

Notes on Quantum Field Theory I

Marco Serone

SISSA, via Bonomea 265, I-34136 Trieste, Italy

Last update: August 17, 2022

Contents

1	Introduction	3
2	General Properties of QFT	7
2.1	The Källén-Lehmann Spectral Representation	8
2.1.1	Asymptotic Theory: a Brief Reminder	8
2.1.2	Spectral Representation	10
2.2	The Cluster Decomposition Principle and the Connected S -Matrix	14
2.3	The Reduction Formula for Connected S -Matrix Elements	16
2.4	The Optical Theorem	19
2.4.1	Perturbative Unitarity	21
2.5	Unstable Particles	24
2.6	Causality and Analyticity*	26
2.7	Bound States and Resonances**	32
3	Renormalization Theory	38
3.1	Superficial Degree of Divergence	39
3.2	Absolute Convergence of a Graph	43
3.3	Cancellation of Divergences and Local Counterterms	44
3.4	Regularization and Renormalization	47
3.4.1	QED Case	50
3.5	Dimensional Regularization	54
4	External Fields and Generating Functionals	57
4.1	The 1PI Effective Action	57
4.2	The Coleman-Weinberg Effective Potential	61
4.3	A Subtlety about Effective Potentials**	63
4.4	Functional Relations	65
4.4.1	Schwinger-Dyson Equations	65

4.4.2	Symmetries and Ward-Takahashi Identities	66
4.4.3	WT Identities in QED	69
5	The Renormalization Group	74
5.1	Relevant, Marginal and Irrelevant Couplings	75
5.2	The Sliding Scale and the Summation of Leading Logs	79
5.3	Minimal Subtraction	82
5.4	Renormalization Scheme Dependence	84
5.4.1	Mass Threshold Effects	86
5.5	Asymptotic Behaviours of β -Functions	87
5.5.1	CFTs and RG Flows	90
5.6	The Callan-Symanzik RG Equations	93
5.6.1	Leading Logs and Callan-Symanzik Equations	96
5.7	Anomalous Dimension of the Photon and QED β -function	97
5.8	“Irrelevant” RG Flow of Dimensionful Couplings	99
5.9	Renormalization of Composite Operators**	103
5.10	“Relevant” RG Flow of Dimensionful Couplings	110
5.11	RG Improved Effective Potential	114
6	Non-Abelian Gauge Theories	118
6.1	Introduction and Classical Analysis	118
6.2	Quantum Treatment: the Faddeev-Popov Method	125
6.3	BRST Symmetry	132
6.4	The Background Field Method	138
6.4.1	The method	139
6.4.2	Two-Point Function of the Background Field: Feynman Rules	140
6.5	Proof of the Renormalizability of Non-Abelian Gauge Theories**	144
6.5.1	The Master Equation**	144
6.5.2	Structure of Divergences**	145
7	Effective Field Theories*	151
7.1	Matching	152
7.2	1PI vs Wilsonian Actions*	153
7.3	Two Scalars*	154
7.4	Yukawa Theory I: Heavy Scalar, Light Fermion*	156
7.5	Yukawa Theory II: Heavy Fermion, Light Scalar*	158
7.6	Naturalness and the Hierarchy Problem*	159

7.7	Non-Leptonic Decays**	160
7.7.1	Useful Color and Spinor identities**	165
7.8	(Ir)relevance of Higher Dimensional Operators*	167
7.9	Redundant Operators*	168
8	Spontaneously Broken Symmetries*	172
8.1	Why Spontaneous Symmetry Breaking?*	173
8.2	The Goldstone Theorem*	175
8.3	Vacuum Alignment and Pseudo-Goldstone Bosons*	178
8.4	Spontaneously Broken Gauge Symmetries: the Higgs Mechanism*	180
8.5	The Goldstone Boson Equivalence Theorem**	184
8.6	Effective Field Theories for Broken Symmetries*	186
8.6.1	Adding Gauge Fields*	190
8.7	$SU(3)_V \times SU(3)_A \rightarrow SU(3)_V$: Mesons in QCD*	191
8.8	$SO(5) \rightarrow SO(4)$: A Composite Higgs?**	195
8.9	Effective Field Theories for Broken Symmetries: Generale Case**	200
9	Anomalies*	203
9.1	The $U(1)_A$ Chiral Anomaly from One-Loop Graphs*	204
9.2	Gauge Anomalies*	211
9.3	A Relevant Example: Cancellation of Gauge Anomalies in the SM*	214
9.4	Path Integral Derivation of the Chiral Anomaly**	216
9.5	The Wess-Zumino Consistency Conditions**	219
9.6	't Hooft Anomaly Matching and the Wess-Zumino-Witten Term*	221
9.7	Anomalous Breaking of Scale Invariance**	223
9.8	The Strong CP Problem and a Possible Solution: Axions*	227
10	Some Formal Developments*	231
10.1	Asymptotic Nature of Perturbation Theory*	231
10.1.1	Asymptotic Series and Optimal Truncation*	233
10.1.2	Borel Summation*	235
10.2	Euler-Heisenberg Effective Action and Vacuum Decay*	236
10.2.1	Landau Levels by Path Integral*	237
10.2.2	Effective Action for Scalar QED*	240
10.2.3	Effective Action for Fermion QED**	247
10.2.4	Vacuum Instability for a Constant Electric Field*	249

11 Final Project: The Abelian Higgs Model**	252
11.1 The Abelian Higgs Model in a generic ξ -gauge	252
11.2 One-loop Effective Potential**	253
11.3 The Quantum Effective Action**	255
11.4 RG equations and Their Solutions**	259

Chapter 1

Introduction

Quantum Field Theory (QFT) is the fundamental tool used for the description of physics at very short distances and high energies. Since high energy implies relativistic motion, QFT has to nicely combine special relativity with quantum mechanics. The description in terms of “fields”, indeed, arise to have a manifestly relativistic invariant description, where time and space are treated (almost) on equal footing. One might think that including special relativity in quantum mechanics should be possible without drastic consequences. This is not true, and its reason is intuitively clear. At very short time scales, the energy-time uncertainty principle tells us that particles with energy $E \geq mc^2$ could be created from the vacuum for a time $t \sim \hbar/E$ (virtual particles), before disappearing again in the vacuum. This effect is totally negligible in studying physical systems at low energies and long time scales, but it becomes relevant for physical processes whose time scale is $t \sim \hbar/E$ or less. A quantum description in terms of single-particle wave function is then inadequate and a more powerful description is needed. This formulation is in fact Quantum Field Theory. It leads to striking consequences, such as the prediction of anti-particles and an understanding of the spin-statistic relation between particles. The experimental successes of QFT are impressive, in particular when applied to Quantum Electro Dynamics (QED).

These notes *do assume* that the reader has a basic knowledge of QFT. Quantizations of spin 0, spin 1/2 and abelian spin 1 fields are assumed, as well as basic notions of the path integral formulation of QFT (including Berezin integration for fermions), Feynman rules, basic knowledge of renormalization and the notion of functional generators of disconnected, connected and one-particle irreducible (1PI) Green functions.

Most of the considerations in these lectures are devoted to the study of fields which are weakly interacting, namely in which the interactions can be studied in a perturbative fashion, starting from the description of free fields. Non-perturbative phenomena, as well

as some more advanced topics, are discussed in the QFT II course.

These notes are structured as follows. We start in chapter 2 by discussing general properties of QFT, in particular cluster decomposition, the reduction formulae, the optical theorem and the relation between causality and analyticity of scattering amplitudes. Most of the results in this chapter do not make use of perturbation theory and are valid at the non-perturbative level. Chapter 3 is devoted to the regularization and renormalization of a QFT. Assuming that the reader is familiar with the basics of the standard treatment of perturbation theory in QFT, and of Feynman diagrams, we discuss how UV divergences are fixed in an arbitrary QFT. In chapter 4 we introduce generating functionals and external fields, the Coleman-Weinberg potential, and the consequences of symmetries in QFT (Ward-Takahashi identities). Chapter 5 is dedicated to the renormalization group (RG), which is treated in some detail. After having introduced the notion of the RG evolution, we discuss its uses from a perturbative high energy point of view, the notion of beta-functions and anomalous dimensions, the latter for both elementary and composite operators, RG flows, and renormalization scheme dependences. Non-abelian gauge theories are discussed in chapter 6. After a brief recap of basic classical properties, we study their quantization, introduce the notion of BRST symmetry and of the background field method, and finally prove the renormalizability of non-abelian gauge theories by means of the so called master equation. Chapter 7 is dedicated to effective field theories. After having introduced the basics of effective field theories, a series of examples are illustrated to explain various notions, including matching, naturalness, the special role played by dimensional regularization in effective field theories, redundant operators. The phenomenon of spontaneous symmetry breaking is discussed in chapter 8. We discuss why it can occur in a QFT, the Goldstone theorem, the Higgs mechanism, and the equivalence theorem. The use of effective field theories for Goldstone bosons is described, with two examples worked out, mesons in QCD and a possible origin of the Higgs field as a pseudo Goldstone boson. Anomalies are studied in chapter 9. We consider in order chiral and gauge anomalies, the Wess-Zumino consistency conditions, 't Hooft anomaly matching conditions, anomalies in scale invariance, and finally the strong CP problem in QCD and its possible solution in terms of axions. Chapter 10 is dedicated to two more formal aspects of QFT: i) a brief discussion about the asymptotic nature of the perturbative series in QFT and how to deal with it, and ii) an example of a non-perturbative computation in QFT without the use of Feynman diagrams (but rather by means of an auxiliary quantum mechanical system) to get the Euler-Heisenberg Lagrangian and the decay rate of the vacuum in presence of an external electric field. Finally, chapter 11 discusses various aspects of the abelian Higgs model, in particular the phenomenon of dimensional transmutation. This last chapter should be seen

as a sort of long exercise in which many of the notions and techniques introduced in the lecture notes (effective potential, background field method, ghosts, gauge-fixing, Callan-Symanzik equations, beta-functions and anomalous dimensions) are considered together in a notable QFT model.

The tutorials and exercises, essential parts of the QFT I course, are not reported in these notes. This implies that some technical topics closely associated with the exercises, such as details of the renormalization of QED, Yukawa and other theories, are *intentionally not included*.

The notes are far from being comprehensive. Due to lack of time, several important topics are not covered at all, or just mentioned. Infra-red divergences, soft theorems, the deep inelastic scattering and the operator product expansion, connection to critical phenomena and statistical field theory are examples of important topics currently missing. Phenomenological applications are limited to a minimum, since they are more systematically considered in the Standard Model course.

The topics marked with a * in the text are optional for students in the Astroparticle curriculum. The topics marked with ** are optional for *all* the students.

In writing these notes we have often consulted refs. [49] and especially [60]. Some parts of these notes follow either ref. [49] or ref. [60]. When this is the case, we warn the reader with a footnote. *Notice that in these notes the metric convention is mostly minus, like in ref. [49], in contrast to ref. [60], where it is mostly plus. This implies a multitude of sign changes with respect to ref. [60].* Moreover, we warn the reader that we do not always follow the same notation as refs. [49, 60], in order to have a common and light notation throughout the notes.

These notes cover most of the QFT I course given at SISSA by M.S., in collaboration with Roberto Iengo until the academic year 2010-2011, in collaboration with Andrea Gambassi until 2014-2015, in collaboration with Joan Elias-Miró until 2017-2018 and in collaboration with David Marzocca until 2021-2022.

These notes are *preliminary* and surely contain many typos, imprecisions, etc. We hope that the students will help us in improving the notes and in spotting the many mistakes in there.

Acknowledgments

I thank Roberto Iengo and Andrea Gambassi for their contribution in the elaboration of part of these lecture notes at their early stage. In particular, I am grateful to Andrea for having contributed to the writing of the early versions of sections 2.1, 2.3, 2.4, 3.5, 4.4,

6.4, 10.2 and for figs. 2.1, 2.2 and 4.2. I thank Roberto Iengo, Andrea Gambassi, Joan Elias-Miró and David Marzocca for their precious collaboration over the years in giving this course. I thank Matteo Bertolini, Marcello D. Caio, Lorenzo Di Pietro, Matthijs Hogervorst, Marco Letizia, Himanshu Raj, Andrea Romanino, Giacomo Sberverglieri for useful comments and help in debugging these notes. I am particularly grateful to Marco Gorghetto for having found so many typos!

Chapter 2

General Properties of QFT

Quantum field theory aims at describing the processes which occur as a result of the interaction between elementary particles, mediated by a certain number of force carriers. The very same requirement of Lorentz covariance (i.e., compatibility with the theory of special relativity) is a fundamental ingredient of the theory and actually motivated the introduction of concepts such as antiparticles which were later on detected experimentally. Most of our knowledge of the interaction between elementary particles comes from the prototypical scattering experiment in which incoming particles are well-separated in space before getting close to each other in such a way to interact and scatter. After scattering has occurred, the particles resulting from the process typically separate again in space in such a way that their interaction is negligible and their properties as isolated particles, e.g., mass, can be determined via suitable detectors. In this chapter we discuss QFT keeping in mind the framework of scattering experiments in order to define the concept of asymptotic states and assign a specific meaning to the mass of particles from their dynamical properties (e.g., the propagator).

The structure of this chapter is as follows. We start in section 2.1 by discussing the Källén-Lehmann spectral representation of two-point functions. Cluster decomposition and the notion of connected part of the S -matrix are introduced in section 2.2. In section 2.3 we derive the fundamental reduction formulae, relating connected S -matrix elements to n -point correlation functions. The optical theorem and unstable particles are discussed next in sections 2.4 and 2.5, respectively. The important relation between causality and analyticity of S -matrix elements is introduced in section 2.6. Finally, in section 2.7 we consider a simple one-dimensional quantum mechanical model where the analytic structure of the S -matrix, as well as bound states and resonances (unstable states), can be computed explicitly. Most of the results in this chapter do not make use of perturbation theory and

are valid at the non-perturbative level.

2.1 The Källén-Lehmann Spectral Representation

2.1.1 Asymptotic Theory: a Brief Reminder

In a nutshell, quantum field theory can be thought of as based on the following elements: (a) the possible (quantum) states of the theory are generated from a (unique) vacuum state $|0\rangle$ by the action of *free fields* $\phi_{in}(x)$, which generates the Fock space of states (and which we assume to be a real scalar); (b) physical observables — such as the interacting field $\phi(x)$ can be all expressed in terms of $\phi_{in}(x)$. The basic idea behind this setting is that $\phi(x)$ as well as all the other observables can be actually obtained from the free fields $\phi_{in}(x)$ by switching adiabatically on and off the interaction as $|x_0| \rightarrow \infty$. This construction is clearly relevant to scattering processes which originate from particles (wavepackets) well-separated in space. Starting from $\phi(x_0, \vec{x})$ one should recover $\phi_{in}(x_0, \vec{x})$ as $x_0 \rightarrow -\infty$, but this generally occurs up to the wave-function renormalization constant $Z_{in}^{1/2}$:

$$\phi(x_0 \rightarrow -\infty, \vec{x}) \rightarrow Z_{in}^{1/2} \phi_{in}(x_0, \vec{x}). \quad (2.1.1)$$

Analogously, $\phi(x_0 \rightarrow +\infty, \vec{x}) \rightarrow Z_{out}^{1/2} \phi_{out}(x_0, \vec{x})$, where ϕ_{out} is a *free* field. The fields $\phi_{in/out}$ satisfy the free Klein-Gordon equation

$$(\square + m^2)\phi_{in/out}(x) = 0, \quad (2.1.2)$$

while the interacting field does not, $(\square + m^2)\phi(x) \neq 0$. In and out free fields can be expressed in terms of the corresponding creation and annihilation operators as

$$\phi_{in/out}(x) = \int \frac{d^3\vec{p}}{\sqrt{(2\pi)^3 2\omega_{\vec{p}}}} \left(a(\vec{p})_{in/out} e^{-ipx} + a^\dagger(\vec{p})_{in/out} e^{ipx} \right), \quad (2.1.3)$$

where $px = \omega_{\vec{p}}x_0 - \vec{p} \cdot \vec{x}$, $p_0 = \omega_{\vec{p}} = \sqrt{\vec{p}^2 + m^2}$. The canonical quantization condition

$$\left[\dot{\phi}_{in/out}(\vec{x}, t), \phi_{in/out}(\vec{y}, t) \right] = -i\delta^{(3)}(\vec{x} - \vec{y}), \quad (2.1.4)$$

where the dot stands for a time derivative, implies the commutation relations

$$[a(\vec{p})_{in/out}, a^\dagger(\vec{q})_{in/out}] = \delta^3(\vec{p} - \vec{q}). \quad (2.1.5)$$

Inverting eq.(2.1.3), we can express the annihilation and creation operators in terms of ϕ and $\dot{\phi}$:

$$a(\vec{p})_{in/out} = i \int d^3\vec{x} \underbrace{\frac{e^{ipx}}{\sqrt{(2\pi)^3 2\omega_{\vec{p}}}}}_{h_{\vec{p}}(x)} \overset{\leftrightarrow}{\partial}_0 \phi_{in(out)}(x), \quad (2.1.6)$$

where $\overleftrightarrow{\partial}_0 \equiv \overrightarrow{\partial}_0 - \overleftarrow{\partial}_0$. The hermitian conjugate of the right-hand side of eq.(2.1.6) clearly gives $a^\dagger(\vec{p})_{in/out}$. Notice that the time-dependence of this expression is only apparent. In general, for any function $h(x)$ satisfying the free Klein-Gordon equation $(\square + m^2)h(x) = 0$,

$$\int d^3\vec{x} h(x) \overleftrightarrow{\partial}_0 \phi_{in(out)}(x) \quad (2.1.7)$$

is time-independent. This is easily shown by taking a time derivative of this expression, the use of the Klein-Gordon equation and an integration by parts. The same result applies if we substitute $\phi_{in(out)}(x)$ by some other solution of the free Klein-Gordon equation.

The *incoming* multiparticle states are generated from the corresponding vacuum state $|0\rangle_{in}$ by the action of the creation operators $a^\dagger(\vec{p})_{in}$ of the free theory, e.g., $|\vec{p}\rangle_{in} = a^\dagger(\vec{p})_{in}|0\rangle_{in}$ for the single-particle state with three-momentum $\vec{p}$.¹ Analogous construction can be done for the *outgoing* states, starting from the corresponding vacuum $|0\rangle_{out}$: $|\vec{p}\rangle_{out} = a^\dagger(\vec{p})_{out}|0\rangle_{out}$. Given that both the incoming and outgoing particle states are nothing but different representations of the free theories, there should be an unitary isomorphism between them, the so-called *S*-matrix: $S|i\rangle_{out} = |i\rangle_{in}$. We will discuss the consequences of the general properties of the *S*-matrix in secs. 2.3 and 2.4. Accordingly, the transition amplitude from an initial state i ($|i\rangle_{in}$) to a final state f ($|f\rangle_{out}$) can be expressed only in terms of incoming states as

$${}_{out}\langle f|i\rangle_{in} = {}_{in}\langle f|S|i\rangle_{in} = S_{f,i}. \quad (2.1.8)$$

The stability of the vacuum requires that $|0\rangle_{in} = |0\rangle_{out} \equiv |0\rangle$ while the kinematic stability of the massive one-particle state implies, in addition, $|\vec{p}\rangle_{in} = |\vec{p}\rangle_{out} \equiv |\vec{p}\rangle$. The isomorphism induced by *S* implies that, at the level of field operators, $\phi_{in}(x) = S\phi_{out}(x)S^{-1}$. The *S*-matrix can be used in order to express all the states in terms of the incoming ones, which we always refer to in what follows, unless specified differently.

