermitian $N_f \times N_f$ matrices in flavour space, depending on x. In terms of chiral components,

$$\mathcal{L} = \mathcal{L}^0_{\rm QCD} + \bar{q}_R \gamma^\mu r_\mu q_R + \bar{q}_L \gamma^\mu l_\mu q_L - \bar{q}_R \left(s + ip\right) q_L - \bar{q}_L \left(s - ip\right) q_R, \tag{2.56}$$

with $r_\mu = v_\mu + a_\mu$ and $l_\mu = v_\mu - a_\mu$. The external fields serve two purposes.

First, the photon and the W boson are among them. In the convention $D_\mu = \partial_\mu - ieQA_\mu$ of Parts I and II, the quark couplings to the photon and to the W are reproduced by

$$\begin{aligned} r_\mu &= e\,Q\,A_\mu, \\ l_\mu &= e\,Q\,A_\mu + \frac{e}{\sqrt{2}\sin\theta_W}\left(W^+_\mu T_+ + \text{h.c.}\right), \end{aligned} \tag{2.57}$$

with, for $N_f = 3$,

$$Q = \frac{1}{3}\,\mathrm{diag}(2,-1,-1), \qquad T_+ = \begin{pmatrix} 0 & V_{ud} & V_{us} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{2.58}$$

where V_{ud}, V_{us} are elements of the Cabibbo-Kobayashi-Maskawa matrix. (With the convention $D_\mu = \partial_\mu + ieQA_\mu$ used by Gasser and Leutwyler and by Peskin and Schroeder one replaces $e \to -e$.) The W only couples to left-handed quarks and changes d or s into u. The Z boson can be included in the same way. Green's functions of the electromagnetic and weak currents are obtained as functional derivatives of the generating functional $Z[v,a,s,p]$ with respect to the external fields.

Second, the scalar field s contains the quark masses. The physical Green's functions of QCD are obtained at

$$v_\mu = a_\mu = p = 0, \qquad s = \mathcal{M} = \mathrm{diag}(m_u, m_d, m_s). \tag{2.59}$$

In other words, we treat the quark mass matrix as the value of a spurion field. The point is that $Z[v,a,s,p]$ can now be computed (or rather, constrained) in a manifestly chirally invariant way, and the breaking of the symmetry enters only when we set $s = \mathcal{M}$ at the end.

The Lagrangian of Eq. 2.55 is invariant under **local** chiral transformations, $g_{L,R} = g_{L,R}(x)$, provided that the external fields transform as

$$\begin{aligned} q &\to g_R \frac{1+\gamma_5}{2}\, q + g_L \frac{1-\gamma_5}{2}\, q, \\ r_\mu &\to g_R\, r_\mu\, g_R^\dagger + i g_R\, \partial_\mu g_R^\dagger, \\ l_\mu &\to g_L\, l_\mu\, g_L^\dagger + i g_L\, \partial_\mu g_L^\dagger, \\ s + ip &\to g_R \left(s + ip\right) g_L^\dagger. \end{aligned} \tag{2.60}$$

The external vector fields r_μ and l_μ behave like gauge fields. The effective Lagrangian must therefore be invariant under local chiral transformations too, and derivatives of U must appear in the covariant combination

$$D_\mu U = \partial_\mu U - i l_\mu U + i U r_\mu, \qquad D_\mu U \to g_L\, D_\mu U\, g_R^\dagger, \tag{2.61}$$

together with the field strengths

$$F_R^{\mu\nu} = \partial^\mu r^\nu - \partial^\nu r^\mu - i\,[r^\mu, r^\nu]\,, \qquad F_L^{\mu\nu} = \partial^\mu l^\nu - \partial^\nu l^\mu - i\,[l^\mu, l^\nu]\,. \tag{2.62}$$

The external fields have no kinetic terms; they are classical sources and are not affected by the spontaneous breaking.

Exercise: *Show that Eq. 2.55 is invariant under the transformations of Eq. 2.60, and that $D_\mu U$ transforms as stated in Eq. 2.61.*

2.3.2 Power counting and the leading-order Lagrangian

At low energies the Goldstone bosons interact weakly, since every interaction carries derivatives. We therefore organise the effective Lagrangian as an expansion in derivatives, counting

$$U = O(p^0), \qquad D_\mu U,\ v_\mu,\ a_\mu = O(p), \qquad F_{L,R}^{\mu\nu} = O(p^2). \tag{2.63}$$

Chiral perturbation theory is also an expansion in the quark masses. We could treat the two expansions independently, but it is convenient to combine them. Since m_π^2 is proportional to the quark masses, Eq. 2.16, the standard choice is

$$s,\ p = O(p^2). \tag{2.64}$$

The most general Lagrangian of order p^2 which is invariant under local chiral transformations, parity and charge conjugation is

$$\mathcal{L}_2 = \frac{f^2}{4}\,\mathrm{tr}\left(D_\mu U\, D^\mu U^\dagger + \chi U^\dagger + \chi^\dagger U\right), \qquad \chi = 2B\,(s - ip) \to g_L\,\chi\, g_R^\dagger. \tag{2.65}$$

It has two low-energy constants (LECs), f and B, which symmetry cannot fix. They are determined by QCD and must be taken from experiment (or from lattice QCD). Let us see what they mean. Setting the external fields to their physical values, $\chi = 2B\mathcal{M}$, and expanding in the meson fields, the kinetic term is canonically normalised and the term $\chi U^\dagger + \chi^\dagger U$ gives the meson masses. For $N_f = 2$ and $U + U^\dagger = 2\cos(|\vec\phi|/f)\,\mathbf{1}$,

$$\frac{f^2}{4}\,2B\,\mathrm{tr}\left[\mathcal{M}\left(U + U^\dagger\right)\right] = \text{const} - \frac{1}{2}\,B\,(m_u + m_d)\,\vec\phi^{\,2} + O(\phi^4), \tag{2.66}$$

so that

$$m_\pi^2 = B\,(m_u + m_d). \tag{2.67}$$

Similarly, the axial current obtained from $\mathcal{L}_2$ by differentiation with respect to a_μ is $+f\,\partial_\mu\vec\phi + \ldots$, whose matrix element reproduces Eq. 2.7, and the vacuum expectation value $\langle\bar q q\rangle$ is obtained by differentiating with respect to s:

$$f_\pi = f\,\left(1 + O(m_q)\right), \qquad \langle\bar u u\rangle = -f^2 B\,\left(1 + O(m_q)\right). \tag{2.68}$$

The two LECs are the pion decay constant and the quark condensate in the chiral limit, and Eq. 2.67 is again the Gell-Mann-Oakes-Renner relation. At p^2 the effective Lagrangian

reproduces everything we have derived so far from current algebra, in a much more economical way: the $\pi\pi$ amplitude of Eq. 2.50 (with $v \to f$, plus mass corrections), the coupling to the W, and the pion mass.

Exercise: *For $N_f = 3$, derive the meson masses from $\mathcal{L}_2$ in the isospin limit $m_u = m_d = \hat{m}$: $m_\pi^2 = 2B\hat{m}$, $m_K^2 = B(\hat{m} + m_s)$, $m_{\eta_8}^2 = \frac{2}{3}B(\hat{m} + 2m_s)$. Show that they satisfy the Gell-Mann-Okubo relation $4m_K^2 = m_\pi^2 + 3m_{\eta_8}^2$, and check it against the measured masses.*

2.3.3 Loops and the Lagrangian of order p^4

The Lagrangian $\mathcal{L}_2$ is not renormalisable. Can we use it beyond tree level? Yes, in the sense of Chapter 8 of Part IV. The key observation, due to Weinberg, is that loops are suppressed by powers of the momenta. Consider a connected diagram with L loops and N_d vertices from the part of the Lagrangian of order p^d. If we scale all external momenta and masses by a factor t, the amplitude scales as t^D, with the **chiral dimension**

$$D = 2 + 2L + \sum_d N_d\,(d-2). \tag{2.69}$$

Since $d \geq 2$, every loop costs two more powers of p. At order p^2 we have tree diagrams of $\mathcal{L}_2$ only. At order p^4 we have one-loop diagrams with vertices from $\mathcal{L}_2$, and tree diagrams with exactly one vertex from a Lagrangian $\mathcal{L}_4$ of order p^4. In dimensional regularisation the divergences of the one-loop diagrams are polynomials in momenta and masses of order p^4, with the symmetries of $\mathcal{L}_2$, and they can be absorbed into the coefficients of $\mathcal{L}_4$. The theory is renormalisable order by order in the momentum expansion, at the price of new LECs at each order. The expansion parameter is $p^2/(4\pi f)^2$, with $4\pi f \approx 1.2$ GeV.

Exercise: *Prove Eq. 2.69. Use that a vertex of order p^d scales as t^d, a meson propagator as t^{-2}, and each loop integration as t^4, together with the topological identity $L = I - V + 1$ for I internal lines and V vertices.*

For $N_f = 2$, the most general $\mathcal{L}_4$ was written down by Gasser and Leutwyler (1984). It contains seven LECs $l_1, \ldots, l_7$ (and some contact terms of the external fields alone). The first three are

$$\mathcal{L}_4 = \frac{l_1}{4}\left[\mathrm{tr}\left(D_\mu U D^\mu U^\dagger\right)\right]^2 + \frac{l_2}{4}\,\mathrm{tr}\left(D_\mu U D_\nu U^\dagger\right)\mathrm{tr}\left(D^\mu U D^\nu U^\dagger\right) + \frac{l_3}{16}\left[\mathrm{tr}\left(\chi U^\dagger + \chi^\dagger U\right)\right]^2 + \ldots \tag{2.70}$$

The first two contribute to $\pi\pi$ scattering at order p^4, the third to the pion mass. For $N_f = 3$ there are ten such constants, $L_1, \ldots, L_{10}$.

Let us see how this works for the pion mass. We use, for this computation, the parametrisation $U = \left(\sigma + i\vec{\tau}\cdot\vec{\phi}\right)/f$ with $\sigma = \sqrt{f^2 - \vec{\phi}^{\,2}}$, and set $m_u = m_d = \hat{m}$, $M^2 \equiv 2B\hat{m}$. Expanding $\mathcal{L}_2$ to fourth order in $\vec{\phi}$, we find

$$\mathcal{L}_2 = \frac{1}{2}\left(\partial_\mu\vec{\phi}\right)^2 - \frac{1}{2}M^2\vec{\phi}^{\,2} + \frac{1}{2f^2}\left(\vec{\phi}\cdot\partial_\mu\vec{\phi}\right)^2 - \frac{M^2}{8f^2}\left(\vec{\phi}^{\,2}\right)^2 + O(\phi^6). \tag{2.71}$$

At one loop, the pion self-energy comes from the tadpole diagram, in which two of the four fields of the quartic terms are contracted into a closed loop. The contractions we need are

$$\left\langle \phi^a \phi^b \right\rangle = \delta^{ab} I, \qquad \left\langle \partial_\mu\phi^a\, \partial^\mu\phi^b \right\rangle = \delta^{ab} M^2 I, \qquad \left\langle \phi^a\, \partial_\mu\phi^b \right\rangle = 0, \tag{2.72}$$

with

$$I = \int \frac{d^dk}{(2\pi)^d}\,\frac{i}{k^2 - M^2 + i\epsilon} = \frac{M^2}{16\pi^2}\left[-\frac{1}{\epsilon} + \gamma_E - \ln 4\pi + \ln\frac{M^2}{\mu^2} - 1\right]. \tag{2.73}$$

(The middle contraction follows from $\int d^dk\, k^2/(k^2-M^2) = M^2 \int d^dk/(k^2-M^2)$ in dimensional regularisation.) Contracting all pairs in the quartic terms, with three pion fields, we obtain the one-loop correction to the quadratic Lagrangian,

$$\delta\mathcal{L} = \frac{I}{2f^2}\left[\left(\partial_\mu\vec{\phi}\right)^2 + M^2\vec{\phi}^{\,2}\right] - \frac{10\,M^2 I}{8f^2}\,\vec{\phi}^{\,2} = \frac{1}{2}\,\frac{I}{f^2}\left(\partial_\mu\vec{\phi}\right)^2 - \frac{1}{2}\,\frac{3M^2 I}{2f^2}\,\vec{\phi}^{\,2}. \tag{2.74}$$

The first term renormalises the wave function, the second the mass. Adding the tree-level contribution of the l_3 term, $-\frac{l_3 M^4}{f^2}\vec{\phi}^{\,2}$, the physical mass is

$$m_\pi^2 = \frac{M^2 + \frac{3M^2 I}{2f^2} + \frac{2l_3M^4}{f^2}}{1+\frac{I}{f^2}} = M^2\left[1 + \frac{M^2}{32\pi^2 f^2}\ln\frac{M^2}{\mu^2} + \frac{2\,l_3^r(\mu)M^2}{f^2}\right] + O(M^6), \tag{2.75}$$

where the pole and the constants of I have been absorbed into the renormalised coupling $l_3^r(\mu)$. The dependence on the arbitrary scale μ cancels between the logarithm and $l_3^r(\mu)$. The logarithm, $M^4\ln M^2$, is a **chiral logarithm**. It is non-analytic in the quark masses, and no term in a Lagrangian could produce it. It is a genuine prediction of the loop expansion, with a coefficient fixed by f alone. The Gell-Mann-Oakes-Renner relation, Eq. 2.67, is the first term of this expansion.

Exercise: *Repeat the computation with the exponential parametrisation $U = \exp(i\vec{\sigma}\cdot\vec{\phi}/f)$. The quartic terms, and the contributions to wave-function and mass renormalisation separately, are different. Show that m_π^2 is the same.*

2.3.4 Baryons

Finally, let us include the nucleons. In the non-linear sigma model we have learned how to do it. We need the square root u of U, $u^2 = U$, which transforms as $u \to g_L\, u\, h^\dagger(x) = h(x)\, u\, g_R^\dagger$, and nucleon fields which transform with the field-dependent matrix $h(x)$ of the unbroken subgroup, $N \to h(x)N$. With external fields, the building blocks are the connection and the axial vector field

$$\Gamma_\mu = \frac{1}{2}\left[u\left(\partial_\mu - ir_\mu\right)u^\dagger + u^\dagger\left(\partial_\mu - il_\mu\right)u\right], \qquad u_\mu = i\left[u\left(\partial_\mu - ir_\mu\right)u^\dagger - u^\dagger\left(\partial_\mu - il_\mu\right)u\right], \tag{2.76}$$

which transform as $\Gamma_\mu \to h\Gamma_\mu h^\dagger + h\partial_\mu h^\dagger$ and $u_\mu \to h\,u_\mu h^\dagger$. Without external fields, Γ_μ and $u_\mu = 2i\Delta_\mu$ are those of Section 2.2.2. Counting the nucleon mass as of order one and the derivative of the nucleon field, $(i\partial\!\!\!/ - m_N)N$, as of order p, the chiral Lagrangian at lowest order is

$$\mathcal{L}_{\pi N}^{(1)} = \bar{N}\left(i\not{D} - m_N + \frac{g_A}{2}\,\gamma^\mu\gamma_5\,u_\mu\right)N, \qquad D_\mu = \partial_\mu + \Gamma_\mu, \qquad (N_f = 2). \tag{2.77}$$

This has exactly the form we found in the linear sigma model, Eq. 2.45, but now m_N and g_A are two independent LECs: the nucleon mass and the axial coupling in the chiral limit. The sigma model fixes $g_A = 1$; QCD does not, and neutron β decay gives $g_A = 1.27$. Expanding $u_\mu = +\vec{\sigma}\cdot\partial_\mu\vec{\phi}/f + \ldots$ we recover the pseudovector coupling and the Goldberger-Treiman relation $g_{\pi NN} = g_A m_N/f$.

For $N_f = 3$ the lightest baryons form an octet, which we collect in a traceless 3×3 matrix,

$$B = \begin{pmatrix} \frac{\Sigma^0}{\sqrt{2}} + \frac{\Lambda}{\sqrt{6}} & \Sigma^+ & p \\ \Sigma^- & -\frac{\Sigma^0}{\sqrt{2}} + \frac{\Lambda}{\sqrt{6}} & n \\ \Xi^- & \Xi^0 & -\frac{2\Lambda}{\sqrt{6}} \end{pmatrix}, \qquad B \to h\,B\,h^\dagger, \tag{2.78}$$

and the lowest-order Lagrangian is

$$\mathcal{L}^{(1)}_{MB} = \mathrm{tr}\left[\bar{B}\left(i\not{D} - m_0\right)B + \frac{D}{2}\,\bar{B}\gamma^\mu\gamma_5\left\{u_\mu, B\right\} + \frac{F}{2}\,\bar{B}\gamma^\mu\gamma_5\left[u_\mu, B\right]\right], \tag{2.79}$$

with $D_\mu B = \partial_\mu B + \left[\Gamma_\mu, B\right]$. There are now two axial couplings, $D \approx 0.80$ and $F \approx 0.46$, which describe the semileptonic decays of all the octet baryons. Restricting to the nucleon, one finds $g_A = D + F$.

Exercise: *Show that the pseudovector coupling in Eq. 2.77, at first order in the pion field, reproduces the Goldberger-Treiman relation. Then show that the D and F terms of $\mathcal{L}^{(1)}_{MB}$, restricted to the proton and neutron and to the pion fields, reduce to Eq. 2.77 with $g_A = D + F$.*

Notice that the nucleon mass $m_N \approx 940$ MeV does not vanish in the chiral limit. It is a "dynamical" mass, generated by the strong interactions and by the spontaneous breaking of chiral symmetry, as in the sigma model. The quark masses contribute only a small correction, which can again be computed in chiral perturbation theory. Unlike the meson case, the baryon mass is not small compared to the chiral scale $4\pi f$, and the power counting for baryons requires more care; we will not pursue it here.

2.4 The θ vacuum

In Chapter 1 we have found that the axial $U(1)$ current is not conserved (Section 1.6), and in Section 2.1.3 we have seen that the same is true for the flavour-singlet axial current of QCD, Eq. 2.20. Its divergence is proportional to the operator $G^a_{\mu\nu}\tilde{G}^{a\mu\nu}$. This operator is a total derivative, and a total derivative in the Lagrangian does not change the classical equations of motion. It is very tempting to conclude that it does not matter in the quantum theory either: the action changes by a surface term, which we usually set to zero by demanding that the fields vanish at infinity. Is this right? For QED it is. For a non-abelian gauge theory such as QCD it is not! The gauge fields of the vacuum need not vanish at infinity; they only need to be a pure gauge, and pure gauges of a non-abelian group come in distinct classes which cannot be deformed into each other. In this section we will first classify the vacuum configurations of the gluon field, and then we will see that the true vacuum of QCD carries a new parameter, an angle θ, which acts as a new coupling constant of the theory.

Small and large gauge transformations

We recall the gauge transformation of the gluon field (Eq. 3.26),

$$A_\mu \to A'_\mu = U(x)A_\mu U^\dagger(x) + \frac{i}{g}U(x)\left(\partial_\mu U^\dagger(x)\right), \qquad U(x) \in SU(3). \tag{2.80}$$

Usually we think of $U(x)$ as the exponential of generators, $U(x) = \exp\left[i\alpha^a(x)T^a\right]$. By rescaling $\alpha^a \to \lambda\alpha^a$, $0 \le \lambda \le 1$, such a transformation can be deformed continuously into the identity. We will call these transformations "small". Are there gauge transformations which are not of this type? A familiar example of group elements which are not connected to the identity is the parity matrix $P = \mathrm{diag}(-1,-1,-1)$, which is an element of $O(3)$ but not of $SO(3)$: no continuous path of rotations leads from $\mathbf{1}$ to P. For gauge transformations the analogous phenomenon is subtler. The group $SU(3)$ itself is connected, but the space of **functions** $U(x)$ with fixed behaviour at infinity is not.

Let us choose the temporal gauge, $A_0 = 0$. The remaining freedom consists of time-independent gauge transformations $U(\vec{x})$. Configurations of zero energy ($G_{\mu\nu} = 0$) are

pure gauges,

$$A_i(\vec{x}) = \frac{i}{g} U(\vec{x})\, \partial_i U^\dagger(\vec{x}), \qquad A_0 = 0, \tag{2.81}$$

which are obtained from $A_\mu = 0$ by a gauge transformation. We shall consider gauge transformations which approach a constant at spatial infinity, which we can take to be $U(\vec{x}) \to \mathbf{1}$ for $|\vec{x}| \to \infty$. All points at infinity are then mapped to the same group element and we may regard them as one single point. Space $\mathbb{R}^3$ with its point at infinity is topologically a three-dimensional sphere S^3. It is sufficient to consider an $SU(2)$ subgroup of $SU(3)$, for example the one generated by $\lambda_1, \lambda_2, \lambda_3$. A general $SU(2)$ matrix,

$$U = a\,\mathbf{1} + i\,\vec{b}\cdot\vec{\tau}, \qquad a^2 + \vec{b}^2 = 1, \tag{2.82}$$

is parametrised by a point on a sphere S^3 in four dimensions. A gauge transformation with $U \to \mathbf{1}$ at infinity is therefore a map $S^3 \to S^3$. Such maps are characterised by an integer, the number of times the second sphere is covered when we move once over the first sphere. Two maps with different covering numbers cannot be deformed into each other continuously. In mathematical parlance, $\pi_3(S^3) = \pi_3(SU(3)) = \mathbb{Z}$. Gauge transformations with a non-zero covering number are called "large".

An explicit example is

$$U_1(\vec{x}) = \frac{\vec{x}^{\,2} - d^2}{\vec{x}^{\,2} + d^2}\,\mathbf{1} + \frac{2id\;\vec{\tau}\cdot\vec{x}}{\vec{x}^{\,2} + d^2}, \tag{2.83}$$

where $d > 0$ is an arbitrary length. It is easy to check that $a^2 + \vec{b}^2 = 1$. At the origin $U_1 = -\mathbf{1}$, and at infinity $U_1 \to \mathbf{1}$. As $|\vec{x}|$ grows from 0 to ∞ along all directions, the image point $(a, \vec{b})$ sweeps the whole sphere S^3 exactly once. Substituting into Eq. 2.81 we find the pure gauge field

$$A_j^{(1)}(\vec{x}) = -\frac{2d}{g\left(\vec{x}^{\,2} + d^2\right)^2}\left[\tau_j\left(d^2 - \vec{x}^{\,2}\right) + 2x_j\left(\vec{\tau}\cdot\vec{x}\right) - 2d\left(\vec{x}\times\vec{\tau}\right)_j\right], \tag{2.84}$$

with $\partial_j \equiv \partial/\partial x^j$ and x_j the components of $\vec{x}$. Notice that this field does not vanish as $d \to 0$; rather, it becomes singular at the origin. We cannot remove it by any continuous deformation.

We can attach to each pure gauge configuration a number which counts the coverings, the **winding number**:

$$n = \frac{ig^3}{24\pi^2}\int d^3x\;\epsilon^{ijk}\,\mathrm{Tr}\left[A_iA_jA_k\right] = \frac{1}{24\pi^2}\int d^3x\;\epsilon^{ijk}\,\mathrm{Tr}\left[(U\partial_iU^\dagger)(U\partial_jU^\dagger)(U\partial_kU^\dagger)\right]. \tag{2.85}$$

For the field of Eq. 2.84 the integrand is spherically symmetric,

$$\frac{ig^3}{24\pi^2}\,\epsilon^{ijk}\,\mathrm{Tr}\left[A_i^{(1)}A_j^{(1)}A_k^{(1)}\right] = \frac{4d^3}{\pi^2\left(\vec{x}^{\,2} + d^2\right)^3},$$

and we obtain $n = 1$.

Exercise: *Derive Eq. 2.84 from Eq. 2.83, compute the winding number density and verify that $n = 1$. What do you obtain for $d < 0$?*

Exercise: *Show that the winding number is invariant under small gauge transformations and that it is additive, $n[UV] = n[U] + n[V]$. (Hint: write the variation of the integrand as a total derivative.) Conclude that $U_m(\vec{x}) \equiv \left[U_1(\vec{x})\right]^m$ has winding number m.*

Therefore, the zero-energy configurations of the gluon field fall into disjoint classes, labelled by an integer n. Within a class we can move with small gauge transformations; from one class to the next we need a large gauge transformation such as U_1.

Tunnelling between classes and the topological charge

Let us now consider a configuration of the gluon field which at $t \to -\infty$ is a pure gauge of winding number n_-, evolves in some way in the interval in between, and at $t \to +\infty$ is a pure gauge of winding number n_+. In between, the field cannot be a pure gauge all the time, since we cannot deform continuously one class into the other. It must pass over an energy barrier. We will see in Section 2.6 that quantum mechanics allows such transitions by tunnelling.

Consider the integral of the topological charge density

$$q(x) \equiv \frac{g^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{a\mu\nu} = \frac{g^2}{16\pi^2} \mathrm{Tr}\left[G_{\mu\nu}\tilde{G}^{\mu\nu}\right], \qquad \tilde{G}^{a\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} G^a_{\rho\sigma}, \tag{2.86}$$

with $\epsilon^{0123} = +1$. In terms of the colour electric and magnetic fields, $E^{ai} = G^{a\,i0}$, $B^{ai} = -\frac{1}{2}\epsilon^{ijk}G^a_{jk}$, we have $G^a_{\mu\nu}\tilde{G}^{a\mu\nu} = -4\,\vec{E}^a \cdot \vec{B}^a$. The density $q(x)$ is a total derivative,

$$q(x) = \partial_\mu \mathcal{K}^\mu, \qquad \mathcal{K}^\mu = \frac{g^2}{8\pi^2}\,\epsilon^{\mu\nu\rho\sigma}\,\mathrm{Tr}\left[A_\nu \partial_\rho A_\sigma - \frac{2ig}{3} A_\nu A_\rho A_\sigma\right]. \tag{2.87}$$

$\mathcal{K}^\mu$ is known as the **Chern-Simons current**. Notice that it is not gauge invariant.
Exercise: *Verify Eq. 2.87 using* $G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig\left[A_\mu, A_\nu\right]$, *the cyclicity of the trace and the antisymmetry of* $\epsilon^{\mu\nu\rho\sigma}$.

We now integrate over all space-time. In the temporal gauge, and with fields that become a pure gauge with $U \to \mathbf{1}$ at spatial infinity, only the surfaces $t = \pm\infty$ contribute:

$$\int d^4x\; q(x) = \int d^4x\; \partial_\mu \mathcal{K}^\mu = \int d^3x\; \mathcal{K}^0\Big|_{t=-\infty}^{t=+\infty}. \tag{2.88}$$

At $t = \pm\infty$ the field is a pure gauge, $G_{\mu\nu} = 0$, so that $\partial_j A_k - \partial_k A_j = ig\left[A_j, A_k\right]$. Then

$$\epsilon^{ijk}\,\mathrm{Tr}\left[A_i \partial_j A_k\right] = \frac{1}{2}\epsilon^{ijk}\,\mathrm{Tr}\left[A_i\left(\partial_j A_k - \partial_k A_j\right)\right] = ig\,\epsilon^{ijk}\,\mathrm{Tr}\left[A_i A_j A_k\right],$$

and

$$\int d^3x\,\mathcal{K}^0 = \frac{g^2}{8\pi^2}\left(ig - \frac{2ig}{3}\right)\int d^3x\,\epsilon^{ijk}\,\mathrm{Tr}\left[A_i A_j A_k\right] = n. \tag{2.89}$$

The final result reads,

$$\frac{g^2}{32\pi^2}\int d^4x\; G^a_{\mu\nu}\tilde{G}^{a\mu\nu} = n_+ - n_-. \tag{2.90}$$

This is a remarkable statement. The integral of a local density, which is a total derivative, is not zero. It is an integer which counts how many times the gluon field has changed its class between the far past and the far future.

The vacuum of QCD

What does this imply for the vacuum of QCD? Let us denote by $|n\rangle$ a state of lowest energy whose gluon field is concentrated around the pure gauges of winding number n. The large gauge transformation U_1 is implemented on the Hilbert space by a unitary operator $\mathcal{T}_1$, with

$$\mathcal{T}_1\,|n\rangle = |n+1\rangle\,. \tag{2.91}$$

Since U_1 is a gauge transformation, $\mathcal{T}_1$ commutes with the Hamiltonian and with every gauge invariant operator. The states $|n\rangle$ are therefore not suitable vacuum states; they are not even eigenstates of $\mathcal{T}_1$. We can however form the superposition

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{-in\theta} |n\rangle , \tag{2.92}$$

which yields

$$\mathcal{T}_1 |\theta\rangle = \sum_n e^{-in\theta} |n+1\rangle = e^{i\theta} |\theta\rangle . \tag{2.93}$$

The state $|\theta\rangle$ is gauge invariant up to a phase, which is all we need for a physical state. The angle $\theta \in [0, 2\pi)$ is a free parameter. Notice that a gauge invariant operator $\mathcal{O}$ cannot connect different θ vacua: since $\mathcal{T}_1$ commutes with $\mathcal{O}$,

$$e^{i\theta} \langle\theta'| \mathcal{O} |\theta\rangle = \langle\theta'| \mathcal{O}\, \mathcal{T}_1 |\theta\rangle = \langle\theta'| \mathcal{T}_1 \mathcal{O} |\theta\rangle = e^{i\theta'} \langle\theta'| \mathcal{O} |\theta\rangle ,$$

where we have used $\mathcal{T}_1^\dagger |\theta'\rangle = e^{-i\theta'} |\theta'\rangle$. Thus $\langle\theta'| \mathcal{O} |\theta\rangle = 0$ for $\theta' \neq \theta$. Each value of θ defines its own world, which never communicates with the worlds of other values of θ.

This is very different from the degenerate vacua of a spontaneously broken symmetry, which we recalled in Part IV, Chapter 1. There, the different vacua $\langle\phi\rangle$ belong to the same theory and are related by a symmetry of the Lagrangian. Here, each value of θ specifies a different theory, in the same way as different values of Λ_{QCD} (or of the quark masses) specify different theories. To define QCD we need the Lagrangian, the coupling constant and, in addition, the vacuum angle θ.

The θ term in the Lagrangian

How can we compute with a θ vacuum? Consider a transition amplitude between the vacuum at $t \to -\infty$ and the vacuum at $t \to +\infty$, with the insertion of some operator X. Inserting Eq. 2.92 we have

$${}_{\text{out}}\langle\theta| X |\theta\rangle_{\text{in}} = \sum_{n_+, n_-} e^{i(n_+ - n_-)\theta} {}_{\text{out}}\langle n_+| X |n_-\rangle_{\text{in}} . \tag{2.94}$$

In the path integral, the amplitude ${}_{\text{out}}\langle n_+| X |n_-\rangle_{\text{in}}$ is an integral over gauge fields which interpolate between the winding numbers n_- and n_+. By Eq. 2.90, we can trade the phase $e^{i(n_+ - n_-)\theta}$ for the exponential of a local integral. The sum over all $n_\pm$ then becomes an unrestricted path integral:

$${}_{\text{out}}\langle\theta| X |\theta\rangle_{\text{in}} = \int \mathcal{D}A\, \mathcal{D}\bar{\psi}\, \mathcal{D}\psi\; X \;\exp\left\{ iS_{\text{QCD}} + i\theta \frac{g^2}{32\pi^2} \int d^4x\, G^a_{\mu\nu} \tilde{G}^{a\mu\nu} \right\} . \tag{2.95}$$

The effect of the θ vacuum is therefore equivalent to a new term in the QCD Lagrangian,

$$\mathcal{L}_{\text{QCD}} = \mathcal{L}^{(\theta=0)}_{\text{QCD}} + \theta\, \frac{g^2}{32\pi^2}\, G^a_{\mu\nu} \tilde{G}^{a\mu\nu} , \tag{2.96}$$

with θ playing the role of a coupling constant. We can now forget about the vacuum structure and use the ordinary path integral with this Lagrangian.

Notice that $\vec{E}^a$ is odd and $\vec{B}^a$ is even under parity, while under time reversal $\vec{E}^a$ is even and $\vec{B}^a$ is odd. The operator $G\tilde{G} \propto \vec{E}^a \cdot \vec{B}^a$ is therefore odd under P and under T (and, by the CPT theorem, odd under CP). A non-zero θ violates CP in the strong interaction.

This has a measurable consequence: the neutron, which is a C-even and P-even bound state of quarks only, acquires an electric dipole moment d_n. A rough estimate gives $d_n \sim 10^{-16}\,\theta\; e$ cm. The experimental bound is $|d_n| < 1.8\times10^{-26}\, e$ cm (PSI, 2020). Therefore,

$$|\theta| \lesssim 10^{-10}. \tag{2.97}$$

Exercise: *The electric dipole moment of the neutron is defined through*

$$\left\langle n(p')\right| J^\mu_{\rm em} \left|n(p)\right\rangle_{\rm edm} = i d_n\, \bar u(p')\sigma^{\mu\nu}(p'-p)_\nu \gamma_5 u(p).$$

Remove the θ term with a chiral rotation of the quarks (see below), so that θ appears as a CP-violating term $\propto \theta\,\bar\psi i\gamma_5\psi$ with coefficient $-\theta\left(m_u^{-1}+m_d^{-1}+m_s^{-1}\right)^{-1}$ (up to a convention-dependent sign). Estimate d_n in first-order perturbation theory, keeping only the lowest negative-parity nucleon resonance $N(1535)$ as intermediate state and approximating the matrix elements by $e\,r_p$ and 1 respectively. You should find d_n of order $10^{-16}\,\theta\; e$ cm.

Connection with chiral rotations

We now come back to the axial anomaly. We recall from Eq. 2.20 that, for N_f massless quarks, the flavour-singlet axial current

$$j_5^\mu = \sum_{f=1}^{N_f} \bar\psi_f \gamma^\mu\gamma_5\psi_f \tag{2.98}$$

satisfies

$$\partial_\mu j_5^\mu = -N_f \frac{g^2}{16\pi^2}\, G^a_{\mu\nu}\tilde G^{a\mu\nu} = -2N_f\, q(x) = N_f\,\frac{g^2}{4\pi^2}\,\vec E^a\cdot\vec B^a. \tag{2.99}$$

This is the non-abelian version of the Adler-Bell-Jackiw anomaly, Eq. 1.66: we replace $e^2F_{\alpha\beta}F_{\gamma\delta}$ by $g^2G^a_{\alpha\beta}G^b_{\gamma\delta}\,\mathrm{Tr}(T^aT^b) = \frac{g^2}{2}G^a_{\alpha\beta}G^a_{\gamma\delta}$ and sum over the N_f flavours. Since $q = \partial_\mu\mathcal{K}^\mu$, we can define a current which is conserved,

$$\tilde j_5^\mu = j_5^\mu + 2N_f\mathcal{K}^\mu, \qquad \partial_\mu\tilde j_5^\mu = 0, \tag{2.100}$$

with the conserved charge

$$\tilde Q_5 = \int d^3x\, \tilde j_5^0 = Q_5 + 2N_f\int d^3x\,\mathcal{K}^0. \tag{2.101}$$

Have we then found a hidden conserved axial charge? Not quite. Q_5 is gauge invariant, but $\tilde Q_5$ is not, because $\mathcal{K}^0$ is not. Under a large gauge transformation the last term, which is the winding number operator, changes by one unit. The winding number operator takes the value n on $|n\rangle$; on $\mathcal{T}_1|n\rangle = |n+1\rangle$ it takes the value $n+1$. Therefore $\mathcal{T}_1^{-1}\,\hat n\,\mathcal{T}_1 = \hat n + 1$, or $\mathcal{T}_1\,\hat n\,\mathcal{T}_1^{-1} = \hat n - 1$, and

$$\mathcal{T}_1\,\tilde Q_5\,\mathcal{T}_1^{-1} = \tilde Q_5 - 2N_f. \tag{2.102}$$

Let us apply a rotation generated by $\tilde Q_5$ on a θ vacuum. We find:

$$\mathcal{T}_1\, e^{i\alpha\tilde Q_5}|\theta\rangle = \mathcal{T}_1\, e^{i\alpha\tilde Q_5}\,\mathcal{T}_1^{-1}\,\mathcal{T}_1|\theta\rangle = e^{i\alpha(\tilde Q_5-2N_f)}\, e^{i\theta}|\theta\rangle = e^{i(\theta-2N_f\alpha)}\; e^{i\alpha\tilde Q_5}|\theta\rangle\,, \tag{2.103}$$

i.e.

$$e^{i\alpha\tilde Q_5}|\theta\rangle = \left|\theta - 2N_f\alpha\right\rangle. \tag{2.104}$$

In a world of massless quarks, $\tilde{Q}_5$ is conserved and commutes with the Hamiltonian. All θ vacua are then related by a symmetry, they have the same energy and the same physics, and θ can be rotated away. The angle θ is unphysical.

It is instructive to see the same result in the Lagrangian language. The operator $e^{i\alpha Q_5}$ acts on the quark fields as $e^{i\alpha Q_5}\psi e^{-i\alpha Q_5} = e^{-i\alpha\gamma_5}\psi$. Consider therefore a change of variables $\psi \to e^{i\beta\gamma_5}\psi$, $\bar{\psi} \to \bar{\psi}e^{i\beta\gamma_5}$ in the path integral, with constant β. The massless Lagrangian is invariant, but the fermion measure is not. Its Jacobian is precisely what produces the anomaly of Eq. 2.99:

$$\mathcal{D}\bar{\psi}\,\mathcal{D}\psi \to \mathcal{D}\bar{\psi}\,\mathcal{D}\psi\ \exp\left\{i\,2N_f\beta\int d^4x\, q(x)\right\} \quad \rightsquigarrow \quad \theta \to \theta + 2N_f\beta. \tag{2.105}$$

Exercise: *Prove this statement. Perform the change of variables with a space-time dependent $\beta(x)$, show that the classical Lagrangian changes by $-(\partial_\mu\beta)j_5^\mu$, and demand that the path integral does not depend on the integration variables. Show that this reproduces Eq. 2.99 only if the Jacobian is the one given above.*

If quarks have masses, the mass terms are not invariant. With $\psi_{R,L} = \frac{1\pm\gamma_5}{2}\psi$ we have $\psi_R \to e^{i\beta}\psi_R$, $\psi_L \to e^{-i\beta}\psi_L$, and

$$-m\left(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L\right) \to -m\left(e^{2i\beta}\bar{\psi}_L\psi_R + e^{-2i\beta}\bar{\psi}_R\psi_L\right). \tag{2.106}$$

The rotation removes θ from the gluonic term but puts a complex phase into the quark masses. The CP violation does not disappear; it has only moved from one place of the Lagrangian to another.

In summary, the existence of large gauge transformations and of gluon configurations which interpolate between the different classes forces a new term, Eq. 2.96, into the QCD Lagrangian. Chiral rotations can change the value of θ. If at least one quark is massless, we can rotate θ away. For massive quarks, a combination of θ and the phases of the quark masses remains and violates CP. We will identify this combination in the next section.

2.5 The axion

The strong CP problem

We have seen that a chiral rotation shifts θ and the phases of the quark masses at the same time. Which combination of them is physical? Let us write the most general mass term for N_f quark flavours,

$$\mathcal{L}_{\text{mass}} = -\bar{\psi}_L\, M\, \psi_R - \bar{\psi}_R\, M^\dagger \psi_L, \tag{2.107}$$

where M is a complex $N_f \times N_f$ matrix (in the Standard Model it is $v/\sqrt{2}$ times the Yukawa matrix). Any complex matrix can be diagonalised with two unitary matrices,

$$S_L^\dagger\, M\, S_R = M_D = \text{diag}(m_1, \ldots, m_{N_f}), \qquad m_f \text{ real and } \geq 0, \tag{2.108}$$

and we obtain real quark masses with the change of variables $\psi_L \to S_L\psi_L$, $\psi_R \to S_R\psi_R$. We write $S_{L,R} = e^{i\phi_{L,R}}\bar{S}_{L,R}$, with $\bar{S}_{L,R} \in SU(N_f)$. The $SU(N_f)$ parts are anomaly free, since their generators are traceless in flavour space. The phases amount to a vector rotation, which is harmless, and to an axial rotation $\psi \to e^{i\beta\gamma_5}\psi$ of all flavours with $\beta = (\phi_R - \phi_L)/2$. From the previous section, it shifts

$$\theta \to \theta + 2N_f\beta = \theta + N_f(\phi_R - \phi_L). \tag{2.109}$$

On the other hand, taking the phase of the determinant of $S_L^\dagger M S_R = M_D$ we find:

$$\begin{aligned} 0 = \arg\det M_D &= \arg\det S_L^\dagger + \arg\det M + \arg\det S_R \\ &= N_f\left(\phi_R - \phi_L\right) + \arg\det M. \end{aligned} \tag{2.110}$$

After the rotation, all quark masses are real and the coefficient of $q(x)$ is

$$\bar\theta = \theta - \arg\det M. \tag{2.111}$$

This is the physical, rephasing invariant, CP-violating parameter of the strong interaction. The bound of the previous section applies to $\bar\theta$: $|\bar\theta| \lesssim 10^{-10}$.

Why is $\bar\theta$ so small? This is known as the **strong CP problem**. The two terms in Eq. 2.111 have nothing to do with each other: θ is a property of the QCD vacuum, while $\arg\det M$ comes from the Yukawa couplings of the Higgs field to the quarks, which also generate the CP-violating phase of the CKM matrix. That phase is of order one. A cancellation between the two terms to ten digits would be a pure "accident". One possible way out would be a massless quark: then $\det M = 0$ and, as we have seen, θ can be rotated away. For a long time it was debated whether the up quark could be massless. Lattice QCD has now settled the question: $m_u/m_d \approx 0.47$, and $m_u \approx 2.2$ MeV is not zero.

The Peccei-Quinn mechanism

The most popular solution of the strong CP problem is due to Peccei and Quinn. The idea is to promote $\bar\theta$ to a dynamical field, which will relax to zero by itself. Suppose that the theory has an additional global $U(1)$ symmetry, $U(1)_{\rm PQ}$, under which quarks (and new scalars) carry chiral charges, so that it has a QCD anomaly. Suppose further that $U(1)_{\rm PQ}$ is spontaneously broken at some scale f_a. Then there is a Goldstone boson, the **axion** $a(x)$. As for every Goldstone boson, the symmetry acts on it as a shift, $a \to a + \alpha f_a$. Because of the anomaly, this shift changes the Lagrangian by a term proportional to $q(x)$. The low-energy effective Lagrangian is then of the form

$$\mathcal{L} = \mathcal{L}_{\rm QCD}\Big|_{\bar\theta} + \frac{1}{2}\left(\partial_\mu a\right)\left(\partial^\mu a\right) + \frac{a}{f_a}\frac{g^2}{32\pi^2}G^a_{\mu\nu}\tilde G^{a\mu\nu} + \frac{\partial_\mu a}{f_a}\sum_f \frac{\kappa_f}{2}\,\bar\psi_f\gamma^\mu\gamma_5\psi_f, \tag{2.112}$$

where the last term collects the derivative couplings allowed by the shift symmetry, with model-dependent constants κ_f. The gluonic terms combine into

$$\theta_{\rm eff}(x) = \bar\theta + \frac{a(x)}{f_a}. \tag{2.113}$$

The constant $\bar\theta$ can now be removed by a shift of the field, $a \to a - f_a\bar\theta$. Since the axion has only derivative couplings apart from the anomaly term, nothing else changes. Equivalently, by a chiral rotation of the quarks we can move $\theta_{\rm eff}$ from the gluonic term into the phase of the quark mass term, $-\bar\psi_L M e^{i\theta_{\rm eff}/N_f}\psi_R$ + h.c., and absorb it there by the same shift.

But which value will the field $a(x)$ take in the vacuum? It will take the value which minimises the vacuum energy. The vacuum energy of QCD depends on $\theta_{\rm eff}$, and the essential fact is that it has its minimum at $\theta_{\rm eff} = 0$. The vacuum then chooses $\langle a\rangle = -f_a\bar\theta$, and the effective CP violating angle vanishes dynamically. The θ term has been taken into the Lagrangian and has been made dynamical, and the dynamics sets it to zero.

Exercise: *(Vafa-Witten.) Consider the Euclidean path integral for the ground state energy of QCD in a volume V, $e^{-VE(\theta)} = \int \mathcal{D}A\,\mathcal{D}\bar\psi\,\mathcal{D}\psi\,\exp\left(-\int d^4x_E\,\mathcal{L}_E\right)$, with real quark masses.*

Show that the θ term is imaginary in the Euclidean action. Integrate out the quarks and show, using $\gamma_5\gamma_\mu = -\gamma_\mu\gamma_5$, that the non-zero eigenvalues of $i\not{D}$ come in pairs $\pm\lambda$, so that $\det(\not{D}+m) = \prod_{\lambda>0}(\lambda^2+m^2) > 0$. *Conclude that $|e^{-VE(\theta)}| \le e^{-VE(0)}$, i.e. $E(\theta) \ge E(0)$. Why does the argument apply to any P-odd operator, and in particular to the axion?*

The axion mass

The axion is not exactly massless: its shift symmetry is broken by the anomaly, and the potential $E(\theta_{\rm eff})$ gives it a mass. We can compute it with the chiral Lagrangian of the pions, which we have constructed in Part IV, Chapter 7 (and in Section 2.3.2). Let us take $N_f = 2$. First, we remove the gluonic term by an axial rotation of the up and down quarks,

$$u \to e^{-ic_u\theta_{\rm eff}\gamma_5/2}\,u, \qquad d \to e^{-ic_d\theta_{\rm eff}\gamma_5/2}\,d, \qquad c_u + c_d = 1. \tag{2.114}$$

The rotation of each flavour by an angle β_f shifts the coefficient of $q(x)$ by $+2\beta_f$, so the total shift is $-(c_u+c_d)\theta_{\rm eff}$ and the gluonic term is gone. The phases reappear in the mass terms,

$$\mathcal{L}_m = -m_u\,\bar u\,e^{-ic_u\theta_{\rm eff}\gamma_5}\,u - m_d\,\bar d\,e^{-ic_d\theta_{\rm eff}\gamma_5}\,d. \tag{2.115}$$

Since $\theta_{\rm eff}$ depends on x through $a(x)$, the kinetic terms of the quarks produce, in addition, derivative couplings

$$\frac{c_u}{2f_a}\left(\partial_\mu a\right)\bar u\gamma^\mu\gamma_5 u + \frac{c_d}{2f_a}\left(\partial_\mu a\right)\bar d\gamma^\mu\gamma_5 d, \tag{2.116}$$

which simply redefine the constants κ_f of Eq. 2.112. The freedom in choosing c_u, c_d is a freedom in defining the axion field; physical results do not depend on it.

In the effective theory the quark bilinears are replaced by the pion field. For the neutral pion, $U = \mathrm{diag}\left(e^{i\pi^0/f_\pi}, e^{-i\pi^0/f_\pi}\right)$ with $f_\pi \approx 92$ MeV, and the mass terms of Eq. 2.65, $\frac{f_\pi^2}{4}\,\mathrm{tr}\left(\chi U^\dagger + \chi^\dagger U\right)$ with $\chi = 2B\,\mathrm{diag}\left(m_u e^{-ic_u\theta_{\rm eff}}, m_d e^{-ic_d\theta_{\rm eff}}\right)$, yield the potential

$$V(\pi^0, a) = -c\left[m_u\cos\left(c_u\frac{a}{f_a} + \frac{\pi^0}{f_\pi}\right) + m_d\cos\left(c_d\frac{a}{f_a} - \frac{\pi^0}{f_\pi}\right)\right], \tag{2.117}$$

with $c = f_\pi^2 B = -\langle\bar u u\rangle = -\left\langle\bar d d\right\rangle > 0$, and where we have already shifted a such that $\bar\theta$ is absorbed. Expanding to second order in the fields, we find the mass matrix

$$\mathcal{M}^2 = c\begin{pmatrix} \dfrac{m_u+m_d}{f_\pi^2} & \dfrac{m_u c_u - m_d c_d}{f_\pi f_a} \\[2ex] \dfrac{m_u c_u - m_d c_d}{f_\pi f_a} & \dfrac{m_u c_u^2 + m_d c_d^2}{f_a^2} \end{pmatrix} \tag{2.118}$$

in the basis (π^0, a). The upper diagonal entry is the pion mass of the Gell-Mann-Oakes-Renner relation (Section 7.3), $m_\pi^2 f_\pi^2 = c\,(m_u+m_d)$. For $f_a \gg f_\pi$ the eigenvalues are m_π^2 and $m_a^2 \approx \det\mathcal{M}^2/m_\pi^2$. The determinant is

$$\begin{aligned}\det\mathcal{M}^2 &= \frac{c^2}{f_\pi^2 f_a^2}\left[(m_u+m_d)(m_u c_u^2 + m_d c_d^2) - (m_u c_u - m_d c_d)^2\right] \\ &= \frac{c^2\,m_u m_d}{f_\pi^2 f_a^2}\,(c_u+c_d)^2\,,\end{aligned}$$

and with $c_u + c_d = 1$ we obtain

$$m_a^2 = \frac{c}{f_a^2}\,\frac{m_u m_d}{m_u + m_d} = \frac{f_\pi^2}{f_a^2}\,\frac{m_u m_d}{(m_u + m_d)^2}\,m_\pi^2. \tag{2.119}$$

Notice that c_u, c_d have dropped out, as they should. Numerically, with $m_u = 2.16$ MeV, $m_d = 4.70$ MeV, $m_\pi = 135$ MeV, we find

$$m_a \approx 5.8\ \text{MeV} \times \frac{1\ \text{GeV}}{f_a} = 5.8\ \mu\text{eV} \times \frac{10^{12}\ \text{GeV}}{f_a}. \tag{2.120}$$

The axion is a very light particle when f_a is large.
Exercise: *Minimise Eq. 2.117 with respect to π^0 for a constant θ_{eff} (choose $c_u = 1$, $c_d = 0$) and show that the vacuum energy of QCD is*

$$E(\theta_{\text{eff}}) = -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4 m_u m_d}{(m_u + m_d)^2}\sin^2\frac{\theta_{\text{eff}}}{2}}.$$

Verify that it is minimal at $\theta_{\text{eff}} = 0$, as required by the Vafa-Witten argument, and rederive Eq. 2.119 from its curvature.
Exercise: *Show that for the choice $c_u = m_d/(m_u + m_d)$, $c_d = m_u/(m_u + m_d)$ there is no mass mixing between the axion and the neutral pion.*

Where is the axion?

The eigenvector of $\mathcal{M}^2$ with eigenvalue m_a^2 contains a small admixture of the neutral pion,

$$a_{\text{phys}} \approx a - \frac{(m_u c_u - m_d c_d)}{m_u + m_d}\,\frac{f_\pi}{f_a}\,\pi^0. \tag{2.121}$$

Through this admixture the axion inherits the couplings of the pion to hadrons and to photons, suppressed by f_π/f_a. Amplitudes for producing an axion are therefore smaller than the corresponding pion amplitudes by a factor of order f_π/f_a.

In the original proposal of Peccei, Quinn, Weinberg and Wilczek, the $U(1)_{\text{PQ}}$ symmetry was broken by the vacuum expectation values of the Higgs doublets, $f_a \sim 250$ GeV. Such an axion, with a mass of about 100 keV, would have been seen in decays of kaons and of heavy quarkonia; it is excluded. The way out is to break $U(1)_{\text{PQ}}$ at a much higher scale with a new scalar field which is a singlet of the Standard Model gauge group. The axion becomes very light and very weakly coupled, the so-called "invisible axion". It is, however, not invisible to stars. Axions with $m_a \lesssim$ keV can be produced in the hot interior of stars and escape freely, carrying away energy. The observed lifetime of red giants and of horizontal branch stars, and above all the duration of the neutrino burst of the supernova SN1987A, require

$$f_a \gtrsim 10^8 - 10^9\ \text{GeV}. \tag{2.122}$$

How long does such an axion live? It decays into two photons, like the π^0. The amplitude is suppressed by f_π/f_a with respect to that of the pion, and the phase space by $(m_a/m_\pi)^3 \sim (f_\pi/f_a)^3$. Therefore

$$\frac{\Gamma(a \to \gamma\gamma)}{\Gamma(\pi^0 \to \gamma\gamma)} \sim \left(\frac{f_\pi}{f_a}\right)^2 \left(\frac{m_a}{m_\pi}\right)^3 \sim \left(\frac{f_\pi}{f_a}\right)^5. \tag{2.123}$$

With $\tau_{\pi^0} = 8.4 \times 10^{-17}$ s, already for $f_a = 10^7$ GeV we find $\tau_a \sim 10^{24}$ s, much longer than the age of the universe (4×10^{17} s). Axions are practically stable. On the other hand, axions are produced in the early universe as a coherent oscillation of the field $a(x)$ around its minimum, and for $f_a \gtrsim 10^{12}$ GeV they would carry more energy density than is observed for the dark matter (for an initial misalignment angle of order one). We are left with a window

$$10^8 \text{ GeV} \lesssim f_a \lesssim 10^{12} \text{ GeV}, \qquad 6\ \mu\text{eV} \lesssim m_a \lesssim 60 \text{ meV}. \tag{2.124}$$

In the upper part of the window the axion is an excellent candidate for the dark matter of the universe. It is searched for in "haloscopes", microwave cavities in strong magnetic fields where dark matter axions of our galaxy would convert into photons. So far, it has not been found.

2.6 Instantons

In the previous two sections we have assumed that the gluon field can make transitions between the vacuum classes $|n\rangle$. Such a transition must pass over an energy barrier, since between two different classes the field cannot be a pure gauge. Classically it is forbidden; quantum mechanically it can happen by tunnelling. How large is its amplitude? Tunnelling is a non-perturbative phenomenon, and we need a new tool: semiclassical solutions of the field equations in Euclidean space-time, the **instantons**. We will first remind ourselves how tunnelling works in quantum mechanics, and then we will find the instanton of the Yang-Mills theory.

Tunnelling in quantum mechanics

Consider a particle of mass m in one dimension, in a potential $V(x)$ with two minima at x_0 and x_1, $V(x_0) = V(x_1) = 0$, separated by a barrier. For an energy E below the top of the barrier there is no classical trajectory from x_0 to x_1. The WKB approximation tells us that the wave function under the barrier is not oscillating but decaying, and that the amplitude to go through is

$$\langle x_1|\, x_0\rangle \propto \exp\left\{-\int_{\xi_0}^{\xi_1} dx\, \sqrt{2m\,(V(x)-E)}\right\}, \tag{2.125}$$

where $\xi_{0,1}$ are the classical turning points ($\hbar = 1$). Compare this with the oscillating factor $e^{iS_{\rm cl}}$ of an allowed transition. We can write Eq. 2.125 in the same form if we replace real time by imaginary time, $t = -i\tau$. Then the action becomes

$$iS = i\int dt \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 - V(x)\right] \;\to\; -\int d\tau \left[\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2 + V(x)\right] \equiv -S_E, \tag{2.126}$$

and the equation of motion becomes $m\, d^2x/d\tau^2 = +V'(x)$: the particle moves in the upside-down potential $-V(x)$. The conserved energy along such a path is

$$\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2 - V(x) = -E. \tag{2.127}$$

Let us take $E = 0$. Along the Euclidean classical path $m(dx/d\tau)^2 = 2V(x)$, and the Euclidean action is

$$S_E = \int d\tau\, m\left(\frac{dx}{d\tau}\right)^2 = \int_{x_0}^{x_1} dx\, \sqrt{2mV(x)},$$

which is exactly the exponent of Eq. 2.125. The tunnelling amplitude is e^{-S_E} evaluated on a solution of the Euclidean equations of motion, which goes from one minimum at $\tau = -\infty$ to the other at $\tau = +\infty$. Such a solution is called an instanton: it is localised in (imaginary) time, around the moment where the tunnelling happens.

We know from QFT II (Eq. 1.22) that the path integral gives $\langle x_1 | e^{-iHT} | x_0 \rangle$. Continuing to imaginary time,

$$\langle x_1 | e^{-HT} | x_0 \rangle = \int \mathcal{D}x \; e^{-S_E[x]}, \tag{2.128}$$

and for small $\hbar$ the integral is dominated by the stationary points of S_E, i.e. by the Euclidean classical solutions; the Gaussian integral over the fluctuations around them provides a prefactor. In field theory exactly the same holds with $x(\tau)$ replaced by $\phi(\vec{x}, x_4)$, $x_4 = ix^0$; the Minkowski metric becomes (minus) the Euclidean one.

There is another important lesson from quantum mechanics. If tunnelling is possible, the ground state is not localised in one of the minima. For a symmetric double well, the ground state is the symmetric combination of the states localised in the two wells, and the energy splitting to the antisymmetric state is proportional to e^{-S_E}. Even more relevant for us is a **periodic** potential, with minima at $x = na$, $n \in \mathbb{Z}$. Let $|n\rangle$ be the state localised in the n-th minimum. Let us compute $\langle n_+ | e^{-HT} | n_- \rangle$ in the approximation where the Euclidean paths are made of widely separated instantons (jumps $n \to n+1$) and anti-instantons (jumps $n \to n-1$), each with action S_0. Every instanton can sit anywhere in the interval T, which gives a factor T, and the fluctuations around it a factor K. For N instantons and $\bar{N}$ anti-instantons we have

$$\langle n_+ | e^{-HT} | n_- \rangle = e^{-E_0 T} \sum_{N,\bar{N}=0}^{\infty} \frac{\left(KTe^{-S_0}\right)^{N+\bar{N}}}{N!\,\bar{N}!} \, \delta_{N-\bar{N},\, n_+ - n_-}, \tag{2.129}$$

where E_0 is the energy of a single well. Writing $\delta_{N-\bar{N},n_+-n_-} = \int_0^{2\pi} \frac{d\theta}{2\pi} \, e^{i\theta(N-\bar{N}-n_++n_-)}$, the sums exponentiate:

$$\langle n_+ | e^{-HT} | n_- \rangle = \int_0^{2\pi} \frac{d\theta}{2\pi} \, e^{-i\theta(n_+-n_-)} \exp\left\{-T\left(E_0 - 2Ke^{-S_0}\cos\theta\right)\right\}. \tag{2.130}$$

Comparing with $\langle n_+ | e^{-HT} | n_- \rangle = \sum_E e^{-ET} \langle n_+ | E \rangle \langle E | n_- \rangle$, we read off that the energy eigenstates are $|\theta\rangle \propto \sum_n e^{-in\theta} |n\rangle$, with energies

$$E(\theta) = E_0 - 2Ke^{-S_0}\cos\theta. \tag{2.131}$$

These are the Bloch waves of solid-state physics, and they have exactly the form of the θ vacua of Eq. 2.92. This is the picture we have to keep in mind for QCD, with one difference. In a crystal the different θ correspond to different crystal momenta of the same system. In QCD the translation $n \to n+1$ is a gauge transformation, all physical states are eigenstates of it with the same θ, and θ is a constant of nature. (See Coleman, "Aspects of Symmetry", chapter 7, for a beautiful exposition.)

Euclidean Yang-Mills theory and self-dual fields

We now look for the tunnelling configurations of the pure Yang-Mills theory; quarks do not matter for the moment. Why must we go to Euclidean space? In Minkowski space, the energy density of the gluon field is

$$\Theta^{00} = \frac{1}{2} \sum_a \left(\vec{E}^a \cdot \vec{E}^a + \vec{B}^a \cdot \vec{B}^a \right) \geq 0, \tag{2.132}$$

and it vanishes only for $G_{\mu\nu} = 0$, i.e. for pure gauges. A configuration which interpolates between two classes has positive energy along the way, and there is no zero-energy real solution which does it. This is the barrier. The energy density could vanish for $\vec{E}^a = \pm i\vec{B}^a$, i.e. for complex Minkowski fields. A complex electric field in Minkowski space becomes a real field in Euclidean space, since $E^{ai} \propto G^a_{0i}$ and $\partial_0 = i\partial_4$. This is the signal that the tunnelling solutions live in Euclidean space.