Despite the interacting field does not satisfy the free Klein-Gordon equation, the amplitude $\langle 0|\phi(x)|\vec{p}\rangle$ does. Indeed, invariance under translations implies

$$\phi(x) = e^{i\hat{p}x}\phi(0)e^{-i\hat{p}x}, \quad (2.1.9)$$

where $\hat{p}$ is the four-momentum operator, and thus

$$\langle 0|\phi(x)|\vec{p}\rangle = \langle 0|\phi(0)|\vec{p}\rangle e^{-ipx}. \quad (2.1.10)$$

It then follows that

$$(\square + m^2)\langle 0|\phi(x)|\vec{p}\rangle = 0, \quad (2.1.11)$$

¹Note that the normalizations of the creation/annihilation operators and of the single- and multi-particle states introduced here differ from the corresponding ones of ref. [49].

even if $(\square + m^2)\phi(x) \neq 0$. By using eq.(2.1.7) we notice that

$$\int d^3\vec{x} h_{\vec{p}}(x) \overset{\leftrightarrow}{\partial}_0 \langle 0|\phi(x)|\vec{p}\rangle \quad (2.1.12)$$

is time-independent. Equation (2.1.3) fixes the analogue amplitude for in and out fields to be

$$\langle 0|\phi_{in}(x)|\vec{p}\rangle = \frac{e^{-ipx}}{\sqrt{(2\pi)^3 2\omega_{\vec{p}}}}. \quad (2.1.13)$$

Taking the limit $x^0 \rightarrow \pm\infty$ of eq.(2.1.12) and using eq.(2.1.13) we get

$$Z_{in} = Z_{out} \equiv Z, \quad (2.1.14)$$

and hence

$$\langle 0|\phi(x)|\vec{p}\rangle = Z^{1/2} \frac{e^{-ipx}}{\sqrt{(2\pi)^3 2\omega_{\vec{p}}}}. \quad (2.1.15)$$

Note that $\phi_{in}(x)|0\rangle$ generates only the one-particle state, whereas $\phi(x)|0\rangle$ generates, in addition, multiparticle states because of the interaction. As a result, one expects $|\langle \vec{p}|\phi(x)|0\rangle|^2 \leq |\langle \vec{p}|\phi_{in}(x)|0\rangle|^2$ and therefore $|Z| \leq 1$ and, in particular, $|Z|$ can be interpreted as the probability of generating the one-particle state when applying the interacting field $\phi(x)$ to the vacuum state $|0\rangle$. We will determine Z in sec. 2.1.2.

The limit in eq. (2.1.1) has to be understood in a weak sense because at best it can hold at the level of matrix elements between fields well separated in space (in the form of suitable wavepackets), while it cannot hold for operators. Indeed, both the interacting and free fields ϕ have to satisfy the equal-time canonical commutation relations $[\dot{\phi}(x_0, \vec{x}), \phi(x_0, \vec{x}')] = -i\delta^3(\vec{x} - \vec{x}')$ which would imply $Z = 1$ by taking the limit $x_0 \rightarrow -\infty$.

2.1.2 Spectral Representation

Important properties of an interacting quantum field theory emerge from the so-called spectral representation of vacuum expectation values of suitable quantities. Here we consider both the commutator $\langle 0|[\phi(x), \phi(x')]|0\rangle$ and the time-ordered product $\langle 0|T\phi(x)\phi(x')|0\rangle$ of the fields, i.e., the propagator of the theory, where the time-ordering is defined by $T\phi(x)\phi(x') = \theta(x_0 - x'_0)\phi(x)\phi(x') + \theta(x'_0 - x_0)\phi(x')\phi(x)$, being $\theta(x_0 > 0) = 1$ and 0 otherwise (note that in case of Fermionic fields, the exchange of their order introduces an

additional – sign). Both these quantities can be expressed in terms of

$$\begin{aligned}
\langle 0 | \phi(x) \phi(x') | 0 \rangle &= \sum_{n=0}^{\infty} \langle 0 | \phi(x) | n \rangle \langle n | \phi(x') | 0 \rangle = \sum_{n=1}^{\infty} e^{-ip_n(x-x')} |\langle 0 | \phi(0) | n \rangle|^2, \\
&= \int d^4q e^{-iq(x-x')} \sum_{n=1}^{\infty} |\langle 0 | \phi(0) | n \rangle|^2 \delta^4(q - p_n) \\
&= \int \frac{d^4q}{(2\pi)^3} e^{-iq(x-x')} \rho(q).
\end{aligned} \tag{2.1.16}$$

Let us explain the various steps performed in (2.1.16). In the first identity in the first row we introduced the completeness relation

$$\mathbb{I} = \sum_{n=0}^{\infty} |n\rangle \langle n|, \tag{2.1.17}$$

where $|n\rangle$ denotes the particle states with definite (total) momentum p_n (such that $\hat{p}|n\rangle = p_n|n\rangle$). The term $n = 0$ corresponds to the vacuum state, $n = 1$ with the single-particle state, while $n \geq 1$ are multi-particle states. The sum over n stands both for the sum over the particles number and by the integral on the corresponding phase space. More explicitly,

$$\sum_n |n\rangle \langle n| = |0\rangle \langle 0| + \int d^3\vec{p} |\vec{p}\rangle \langle \vec{p}| + \frac{1}{2!} \int d^3\vec{p}_1 d^3\vec{p}_2 |\vec{p}_1 \vec{p}_2\rangle \langle \vec{p}_1 \vec{p}_2| + \dots \tag{2.1.18}$$

In the second identity in the second row we used eq. (2.1.9) and assumed that $\langle 0 | \phi(0) | 0 \rangle = 0$, i.e. that no spontaneous symmetry breaking occurs. In the second line we introduced the identity $1 = \int d^4q \delta^4(q - p_n)$ and, in the third we defined the so-called *spectral density*

$$\rho(q) \equiv (2\pi)^3 \sum_{n=1}^{\infty} |\langle 0 | \phi(0) | n \rangle|^2 \delta^4(q - p_n). \tag{2.1.19}$$

This function is a Lorentz-invariant, positive scalar and therefore can actually be only a function of q^2 . In addition, $p_n = \sum_{i=1}^n p_{n,i}$ where $p_{n,i}$ is the momentum of the i -th particle within the n -particle state $|n\rangle$ and therefore $(p_{n,i})_0 > 0$ with $p_{n,i}^2 > 0$; this implies that $(p_n)_0 > 0$ and $p_n^2 > 0$ and, in turn, that $\rho(q)$ vanishes in the backward light cone: $\rho(q) \mapsto \rho(q^2)\theta(q_0)$ with $\rho(q^2 < 0) = 0$. In eq. (2.1.19), it is convenient to isolate the contribution of the one-particle state $n = 1$, which can be readily calculated by taking into account the normalization (2.1.15):

$$\begin{aligned}
\rho(q^2)|_{n=1} \theta(q^0) &= (2\pi)^3 \theta(q^0) \int d^3\vec{p} |\langle 0 | \phi(0) | \vec{p} \rangle|^2 \delta^4(q - p) \\
&= Z \theta(q^0) \int \frac{d^3\vec{p}}{2\omega_p} \delta^4(q - p) = Z \theta(q_0) \delta(q^2 - m^2)
\end{aligned} \tag{2.1.20}$$

The resulting expression for ρ is

$$\rho(q^2)\theta(q_0) = Z\delta(q^2 - m^2)\theta(q_0) + (2\pi)^3 \sum_{n>1} |\langle 0|\phi(0)|n\rangle|^2 \delta^4(q - p_n), \quad (2.1.21)$$

which reveals its generic structure: it consists of an isolated δ corresponding to the mass of the asymptotic field (including the self-interaction) and of a continuum starting from the threshold m_{th}^2 for the production of multi-particle states. In addition, there might be some bound states in the gap, as depicted in fig. 2.1(a).

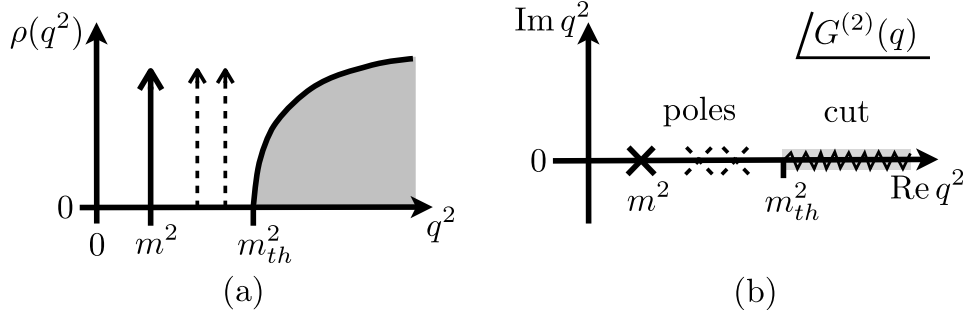


Figure 2.1: Sketches (a) of the spectral density $\rho(q^2)$ as a function of q^2 , which highlights the presence of a δ corresponding to the physical mass, of possible additional peaks related to bound states (dashed) and of a continuum associated to multi-particle states. (b) Analytic structure of the propagator $G^{(2)}(q)$ of a scalar field, as a function of $q^2 \in \mathbb{C}$ which highlights the presence of an isolated pole corresponding to the physical mass, possible additional poles due to bound states and a branch cut related to multi-particle states.

In terms of the spectral density, $\langle 0|[\phi(x), \phi(0)]|0\rangle$ can be written as

$$\begin{aligned} \langle 0|[\phi(x), \phi(0)]|0\rangle &= \int \frac{d^4 q}{(2\pi)^3} \rho(q^2)\theta(q_0) [e^{-iqx} - e^{iqx}] \\ &= \int_0^\infty d\sigma \rho(\sigma) \underbrace{\int \frac{d^4 q}{(2\pi)^3} \delta(q^2 - \sigma)\theta(q_0) [e^{-iqx} - e^{iqx}]}_{i\Delta_c^0(x;\sigma)}, \end{aligned} \quad (2.1.22)$$

where we introduced the identity in the form $1 = \int_0^\infty d\sigma \delta(q^2 - \sigma)$. In the previous equation one recognizes the expression of $\langle 0|[\phi_0(x), \phi_0(0)]|0\rangle \equiv i\Delta_c^0(x; m^2)$ for a *free* scalar field

$\phi_0(x)$ of mass m . The analogous expression for $\langle 0|T\phi(x)\phi(0)|0\rangle$ is

$$\begin{aligned}\langle 0|T\phi(x)\phi(0)|0\rangle &= \int \frac{d^4q}{(2\pi)^3} \rho(q^2) \theta(q_0) [\theta(x_0)e^{-iqx} + \theta(-x_0)e^{iqx}], \\ &= \int_0^\infty d\sigma \rho(\sigma) \underbrace{\int \frac{d^4q}{(2\pi)^3} \delta(q^2 - \sigma) \theta(q_0) [\theta(x_0)e^{-iqx} + \theta(-x_0)e^{iqx}]}_{i\Delta_F^0(x;\sigma)},\end{aligned}\tag{2.1.23}$$

and involves, instead, the Feynman propagator $i\Delta_F^0(x; m^2) \equiv \langle 0|T\phi_0(x)\phi_0(0)|0\rangle$ of the same field. Taking into account eq. (2.1.21) one finds

$$\langle 0|[\phi(x), \phi(0)]|0\rangle = Zi\Delta_c^0(x; m^2) + \int_{m_{th}^2}^\infty d\sigma \rho(\sigma) i\Delta_c^0(x; \sigma),\tag{2.1.24}$$

$$\langle 0|T\phi(x)\phi(0)|0\rangle = Zi\Delta_F^0(x; m^2) + \int_{m_{th}^2}^\infty d\sigma \rho(\sigma) i\Delta_F^0(x; \sigma),\tag{2.1.25}$$

which provide the Källén-Lehmann spectral representation of the commutator and the propagator, respectively [39, 42]. Occasionally, it might be useful to introduce *renormalized* fields $\phi_R(x) \equiv Z^{-1/2}\phi(x)$ and the renormalized spectral density ρ_R via $\rho = Z\rho_R$. The consequences of these relations can be easily worked out: for example, by taking the derivative $\partial_0 = \partial/\partial x_0$ of the l.h.s. of eq. (2.1.24) and by eventually setting $x_0 = 0$ one recovers the canonical commutation relation for the field ϕ ; the same relation for the free field ϕ_0 implies that $\partial_0 i\Delta_c^0(x; m^2)|_{x_0=0} = -i\delta^3(\vec{x})$, which appears on the r.h.s. of eq. (2.1.24) and implies

$$1 = Z \left[1 + \int_{m_{th}^2}^\infty d\sigma \rho_R(\sigma) \right].\tag{2.1.26}$$

This expression shows that $Z \leq 1$ ($\rho_R \geq 0$) and allows one to calculate Z from the spectral density. Consider now the Fourier transform of the renormalized propagator $G_R^{(2)}(x) = \langle 0|T\phi_R(x)\phi_R(0)|0\rangle = G^{(2)}(x)/Z$. Recalling that for the free theory

$$\Delta_F^0(x, m^2) = \int \frac{d^4q}{(2\pi)^4} \frac{e^{-iqx}}{q^2 - m^2 + i\epsilon},\tag{2.1.27}$$

one finds

$$G_R^{(2)}(q^2) = \int_0^\infty d\sigma \rho_R(\sigma) \frac{i}{q^2 - \sigma + i\epsilon} = \frac{i}{q^2 - m^2 + i\epsilon} + \int_{m_{th}^2}^\infty d\sigma \rho_R(\sigma) \frac{i}{q^2 - \sigma + i\epsilon}.\tag{2.1.28}$$

As a function of $q^2 \in \mathbb{C}$, $G_R^{(2)}(q^2)$ is characterized by an isolated pole for $q^2 = m^2$, with residue i and by a branch cut on the positive real axis starting from $q^2 = m_{th}^2$ and controlled by the spectral density, with possible additional poles due to bound states, as depicted in

fig. 2.1(b). Note in particular that $G_R^{(2)}(z)$ is analytic in the complex plane away from the real axis. The $i\epsilon$ prescription in the propagator tells us that the physical sheet lies above the branch-cut. In sec. 2.4 we provide a physical interpretation of the discontinuity which arises in $G_R^{(2)}(q)$ upon crossing the branch cut. Another relevant relation between the propagator of the theory and the spectral density can be found from eq. (2.1.28), taking into account that

$$\frac{1}{x + i\epsilon} = P \frac{1}{x} - i\pi\delta(x) \quad \text{for } \epsilon \rightarrow 0^+ \quad (2.1.29)$$

(in the sense of distributions, being P the principal part) and therefore

$$\text{Im } iG_R^{(2)}(q^2) = \pi\rho_R(q^2). \quad (2.1.30)$$

2.2 The Cluster Decomposition Principle and the Connected S -Matrix

In the previous section we briefly mentioned that the transition amplitudes from an initial state $|i\rangle$ to a final state $|f\rangle$ can be calculated as a matrix element of the S -matrix, as in eq. (2.1.8).² Let us denote by $S_{m \rightarrow n}$ a generic S -matrix element for the transition from an m -particle incoming state to an n -particle outgoing state. A fundamental principle in physics states that two spatially separated events should have no correlation among each other.³ In the context of S -matrix, this principle is called cluster decomposition. If the initial and final m and n particles are grouped in sets $\{m_1, \dots, m_N\}$, $\{n_1, \dots, n_N\}$ and the process $m \rightarrow n$ consists of a set of N different sub-processes $m_i \rightarrow n_i$ ($i = 1, \dots, N$), occurring all far away from each other, then we should demand that

$$S_{m \rightarrow n} = \prod_{i=1}^N S_{m_i \rightarrow n_i}. \quad (2.2.1)$$

We can rephrase eq.(2.2.1) in another way by defining the connected part of the S -matrix, S^C . In perturbation theory, S^C is defined in terms of connected Feynman diagrams, namely $S_{m \rightarrow n}^C$ is given, order by order in perturbation theory, by the diagrams where all

²The analysis in this section closely follows section 4.3 of ref. [60].

³Strictly speaking this statement is not correct, since quantum entanglement among states provide correlations that do not decrease as the distance increases (this was at the base of the so called Einstein-Podolski-Rosen paradox claiming possible violations of causality in quantum mechanics). Such states are however fragile and short-lived, since decoherence effects given by arbitrarily small external perturbations will destroy the entanglement among the states asymptotically. A more precise statement would be that stable long-lived states should satisfy cluster decomposition. In general, cluster decomposition applies when the vacuum is a pure state, and it fails when it is a mixed state. The phenomenon of spontaneous symmetry breaking is also related to cluster decomposition, see section 8.1.

m and n particles are connected with each other. A non-perturbative definition for $S_{m \rightarrow m}^C$ can be given iteratively starting from the $1 \rightarrow 1$ process for which

$$S_{p \rightarrow p'} = S_{p \rightarrow p'}^C = \delta(p - p'), \quad (2.2.2)$$

where $\delta(p - p') \equiv \delta(\vec{p} - \vec{p}')\delta_{s,s'}\delta_{n,n'}$ includes Kronecker or Dirac delta's in momentum, spin and particle species. The $2 \rightarrow 2$ connected matrix element $S_{p_1, p_2 \rightarrow p'_1, p'_2}^C$ is given as

$$\begin{aligned} S_{p_1, p_2 \rightarrow p'_1, p'_2} &= S_{p_1, p_2 \rightarrow p'_1, p'_2}^C + S_{p_1 \rightarrow p'_1}^C S_{p_2 \rightarrow p'_2}^C \pm S_{p_1 \rightarrow p'_2}^C S_{p_2 \rightarrow p'_1}^C \\ &= S_{p_1, p_2 \rightarrow p'_1, p'_2}^C + \delta(p_1 - p'_1)\delta(p_2 - p'_2) \pm \delta(p_1 - p'_2)\delta(p_2 - p'_1), \end{aligned} \quad (2.2.3)$$

where $\pm$ refers to an even or odd exchange of fermion particles. Similarly, the $3 \rightarrow 3$ connected matrix element $S_{p_1, p_2, p_3 \rightarrow p'_1, p'_2, p'_3}^C$ is given as

$$\begin{aligned} S_{p_1, p_2, p_3 \rightarrow p'_1, p'_2, p'_3} &= S_{p_1, p_2, p_3 \rightarrow p'_1, p'_2, p'_3}^C + (S_{p_1, p_2 \rightarrow p'_1, p'_2}^C \delta(p_3 - p'_3) \pm 8 \text{ perms.}) \\ &\quad + (\delta(p_1 - p'_1)\delta(p_2 - p'_2)\delta(p_3 - p'_3) \pm 5 \text{ perms.}), \end{aligned} \quad (2.2.4)$$

where perms. refer to all other possible ways to combine the in-coming and the out-coming particles. By further iterating, we can define all higher connected matrix elements $S_{m \rightarrow m}^C$ with $m > 3$. Similarly, we could define the “off-diagonal” connected elements $S_{m \rightarrow n}^C$ with $m \neq n$. In terms of S^C , cluster decomposition implies that

$$S_{m \rightarrow n}^C = 0, \quad (2.2.5)$$

if any among the $m + n$ particles is far away from the remaining ones. For instance, suppose that a $3 \rightarrow 3$ scattering consists of an actual $2 \rightarrow 2$ scattering involving particles 1 and 2, with particle 3 far away from the interaction place. In this case, we would expect

$$S_{3 \rightarrow 3} = S_{2 \rightarrow 2} S_{1 \rightarrow 1} = S_{p_1, p_2 \rightarrow p'_1, p'_2} \delta(p_3 - p'_3), \quad (2.2.6)$$

with $S_{p_1, p_2 \rightarrow p'_1, p'_2}$ given by eq.(2.2.3). When we impose eq.(2.2.5), namely we demand that all connected S matrix elements where particle 3 enters together with particles 1 and 2 vanish, leaves only three terms in eq.(2.2.4), that indeed recombine in the right-hand-side of eq.(2.2.6).