In Euclidean space ($\mu, \nu = 1, \ldots, 4$, all indices down) the action is

$$S_E = \frac{1}{4}\int d^4x_E \; G^a_{\mu\nu}G^a_{\mu\nu}, \tag{2.133}$$

and we define the dual field $\tilde{G}^a_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}G^a_{\rho\sigma}$, with $\epsilon_{1234} = 1$. The topological charge is (with a suitable orientation of Euclidean space) the same integer as in Eq. 2.90,

$$n = \frac{g^2}{32\pi^2}\int d^4x_E \; G^a_{\mu\nu}\tilde{G}^a_{\mu\nu}. \tag{2.134}$$

Notice a difference between the two signatures. In Minkowski space the dual of the dual is $\tilde{\tilde{G}}_{\mu\nu} = -G_{\mu\nu}$, so a real field cannot be self-dual: $\tilde{G} = \pm G$ implies $G = -G = 0$. In Euclidean space $\tilde{\tilde{G}}_{\mu\nu} = +G_{\mu\nu}$, and nontrivial real fields with

$$\tilde{G}^a_{\mu\nu} = \pm G^a_{\mu\nu} \tag{2.135}$$

can exist. We call them **self-dual** (+) and **anti-self-dual** (−). Since $\tilde{G}\tilde{G} = GG$ in Euclidean space, we can write the action as

$$\begin{aligned} S_E &= \frac{1}{8}\int d^4x_E \left(G^a_{\mu\nu}G^a_{\mu\nu} + \tilde{G}^a_{\mu\nu}\tilde{G}^a_{\mu\nu}\right) \\ &= \frac{1}{8}\int d^4x_E \left(G^a_{\mu\nu} \mp \tilde{G}^a_{\mu\nu}\right)^2 \pm \frac{1}{4}\int d^4x_E \; G^a_{\mu\nu}\tilde{G}^a_{\mu\nu} \\ &\geq \pm\frac{8\pi^2}{g^2}\, n. \end{aligned} \tag{2.136}$$

Choosing the sign which makes the right-hand side positive, we obtain the bound

$$S_E \geq \frac{8\pi^2}{g^2}\,|n|, \tag{2.137}$$

with equality if and only if the field is self-dual ($n > 0$) or anti-self-dual ($n < 0$). In each topological sector the action is minimised by the (anti-)self-dual configurations.

There is one more beautiful property. The Yang-Mills equations of motion are

$$D_\mu G_{\mu\nu} = \partial_\mu G_{\mu\nu} - ig\left[A_\mu, G_{\mu\nu}\right] = 0. \tag{2.138}$$

On the other hand, every field strength which is derived from a potential satisfies the Bianchi identity

$$D_\mu \tilde{G}_{\mu\nu} = 0, \tag{2.139}$$

whether or not A_μ solves the equations of motion.

Exercise: *Prove the Bianchi identity from the Jacobi identity* $\left[D_\mu, \left[D_\nu, D_\rho\right]\right]$ + cyclic = 0.

For a self-dual field, $\tilde{G} = \pm G$, the Bianchi identity **is** the equation of motion. We only need to solve the first-order equations $\tilde{G} = \pm G$. Finally, the Euclidean energy-momentum tensor, $\Theta_{\mu\nu} = G^a_{\mu\lambda}G^a_{\nu\lambda} - \frac{1}{4}\delta_{\mu\nu}G^a_{\rho\sigma}G^a_{\rho\sigma} = \frac{1}{2}\left(G^a_{\mu\lambda}G^a_{\nu\lambda} - \tilde{G}^a_{\mu\lambda}\tilde{G}^a_{\nu\lambda}\right)$, vanishes for self-dual fields. In this sense self-dual fields are "vacuum-like" configurations, and the leading tunnelling amplitude between the classes n and $n+1$ is given by them.

The BPST instanton

A self-dual solution with $n = 1$ was found by Belavin, Polyakov, Schwartz and Tyupkin in 1975. For $SU(2)$ (embedded in $SU(3)$ as in Section 2.4) it reads

$$A^a_\mu(x) = \frac{2}{g}\frac{\eta^a_{\mu\nu}\, x_\nu}{x^2+\lambda^2}, \qquad x^2 = x_\mu x_\mu, \tag{2.140}$$

where λ is an arbitrary length, the size of the instanton, and $\eta^a_{\mu\nu}$ are the 't Hooft symbols, which mix colour and space-time indices,

$$\eta^a_{\mu\nu} = \epsilon_{a\mu\nu 4} + \delta_{a\mu}\delta_{\nu 4} - \delta_{a\nu}\delta_{\mu 4}. \tag{2.141}$$

They are antisymmetric in $\mu\nu$, self-dual, $\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\eta^a_{\rho\sigma} = \eta^a_{\mu\nu}$, and satisfy $\eta^a_{\mu\nu}\eta^a_{\mu\nu} = 12$.
Exercise: *With $G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\epsilon^{abc}A^b_\mu A^c_\nu$ (which is our convention for $SU(2)$, $T^a = \tau^a/2$), show that the field strength of Eq. 2.140 is*

$$G^a_{\mu\nu} = -\frac{4}{g}\frac{\lambda^2\,\eta^a_{\mu\nu}}{(x^2+\lambda^2)^2}.$$

Conclude that it is self-dual, that $G^a_{\mu\nu}G^a_{\mu\nu} = \frac{192\,\lambda^4}{g^2(x^2+\lambda^2)^4}$, and that

$$S_E = \frac{1}{4}\int d^4x_E\, G^a_{\mu\nu}G^a_{\mu\nu} = \frac{8\pi^2}{g^2}, \qquad n = 1.$$

You will need $\epsilon^{abc}\eta^b_{\mu\nu}\eta^c_{\rho\sigma} = \delta_{\mu\rho}\eta^a_{\nu\sigma} - \delta_{\mu\sigma}\eta^a_{\nu\rho} - \delta_{\nu\rho}\eta^a_{\mu\sigma} + \delta_{\nu\sigma}\eta^a_{\mu\rho}$ and the volume of the unit three-sphere, $2\pi^2$.

The field strength falls off as $1/x^4$, so the action is finite. The potential itself falls off only as $1/|x|$: at large $|x|$ it becomes a pure gauge,

$$A_\mu \to \frac{i}{g}\,U(\hat{x})\,\partial_\mu U^\dagger(\hat{x}), \qquad U(\hat{x}) = \frac{x_4 + i\,\vec{x}\cdot\vec{\tau}}{|x|} \quad \text{(up to orientation conventions)},$$

with a gauge function on the sphere S^3 at Euclidean infinity, which covers $SU(2)$ exactly once. This is where the integer n comes from. After a gauge transformation to $A_4 = 0$, the instanton is a pure gauge of winding number n_- at $x_4 \to -\infty$ and of winding number $n_- + 1$ at $x_4 \to +\infty$: it is the tunnelling event $|n\rangle \to |n+1\rangle$ of Section 2.4. Notice also that the action does not depend on λ. This is a consequence of the scale invariance of the classical Yang-Mills theory; in the quantum theory, the coupling in $e^{-8\pi^2/g^2}$ must be taken at the scale $1/\lambda$, and the integral over large instanton sizes is not under control, since $g(1/\lambda)$ grows.

Physical consequences

By Eq. 2.137, the amplitude for tunnelling from $|n\rangle$ to $|n+1\rangle$ is proportional to

$$e^{-S_E} = \exp\left(-\frac{8\pi^2}{g^2}\right) = \exp\left(-\frac{2\pi}{\alpha_s}\right). \tag{2.142}$$

This function has a vanishing Taylor expansion around $g = 0$; we will **never** see it in perturbation theory, at any order. It is a genuinely non-perturbative effect. As in the periodic potential, a dilute gas of instantons and anti-instantons makes the vacuum energy

depend on θ, $E(\theta) - E(0) \propto e^{-8\pi^2/g^2}(1 - \cos\theta)$, minimal at $\theta = 0$ in accordance with the Vafa-Witten argument of Section 2.5.

What happens if there are massless quarks? The Dirac operator in the background of an instanton has normalisable zero modes, one for each massless flavour, and the fermion determinant vanishes: the vacuum-to-vacuum tunnelling amplitude from $|n\rangle$ to $|n+1\rangle$ is zero, and the θ dependence disappears, as we found in Section 2.4. The instanton contributes, however, to amplitudes with $2N_f$ external quarks, one left-handed and one right-handed of each flavour. 't Hooft showed that this is described by an effective interaction of the form $\det_{fg}\left(\bar{\psi}_{Rf}\psi_{Lg}\right)$, which changes the axial charge by $|\Delta Q_5| = 2N_f$, exactly as the anomaly equation 2.99 requires for $|\Delta n| = 1$. This interaction breaks $U(1)_A$ explicitly and explains why there is no light ninth pseudo-Goldstone boson in QCD: the η' meson is much heavier than the pions and kaons.

The same physics exists in the electroweak theory (Section 3.6). There, instantons of the $SU(2)_W$ gauge field change the baryon and lepton numbers, which have an electroweak anomaly. The rate is proportional to $e^{-16\pi^2/g_W^2} = e^{-4\pi/\alpha_W} \approx 10^{-162}$. It is an amazing fact that the Standard Model violates baryon number, and it is reassuring that the protons of our world do not care.

3

Topological aspects of field theory

So far, a field theory has been for us a theory of particles. We expand the fields around a constant vacuum value, we identify the quadratic part of the Lagrangian with free particles and we treat everything else as a perturbation. However, the classical field equations of an interacting theory are non-linear, and non-linear equations may have solutions which look nothing like small waves around the vacuum. Of particular interest are solutions which are static, have a finite energy which is localised in a region of space, and are nowhere singular. Such solutions are known as **solitons**. A soliton behaves like an extended particle; its energy is its mass.

Are solitons relevant for particle physics? Gauge theories with spontaneous symmetry breaking are non-linear, and we will see that they possess solitons with remarkable properties: **vortex lines** which carry quantised magnetic flux, and **magnetic monopoles**. They are the static cousins of the **instantons** of Section 2.6, which are localised in space *and* in (Euclidean) time.

What makes these solutions stable? We will find that the answer is not dynamical but topological. A field configuration with finite energy must approach a vacuum at spatial infinity. When the vacuum is degenerate, the boundary values of the field define a map from the "boundary" of space into the manifold of vacua. Such maps fall into classes which cannot be continuously deformed into each other, and the vacuum itself belongs only to one of these classes. A configuration in a different class can never decay into the vacuum by a continuous time evolution. In this chapter we will first study this mechanism in the simplest possible example, a scalar field in one space dimension, and then proceed to vortices in the abelian Higgs model and to magnetic monopoles, first in Dirac's form and then as smooth solitons of a non-abelian gauge theory with spontaneous symmetry breaking. In the second half of the chapter we ask where such objects come from in the early universe, and we study the phase transitions of gauge theories at finite temperature, the violation of baryon and lepton number by the electroweak anomaly, the sphalerons, and baryogenesis. The required notions of topology will be introduced as we go along. A good textbook account is in Ryder, *Quantum Field Theory*, chapter 10; see also Coleman, *Aspects of Symmetry*, chapter 6, and Srednicki, chapter 92.

3.1 Topological objects in field theory

3.1.1 The sine-Gordon kink

Let us consider a real scalar field $\phi(t,x)$ in one space and one time dimension, with Lagrangian density

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - V(\phi) = \frac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2 - \frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 - V(\phi), \qquad V(\phi) = \frac{m^2}{b^2}\left[1-\cos(b\phi)\right]. \quad (3.1)$$

This is the **sine-Gordon** model. Expanding the potential for small fields,

$$V(\phi) = \frac{m^2}{2}\phi^2 - \frac{\lambda}{4!}\phi^4 + O(\phi^6), \qquad \lambda = m^2 b^2, \tag{3.2}$$

we recognise a particle of mass m with a quartic self-interaction of strength λ; the parameter b is therefore a coupling constant. The equation of motion reads

$$\frac{\partial^2\phi}{\partial t^2} - \frac{\partial^2\phi}{\partial x^2} + \frac{m^2}{b}\sin(b\phi) = 0, \tag{3.3}$$

and the energy of a field configuration is

$$E = \int_{-\infty}^{\infty} dx \left[\frac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2 + \frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 + V(\phi)\right]. \tag{3.4}$$

All three terms are positive. The energy vanishes only for a constant field sitting at a zero of the potential,

$$\phi = \frac{2\pi n}{b}, \qquad n \in \mathbb{Z}. \tag{3.5}$$

The classical vacuum is infinitely degenerate, see Fig. 3.1. (By "vacuum" we mean here a classical field configuration of zero energy, not yet a state in a Hilbert space.) In perturbation theory we choose one of them, say $n = 0$, and we never see the others.

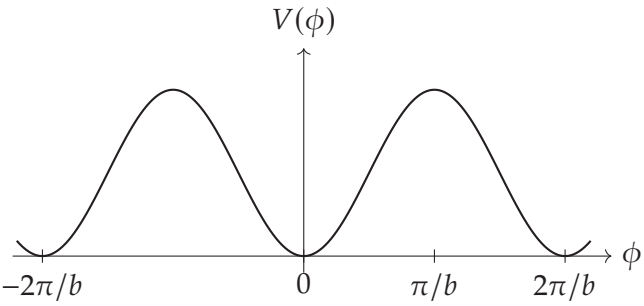


Figure 3.1. The sine-Gordon potential. Every zero $\phi = 2\pi n/b$ is a classical vacuum.

Let us now look for a static configuration which approaches one vacuum, $n = 0$, for $x \to -\infty$ and a different one, $n = 1$, for $x \to +\infty$. Somewhere in between the field must leave the zeros of the potential and it must have a non-vanishing gradient, so its energy is certainly not zero. Is it finite? For a static field we can complete the square in the energy functional:

$$\begin{aligned} E &= \int dx \left[\frac{1}{2}\left(\frac{d\phi}{dx}\right)^2 + V(\phi)\right] \\ &= \int dx\, \frac{1}{2}\left[\frac{d\phi}{dx} - \sqrt{2V(\phi)}\right]^2 + \int dx\, \frac{d\phi}{dx}\sqrt{2V(\phi)} \\ &\geq \int_{\phi(-\infty)}^{\phi(+\infty)} d\phi\, \sqrt{2V(\phi)}. \end{aligned} \tag{3.6}$$

Notice that the lower bound depends **only** on the boundary values of the field and not on the details of the configuration in between. It is saturated when the first-order equation

$$\frac{d\phi}{dx} = \sqrt{2V(\phi)} = \frac{2m}{b}\sin\left(\frac{b\phi}{2}\right) \tag{3.7}$$

holds. Differentiating it once more we obtain $\phi'' = V'(\phi)$, which is precisely the static equation of motion. Equivalently, multiplying the static equation of motion by ϕ' and integrating once we find $\frac{1}{2}\phi'^2 = V(\phi)$, where the integration constant is zero because both sides vanish at infinity. The solution of Eq. 3.7 is

$$\phi(x) = \frac{4}{b} \arctan\left(e^{m(x-x_0)}\right), \tag{3.8}$$

with an arbitrary centre x_0. It interpolates between $\phi(-\infty) = 0$ and $\phi(+\infty) = 2\pi/b$; the transition happens in a region of size $1/m$ around x_0. Its energy is

$$E = \int_0^{2\pi/b} d\phi\, \sqrt{2V(\phi)} = \frac{\sqrt{2}\,m}{b^2} \int_0^{2\pi} d\alpha\, \sqrt{1-\cos\alpha} = \frac{8m}{b^2} = \frac{8m^3}{\lambda}. \tag{3.9}$$

We have found a localised lump of energy, the **kink** (Fig. 3.2). It is finite, and it is **inversely** proportional to the coupling constant. Such a dependence can never arise at any finite order of perturbation theory, which produces only positive powers of λ. For a weak coupling the kink is very heavy compared to the elementary particle of mass m.

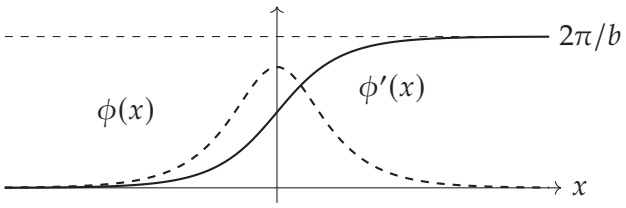


Figure 3.2. The kink of Eq. 3.8 (solid) and its energy-carrying gradient (dashed).

The kink is a static solution, but since the theory is Lorentz invariant we obtain a moving kink by a boost:

$$\phi(t,x) = \frac{4}{b} \arctan\left(e^{m\gamma(x-vt-x_0)}\right), \qquad \gamma = \frac{1}{\sqrt{1-v^2}}. \tag{3.10}$$

It moves with velocity v without changing its shape and without spreading out, unlike a wave packet of the linear Klein-Gordon equation.

Exercise: *Verify by direct substitution that the moving kink solves Eq. 3.3, and show that its energy is $\gamma\, 8m/b^2$, as expected for a particle of mass $8m/b^2$.*

Since the field equation is non-linear, the superposition principle does not hold. When two kinks collide the field looks complicated, but in the sine-Gordon model they emerge from the collision with their shape intact. Also the quantisation of a soliton is not the quantisation of small oscillations around it. We will not pursue these issues here.

Why is the kink stable? Its energy is larger than the energy of the vacuum, $E = 0$. However, to decay into the vacuum $\phi = 0$ the field at $x \to +\infty$ would have to move from $2\pi/b$ to 0, passing over the hill of the potential in a semi-infinite region of space. This costs an infinite amount of energy. A mechanical model makes this transparent: think of an infinitely long horizontal rubber band, to which pendula are attached at equal intervals, each pendulum coupled to its neighbours by the twisting of the band and pulled down by gravity. The angle of the pendula is $b\phi$. In the vacuum all pendula hang down. The kink is a configuration in which, going from left to right, the pendula make one full turn. To undo the turn we would have to flip an infinite number of pendula.

Mathematically, the stability is a consequence of the boundary conditions. "Space" is a line whose boundary consists of two points, $x = \pm\infty$. A finite-energy configuration maps

these two points onto vacua, and the integer

$$N = n(+\infty) - n(-\infty) \tag{3.11}$$

cannot change under a continuous time evolution with finite energy. The kink is a **topological** soliton. A configuration with $N = 2$ is not a static solution; it consists of two kinks, and configurations with $N < 0$ are anti-kinks.

The stability of the kink must be expressed as a conservation law. Indeed, let us define the current

$$j^\mu = \frac{b}{2\pi}\epsilon^{\mu\nu}\partial_\nu\phi, \qquad \epsilon^{01} = 1. \tag{3.12}$$

It is conserved, $\partial_\mu j^\mu = \frac{b}{2\pi}\epsilon^{\mu\nu}\partial_\mu\partial_\nu\phi = 0$, because $\epsilon^{\mu\nu}$ is antisymmetric. The charge is

$$Q = \int_{-\infty}^{\infty} dx\, j^0 = \frac{b}{2\pi}\int_{-\infty}^{\infty} dx\, \frac{\partial\phi}{\partial x} = \frac{b}{2\pi}\left[\phi(+\infty) - \phi(-\infty)\right] = N. \tag{3.13}$$

Notice that this current is very different from the Noether currents that we have met so far. It does not originate from a symmetry of the Lagrangian, and its conservation holds identically, without using the equations of motion. Such currents are called **topological currents**.

Is the periodic potential essential? No! What we need is a degenerate vacuum.
Exercise: *Consider* $\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)$ *with* $V(\phi) = \frac{1}{2}m^2\phi^2 + \frac{\lambda}{4!}\phi^4$ *and* $m^2 < 0$, *in one space dimension.*

- *Find the constant solutions of the equations of motion and add a constant to* V *such that they have zero energy.*
- *Find the static solutions which interpolate between the two vacua, using the ansatz* $\phi(x) = a\tanh(\beta x)$.
- *Show that the energy of the soliton is* $E = 4\sqrt{2}\,|m|^3/\lambda$.
- *Show that* $J^\mu = \epsilon^{\mu\nu}\partial_\nu\phi$ *is conserved. Which values can* $\int dx J^0$ *take for a finite-energy configuration?*

3.1.2 A scalar field in two space dimensions

Can we generalise the kink to two space dimensions? The "boundary" of the plane is now a circle at infinity, S^1. Consider a complex scalar field with a potential which has a whole circle of minima,

$$\mathcal{L} = \partial_\mu\phi^*\partial^\mu\phi - V(\phi), \qquad V(\phi) = \frac{\lambda}{4}\left(|\phi|^2 - v^2\right)^2. \tag{3.14}$$

This is the model with a spontaneously broken global $U(1)$ symmetry that we studied in QFT II (see Chapter 1 for a recap). The analogue of the kink boundary condition is that the field winds around the circle of vacua as we go around the circle at infinity:

$$\phi(r,\theta) \to v\, e^{in\theta}, \qquad r \to \infty, \tag{3.15}$$

where (r,θ) are polar coordinates in the plane. The field must be single valued, therefore n is an integer. The potential vanishes at infinity, but the gradient does not:

$$\vec{\nabla}\phi \to \frac{inv}{r}e^{in\theta}\,\hat{\theta}, \qquad |\vec{\nabla}\phi|^2 \to \frac{n^2v^2}{r^2}. \tag{3.16}$$

The static energy outside a large disc of radius R_0 is then

$$E \geq \int_0^{2\pi} d\theta \int_{R_0}^{R} r\, dr\, |\vec{\nabla}\phi|^2 = 2\pi n^2 v^2 \ln \frac{R}{R_0}, \tag{3.17}$$

which diverges logarithmically when the size R of the system goes to infinity. Our naive attempt has failed: there is no finite-energy vortex of a scalar field with a global symmetry. (In a finite system, such as a superfluid, the logarithm is cut off by the size of the container and such vortices exist.)

This failure is not an accident of our example.

Exercise (Derrick's theorem): *Consider real scalar fields with* $\mathcal{L} = \frac{1}{2}(\partial_\mu \phi_i)(\partial^\mu \phi_i) - V(\phi_i)$, $V \geq 0$, *in D space dimensions, and a static solution* $\phi_i(\vec{x})$ *with energy* $E = T + U$, $T = \frac{1}{2}\int d^D x (\vec{\nabla}\phi_i)^2$, $U = \int d^D x V(\phi_i)$.

- *Show that the rescaled configuration* $\phi_i(\vec{x}/\alpha)$ *has energy* $E(\alpha) = \alpha^{D-2} T + \alpha^D U$.
- *Argue why a solution must satisfy* $dE/d\alpha|_{\alpha=1} = 0$.
- *Show that this is impossible for* $D \geq 2$ *unless the solution is a vacuum. What happens for* $D = 1$*?*

Scalar fields alone give no stable solitons localised in more than one direction. Gauge fields change the scaling argument, since the energy density $\frac{1}{2}(\vec{E}^2 + \vec{B}^2)$ scales differently from $(\vec{\nabla}\phi)^2$. More importantly, a gauge field can compensate the angular gradient of the scalar at infinity. This is our next step.

3.2 Vortices in the abelian Higgs model

Let us gauge the $U(1)$ symmetry of the previous example. The Lagrangian of scalar electrodynamics with a Higgs potential is

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \left(D_\mu \phi\right)^* D^\mu \phi - m^2 |\phi|^2 - \lambda |\phi|^4, \qquad D_\mu = \partial_\mu - ieA_\mu. \tag{3.18}$$

For $m^2 < 0$ the minimum of the potential is at

$$|\phi| = v \equiv \sqrt{\frac{-m^2}{2\lambda}}, \tag{3.19}$$

and we know from QFT II (chapter on spontaneous symmetry breaking, section on local gauge symmetries) that the photon acquires a mass $m_A = \sqrt{2} e v$ and the would-be Goldstone boson disappears from the spectrum (Higgs mechanism). The field equations are

$$\begin{aligned} D_\mu D^\mu \phi &= -m^2 \phi - 2\lambda |\phi|^2 \phi, \\ \partial_\nu F^{\nu\mu} &= ie\left[\phi \left(D^\mu \phi\right)^* - \phi^* D^\mu \phi\right] = ie\left(\phi\, \partial^\mu \phi^* - \phi^* \partial^\mu \phi\right) - 2e^2 A^\mu |\phi|^2. \end{aligned} \tag{3.20}$$

Exercise: *Derive the field equations 3.20 and verify that the current on the right-hand side is conserved when the scalar field equation holds.*

We are interested in static configurations in the (x, y) plane with $A_0 = 0$. Their energy is

$$E = \int d^2 x \left[\frac{1}{2} B^2 + \left|D_i \phi\right|^2 + \lambda \left(|\phi|^2 - v^2\right)^2\right], \qquad B = F_{12}, \tag{3.21}$$

where we have shifted the potential by a constant so that it vanishes in the vacuum. We impose again the winding boundary condition of Eq. 3.15, $\phi \to v e^{in\theta}$. Now, however, the angular derivative of the scalar field can be cancelled by the gauge field. Indeed, the covariant derivative vanishes at infinity,

$$D_i\phi = (in\,\partial_i\theta - ieA_i)\,v e^{in\theta} = 0, \tag{3.22}$$

if the gauge field approaches a pure gauge:

$$A_i \to \frac{n}{e}\partial_i\theta \quad \rightsquigarrow \quad \vec{A} \to -\frac{n}{e}\vec{\nabla}\theta = -\frac{n}{er}\,\hat{\theta}, \qquad r \to \infty. \tag{3.23}$$

A pure gauge has a vanishing field strength, so all three terms of the energy density vanish at infinity. The logarithmic divergence of the global case has disappeared: the integrand can now fall off exponentially, with the masses of the photon and of the Higgs boson setting the scale. This does not yet prove that a solution exists, but the obstruction of the previous section has been removed.

The gauge field cannot be a pure gauge everywhere. Let us compute the magnetic flux through the plane with Stokes' theorem along the circle at infinity:

$$\Phi = \int d^2x\, B = \oint_{r\to\infty} \vec{A}\cdot d\vec{l} = \int_0^{2\pi} d\theta\, r A_\theta = -\frac{2\pi n}{e}. \tag{3.24}$$

The flux is quantised in units of $2\pi/e$! Notice that we did not solve any field equation to obtain this result. The flux is fixed by the winding number of the scalar field at infinity alone. If we add a third dimension along which the fields do not depend, the configuration becomes a tube of magnetic flux, a **vortex line** or "string". Its energy per unit length is finite, and of the order of v^2 times a function of λ/e^2.

To find the solution explicitly we use the rotationally symmetric ansatz

$$\phi = v f(r) e^{in\theta}, \qquad A_i = \frac{n}{e} a(r)\partial_i\theta, \tag{3.25}$$

with $f(\infty) = a(\infty) = 1$ to reproduce the boundary conditions and $f(0) = a(0) = 0$ for the fields to be regular at the origin. Inserting it into the field equations we obtain two coupled ordinary differential equations for f and a, which must be solved numerically. This is the **Nielsen-Olesen vortex**.

Exercise: *Show that for this ansatz $B = \frac{n}{er}a'(r)$ (up to the orientation sign) and that the energy per unit length is*

$$E = 2\pi v^2 \int_0^\infty r\,dr \left[\frac{n^2 a'^2}{2e^2v^2r^2} + f'^2 + \frac{n^2}{r^2}(1-a)^2 f^2 + \lambda v^2 (f^2-1)^2\right].$$

Vortex lines are not a theorist's fantasy. The abelian Higgs model is the relativistic version of the Landau-Ginzburg theory of superconductivity, where ϕ is the wave function of the Cooper pairs, with charge $2e$. A magnetic field penetrates a type II superconductor in the form of Abrikosov flux tubes, each carrying one flux quantum $2\pi/(2e) = \pi/e$, or, restoring units, $h/(2e)$. This has been observed in experiments.

Why are the vortices stable? The reason is the same as for the kink. The vacua of the theory, $\phi = v e^{i\alpha}$, form a circle S^1, the same as the group manifold of $U(1)$. The boundary of the plane is also a circle S^1. A finite-energy configuration therefore defines a map

$$\phi_\infty : S^1 \text{ (space)} \to S^1 \text{ (vacua)}, \qquad \theta \mapsto v e^{in\theta}, \tag{3.26}$$

and n counts how many times the field winds around the circle of vacua as we go once around the circle at infinity. A map with winding n cannot be continuously deformed into one with winding $n' \neq n$: think of a rubber band wound twice around a cylinder, which cannot be wound once around it without cutting it. Maps of a circle into a manifold, modulo continuous deformations, form the **first homotopy group**, and we have found

$$\pi_1(S^1) = \mathbb{Z}. \tag{3.27}$$

The integer n is a conserved topological charge, and because the flux is $-2\pi n/e$, magnetic flux is conserved for the same topological reason.

Let us summarise the logic, which will be the same in all examples of this chapter. Topology tells us in which sectors a soliton **may** exist, by classifying the boundary conditions. Whether a finite-energy solution exists in a given sector must then be decided with the field equations, as we did with the Nielsen-Olesen ansatz.

3.3 The Dirac monopole

Maxwell's equations are almost symmetric under the exchange of electric and magnetic fields; only magnetic charges are missing. Dirac asked whether a magnetic monopole is compatible with quantum mechanics. His answer was yes, provided that electric charge is quantised. This is a remarkable statement, since charge quantisation is observed in nature but has no explanation in QED.

Consider a monopole of magnetic charge g at the origin,

$$\vec{B} = \frac{g}{4\pi}\frac{\hat{r}}{r^2} = -\frac{g}{4\pi}\vec{\nabla}\frac{1}{r}, \qquad \vec{\nabla}\cdot\vec{B} = g\,\delta^{(3)}(\vec{x}), \tag{3.28}$$

where we have used $\nabla^2\frac{1}{r} = -4\pi\delta^{(3)}(\vec{x})$. The total flux through any sphere around the origin is $\Phi = g$. We would like to describe the coupling to a charged particle, of charge e, with the vector potential, $\vec{B} = \vec{\nabla}\times\vec{A}$. However, if $\vec{A}$ were regular everywhere, $\vec{\nabla}\cdot\vec{B} = \vec{\nabla}\cdot(\vec{\nabla}\times\vec{A}) = 0$ and the magnetic charge would vanish. The vector potential of a monopole must be singular somewhere, and not only at the origin.

Recall how a vector potential enters the wave function. In a region where the field strength vanishes, $A_\mu = \frac{1}{e}\partial_\mu\Lambda$ is a pure gauge and the solution of the Schrödinger (or Klein-Gordon or Dirac) equation with covariant derivative $D_\mu = \partial_\mu - ieA_\mu$ is $\psi = e^{i\Lambda}\psi_0$, with ψ_0 the solution without a field. More generally, transporting a charged particle along a closed loop C multiplies its wave function by the phase

$$\exp\left(-ie\oint_C \vec{A}\cdot d\vec{l}\right) = \exp\left(-ie\Phi_C\right), \tag{3.29}$$

where Φ_C is the magnetic flux through a surface bounded by C. This is the Aharonov-Bohm phase.

Let us now take for C a circle at fixed r and polar angle θ (Fig. 3.3), and for the surface the cap above it. The flux $\Phi(r,\theta)$ through the cap grows from 0 at $\theta = 0$, where the loop shrinks to a point, to the total flux g at $\theta \to \pi$. But at $\theta \to \pi$ the loop shrinks again to a point, and $\oint \vec{A}\cdot d\vec{l}$ can remain equal to g only if $\vec{A}$ is singular at $\theta = \pi$. Since this holds for all r, the vector potential is singular along the whole negative z-axis, the **Dirac string**. It is like an infinitely thin solenoid which brings the flux g from infinity to the origin, where it spreads out radially. An explicit vector potential is

$$A_x = -\frac{g}{4\pi}\frac{y}{r(r+z)}, \qquad A_y = \frac{g}{4\pi}\frac{x}{r(r+z)}, \qquad A_z = 0, \tag{3.30}$$

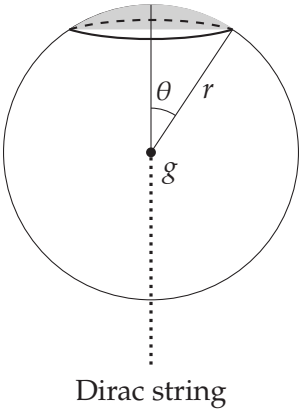


Figure 3.3. A loop at fixed r and polar angle θ around a monopole g, and the cap above it. Lowering the loop to the south pole forces the vector potential to be singular on the negative z-axis.

or, in spherical coordinates,

$$A_r = A_\theta = 0, \qquad A_\varphi = \frac{g}{4\pi}\frac{1-\cos\theta}{r\sin\theta}. \tag{3.31}$$

Exercise: *Show that the two expressions are the same, that* $\vec{\nabla}\times\vec{A}$ *reproduces Eq. 3.28 away from the negative* z*-axis, and that* A_φ *is singular only at* $\theta=\pi$.

Is the string observable? A charged particle encircling the string at a tiny distance picks up the phase of Eq. 3.29 with $\Phi_C \to g$. The string is unobservable, as it should be for a point-like monopole, only if this phase is trivial:

$$e\,g = 2\pi n, \qquad n\in\mathbb{Z}. \tag{3.32}$$

This is the **Dirac quantisation condition**. In Gaussian units it reads $eg = n\hbar c/2$, the form in which it is often quoted. Notice that it is a quantum condition: it contains $\hbar$. Its consequence is dramatic. If a single magnetic monopole exists anywhere in the universe, then every electric charge must be an integer multiple of $2\pi/g$. This is a possible explanation of the observed quantisation of electric charge. We will see in the next section that in some non-abelian gauge theories charge quantisation and the existence of monopoles go together.

The position of the string is a matter of choice. With the string along the positive z-axis we can use

$$A'_\varphi = -\frac{g}{4\pi}\frac{1+\cos\theta}{r\sin\theta}, \tag{3.33}$$

which gives the same magnetic field. Only the singularity at the origin, where $\vec{\nabla}\cdot\vec{B}\neq 0$, is physical. This suggests a construction due to Wu and Yang which avoids strings altogether. We cover the sphere around the monopole (and thus space without the origin) with two overlapping patches: R_N, which contains the north pole and excludes a neighbourhood of the south pole, and R_S, which contains the south pole and excludes the north pole. In R_N we use $\vec{A}_N$ of Eq. 3.31, in R_S we use $\vec{A}_S$ of Eq. 3.33. Each of them is regular in its own patch. In the overlap they differ by a gradient,

$$\vec{A}_S - \vec{A}_N = -\frac{g}{2\pi}\frac{1}{r\sin\theta}\hat{\varphi} = -\frac{g}{2\pi}\vec{\nabla}\varphi, \tag{3.34}$$

i.e. by a gauge transformation. In our conventions a gauge transformation acts as $\psi \to e^{i\alpha}\psi$, $A_\mu \to A_\mu + \frac{1}{e}\partial_\mu\alpha$, i.e. $\vec{A} \to \vec{A} - \frac{1}{e}\vec{\nabla}\alpha$. Therefore

$$\psi_S = S\,\psi_N, \qquad S = e^{i\alpha} = \exp\left(i\frac{eg}{2\pi}\varphi\right), \tag{3.35}$$

in the overlap region. The wave function of the charged particle must be single-valued in both patches, and so must the transition function S when we go around the equator, $\varphi \to \varphi + 2\pi$. This is again the Dirac condition $eg = 2\pi n$. Covariantly, the two potentials are related by $A^S_\mu = A^N_\mu + \frac{i}{e}S\partial_\mu S^{-1}$.

The magnetic charge is now hidden in the topology of the patching. Indeed, let us split the sphere at the equator and apply Stokes' theorem on each hemisphere. The equator bounds the northern hemisphere with positive and the southern with negative orientation:

$$\Phi = \int_{\text{north}} \vec{\nabla}\times\vec{A}_N\cdot d\vec{S} + \int_{\text{south}} \vec{\nabla}\times\vec{A}_S\cdot d\vec{S} = \oint_{\theta=\pi/2}\left(\vec{A}_N - \vec{A}_S\right)\cdot d\vec{l} = \frac{g}{2\pi}\int_0^{2\pi} d\varphi = g. \tag{3.36}$$

The flux is given by the winding of the transition function $S : S^1 \to U(1)$ around the equator, which is the integer $n = eg/2\pi$. Once more, $\pi_1(S^1) = \mathbb{Z}$ is at work.

In mathematical language, the Wu-Yang construction describes the monopole as a non-trivial **fibre bundle**: at every point of the base space, essentially the sphere S^2 surrounding the monopole, we attach the group manifold S^1 of $U(1)$ (the phase of the wave function). Locally the bundle looks like $S^2 \times S^1$, but globally it is twisted, like a Möbius strip, which locally is a product of an interval and a circle but cannot be described with one set of coordinates. For the minimal charge $n = 1$ the bundle is the three-sphere S^3 (the Hopf fibration).

The Dirac monopole is a consistent addition to electrodynamics, but nothing in QED requires it. Electrodynamics without monopoles is perfectly consistent, and the monopole is a point particle with an infinite self-energy, just like the classical electron. We now turn to theories where monopoles are smooth solitons with a finite and calculable mass, and where they **must** exist.

3.4 The 't Hooft-Polyakov monopole

Consider an $SU(2)$ gauge theory with a real scalar triplet ϕ^a, $a = 1,2,3$, in the adjoint representation:

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{a\,\mu\nu} + \frac{1}{2}(D_\mu\phi)^a(D^\mu\phi)^a - \frac{\lambda}{8}\left(\phi^a\phi^a - v^2\right)^2, \tag{3.37}$$

with

$$\begin{aligned}(D_\mu\phi)^a &= \partial_\mu\phi^a + e\,\epsilon^{abc}A^b_\mu\phi^c,\\ F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + e\,\epsilon^{abc}A^b_\mu A^c_\nu.\end{aligned} \tag{3.38}$$

These are the expressions of QFT II for $D_\mu = \partial_\mu - ieA^a_\mu T^a$ with the adjoint generators $(T^a)_{bc} = -i\epsilon^{abc}$ and structure constants $f^{abc} = \epsilon^{abc}$. Equivalently, ϕ^a is a vector of $SO(3)$.

The vacua satisfy $\phi^a\phi^a = v^2$; they form a two-sphere S^2. We may pick $\langle\phi^a\rangle = v\delta^{a3}$. Rotations around the third axis leave the vacuum invariant, so $SU(2)$ is spontaneously broken to $U(1)$. Writing $\phi^a = \langle\phi^a\rangle + \chi^a$, the mass terms of the gauge fields come from

$$\frac{1}{2}(D_\mu\phi)^a(D^\mu\phi)^a \supset \frac{e^2v^2}{2}\left(A^1_\mu A^{1\mu} + A^2_\mu A^{2\mu}\right) = e^2v^2\,W^+_\mu W^{-\mu}, \qquad W^\pm_\mu = \frac{A^1_\mu \mp iA^2_\mu}{\sqrt{2}}. \tag{3.39}$$

The field A^3_μ stays massless and is the photon of the unbroken $U(1)$. The $W^\pm$ bosons have electric charge $\pm e$ under this $U(1)$ and mass

$$m_W = ev. \tag{3.40}$$

The fluctuations $\chi^{1,2}$ are would-be Goldstone bosons which are eaten by the $W^\pm$, and χ^3 is a physical Higgs boson with mass $m_H = \sqrt{\lambda}\, v$.
Exercise: *Verify the masses m_W and m_H.*
This is the **Georgi-Glashow model**. It was proposed in 1972 as a model of the weak and electromagnetic interactions, alternative to the $SU(2) \times U(1)$ model of Glashow, Weinberg and Salam. It has no neutral weak current, and it was excluded by the discovery of neutral currents at CERN in 1973. Here we use it as the simplest theory with monopoles. Notice that in this model the gauge boson mass, the electric charge and the vacuum expectation value are tied together; if m_W and e are measured, v is known.

3.4.1 Topology and magnetic charge

A finite-energy configuration must approach a vacuum at spatial infinity, $\phi^a\phi^a \to v^2$. The boundary of three-dimensional space is a sphere S^2_∞, and the vacua form a sphere S^2. The field at infinity therefore defines a map $S^2 \to S^2$. As for circles, such maps are classified by an integer, the number of times the image covers the target sphere,

$$\pi_2(S^2) = \mathbb{Z}. \tag{3.41}$$

With $\hat\phi^a = \phi^a/|\phi|$, the winding number can be written as an integral over the sphere at infinity,

$$n = \frac{1}{8\pi}\oint_{S^2_\infty} dS_i\, \epsilon_{ijk}\epsilon^{abc}\hat\phi^a\partial_j\hat\phi^b\partial_k\hat\phi^c. \tag{3.42}$$

The integrand is the Jacobian of the map, i.e. the area element of the target sphere, and the total area of the unit sphere is 4π.
Exercise: *Show that the "hedgehog" $\hat\phi^a = x^a/r$ (the identity map) has $n = 1$.*
What is the physical meaning of n? Far from the core, the unbroken $U(1)$ points along $\hat\phi^a$, which now varies in space. To identify the electromagnetic field in a gauge invariant way, 't Hooft proposed the field strength

$$\mathcal{F}_{\mu\nu} = \hat\phi^a F^a_{\mu\nu} - \frac{1}{e}\epsilon^{abc}\hat\phi^a(D_\mu\hat\phi)^b(D_\nu\hat\phi)^c. \tag{3.43}$$

It is gauge invariant and it is defined wherever $|\phi| \neq 0$. In the vacuum $\phi^a = v\delta^{a3}$ the second term vanishes and $\mathcal{F}_{\mu\nu} = \partial_\mu A^3_\nu - \partial_\nu A^3_\mu$, the field strength of the photon. Using $\hat\phi^a\hat\phi^a = 1$ and hence $\hat\phi^a\partial_\mu\hat\phi^a = 0$ repeatedly, we can bring it to the form

$$\mathcal{F}_{\mu\nu} = \partial_\mu\left(\hat\phi^a A^a_\nu\right) - \partial_\nu\left(\hat\phi^a A^a_\mu\right) - \frac{1}{e}\epsilon^{abc}\hat\phi^a\partial_\mu\hat\phi^b\partial_\nu\hat\phi^c. \tag{3.44}$$

Exercise: *Prove Eq. 3.44. Hint: the vectors $\hat\phi$, $\partial_\mu\hat\phi$ and $\partial_\nu\hat\phi$ satisfy $\hat\phi\cdot\partial\hat\phi = 0$, so $\partial_\mu\hat\phi\times\partial_\nu\hat\phi$ is parallel to $\hat\phi$.*
The first part is a curl, exactly as in electrodynamics. The second part contains no gauge field at all; it is the topological density of Eq. 3.42. The magnetic field is $B_i = -\frac{1}{2}\epsilon_{ijk}\mathcal{F}_{jk}$ (so that $\vec B = \vec\nabla\times\vec A$ for $A^i = -A_i$),

$$B_i = -\epsilon_{ijk}\partial_j\left(\hat\phi^a A^a_k\right) + \frac{1}{2e}\epsilon_{ijk}\epsilon^{abc}\hat\phi^a\partial_j\hat\phi^b\partial_k\hat\phi^c. \tag{3.45}$$

The divergence of a curl integrates to zero over a closed surface, and the magnetic flux through the sphere at infinity is

$$\Phi = \oint_{S^2_\infty} d\vec{S}\cdot\vec{B} = \frac{1}{2e}\oint_{S^2_\infty} dS_i\,\epsilon_{ijk}\epsilon^{abc}\hat{\phi}^a\partial_j\hat{\phi}^b\partial_k\hat{\phi}^c = \frac{4\pi n}{e}. \tag{3.46}$$

Every finite-energy configuration with $n \neq 0$ is a **magnetic monopole** with magnetic charge $g = 4\pi n/e$. The magnetic charge is topological: it cannot change in time, and no equation of motion was needed to compute it.

Let us compare with the Dirac condition, Eq. 3.32. The W bosons have charge e, and $e\cdot 4\pi/e = 4\pi = 2\pi\cdot 2$. If we add matter in the fundamental (doublet) representation of $SU(2)$, its components have charges $\pm e/2$, and then $(e/2)\cdot(4\pi/e) = 2\pi$: the 't Hooft-Polyakov monopole has the smallest magnetic charge allowed by Dirac's condition. In this theory electric charge is quantised because the $U(1)$ is embedded in the compact non-abelian group $SU(2)$, and the same embedding produces the monopoles. The two facts go together.

3.4.2 The monopole solution and its mass

Let us now construct the soliton in the sector $n = 1$. We start with the hedgehog boundary condition

$$\phi^a(\vec{x}) \to v\frac{x^a}{r}, \qquad r\to\infty. \tag{3.47}$$

As for the vortex, the gradient energy of the scalar would diverge (here linearly, $\int r^2 dr/r^2$) unless the covariant derivative vanishes at infinity:

$$(D_i\phi)^a = v\left[\partial_i\frac{x^a}{r} + e\,\epsilon^{abc}A^b_i\frac{x^c}{r}\right] = v\left[\frac{\delta^{ai}}{r} - \frac{x^a x^i}{r^3} + e\,\epsilon^{abc}A^b_i\frac{x^c}{r}\right] \to 0. \tag{3.48}$$

Let us try $A^b_i = \kappa\,\epsilon^{bij}x^j/r^2$. With $\epsilon^{abc}\epsilon^{bij} = \delta^{ci}\delta^{aj} - \delta^{cj}\delta^{ai}$ the last term becomes $e\kappa\,(x^a x^i/r^3 - \delta^{ai}/r)$, and the covariant derivative vanishes for $\kappa = 1/e$:

$$A^a_i \to \frac{1}{e}\epsilon^{aij}\frac{x^j}{r^2}, \qquad r\to\infty. \tag{3.49}$$

For this configuration the non-abelian field strength at large r points along $\hat{x}^a$, and the magnetic field of Eq. 3.46 is $\vec{B} = \hat{r}/(er^2)$, a Coulomb field of magnetic charge $4\pi/e$. The full solution is found with the ansatz

$$\phi^a = v\,\frac{H(\xi)}{\xi}\,\frac{x^a}{r}, \qquad A^a_i = \frac{1}{e}\epsilon^{aij}\frac{x^j}{r^2}\,[1-K(\xi)], \qquad \xi = evr, \tag{3.50}$$

with $K\to 0$, $H/\xi\to 1$ for $\xi\to\infty$ and $K\to 1$, $H/\xi\to 0$ for $\xi\to 0$, so that the fields are regular at the origin. This is the **'t Hooft-Polyakov monopole**. Unlike the Dirac monopole, it is smooth everywhere: the core, of size $1/m_W$, is a region where the $SU(2)$ symmetry is restored and the massive fields are excited.

Exercise: *Insert the ansatz 3.50 into the static energy and show that*

$$E = \frac{4\pi v}{e}\int_0^\infty \frac{d\xi}{\xi^2}\left[\frac{1}{2}(\xi H' - H)^2 + H^2K^2 + (\xi K')^2 + \frac{1}{2}(K^2-1)^2 + \frac{\lambda}{8e^2}(H^2-\xi^2)^2\right].$$

Derive the equations for K and H which minimise it: $\xi^2K'' = KH^2 + K(K^2-1)$, $\xi^2H'' = 2K^2H + \frac{\lambda}{2e^2}H(H^2-\xi^2)$.

Even without solving these equations we can find a lower bound for the mass, with the same trick that we used for the kink. For a static configuration with $A_0^a = 0$ the energy is

$$M = \int d^3x \left[\frac{1}{2}B_i^a B_i^a + \frac{1}{2}(D_i\phi)^a(D_i\phi)^a + V(\phi)\right], \qquad B_i^a = -\frac{1}{2}\epsilon_{ijk}F_{jk}^a. \tag{3.51}$$

We complete the square,

$$M = \int d^3x \left[\frac{1}{2}\left(B_i^a - (D_i\phi)^a\right)^2 + V(\phi)\right] + \int d^3x\, B_i^a(D_i\phi)^a. \tag{3.52}$$

The last term is a total derivative. The covariant derivative obeys the Leibniz rule for the invariant product $B_i^a\phi^a$, and the Bianchi identity, $\epsilon_{ijk}D_iF_{jk} = 0$, gives $(D_iB_i)^a = 0$. Therefore

$$B_i^a(D_i\phi)^a = \partial_i\left(B_i^a\phi^a\right) - (D_iB_i)^a\phi^a = \partial_i\left(B_i^a\phi^a\right). \tag{3.53}$$

The first integral is positive, and we obtain

$$M \geq \int d^3x\, \partial_i\left(B_i^a\phi^a\right) = \oint_{S^2_\infty} dS_i\, B_i^a\phi^a. \tag{3.54}$$

At infinity $\phi^a = v\hat{\phi}^a$ and $D_i\phi = 0$, so $B_i^a\phi^a = vB_i$ with B_i the magnetic field of the unbroken $U(1)$ (Eq. 3.43). The surface integral is v times the magnetic flux, $4\pi v/e$. (Completing the square with the opposite sign gives the bound $M \geq -v\Phi$, relevant for $n < 0$.) We have found the **Bogomolny bound**

$$M \geq \frac{4\pi v}{e}|n| = \frac{m_W}{\alpha}|n|, \qquad \alpha = \frac{e^2}{4\pi}. \tag{3.55}$$

It is saturated when $V = 0$, i.e. in the limit $\lambda \to 0$ with v fixed, and when the first-order Bogomolny equations $B_i^a = (D_i\phi)^a$ hold. In this limit the solution is known in closed form, $K = \xi/\sinh\xi$, $H = \xi\coth\xi - 1$ (Prasad and Sommerfield).

Exercise: *Verify that the Prasad-Sommerfield functions solve the equations of the previous exercise for $\lambda = 0$.*

For $\lambda > 0$ the mass is $M = (4\pi v/e)\,C(\lambda/e^2)$ with C a slowly varying function of order one. As for the kink, the soliton mass is inversely proportional to the coupling: for weak coupling the monopole is much heavier than the W boson, $M \sim m_W/\alpha$.

Does nature have such monopoles? The Georgi-Glashow model is not realised in nature, and the Standard Model, which is, has no 't Hooft-Polyakov monopoles. In the Standard Model the photon is a combination of an $SU(2)$ generator and the hypercharge of the abelian $U(1)_Y$ factor. Nothing in the theory quantises hypercharge: we could add a singlet field with an arbitrarily small hypercharge, hence an arbitrarily small electric charge, and then Dirac's condition would forbid monopoles altogether. The situation changes in grand unified theories, where the gauge group has a single simple factor, such as $SU(5)$, broken at a scale $m_X \sim 10^{15}$ GeV to $SU(3)\times SU(2)\times U(1)$. Electric charge is then quantised and monopoles with masses $M \sim m_X/\alpha \sim 10^{17}$ GeV must exist. They cannot be produced in any accelerator. However, in the early universe, at temperatures where the grand unified symmetry was restored, a phase transition to the broken phase produces regions of space with uncorrelated orientations of the vacuum expectation value of ϕ, and monopoles form at the boundaries of such regions. We will return to phase transitions in the early universe in the next section.

3.5 Phase transitions in the early universe

We have found in this chapter a whole zoo of classical field configurations with finite energy: kinks, vortices, and magnetic monopoles. Their existence is guaranteed by topology, i.e. by the structure of the vacuum manifold of a theory with a spontaneously broken symmetry. Their masses, however, are set by the scale of symmetry breaking, and we have seen at the end of Section 3.4 that grand unified monopoles, with masses of order 10^{17} GeV, can only have been produced in the early universe. Let us now understand how.

The universe was once very hot, and it has been cooling down ever since. We will see in this section that at high temperatures spontaneously broken symmetries are restored: the minimum of the free energy sits at $\langle\phi\rangle = 0$. As the temperature drops below a critical value, the scalar field acquires a vacuum expectation value. In the Standard Model this happens at temperatures of order 100 GeV, where $SU(2)_L \times U(1)_Y \to U(1)_{\rm em}$. If the Standard Model is embedded in a grand unified theory, e.g. with gauge group $SU(5)$, there is an earlier transition $SU(5) \to SU(3)_c \times SU(2)_L \times U(1)_Y$ at temperatures of order 10^{15} GeV, and possibly a sequence of intermediate ones. Notice that in regions of space which are not in causal contact the field cannot "know" in which direction of the vacuum manifold the neighbouring regions have chosen to fall. Topologically non-trivial configurations, domain walls, strings or monopoles, are then unavoidably frozen in at the boundaries of such regions. Phase transitions can affect the history of the universe in other ways too: if the universe supercools in a false vacuum, the latent heat released at the end of the transition reheats the plasma, and if the universe is dominated for a while by the constant energy density of a false vacuum, it expands exponentially. We will study this last possibility, inflation, in the next chapter.

In what follows we will first learn how to compute the effective potential at finite temperature. We will then apply it to a simple gauge theory with a Higgs field and find under which conditions the transition is of first or of second order.

3.5.1 Thermal field theory in one page

In quantum statistical mechanics, a system in thermal equilibrium at temperature T is described by the partition function

$$Z = \mathrm{tr}\, e^{-\beta H} = \sum_n \langle n|\, e^{-\beta H}\, |n\rangle\,, \qquad \beta \equiv \frac{1}{T}, \tag{3.56}$$

where we set $k_B = 1$. Let us take for simplicity a quantum mechanical particle and evaluate the trace in the basis of position eigenstates. Inserting twice the unit operator, $1 = \int d^3x\, |x\rangle\langle x|$, we find:

$$\begin{aligned} Z &= \sum_n \int d^3x\, d^3y\, \langle n|\, x\rangle \langle x|\, e^{-\beta H}\, |y\rangle \langle y|\, n\rangle = \int d^3x\, d^3y\, \langle x|\, e^{-\beta H}\, |y\rangle \sum_n \langle y|\, n\rangle \langle n|\, x\rangle \\ &= \int d^3x\, \langle x|\, e^{-\beta H}\, |x\rangle\,. \end{aligned} \tag{3.57}$$

This is a familiar object! The matrix element $\langle x|\, e^{-iHt}\, |x\rangle$ is the amplitude for a particle to start at x and to return to the same point x after a time t. The partition function is the same amplitude at the imaginary time $t = -i\beta$, integrated over all starting points. We have learned in QFT II how to write such amplitudes as path integrals, and we have also learned that the path integral is best defined after a Wick rotation $t = -i\tau$. There are only two differences from the vacuum amplitudes of field theory: the Euclidean time τ runs over a

finite interval $[0, \beta]$, and the paths must return to their starting point. We can therefore write:

$$Z = \int_{q(\beta)=q(0)} \mathcal{D}q \; e^{-S_E[q]}, \qquad S_E[q] = \int_0^\beta d\tau \left[\frac{m}{2} \left(\frac{dq}{d\tau} \right)^2 + V(q) \right]. \tag{3.58}$$

The generalisation to a scalar field theory is immediate:

$$Z = \operatorname{tr} e^{-\beta H} = \int_{\phi(\beta,\vec{x})=\phi(0,\vec{x})} \mathcal{D}\phi \; \exp\left\{ - \int_0^\beta d\tau \int d^3x \left[\frac{1}{2} \left(\partial_\tau \phi \right)^2 + \frac{1}{2} \left(\vec{\nabla} \phi \right)^2 + V(\phi) \right] \right\}. \tag{3.59}$$

Statistical field theory is thus nothing else than Euclidean field theory with a compact time direction of circumference β. In the limit $T \to 0, \beta \to \infty$, we recover the Wick-rotated vacuum amplitudes of ordinary field theory.

Periodicity in τ means that the Fourier decomposition of the field contains only discrete frequencies,

$$\phi(\tau, \vec{x}) = T \sum_{n=-\infty}^{\infty} \int \frac{d^3k}{(2\pi)^3} \, e^{i\omega_n \tau + i\vec{k}\cdot\vec{x}} \, \tilde{\phi}_n(\vec{k}), \qquad \omega_n = 2\pi n T, \tag{3.60}$$

which are known as **Matsubara frequencies**. For fermion fields, the trace over Grassmann variables requires **antiperiodic** boundary conditions, $\psi(\beta, \vec{x}) = -\psi(0, \vec{x})$, and the frequencies are $\omega_n = (2n+1)\pi T$. In practice, the Feynman rules of finite temperature field theory are those of Euclidean field theory with the replacement

$$\int \frac{d^4k_E}{(2\pi)^4} \; \to \; T \sum_n \int \frac{d^3k}{(2\pi)^3}, \qquad k_E^0 \to \omega_n. \tag{3.61}$$

With the partition function at hand we can compute all thermodynamic quantities, e.g. the free energy $F = -T \log Z$, the pressure and the entropy. Here we are interested in something else: the expectation value of the scalar field in the thermal ensemble. As in the vacuum, this is obtained by coupling the field to a source, differentiating $\log Z[J]$ with respect to J, and Legendre transforming. Nothing changes in this formal procedure, and we are led again to an effective potential, whose minimum is now the thermal expectation value of ϕ.

3.5.2 The effective potential at one loop

We recall from QFT II (Eq. 6.25) and from the recap in Section 1.1 that the effective action $\Gamma[\langle\phi\rangle]$ is the Legendre transform of $W[J] = -i \log Z[J]$,

$$\Gamma[\langle\phi\rangle] = W[J] - \int d^4x \, J(x) \langle\phi(x)\rangle, \qquad \frac{\delta\Gamma}{\delta \langle\phi(x)\rangle} = -J(x), \tag{3.62}$$

and that for a constant field $\Gamma[\phi] = -\left(\int d^4x\right) V_{\text{eff}}(\phi)$. In the absence of sources the vacuum expectation value of the field is a minimum of the effective potential. In QFT II we computed V_{eff} at tree level, finding the classical potential. Let us now compute the first quantum correction, which we will need at finite temperature.

Consider a real scalar field with action $S[\phi] = \int d^4x \left[\frac{1}{2}(\partial\phi)^2 - V(\phi) \right]$. We evaluate the path integral

$$e^{\frac{i}{\hbar}W[J]} = \int \mathcal{D}\phi \; e^{\frac{i}{\hbar}(S[\phi] + J\phi)}, \qquad J\phi \equiv \int d^4x J(x)\phi(x), \tag{3.63}$$

with the method of steepest descent, keeping $\hbar$ explicitly to count the loops. The exponent is stationary at the classical solution ϕ_s in the presence of the source,

$$\partial^2\phi_s(x) + V'(\phi_s(x)) = J(x). \tag{3.64}$$

We write $\phi = \phi_s + \tilde{\phi}$ and expand the exponent to quadratic order in the fluctuation $\tilde{\phi}$. The linear term vanishes by construction, and we find:

$$\begin{aligned} e^{\frac{i}{\hbar}W[J]} &\approx e^{\frac{i}{\hbar}\left(S[\phi_s]+J\phi_s\right)} \int \mathcal{D}\tilde{\phi}\ \exp\left\{-\frac{i}{2\hbar}\int d^4x\ \tilde{\phi}\left[\partial^2 + V''(\phi_s)\right]\tilde{\phi}\right\} \\ &= \exp\left\{\frac{i}{\hbar}\left(S[\phi_s] + J\phi_s\right) - \frac{1}{2}\operatorname{Tr}\log\left[\partial^2 + V''(\phi_s)\right]\right\}, \end{aligned} \tag{3.65}$$

where we have used that a Gaussian integral gives the inverse square root of the determinant of the operator, and det = exp Tr log. Therefore,

$$W[J] = S[\phi_s] + J\phi_s + \frac{i\hbar}{2}\operatorname{Tr}\log\left[\partial^2 + V''(\phi_s)\right] + O(\hbar^2). \tag{3.66}$$

Differentiating with respect to J, the dependence of ϕ_s on J drops out because of the stationarity condition, and $\langle\phi\rangle = \phi_s + O(\hbar)$. The Legendre transform is then simply:

$$\Gamma[\phi] = S[\phi] + \frac{i\hbar}{2}\operatorname{Tr}\log\left[\partial^2 + V''(\phi)\right] + O(\hbar^2). \tag{3.67}$$

For an arbitrary $\phi(x)$ we would need all the eigenvalues of the operator $\partial^2 + V''(\phi(x))$, which is hopeless. For a constant ϕ, however, the operator is translation invariant and diagonal in momentum space:

$$\operatorname{Tr}\log\left[\partial^2 + V''(\phi)\right] = \int d^4x \int \frac{d^4k}{(2\pi)^4}\ \log\left[-k^2 + V''(\phi)\right]. \tag{3.68}$$

We add a ϕ-independent constant to make the argument of the logarithm dimensionless, and we Wick rotate, $k^0 = ik_E^0$, $k^2 = -k_E^2$, $d^4k = i\,d^4k_E$. We obtain:

$$V_{\text{eff}}(\phi) = V(\phi) + \frac{\hbar}{2}\int \frac{d^4k_E}{(2\pi)^4}\ \log\left[\frac{k_E^2 + V''(\phi)}{k_E^2}\right] + O(\hbar^2). \tag{3.69}$$

The physical meaning is transparent: each mode of the fluctuation field with momentum k is a harmonic oscillator of frequency $\sqrt{\vec{k}^2 + V''(\phi)}$, and the integral is the sum of the zero-point energies of all these oscillators (exercise below). The background field ϕ acts as a field-dependent mass, $M^2(\phi) \equiv V''(\phi)$.