Since cluster decomposition is formulated in configuration space, we can define a configuration space connected S -matrix $S_{x, x'}^C$ by taking the Fourier transform of $S_{m \rightarrow n}^C$:

$$S_{m \rightarrow n}^C(x_1, \dots, x_m, x'_1, \dots, x'_n) \equiv \int \prod_{i=1}^m d^3 p_i \prod_{j=1}^n d^3 p'_j e^{i\vec{p}_i \cdot \vec{x}_i - i\vec{p}'_j \cdot \vec{x}'_j} S_{m \rightarrow n}^C(p_1, \dots, p_m, p'_1, \dots, p'_n). \quad (2.2.7)$$

Invariance under space-time translations imply that $S_{m \rightarrow n}^C(x, x')$ cannot vary if we shift by a constant four-vector all coordinates. Correspondingly the connected S -matrix should be proportional to a delta function imposing energy-momentum conservation:

$$S_{m \rightarrow n}^C(p_1, \dots, p_m, p'_1, \dots, p'_n) = (2\pi)^4 \delta^{(4)}\left(\sum_i p_i - \sum_j p'_j\right) A^C(p_1, \dots, p_m, p'_1, \dots, p'_n), \quad (2.2.8)$$

with A^C encoding the remaining part of the scattering process. Cluster decomposition implies that

$$S_{m \rightarrow n}^C(x_1, \dots, x_m, x'_1, \dots, x'_n) \rightarrow 0 \quad (2.2.9)$$

if we move apart the in-coming particles $|\vec{x}_i - \vec{x}_j| \rightarrow \infty$, $i = 1, 2, \dots, m_1$, $j = 1, 2, \dots, m - m_1$, and their associated out-going particles $|\vec{x}_k' - \vec{x}_l'| \rightarrow \infty$, $k = 1, 2, \dots, n_1(m_1)$, $l = 1, 2, \dots, n - n_1(m_1)$, for any choice of m_1 .

No other delta functions imposing further constraints on the initial and outgoing momenta can enter in A^C . If such a constraint would exist, we could move all the particles subject to the constraint away from the others without affecting eq.(2.2.7), violating the cluster decomposition property (2.2.9). In contrast, the disconnected element of the full S -matrix can and do have multiple delta functions. This shows that the connected and disconnected elements of the S -matrix cannot interfere among each other. We will mostly focus in what follows to the connected part of the S -matrix, but it should be emphasized that disconnected S -matrix elements are not negligible, since in actual experiments the scattering process is often of this type.

2.3 The Reduction Formula for Connected S-Matrix Elements

Scattering experiments are determined by the S -matrix. However, quantum field theory naturally provides access to time-ordered products of fields and operators and therefore it is important to understand how to extract from the latter the S -matrix elements. This connection is provided by the so-called Lehmann-Symanzik-Zimmermann (LSZ) reduction formulas [43, 44] which we discuss here for the case of a neutral scalar field.

Consider a generic S -matrix element for the transition from the m -particle incoming state $|\vec{p}_1, \dots, \vec{p}_m\rangle_{in}$ to the n -particle outgoing state $|\vec{q}_1, \dots, \vec{q}_n\rangle_{out}$:

$$S_{m \rightarrow n} = {}_{out}\langle \vec{q}_1, \dots, \vec{q}_n | \vec{p}_1, \dots, \vec{p}_m \rangle_{in} = {}_{in}\langle \vec{q}_1, \dots, \vec{q}_n | S | \vec{p}_1, \dots, \vec{p}_m \rangle_{in}, \quad (2.3.1)$$

where $|\vec{p}_1, \dots, \vec{p}_m\rangle_{in} = a^\dagger(\vec{p}_1)_{in} |\vec{p}_2, \dots, \vec{p}_m\rangle_{in}$ (see sec. 2.1.1). In turn, $a(\vec{p})_{in(out)}$ (and $a^\dagger(\vec{p})_{in(out)}$) can be expressed in terms of the incoming (outgoing) free field $\phi_{in(out)}$ as

in eq.(2.1.6). We now write $a_{\vec{p},in}^\dagger = (a_{\vec{p},in}^\dagger - a_{\vec{p},out}^\dagger) + a_{\vec{p},out}^\dagger$ and therefore

$$S_{m \rightarrow n} = {}_{out} \langle \vec{q}_1, \dots, \vec{q}_n | (a^\dagger(\vec{p}_1)_{in} - a^\dagger(\vec{p}_1)_{out}) | \vec{p}_2, \dots, \vec{p}_m \rangle_{in} + {}_{out} \langle \vec{q}_1, \dots, \vec{q}_n | a^\dagger(\vec{p}_1)_{out} | \vec{p}_2, \dots, \vec{p}_m \rangle_{in}, \quad (2.3.2)$$

where the second term on the r.h.s. does not vanish only if $\vec{p}_1 \in \{\vec{q}_1, \dots, \vec{q}_n\}$, i.e., if it coincides with one of the outgoing momenta. This contribution represents a process in which one of the incoming particles does not participate in the scattering and can be neglected if we are only interested in connected S -matrix elements. In the first contribution, instead, one can use eq. (2.1.6) in order to express $a^\dagger(\vec{p}_1)_{in}$ and $a^\dagger(\vec{p}_1)_{out}$ in terms of $\phi_{in}(x)$ and $\phi_{out}(x)$, respectively, and the latter, in turn, as $\lim_{x_0 \rightarrow -\infty} Z^{-1/2} \phi(x)$ and $\lim_{x_0 \rightarrow +\infty} Z^{-1/2} \phi(x)$ [see eq. (2.1.1)]:

$$S_{m \rightarrow n}^C = -iZ^{-1/2} \int d^3 \vec{x} \left(\lim_{x_0 \rightarrow -\infty} - \lim_{x_0 \rightarrow +\infty} \right) {}_{out} \langle \vec{q}_1, \dots, \vec{q}_n | h_{\vec{p}_1}^*(x) \overset{\leftrightarrow}{\partial}_0 \phi(x) | \vec{p}_2, \dots, \vec{p}_m \rangle_{in}^C, \quad (2.3.3)$$

where the superscript C in the bracket stands for the connected part of the amplitude, which for simplicity will be omitted in what follows. The difference between the limits can be obtained as $-\int_{-\infty}^{+\infty} dx_0 \partial_0$ which renders an integration over space-time $\int d^4x$, while $\partial_0[h_{\vec{p}_1}^*(x) \overset{\leftrightarrow}{\partial}_0 \phi(x)] = h_{\vec{p}_1}^*(x)(\square_x + m^2)\phi(x)$ and therefore

$$S_{m \rightarrow n}^C = iZ^{-1/2} \int d^4x h_{\vec{p}_1}^*(x)(\square_x + m^2) {}_{out} \langle \vec{q}_1, \dots, \vec{q}_n | \phi(x) | \vec{p}_2, \dots, \vec{p}_m \rangle_{in}, \quad (2.3.4)$$

which has been now reduced to the matrix element of the field $\phi(x)$ with one less particle in the incoming state. An analogous reduction can be done for the outgoing state by expressing ${}_{out} \langle \vec{q}_1, \dots, \vec{q}_n |$ as ${}_{out} \langle \vec{q}_2, \dots, \vec{q}_n | a(\vec{q}_1)_{out}$ and by using the fact that $a(\vec{q}_1)_{out} \phi(x) = [a(\vec{q}_1)_{out} \phi(x) - \phi(x) a(\vec{q}_1)_{in}] + \phi(x) a(\vec{q}_1)_{in}$; as before, when inserted in eq. (2.3.4), the second term on the r.h.s. of this identity corresponds to a disconnected amplitude and can be neglected. The first, instead, can be simplified by expressing $a(\vec{q}_1)_{out/in}$ via eq. (2.1.6) and then by introducing the suitable (weak) limits in order to replace ϕ by $\phi_{in/out}$

$$a(\vec{q}_1)_{out} \phi(x) - \phi(x) a(\vec{q}_1)_{in} = iZ^{-1/2} \int d^3 \vec{y} \left[\lim_{y_0 \rightarrow +\infty} h_{\vec{q}_1}(y) \overset{\leftrightarrow}{\partial}_{y_0} \phi(y) \phi(x) - \lim_{y_0 \rightarrow -\infty} h_{\vec{q}_1}(y) \overset{\leftrightarrow}{\partial}_{y_0} \phi(x) \phi(y) \right]; \quad (2.3.5)$$

due to the presence of the limits, both $\phi(y)\phi(x)$ and $\phi(x)\phi(y)$ can be replaced by their time-ordered product $T\phi(x)\phi(y)$ and therefore the difference between the limits, as before, can be expressed as $\int_{-\infty}^{+\infty} dy_0 \partial_{y_0}$, while $\partial_{y_0}[h_{\vec{q}_1}(y) \overset{\leftrightarrow}{\partial}_{y_0} T\phi(x)\phi(y)] = h_{\vec{q}_1}(y)(\square_y + m^2)T\phi(x)\phi(y)$, with the result

$$\begin{aligned} {}_{out} \langle \vec{q}_1, \dots, \vec{q}_n | \phi(x) | \vec{p}_2, \dots, \vec{p}_m \rangle_{in} &= \\ &= iZ^{-1/2} \int d^4y h_{\vec{q}_1}(y)(\square_y + m^2) {}_{out} \langle \vec{q}_2, \dots, \vec{q}_n | T\phi(x)\phi(y) | \vec{p}_2, \dots, \vec{p}_m \rangle_{in}, \end{aligned} \quad (2.3.6)$$

in which the number of particles in the outgoing state has been reduced by one. Together with eq. (2.3.4) one finds

$$S_{m \rightarrow n}^C = (iZ^{-1/2})^2 \int d^4x d^4y h_{\vec{p}_1}^*(x) h_{\vec{q}_1}(y) (\square_x + m^2) (\square_y + m^2) \times_{out} \langle \vec{q}_2, \dots, \vec{q}_n | T\phi(x)\phi(y) | \vec{p}_2, \dots, \vec{p}_m \rangle_{in}, \quad (2.3.7)$$

in which the procedure outlined above can be iteratively repeated till all the particles in the incoming and outgoing states are reduced to zero and one is left with the vacuum state $|0\rangle_{in} = |0\rangle_{out} = |0\rangle$:

$$S_{m \rightarrow n}^C = (iZ^{-1/2})^{m+n} \int \prod_{i=1}^m d^4x_i h_{\vec{p}_i}^*(x_i) (\square_{x_i} + m^2) \prod_{j=1}^n d^4y_j h_{\vec{q}_j}(y_j) (\square_{y_j} + m^2) \times \langle 0 | T\phi(x_1) \dots \phi(x_m) \phi(y_1) \dots \phi(y_n) | 0 \rangle, \quad (2.3.8)$$

which is the LSZ reduction formula.⁴

In order to connect the r.h.s. of this equation to the N -point correlation function $\langle 0 | T\phi(x_1) \dots \phi(x_N) | 0 \rangle$ of the fields which is usually calculated in quantum field theory, we focus on its Fourier transform

$$G^{(N)}(p_1, \dots, p_N) = \int \prod_{k=1}^N d^4x_k e^{-ip_k x_k} \langle 0 | T\phi(x_1) \dots \phi(x_N) | 0 \rangle; \quad (2.3.9)$$

Due to the space-time translational invariance of the theory, the correlation function on the r.h.s. depends only on the difference of the coordinates and therefore one can always factor out $(2\pi)^4 \delta^4(\sum_{k=1}^N p_k)$ and actually indicate by $G^{(N)}$ the remaining factor. Taking into account the definition of $h_{\vec{p}}(x)$ in eq. (2.1.6) one easily recognizes in eq. (2.3.8) $(2\pi)^4 \delta^4(\sum_{i=1}^m p_i - \sum_{j=1}^n q_j) G^{(n+m)}(p_1, \dots, p_m, -q_1, \dots, -q_n)$, while $\square_{x_i}$ and $\square_{y_j}$ turn into $-p_i^2$ and $-q_j^2$ when acting to the left on $h_{\vec{p}_i}^*(x_i)$ and $h_{\vec{q}_j}(y_j)$, respectively. Note, however, that according to eq. (2.1.6) $p_i^2 = q_j^2 = m^2$, being the incoming and outgoing particles on-shell, and therefore the factors $\square_{x_i/y_j} + m^2$ vanish; it is then convenient to think of them as limits:

$$S_{m \rightarrow n}^C = (2\pi)^4 \delta^4(\sum_{i=1}^m p_i - \sum_{j=1}^n q_j) \prod_{i=1}^m \frac{1}{\sqrt{(2\pi)^3 2\omega_{\vec{p}_i}}} \prod_{j=1}^n \frac{1}{\sqrt{(2\pi)^3 2\omega_{\vec{q}_j}}} \times \prod_{i=1}^m \lim_{p_i^2 \rightarrow m^2} \frac{p_i^2 - m^2}{iZ^{1/2}} \prod_{j=1}^n \lim_{q_j^2 \rightarrow m^2} \frac{q_j^2 - m^2}{iZ^{1/2}} G^{(n+m)}(p_1, \dots, p_m, -q_1, \dots, -q_n). \quad (2.3.10)$$

⁴Note that in some textbooks, e.g., ref. [49] the factors $h_{\vec{p}}$ are actually absorbed in the definition of the multiparticle states and therefore they are absent in the formula above.

This equation clearly shows that connected S -matrix elements are basically given by the residues of the poles of the Fourier transform of the T -product of fields (i.e., of their Green functions) as the incoming and outgoing particles go on-shell. As we discussed in sec. 2.1.2, $G^{(2)}(p^2 \rightarrow m^2) = iZ/(p^2 - m^2)$, which allows one to identify the factors $\lim_{p^2 \rightarrow m^2} (p^2 - m^2)/(iZ^{1/2})$ in eq. (2.3.10) as $\sqrt{Z}[G^{(2)}(p^2 \rightarrow m^2)]^{-1}$. In view of this fact it is convenient to introduce the notion of *amputated* Green function $G_{amp}^{(N)}$, which in terms of Feynman diagrams, is obtained by stripping the diagram corresponding to $G^{(N)}$ of the external propagators (the latter being the cause of the presence of the poles mentioned above):

$$G^{(N)}(p_1, \dots, p_N) = \left[\prod_{i=1}^N G^{(2)}(p_i) \right] G_{amp}^{(N)}(p_1, \dots, p_N). \quad (2.3.11)$$

Accordingly, eq. (2.3.10) becomes

$$\begin{aligned} S_{m \rightarrow n}^C &= (2\pi)^4 \delta^4 \left(\sum_{i=1}^m p_i - \sum_{j=1}^n q_j \right) \prod_{i=1}^m \frac{1}{\sqrt{(2\pi)^3 2\omega_{\vec{p}_i}}} \prod_{j=1}^n \frac{1}{\sqrt{(2\pi)^3 2\omega_{\vec{q}_j}}} \\ &\times Z^{(n+m)/2} G_{amp}^{(n+m)}(p_1, \dots, p_m, -q_1, \dots, -q_n) \big|_{p_i^2 = q_j^2 = m^2}. \end{aligned} \quad (2.3.12)$$

Quantum field theory provides access (e.g., perturbatively) to $G_{amp}^{(N)}$ on the r.h.s. and therefore, via this relation, to $S_{m \rightarrow n}^C$ which has to be supplemented by appropriate phase space (i.e., kinematic) factors in order to calculate cross sections.

An important consequence of eq. (2.3.12) is the *crossing symmetry*: in fact, incoming particles are distinguished from the outgoing ones only by the sign of the four-momentum of the argument of $G_{amp}^{(n+m)}$. When the theory involves also antiparticles, the construction can be straightforwardly generalized in order to conclude that an incoming antiparticle with momentum $\vec{p}$ [and four-momentum $p = (-\omega_{\vec{p}}, \vec{p})$] is equivalent to an outgoing particle with momentum $-\vec{p}$ [i.e., with four momentum $(\omega_{\vec{p}}, -\vec{p}) = -p$].

2.4 The Optical Theorem

The S -matrix is a unitary mapping between the space of the incoming asymptotic states and the one of the outgoing states. As such, it satisfies $S^\dagger S = S S^\dagger = 1$. This unitarity has important consequences, which we investigate here.

In order to factor out from the S -matrix the contribution corresponding to the absence of interaction between the incoming and the outgoing particles, one usually introduces the T -matrix:

$$S = 1 + iT, \quad (2.4.1)$$

in terms of which, the condition of unitarity becomes

$$T - T^\dagger = iT^\dagger T. \quad (2.4.2)$$

Consider now a scattering process from an initial state $|i\rangle$ to a final state $|f\rangle$: the previous relation implies that

$$T_{f,i} - T_{f,i}^\dagger = i \sum_n T_{f,n}^\dagger T_{n,i} \quad (2.4.3)$$

for the T -matrix element $T_{f,i} \equiv \langle f|T|i\rangle$ between these states, where the sum runs over all possible intermediate states with definite momentum. Note that the conservation of energy and momentum requires generically

$$T_{a,b} = \langle a|T|b\rangle = (2\pi)^4 \delta^4(p_a - p_b) A_{a,b} \quad (2.4.4)$$

where p_a and p_b are the momenta of states $|a\rangle$ and $|b\rangle$, respectively, and $A_{a,b}$ is the remaining factor usually calculated via Feynman diagrams. In terms of $A_{a,b}$, eq. (2.4.3) becomes

$$A_{f,i} - A_{i,f}^* = i \sum_n (2\pi)^4 \delta^4(p_n - p_{f(i)}) A_{n,f}^* A_{n,i}. \quad (2.4.5)$$

This relation has a very simple interpretation for $f = i$, i.e., when focussing on the forward scattering amplitude because in this case the l.h.s. is $2i \operatorname{Im} A_{i,i}$, while the r.h.s. involves $|A_{n,i}|^2$:

$$2 \operatorname{Im} A_{i,i} = \sum_n (2\pi)^4 \delta^4(p_n - p_i) |A_{n,i}|^2. \quad (2.4.6)$$

Consider the case in which the initial state $|i\rangle = |\vec{k}_1, \vec{k}_2\rangle$ is a two-particle state: the r.h.s of eq. (2.4.6) is then the sum over all possible final states of the probability to obtain these states from the scattering of the two initial particles. This is related to the total cross section σ_T . In our conventions, we have⁵

$$\sigma_T = \frac{(2\pi)^6 E_1 E_2}{\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}} \sum_n \prod_{i=1}^n \int d^3 p'_i (2\pi)^4 \delta(p_1 + p_2 - \sum_i p'_i) |A_{i,n}|^2. \quad (2.4.7)$$

In the center of mass frame of the two particles we get

$$\frac{E_{cm} |\vec{k}_{cm}|}{(2\pi)^6 E_1 E_2} \sigma_T = \sum_n (2\pi)^4 \delta^4(p_1 + p_2 - p'_n) |A_{n,i}|^2, \quad (2.4.8)$$

where we have rewritten more compactly the phase space integral as $\sum_n$. Accordingly, the optical theorem implies that

$$2 \operatorname{Im} A_{i,i} = \frac{E_{cm} |\vec{k}_{cm}|}{(2\pi)^6 E_1 E_2} \sigma_T. \quad (2.4.9)$$

⁵This expression differ from the one presented, e.g., in eq. (4.79) of ref. [49] because of the different conventions for the normalization of the states.