The integral is ultraviolet divergent, as zero-point energies always are. We regulate it in $d = 4 - 2\epsilon$ dimensions. Differentiating with respect to M^2 we find a tadpole integral,

$$\frac{\partial}{\partial M^2}\frac{1}{2}\int \frac{d^dk_E}{(2\pi)^d}\log\left(k_E^2 + M^2\right) = \frac{1}{2}\int \frac{d^dk_E}{(2\pi)^d}\frac{1}{k_E^2 + M^2} = \frac{\Gamma(1-d/2)}{2(4\pi)^{d/2}}\left(M^2\right)^{d/2-1}, \tag{3.70}$$

and, integrating back,

$$\frac{1}{2}\int \frac{d^dk_E}{(2\pi)^d}\log\left(k_E^2 + M^2\right) = -\frac{\Gamma(-d/2)}{2(4\pi)^{d/2}}M^d. \tag{3.71}$$

With $\Gamma(-2+\epsilon) = \frac{1}{2\epsilon} + \frac{3}{4} - \frac{\gamma_E}{2} + O(\epsilon)$ and a scale μ to keep the dimension of the potential, the pole is removed by a counterterm proportional to $M^4 = (V'')^2$; for a renormalisable potential this is a counterterm for the mass and the quartic coupling. In the $\overline{\text{MS}}$ scheme we obtain the famous **Coleman-Weinberg** potential:

$$V_{\text{eff}}(\phi) = V(\phi) + \frac{\left[V''(\phi)\right]^2}{64\pi^2}\left[\log\frac{V''(\phi)}{\mu^2} - \frac{3}{2}\right] + O(\hbar^2). \tag{3.72}$$

Exercise: *Perform the k_E^0 integration in Eq. 3.69 and show that, up to a ϕ-independent constant, the one-loop correction is $\int \frac{d^3k}{(2\pi)^3}\frac{1}{2}\sqrt{\vec{k}^2 + V''(\phi)}$, the sum of the zero-point energies of the fluctuations. (Hint: differentiate with respect to V'' first.)*
Exercise: *For several fields, V'' is replaced by the mass matrix in the background ϕ. Show that a massive vector boson contributes like three real scalars and a Dirac fermion of mass $m(\phi)$ like -4 real scalars, i.e. $V_1 = \sum_i \frac{n_i}{64\pi^2} m_i^4(\phi)\left[\log\frac{m_i^2(\phi)}{\mu^2} - c_i\right]$ with $n_i = +1, +3, -4$ and $c_i = 3/2, 5/6, 3/2$ in $\overline{\text{MS}}$.*

3.5.3 The effective potential at finite temperature

At finite temperature we repeat the computation of the previous subsection with the Euclidean time compactified. The only change in Eq. 3.69 is the replacement of the k_E^0 integral by the Matsubara sum:

$$V_1(\phi, T) = \frac{T}{2}\sum_{n=-\infty}^{\infty}\int\frac{d^3k}{(2\pi)^3}\log\left[\omega_n^2 + E_k^2\right], \qquad E_k = \sqrt{\vec{k}^2 + M^2(\phi)}. \tag{3.73}$$

The sum is performed with a standard trick. We differentiate with respect to E_k,

$$\frac{\partial}{\partial E_k}\frac{T}{2}\sum_n \log\left[(2\pi nT)^2 + E_k^2\right] = T\sum_n \frac{E_k}{(2\pi nT)^2 + E_k^2} = \frac{1}{2}\coth\frac{E_k}{2T} = \frac{1}{2} + \frac{1}{e^{E_k/T} - 1}, \tag{3.74}$$

where we have used $\sum_n \frac{1}{n^2+x^2} = \frac{\pi}{x}\coth\pi x$. Integrating back with respect to E_k, we find:

$$V_1(\phi, T) = \int\frac{d^3k}{(2\pi)^3}\left[\frac{E_k}{2} + T\log\left(1 - e^{-E_k/T}\right)\right] = V_1(\phi, 0) + \frac{T^4}{2\pi^2}J_B\left(\frac{M^2(\phi)}{T^2}\right), \tag{3.75}$$

with

$$J_B(y^2) = \int_0^\infty dx\, x^2 \log\left(1 - e^{-\sqrt{x^2+y^2}}\right). \tag{3.76}$$

The first term is the zero-temperature result of Eq. 3.72. The second term is the free energy of a gas of particles with mass $M(\phi)$; this is the term which knows about the temperature. For fermions the antiperiodic boundary conditions lead to $-\frac{T^4}{2\pi^2}J_F$ per degree of freedom, with $J_F(y^2) = \int_0^\infty dx\, x^2 \log(1 + e^{-\sqrt{x^2+y^2}})$.

In the early universe the temperature is high compared to the masses. We then need the high-temperature expansion of the thermal functions:

$$\frac{T^4}{2\pi^2}J_B\left(\frac{M^2}{T^2}\right) = -\frac{\pi^2T^4}{90} + \frac{M^2T^2}{24} - \frac{M^3T}{12\pi} + O\left(M^4\log\frac{M^2}{T^2}\right), \tag{3.77}$$

$$-\frac{T^4}{2\pi^2}J_F\left(\frac{M^2}{T^2}\right) = -\frac{7}{8}\frac{\pi^2T^4}{90} + \frac{M^2T^2}{48} + O\left(M^4\log\frac{M^2}{T^2}\right). \tag{3.78}$$

The first term is the familiar Stefan-Boltzmann free energy of a relativistic gas. Notice the term M^3T. It is present only for bosons; it comes from the $n = 0$ Matsubara mode, which behaves as a field in three dimensions, and $M^3 = \left(M^2\right)^{3/2}$ is not analytic in the field.
Exercise: *Verify numerically the expansion of Eq. 3.77 for $M/T = 0.1$ and 0.3. Then derive it: show that $\frac{\partial J_B}{\partial y^2} = \frac{1}{2}\int_0^\infty dx\, \frac{x^2}{\sqrt{x^2+y^2}} \frac{1}{e^{\sqrt{x^2+y^2}}-1}$ and that the small-y behaviour of this integral is $\frac{\pi^2}{12} - \frac{\pi}{4}y + \ldots$ (the linear term comes from the region $x \sim y$, where the Bose distribution can be replaced by $1/\sqrt{x^2+y^2}$).*

Let us now apply these results to the simplest gauge theory with a Higgs mechanism, the abelian Higgs model:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \left(D_\mu\phi\right)^* D^\mu\phi - V(\phi), \qquad V(\phi) = \lambda\left(\phi^*\phi - \frac{v^2}{2}\right)^2, \qquad D_\mu = \partial_\mu - ieA_\mu. \tag{3.79}$$

We take the background real, $\phi = \varphi/\sqrt{2}$, so that at tree level $V_0(\varphi) = -\frac{\mu^2}{2}\varphi^2 + \frac{\lambda}{4}\varphi^4$ with $\mu^2 = \lambda v^2$. In the background φ the field-dependent masses are

$$m_h^2(\varphi) = -\mu^2 + 3\lambda\varphi^2, \qquad m_G^2(\varphi) = -\mu^2 + \lambda\varphi^2, \qquad m_A^2(\varphi) = e^2\varphi^2, \tag{3.80}$$

for the Higgs, the would-be Goldstone boson and the three polarisations of the vector boson. Inserting them into the high-temperature expansion, we find, up to a φ-independent term:

$$V_{\text{eff}}(\varphi, T) \approx \frac{1}{2}m^2(T)\,\varphi^2 - \frac{C}{3}\,T\,\varphi^3 + \frac{\lambda}{4}\varphi^4, \tag{3.81}$$

with

$$m^2(T) = -\mu^2 + \frac{4\lambda + 3e^2}{12}\,T^2, \qquad 4\pi C = (3\lambda)^{3/2} + \lambda^{3/2} + 3e^3. \tag{3.82}$$

In the cubic term we have neglected μ^2 in the scalar masses. This is an approximation which is valid if the scalar contributions to C are small anyway, i.e. for $\lambda^{3/2} \ll e^3$. We also need $\lambda \gg e^4$, so that the zero-temperature Coleman-Weinberg correction, of order $e^4\varphi^4 \log$, is small compared to the tree-level quartic term. Our formula is thus trustworthy for $e^4 \ll \lambda \ll e^2$.
Exercise: *Derive Eq. 3.81 from Eq. 3.77.*

What does this potential teach us? At zero temperature $m^2(0) = -\mu^2 < 0$ and the symmetry is broken. At very high temperatures the thermal mass term wins, $m^2(T) > 0$, and the minimum is at $\varphi = 0$: **the symmetry is restored**. Physically, the particles of the hot plasma prefer the phase where they are light. If the cubic term were absent, the minimum would move continuously away from the origin when $m^2(T)$ changes sign; this would be a second-order transition. The cubic term changes the picture, as we show in Fig. 3.4. At a temperature T_D the potential has only the symmetric minimum. When the temperature drops, a second minimum appears at $\varphi \neq 0$ (curve C). At a temperature T_c the two minima are degenerate (curve B), and below T_c the broken minimum is the true vacuum (curve A), separated from the false vacuum at $\varphi = 0$ by a barrier. The universe can then only reach the true vacuum by tunnelling, i.e. by nucleating bubbles of the broken phase which expand and fill the space. This is a **first-order phase transition**.

The critical temperature is easily found. At $T = T_c$ the potential must have the form $\frac{\lambda}{4}\varphi^2(\varphi - \varphi_c)^2$. Comparing with Eq. 3.81, we find:

$$\varphi_c = \frac{2C}{3\lambda}\,T_c, \qquad m^2(T_c) = \frac{2C^2}{9\lambda}\,T_c^2 \qquad \rightsquigarrow \qquad \frac{\varphi_c}{T_c} = \frac{2C}{3\lambda}. \tag{3.83}$$

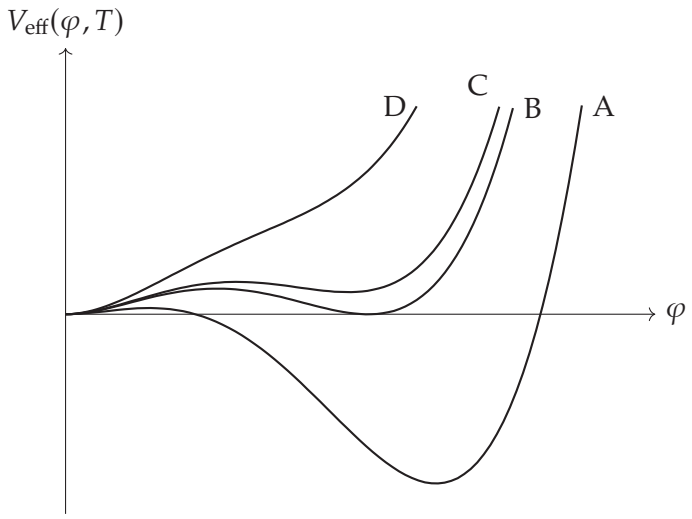


Figure 3.4. The effective potential of Eq. 3.81 at four temperatures $T_D > T_C > T_B = T_c > T_A$ (curves D, C, B, A): only the symmetric minimum (D); a metastable broken minimum (C); degenerate minima at the critical temperature (B); the broken minimum is the true vacuum (A).

The ratio φ_c/T_c measures the strength of the transition: the jump of the order parameter in units of the temperature. The transition is strongly first order only for a small scalar self-coupling, i.e. for a light Higgs boson compared to the vector boson, $\lambda \lesssim e^3$. We will see in the section on baryogenesis why this ratio matters a lot.

Exercise: *Verify Eq. 3.83. Estimate φ_c/T_c for the Standard Model, where in Eq. 3.81 the cubic term is dominated by the W and Z bosons, $4\pi C \simeq 3(2m_W^3 + m_Z^3)/v^3$, and $\lambda = m_h^2/(2v^2)$ with $v = 246$ GeV. What upper bound on m_h do you obtain from the requirement $\varphi_c/T_c \gtrsim 1$?*

We must add a word of caution. The perturbative picture breaks down near T_c for a heavy Higgs boson, because the expansion parameter of the theory of the $n = 0$ Matsubara modes is $\sim g^2 T/m_W(T)$, which is not small. Lattice simulations of the electroweak theory show that the first-order line ends at $m_h \simeq 72$ GeV. For the measured $m_h = 125$ GeV the electroweak "transition" of the Standard Model is a smooth **crossover**, at a pseudo-critical temperature $T \simeq 160$ GeV, with no bubbles and no departure from equilibrium. A first-order electroweak transition therefore requires physics beyond the Standard Model (e.g. extra scalars).

3.5.4 Topological defects in a cooling universe

We are now in a position to understand why topological defects are produced in the early universe. When the universe cools through a symmetry-breaking transition, the expectation value of ϕ is chosen independently in regions which are not causally connected. The correlation length of the field can at most be as large as the horizon, $d_H \sim t \sim M_{\text{Pl}}/T^2$ in a radiation-dominated universe (we will derive this in the next chapter). If the vacuum manifold has non-trivial topology, defects form at the borders of regions where the field has chosen different vacua: domain walls if π_0 of the vacuum manifold is non-trivial (the kinks of this chapter in three dimensions), strings if π_1 is non-trivial (the vortices), monopoles if π_2 is non-trivial. This is the **Kibble mechanism**. We expect at least about one defect per horizon volume.

Let us estimate what this means for the grand unified monopoles of Section 3.4. At $T_c \sim 10^{15}$ GeV the horizon contains about one monopole, so the monopole number

density is $n_M \gtrsim T_c^6/M_{\rm Pl}^3$. Compared to the number density of photons, $n_\gamma \sim T_c^3$, we find $n_M/n_\gamma \gtrsim (T_c/M_{\rm Pl})^3 \sim 10^{-12}$. This ratio stays roughly constant during the expansion if monopoles do not annihilate efficiently (they do not). Today $n_\gamma \simeq 410\,\mathrm{cm}^{-3}$, and monopoles of mass 10^{17} GeV would give an energy density many orders of magnitude larger than the critical density of the universe. The universe would have collapsed long ago! Domain walls are even worse: a single wall of tension $\sigma \sim v^3$ across the horizon carries an energy growing with the size of the universe. This is the **monopole problem** (or more generally the defect problem).

Either the defects did not form, which topology forbids if the theory and the transition are as we described, or they formed and were later diluted. A period of exponential expansion of the universe after their formation, inflation, dilutes their density by an enormous factor. This is one of the main motivations for the next chapter.

Exercise: *Using $n_M/n_\gamma \sim 10^{-12}$, compute the present monopole energy density for $m_M \sim 10^{17}$ GeV and compare it to the critical density $\rho_c \simeq 10^{-5}\,\mathrm{GeV\,cm}^{-3}$.*

3.6 Baryon and lepton number violation

In the Standard Model, baryon number and lepton number are conserved in every Feynman diagram. This is not a consequence of a principle; it is an "accident" of the field content and of renormalisability: there is no gauge invariant operator of dimension four or less which violates them. Nevertheless, the proton could decay if B were violated, and the universe contains baryons but almost no antibaryons. Is B really conserved? We will see that it is not, already in the Standard Model, because of the anomaly.

The baryon and lepton number currents are

$$j_B^\mu = \frac{1}{3}\sum_{\text{generations}} \left(\bar q_L \gamma^\mu q_L + \bar u_R \gamma^\mu u_R + \bar d_R \gamma^\mu d_R\right), \qquad j_L^\mu = \sum_{\text{generations}} \left(\bar \ell_L \gamma^\mu \ell_L + \bar e_R \gamma^\mu e_R\right), \tag{3.84}$$

where q_L and ℓ_L are the left-handed quark and lepton doublets, a sum over colour is understood, and $B = \int d^3x\, j_B^0$, $L = \int d^3x\, j_L^0$. Classically both currents are conserved. The baryon current, however, is a vector current for the gluons but a chiral current for the W bosons, which couple only to the left-handed doublets. We have learned in Chapter 1 that a current which is chiral with respect to the gauge fields is anomalous. Applying the result for the triangle diagram, Eq. 1.66, one left-handed $SU(2)$ doublet contributes $\frac{g^2}{32\pi^2} W^a_{\mu\nu}\tilde W^{a\,\mu\nu}$ to the divergence of its fermion number current. Each generation has three coloured quark doublets with $B = 1/3$ and one lepton doublet with $L = 1$. We therefore find:

$$\partial_\mu j_B^\mu = \partial_\mu j_L^\mu = \frac{N_f}{32\pi^2}\left(g^2\, W^a_{\mu\nu}\tilde W^{a\,\mu\nu} - g'^2\, B_{\mu\nu}\tilde B^{\mu\nu}\right), \qquad \tilde W^{a\,\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} W^a_{\rho\sigma}, \tag{3.85}$$

where $N_f = 3$ is the number of generations, $W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc} W^b_\mu W^c_\nu$, and $B_{\mu\nu}$ is the hypercharge field strength. (The overall sign depends on the convention for ϵ^{0123}; with $\epsilon^{0123} = +1$ it is the one which follows from Eq. 1.66, as for the $U(1)_A$ anomaly of Eq. 2.99. It is irrelevant for what follows.) Remarkably, the anomalies of B and L are equal. Therefore,

$$\partial_\mu\left(j_B^\mu - j_L^\mu\right) = 0, \qquad \partial_\mu\left(j_B^\mu + j_L^\mu\right) = \frac{N_f}{16\pi^2}\left(g^2\, W\tilde W - g'^2 B\tilde B\right). \tag{3.86}$$

$B - L$ is an exact global symmetry of the Standard Model, while $B + L$ is violated.

Exercise: *Starting from the anomaly of the axial current of a Dirac fermion in a representation*

R of a gauge group, $\partial_\mu j^\mu_5 = -\frac{g^2 T(R)}{8\pi^2} F^a_{\mu\nu}\tilde{F}^{a\,\mu\nu}$ (Eq. 1.66 and Eq. 2.99), with $\mathrm{tr}(t^a t^b) = T(R)\delta^{ab}$, derive the coefficient $g^2/(32\pi^2)$ for the number current of a single left-handed doublet. Why do the right-handed fields not contribute to the $W\tilde{W}$ term?

The right-hand side is a total derivative. As in QCD (Eq. 2.87), $W^a\tilde{W}^a = \partial_\mu K^\mu$ with the Chern-Simons current

$$K^\mu = 2\epsilon^{\mu\nu\rho\sigma}\left(W^a_\nu \partial_\rho W^a_\sigma + \frac{g}{3}\epsilon^{abc} W^a_\nu W^b_\rho W^c_\sigma\right). \tag{3.87}$$

Integrating the anomaly equation over space-time, and neglecting the abelian part which cannot change the topology, we find that the change of baryon number is related to the change of the **Chern-Simons number**

$$N_{CS}(t) = \frac{g^2}{32\pi^2}\int d^3x\, K^0(t,\vec{x}) = \frac{g^2}{16\pi^2}\int d^3x\, \epsilon^{ijk}\left(W^a_i \partial_j W^a_k + \frac{g}{3}\epsilon^{abc} W^a_i W^b_j W^c_k\right), \tag{3.88}$$

as

$$\Delta B = \Delta L = N_f\left[N_{CS}(t_f) - N_{CS}(t_i)\right]. \tag{3.89}$$

Exercise: *Verify that $W^a_{\mu\nu}\tilde{W}^{a\,\mu\nu} = \partial_\mu K^\mu$.*

The classical vacua of a non-abelian gauge theory are pure gauge configurations, $W_i = -\frac{2i}{g}(\partial_i U)U^{-1}$ in matrix notation, with $U(\vec{x}) \to 1$ at spatial infinity. The maps $U : S^3 \to SU(2)$ are classified by an integer winding number, exactly as the maps $S^2 \to S^2$ which classified the monopoles (Eq. 3.42) and as the large gauge transformations of QCD in Section 2.4 ($\pi_3(S^3) = \mathbb{Z}$), and N_{CS} is precisely this integer for a pure gauge. A transition between neighbouring vacua changes B and L by N_f units each, and $B + L$ by $2N_f$. For three generations, a single transition creates (or destroys) one lepton from every generation and three quarks, one of each colour, from every generation: twelve fermions in total. This was discovered by 't Hooft in 1976.

How often do such transitions happen? At zero temperature, they are tunnelling processes, mediated by the instantons of the $SU(2)$ theory (Section 2.6, $S_{\mathrm{inst}} = 8\pi^2/g^2$), and the rate is suppressed by the square of the tunnelling amplitude,

$$\Gamma \sim e^{-2S_{\mathrm{inst}}} = e^{-\frac{16\pi^2}{g^2}} = e^{-\frac{4\pi}{\alpha_W}} \sim 10^{-162}. \tag{3.90}$$

This is ridiculously small: not a single proton in the visible universe has decayed by this mechanism. So, is the anomalous violation of $B + L$ a mathematical curiosity? We have oversimplified an important point. We have treated the weak interactions as an unbroken gauge theory, like QCD, ignoring the Higgs field. The Higgs field carries no B or L, but it carries $SU(2)$ charge, and it changes the energy landscape between the vacua. We will now see that the barrier between them has a finite height, and that at high enough energy or temperature one can go **over** the barrier instead of through it.

3.7 Sphalerons

Let us consider the electroweak theory without the hypercharge interaction, which is a small perturbation ($\sin^2\theta_W \simeq 0.23$), i.e. an $SU(2)$ gauge theory with a Higgs doublet:

$$\mathcal{L} = -\frac{1}{4}W^a_{\mu\nu}W^{a\,\mu\nu} + \left(D_\mu\phi\right)^\dagger D^\mu\phi - V(\phi), \tag{3.91}$$

with

$$D_\mu = \partial_\mu - \frac{ig}{2}\tau^a W^a_\mu, \qquad V(\phi) = \lambda\left(\phi^\dagger\phi - \frac{v^2}{2}\right)^2. \tag{3.92}$$

In the usual treatment of the Higgs mechanism (Part III (QFT II) and Part IV of this script), we expand around the constant vacuum $\phi = \frac{v}{\sqrt{2}}(0,1)^T$, $W^a_\mu = 0$. Here we ask a different question: what do **all** the vacua of this theory look like, and what is the energy needed to go from one vacuum to another?

We choose the temporal gauge $W^a_0 = 0$. The Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}\dot{W}^a_i \dot{W}^a_i + \dot{\phi}^\dagger \dot{\phi} - \frac{1}{4} W^a_{ij} W^a_{ij} - \left(D_i\phi\right)^\dagger D_i\phi - V(\phi), \tag{3.93}$$

where we sum over repeated spatial indices with a Euclidean metric. The canonical momenta are $\pi^a_i = \dot{W}^a_i$, $\pi_\phi = \dot{\phi}^\dagger$, $\pi_{\phi^\dagger} = \dot{\phi}$, and the energy is

$$E = \int d^3x \left[\frac{1}{2}\dot{W}^a_i \dot{W}^a_i + \dot{\phi}^\dagger \dot{\phi} + \frac{1}{4} W^a_{ij} W^a_{ij} + \left(D_i\phi\right)^\dagger D_i\phi + V(\phi)\right]. \tag{3.94}$$

This is a sum of non-negative terms. The energy vanishes if and only if each term vanishes:

$$\dot{W}^a_i = 0, \qquad \dot{\phi} = 0, \qquad W^a_{ij} = 0, \qquad D_i\phi = 0, \qquad V(\phi) = 0. \tag{3.95}$$

These conditions are solved by

$$\hat{W}_i(\vec{x}) \equiv W^a_i \tau^a = -\frac{2i}{g}\left(\partial_i U(\vec{x})\right) U^{-1}(\vec{x}), \qquad \phi(\vec{x}) = \frac{v}{\sqrt{2}} U(\vec{x}) \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{3.96}$$

where $U(\vec{x})$ is an arbitrary $SU(2)$ matrix. Indeed, $W_{ij} = 0$ for a pure gauge, and

$$D_i\phi = \frac{v}{\sqrt{2}}\left[\partial_i U - \frac{ig}{2}\left(-\frac{2i}{g}\right)(\partial_i U)U^{-1}U\right]\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0.$$

Notice that the Higgs field must be rotated together with the gauge field: a non-trivial pure gauge with the Higgs field in the constant vacuum would cost gradient energy.

As in the pure gauge theory, we require $U(\vec{x}) \to 1$ (or any constant) at spatial infinity. The maps $U : S^3 \to SU(2)$ are classified by their winding number, and for the configurations of Eq. 3.96 the Chern-Simons number of Eq. 3.88 equals this integer. We have an infinite set of degenerate vacua, labelled by $N_{CS} \in \mathbb{Z}$, and neighbouring vacua differ by N_f units of B and of L. Between two such vacua there are configurations with non-integer N_{CS} and positive energy. The question is: what is the lowest energy barrier which we must overcome to go from $N_{CS} = 0$ to $N_{CS} = 1$?

Let us draw the energy as a function of the Chern-Simons number, minimising over all other directions in the space of field configurations. We obtain a periodic curve, with zeros at the integers (Fig. 3.5). The maximum of this curve, at $N_{CS} = 1/2$, is a saddle point of the energy functional: a static, unstable solution of the field equations. It is called the **sphaleron**, from the Greek "ready to fall". It is not a soliton; it has exactly one unstable direction, along which it falls into either of the two neighbouring vacua.

To find the sphaleron we use the field equations of the static theory,

$$\begin{aligned} D_i W^a_{ij} &= \frac{ig}{2}\left[\phi^\dagger \tau^a D_j\phi - \left(D_j\phi\right)^\dagger \tau^a \phi\right], \\ D_i D_i \phi &= 2\lambda\left(\phi^\dagger\phi - \frac{v^2}{2}\right)\phi, \end{aligned} \tag{3.97}$$

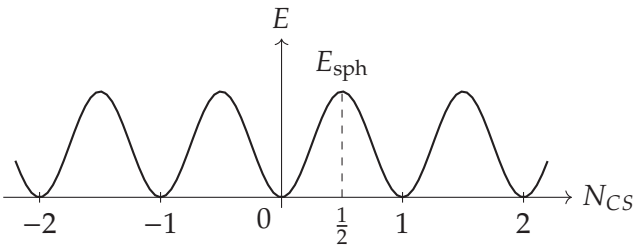


Figure 3.5. The static energy of the $SU(2)$-Higgs theory along the minimum-energy path as a function of the Chern-Simons number (schematic). Integer values are vacua; the sphaleron sits at the top of the barrier, $N_{CS} = 1/2$.

together with a clever ansatz due to Klinkhamer and Manton. We take the vacuum configuration of Eq. 3.96 with the "hedgehog" matrix

$$U^\infty(\vec{x}) = \frac{i\,\vec{\tau}\cdot\vec{x}}{r}, \qquad r = |\vec{x}|, \tag{3.98}$$

which has winding number one and identifies a direction in isospin space with the direction in space, and we switch it off towards the origin with two radial profile functions:

$$W_i^a = \frac{2}{g}\,f(r)\,\frac{\epsilon_{ija}x_j}{r^2}, \qquad \phi = \frac{v}{\sqrt{2}}\,h(r)\,\frac{i\,\vec{\tau}\cdot\vec{x}}{r}\begin{pmatrix}0\\1\end{pmatrix}, \tag{3.99}$$

with

$$f(0) = h(0) = 0, \qquad f(\infty) = h(\infty) = 1. \tag{3.100}$$

At large r the configuration is the pure gauge of U^∞, so the energy is finite; at the origin the fields are regular.

Exercise: *Show that* $-\frac{2i}{g}(\partial_i U^\infty)(U^\infty)^{-1} = \frac{2}{g}\frac{\epsilon_{ija}x_j}{r^2}\tau^a$, *so that* $f = 1$ *is a pure gauge.*

Inserting the ansatz into Eq. 3.94, the energy becomes a functional of f and h, and the field equations become two coupled ordinary differential equations, which are solved numerically. The profiles are shown in Fig. 3.6. The Chern-Simons number of the sphaleron is exactly 1/2, and its energy is

$$E_{\rm sph} = \frac{4\pi v}{g}\,B\!\left(\frac{\lambda}{g^2}\right) = \frac{8\pi m_W}{g^2}\,B\!\left(\frac{\lambda}{g^2}\right), \tag{3.101}$$

where the function B grows slowly from $B(0) = 1.56$ to $B(\infty) = 2.72$. The structure m_W/α_W is the same as for the monopole: it is the typical mass of a classical lump made of weakly coupled fields. With $\lambda/g^2 \simeq 0.3$ for the measured Higgs mass, $B \simeq 1.9$ and $E_{\rm sph} \simeq 9$ TeV.

Exercise: *Insert the ansatz of Eq. 3.99 into the energy and show that, with* $\xi = gvr$, *it takes the form* $E = \frac{4\pi v}{g}\int_0^\infty d\xi\left[4f'^2 + \frac{8}{\xi^2}f^2(1-f)^2 + \frac{\xi^2}{2}h'^2 + h^2(1-f)^2 + \frac{\lambda}{4g^2}\xi^2(h^2-1)^2\right]$.

How do the fermions know about all this? The fermion number changes through the **level crossing** phenomenon, which we illustrate in Fig. 3.7. Consider the eigenvalues of the Dirac Hamiltonian of a left-handed doublet in the background of a slowly varying gauge-Higgs configuration which interpolates from the vacuum $N_{CS} = 0$ to the vacuum $N_{CS} = 1$. The spectrum at the beginning and at the end is the same, since the two vacua are gauge equivalent. However, along the way one level crosses zero energy from below. If initially the Dirac sea was filled, at the end one level with positive energy is filled: one fermion was created out of the vacuum. This happens for every left-handed doublet,

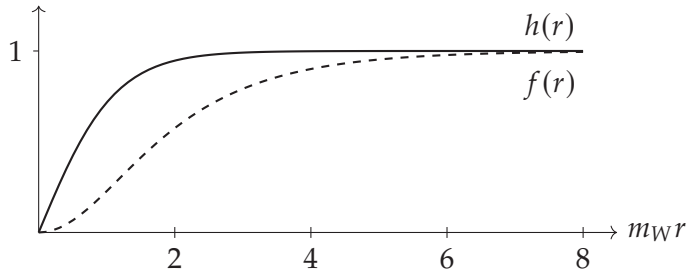


Figure 3.6. Qualitative shape of the sphaleron profile functions $f(r)$ and $h(r)$ of Eq. 3.99: both rise from zero at the origin to one at a distance of a few m_W^{-1} (schematic; the exact profiles are obtained numerically).

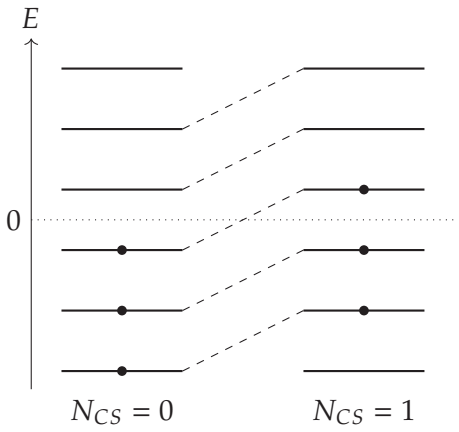


Figure 3.7. Level crossing: along a path from $N_{CS} = 0$ to $N_{CS} = 1$ every level of a left-handed doublet moves up by one unit. The filled Dirac sea (dots) ends with one occupied positive-energy level: one fermion has been created.

i.e. for three colours of quarks and one lepton in each generation, in agreement with the anomaly equation.

What is the physical significance of the sphaleron? It sets the height of the barrier between vacua with different $B + L$. At zero temperature, we can only tunnel through it, and we have found that the tunnelling rate is $\sim 10^{-162}$. In a hot plasma, however, the system can be thermally excited over the barrier. The rate per unit volume of such transitions is controlled by a Boltzmann factor,

$$\frac{\Gamma_{\rm sph}}{V} \propto T^4 \, e^{-E_{\rm sph}(T)/T}, \tag{3.102}$$

where $E_{\rm sph}(T)$ is the sphaleron energy computed with the temperature-dependent Higgs expectation value $v(T)$ of the previous section. As the temperature approaches the electroweak scale, $v(T)$ decreases and the suppression becomes weak. In the symmetric phase, $v(T) = 0$, there is no barrier at all and the rate is unsuppressed, $\Gamma_{\rm sph}/V \sim \alpha_W^5 T^4$. Comparing with the expansion rate of the universe, $B + L$ violating processes are in thermal equilibrium for temperatures between about 10^{12} GeV and the electroweak scale. Notice that this is the same physics as the instanton tunnelling, but now the barrier is crossed classically: the electroweak theory, which is weakly coupled, violates baryon number at a fast rate in the early universe!

Could sphaleron transitions be induced in a high-energy collision, with a centre-of-mass energy above $E_{\rm sph} \simeq 9$ TeV? The signature would be spectacular: not a resonance,

but a final state of many gauge and Higgs bosons together with three leptons (one per generation) and nine quarks, violating B and L by three units each.

Unfortunately, being above the barrier energy is not sufficient: a two-particle initial state has a very small overlap with the coherent, multi-particle sphaleron configuration, and the cross-section remains exponentially suppressed at LHC energies in all controlled calculations. Searches for multi-object final states by CMS have found nothing, and sphaleron processes are not expected to be observable at the LHC.

3.8 Baryogenesis

Let us now return to a question which we have raised at the beginning of the section on B and L violation: why is there matter but almost no antimatter in the universe? Antiprotons are found in cosmic rays only at the level expected from secondary production, and there is no sign of the annihilation radiation which would come from regions of antimatter in contact with matter. The asymmetry is conveniently quoted as the ratio of the baryon number density to the photon number density of the cosmic microwave background,

$$\eta = \frac{n_B - n_{\bar{B}}}{n_\gamma} \simeq 6 \times 10^{-10}. \tag{3.103}$$

It is measured independently from the abundances of light elements produced in Big Bang nucleosynthesis and from the acoustic peaks of the microwave background. In the very early universe, at temperatures above a GeV, quarks and antiquarks were as abundant as photons. The number above means that for every 10^9 antiquarks there were $10^9 + 1$ quarks. All the rest annihilated, and the tiny excess is the matter of which we and the galaxies are made. Could this excess be an initial condition? It could, but it would be destroyed by inflation, which dilutes any initial charge density. We would like to explain η dynamically.

Sakharov formulated in 1967 three necessary conditions for a theory to produce a baryon asymmetry from a symmetric initial state:

- interactions which violate baryon number;
- violation of C and of CP;
- departure from thermal equilibrium.

The first condition is obvious. The second is necessary because otherwise for every process producing an excess of baryons, the C or CP conjugate process would produce the same excess of antibaryons at the same rate. The third condition follows from CPT invariance. In thermal equilibrium the density matrix is $e^{-\beta H}/Z$, and since H commutes with $\Theta = CPT$ while B is odd under Θ,

$$\langle B \rangle_T = \frac{1}{Z} \mathrm{tr}\left(e^{-\beta H} B\right) = \frac{1}{Z} \mathrm{tr}\left(\Theta^{-1}\Theta\, e^{-\beta H} B\right) = \frac{1}{Z} \mathrm{tr}\left(e^{-\beta H} \Theta B \Theta^{-1}\right) = -\langle B \rangle_T\,, \tag{3.104}$$

so that $\langle B \rangle_T = 0$. In other words, in equilibrium the rates of the processes creating and destroying baryons are equal.

Does the Standard Model satisfy these conditions? Surprisingly, it has all the ingredients:

- C is maximally violated, since only left-handed fermions couple to the W;
- CP is violated by the phase of the CKM matrix;

- B is violated by the sphaleron transitions of the previous section;
- the universe expands, and the electroweak phase transition could provide the departure from equilibrium.

A sphaleron transition with $\Delta N_{CS} = \pm 1$ creates or destroys the combination

$$\prod_{i=1}^{N_f} (u_{Li}\, d_{Li}\, d_{Li}\, \nu_{Li}) \tag{3.105}$$

(with the colours of the three quarks antisymmetrised). Let us check its quantum numbers, with $Q = T_3 + Y/2$:

	u_L	d_L	d_L	ν_L	total
T_3	$+1/2$	$-1/2$	$-1/2$	$+1/2$	0
Y	$1/3$	$1/3$	$1/3$	-1	0
Q	$2/3$	$-1/3$	$-1/3$	0	0
B	$1/3$	$1/3$	$1/3$	0	1
L	0	0	0	1	1
$B-L$	$1/3$	$1/3$	$1/3$	-1	0

All gauge charges are conserved, B and L change by one unit per generation, and $B - L$ is conserved.

We can now draw an important conclusion. Sphaleron transitions are in thermal equilibrium above the electroweak scale. If the universe has $B - L = 0$ at high temperatures, any $B + L$ asymmetry is washed out by the sphalerons, since in equilibrium they drive $B + L$ to zero. There are then only two ways out:

- The asymmetry is generated at the electroweak phase transition, when the sphalerons go out of equilibrium: **electroweak baryogenesis**.
- A non-zero $B - L$ is generated at higher temperatures by some other mechanism. Sphalerons do not destroy it; instead, they redistribute it between baryons and leptons.

In equilibrium, the chemical potentials of all species are related by the reactions which are fast, and one finds that the final baryon asymmetry is a fixed fraction of $B - L$,

$$B = \frac{8N_f + 4N_H}{22N_f + 13N_H}\,(B - L) = \frac{28}{79}\,(B - L), \tag{3.106}$$

for $N_f = 3$ generations and $N_H = 1$ Higgs doublet.

Electroweak baryogenesis is an attractive idea, since all the ingredients are already in the Standard Model and can be tested at colliders. It needs a strongly first-order transition: bubbles of the broken phase expand into the symmetric phase, CP violating interactions of the fermions with the bubble walls create a chiral asymmetry in front of the walls, and sphalerons, fast in the symmetric phase, convert it into a baryon asymmetry. Once inside the bubble, the baryons must survive: the sphaleron rate in the broken phase, $\propto e^{-E_{\rm sph}(T_c)/T_c}$, must be slower than the expansion rate. Using Eq. 3.101 with $v \to \varphi_c$ this requires

$$\frac{\varphi_c}{T_c} \gtrsim 1. \tag{3.107}$$

This is the strength of the transition which we computed in Eq. 3.83! In the Standard Model it fails twice. First, $\varphi_c/T_c \gtrsim 1$ requires a light Higgs boson. Second, CP violation through the CKM matrix requires all three generations, and it is proportional to the Jarlskog invariant times the differences of the quark masses squared; at $T \sim 100$ GeV this gives an effect of order 10^{-20}, many orders of magnitude below η. Electroweak baryogenesis needs new physics: additional scalars and new sources of CP violation.

The most precise value of the asymmetry comes from the Planck 2018 measurement of the baryon density, $\Omega_b h^2 = 0.0224 \pm 0.0001$, which gives $\eta = (6.1 \pm 0.04) \times 10^{-10}$, in agreement with Big Bang nucleosynthesis. With $m_h = 125$ GeV, the electroweak transition of the Standard Model is a crossover (see the end of Section 3.5.3), so Standard Model electroweak baryogenesis is excluded. Electroweak baryogenesis in extensions of the Standard Model is constrained by electric dipole moment searches (the ACME bound on the electron EDM, $|d_e| < 1.1 \times 10^{-29}\, e$ cm) and by the Higgs measurements at the LHC.

3.8.1 Leptogenesis

The second possibility is that a lepton asymmetry was produced first, with $B = 0$, $L \neq 0$, so that $B - L = -L \neq 0$. Sphalerons then convert a fraction of it into a baryon asymmetry, Eq. 3.106. This is **leptogenesis**, proposed by Fukugita and Yanagida. It is attractive because the lepton sector is much less constrained than the quark sector, and because it is linked to the masses of the neutrinos.

Let us add to the Standard Model heavy right-handed neutrinos N_i, which are singlets under all gauge interactions and can therefore have Majorana masses M_i. They couple to the lepton doublets and the Higgs field through Yukawa couplings $h_{\alpha i}$,

$$\mathcal{L}_N = \frac{1}{2}\bar{N}_i \left(i\partial\!\!\!/ - M_i\right) N_i - \left(h_{\alpha i}\, \bar{\ell}_{L\alpha}\, \tilde{\phi}\, N_i + \text{h.c.}\right), \qquad \tilde{\phi} = i\tau^2 \phi^*. \tag{3.108}$$

A Majorana particle is its own antiparticle, $N = N^c$. It can therefore decay both into a lepton and a Higgs boson and into an antilepton and an anti-Higgs:

$$\Gamma_i = \Gamma\left(N_i \to \ell\, \phi\right) + \Gamma\left(N_i \to \bar{\ell}\, \phi^\dagger\right) = \frac{\left(h^\dagger h\right)_{ii}}{8\pi} M_i \tag{3.109}$$

at tree level. Lepton number is violated by the coexistence of the Majorana mass and of the Yukawa coupling. Now we check Sakharov's conditions:

- L is violated, as we have just seen, and C is violated since only ℓ_L couples to N.
- CP is violated if the Yukawa matrix has complex phases which cannot be removed by field redefinitions. The interference of the tree-level decay with one-loop diagrams (vertex and self-energy corrections with another N_j in the loop) gives a CP asymmetry

  $$\varepsilon_1 = \frac{\Gamma(N_1 \to \ell\phi) - \Gamma(N_1 \to \bar{\ell}\phi^\dagger)}{\Gamma(N_1 \to \ell\phi) + \Gamma(N_1 \to \bar{\ell}\phi^\dagger)} \simeq -\frac{3}{16\pi}\frac{1}{\left(h^\dagger h\right)_{11}} \sum_{j\neq 1} \text{Im}\left[\left(h^\dagger h\right)^2_{1j}\right] \frac{M_1}{M_j}, \tag{3.110}$$

  for $M_1 \ll M_j$, which is non-zero only if $\text{Im}\, h \neq 0$.
- The decays are out of equilibrium if the decay rate is smaller than the expansion rate of the universe at $T \simeq M_1$, $\Gamma_1 \lesssim H(T = M_1)$. The heavy neutrinos then decay when their equilibrium density is already Boltzmann suppressed, and the inverse decays cannot erase the asymmetry.

The lepton asymmetry is then of order ε_1 times the abundance of N_1 relative to photons, reduced by an efficiency factor $\kappa < 1$ which accounts for washout, and the baryon asymmetry follows from Eq. 3.106. Roughly, $\eta \sim 10^{-2}\,\varepsilon_1\kappa$, so we need $\varepsilon_1 \gtrsim 10^{-7}$.

The same Yukawa couplings and Majorana masses give, after electroweak symmetry breaking, masses to the light neutrinos through the **seesaw mechanism**: $m_\nu \simeq -m_D M^{-1} m_D^T$ with $m_D = h\,v/\sqrt{2}$. Light neutrino masses of order 0.05 eV with Yukawa couplings of order one point to $M \sim 10^{14}$ GeV. With the light neutrino masses one can bound the CP asymmetry, $|\varepsilon_1| \lesssim \frac{3}{8\pi}\frac{M_1 m_3}{v^2}$, and successful leptogenesis requires $M_1 \gtrsim 10^9$ GeV. In this way, the smallness of neutrino masses and the matter-antimatter asymmetry of the universe may have a common origin.

Exercise: *Show that the condition* $\Gamma_1 < H(T = M_1)$*, with* $H = 1.66\sqrt{g_*}\,T^2/M_{\rm Pl}$*, can be written as a condition on the effective neutrino mass* $\tilde{m}_1 = (h^\dagger h)_{11} v^2/(2M_1) \lesssim 10^{-3}$ *eV, independently of* M_1 *(use* $g_* \simeq 107$*).*

What do we know about these ingredients experimentally? Neutrino oscillations have established that at least two light neutrinos are massive, with $\Delta m^2_{\rm atm} \simeq 2.5 \times 10^{-3}\,{\rm eV}^2$, i.e. $m_3 \gtrsim 0.05$ eV. Leptonic CP violation (the Dirac phase of the PMNS matrix) is hinted at by T2K and NOvA but not established; in any case it is not directly the phase entering ε_1. The Majorana nature of neutrinos is being tested by neutrinoless double beta decay searches, with no signal so far.

4

(Quantum) fields in curved spacetime

In the previous chapter we found that particle physics leaves fingerprints in the early universe: topological defects formed at phase transitions, and a baryon asymmetry generated by B-violating interactions. In this chapter we look at the problem from the other side. The standard cosmological model, a hot universe expanding according to Einstein's equations, describes very successfully the abundances of light elements and the cosmic microwave background. It leaves, however, a number of puzzles unexplained:

- **Monopoles and other defects.** We estimated in Section 3.5.4 that grand unified monopoles would be produced in such numbers that they would dominate the energy density of the universe by many orders of magnitude. None has been observed.
- **Flatness.** The spatial curvature of the universe today is compatible with zero. We will see that a flat universe is an unstable solution of the Friedmann equations: any initial deviation from flatness grows with time. Why was the universe so incredibly flat at early times?
- **Homogeneity and the horizon.** The temperature of the cosmic microwave background is the same in all directions, up to fluctuations $\delta T/T \sim 10^{-5}$. In the standard hot Big Bang, however, regions of the sky separated by more than about one degree have never been in causal contact. How could they agree on their temperature?

All three puzzles are solved if the universe went through an early epoch of exponential expansion, driven by the energy density of a scalar field which behaves like a vacuum energy: **inflation**. A tiny causally connected region is blown up to a size larger than the observable universe, the curvature is diluted, and so are the monopoles. In this chapter we will first review the evolution of a homogeneous universe. Then we will study a scalar field theory in an expanding universe, first classically and then quantum mechanically. We will find a remarkable byproduct: quantum fluctuations of the scalar field are stretched to cosmological size and provide the seeds of the density fluctuations which we observe in the microwave background and in the distribution of galaxies. A good textbook for this chapter is Bergstrom and Goobar, "Cosmology and Particle Astrophysics"; for the gravitational background see Weinberg, "Gravitation and Cosmology".

4.1 Evolution of the universe

On large scales, the universe is homogeneous and isotropic. The most general metric with these properties is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric,

$$ds^2 = dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta\, d\varphi^2\right], \tag{4.1}$$

where $k = +1, 0, -1$ for a closed, flat or open universe, and $a(t)$ is the **scale factor**. Physical distances between objects at fixed comoving coordinates grow proportionally to $a(t)$. The rate of expansion is the **Hubble parameter** $H(t) = \dot{a}/a$. The matter content compatible with the symmetries is a perfect fluid with energy density $\rho(t)$ and pressure $p(t)$,

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - p\, g_{\mu\nu}, \tag{4.2}$$

where $u^\mu = dx^\mu/d\tau$ is the four-velocity of the fluid, $u^\mu = (1, \vec{0})$ in comoving coordinates. Inserting the metric into the Einstein equations,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G\, T_{\mu\nu}, \tag{4.3}$$

where we have included a possible cosmological constant in $T_{\mu\nu}$, we find two independent equations, the **Friedmann equations**:

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho, \tag{4.4}$$

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = -8\pi G\, p. \tag{4.5}$$

The first comes from the 00 component, the second from the ii components of Einstein's equations.

Exercise: *Compute the Christoffel symbols and the Ricci tensor of the FLRW metric for $k = 0$ in Cartesian comoving coordinates, $ds^2 = dt^2 - a^2 d\vec{x}^2$, and derive the Friedmann equations.*

Instead of the second Friedmann equation it is often more convenient to use energy conservation. A comoving volume $V \propto a^3$ contains the energy $E = \rho a^3$, and it expands adiabatically, $dE = -p\, dV$:

$$\frac{d}{dt}\left(\rho a^3\right) = -p\,\frac{d}{dt}a^3 \qquad \rightsquigarrow \qquad \dot{\rho} + 3H\left(\rho + p\right) = 0. \tag{4.6}$$

This is the time component of $\nabla_\mu T^{\mu\nu} = 0$, and it follows from the two Friedmann equations (**exercise**). We need an equation of state. Three cases are important:

- **Vacuum energy** (a cosmological constant), $T_{\mu\nu} = \rho_\Lambda g_{\mu\nu}$, i.e. $p = -\rho$. Then $\rho =$ constant: the energy density of the vacuum does not dilute.
- **Non-relativistic matter** ("dust"), $p = 0$. Then $\rho \propto a^{-3}$: the number of particles in a comoving volume is constant.
- **Radiation**, $p = \rho/3$ (the energy-momentum tensor is traceless). Then $\rho \propto a^{-4}$: the number density dilutes as a^{-3}, and in addition every quantum is redshifted, $E \propto a^{-1}$.

Let us now solve the first Friedmann equation for a flat universe, $k = 0$, dominated by one of these components. For the vacuum,

$$H^2 = \frac{8\pi G}{3}\rho_\Lambda \equiv \frac{\Lambda}{3} = \text{constant} \qquad \rightsquigarrow \qquad a(t) = a_0\, e^{Ht}. \tag{4.7}$$

This is the **de Sitter** universe: it expands exponentially. For radiation and matter we write $\rho = \rho_0 (a_0/a)^n$ with $n = 4$ or 3, and we find $\dot{a} \propto a^{1-n/2}$, i.e.

$$\begin{aligned} &\text{radiation:} \quad a(t) \propto t^{1/2}, \quad H = \frac{1}{2t}, \quad \rho = \frac{3}{32\pi G\, t^2}, \\ &\text{matter:} \quad a(t) \propto t^{2/3}, \quad H = \frac{2}{3t}, \quad \rho = \frac{1}{6\pi G\, t^2}. \end{aligned} \tag{4.8}$$

In both cases the energy density is of order $1/(Gt^2)$. In a radiation-dominated universe with g_* relativistic degrees of freedom, $\rho = g_* \frac{\pi^2}{30} T^4$, and therefore $t \sim M_{\rm Pl}/T^2$, with $G = 1/M_{\rm Pl}^2$, $M_{\rm Pl} = 1.22 \times 10^{19}$ GeV.
Exercise: *Show that in a radiation-dominated universe* $t = 0.30\, g_*^{-1/2}\, M_{\rm Pl}/T^2$ *and estimate the age of the universe at the electroweak phase transition,* $T \simeq 160$ *GeV, with* $g_* \simeq 107$.

4.1.1 The flatness problem

Let us now keep the curvature term. It is customary to define the **critical density** $\rho_c = 3H^2/(8\pi G)$, i.e. the density of a flat universe with the same expansion rate, and $\Omega = \rho/\rho_c$. The first Friedmann equation becomes

$$\Omega - 1 = \frac{k}{a^2 H^2} = \frac{k}{\dot{a}^2}. \tag{4.9}$$

A flat universe has $\Omega = 1$ at all times. What happens if Ω deviates slightly from one? In a universe dominated by radiation or matter, $\dot{a}$ decreases with time, and we find:

$$|\Omega - 1| \propto \begin{cases} t & \text{(radiation)}, \\ t^{2/3} & \text{(matter)}, \\ e^{-2Ht} & \text{(vacuum)}. \end{cases} \tag{4.10}$$

In the standard hot Big Bang any deviation from flatness grows. Today, measurements of the microwave background together with the large-scale distribution of galaxies give $|\Omega_0 - 1| \lesssim 10^{-2}$ (more precisely, the curvature contribution is $\Omega_k = 0.001 \pm 0.002$). Running the evolution backwards, the universe was radiation dominated since a redshift of about 3400, and $|\Omega - 1| \propto t \propto a^2 \propto T^{-2}$. At the Planck time, $T \sim M_{\rm Pl}$, we need $|\Omega - 1| \lesssim 10^{-60}$. Such a fine-tuning of the initial conditions calls for an explanation. Notice that a vacuum dominated epoch does exactly the opposite: it drives Ω towards one exponentially fast. About 60 e-folds of exponential expansion, $e^{-2\cdot 60} \simeq 10^{-52}$, are enough to make the universe as flat as we observe it, whatever the curvature before inflation.
Exercise: *Verify the estimate* $|\Omega - 1| \lesssim 10^{-60}$ *at* $T \sim M_{\rm Pl}$, *using a matter dominated epoch from* $T_{\rm eq} \simeq 0.8$ *eV until today and a radiation dominated epoch before.*

4.1.2 The horizon problem

Light travels on $ds^2 = 0$, i.e. in comoving coordinates $dr = dt/a(t)$ (for $k = 0$). The comoving distance which a light signal can travel from the Big Bang to the time t is

$$r_H(t) = \int_0^t \frac{dt'}{a(t')} = \int_0^a \frac{da'}{a'^2 H(a')}. \tag{4.11}$$

For radiation or matter domination $a^2 H \propto a^{1-n/2}$ grows with a, so the integral is dominated by late times: $r_H \propto 1/(aH)$, the comoving Hubble radius, which grows with time. Regions which come into view today have never been in causal contact before. The comoving horizon at the time of the emission of the microwave background, seen from today, subtends an angle of about one degree on the sky. Yet the microwave background has the same temperature in all directions to one part in 10^5. During inflation, on the other hand, $1/(aH)$ **decreases**: the whole observable universe was once inside a single causally connected patch, which was then blown up by the exponential expansion.

4.2 Inflation

We have seen that a universe dominated by vacuum energy expands exponentially. A true cosmological constant, however, would never stop dominating. We need a vacuum energy which is large for some time and then disappears, converting itself into the hot radiation of the standard Big Bang. A scalar field slowly rolling down its potential does exactly this. We will first discuss the dynamics of such a field classically, and then quantum mechanically in the next section.

Let us consider a real scalar field, the **inflaton**, minimally coupled to gravity:

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi - V(\phi)\right]. \tag{4.12}$$

In the **chaotic inflation** scenario of Linde, we imagine that at the Planck time, $t_{\rm Pl} \simeq 5.4\times10^{-44}$ s, some region of size $l_{\rm Pl}\simeq 1.6\times10^{-35}$ m happens to have an approximately homogeneous scalar field, i.e. $(\vec\nabla\phi)^2 \ll V(\phi)$. In such a region we can use the FLRW metric with $k=0$ (we will see that curvature becomes irrelevant anyway), $\sqrt{-g}=a^3$, and $\phi=\phi(t)$. The action, per unit comoving volume, becomes

$$S = \int dt\; a^3(t)\left[\frac{1}{2}\dot\phi^2 - V(\phi)\right]. \tag{4.13}$$

The Euler-Lagrange equation gives

$$\frac{d}{dt}\left(a^3\dot\phi\right) + a^3V'(\phi) = 0 \qquad \rightsquigarrow \qquad \ddot\phi + 3H\dot\phi + V'(\phi) = 0. \tag{4.14}$$

This is the motion of a ball in the potential V, with a friction term $3H\dot\phi$ due to the expansion. The energy density and the pressure of the homogeneous field are read off from its energy-momentum tensor,

$$\rho = \frac{1}{2}\dot\phi^2 + V(\phi), \qquad p = \frac{1}{2}\dot\phi^2 - V(\phi), \tag{4.15}$$

and the Friedmann equation reads

$$H^2 = \frac{8\pi}{3M_{\rm Pl}^2}\left[\frac{1}{2}\dot\phi^2 + V(\phi)\right]. \tag{4.16}$$

Notice that if the potential energy dominates over the kinetic energy, $p\simeq-\rho$: the field behaves like a vacuum energy. This is the **slow-roll** regime, $\dot\phi^2 \ll V$ and $|\ddot\phi| \ll 3H|\dot\phi|$, in which the equations simplify to

$$3H\dot\phi \simeq -V'(\phi), \qquad H^2 \simeq \frac{8\pi}{3M_{\rm Pl}^2}V(\phi). \tag{4.17}$$

When is the slow-roll regime consistent? Inserting $\dot\phi\simeq -V'/(3H)$ into $\dot\phi^2\ll V$, we find that the dimensionless **slow-roll parameters**

$$\epsilon_V = \frac{M_{\rm Pl}^2}{16\pi}\left(\frac{V'}{V}\right)^2, \qquad \eta_V = \frac{M_{\rm Pl}^2}{8\pi}\frac{V''}{V} \tag{4.18}$$

must be small; the second condition comes from $|\ddot\phi|\ll 3H|\dot\phi|$ (**exercise**). For a polynomial potential $V'/V\sim 1/\phi$, and slow roll requires

$$\phi \gg M_{\rm Pl}. \tag{4.19}$$

The field must be larger than the Planck mass. Is this a problem? Not necessarily: what matters for quantum gravity is the energy density, not the value of the field. For $V = \lambda\phi^4$ we have $V(\phi \sim M_{\rm Pl}) \sim \lambda M_{\rm Pl}^4 \ll M_{\rm Pl}^4$ if $\lambda \ll 1$. If we start at the Planck density, $V(\phi_{\rm in}) \sim M_{\rm Pl}^4$, then $\phi_{\rm in} \sim \lambda^{-1/4} M_{\rm Pl} \gg M_{\rm Pl}$, and there is a lot of room for slow roll.

During slow roll, the universe expands quasi-exponentially. The number of e-folds between a field value ϕ and the end of inflation at $\phi_{\rm end}$ (where $\epsilon_V \simeq 1$) is

$$N(\phi) = \int_t^{t_{\rm end}} H\, dt' = \int_{\phi_{\rm end}}^{\phi} \frac{H}{|\dot\phi|}\, d\phi' \simeq \frac{8\pi}{M_{\rm Pl}^2} \int_{\phi_{\rm end}}^{\phi} \frac{V}{V'}\, d\phi'. \tag{4.20}$$

For $V = \lambda\phi^4$ we find $N \simeq \pi\phi^2/M_{\rm Pl}^2$, and the 60 e-folds which we need to solve the flatness and horizon problems are obtained if inflation starts at $\phi \gtrsim 4.4\, M_{\rm Pl}$.

When ϕ approaches $M_{\rm Pl}$ the slow roll ends, and the field starts to oscillate around the minimum of its potential, with an amplitude damped by the friction term. Through its couplings to other fields the inflaton decays into the particles of the Standard Model (or of its extensions), which thermalise: this is the **reheating** of the universe, after which the standard hot Big Bang begins. The temperature after reheating sets the initial conditions for everything we discussed in the previous chapter. Notice that monopoles produced before or during inflation are diluted by a factor e^{-3N}, and the reheating temperature must stay below the grand unification scale, so that they are not produced again.

Exercise: *For $V \propto \phi^p$ show that, N e-folds before the end of inflation, $\epsilon_V = \frac{p}{4N}$ and $\eta_V = \frac{p-1}{2N}$ (neglecting $\phi_{\rm end}$). The spectral index and the tensor-to-scalar ratio of the primordial fluctuations are, to first order in slow roll, $n_s = 1 - 6\epsilon_V + 2\eta_V$ and $r = 16\epsilon_V$. Evaluate them for $p = 2$ and $p = 4$ with $N = 60$.*

How do these models compare with observation? The Planck 2018 measurements of the microwave background give $n_s = 0.9649 \pm 0.0042$ and, together with BICEP/Keck, $r < 0.036$ (95% C.L.). The simple monomial potentials $\lambda\phi^4$ ($n_s \simeq 0.95$, $r \simeq 0.27$) and $m^2\phi^2$ ($n_s \simeq 0.967$, $r \simeq 0.13$) are therefore excluded by the bound on r. Potentials with a plateau, such as Starobinsky inflation or Higgs inflation with a non-minimal coupling to gravity, predict $n_s \simeq 1 - 2/N$ and $r \simeq 12/N^2 \simeq 0.003$, in agreement with the data. The physics which we have learned from the chaotic-inflation example (slow roll, e-folds, reheating) is the same in these models.

4.3 Quantum theory of the inflaton

We now treat the inflaton as a quantum field. We will keep the spacetime fixed, as a classical background determined by the vacuum energy, and ignore the back-reaction of the quantum fluctuations on the metric. Let us consider the simplest case, a free massive scalar field,

$$\mathcal{L} = \frac{1}{2}\sqrt{-g}\left[g^{\mu\nu}(x)\,\partial_\mu\phi\,\partial_\nu\phi - m^2\phi^2\right], \tag{4.21}$$

in the flat FLRW metric $g_{\mu\nu} = \mathrm{diag}(1, -a^2, -a^2, -a^2)$, $\sqrt{-g} = a^3$. The Euler-Lagrange equation gives the Klein-Gordon equation in curved spacetime,

$$\left(\Box_g + m^2\right)\phi = 0, \qquad \Box_g\phi = \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}\, g^{\mu\nu}\partial_\nu\phi\right), \tag{4.22}$$

which in the FLRW metric becomes

$$\ddot\phi + 3H\dot\phi - \frac{\vec\nabla^2\phi}{a^2} + m^2\phi = 0. \tag{4.23}$$

For a constant field, this is the equation of motion of the previous section with $V = m^2\phi^2/2$. How do we quantise this field? In flat spacetime we expanded the field in plane waves $e^{-i\omega t}$, which are solutions of the field equation with definite frequency. Here the coefficients of the equation depend on time, and there are no such solutions. Still, the canonical procedure goes through.