Further below we shall check explicitly this relation. On this specific case of a two-particle initial state we can discuss a general property of the amplitude $A_{i,i}$ as a function of the Mandelstam variable $s = E_{cm}^2$ which we let be a complex number. First of all we note that the diagrams which contribute to $A_{i,i}$ in a perturbative expansion are generically *real* as long as one can forget about the Feynman prescription of adding $+i\epsilon$ at the denominators of the propagators of the particles involved (think, for example, of scalars). This is the case if none of the intermediate virtual particles involved in the diagram goes on-shell: in fact, being the denominator of the corresponding propagator always non-vanishing, one can safely take $\epsilon \rightarrow 0$, which makes the imaginary part of the corresponding diagram vanish. Accordingly, the amplitude $A_{i,i}(s)$ is a real function of s for real $s < m_{th}^2$, where m_{th}^2 is the energy threshold for the production of on-shell multiparticle states:

$$A_{i,i}^*(s) = A_{i,i}(s^*). \quad (2.4.10)$$

Since $A_{i,i}$ is real on the real axis for $s < m_{th}^2$, its continuation satisfies eq. (2.4.10), i.e.,

$$\begin{cases} \operatorname{Re} A_{i,i}(s) = \operatorname{Re} A_{i,i}(s^*), \\ \operatorname{Im} A_{i,i}(s) = -\operatorname{Im} A_{i,i}(s^*). \end{cases} \quad (2.4.11)$$

As a consequence of the continuation outside the real axis, $A_{i,i}(s)$ acquires an imaginary part: upon approaching the real axis $s \in \mathbb{R}$ from above with $s + i\epsilon$ and $\epsilon = 0^+$ or from below with $s - i\epsilon$, eq. (2.4.11) implies that $A_{i,i}(s)$ has opposite imaginary parts $\operatorname{Im} A_{i,i}(s + i\epsilon) = -\operatorname{Im} A_{i,i}(s - i\epsilon)$ and therefore, upon crossing the real axis, there is a discontinuity

$$\operatorname{Disc} A_{i,i}(s) = A_{i,i}(s + i\epsilon) - A_{i,i}(s - i\epsilon) = 2i \operatorname{Im} A_{i,i}(s + i\epsilon), \quad (2.4.12)$$

which vanishes along the real axis with $s < m_{th}^2$. Accordingly, the imaginary part of the amplitude $A_{i,i}$ is alternatively given by its discontinuity upon crossing the real axis, which might be easier to determine in actual calculations. Note that the presence of this discontinuity signals the presence of a *branch cut* running along the real axis for $s > m_{th}^2$, which is analogous to the one highlighted in fig. 2.1.

2.4.1 Perturbative Unitarity

The results obtained so far are exact, in the sense that they do not rely on a perturbative expansion. Exact results in QFT are however quite rare (at least in $d = 4$ space-time dimensions). In practice we often demand perturbative unitarity, namely we expand in some small coupling constant both the left and right hand side of eq.(2.4.9) (or more

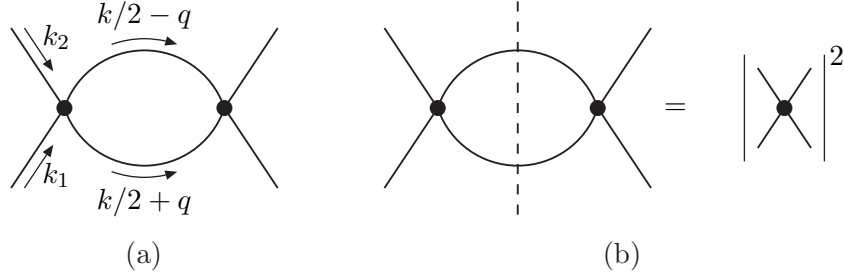


Figure 2.2: (a) Lowest-order correction to the two-particle scattering amplitude $A_{i,i}$ in a $\lambda\phi^4$ theory. (b) Schematic representation of the optical theorem (2.4.6) applied to the diagram in panel (a): the imaginary part of the amplitude in panel (a) can be obtained via the cutting rules by cutting the diagram as indicated in (b) by the dashed line.

generally of eq.(2.4.6)) and demand the equality of both terms at each order in perturbation theory. In order to calculate directly the discontinuity of an amplitude across the cut, one can take advantage of the so-called Cutkosky (or cutting) rules [21]:

1. consider all possible ways of cutting the internal propagators of a diagram in such a way that the diagram is cut into two disconnected parts and that the particles whose propagators have been cut can go simultaneously on-shell;
2. in each cut propagator with momentum p substitute the factor $1/(p^2 - m^2 + i\epsilon)$ with $-2\pi i\delta(p^2 - m^2)$ and calculate the resulting diagram;
3. sum over all possible cuts the contributions obtained above.

As an illustrative example, consider the scattering amplitude $iA_{2,2}(k^2)$ for 2–2 particles with a $\lambda\phi^4$ interaction. At the lowest (tree) non-trivial order in perturbation theory, the amplitude is proportional to the coupling λ and is manifestly real. This is in agreement with eq.(2.4.9), since an imaginary part of order λ would not match with the right hand side which is of order λ^2 . An imaginary part in the amplitude should then arise starting from one-loop. The relevant diagram contribution $i\mathcal{D}(k^2)$ is depicted in fig. 2.2(a). According to the LSZ formula (2.3.12), $iA_{2,2}(k^2) = i\mathcal{D}(k^2)/[(2\pi)^6 k^2]$ (with $k = k_1 + k_2$), where

$$i\mathcal{D}(k^2) = \frac{(i\lambda)^2}{2} \int \frac{d^4q}{(2\pi)^4} \frac{i}{(\frac{k}{2} - q)^2 - m^2 + i\epsilon} \frac{i}{(\frac{k}{2} + q)^2 - m^2 + i\epsilon}. \quad (2.4.13)$$

One can easily realize that among the three possible scattering channels s , t , and u only in the channel s it is possible to exceed the threshold for particle production such that the corresponding amplitude acquires an imaginary part. We shall focus on this case. By Lorentz invariance $A_{2,2}$ can be calculated in the center-of-mass frame with $k = (k_0, \vec{0})$ and it is convenient for the present purposes to perform first the integration over q_0 using Cauchy's theorem, closing the contour at infinity. The integrand is characterized by 4 simple poles at

$$\frac{k_0}{2} + \omega_{\vec{q}} - i\epsilon, \quad -\frac{k_0}{2} + \omega_{\vec{q}} - i\epsilon, \quad -\frac{k_0}{2} - \omega_{\vec{q}} + i\epsilon, \quad \frac{k_0}{2} - \omega_{\vec{q}} + i\epsilon. \quad (2.4.14)$$

After the integration one finds

$$\mathcal{D}(k^2) = \frac{\lambda^2}{2} \int \frac{d^3\vec{q}}{(2\pi)^3} \frac{1}{4k_0\omega_{\vec{q}}} \left(\frac{1}{\omega_{\vec{q}} - \frac{k_0}{2} - i\epsilon} - \frac{1}{\omega_{\vec{q}} + \frac{k_0}{2} - i\epsilon} \right), \quad (2.4.15)$$

Using eq.(2.1.29), an imaginary part can arise only when one of the denominators vanishes. It is evident that only the first of the two terms on the r.h.s. can possibly vanish (we take $k^0 > 0$) when $\omega_{\vec{q}} = k_0/2$ for some values of $|\vec{q}|$. This is possible only if $k_0 \geq 2m \equiv m_{th}$, as expected on kinematic reasons. We then get

$$\begin{aligned} \text{Im } \mathcal{D}(k^2) &= \frac{\lambda^2}{16\pi^2 k_0} \int_m^\infty d\omega_{\vec{q}} \sqrt{\omega_{\vec{q}}^2 - m^2} \times \pi \delta\left(\omega_{\vec{q}} - \frac{k_0}{2}\right) \\ &= \frac{\lambda^2}{16\pi k_0} \sqrt{(k_0/2)^2 - m^2} \theta(k_0 - 2m). \end{aligned} \quad (2.4.16)$$

where we used the fact that $d^3\vec{q} = 4\pi q^2 dq = 4\pi \sqrt{\omega_{\vec{q}}^2 - m^2} \omega_{\vec{q}} d\omega_{\vec{q}}$. Note also that for $k^0 < 2m$ the location of the poles (2.4.14) allows us to perform a Wick's rotation $q_0 = iq_E$. The two propagators are now complex conjugates of each other and the amplitude $A_{2,2}$ is manifestly real, in agreement with eq.(2.4.10).

The same result can be obtained by using Cutvosky rules. In this case we have

$$\text{Disc } \mathcal{D}(k^2) = \frac{\lambda^2}{2i} \int \frac{d^4q}{(2\pi)^4} (-2\pi i)^2 \delta((k/2 - q)^2 - m^2) \delta((k/2 + q)^2 - m^2), \quad (2.4.17)$$

where, as before, it is convenient to integrate first on q_0 in the center-of-mass frame. Using the fact that we can express the second δ above as

$$\delta((k_0/2 + q_0)^2 - \omega_{\vec{q}}^2) = \frac{\delta(q_0 + k_0/2 - \omega_{\vec{q}})}{2\omega_{\vec{q}}} + \frac{\delta(q_0 + k_0/2 + \omega_{\vec{q}})}{2\omega_{\vec{q}}}, \quad (2.4.18)$$

one immediately sees that the product of the first δ on the r.h.s. of eq. (2.4.17) with the second term above can never vanish, while the product with the first term, after integration

on q_0 , is equivalent to $\delta(k_0(k_0 - 2\omega_{\vec{q}})) = \delta(\omega_{\vec{q}} - k_0/2)/(2k_0)$, such that eq. (2.4.17) becomes

$$\begin{aligned} \text{Disc } \mathcal{D}(k^2) &= i \frac{\lambda^2}{8\pi k_0} \int_m^\infty d\omega_{\vec{q}} \sqrt{\omega_{\vec{q}}^2 - m^2} \delta(\omega_{\vec{q}} - k_0/2) \\ &= 2i \text{Im } \mathcal{D}(k^2), \end{aligned} \quad (2.4.19)$$

where on the second line we have taken into account the first line of eq. (2.4.16). This example shows the effectiveness of the rules mentioned above for the calculation of the imaginary part of an amplitude or, alternatively, of its discontinuity across the cut.

As a final check, let us consider the total cross section σ_T for the lowest-order scattering process of these two particles, which is characterized by the amplitude $i\mathcal{D}_0 = -i\lambda$. According to the optical theorem (2.4.9) and to (2.4.12), the imaginary part of the diagram in fig. 2.2(a) — which can be calculated (up to an i) by cutting the diagram as indicated by the dashed line fig. 2.2(b) — is equal to the modulus square of the tree-level scattering amplitude (up to the corresponding phase-space), as schematically indicated by fig. 2.2(b). By using the standard expression for the differential cross-section in a two-body scattering (see, e.g., eq. (4.99) in ref. [49]), one has

$$\left(\frac{d\sigma}{d\Omega} \right)_{cm} = \frac{|\lambda|^2}{64\pi^2 E_{cm}^2} \times \frac{1}{2}, \quad (2.4.20)$$

where the factor $1/2$ accounts for the fact that the particles are assumed indistinguishable and which gives $\sigma_T = \lambda^2/(32\pi E_{cm}^2)$ for the total cross section. By forming the combination (with $E_{cm} = k_0$)

$$\frac{4|\vec{k}_{cm}|}{(2\pi)^6 E_{cm}} \sigma_T = \frac{1}{(2\pi)^6 k_0^2} \times \frac{\lambda^2}{8\pi k_0} \sqrt{(k_0/2)^2 - m^2} \quad (2.4.21)$$

and by comparing with eq. (2.4.16), one recognizes $2 \text{Im } \mathcal{D}(k_0^2)/[(2\pi)^6 k_0^2] = 2 \text{Im } A_{2,2}$ on the r.h.s., in agreement with the optical theorem (2.4.9). Alternatively, one can verify the validity of the optical theorem written in the form of eq. (2.4.6) by comparing the expression of its l.h.s. in eq. (2.4.16) with the one obtained by specializing its r.h.s. to the case in which $G_R^{(4)} = i\lambda$ (and therefore $A_{2,i} = \lambda/[(2\pi)^3 k_0 \sqrt{(2\pi)^3 \omega_{\vec{p}_1}} \sqrt{(2\pi)^3 \omega_{\vec{p}_1}}]$ from eq. (2.3.12)) and by taking into account that the two-particle invariant phase space is given by $\int \frac{d^3 \vec{p}_1}{(2\pi)^3 2\omega_{\vec{p}_1}} \frac{d^3 \vec{p}_2}{(2\pi)^3 2\omega_{\vec{p}_2}} (2\pi)^4 \delta^4(p_n - k) = |\vec{k}_{cm}|/(4\pi k_0)$ (see, e.g., eq. (A.58) of ref. [49]).

2.5 Unstable Particles

The formalism above is also useful in understanding how unstable particles should be considered in QFT. These particles cannot appear as asymptotic states in the Hilbert

space and hence their treatment is a bit trickier.⁶ We can distinguish between two kinds of unstable particles. Those that turn into stable asymptotic states when interactions are switched off: this is the typical instance of weakly coupled unstable states, that turn into free stable states in the limit, and correspond to all unstable elementary particles we know. It also applies to bound states, when their instability is not governed by the same interactions that are at the basis of their formation. For instance the pion mesons in QCD are stable in the limit in which we switch off the electroweak interactions. A more complicated situation occurs with bound states that decay due to the same interactions responsible for their formation. In this case there is not a simple limit where we can consider them as stable. This is for instance the case of the positronium, a bound state composed of an electron and a positron in QED. Let us consider the former simpler case of unstable states admitting a limit in which they are stable. In this case we expect on general grounds that the particle develops a propagator with a Breit-Wigner form. For $p^2 \sim m^2$ we have

$$G_R^{(2)}(p^2) \sim \frac{i}{p^2 - m^2 + \Sigma(p^2)}, \quad (2.5.1)$$

where m is the mass of the stable particle when interactions are switched off. In presence of interactions Σ develops an imaginary part. In the particle rest frame eq.(2.5.1) implies an “effective mass”

$$M \simeq m - i \operatorname{Im} \Sigma(m^2)/(2m), \quad (2.5.2)$$

where we have neglected $\operatorname{Re} \Sigma$ and assumed $\operatorname{Im} \Sigma(m^2) \ll m$. The wave function Ψ of the particle evolves as $\Psi(t) = e^{-iMt} \Psi(0)$, such that the probability $|\Psi(t)|^2 \propto e^{-\Gamma t}$ decreases exponentially in time with a decay rate

$$\Gamma = \frac{\operatorname{Im} \Sigma(p^2 = m^2)}{m}. \quad (2.5.3)$$

In the rest frame of the decaying particle the total decay width equals

$$\Gamma = (2\pi)^3 \sum_n (2\pi)^4 \delta^4(p_n - p) |A_{n,1}|^2, \quad (2.5.4)$$

where $A_{1,n}$ are the amplitudes for the decay in n particles: $1 \rightarrow n$. Despite it does not make sense to talk of S -matrix elements for a single-particle state, the link between $\operatorname{Im} \Sigma$ and Γ can be formally seen as a consequence of eq.(2.4.6), with $|i\rangle$ a single particle state.

When the interactions responsible for the decay are switched off, $G_R^{(2)}$ has a simple isolated pole, as expected from the representation (2.1.28), and the operator $\phi(x)$ associated to this particle has vanishing matrix elements with the decaying product particles.

⁶Clearly, the distinction is dictated by the time scale of the process. Particles with a time decay much longer than the interacting time of a process can be considered stable and described by asymptotic states.

When interactions are switched on, the operator $\phi(x)$ has non-vanishing matrix elements with the decaying particles. As a consequence the threshold mass in eq.(2.1.28), which is given by the lowest energy to produce multi-particle states, decreases and the previously isolated pole falls within the branch-cut singularity associated to the production of the decaying particles. Moreover, the simple pole moves away from the real axis and for small Σ lies close to the branch-cut singularity. We see from eqs.(2.5.1) and (2.5.3) that the pole sits approximately at $p^2 \sim m^2 - im\Gamma$ and moves in the lower complex half-plane. At first sight, this result is in contradiction with the analyticity properties derived from the Källén-Lehmann spectral representation (2.1.28), according to which $G_R^{(2)}$ can have no singularity away from the real axis. Indeed, this complex pole does not appear as a pole of $G_R^{(2)}$ but shows up in its *analytic continuation below the branch-cut singularity*. In other words, the pole is not on the physical Riemann sheet, but on the second (or higher) Riemann sheet. This is understood by recalling that the physical momenta sit above the cut. Since the pole lies instead below the cut, in moving around the cut $\text{Im}\Sigma$ picks up a phase. With such phase the propagator has no longer a pole. In order to not get any phase we have to cross through the cut, which unavoidably leads us to another Riemann sheet. A concrete example of this phenomenon is described in section 2.7.

2.6 Causality and Analyticity*

In subsection 2.1.2 we have briefly discussed the analytic continuation of the exact propagator $G_R^{(2)}(q^2)$ for complex values of q^2 . A similar analytic continuation was assumed in section 2.4 where we briefly discussed the elastic 2-particle scattering. More in general, one could ask whether and in which sense S-matrix elements or the Green functions $G^{(N)}(p_i)$ can be analytically continued for unphysical values of the momenta. We will see that there is a fundamental connection between causality and analyticity, which points towards a positive answer to this question.

The key points are best understood from classical considerations. Let $A_{\text{inc}}(z, t)$ be a wave packet traveling along the z direction:

$$A_{\text{inc}}(z, t) = \int_{-\infty}^{+\infty} d\omega a_{\text{inc}}(\omega) e^{i\omega(z-t)}, \quad (2.6.1)$$

and scattered by a particle at rest at $z = 0$. The asymptotic forward scattered wave packet will be of the form

$$A_{\text{scatt}}(z, t)|_{|z| \rightarrow \infty} \approx \frac{1}{z} \int_{-\infty}^{+\infty} d\omega a_{\text{scatt}}(\omega) e^{i\omega(z-t)}, \quad (2.6.2)$$

where $a_{\text{scatt}}(\omega) = a_{\text{inc}}(\omega)f(\omega)$, with $f(\omega)$ the forward scattering amplitude. Suppose that this packet represents a signal which vanishes for $t < z$, so in particular

$$A_{\text{inc}}(0, t) = 0, \quad \text{for } t < 0. \quad (2.6.3)$$

By taking the inverse Fourier transform of eq.(2.6.1), we get

$$a_{\text{inc}}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dt A_{\text{inc}}(0, t) e^{i\omega t}. \quad (2.6.4)$$

Let us now see what happens if we take complex ω in (2.6.4). In general, the presence of the exponential factor proportional to $i\omega t$ is problematic. Since we have to integrate over all times, no matter the sign of the imaginary part of ω , we would in general get a dangerous growing exponential and a divergent result for $a_{\text{inc}}(\omega)$. However, $A_{\text{inc}}(0, t)$ is not an arbitrary function. Crucially, it has to satisfy the condition (2.6.3) imposed by causality. Thanks to (2.6.3) the integral in (2.6.4) can be convergent in the upper ω half-plane. Indeed, for $t > 0$, the phase factor is exponentially suppressed, while for $t < 0$, where we would have an exponentially increasing factor, the wave packet vanishes. So causality implies that $a_{\text{inc}}(\omega)$ is an *analytic* function in the upper ω half-plane. Consider now eq.(2.6.1) in the complex ω -plane. If we assume that $a_{\text{inc}}(\omega)$ does not grow too much at infinity, so as to compensate the exponential suppression given by $e^{i\omega(z-t)}$, we can close the contour at infinity. This is automatically satisfied if we demand that $a_{\text{inc}}(\omega)$ grows at most as a polynomial in ω . Such condition is often denoted as polynomial boundedness. For $t < z$, we get

$$\oint_C d\omega a_{\text{inc}}(\omega) e^{i\omega(z-t)} = 0, \quad (2.6.5)$$

The causality principle requires that the scattered wave should also vanish for $t < z$, since no wave can be ahead of the incident wave packet:

$$A_{\text{scatt}}(z, t) = 0, \quad \text{for } t < z. \quad (2.6.6)$$

We can now take the inverse Fourier transform of eq.(2.6.2) and repeat the argument above to conclude that the *causality condition (2.6.6) implies the analyticity of $a_{\text{scatt}}(\omega)$ — and hence of $f(\omega)$ — in the upper half-plane.*⁷

Let us see some implications coming from such analyticity properties. Given a point z in the upper half-plane we can use Cauchy theorem to write

$$f(z) = \frac{1}{2i\pi} \int_{C_z} \frac{d\omega' f(\omega')}{\omega' - z}, \quad (2.6.7)$$

⁷Unless $a_{\text{inc}}(\omega)$ has zeros that induce poles in $f(\omega)$, but these are avoidable by simply changing the (quite arbitrary) incident wave packet.