We consider a de Sitter background, $a(t) = a_0 e^{H(t-t_0)}$ with H constant, which is a good approximation during slow-roll inflation. It is convenient to introduce the **conformal time**

$$\eta = \int \frac{dt}{a(t)} = -\frac{1}{a(t)H}, \qquad -\infty < \eta < 0, \tag{4.24}$$

where we have chosen the integration constant such that $\eta \to 0^-$ at late times. In conformal time the metric is proportional to the Minkowski metric,

$$ds^2 = a^2(\eta)\left[d\eta^2 - d\vec{x}^2\right], \qquad a(\eta) = -\frac{1}{H\eta}, \tag{4.25}$$

and the action becomes

$$S = \int d\eta\, d^3x \left[\frac{a^2}{2}\left(\phi'^2 - (\vec{\nabla}\phi)^2\right) - \frac{a^4}{2}m^2\phi^2\right], \qquad \phi' \equiv \frac{\partial\phi}{\partial\eta}. \tag{4.26}$$

The canonical momentum is $\pi = a^2\phi'$, and we impose the canonical commutation relation at equal conformal time,

$$\left[\phi(\eta,\vec{x}), \pi(\eta,\vec{y})\right] = a^2(\eta)\left[\phi(\eta,\vec{x}), \phi'(\eta,\vec{y})\right] = i\,\delta^{(3)}(\vec{x}-\vec{y}). \tag{4.27}$$

Space is still homogeneous, so we can expand in plane waves in $\vec{x}$, with time-dependent mode functions:

$$\phi(\eta,\vec{x}) = \int \frac{d^3k}{(2\pi)^{3/2}}\left[a_{\vec{k}}\,\phi_k(\eta)\,e^{i\vec{k}\cdot\vec{x}} + a^\dagger_{\vec{k}}\,\phi^*_k(\eta)\,e^{-i\vec{k}\cdot\vec{x}}\right], \qquad \left[a_{\vec{k}}, a^\dagger_{\vec{q}}\right] = \delta^{(3)}(\vec{k}-\vec{q}). \tag{4.28}$$

Inserting into the commutation relation we find that the mode functions must satisfy the Wronskian condition

$$a^2(\eta)\left[\phi_k\,\phi^{*\prime}_k - \phi'_k\,\phi^*_k\right] = i. \tag{4.29}$$

Exercise: *Verify Eq. 4.29. Show that the field equation guarantees that the left-hand side is independent of η.*

The extra factor a^2 suggests a rescaling, $\phi_k(\eta) = \chi_k(\eta)/a(\eta)$. A short computation gives

$$\chi''_k + \left[k^2 + m^2a^2 - \frac{a''}{a}\right]\chi_k = 0, \qquad \chi_k\chi^{*\prime}_k - \chi'_k\chi^*_k = i. \tag{4.30}$$

This is the equation of a harmonic oscillator with a time-dependent frequency. In de Sitter space, $a''/a = 2/\eta^2$ and $m^2a^2 = m^2/(H^2\eta^2)$, so that

$$\chi''_k + \left[k^2 - \frac{\nu^2 - 1/4}{\eta^2}\right]\chi_k = 0, \qquad \nu^2 = \frac{9}{4} - \frac{m^2}{H^2}. \tag{4.31}$$

This is Bessel's equation. The general solution is

$$\chi_k(\eta) = \sqrt{-\eta}\left[c_1\,H^{(1)}_\nu(-k\eta) + c_2\,H^{(2)}_\nu(-k\eta)\right], \tag{4.32}$$

with the Hankel functions $H_\nu^{(1)}$ and $H_\nu^{(2)} = (H_\nu^{(1)})^*$ (for real ν).

Which solution should we choose? At early times, $k|\eta| \to \infty$, the physical wavelength of the mode, a/k, is much shorter than the Hubble radius $1/H$: the mode does not feel the expansion. Indeed, the term k^2 dominates in Eq. 4.30, and with the asymptotic form of the Hankel functions we find

$$\chi_k \to \sqrt{\frac{2}{\pi k}} \left[c_1 \, e^{-ik\eta - i\frac{\pi}{2}(\nu+\frac{1}{2})} + c_2 \, e^{+ik\eta + i\frac{\pi}{2}(\nu+\frac{1}{2})} \right]. \tag{4.33}$$

We choose the solution which at early times is a positive frequency plane wave, $\chi_k \to e^{-ik\eta}/\sqrt{2k}$, exactly as in Minkowski space. This requires $c_2 = 0$ and $|c_1| = \sqrt{\pi}/2$, and the Wronskian condition is then satisfied. The resulting state, annihilated by all $a_{\vec{k}}$, is called the **Bunch-Davies vacuum**. The inflaton field operator reads:

$$\phi(\eta, \vec{x}) = \frac{\sqrt{-\pi\eta}}{2\,a(\eta)} \int \frac{d^3k}{(2\pi)^{3/2}} \left[a_{\vec{k}} \, H_\nu^{(1)}(-k\eta) \, e^{i\vec{k}\cdot\vec{x}} + a_{\vec{k}}^\dagger \, H_\nu^{(2)}(-k\eta) \, e^{-i\vec{k}\cdot\vec{x}} \right], \tag{4.34}$$

up to an irrelevant constant phase.

What happens at late times? Let us take a light field, $m \ll H$, i.e. $\nu \simeq 3/2$. For $\nu = 3/2$ the Hankel function is elementary, and the normalised mode function is

$$\chi_k(\eta) = \frac{e^{-ik\eta}}{\sqrt{2k}} \left(1 - \frac{i}{k\eta} \right). \tag{4.35}$$

Exercise: *Verify that this function solves Eq. 4.30 for $m = 0$ and satisfies the Wronskian condition.*

As $\eta \to 0^-$, the physical wavelength of the mode grows beyond the Hubble radius (the mode "exits the horizon"), and

$$|\phi_k|^2 = \frac{|\chi_k|^2}{a^2} = \frac{H^2\eta^2}{2k} \left(1 + \frac{1}{k^2\eta^2} \right) \to \frac{H^2}{2k^3}. \tag{4.36}$$

The mode stops oscillating and freezes with a constant amplitude. The variance of the field in the Bunch-Davies vacuum is then

$$\langle 0 | \, \phi^2(\eta, \vec{x}) \, | 0 \rangle = \int \frac{d^3k}{(2\pi)^3} |\phi_k|^2 = \int \frac{dk}{k} \left(\frac{H}{2\pi} \right)^2, \tag{4.37}$$

i.e. every logarithmic interval of wavelengths carries fluctuations of size $\delta\phi \simeq H/(2\pi)$. This is a genuinely new phenomenon of quantum fields in curved spacetime. The vacuum at early times is not the vacuum at late times: the expansion of the universe amplifies the zero-point fluctuations of the field and converts them into classical-looking, frozen field configurations on super-horizon scales. In a different language, quanta of the field are created by the time-dependent gravitational background.

These frozen fluctuations make inflation end at slightly different times at different places, $\delta t \simeq \delta\phi/\dot{\phi}$, and therefore leave behind small fluctuations of the energy density. Their size is

$$\frac{\delta\rho}{\rho} \sim H \, \delta t \simeq \frac{H^2}{2\pi\dot{\phi}}, \tag{4.38}$$

evaluated when the scale of interest exits the horizon. Since H and $\dot{\phi}$ change slowly during inflation, the spectrum is almost the same on all scales. This is the origin of the fluctuations of the microwave background and of the seeds of the galaxies. We have not included

the feedback of the field on the background metric; a consistent treatment requires the quantisation of the coupled scalar-metric perturbations, which goes beyond the scope of these lectures.

These predictions have been confirmed spectacularly. The amplitude of the primordial curvature fluctuations is measured to be $\Delta^2_{\mathcal{R}} = \frac{1}{8\pi^2\epsilon_V}\frac{H^2}{M^2_{\rm Pl}/(8\pi)} \simeq 2.1\times10^{-9}$ (Planck 2018), with a slight red tilt $n_s = 0.9649 \pm 0.0042$, as predicted by slow-roll inflation. Primordial gravitational waves, i.e. the same frozen vacuum fluctuations of the two polarisations of the graviton with amplitude $\propto H/M_{\rm Pl}$, have not been detected; the bound $r < 0.036$ implies $V^{1/4} \lesssim 1.4\times10^{16}$ GeV during inflation.

A

Special relativity

Special relativity is based on the assumption that the laws of nature are the same for inertial observers whose coordinates are related via Lorentz transformations,

$$x^\mu \longrightarrow x^{\mu\prime} = \Lambda^\mu_{\ \nu} x^\nu + \rho^\mu, \tag{A.1}$$

where

$$x^\mu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{A.2}$$

is a "four-vector" comprising space-time coordinates with $x^0 = ct$, ρ^μ is a constant four-vector and $\Lambda^\mu_{\ \nu}$ satisfies:

$$\Lambda^\mu_{\ \rho} \Lambda^\nu_{\ \sigma} g_{\mu\nu} = g_{\rho\sigma}. \tag{A.3}$$

The 4×4 matrix $g_{\mu\nu}$ is the so-called *metric*, defined as:

$$g_{\mu\nu} = \begin{cases} +1, & \mu = \nu = 0, \\ -1, & \mu = \nu = 1, 2, 3, \\ 0, & \mu \neq \nu. \end{cases} \tag{A.4}$$

In the above we have used Einstein's summation convention. For example, one would write explicitly

$$\Lambda^\mu_{\ \nu} x^\nu = \Lambda^\mu_{\ 0} x^0 + \Lambda^\mu_{\ 1} x^1 + \Lambda^\mu_{\ 2} x^2 + \Lambda^\mu_{\ 3} x^3. \tag{A.5}$$

This is a convention that we will use extensively from now on.

A.1 Proper time

Lorentz transformations leave "proper-time" intervals invariant. These are defined as

$$d\tau^2 \equiv c^2 dt^2 - d\vec{x}^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{A.6}$$

Indeed, in a different reference frame we have from Eq. A.1:

$$dx^{\mu\prime} = \Lambda^\mu_{\ \nu} dx^\nu. \tag{A.7}$$

A proper-time interval in the new frame is

$$\begin{aligned} d\tau'^2 &= g_{\mu\nu} dx'^\mu dx'^\nu \\ &= g_{\mu\nu} \left(\Lambda^\mu_{\ \rho} dx^\rho \right) \left(\Lambda^\nu_{\ \sigma} dx^\sigma \right) \\ &= \left(g_{\mu\nu} \Lambda^\mu_{\ \rho} \Lambda^\nu_{\ \sigma} \right) dx^\rho dx^\sigma \\ &= g_{\rho\sigma} dx^\rho dx^\sigma = d\tau^2. \end{aligned} \tag{A.8}$$

As a consequence of the invariance of proper-time intervals, *the speed of light is the same in all inertial frames*. Indeed, for light we have

$$\left|\frac{d\vec{x}}{dt}\right| = c \rightsquigarrow d\tau^2 = c^2dt^2 - d\vec{x}^2 = 0. \tag{A.9}$$

In a new frame,

$$d\tau^{2\prime} = d\tau^2 = 0 \rightsquigarrow \left|\frac{d\vec{x}'}{dt'}\right| = c. \tag{A.10}$$

Lorentz transformations are the only non-singular transformations which preserve proper-time intervals:

$$\begin{aligned} d\tau^2 &= d\tau'^2 \\ \rightsquigarrow g_{\rho\sigma}dx^\rho dx^\sigma &= g_{\mu\nu}dx'^\mu dx'^\nu \\ \rightsquigarrow g_{\rho\sigma}dx^\rho dx^\sigma &= g_{\mu\nu}\frac{\partial x'^\mu}{\partial x^\rho}\frac{\partial x'^\nu}{\partial x^\sigma}dx^\rho dx^\sigma, \end{aligned} \tag{A.11}$$

from which we conclude that

$$g_{\rho\sigma} = g_{\mu\nu}\frac{\partial x'^\mu}{\partial x^\rho}\frac{\partial x'^\nu}{\partial x^\sigma}. \tag{A.12}$$

Differentiating with respect to x^ϵ, we obtain

$$0 = g_{\mu\nu}\left[\frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\rho}\frac{\partial x'^\nu}{\partial x^\sigma} + \frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\sigma}\frac{\partial x'^\nu}{\partial x^\rho}\right]. \tag{A.13}$$

To this, we add the same equation with $\epsilon \leftrightarrow \rho$ and subtract the same equation with $\epsilon \leftrightarrow \sigma$. We obtain

$$0 = 2g_{\mu\nu}\frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\rho}\frac{\partial x'^\nu}{\partial x^\sigma}. \tag{A.14}$$

Assuming that the transformation $x^\mu \to x'^\mu$ is a well-behaved differentiable function and that the inverse of the transformation also exists,

$$\frac{\partial x'^\mu}{\partial x^\sigma}\frac{\partial x^\sigma}{\partial x'^\nu} = \delta^\mu_{\ \nu}, \tag{A.15}$$

we obtain that

$$\frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\rho} = 0. \tag{A.16}$$

Therefore, the transformation $x^\mu \to x'^\mu$ ought to be linear:

$$x'^\mu = \Lambda^\mu_{\ \nu}x^\nu + \rho^\mu. \tag{A.17}$$

A.2 Subgroups of Lorentz transformations

The set of all Lorentz transformations

$$x'^\mu = \Lambda^\mu_{\ \nu}x^\nu + \rho^\mu, \quad g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma} = g_{\rho\sigma}, \tag{A.18}$$

form a group (**exercise:** prove it), which is known as the *inhomogeneous Lorentz group* or the Poincaré group. The subset of transformations with $\rho^\mu = 0$ is known as the *homogeneous Lorentz group*.

From

$$g_{\mu\nu}\Lambda^{\mu}_{\ \rho}\Lambda^{\nu}_{\ \sigma} = g_{\rho\sigma}$$

and for $\rho = \sigma = 0$, we have:

$$(\Lambda^0_{\ 0})^2 - \sum_{i=1}^{3}(\Lambda^i_{\ 0})^2 = 1 \rightsquigarrow (\Lambda^0_{\ 0})^2 \geq 1. \tag{A.19}$$

Also, in matrix form the definition of the Lorentz transformation becomes

$$g = \Lambda^T g \Lambda \rightsquigarrow \det g = \det(\Lambda^T g \Lambda) \rightsquigarrow (\det \Lambda)^2 = 1. \tag{A.20}$$

The subgroup of transformations with

$$\det \Lambda = 1,\ \Lambda^0_{\ 0} \geq 1,$$

which contains the identity $\mathbf{1} = \delta^{\mu}_{\nu}$, is known as the group of proper Lorentz transformations. All other transformations are known as *improper Lorentz transformations*. It is impossible to change, with a continuous change of parameters,

$$\det \Lambda = 1 \rightarrow \det \Lambda = -1 \text{ or } \Lambda^0_{\ 0} \geq 1 \rightarrow \Lambda^0_{\ 0} \leq -1.$$

Improper Lorentz transformations involve either space reflection ($\det \Lambda = -1, \Lambda^0_{\ 0} \geq 1$) or time inversion ($\det \Lambda = -1, \Lambda^0_{\ 0} \leq -1$) or both ($\det \Lambda = 1, \Lambda^0_{\ 0} \leq -1$).

Proper homogeneous or inhomogeneous Lorentz transformations have a further subgroup: the group of rotations,

$$\Lambda^0_{\ 0} = 1,\ \Lambda^i_{\ 0} = \Lambda^0_{\ i} = 0,\ \Lambda^i_{\ j} = R_{ij}, \tag{A.21}$$

with

$$\det R = 1,\ R^T R = 1. \tag{A.22}$$

Thus, for rotations and translations ($x^{\mu} \rightarrow x'^{\mu} = x^{\mu} + \rho^{\mu}$), Lorentz transformations are no different from Galilei transformations.

A difference from Galilei transformations arises for *boosts*. Assume a reference frame O in which a certain particle appears at rest, and O' a reference frame where the particle appears to move with a velocity $\vec{v}$. Space-time intervals in the two frames are related via

$$dx'^{\mu} = \Lambda^{\mu}_{\ \nu} dx^{\nu} = \Lambda^{\mu}_{\ 0} c dt, \tag{A.23}$$

given that $d\vec{x} = 0$ in the frame O. For $\mu = 0$, this equation gives

$$dt' = \Lambda^0_{\ 0} dt. \tag{A.24}$$

For $\mu = i = 1, 2, 3$, we have

$$dx'^{i} = \Lambda^i_{\ 0} c dt. \tag{A.25}$$

Dividing the two, we have

$$v^i \equiv \frac{dx'^i}{dt'} = c\frac{\Lambda^i_{\ 0}}{\Lambda^0_{\ 0}} \rightsquigarrow \Lambda^i_{\ 0} = \frac{v^i}{c}\Lambda^0_{\ 0}. \tag{A.26}$$

From

$$g_{\mu\nu}\Lambda^{\mu}_{\ \rho}\Lambda^{\nu}_{\ \sigma} = g_{\rho\sigma}$$

and for $\rho = \sigma = 0$, we have

$$\begin{aligned} & (\Lambda^0_{\ 0})^2 - (\Lambda^i_{\ 0})^2 = 1 \\ \rightsquigarrow \qquad & \Lambda^0_{\ 0} = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}, \end{aligned} \tag{A.27}$$

and thus

$$\Lambda^i_{\ 0} = \gamma \frac{v^i}{c}. \tag{A.28}$$

The remaining components are not determined uniquely by knowing the velocity $\vec{v}$ of the particle. Indeed, two Lorentz transformations

$$\Lambda^\mu_{\ \nu} \quad \text{and} \quad \Lambda^\mu_{\ \rho} R^\rho_{\ \nu}$$

where R is a rotation, boost a particle to the same velocity. For coordinate systems O and O' with parallel axes we find that (**exercise**)

$$\Lambda^i_{\ j} = \delta^i_j + \frac{v^i v^j}{v^2}(\gamma - 1) \tag{A.29}$$

and

$$\Lambda^0_{\ j} = \gamma \frac{v^j}{c}. \tag{A.30}$$

A.3 Time dilation

Consider an inertial observer O who looks at a clock at rest. Two ticks of the clock correspond to a space-time interval

$$d\vec{x} = 0, \quad dt = \Delta t. \tag{A.31}$$

The proper time interval is

$$d\tau = (c^2 dt^2 - d\vec{x}^2)^{\frac{1}{2}} = c\Delta t. \tag{A.32}$$

A second observer sees the clock with velocity $\vec{v}$. Two ticks of the clock define a space-time interval

$$dt' = \Delta t', \quad d\vec{x}' = \vec{v} dt'. \tag{A.33}$$

The proper-time interval in the new frame is

$$d\tau' = (c^2 dt'^2 - d\vec{x}'^2)^{\frac{1}{2}} = c\Delta t' \sqrt{1 - \left|\frac{d\vec{x}'}{c dt'}\right|^2} = c\Delta t' \sqrt{1 - \frac{\vec{v}^2}{c^2}}. \tag{A.34}$$

The proper time is invariant under changes of inertial reference frame. Thus, we conclude that

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}} = \gamma \Delta t. \tag{A.35}$$

A.4 Doppler effect

Take our clock to be a source of light with a frequency

$$\omega = \frac{2\pi}{\Delta t}.$$

For an observer relative to whom the light source is moving with velocity $\vec{v}$, this time interval is measured to be

$$dt' = \gamma \Delta t.$$

In the same period, the distance of the observer from the light source increases by

$$v_r dt',$$

where v_r is the component of the velocity of the light source along the line of sight of the observer. The time dt_0 elapsing between the reception of two successive light wave fronts by the observer is given by

$$c dt_0 = c dt' + v_r dt'. \tag{A.36}$$

The frequency measured by the observer is

$$\omega' = \frac{2\pi}{dt_0} = \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{v_r}{c}} \omega. \tag{A.37}$$

If the light source is moving along the line of sight, $v_r = v$, we have

$$\omega' = \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} \omega. \tag{A.38}$$

If the light source moves away from the observer, $v_r > 0$, the frequency decreases and the light appears to be more red (red shift). If the source moves towards the observer, the frequency increases (blue shift).
Exercise: Calculate the angle of the direction of motion of the light source with respect to the line of sight of the observer for which there is no shift in the frequency.

For an application of the Doppler effect in cosmology, read about Hubble's law.

A.5 Particle dynamics

How can we compute the force on a particle which moves with a relativistic velocity $\vec{v}$? We should expect that our classical formulae from Newtonian mechanics need to be modified. Nevertheless, Newtonian expressions for the force should be valid if a particle is at rest. We can always change reference frame with Lorentz transformations to bring a particle to rest and calculate the change in its velocity for a small time interval using Newtonian mechanics. However, we will need to perform these changes of reference frame at every small increase of the velocity of the particle during its acceleration due to the force.

As a more elegant solution to the problem, we define a relativistic force acting on a particle as

$$f^\mu = mc^2 \frac{d^2 x^\mu}{d\tau^2}, \tag{A.39}$$

where m is the mass of a particle[1]. If the particle is at rest, the proper-time interval $d\tau$ coincides with the common time interval dt (times c),

$$d\tau = cdt.$$

Therefore, in the rest frame of the particle, the "space" components of the force four-vector become

$$f^i_{\text{rest}} = m\frac{d^2x^i}{dt^2} = F^i_{\text{Newton}}, \quad \text{for } i = 1, 2, 3, \tag{A.40}$$

where $\vec{F}_{\text{Newton}}$ is the force vector as we know it from Newtonian mechanics. The "time" component of the force four-vector vanishes:

$$f^0_{\text{rest}} = mc\frac{d^2t}{dt^2} = 0. \tag{A.41}$$

Under a Lorentz transformation, f^μ transforms as

$$f^\mu = mc^2\frac{d^2x^\mu}{d\tau^2} \to mc^2\frac{d^2x'^\mu}{d\tau^2} = mc^2\frac{d^2\left(\Lambda^\mu_{\ \nu}x^\nu + \rho^\mu\right)}{d\tau^2} = \Lambda^\mu_{\ \nu}\, mc^2\frac{d^2x^\nu}{d\tau^2}. \tag{A.42}$$

Therefore,

$$f'^\mu = \Lambda^\mu_{\ \nu} f^\nu. \tag{A.43}$$

The components of f^μ transform under Lorentz transformations in exactly the same way as the components of space-time coordinates. It is therefore a four-vector as well.

For a specific transformation from the rest frame of a particle to a frame where the particle moves with a velocity $\vec{v}$, we have

$$f^\mu = \Lambda^\mu_{\ \nu}(\vec{v}) f^\nu_{\text{rest}}, \tag{A.44}$$

where we have found that

$$\begin{aligned}\Lambda^0_{\ 0}(\vec{v}) = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}, \quad \Lambda^i_{\ 0}(\vec{v}) = \Lambda^0_{\ i}(\vec{v}) = \gamma\frac{v^i}{c},\\ \Lambda^i_{\ j}(\vec{v}) = \delta^i_j + \frac{v^i v^j}{v^2}(\gamma - 1).\end{aligned} \tag{A.45}$$

Therefore, the force on a moving particle is:

$$\vec{f} = \vec{F}_{\text{Newton}} + (\gamma - 1)\frac{\vec{v}\left(\vec{F}_{\text{Newton}} \cdot \vec{v}\right)}{v^2}, \tag{A.46}$$

and

$$f^0 = \gamma\frac{\vec{v} \cdot \vec{F}_{\text{Newton}}}{c} = \frac{\vec{v}}{c} \cdot \vec{f}. \tag{A.47}$$

In Newtonian mechanics, if the force $\vec{F}$ is given, we can compute the trajectory $\vec{x}(t)$ by solving the second-order differential equation

$$\frac{d^2\vec{x}}{dt^2} = \frac{\vec{F}(\vec{x}, t)}{m}. \tag{A.48}$$

[1] By mass, we mean the mass of a particle as it is measured in its rest frame. We will refrain from using the "relativistic", velocity-dependent, mass.

Similarly, in special relativity, when the relativistic force f^μ is known, the differential equation of Eq. A.39 can, in principle, be solved to give the space-time coordinates as a function of the proper time τ:

$$x^\mu = x^\mu(\tau). \tag{A.49}$$

To calculate the trajectory, we then need to calculate the proper time in terms of the time coordinate by inverting

$$x^0 = x^0(\tau) \rightsquigarrow \tau = \tau(x^0), \tag{A.50}$$

which we can use to cast the space components directly as functions of the time coordinate.

We should not forget a second constraint that must be satisfied by our solutions $x^\mu(\tau)$, namely

$$\Omega \equiv g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 1. \tag{A.51}$$

For the derivative of Ω with respect to proper time, we have

$$\frac{d\Omega}{d\tau} = 2g_{\mu\nu}\frac{d^2x^\mu}{d\tau^2}\frac{dx^\nu}{d\tau} = \frac{2}{mc^2}g_{\mu\nu}f^\mu\frac{dx^\nu}{d\tau}. \tag{A.52}$$

The rhs is a Lorentz-invariant quantity. In a new frame,

$$g_{\mu\nu}f'^\mu\frac{dx'^\nu}{d\tau} = g_{\mu\nu}\left(\Lambda^\mu_{\ \rho}f^\rho\right)\frac{(\Lambda^\nu_{\ \sigma}dx^\sigma)}{d\tau} = \left(g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma}\right)f^\rho\frac{dx^\sigma}{d\tau} = g_{\rho\sigma}f^\rho\frac{dx^\sigma}{d\tau}.$$

We are therefore allowed to compute $\frac{d\Omega}{d\tau}$ in any reference frame we wish. Let us choose the rest frame of the particle, where

$$x^\mu = (ct, \vec{0}), \quad f^\mu = (0, \vec{F}_{\text{Newton}}).$$

We obtain

$$\frac{d\Omega}{d\tau} = \frac{2}{mc^2}\left(f^0\frac{dx^0}{d\tau} - \vec{f}\cdot\frac{d\vec{x}}{d\tau}\right) = 0. \tag{A.53}$$

Therefore, the quantity Ω is always a constant:

$$\Omega(\tau) = \text{constant}. \tag{A.54}$$

If for some initial value τ_0 we choose the constant to be one, we will have

$$\Omega(\tau) = \Omega(\tau_0) = 1, \qquad \forall\ \tau. \tag{A.55}$$

Exercise: Calculate the trajectory of a particle on which the exerted four-vector force is $f^\mu = 0$.

A.6 Energy and momentum

We can define a relativistic four-vector analogue of momentum as

$$p^\mu = mc\frac{dx^\mu}{d\tau}. \tag{A.56}$$

We have that

$$d\tau = \left(c^2dt^2 - d\vec{x}^2\right)^{\frac{1}{2}} = cdt\left[1 - \left(\frac{d\vec{x}}{cdt}\right)^2\right]^{\frac{1}{2}} = cdt\left[1 - \frac{\vec{v}^2}{c^2}\right]^{\frac{1}{2}} = \frac{cdt}{\gamma}. \tag{A.57}$$

Thus, for the time component ($\mu = 0$) of the four-momentum we have

$$p^0 = mc\frac{dx^0}{d\tau} = m\gamma c. \tag{A.58}$$

For the space components ($\mu = i = 1, 2, 3$) we have

$$p^i = mc\frac{dx^i}{d\tau} = m\gamma\frac{dx^i}{dt} = m\gamma v^i. \tag{A.59}$$

For small velocities, we can expand the factor γ as

$$\gamma = \left[1 - \frac{v^2}{c^2}\right]^{-\frac{1}{2}} \approx 1 + \frac{1}{2}\frac{v^2}{c^2} + \mathcal{O}\left(\frac{v^4}{c^4}\right). \tag{A.60}$$

Therefore, for small velocities the space components of the four-momentum become the classical momentum,

$$p^i \approx mv^i + \ldots, \tag{A.61}$$

while the time component becomes

$$p^0 \approx mc + \frac{1}{2c}mv^2 + \ldots \tag{A.62}$$

In the second term of the above expansion we recognise the kinetic energy $\frac{1}{2}mv^2$ of the particle. We then identify the relativistic energy of a particle with

$$E = cp^0 = m\gamma c^2. \tag{A.63}$$

Eliminating the velocity $\vec{v}$ from Eqs. A.59 and A.63, we obtain the relation

$$E = \sqrt{\vec{p}^2c^2 + m^2c^4}. \tag{A.64}$$

A.7 The inverse of a Lorentz transformation

Recall the metric matrix

$$g_{\mu\nu} = \text{diag}(1, -1, -1, -1). \tag{A.65}$$

We define its inverse

$$g^{\mu\nu}: \quad g^{\mu\nu}g_{\nu\rho} = \delta^\mu_{\ \rho}, \tag{A.66}$$

where $\delta^\mu_{\ \nu}$ is the Kronecker delta. It is easy to verify that the inverse of the metric is the metric itself:

$$g^{\mu\nu} = g_{\mu\nu} = \text{diag}(1, -1, -1, -1). \tag{A.67}$$

Now consider a Lorentz transformation $\Lambda^\mu_{\ \nu}$, which satisfies the identity

$$\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma}g_{\mu\nu} = g_{\rho\sigma}. \tag{A.68}$$

We can prove that the matrix

$$\Lambda_\mu^{\ \nu} \equiv g_{\mu\rho}g^{\nu\sigma}\Lambda^\rho_{\ \sigma} \tag{A.69}$$

is the inverse of $\Lambda^\mu_{\ \nu}$. Indeed,

$$\Lambda^\mu_{\ \lambda}\Lambda_\mu^{\ \nu} = g_{\mu\rho}g^{\nu\sigma}\Lambda^\rho_{\ \sigma}\Lambda^\mu_{\ \lambda} = g_{\sigma\lambda}g^{\nu\sigma} = \delta^\nu_{\ \lambda}. \tag{A.70}$$

If Λ^{μ}_{ν} is the boost with velocity $\vec{v}$ of Eq. A.45, then

$$\Lambda^{0}_{0}(\vec{v}) = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}, \quad \Lambda^{0}_{i}(\vec{v}) = \Lambda^{i}_{0}(\vec{v}) = -\gamma\frac{v^i}{c},$$
$$\Lambda^{j}_{i}(\vec{v}) = \delta^{j}_{i} + \frac{v^i v^j}{v^2}(\gamma - 1). \tag{A.71}$$

We therefore have that the inverse of a boost is

$$\Lambda^{\nu}_{\mu}(\vec{v}) = \Lambda^{\mu}_{\nu}(-\vec{v}), \tag{A.72}$$

as we also expect physically.