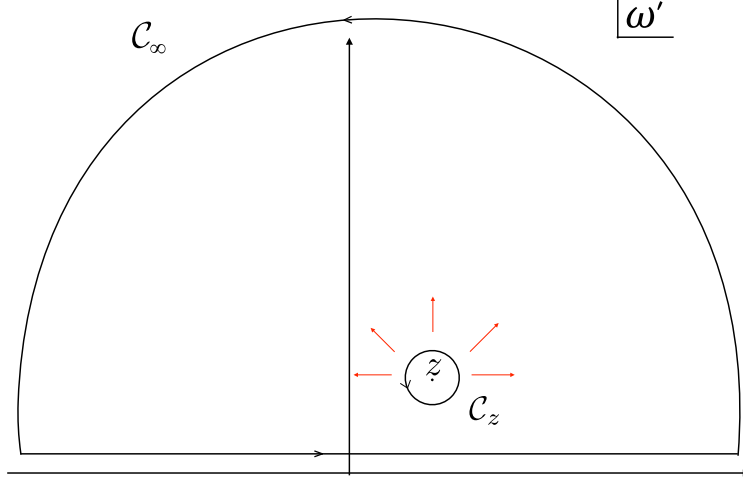


Figure 2.3: Thanks to the analyticity of $f(\omega')$, the contour of integration C_z can be blown up in a large semi-circle approaching the real line from above and going at infinity (C_∞) in the upper half-plane.

where C_z is a small circle around the point z . Since f is analytic over the whole upper ω' half-plane, we can blow the contour C_z until it approaches the real axis and closes at infinity, like in fig.2.3. When z approaches the real line from above, $z = \omega + i\epsilon$, we have

$$\begin{aligned} f(\omega) &\equiv \lim_{\epsilon \rightarrow 0} f(\omega + i\epsilon) = \lim_{\epsilon \rightarrow 0} \frac{1}{2i\pi} \int_{-\infty}^{+\infty} \frac{d\omega' f(\omega')}{\omega' - \omega - i\epsilon} + \frac{1}{2} f_\infty \\ &= \frac{1}{2i\pi} \text{P} \int_{-\infty}^{+\infty} \frac{d\omega' f(\omega')}{\omega' - \omega} + \frac{1}{2} f(\omega) + \frac{1}{2} f_\infty. \end{aligned} \quad (2.6.8)$$

In the last relation we have used eq.(2.1.29), P stands for the principal part of the integral and $f_\infty/2$ denotes the contribution to the integral coming from C_∞ . If we assume for the moment that $f(\omega) \rightarrow 0$ at infinity, so that $f_\infty = 0$, eq.(2.6.8) gives

$$f(\omega) = \frac{1}{i\pi} \text{P} \int_{-\infty}^{+\infty} \frac{d\omega' f(\omega')}{\omega' - \omega}. \quad (2.6.9)$$

The real and imaginary parts of eq.(2.6.9) give

$$\text{Re} f(\omega) = \frac{1}{\pi} \text{P} \int_{-\infty}^{+\infty} \frac{d\omega' \text{Im} f(\omega')}{\omega' - \omega}, \quad \text{Im} f(\omega) = -\frac{1}{\pi} \text{P} \int_{-\infty}^{+\infty} \frac{d\omega' \text{Re} f(\omega')}{\omega' - \omega}. \quad (2.6.10)$$

Another commonly used rewriting of these equations is

$$f(\omega) = \lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{d\omega' \text{Im} f(\omega')}{\omega' - \omega - i\epsilon}, \quad (2.6.11)$$

whose real part coincides with the first relation in eq.(2.6.10). We see that, by knowing the imaginary part of the forward scattering amplitude, we can reconstruct the whole amplitude. The original derivation of eq.(2.6.11) dates back to Kramers and Kronig that applied it in classical optics with $f(\omega)$ being the index of refraction. A frequency dependence of the index of refraction leads to the dispersion of light (like in a rainbow) and hence eq.(2.6.11) (and their quantum generalizations) is called a dispersion relation. An imaginary component of the refractive index leads to dissipation, so the dispersion relations allow us to compute the refractive index of a material (i.e. $\text{Re } f$) from its dissipation properties ($\text{Im } f$). The latter is also related, through a classical optical theorem, to the total absorption cross section.

If $f(\omega)$ does not vanish at infinity, as we assumed, but rather

$$f(\omega)|_{|\omega| \rightarrow \infty} \propto |\omega|^{N-1} \quad (2.6.12)$$

for some $N \geq 0$, we can divide $f(\omega)$ by a given polynomial $P_n(\omega)$ of degree $n = [N]$ and apply Cauchy theorem to f/P_n . By construction, the contribution at infinity vanishes, but we have to be careful because the zeros of P_n give rise to poles whose contribution should be considered. Choosing for definiteness $P_n = \omega^n$ and repeating the same steps as before, we get

$$f(\omega) = \omega^n \lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{d\omega' \text{Im} f(\omega')}{(\omega' - \omega - i\epsilon)(\omega')^n} + Q_n(\omega). \quad (2.6.13)$$

In eq.(2.6.13), the residues due to the poles of P_n are encoded in the factor Q_n , which is a polynomial of degree n in ω with coefficients determined in terms of the values of $f(\omega)$ and its first $(n-1)$ -derivatives at $\omega = 0$ (more generally, at the specific values of ω where $P_n = 0$). The function f is thus no longer determined by $\text{Im } f$ only, but requires also the knowledge of such “boundary” data. This is the price we have to pay in order to have a sufficiently well-behaved function at infinity. Eq.(2.6.13) is called a dispersion relation with n subtractions, or n -subtracted dispersion relation.

The classical causality condition (2.6.6) in QFT is replaced by the microcausality condition [26]

$$[\mathcal{O}_1(x), \mathcal{O}_2(y)] = 0, \quad \text{for } (x - y)^2 < 0, \quad (2.6.14)$$

where $\mathcal{O}_1$ and $\mathcal{O}_2$ are arbitrary bosonic local operators in the QFT. Using eq.(2.6.14) and the LSZ reduction formulae, some analyticity properties of certain amplitudes can be studied axiomatically, i.e. at the non-perturbative level and using first principles only, such as the completeness relation of physical states and Lorentz invariance (in addition to microcausality itself, of course).

For concreteness, let us consider the elastic scattering of two identical particles of mass m , that we assume to be the lightest particles present in the theory. The scalar amplitude $\mathcal{M}$ is a function of the four particle momenta p_i ($i = 1, 2, 3, 4$). Particles 1 and 2 are in-going, particles 3 and 4 are out-going. Momentum conservation and the on-shell conditions fix $p_1^\mu + p_2^\mu = p_3^\mu + p_4^\mu$, $p_i^2 = m^2$. Lorentz invariance requires that the scalar quantity $\mathcal{M}$ can only depend on Lorentz invariant combinations of the momenta. A useful parametrization is given by the Mandelstam variables

$$s = (p_1 + p_2)^2, \quad t = (p_1 - p_3)^2, \quad u = (p_1 - p_4)^2. \quad (2.6.15)$$

It is immediate to verify that $s + t + u = 4m^2$, so that the amplitude is effectively a function of two independent variables, that we can take to be s and t : $\mathcal{M} = \mathcal{M}(s, t)$. Going to the center of mass frame of the two particles, it is straightforward to see that the kinematically allowed range of the Mandelstam variables is

$$s \geq 4m^2, \quad t \leq 0, \quad u \leq 0, \quad (1 + 2 \rightarrow 3 + 4). \quad (2.6.16)$$

We can then ask what are the analyticity properties of the amplitude as a function of arbitrary complex values of s and t . Studying the analyticity properties of a function of two variables is a difficult task, so let us take t fixed and consider $\mathcal{M}$ as a function of s . Some expectations on the analytic properties of $\mathcal{M}$ can be deduced from perturbation theory. We expect simple poles at $s = \mu^2$, where μ is the mass of the possible particle exchanged in the 1-2 interaction, and then a branch-cut singularity starting from the 2-particle threshold production, i.e. $s \geq 4m^2$. Using crossing symmetry, the same amplitude $\mathcal{M}$ also describes the scattering process $1 + \bar{3} \rightarrow \bar{2} + 4$ and $1 + \bar{4} \rightarrow \bar{2} + 3$. In terms of the variables (2.6.15), the allowed range of s , t and u in the different channels is

$$\begin{aligned} t \geq 4m^2, \quad s \leq 0, \quad u \leq 0, \quad (1 + \bar{3} \rightarrow \bar{2} + 4), \\ u \geq 4m^2, \quad s \leq 0, \quad t \leq 0, \quad (1 + \bar{4} \rightarrow \bar{2} + 3). \end{aligned} \quad (2.6.17)$$

This implies that, at fixed t , if there is a pole in $\mathcal{M}(s)$ at $s = \mu^2$ in the $1 + 2 \rightarrow 3 + 4$ scattering process (s -channel), a pole should also occur at $s = 4m^2 - \mu^2 - t$, which corresponds to the same pole μ^2 occurring at $4m^2 - s - t = u = \mu^2$ in the $1 + \bar{4} \rightarrow \bar{2} + 3$ process (u -channel).⁸ Similarly, the branch-cut at $u \geq 4m^2$ in the $1 + \bar{4} \rightarrow \bar{2} + 3$ process implies a branch-cut at $s = -t$ in $\mathcal{M}(s)$. More in general we expect from causality that all the singularities of $\mathcal{M}(s)$ should be associated to physical processes and, as such, they should all lie on the real s -axis. In particular, we might have other poles at $s = \mu_1^2, \mu_2^2, \dots$

⁸If the same particle can be exchanged in the process, we have $\mu = m$, but this is not necessary.

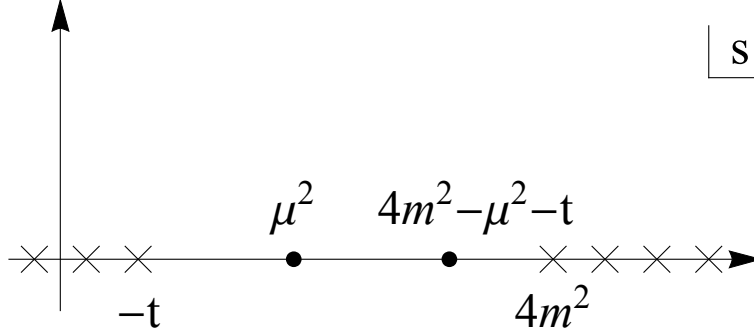


Figure 2.4: The expected singularity structure of the $2 \rightarrow 2$ scattering amplitude $\mathcal{M}(s)$ at fixed $t < 0$ in the complex s -plane. The black points indicate the simple poles associated to the particles (bound states) exchanged in the s -channel (μ^2) and in the u -channel ($4m^2 - \mu^2 - t$). The crosses indicate the branch-cut singularities, starting at $s = 4m^2$ in the s -channel and at $s = -t$ in the u -channel. Poles associated to the possible presence of other bound states have not been indicated.

associated to the exchange of other particles (bound states in general), or branch-cut singularities due to the opening of other multi-particle processes. For example, another branch-cut is expected to occur for $s \geq 9m^2$ in association to the production of three-particles (if allowed by the symmetries of the system). These other branch-cuts will occur on top of the first one that starts at $s \geq 4m^2$ (and $s \leq -t$).

Working out the consequences of the causality conditions like eq.(2.6.14), the analyticity domain of $\mathcal{M}(s)$, for a fixed physical value of $t < 0$ within a certain range, have been rigorously proven to be as in fig.2.4. Aside from branch-cuts on the real axis and simple poles associated to bound states, causality implies that no other singularities occur. The physical region in the s channel for $s \geq 4m^2$ lies *above* the cut. This can be seen from the $i\epsilon$ prescription that amounts to a shift $m^2 \rightarrow m^2 - i\epsilon$. Since $s + t + u = 4m^2$, the physical region in the u -channel sits *below* the cut $s < -t$ in fig.2.4. Causality also implies analyticity properties in t at fixed s for physical values $s > 4m^2$ which we will not discuss.

It is important to emphasize that we did not make use of unitarity so far, as in the classical example of the wave packet considered at the beginning of this section. If unitarity is assumed, the range of analyticity of the amplitude is significantly extended. In particular, we can get analyticity in s as depicted in fig.2.4 in a entire complex domain in t which extends up to positive values of $t \leq 4m^2$. When $t > 4m^2$ other singularities not associated to multi-particle production, known as anomalous thresholds, can and generally do occur. For complex values of t we can also get off-axis singularities in the s -plane. When we can

determine an analytic domain of $\mathcal{M}(s, t)$, dispersion relations like eq.(2.6.11) can be written down, assuming that the amplitude does not grow too much at infinity (polynomial boundedness) as in eq.(2.6.12). Among other things, these allow us to relate $\mathcal{M}$ to the total cross section, that through the optical theorem can be expressed in terms of $\text{Im } \mathcal{M}$.

The methods used to establish analytic properties of amplitudes are state-dependent, and not all $2 \rightarrow 2$ amplitudes admit a dispersion relation. Inelastic amplitudes of the form $2 \rightarrow n$ with $n > 2$ are much more complicated to study (more kinematical invariants) and their analytic properties are to a large extent unknown.

The analytical properties of Green functions received great attention starting from the mid-50's until about the late 60's, in an attempt to describe the strong interactions from first principles, without relying on a weakly coupled Lagrangian description (S-matrix bootstrap). This activity quickly faded away after the discovery of Quantum ChromoDynamics (QCD) as the theory of strong interactions, but left us with several important results. In the impossibility to review here all such results, let us just mention a general bound on the cross-section (Froissart-Martin bound [25, 45]), an understanding of certain high energy scattering limits in strongly coupled gauge theories in terms of so called Regge trajectories [51, 52], and the birth of what we now call string theory. More recently, among other developments, arguments based on the analyticity of amplitudes have allowed us to put constraint on effective field theories [2] and played an essential role in proving an important property satisfied by the renormalization group flow of QFTs in four space-time dimensions (the so called a-theorem) [40].

A proper understanding of the analytic structure of amplitudes in QFT would also be important to rigorously establish important properties of amplitudes, such as crossing symmetry.⁹ It is not excluded that a fresh look at some of the old problems left open in the S-matrix bootstrap (together with the vastly increased numerical power we have today with respect to half-century ago) would lead to a revival and further progress in this direction.

⁹We briefly discussed crossing symmetry as a “simple” consequence of the LSZ formula (2.3.12) below that equation. In fact, this property is far from trivial because the crossed amplitudes are functions defined in disconnected regions in the Mandelstam variables with respect to the original amplitude function. Proving crossing symmetry is equivalent to prove that the crossed amplitude is the analytic continuation in the crossed region of the original amplitude. Crossing for general two-particle amplitudes has been proved long ago in ref.[11], but more in general it is just assumed.

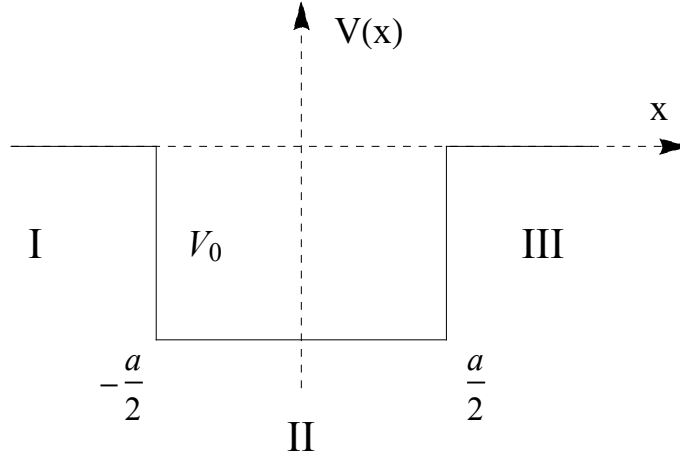


Figure 2.5: One-dimensional potential well.

2.7 Bound States and Resonances**

We have seen in the previous sections that simple poles of the two-point function or of the scattering amplitude (S -matrix) are a signal of one-particle states. These can arise also in correspondence of bound states. We have also seen at the end of section 2.4 that unstable states, resonances, correspond to simple poles with a non-vanishing imaginary component, i.e. complex poles. In this section we will show that

- bound states correspond to simple poles of the scattering amplitude
- resonances give rise to poles hidden below the branch-cut.

In order to keep the analysis as simple as possible, we will just work out a simple example in quantum mechanics (relativistic bound states are a notoriously difficult subject).

The system we consider is a one-dimensional quantum system in a potential well, see fig.2.5. The Schrödinger equation is

$$\left(\frac{-1}{2m} \frac{d^2}{dx^2} + V(x) \right) \psi_n(x) = E_n \psi_n(x), \quad (2.7.1)$$

where ψ_n and E_n are the eigenfunctions and eigenvalues of the system, respectively. The system is invariant under the $\mathbf{Z}_2$ parity transformation $x \rightarrow -x$, so we can separately consider even and odd eigenstates. Let us first discuss the bound states, i.e. the states with energy $-V_0 < E_n < 0$, where $V_0 > 0$ is the depth of the well. For the different regions

in fig.2.5 we have, for the parity even states,

$$\begin{aligned}\psi_{\text{I}}(x) &= Ce^{kx}, \\ \psi_{\text{II}}(x) &= A \cos(px), \\ \psi_{\text{III}}(x) &= Ce^{-kx},\end{aligned}\tag{2.7.2}$$

where

$$k = \sqrt{-2mE}, \quad p = \sqrt{2m(E + V_0)},\tag{2.7.3}$$

and A and C are undetermined constants so far. Demanding the continuity of the wave function and its first derivative at $x = a/2$ gives the equation that the even eigenfunctions should satisfy:

$$k = p \tan \frac{pa}{2}, \quad (\text{even}).\tag{2.7.4}$$

Analogously, the equation for the odd eigenfunctions is

$$k = -p \cot \frac{pa}{2}, \quad (\text{odd}).\tag{2.7.5}$$

For illustration, we plot in fig.2.6 the solutions of eqs.(2.7.4) and (2.7.5) for some values of m , a , and V_0 . At fixed m and a , the number of bound states depends on V_0 and increases with V_0 , as expected. The deepest bound state (i.e. with the largest negative energy) is always parity even. For small enough V_0 there are no odd bound states and only one even. Its energy can be computed analytically by expanding eq.(2.7.4). For small V_0 , the argument of the tangent is small, $\tan x \approx x$. Taking the square of that equation, we see that $E \sim \mathcal{O}(V_0^2)$ and hence at leading order we get the bound state energy

$$E \approx -\frac{ma^2V_0^2}{2}.\tag{2.7.6}$$

Let us now consider the states with $E > 0$. The spectrum is a continuum of waves and we look for solutions with an incident wave coming from $x \rightarrow -\infty$. In the different regions we have

$$\begin{aligned}\psi_{\text{I}}(x) &= e^{iqx} + Re^{-iqx}, \\ \psi_{\text{II}}(x) &= B_1 e^{ipx} + B_2 e^{-ipx}, \\ \psi_{\text{III}}(x) &= Te^{iq(x-a)},\end{aligned}\tag{2.7.7}$$

where $q = \sqrt{2mE}$. The coefficients $B_{1,2}$, R and T are to be determined, with the latter two being identified with the reflection and transmission coefficients of the wave, respectively. Demanding the continuity of the wave function and its first derivative at $x = \pm a/2$ allows

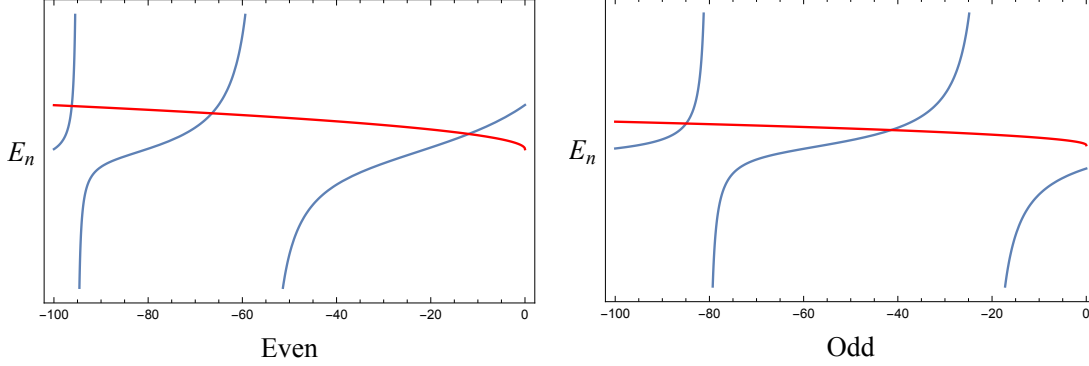


Figure 2.6: Solutions of the bound states equations (2.7.4) and (2.7.5) for even and odd parity, respectively. We plot in red the function k and in blue $p \tan(pa/2)$ (even) and $-p \cot(pa/2)$ (odd). We have taken $m = a = 1$ and $V_0 = 100$.

us, after some algebra, to determine these coefficients. In particular, we are interested here to the transmission coefficient that reads

$$T(E) = \left[\cos(pa) - \frac{i}{2} \left(\frac{q}{p} + \frac{p}{q} \right) \sin(pa) \right]^{-1}, \quad (2.7.8)$$

which is the one-dimensional analogue of a scattering amplitude. Having the luxury of an explicit expression, the analytic properties of $T(E)$ in the complex E -plane are easily determined. The transmission coefficient is analytic over the whole complex plane except a branch-cut at $E = 0$ given by the variable q ¹⁰ and possible poles whenever

$$\cos pa = \frac{i}{2} \left(\frac{q}{p} + \frac{p}{q} \right) \sin pa. \quad (2.7.9)$$

When properly normalized, poles in $T(E)$ corresponds to a transmitted wave without the corresponding incident wave, that is bound states. Dividing by $\cos(pa)$ and using the trigonometric relation $\tan 2x = 2(\cot x - \tan x)^{-1}$, we can recast eq.(2.7.9) in the form

$$\cot \frac{pa}{2} - \tan \frac{pa}{2} = i \left(\frac{q}{p} + \frac{p}{q} \right), \quad (2.7.10)$$

which has two possible set of solutions:

$$\cot \frac{pa}{2} = i \frac{p}{q}, \quad \cot \frac{pa}{2} = i \frac{q}{p}. \quad (2.7.11)$$

Let us now write

$$E = |E| e^{i\phi} \quad (2.7.12)$$

¹⁰Note that the possible branch-cut singularities around $p = 0$ ($E = -V_0$) eventually cancel between each other, so that $T(E)$ is single-valued around $p = 0$.

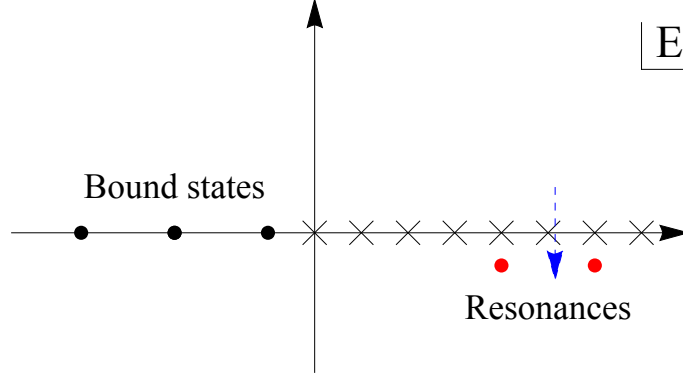


Figure 2.7: Analytic structure of the transmission coefficient in the complex E -plane. The black dots for $E < 0$ represents the stable bound states of the system, while the red points are the resonances, that appear as poles in the second Riemann sheet when crossing the branch-cut.

and take the branch of the square root where

$$\sqrt{E} = \sqrt{|E|}e^{i\phi/2}. \quad (2.7.13)$$

In this branch, negative real values of E correspond to $\phi = \pi$, and thus

$$q = \sqrt{2mE} \rightarrow i\sqrt{2m|E|} = ik, \quad (2.7.14)$$

with k as in eq.(2.7.3). It is now immediate to see that eqs.(2.7.11) turn into the bound state equations (2.7.4) and (2.7.5), confirming the expectations that simple poles of the scattering amplitude are associated to stable bound states in the system.

Let us finally discuss unstable resonances. The analogue of the S -matrix is given by

$$S(E) \equiv |T|^2 = \left(1 + \frac{\sin^2 pa}{4(\frac{E}{V_0})(1 + \frac{E}{V_0})}\right)^{-1}. \quad (2.7.15)$$

Resonances occur when S is maximal, in this case at

$$\sin pa = 0, \quad \text{with } E > 0, \quad (2.7.16)$$

corresponding to

$$a\sqrt{2m(E + V_0)} = n\pi \rightarrow E_n = \frac{\pi^2 n^2}{2a^2 m} - V_0 > 0, \quad (2.7.17)$$

for all positive integer values of n for which the last relation in eq(2.7.17) holds. For such values of E_n the reflection coefficient vanishes and the wave is totally transmitted. The

transmission coefficient can be written as

$$T(E) = \frac{1}{\cos pa} \left(1 - \frac{i}{2} f(E) \right)^{-1}, \quad (2.7.18)$$

where $f(E) = (q/p + p/q) \tan pa$. Expanding around E_n , we have

$$f(E) = 0 + \left. \frac{df}{dE} \right|_{E_n} (E - E_n) + \mathcal{O}(E - E_n)^2, \quad \cos pa = 1 + \mathcal{O}(E - E_n)^2, \quad (2.7.19)$$

where

$$\left. \frac{df}{dE} \right|_{E_n} = \left(\frac{q}{p} + \frac{p}{q} \right) \left. \frac{dp}{dE} \right|_{E_n} \equiv \frac{4}{\Gamma_n}, \quad (2.7.20)$$

is real and positive. Plugging in $T(E)$, we finally have, for $E \approx E_n$,

$$T(E) \approx \frac{1}{1 - \frac{i}{2} \frac{4}{\Gamma_n} (E - E_n)} = \frac{i\Gamma/2}{(E - E_n) + i\Gamma_n/2}, \quad (2.7.21)$$

which reproduces the expected Breit-Wigner form of a resonance, with Γ_n its decay width. For $\Gamma_n \ll E_n$, eq.(2.7.21) would indicate that $T(E)$ has simple poles at

$$E = E_n - \frac{i}{2} \Gamma_n, \quad (2.7.22)$$

i.e. in the complex plane just below the branch-cut, see fig.2.7. However, a closer inspection shows that these poles are *not* associated to the original function $T(E)$, but to its analytic continuation below the cut, namely they sit at the second Riemann sheet of the function. Indeed, the pole equation for T is given by eq.(2.7.9). Expanding around E_n , where $\sin p_n a = 0$, we get

$$1 + \mathcal{O}(E - E_n)^2 = \frac{i}{2} \left(\frac{q}{p} + \frac{p}{q} \right) \left. \frac{dp}{dE} \right|_{E_n} (E - E_n) + \mathcal{O}(E - E_n)^2. \quad (2.7.23)$$

With the choice (2.7.13), a pole in the first Riemann sheet just below the positive real axis is reached by taking $\phi = 2\pi$, which would give in eq.(2.7.23) $q \rightarrow -q$. In this way, using eqs.(2.7.20) and (2.7.22), the r.h.s. of eq.(2.7.23) would give -1 instead of $+1$. The equation is instead satisfied if we make a clock-wise rotation with ϕ small and negative, crossing in this way the branch-cut at $E > 0$. In conclusion, the complex poles associated to resonances occur in the second (more in general higher) Riemann sheet of scattering amplitude functions.

Chapter 3

Renormalization Theory

In chapter 2 we have seen the basic axioms that any QFT is expected to satisfy, such as cluster decomposition, unitarity, causality, and some of the consequences that derive by them, like the optical theorem or some analyticity properties of S-matrix elements. It is generally difficult to make further progress in QFT, because the exact computation of S-matrix elements is prohibitively hard. We are then forced to leave the “axiomatic” world where exact results can be established and use approximation procedures. A notable and useful approximation assumes that the interactions among particles are weak. In this way we can systematically use a perturbative expansion around free particles, where correlation functions can be computed. We assume the reader is familiar with the basics of the standard treatment of perturbation theory in QFT, which makes use of Feynman diagrams. Beyond tree-level, the latter are known to be generally divergent and the purpose of this chapter is to discuss how to fix these divergences.

There are at least three kinds of divergences in QFT amplitudes: infra-red, collinear and ultra-violet divergences. Here we only deal with the last kind, the ultra-violet (UV) divergences. They occur when the momenta of the virtual particles running in the loops go to infinity (hence the name UV). In order to make sense of these otherwise ill-defined amplitudes, we have to “renormalize” the theory. Schematically, this process requires two steps. First, we change the theory so that the amplitudes become finite (regularization) and then we redefine the parameters of the Lagrangian in order to “hide” the divergences in unphysical quantities. All this process is called renormalization of the theory.

The structure of this chapter is as follows. We introduce in section 3.1 the notion of superficial degree of divergence of an arbitrary Feynman diagram, which allows us to classify QFTs as renormalizable and non-renormalizable. In section 3.2 we discuss the notion of absolute convergence of a Feynman diagram. In section 3.3 we explain how UV

divergences are related to local counterterms in a Lagrangian and show (without proving it) how they cancel to all orders in perturbation theory. The notions of regularization and renormalization are discussed in section 3.4, where as an example the structure of one-loop divergences in QED are reviewed. Finally, in section 3.5 we introduce dimensional regularization, the main regularization procedure we will be using in these notes.

3.1 Superficial Degree of Divergence

Before discussing the details of the renormalization procedure, it is useful to first classify which amplitudes are potentially divergent.¹ In this chapter we will consider a generic QFT, with an arbitrary number of scalar, fermion and gauge fields. The latter are taken in the unbroken phase, when they are massless.²

Consider a generic graph $\mathcal{G}$ contributing to a given connected one-particle irreducible (1PI) amplitude. Our aim will be to compute the superficial degree of divergence of $\delta(\mathcal{G})$.³ δ is defined as the degree of divergence of the graph when all virtual momenta go to infinity at the same time. Denoting by q_i all the virtual momenta running and by $\mathcal{G}(q_i)$ the integrand of the L -loop graph, we can define δ as

$$\lim_{\lambda \rightarrow \infty} \lambda^{4L} \mathcal{G}(\lambda q_i) \propto \lambda^{\delta(\mathcal{G})}. \quad (3.1.1)$$

Notice that the rescaling $q_i \rightarrow \lambda q_i$ has to be performed also to the loop integrals, and this explains the λ^{4L} factor in eq. (3.1.1). For instance, a one-loop graph of the form

$$\int \frac{d^4 q}{(2\pi)^4} \frac{i}{(p+q)^2 - m^2} \frac{i}{q^2 - m^2} \quad (3.1.2)$$

gives

$$\lim_{\lambda \rightarrow \infty} \lambda^4 \frac{i}{(p+\lambda q)^2 - m^2} \frac{i}{\lambda^2 q^2 - m^2} \propto \lambda^0 \quad (3.1.3)$$

and hence $\delta = 0$. A graph is superficially divergent if $\delta \geq 0$ and superficially convergent if $\delta < 0$. In particular, $\delta = 0$ corresponds to a logarithmic divergence, while $\delta > 0$ gives rise to a power-like divergence. The superficial degree of divergence of a graph depends clearly on the structure of the graph, number of internal propagators, vertices, etc. In order to be as general as possible, let us introduce some notation. We label with $i = 1, \dots, n_V$ the

¹The analysis in the following three sections follows in part chapter 12 of ref. [60], with a straightforward generalization to space-time dimensions different than 4.

²The renormalization of spontaneously broken gauge theories will not be discussed in detail. However, see the end of section 8.4 for a brief comment.

³This is called D in ref. [60]. We call it δ , because we will denote with the symbol D the number of space-time dimensions in what follows.

n_V different interactions present in the Lagrangian of the theory, and by d_i the number of derivatives present in each interaction vertex i . We also denote by $f = e, \gamma, \dots$ the different fields present in the theory and by n_{if} the number of fields of type f present in the interaction vertex i . Any graph $\mathcal{G}$ is then characterized by the following quantities:

$$\begin{aligned} I_f &\equiv \text{number of propagators of type } f \\ E_f &\equiv \text{number of external lines of type } f \\ N_i &\equiv \text{number of interaction vertices } i \end{aligned} \tag{3.1.4}$$

The momentum space propagator behaves as

$$\Delta_f(p) \propto p^{2s_f-2} \tag{3.1.5}$$

where s_f is the spin of the particle type f . This formula is correct for scalars ($s_f = 0$) and spin 1/2 fermions ($s_f = 1/2$). For gauge fields with $s_f = 1$, eq. (3.1.5) would give $\Delta_f(p) \sim \text{constant}$, which is in fact the correct result for a massive vector, but it is not the correct one, due to gauge invariance, for massless gauge fields, where one has $s_f = 0$. In the following we will use eq. (3.1.5), with the understanding that $s_f = 0$ for massless gauge fields. The powers of virtual momenta entering $\mathcal{G}$ is easily determined. We have

$$\delta(\mathcal{G}) = \sum_f I_f(2s_f - 2) + \sum_i d_i N_i + 4 \left(\sum_f I_f - \sum_i N_i + 1 \right). \tag{3.1.6}$$

The first, second and third terms in the right-hand side of eq. (3.1.6) represent the propagator, the vertex and the loop contributions, respectively. The loop contribution has been written by considering each internal propagator with an independent virtual momentum and imposing momentum conservation at each vertex. The +1 takes into account the overall momentum conservation of the external lines. We can rewrite eq. (3.1.6) as

$$\delta(\mathcal{G}) = \sum_f I_f(2s_f + 2) + \sum_i N_i(d_i - 4) + 4. \tag{3.1.7}$$

Since each internal (propagator) and external line has to end to a vertex, the internal to both ends, the external to one end only, the following identities must hold, for any field type f :

$$2I_f + E_f = \sum_i N_i n_{if}. \tag{3.1.8}$$

We can use eq. (3.1.8) to solve for I_f and plug the result in eq. (3.1.7). In so doing we get

$$\delta(\mathcal{G}) = 4 - \sum_f E_f(s_f + 1) - \sum_i N_i \left[4 - d_i - \sum_f n_{if}(s_f + 1) \right]. \tag{3.1.9}$$

We now show that the expression in square brackets in the second term of the right-hand side of eq. (3.1.9) coincides with the dimensionality of the coupling constant, call it g_i , multiplying the vertex i in the Lagrangian. The latter, indeed, has, by definition, the schematic form

$$\mathcal{L} \supset g_i \partial^{d_i} \prod_f \phi_f^{n_{if}}, \quad (3.1.10)$$

omitting unnecessary details, such as the Lorentz indices and how the derivatives act on the various fields. Since

$$[\phi_f] = s_f + 1, \quad (3.1.11)$$

we immediately get

$$[g_i] \equiv \Delta_i = 4 - d_i - \sum_f n_{if}(s_f + 1). \quad (3.1.12)$$

Using eq. (3.1.12) we finally arrive to the desired expression for the superficial degree of divergence of a graph:

$$\delta(\mathcal{G}) = 4 - \sum_f E_f(s_f + 1) - \sum_i N_i \Delta_i. \quad (3.1.13)$$

Remarkably, the key factors determining δ are the dimensionality of the coupling constants Δ_i . Notice that we are here computing the degree of divergence of single graphs. A given connected 1PI Green function Γ_{E_f} is uniquely determined by E_f . In perturbation theory, $\Gamma_{E_f} = \sum \mathcal{G}$, where the sum runs over all graphs with E_f external lines. A Green function $\Gamma(E_f)$ is superficially finite only if all the graphs $\mathcal{G}$ entering this sum have $\delta(\mathcal{G}) < 0$. We immediately learn from eq. (3.1.13) that increasing the number of external lines tends to decrease δ . We classify all QFT's as follows:

- If $\Delta_i \geq 0$ for all $i = 1, \dots, n_V$, only a finite number of Green functions Γ_{E_f} , those for which at most $4 - \sum_f E_f(s_f + 1) \geq 0$, is divergent. We call such theories *renormalizable*.
- If at least one $\Delta_i < 0$, an infinite number of Green functions $\Gamma(E_f)$ is divergent. We call such theories *non-renormalizable*.

Among the renormalizable theories, a further distinction is possible: if $\Delta_i > 0$ for all $i = 1, \dots, n_V$, only a finite number of individual graphs diverges. We call such theories *super-renormalizable*. When this distinction occurs, a renormalizable theory is defined as one in which at least one coupling is dimensionless. The terminology “renormalizable” and “non-renormalizable” applies also to individual operators or couplings. We denote a coupling constant as renormalizable if its dimension is positive and non-renormalizable if

it is strictly negative. Similarly, an operator is denoted renormalizable if its dimension is less or equal to four and non-renormalizable if it is higher than four.

Let us list the possible renormalizable couplings. In a QFT with scalars only, we can have ϕ^3 and ϕ^4 interactions only. No renormalizable derivative interactions are allowed, compatibly with Lorentz invariance. In a QFT with scalars and fermions, the only possible renormalizable interactions are the Yukawa-like of the form $\phi\bar{\psi}\psi$ or $\phi\bar{\psi}\gamma_5\psi$. Again, no Lorentz invariant renormalizable derivative interactions are allowed. When gauge fields are included, we have $\bar{\psi}\gamma^\mu\psi A_\mu$, $\bar{\psi}\gamma^\mu\gamma_5\psi A_\mu$, $(\phi^\dagger\partial_\mu\phi - \phi\partial_\mu\phi^\dagger)A^\mu$ and $\phi^2 A_\mu A^\mu$. Other possibly renormalizable interactions like $(A_\mu A^\mu)^2$ or $(\phi^\dagger\partial_\mu\phi + \phi\partial_\mu\phi^\dagger)A^\mu$ are forbidden by gauge invariance.