A.8 Vectors and tensors

It is now time to give a formal definition of vectors in special relativity. We call any set of four components which transform according to the rule

$$V^{\mu} \to V'^{\mu} = \Lambda^{\mu}_{\nu} V^{\nu}, \tag{A.73}$$

a *contravariant* vector. Contravariant vectors transform in the same way as space-time coordinates x^{μ} do under homogeneous Lorentz transformations.

Not all vectors transform as contravariant vectors. Consider the derivative $\frac{\partial}{\partial x^{\mu}}$. Under a Lorentz transformation, it transforms as

$$\frac{\partial}{\partial x^{\mu}} \to \frac{\partial}{\partial x'^{\mu}} = \frac{\partial x^{\rho}}{\partial x'^{\mu}}\frac{\partial}{\partial x^{\rho}}. \tag{A.74}$$

We have that

$$\left(\frac{\partial x^{\rho}}{\partial x'^{\mu}}\right)\left(\frac{\partial x'^{\mu}}{\partial x^{\nu}}\right) = \delta^{\rho}_{\nu} \rightsquigarrow \left(\frac{\partial x^{\rho}}{\partial x'^{\mu}}\right)\Lambda^{\mu}_{\nu} = \delta^{\rho}_{\nu}. \tag{A.75}$$

Therefore, $\left(\frac{\partial x^{\rho}}{\partial x'^{\mu}}\right)$ is the inverse of a Lorentz transformation Λ^{μ}_{ν}:

$$\frac{\partial x^{\nu}}{\partial x'^{\mu}} = \Lambda^{\nu}_{\mu}. \tag{A.76}$$

Substituting into Eq. A.74, we find

$$\frac{\partial}{\partial x^{\mu}} \to \frac{\partial}{\partial x'^{\mu}} = \Lambda^{\rho}_{\mu}\frac{\partial}{\partial x^{\rho}}. \tag{A.77}$$

We find that the derivative does not transform according to the Lorentz transformation but according to its inverse. All vectors which transform with the inverse Lorentz transformation,

$$U_{\mu} \to U'_{\mu} = \Lambda^{\nu}_{\mu} U_{\nu}, \tag{A.78}$$

are called covariant vectors.

For every contravariant vector U^{μ} there is a dual vector

$$U_{\mu} = g_{\mu\nu}U^{\nu}. \tag{A.79}$$

We can invert the above equation by multiplying with $g^{\rho\mu}$,

$$g^{\rho\mu}U_\mu = g^{\rho\mu}g_{\mu\nu}U^\nu = \delta^\rho_\nu U^\nu = U^\rho. \tag{A.80}$$

The dual vector U_μ is a covariant vector. Indeed, under a Lorentz transformation we have

$$U_\mu \to U'_\mu = g_{\mu\nu}U'^\nu = g_{\mu\nu}\Lambda^\nu_{\ \rho}U^\rho = g_{\mu\nu}\Lambda^\nu_{\ \rho}g^{\rho\sigma}U_\sigma = \Lambda^\sigma_{\ \mu}U_\sigma. \tag{A.81}$$

The scalar product of a contravariant and a covariant vector

$$A \cdot B \equiv A^\mu B_\mu = A_\mu B^\mu = g_{\mu\nu}A^\mu B^\nu = g^{\mu\nu}A_\mu B_\nu, \tag{A.82}$$

is invariant under Lorentz transformations. Indeed,

$$A \cdot B \to A' \cdot B' = A'^\mu B'_\mu = \Lambda^\mu_{\ \rho}A^\rho\Lambda^\sigma_{\ \mu}B_\sigma = \delta^\sigma_{\ \rho}A^\rho B_\sigma = A^\rho B_\rho = A \cdot B. \tag{A.83}$$

Let us define, for short,

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu}, \tag{A.84}$$

and the dual contravariant vector

$$\partial^\mu = \frac{\partial}{\partial x_\mu} = g^{\mu\nu}\partial_\nu. \tag{A.85}$$

The d'Alembert operator is the scalar product

$$\Box \equiv \partial^2 \equiv \partial_\mu\partial^\mu = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2. \tag{A.86}$$

Being a scalar product, the d'Alembert operator is invariant under Lorentz transformations.

Finally, we define a tensor with multiple "up" and/or "down" indices to be an object

$$T^{\mu_1\mu_2\cdots}_{\nu_1\nu_2\cdots} \tag{A.87}$$

which transforms as

$$T^{\mu_1\mu_2\cdots}_{\nu_1\nu_2\cdots} \to \Lambda^{\mu_1}_{\rho_1}\Lambda^{\mu_2}_{\rho_2}\ldots\Lambda^{\sigma_1}_{\nu_1}\Lambda^{\sigma_2}_{\nu_2}\ldots T^{\rho_1\rho_2\cdots}_{\sigma_1\sigma_2\ldots}. \tag{A.88}$$

A.9 Currents and densities

Consider a set of particles $\{n\}$ with charges q_n at positions $\vec{r}_n(t)$. The charge and current densities are

$$\rho(\vec{x},t) = \sum_n q_n\delta(\vec{x} - \vec{r}_n(t)), \tag{A.89}$$

$$\vec{j}(\vec{x},t) = \sum_n q_n\frac{d\vec{r}_n(t)}{dt}\delta(\vec{x} - \vec{r}_n(t)) = \sum_n q_n\frac{d\vec{x}}{dt}\delta(\vec{x} - \vec{r}_n(t)). \tag{A.90}$$

Recall now that for the vector $x^\mu = (ct, \vec{x})$, we have

$$\frac{dx^\mu}{dt} = \frac{d}{dt}(ct,\vec{x}) = \left(c, \frac{d\vec{x}}{dt}\right). \tag{A.91}$$

We can then combine the charge and current densities into one object:

$$j^\mu \equiv \left(c\rho, \vec{j}\right), \tag{A.92}$$

with

$$j^\mu(\vec{x},t) = \sum_n q_n \frac{dx^\mu}{dt}\delta(\vec{x}-\vec{r}_n(t)). \tag{A.93}$$

We can now cast j^μ in a form which manifestly shows that it is a contravariant four-vector. First, we define a delta function in four dimensions as

$$\delta\left(x^\mu - y^\mu\right) = \delta(x^0-y^0)\delta(\vec{x}-\vec{y}) = \frac{1}{c}\delta(t_x-t_y)\delta(\vec{x}-\vec{y}). \tag{A.94}$$

Notice that the δ function of a four-vector is a scalar. Under a Lorentz transformation,

$$\delta\left(U^\mu\right) \to \delta\left(U'^\mu\right) = \delta\left(\Lambda^\mu_{\ \nu}U^\nu\right) = \frac{\delta(U^\nu)}{|\det\Lambda|} = \delta(U^\nu). \tag{A.95}$$

With the use of the delta function, we can write the current-density four-vector of Eq. A.93 as an integral

$$j^\mu(\vec{x},t) = \sum_n q_n \int dt' \frac{dr_n^\mu}{dt'}\delta(\vec{x}-\vec{r}_n(t'))\delta(t'-t) = c\sum_n q_n \int dt' \frac{dr_n^\mu}{dt'}\delta(x^\mu - r_n^\mu(t')), \tag{A.96}$$

where

$$x^\mu \equiv (ct,\vec{x}), \quad r_n^\mu(t') \equiv (ct',\vec{r}_n(t')).$$

Now, we change the integration variable from t' to τ. Recall that

$$dt' = \frac{d\tau}{c}\gamma. \tag{A.97}$$

Then, we have

$$j^\mu(\vec{x},t) = c\sum_n q_n \int d\tau \frac{dr_n^\mu}{d\tau}\delta(x^\mu - r_n^\mu(\tau)), \tag{A.98}$$

which shows manifestly that j^μ transforms as $\frac{dx^\mu}{d\tau}$ and is therefore a contravariant four-vector.

Charge conservation implies the continuity equation

$$\frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot\vec{j} = 0. \tag{A.99}$$

In relativistic notation, the continuity equation takes the more elegant form

$$\partial_\mu j^\mu = 0. \tag{A.100}$$

A.10 The energy-momentum tensor

Consider a collection of particles $\{n\}$ at positions $\vec{r}_n(t)$. The energy density is

$$\text{energy density} = \sum_n E_n(t)\delta(\vec{x}-\vec{r}_n(t)). \tag{A.101}$$

Changes in the energy density result in an "energy-current density",

$$\text{energy current density} = \sum_n E_n(t)\frac{d\vec{r}_n}{dt}\delta(\vec{x}-\vec{r}_n(t)). \tag{A.102}$$

As with the charge density and its current density in the last section, we can combine the two together into a single object:

$$\sum_n E_n(t)\frac{dr_n^\nu}{dt}\delta(\vec{x}-\vec{r}_n(t)). \tag{A.103}$$

Similarly to the energy, the densities and current densities for the components of the three-dimensional momentum are

$$\sum_n p_n^i(t)\frac{dr_n^\nu}{dt}\delta(\vec{x}-\vec{r}_n(t)), \quad i=1,2,3. \tag{A.104}$$

Collectively, we can form the so-called "energy-momentum tensor" which encompasses the density and current density for all components of the four-momentum:

$$T^{\mu\nu} \equiv \sum_n p_n^\mu \frac{dr_n^\nu}{dt}\delta(\vec{x}-\vec{r}_n(t)), \tag{A.105}$$

or, in the equivalent form,

$$T^{\mu\nu} = c\sum_n \int d\tau p_n^\mu \frac{dx_n^\nu}{d\tau}\delta(x^\rho - r_n^\rho(\tau)). \tag{A.106}$$

From the last equation, we can see manifestly that this object is justifiably called a tensor since it transforms as the product of two four-vectors under Lorentz transformations:

$$T'^{\mu\nu} = \Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma}T^{\rho\sigma}. \tag{A.107}$$

The energy-momentum tensor is symmetric:

$$T^{\mu\nu} = T^{\nu\mu}. \tag{A.108}$$

To verify this, we recall that

$$p_n^\nu = m_n c\frac{dr_n^\nu}{d\tau} = m_n\gamma\frac{dr_n^\nu}{dt} = \frac{E_n}{c^2}\frac{dr_n^\nu}{dt}. \tag{A.109}$$

The energy-momentum tensor then takes the explicitly symmetric form

$$T^{\mu\nu} \equiv \sum_n c^2\frac{p_n^\mu p_n^\nu}{E_n}\delta(\vec{x}-\vec{r}_n(t)). \tag{A.110}$$

For the charge-density four-vector, we have found the continuity identity

$$\partial_\mu j^\mu = 0.$$

This was a consequence of the conservation of charge. If the total energy and momentum of the system of particles is conserved, we anticipate a similar continuity identity for the energy-momentum tensor:

$$\partial_\nu T^{\mu\nu} = 0.$$

We have:

$$\begin{aligned}
\partial_i T^{\mu i} &= \sum_n p_n^\mu \frac{dr_n^i}{dt}\frac{\partial}{\partial x^i}\delta\left(\vec{x}-\vec{r}_n\right) \\
&= -\sum_n p_n^\mu \frac{dr_n^i}{dt}\frac{\partial}{\partial r_n^i}\delta\left(\vec{x}-\vec{r}_n\right) \\
&= -\sum_n p_n^\mu \frac{\partial}{\partial t}\delta\left(\vec{x}-\vec{r}_n\right) \\
&= -\frac{\partial}{\partial t}\sum_n p_n^\mu \delta\left(\vec{x}-\vec{r}_n\right) + \sum_n \frac{\partial p_n^\mu}{\partial t}\delta\left(\vec{x}-\vec{r}_n\right) \\
&= -\frac{1}{c}\frac{\partial}{\partial t}T^{\mu 0} + \sum_n \frac{\partial p_n^\mu}{\partial t}\delta\left(\vec{x}-\vec{r}_n\right). \qquad \text{(A.111)}
\end{aligned}$$

In the above we have used that

$$\begin{aligned}
\partial_x \delta(x-y) &= \frac{1}{2\pi}\partial_x \int_{-\infty}^{\infty} dw e^{-iw(x-y)} = -\frac{i}{2\pi}\int_{-\infty}^{\infty} dw\, w e^{-iw(x-y)} \\
&= -\frac{1}{2\pi}\partial_y \int_{-\infty}^{\infty} dw e^{-iw(x-y)} = -\partial_y \delta(x-y). \qquad \text{(A.112)}
\end{aligned}$$

We have therefore arrived at the equation

$$\partial_\nu T^{\mu\nu} = G^\mu, \qquad \text{(A.113)}$$

where

$$G^\mu = \sum_n \frac{\partial p_n^\mu}{\partial t}\delta\left(\vec{x}-\vec{r}_n\right) = \sum_n \frac{1}{c}\frac{\partial \tau}{\partial t} f_n^\mu(t)\delta\left(\vec{x}-\vec{r}_n\right) \qquad \text{(A.114)}$$

is the "density of force".

For free particles, where the energy and momentum of each particle are separately conserved, p_n^μ = constant, the energy-momentum tensor satisfies the continuity equation

$$\partial_\nu T^{\mu\nu} = 0. \qquad \text{(A.115)}$$

The energy-momentum tensor is also conserved if the particles interact only at the points where they collide with each other. In that case, the force density is

$$\begin{aligned}
G^\mu &= \sum_n \frac{\partial p_n^\mu}{\partial t}\delta\left(\vec{x}-\vec{r}_n\right) \\
&= \sum_{coll.} \delta(\vec{x}-\vec{x}_{coll}(t))\frac{d}{dt}\sum_{n\in coll.} p_n^\mu(t). \qquad \text{(A.116)}
\end{aligned}$$

We have grouped the sum over all particles contributing to the force density according to the collision points at which they meet. At each collision point, the sum of the momenta of the colliding particles is conserved,

$$\frac{d}{dt}\sum_{n\in coll.} p_n^\mu(t) = 0 \rightsquigarrow \partial_\nu T^{\mu\nu} = 0. \qquad \text{(A.117)}$$

If the continuity equation is satisfied, then

$$\begin{aligned} 0 &= \partial_\nu T^{\mu\nu} \\ \rightsquigarrow 0 &= \partial_0 T^{\mu 0} + \partial_i T^{\mu i} \\ \rightsquigarrow 0 &= \partial_0 \int d^3\vec{x} T^{\mu 0} + \int d^3\vec{x} \partial_i T^{\mu i} \\ 0 &= \partial_0 \int d^3\vec{x} T^{\mu 0}, \end{aligned} \tag{A.118}$$

which means that the vector

$$P^\mu \equiv \int d^3\vec{x} T^{\mu 0} = \text{constant} \tag{A.119}$$

is conserved. We find that the conserved vector is the total four-momentum of the particles in the distribution:

$$P^\mu = \sum_n \int d^3\vec{x} p_n^\mu \delta(\vec{x} - \vec{r}_n(t)) = \sum_n p_n^\mu. \tag{A.120}$$

A.11 Relativistic formulation of electrodynamics

From now on we will set $\epsilon_0 = c = 1$. It will be easy to restore the full dependence on these constants with dimensional analysis when necessary. The Maxwell equations are

$$\vec{\nabla} \cdot \vec{E} = \rho, \tag{A.121}$$

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0, \tag{A.122}$$

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{A.123}$$

$$\vec{\nabla} \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = \vec{j}. \tag{A.124}$$

We construct an "electromagnetic field tensor" $F^{\mu\nu}$ from the components of the electric $\vec{E} \equiv (E^1, E^2, E^3)$ and magnetic field $\vec{B} \equiv (B^1, B^2, B^3)$ as

$$F^{0i} = -E^i, \quad F^{ij} = -\epsilon_{ijk} B^k, \quad F^{\mu\nu} = -F^{\nu\mu}, \tag{A.125}$$

where we use $\epsilon_{123} = +1$. Explicitly,

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix}. \tag{A.126}$$

Conversely, if the tensor $F^{\mu\nu}$ is given, we can obtain the magnetic field via (**exercise**):

$$B^i = -\frac{1}{2} \epsilon_{ijk} F^{jk}. \tag{A.127}$$

Exercise: What is the covariant tensor $F_{\mu\nu}$ in matrix form?
Exercise: Write the matrix:

$$\tilde{F}_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma}. \tag{A.128}$$

Exercise: Prove that:

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{\vec{E}^2 - \vec{B}^2}{2}, \quad \epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma} = 8\vec{E}\cdot\vec{B} \quad (\epsilon_{0123} = +1). \tag{A.129}$$

As we have already noted, the charge and current densities form a four-vector:

$$j^\mu = (\rho, \vec{j}).$$

We can observe that two of the Maxwell equations can be combined into one,

$$\partial_\nu F^{\nu\mu} = j^\mu. \tag{A.130}$$

Indeed, for $\mu = 0$, we have

$$\partial_0 F^{00} + \partial_i F^{i0} = j^0 \rightsquigarrow \vec{\nabla}\cdot\vec{E} = \rho.$$

For $\mu = i = 1, 2, 3$ we have

$$\partial_0 F^{0i} + \partial_j F^{ji} = j^i \rightsquigarrow -\partial_0 E^i + \epsilon_{ijk}\partial_j B^k = j^i \rightsquigarrow \vec{\nabla}\times\vec{B} - \frac{\partial\vec{E}}{\partial t} = \vec{j}.$$

The remaining two Maxwell equations tell us, as we know from classical electrodynamics, that we can derive the electric and magnetic fields by means of the scalar and vector potentials $\phi, \vec{A}$, via

$$\vec{E} = -\vec{\nabla}\phi - \frac{\partial\vec{A}}{\partial t}, \quad \vec{B} = \vec{\nabla}\times\vec{A}. \tag{A.131}$$

We can combine the scalar and vector potentials into a single four-vector:

$$A^\mu \equiv (\phi, \vec{A}) = (\phi, A^1, A^2, A^3). \tag{A.132}$$

Then, the above equations take the elegant form

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{A.133}$$

Indeed, for $\mu = 0, \nu = i = 1, 2, 3$ we have

$$F^{0i} = \partial^0 A^i - \partial^i A^0 \rightsquigarrow -\vec{E} = \frac{\partial\vec{A}}{\partial t} + \vec{\nabla}\phi.$$

For $\mu = i, \nu = j,\ i, j = 1, 2, 3$ we have

$$F^{ij} = \partial^i A^j - \partial^j A^i \rightsquigarrow -\frac{1}{2}\epsilon_{ijk}F^{ij} = \epsilon_{ijk}\partial_i A^j \rightsquigarrow \vec{B} = \vec{\nabla}\times\vec{A}.$$

It is now straightforward to prove the following identities (**exercise**):

$$\epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} = 0, \tag{A.134}$$

$$\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0. \tag{A.135}$$

Substituting Eq. A.133 into Eq. A.130, we find

$$\partial^2 A^\nu - \partial^\nu(\partial_\mu A^\mu) = j^\nu. \tag{A.136}$$

Explicitly, for $\nu = 0$ and $\nu = i = 1, 2, 3$ we recover the known differential equations for the scalar and vector potentials respectively:

$$\Box\phi - \frac{\partial}{\partial t}\left(\vec{\nabla}\cdot\vec{A} + \frac{\partial\phi}{\partial t}\right) = \rho, \tag{A.137}$$

$$\Box\vec{A} + \vec{\nabla}\left[\vec{\nabla}\cdot\vec{A} + \frac{\partial\phi}{\partial t}\right] = \vec{j}. \tag{A.138}$$

The property of gauge invariance becomes more elegant as well in relativistic notation. The gauge transformations of the vector and scalar potentials which leave the Maxwell equations invariant are written as

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu\chi, \tag{A.139}$$

where χ is a scalar function. Indeed, under a gauge transformation the electromagnetic field tensor remains invariant:

$$F_{\mu\nu} \to \partial_\mu(A_\nu + \partial_\nu\chi) - \partial_\nu(A_\mu + \partial_\mu\chi) = \partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu}. \tag{A.140}$$

The Lorenz gauge-fixing condition

$$\vec{\nabla}\cdot\vec{A} + \frac{\partial\phi}{\partial t} = 0 \tag{A.141}$$

can be written simply as

$$\partial_\mu A^\mu = 0. \tag{A.142}$$

In the Lorenz gauge, the Maxwell equations for the four-vector potential become

$$\partial^2 A^\mu = j^\mu. \tag{A.143}$$

From classical electrodynamics, we know that the scalar and vector potentials can be computed via

$$\phi(\vec{r}, t) = \int_{-\infty}^{\infty} d^3\vec{r}'dt' G\left(\vec{r} - \vec{r}'; t - t'\right)\rho(\vec{r}', t') \tag{A.144}$$

and

$$\vec{A}(\vec{r}, t) = \int_{-\infty}^{\infty} d^3\vec{r}'dt' G\left(\vec{r} - \vec{r}'; t - t'\right)\vec{j}(\vec{r}', t'), \tag{A.145}$$

where the Green's function can be written in the form

$$G(\Delta\vec{r}, \Delta t) = \frac{1}{2\pi}\delta\left(\Delta t^2 - \Delta\vec{r}^2\right)\Theta(\Delta t > 0). \tag{A.146}$$

In relativistic notation, the solutions for the four-vector potential take the form

$$A^\nu(x^\mu) = \frac{1}{2\pi}\int d^4x' j^\nu(x'^\mu)\delta\left((x'^\mu - x^\mu)^2\right)\Theta\left(x^0 > x'^0\right). \tag{A.147}$$

The electromagnetic force acting on a particle with charge q is

$$f^\mu = qF^{\mu\nu}\frac{dx_\nu}{d\tau}. \tag{A.148}$$

Indeed, in the rest frame of the particle, where $d\tau = dt$ and $d\vec{x} = 0$, we get

$$f^{\mu}_{\text{rest}} = qF^{\mu 0} \rightsquigarrow f^{0}_{\text{rest}} = 0, \ \vec{f}_{\text{rest}} = q\vec{E}. \tag{A.149}$$

Since our expression for the electromagnetic force is correct in one reference frame, it should hold in every inertial reference frame, since it is written as an equation between four-vectors. In a frame where the charge is moving with a velocity $\vec{v}$, the three-dimensional force is

$$\begin{aligned} f^i &= qF^{i0}\frac{dx^0}{d\tau} - qF^{ij}\frac{dx^j}{d\tau} = q\gamma\left(E^i + \epsilon_{ijk}B^k v^j\right) \\ \rightsquigarrow \vec{f} &= q\gamma\left(\vec{E} + \vec{v}\times\vec{B}\right). \end{aligned} \tag{A.150}$$

Recall that

$$\vec{f} = \frac{d\vec{p}}{d\tau} = \gamma\frac{d\vec{p}}{dt}. \tag{A.151}$$

We have therefore recovered our familiar expression for the Lorentz force:

$$\frac{d\vec{p}}{dt} = q\left(\vec{E} + \vec{v}\times\vec{B}\right). \tag{A.152}$$

A.11.1 The energy-momentum tensor in the presence of an electromagnetic field

Consider a number of charges q_n which interact via the electromagnetic field. The energy-momentum tensor is not conserved:

$$\partial_\nu T^{\mu\nu} = G^{\mu}, \tag{A.153}$$

where the force density is given by

$$\begin{aligned} G^{\mu} &= \sum_n \frac{\partial\tau}{\partial t} f_n^{\mu}(t)\delta\left(\vec{x} - \vec{r}_n\right) \\ &= \sum_n \frac{\partial\tau}{\partial t} q_n F^{\mu}{}_{\nu}\frac{dr_n^{\nu}}{d\tau}\delta\left(\vec{x} - \vec{r}_n\right) \\ &= F^{\mu}{}_{\nu}\sum_n q_n\frac{dr_n^{\nu}}{dt}\delta\left(\vec{x} - \vec{r}_n\right) \\ &= F^{\mu\nu}j_\nu. \end{aligned} \tag{A.154}$$

Consider the tensor

$$T^{\mu\nu}_{em} \equiv F^{\mu}_{\rho}F^{\rho\nu} + \frac{1}{4}g^{\mu\nu}F_{\rho\sigma}F^{\rho\sigma}. \tag{A.155}$$

This tensor is explicitly

- symmetric,
- gauge invariant.

The components of the tensor are (**exercise**)

$$T^{00}_{em} = \frac{\vec{E}^2 + \vec{B}^2}{2}, \quad T^{0i}_{em} = T^{i0}_{em} = (\vec{E}\times\vec{B})_i. \tag{A.156}$$

Exercise: Find the remaining components.

We find

$$\begin{aligned}\partial_\nu T^{\mu\nu}_{em} &= \partial_\nu \left\{ F^{\mu\rho} F_\rho{}^\nu + \frac{g^{\mu\nu}}{4} F_{\rho\sigma} F^{\rho\sigma} \right\} \\ &= F^{\mu\rho} \partial_\nu F_\rho{}^\nu + (\partial^\nu F^{\mu\rho}) F_{\rho\nu} + \frac{1}{2} F_{\rho\sigma} \partial^\mu F^{\rho\sigma} \\ &= -F^{\mu\rho} j_\rho + \frac{1}{2} (\partial^\nu F^{\mu\rho} - \partial^\rho F^{\mu\nu}) F_{\rho\nu} + \frac{1}{2} F_{\rho\sigma} \partial^\mu F^{\rho\sigma} \\ &= -F^{\mu\rho} j_\rho + \frac{1}{2} (\partial^\sigma F^{\mu\rho} + \partial^\mu F^{\rho\sigma} + \partial^\rho F^{\sigma\mu}) F_{\rho\sigma} \end{aligned} \tag{A.157}$$

which, due to Eq. A.135, yields

$$\partial_\nu T^{\mu\nu}_{em} = -F^{\mu\nu} j_\nu . \tag{A.158}$$

$T^{\mu\nu}_{em}$ is purely electromagnetic. While neither $T^{\mu\nu}$ nor $T^{\mu\nu}_{em}$ satisfies a continuity equation, their sum

$$\Theta^{\mu\nu} \equiv T^{\mu\nu} + T^{\mu\nu}_{em} = \sum_n \frac{p^\mu_n p^\nu_n}{E_n} \delta(\vec{x} - \vec{r}_n(t)) + F^{\mu\rho} F_\rho{}^\nu + \frac{g^{\mu\nu}}{4} F_{\rho\sigma} F^{\rho\sigma}, \tag{A.159}$$

does:

$$\partial_\nu \Theta^{\mu\nu} = 0. \tag{A.160}$$

From the continuity equation, we obtain a conserved four-vector:

$$\partial_t P^\mu \equiv \partial_t \int d^3\vec{x} \Theta^{\mu 0} = - \int d^3\vec{x} \partial_i \Theta^{\mu i} = - \Theta^{\mu i}\Big|_\infty = 0. \tag{A.161}$$

The conserved vector is:

$$P^\mu = \int d^3\vec{x} \Theta^{\mu 0} = \sum_n p^\mu_n + \int d^3\vec{x} T^{\mu 0}_{em}. \tag{A.162}$$

The four-momentum of the charges,

$$P^\mu_{\text{charges}} = \sum_n p^\mu_n, \tag{A.163}$$

is not conserved on its own. Some momentum

$$P^\mu_{em} = \int d^3\vec{x} T^{\mu 0}_{em}, \tag{A.164}$$

is also carried by the electromagnetic field itself. This is not conserved either. Momentum can be exchanged between the charges and the field; however, this is done in such a way that the total momentum is always the same:

$$P^\mu = P^\mu_{\text{charges}} + P^\mu_{em} = \text{constant}. \tag{A.165}$$

The time component of the four-vector is the total energy. The energy stored in the electromagnetic field is:

$$E_{em} = \int d^3\vec{x} T^{00}_{em}. \tag{A.166}$$

Therefore, the energy density w of the electromagnetic field is

$$w = T^{00}_{em} = \frac{\vec{E}^2 + \vec{B}^2}{2}, \quad \text{EM field energy density.} \tag{A.167}$$

Similarly, we find that the three-momentum density $\vec{S}$ of the electromagnetic field is

$$S^i = T^{0i}_{em} = T^{i0}_{em} = (\vec{E} \times \vec{B})_i, \quad \text{EM field momentum density.} \tag{A.168}$$

The vector

$$\vec{S} \equiv \vec{E} \times \vec{B}, \tag{A.169}$$

is known as the *Poynting vector*.

Setting $\mu = 0$ in Eq. A.158, we have

$$\partial_0 T^{00}_{em} + \partial_i T^{i0}_{em} = -F^{0i} j_i, \tag{A.170}$$

and equivalently,

$$\frac{\partial w}{\partial t} + \vec{\nabla} \cdot \vec{S} = -\vec{E} \cdot \vec{j}. \tag{A.171}$$

Bibliography

[1] S. Weinberg, *The Quantum Theory of Fields, Volume I: Foundations*, Cambridge University Press (1995).

[2] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley (1995).

[3] A. Zee, *Quantum Field Theory in a Nutshell*, Princeton University Press (2003).

[4] M. Srednicki, *Quantum Field Theory*, Cambridge University Press (2007).

[5] G. Sterman, *An Introduction to Quantum Field Theory*, Cambridge University Press (1993).

[6] H. Goldstein, C. Poole and J. Safko, *Classical Mechanics*, 3rd edition, Addison-Wesley (2002).

[7] A. Grozin, *Lectures on QED and QCD: Practical Calculation and Renormalization of One- and Multi-Loop Feynman Diagrams*, World Scientific (2007).

[8] S. K. Lamoreaux, *Experimental verifications of the Casimir attractive force between solid bodies*, arXiv:quant-ph/9907076.

[9] K. A. Milton, *The Casimir effect: physical manifestations of zero-point energy*, lectures at the 17th Symposium on Theoretical Physics, Seoul National University (1998), arXiv:hep-th/9901011.