It is straightforward to generalize to an arbitrary number of space-time dimensions the computation of the superficial degree of divergence of a graph. Equation (3.1.6) is still valid provided we replace the factor of 4 with D . In so doing we get

$$\delta(\mathcal{G}) = D - \sum_f E_f \left(s_f - 1 + \frac{D}{2} \right) - \sum_i N_i \left(D - d_i - \sum_f n_{if} \left(s_f - 1 + \frac{D}{2} \right) \right). \quad (3.1.14)$$

In D space-time dimensions

$$[\phi_f] = s_f - 1 + \frac{D}{2} \quad (3.1.15)$$

and consequently the interactions (3.1.10) have dimensions

$$[g_i] \equiv \Delta_i = D - d_i - \sum_f n_{if} \left(s_f - 1 + \frac{D}{2} \right). \quad (3.1.16)$$

and eq. (3.1.13) becomes

$$\delta(\mathcal{G}) = D - \sum_f E_f \left(s_f - 1 + \frac{D}{2} \right) - \sum_i N_i \Delta_i. \quad (3.1.17)$$

The renormalizability of a theory in D dimensions is still characterized by the dimensions of their couplings Δ_i . If all $\Delta_i \geq 0$ the theory is renormalizable, otherwise it is non-renormalizable. We notice that when $D > 4$ most four-dimensional renormalizable theories turn into non-renormalizable ones, because the dimension of the couplings Δ_i decreases when D increases. For instance, in $D = 5$ dimensions, the gauge couplings constants have dimension $\Delta = -1/2$. All gauge theories are non-renormalizable for $D > 4$. The only renormalizable interaction up to $D = 6$ is the trilinear ϕ^3 vertex. For $D > 6$ there are no renormalizable vertices. On the contrary, when $D < 4$, non-renormalizable interactions in $D = 4$ turn into renormalizable ones. For instance, ϕ^6 in $D = 3$. Gauge theories become super-renormalizable for $D < 4$.

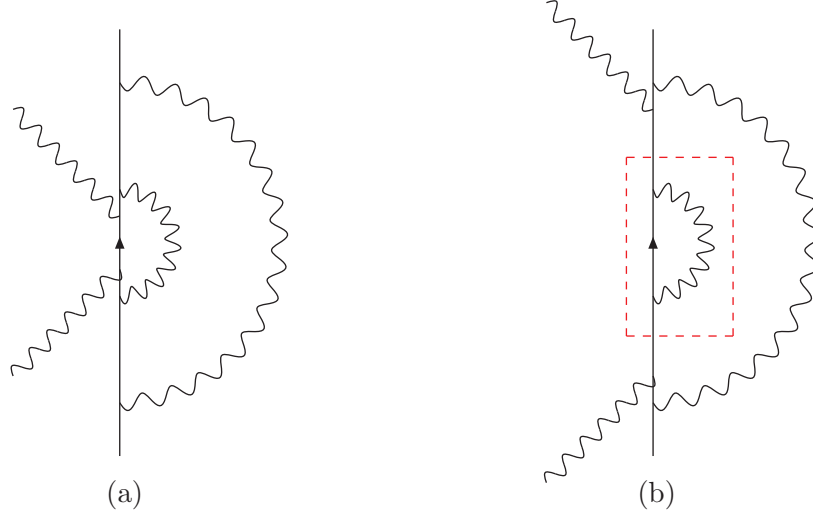


Figure 3.1: Two superficially convergent graphs contributing at two-loop level to the Compton scattering. Graph (a) is absolutely convergent, graph (b) is divergent, due to the subintegration associated to the sub-graph enclosed in the dashed red rectangle.

3.2 Absolute Convergence of a Graph

Multi-loop graphs can be divergent, despite being superficially convergent. This happens when the divergence arises from a sub-set of all virtual momenta becoming large, keeping fixed the others. As an example, consider the two two-loop contributions to the Compton scattering in QED in figure 3.1. In QED we have $f = e, \gamma$ and no index i is needed, since we have one interaction only, with $\Delta = 0$ (the theory is renormalizable). Applying eq. (3.1.13) to these graphs, we get

$$\delta(\mathcal{G}_a) = \delta(\mathcal{G}_b) = 4 - 2 \times 1 - 2 \times \frac{3}{2} = -1. \quad (3.2.1)$$

Both graphs are superficially convergent, but while graph (a) is actually convergent, graph (b) is divergent. This is seen by keeping fixed the virtual momentum running over the big external loop while letting the virtual momenta in the small loop go to infinity. In this situation, the small loop behaves as a sub-graph with effectively $E_e = 2$ and $E_\gamma = 0$, that has $\delta = 1$. Let us see this in some further detail. Neglecting unnecessary factors, the amplitude associated to the graphs (a) and (b) in fig. 3.1 is schematically given by

$$\begin{aligned} \mathcal{M}_a &\sim \int d^4 q_1 d^4 q_2 \frac{1}{\not{p}_1 + \not{q}_1} \frac{1}{q_1^2} \frac{1}{\not{p}_1 + \not{q}_1 + \not{q}_2} \frac{1}{q_2^2} \frac{1}{\not{p}_1 + \not{p}_2 + \not{q}_1 + \not{q}_2} \frac{1}{\not{p}_4 + \not{q}_1 + \not{q}_2} \frac{1}{\not{p}_4 + \not{q}_1}, \\ \mathcal{M}_b &\sim \int d^4 q_1 d^4 q_2 \frac{1}{\not{p}_1 + \not{q}_1} \frac{1}{q_1^2} \frac{1}{(\not{p}_1 + \not{p}_2 + \not{q}_1)^2} \frac{1}{q_2^2} \frac{1}{\not{p}_1 + \not{p}_2 + \not{q}_1 + \not{q}_2} \frac{1}{\not{p}_4 + \not{q}_1}. \end{aligned} \quad (3.2.2)$$

We have denoted by q_1 and q_2 the virtual momenta of the big and small photon loops, respectively, by p_1 and p_4 the electron momenta and by p_2 and p_3 the photon momenta,

and used momentum conservation to write eq. (3.2.2) in terms of p_1 , p_2 and p_4 . By rescaling $q_{1,2} \rightarrow \lambda q_{1,2}$ in eq. (3.2.2) one immediately recovers $\delta = -1$ for both graphs. However, if one keeps q_1 fixed, $\mathcal{M}_a$ is still finite when integrating in q_2 , while $\mathcal{M}_b$ diverges, with an effective degree of divergence of one, in agreement with the analysis above.

In general, a graph is convergent if $\delta < 0$ not only when all the virtual momenta are taken large, but also when any combination of virtual momenta is taken large, with the remaining ones kept fixed. When this is the case, we say that the graph is absolutely convergent.

Luckily enough, we do not need to classify divergent multi-loops graphs that are superficially convergent in order to complete the cancellation of divergences in a QFT.

3.3 Cancellation of Divergences and Local Counterterms

Any graph $\mathcal{G}$ with E_f external lines is a function of the momenta of the external particles. If $\mathcal{G}$ is divergent, one might naively argue that this function is ill-defined and does not make sense. On the contrary, there is a way to disentangle the divergent part from the rest, which remains finite and calculable. In order to understand how divergences can be disentangled, it is enough to consider one-dimensional divergent integrals, toy versions of the more involved integrals appearing in real amplitudes. Take for instance

$$I(q) = \int_0^\infty \frac{dk}{q+k}. \quad (3.3.1)$$

The function $I(q)$ is divergent, with $\delta = 0$. Differentiating with respect to the “external momentum” q gives a finite integral, with $\delta = -1$:⁴

$$I'(q) = - \int_0^\infty \frac{dk}{(q+k)^2} = -\frac{1}{q}. \quad (3.3.2)$$

Integrating back, we get

$$I(q) = -\log q + c \quad (3.3.3)$$

Evidently, from eq. (3.3.3), the divergence of $I(q)$ is all encoded in the constant c , while the $\log q$ term is finite and calculable. Consider another example, with $\delta = 1$:

$$\hat{I}(q) = \int_0^\infty \frac{k dk}{q+k}. \quad (3.3.4)$$

We now need to differentiate twice with respect to q to get a finite result. Integrating back twice gives

$$\hat{I}(q) = q \log q + aq + b. \quad (3.3.5)$$

⁴It is understood here and in the following that these integrals have been “regularized” (i.e. made finite). In this way derivatives commute with the integration and can act directly on the integrand.

As before all the divergences are encoded in the two coefficients a and b , while the non-polynomial $q \log q$ term is perfectly finite. The generalization of these results is obvious. A graph with $\delta > 0$ can be made convergent by taking $\delta + 1$ derivatives with respect to the external momenta and all the divergences are encoded in the coefficients of a polynomial of degree δ in the external momenta. Notice that in general the graph will be only superficially, and not absolutely, convergent. We will ignore for the moment this problem, postponing its solution for later. Terms polynomial in momenta can be written, in configuration space, as derivatives of a local Lagrangian density. In other words, *divergences only affect terms that can be written as local interactions in a Lagrangian*. By “interactions” we also include terms quadratic in the fields. The divergences of a graph with degree of divergence δ can be reabsorbed by adding to the Lagrangian $\delta + 1$ vertices with $n_{if} = E_f$, of the form

$$\mathcal{L} \supset a_0 \prod_f \phi_f^{E_f} + a_1 \partial \prod_f \phi_f^{E_f} + \dots + a_\delta \partial^\delta \prod_f \phi_f^{E_f}. \quad (3.3.6)$$

If the theory already contains such interactions, the divergences simply amount to a redefinitions of the corresponding couplings, or to a redefinition of the field themselves for 2-point functions. If they were not present, they are induced by quantum effects. In a non-renormalizable theory, where Green functions eventually diverge for any E_f , cancellation of the divergences imposes to us to add to the Lagrangian all possible terms of the form (3.3.6). Of course, not every possible term is generated, but only those compatible with the global symmetries of the Lagrangian. In a renormalizable theory, the number of terms in eq. (3.3.6) is finite. Moreover, the dimension of the couplings a_i in eq. (3.3.6) induced by the divergences is

$$[a_n] = 4 - \sum_f E_f (s_f + 1) - n, \quad n = 0, \dots, \delta. \quad (3.3.7)$$

For a renormalizable theory, we have $4 - \sum_f E_f (s_f + 1) \geq \delta$ and hence

$$[a_n] \geq \delta - n \geq 0, \quad n = 0, \dots, \delta. \quad (3.3.8)$$

The couplings induced by quantum effects in a renormalizable theory are themselves renormalizable. These theories are self-consistent at the quantum level, provided we write down all the couplings with $\Delta_i \geq 0$ allowed by symmetries. In renormalizable theories the finite number of local counter-terms a_n is enough to remove all divergences and make the theory finite. Since these terms were already present in the original Lagrangian, all the divergences can be buried into a redefinition of the interactions, masses and fields. The fields, masses and couplings in the initial Lagrangian are called “bare” fields or parameters and are not physical. We will use a subscript B to generally denote a bare parameter or

field. A way to effectively generate, order by order in perturbation theory, the terms in eq. (3.3.6) is obtained by redefining the bare fields and parameters in terms of the physical ones, by means of so called “counter-terms”. Schematically

$$\begin{aligned}(\phi_f)_B &= \sqrt{Z_{\phi_f}} \phi_f, \quad Z_{\phi_f} = 1 + \dots \\(a_n)_B &= Z_n a_n, \quad Z_n = 1 + \dots,\end{aligned}\tag{3.3.9}$$

where $\dots$ are divergent terms, induced by the interactions. The parameters Z_{ϕ_f} and Z_n are denoted counter-terms. The relations in (3.3.9) are quite schematic and more useful reparametrizations can be used, depending on the theory under consideration. For instance, a more standard definition of counter-terms in QED is given in eqs.(3.4.13), (3.4.15) below. Order by order in perturbation theory the $\dots$ appearing in the counter-terms induce new vertices that should themselves be inserted in Feynman graphs. The explicit form of counter-terms in QED and an example on how they should be inserted in loops will be provided in the next section.

It is now time to come back to a problem mentioned below eq. (3.3.5), related to the fact that in general taking derivatives with respect to external momentum does not necessarily make the graph absolutely convergent. An example of a graph of this sort is given by the two-loop graph (b) in fig. 3.1. If we take derivatives with respect to the external momenta of $\mathcal{M}_b$ in eq. (3.2.2), say p_1 or p_4 , we see that when the derivative acts on the q_2 -independent propagators we do not improve the convergence of the q_2 -loop integration, that in fact remains divergent. This problem is solved by noticing that, together with the graph (b) of fig. 3.1, we should also consider the graph in which the one-loop enclosed in the red rectangle is replaced by the fermion propagator counter-term. The divergence in the small loop is cancelled when summing the two diagrams. This example makes clear that counter-terms have to be considered in evaluating Feynman diagrams like ordinary interaction vertices and should be inserted in graphs consistently to the order in perturbation theory one is considering.

The proof that all divergences are cancelled by the counter-terms to all orders in perturbation theory is not trivial and is known as the BPHZ theorem, from the authors Bogoliubov, Parasiuk, [10] Hepp [33] and Zimmermann [64] that at different stages have proved this statement. It is thanks to this powerful theorem that we do not need to worry about superficially convergent, but actually divergent graphs. The BPHZ theorem ensures us that once we have cancelled by means of counterterms the divergences associated to superficially divergent graphs, all other divergences will automatically be cancelled by a systematic use of counterterms.

Renormalizability puts stringent constraints on the possible interactions appearing

in the Lagrangian. However, upon including all possible interactions, all theories become renormalizable, in the sense that we can get rid of all possible infinities. The key difference between renormalizable and non-renormalizable theories is in their predictivity. There is nothing wrong with non-renormalizable theories. If we have a coupling constant g with scaling dimension $\Delta < 0$, we can always write it as

$$g = \frac{1}{M^{|\Delta|}}. \quad (3.3.10)$$

In scattering amplitudes, by dimensional analysis, the coupling g will always appear as $(E/M)^{|\Delta|}$, where E is the energy scale of the process under consideration. For $E \ll M$, the coupling g is irrelevant (in fact, non-renormalizable couplings are also called irrelevant couplings, see section 5.1). Despite the fact that the theory is formally non-renormalizable and requires an infinite number of counter-terms, in practice at sufficiently low energies most of these interactions are negligible and make the non-renormalizable theory useful and predictive enough. A relevant example of non-renormalizable theory is gravity. Seen as a QFT, gravity is among the most non-renormalizable theories. The mass scale entering the non-renormalizable couplings like in eq. (3.3.10) is $M_{Pl} \sim 10^{18}$ GeV. We do not know what is the quantum theory of gravity (although we have some candidates, most notably string theory). Unfortunately (or fortunately, depending on the point of view), in all practical cases, M_{Pl} is so large that gravity effects are always negligible in high energy physics. Non-renormalizable theories might also appear from renormalizable ones when some massive fields are integrated out. A relevant case is the four-fermion electroweak interaction, approximation of the Standard Model when one integrates out the W and Z bosons. At sufficiently low energies $E \ll M_W, M_Z$, the Fermi theory provides an excellent description of the physics.

In the modern approach to QFT we consider all theories (possibly with the exception of the ultimate theory of nature) as effective field theories, namely theories valid up to some energy scales. More on effective field theories in chapter 7.

3.4 Regularization and Renormalization

Dealing with divergent Green functions can lead to subtleties and misleading results. For this reason it is highly desirable to modify the original theory so that divergences no longer appear. This process is called regularization. There are various ways to do that and we will not list all of them. The regularization process should be done in a way such that the modified Lagrangian has all the symmetries present in the original one.⁵

⁵As we will see, there are symmetries that are not respected by any regulator. When this happens, the symmetry is called anomalous, since it is no longer there at the quantum level. Anomalies will be the

We only mention two relevant types of regularization, that will be used throughout the course: cut-off regularization and dimensional regularization. The first is one of the simplest and most physical way to regularize a theory. It consists in truncating the integration over the virtual momenta to a limiting value Λ . This is typically done in a Lorentz-invariant way by rotating the virtual momenta to euclidean space and imposing the cut-off to the modulus of the momentum in radial coordinates. In this way all UV divergences are removed and Green functions become Λ -dependent. Divergences appear as power-like or logarithmic terms in Λ in an expansion for $\Lambda \rightarrow \infty$. The physical and reasonable assumption underlying cut-off regularization is that at short distances $L \sim 1/\Lambda$ the theory is no longer valid, and we parametrize our ignorance by cutting the momenta at that scale. The regularized theory differs from the original one only in the UV, while their IR behaviours are similar, a welcome feature. Cut-off regularization has however a serious drawback: it does not respect gauge invariance.

The second is the most commonly used regularization in high energy physics: Dimensional Regularization (DR). It consists in changing the space-time dimensions from 4 to $4 - \epsilon$, with $\epsilon \ll 1$. We make sense of the notion of a non-integer number of dimensions by analytic continuation. In the dimensionally regulated Lagrangian UV divergences no longer appear. In an expansion for small ϵ they are given by poles in $1/\epsilon^n$, with $n \geq 1$. Dimensional regularization respects gauge invariance and is one of the simplest regulators to use. Most of the computations in the literature, in particular with gauge theories, are performed using DR. In addition, DR has the great virtue of respecting dimensional analysis and is hence the most suitable regularization in the context of effective field theories. More on this point in section 7.8. Dimensional regularization can only be defined perturbatively, in contrast to cut-off regularization, that admits a non-perturbative definition. Moreover, it modifies the theory to all scales, not only in the UV, and for this reason it has to be used with some care, when mass scales are present. More on this point in section 5.4. Further technical details on how to use DR are discussed in subsection 3.5.

We summarize the key pros and cons of the two regularizations in table 3.1.

In the regularized theory we can meaningfully manipulate the Green functions, now finite. We have seen in section 3.3 that divergences only appear in local terms in Green functions and can be removed by redefining the fields, couplings and masses of the original Lagrangian. The process of redefining fields, couplings and masses, so that infinities no longer appear in correlation functions is called *renormalization*. It is important to distinguish the two key steps in the entire renormalization process:

subject of chapter 9.

	Cut-off	DR
Gauge invariance	No	Yes
Ideal in EFT	No	Yes
Non-perturbative	Yes	No

Table 3.1: Schematic comparison of Cut-off and DR. EFT stands for Effective Field Theory, see chapter 7.

- i. regularization, i.e. making the theory finite.
- ii. renormalization, i.e. removing the would-be infinities from the regularized theory and take the limit (e.g. $\Lambda \rightarrow \infty$ or $\epsilon \rightarrow 0$ in cut-off or dimensional regularization, respectively) giving back the original, now finite, theory.

In removing infinities we always have the freedom to add or subtract finite terms, so step ii. is not uniquely defined and needs a prescription. This prescription goes under the name of *renormalization scheme*. When dealing with renormalization, we can choose different regularizations and renormalization schemes. So we can have two different regularizations using the same renormalization scheme, or the same regularization but different renormalization schemes. For instance, we could use the form (2.1.28) of the exact propagator in an interacting theory given by the Källén-Lehmann representation to *define* the physical mass and the properly normalized quantum field:

$$iG_R^{(2)-1}(q^2 = m^2) = 0, \quad i \frac{dG_R^{(2)-1}(q^2)}{dq^2} \Big|_{q^2=m^2} = 1. \quad (3.4.1)$$

One could analogously define a physical coupling constant by the value of a Green function at some specific kinematic point. These conditions define what is called a physical, or on-shell renormalization scheme. Schemes of this kind are useful because they are directly related to physical parameters. On the other hand, important functions that we will discuss later on, such as β -functions and field anomalous dimensions, have complicated expressions in perturbation theory. In particular, they depend on the particle masses. Renormalization schemes where β -functions and field anomalous dimensions depend on masses are denoted mass-dependent. It is useful to introduce renormalization schemes where correlation functions are made finite by choosing the minimal counterterm needed to remove the divergence. In this way β -functions and field anomalous dimensions can be made simpler and independent of particle masses (mass-independent schemes). This comes at the cost that the resulting masses, couplings and fields will no longer be physical, though they can be related to the physical ones with a finite mapping, to be performed

at the end of the computation. A notable renormalization scheme of this kind is denoted minimal subtraction and is mostly adopted when using DR.

3.4.1 QED Case

Let us consider as an example the renormalization of QED, namely one charged massive fermion (electron) interacting with a photon. First of all, let us use eq. (3.1.13) to classify the possible divergent Green functions. We have only one dimensionless coupling constant and $f = e, \gamma$, so

$$\delta = 4 - \frac{3}{2}E_e - E_\gamma. \quad (3.4.2)$$

The divergent Green functions contain graphs with $\delta \geq 0$. Given eq. (3.4.2), they are $(E_e = 0, E_\gamma = 0, 1, 2, 3, 4)$, $(E_e = 1, E_\gamma = 0, 1, 2)$, $(E_e = 2, E_\gamma = 0, 1)$. Lorentz invariance forbids Green functions with an odd number of electron lines, while charge conjugation implies that $(E_e = 0, E_\gamma = 1, 3)$ vanishes. We neglect the quartically divergent vacuum amplitude $(E_0 = E_\gamma = 0)$ since it is irrelevant in a QFT where gravity is not dynamical. In total, we are left with four cases: $(E_e = 0, E_\gamma = 2, 4)$, $(E_e = 2, E_\gamma = 0, 1)$.

$E_e = 2, E_\gamma = 1$ We denote this Green function by $\Gamma_\mu^{ab}(p, p')$, where a, b are spinor indices and p and p' are the momenta of the electron lines. It corresponds to the basic QED interaction given by minimal coupling. Since $\delta(\Gamma_\mu) = 0$, the divergence can only appear in a constant term, independent of p and p' . Given the Lorentz structure of Γ_μ , we necessarily have (omitting spinor indices from now on)

$$i\Gamma_\mu(p, p') = ieL\gamma^\mu + i\Gamma_\mu^{finite}(p, p'), \quad (3.4.3)$$

where L is a divergent constant, possibly reabsorbed in the definition of the electric charge.

$E_e = 2, E_\gamma = 0$ We denote this Green function by $\Sigma(p)$. It corresponds to the inverse of the electron propagator. We have $\delta(\Sigma) = 1$, so divergences proportional to a constant and to a term linear in momentum can appear. The Lorentz structure implies

$$i\Sigma(p) = iA + iB\not{p} + i\Sigma^{finite}(p), \quad (3.4.4)$$

where A and B are divergent constants. Gauge invariance imposes the identity

$$B = L, \quad (3.4.5)$$

to all orders in perturbation theory.

$\mathbf{E}_e = \mathbf{0}, \mathbf{E}_\gamma = \mathbf{2}$ We denote this Green function by $\Pi_{\mu\nu}(q)$. It corresponds to the inverse of the photon propagator. We have $\delta(\Pi) = 2$, so divergences up to terms quadratic in q might appear. Lorentz invariance forbids terms linear in q , so we have

$$i\Pi_{\mu\nu}(q) = iC_1\eta_{\mu\nu} + iC_2q^2\eta_{\mu\nu} + iC_3q_\mu q_\nu + i\Pi_{\mu\nu}^{finite}(q). \quad (3.4.6)$$

Dimensional arguments only would imply the appearance of three independent divergent constants. However, gauge invariance implies the following identity:

$$q^\mu \Pi_{\mu\nu}(q) = q^\nu \Pi_{\mu\nu}(q) = 0. \quad (3.4.7)$$

Using eq. (3.4.7) we have

$$\begin{aligned} C_1 &= 0, \quad C_2 = -C_3 \equiv C, \\ \Pi_{\mu\nu}^{finite}(q) &= (\eta_{\mu\nu}q^2 - q_\mu q_\nu)\Pi^{finite}(q^2) \end{aligned} \quad (3.4.8)$$

and hence

$$i\Pi_{\mu\nu}(q) = (\eta_{\mu\nu}q^2 - q_\mu q_\nu)(iC + i\Pi^{finite}(q^2)). \quad (3.4.9)$$

$\mathbf{E}_e = \mathbf{0}, \mathbf{E}_\gamma = \mathbf{4}$ We denote this Green function by $M_{\mu\nu\rho\sigma}(p_i)$, with p_i ($i = 1, 2, 3, 4$) the momenta of the external photons. It corresponds to the $2 \rightarrow 2$ photon scattering. We have $\delta(M) = 0$, implying a possible divergence in the constant factor. Lorentz invariance and Bose symmetry of the external legs fix the Green function to be

$$M_{\mu\nu\rho\sigma}(p_i) = K(\eta_{\mu\nu}\eta_{\rho\sigma} + \eta_{\mu\rho}\eta_{\nu\sigma} + \eta_{\mu\sigma}\eta_{\nu\rho}) + M_{\mu\nu\rho\sigma}^{finite}(p_i) \quad (3.4.10)$$

with K the divergent constant. Once again, gauge invariance imposes other constraints and requires that

$$p_1^\mu M_{\mu\nu\rho\sigma}(p_i) = p_2^\nu M_{\mu\nu\rho\sigma}(p_i) = p_3^\rho M_{\mu\nu\rho\sigma}(p_i) = p_4^\sigma M_{\mu\nu\rho\sigma}(p_i) = 0. \quad (3.4.11)$$

The above relation implies that $K = 0$ and that a relation similar to eq. (3.4.11) applies with $M \rightarrow M^{finite}$.⁶ We will prove eqs. (3.4.5), (3.4.7) and (3.4.11) in section 4.4.3.

Summarizing, taking gauge invariance into account, three Green functions diverge in QED, the photon and electron propagators and the interaction vertex, for a total of three independent divergences encoded in the constants L , A , C in eqs. (3.4.3), (3.4.4), (3.4.9).

⁶It is easy to understand why gauge invariance forbids the presence of the factor K in eq.(3.4.10) or C_1 in eq.(3.4.6). If present, these divergences would require counter-terms in the Lagrangian of the form $(A_\mu A^\mu)^2$ and $A_\mu A^\mu$ respectively, which do not respect gauge invariance.

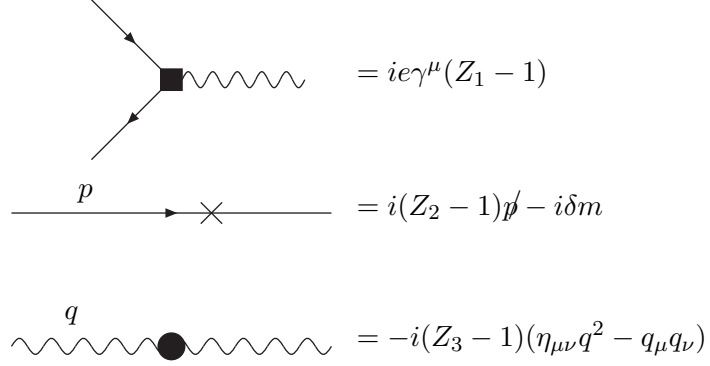


Figure 3.2: Feynman rules for the QED counter-terms defined in eq. (3.4.16).

Let us now see how these divergences can be cancelled, to all orders in perturbation theory, by suitable local counter-terms. The QED Lagrangian reads

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{B,\mu\nu}F_B^{\mu\nu} + \bar{\psi}_B(i\not{D}_B(A_B) - m_B)\psi_B. \quad (3.4.12)$$

Here $D_{B,\mu}(A_B) = \partial_\mu - ie_B A_{B,\mu}$ is the covariant derivative and the subscript B reminds us that the fields, mass and coupling constant are the bare ones. The physical fields are defined as

$$\psi_B = \sqrt{Z_2}\psi, \quad A_{B,\mu} = \sqrt{Z_3}A_\mu, \quad (3.4.13)$$

We then have

$$\mathcal{L}_{QED} = -\frac{1}{4}Z_3F_{\mu\nu}F^{\mu\nu} + Z_2\bar{\psi}i\not{\partial}\psi - m_BZ_2\bar{\psi}\psi + e_BZ_2\sqrt{Z_3}\bar{\psi}A_\mu\gamma^\mu\psi, \quad (3.4.14)$$

and define renormalized mass and electric charge as

$$e_BZ_2\sqrt{Z_3} = Z_1e, \quad m_BZ_2 = m + \delta m. \quad (3.4.15)$$

In terms of the physical fields and parameters, the QED Lagrangian (3.4.12) can be rewritten as

$$\begin{aligned} \mathcal{L}_{QED} = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\not{D}(A) - m)\psi \\ & -\frac{1}{4}(Z_3 - 1)F_{\mu\nu}F^{\mu\nu} + (Z_2 - 1)\bar{\psi}i\not{\partial}\psi - \delta m\bar{\psi}\psi + e(Z_1 - 1)\bar{\psi}A_\mu\gamma^\mu\psi. \end{aligned} \quad (3.4.16)$$

The first line in eq. (3.4.16) is the QED Lagrangian written now in terms of the physical field and parameters, and the second line contains the counter-terms. The latter are unphysical local terms which are there just to cancel the divergences appearing in the

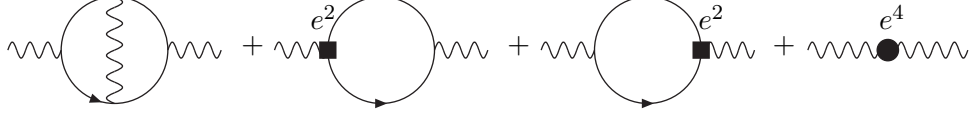


Figure 3.3: Two-loop graph contributing to the photon propagator and its additional graphs including counter-terms canceling all the divergences. The e^2 and e^4 factors denote the order in which the counter-terms Z_2 and Z_3 enter the various graphs.

loop diagrams. They also include terms quadratic in the fields. However, contrary to the canonical kinetic terms in the first line, they should be considered as *new effective vertices*. This is clear by noticing that for $e = 0$ (free theory), no divergences occur and all counter-terms vanish. For $e \ll 1$, all the counter-terms can be Taylor expanded and are given by positive powers of e . As such, they should be considered as interaction terms. In evaluating Feynman diagrams, we should take into account the new contributions given by the counter-terms. The Feynman rules associated to these interactions are reported in fig. 3.2. The four counter-terms Z_1 , Z_2 , Z_3 and δm in the Lagrangian (3.4.16) are not all independent. In particular, gauge invariance requires

$$Z_1 = Z_2, \quad (3.4.17)$$

which is equivalent to eq. (3.4.5). This will be proved in section 4.4.3. Order by order in perturbation theory, the values of the counter-terms $Z_1 = Z_2$, Z_3 and δm are suitably chosen to cancel the divergences appearing in $L = B$, C and A , respectively. For instance, at one-loop level, we have $L = e^2 L_{(1)} + \dots$, $C = e^2 C_{(1)} + \dots$, $A = e^2 A_{(1)} + \dots$. Adding the counter-term contributions, the tree-level + one-loop Green functions read

$$\begin{aligned} \Gamma_\mu(p, p') &= e\gamma^\mu(1 + e^2 L_{(1)} + Z_1 - 1) + \Gamma_\mu^{finite}(p, p'), \\ \Sigma(p) &= (1 + e^2 L_{(1)} + Z_2 - 1)\not{p} - m + e^2 A_{(1)} - \delta m + \Sigma^{finite}(p), \\ \Pi_{\mu\nu}(q) &= (\eta_{\mu\nu}q^2 - q_\mu q_\nu)(1 + e^2 C_{(1)} - Z_3 + 1 + \Pi^{finite}(q^2)). \end{aligned} \quad (3.4.18)$$

from which we get the values of the counter-terms up to order e^2 :

$$Z_1 - 1 = -e^2 L_{(1)}, \quad Z_2 = Z_1, \quad Z_3 = 1 + e^2 C_{(1)}, \quad \delta m = e^2 A_{(1)}. \quad (3.4.19)$$

In any regularisation, the choice of renormalization scheme is implicitly encoded in eq.(3.4.19). Different renormalization schemes would correspond to differences in finite terms in $L_{(1)}$, $C_{(1)}$ and $A_{(1)}$.

Let us now consider the two-loop correction to the photon propagator shown in fig. 3.3. In order to remove all divergences from this graph, one has to consider the one-loop graphs where one vertex is replaced by a counter-term, up to order e^2 . Then a remaining local divergence is finally cancelled by considering the photon propagator counter-term at order e^4 . This is an example of how a systematic use of counter-terms allows us to remove all the UV divergences of a theory.

3.5 Dimensional Regularization

Dimensional Regularization (DR) is a regularization for QFTs introduced by 't Hooft and Veltman [57]. It consists in changing the number of space-time dimensions from 4 to d , where d is generally a non-integer (even complex) number. In this way divergent expressions become finite thanks to analytic continuation. Despite its exotic nature and not so clear physical interpretation, DR is widely used in the renormalization of field theories for a variety of reasons that make it preferable to other regularization schemes, such as cut-off, Pauli-Villars, etc. For example, DR does not break gauge invariance, it is quite practical in actual calculations, and it respects naive dimensional analysis when applied in effective field theories. Dimensional regularization can be formally introduced via a set of axioms, as explained, e.g., in ref. [19]. Here we briefly discuss its practical aspects, starting from the simple example of the (Euclidean) integral (tadpole of ϕ^4 theory)

$$T_d \equiv \int \frac{d^d p_E}{(2\pi)^d} \frac{1}{p_E^2 + m^2} = m^{d-2} \frac{\Gamma(1 - d/2)}{(4\pi)^{d/2}}, \quad (3.5.1)$$

(where m is the mass of the scalar) which converges in the IR as long as $m^2 \neq 0$ and in the UV only for $d < 2$. In this case, the r.h.s. is easily calculated by recalling that $1/A = \int_0^\infty dt e^{-tA}$, which turns T_d into a Gaussian integral in d -dimensions that, once calculated, allows to express T_d in terms of $\Gamma(x) = \int_0^\infty dt t^{x-1} e^{-t}$, which converges for $x > 0$. Note that, while in the intermediate steps of the calculation d has been treated as an integer number, the r.h.s. of eq. (3.5.1) is well defined (via the analytic continuation of the Γ function⁷) for any complex d which is not an even integer. This simple observation constitutes, in practice, a *regularization* of the integral T_d because it now assumes a finite value (given by the r.h.s. of eq. (3.5.1)) for (almost) all d , analogously to what happens when the regularization is done by other techniques, such as the introduction of a large-momentum cutoff Λ . However, a reminder of the divergent nature of the original integral is in the fact that the r.h.s. itself diverges if d approaches even integer values ≥ 2 .

⁷As we discuss below, the analytic continuation of integrals such as eq. (3.5.1) can be done without invoking Γ functions and their properties.

An important and general property of this regularization method is that the regularized integral (3.5.1) in the massless case $m \rightarrow 0$ vanishes for $d > 2$, leading to the relation

$$\int \frac{d^d p_E}{(2\pi)^d} \frac{1}{p_E^2} = 0. \quad (3.5.2)$$

Because of the absence of a mass scale, the IR and UV behaviors of the original integral are the same and there is no value of d for which the integral converges in the absence of suitable cutoffs. In general, DR sets consistently to 0 all the integrals of this kind, i.e., those which are homogeneous functions of the momenta and therefore lack a mass scale.

As stated above, dimensionally regularized integrals are characterized by dimensional poles. In order to understand their meaning consider T_d in $d = 4$ regularized using a UV cutoff Λ . For $\Lambda \gg m$ we have

$$T_4^{(\Lambda)} = a_2^{(4)} \Lambda^2 + a_0^{(4)} \ln\left(\frac{\Lambda}{m}\right) + \mathcal{O}(\Lambda^{-2}), \quad a_2^{(4)} = \frac{1}{16\pi^2}, \quad a_0^{(4)} = -\frac{m^2}{8\pi^2}. \quad (3.5.3)$$

On the other hand, for $d = 4 - \epsilon$, eq.(3.5.1) gives

$$T_{4-\epsilon} = \frac{a_0^{(4)}}{\epsilon} + \mathcal{O}(1), \quad (3.5.4)$$

with $a_0^{(4)}$ as in eq.(3.5.3). In DR the UV divergence of T_4 arises as a simple pole in $1/(4-d)$. Interestingly enough, the residue at this pole equals the coefficient of the $\ln(\Lambda/m)$ term in $T_4^{(\Lambda)}$. There is no remnant of the quadratically divergent term in eq.(3.5.3), which is set to zero in DR, in agreement with eq.(3.5.2). Dimensional regularization can of course be used also for space-time dimensions different from 4. If for some physical space-time dimension an integral regularized with a cutoff shows a logarithmic term $\sim \ln(\Lambda/m)$, a dimensional pole is expected in DR. For instance, for $d = 2$ we have

$$T_2^{(\Lambda)} = a_0^{(2)} \ln\left(\frac{\Lambda}{m}\right) + \mathcal{O}(\Lambda^{-2}), \quad a_0^{(2)} = \frac{1}{2\pi}, \quad (3.5.5)$$

while from eq.(3.5.1) we get

$$T_{2-\epsilon} = \frac{a_0^{(2)}}{\epsilon} + \mathcal{O}(1), \quad (3.5.6)$$

with $a_0^{(2)}$ as in eq.(3.5.5). In $d = 3$, $T_3^{(\Lambda)}$ does not contain log terms and correspondingly T_3 is finite in DR, as we can immediately see by the fact that $\Gamma(-1/2) = -2\sqrt{\pi}$.

For any given fixed space time dimensionality, the divergent part of a one-loop integral in DR, i.e. the residue of the $1/\epsilon$ pole, can be computed without using Γ functions. Let us consider for simplicity T_d in eq.(3.5.1). We can go to radial coordinates and write

$$T_d = \frac{\Omega_d}{(2\pi)^d} \int_0^\infty dp_E \frac{p_E^{d-1}}{p^2 + m^2}, \quad (3.5.7)$$

where

$$\Omega_d = \frac{2\pi^{d/2}}{\Gamma(d/2)} \quad (3.5.8)$$

is the solid angle in d -dimensions. If $d = 2n$ dimensions, the pole in $d = 2n$ will arise from the term of order $1/(p^2)^n$ in the expansion for $p^2 \gg m^2$ of the $1/(p^2 + m^2)$ factor. Since

$$\frac{1}{p_E^2 + m^2} = \frac{1}{p_E^2} \sum_{n=1}^{\infty} \left(\frac{-m^2}{p_E^2} \right)^{n-1} \quad (3.5.9)$$

the residue $a_0^{(2n)}$ of the $1/\epsilon$ pole for $\epsilon = d - 2n$ reads

$$a_0^{(2n)} = (-m^2)^{n-1} \frac{\Omega_{2n}}{(2\pi)^{2n}}, \quad (3.5.10)$$

which generalize for all dimensions the values found in eqs.(3.5.3) and (3.5.5) for $d = 2$ and $d = 4$, respectively. For instance, the divergence of the 1-loop integral in $d = 4 - \epsilon$

$$I_4 \equiv \int \frac{d^d p}{(2\pi)^d} \frac{i}{p^2 - m^2} \frac{i}{(p - q)^2 - m^2} \quad (3.5.11)$$

can be extracted with no need of introducing a Feynman parameter, performing a shift in the momentum p and computing the whole integral. We can simply expand the denominators and look for the term of order $1/(p^2)^2$:

$$I_4|_{\text{div}} = \int \frac{d^d p}{(2\pi)^d} \frac{i^2}{(p^2)^2} = i \int \frac{d^d p_E}{(2\pi)^d} \frac{i^2}{(p_E^2)^2} = -i \frac{\Omega_d}{(2\pi)^d} \int_0^\infty p_E^{d-3} dp_E = -i \frac{\Omega_d}{(2\pi)^d} \frac{1}{\epsilon} = -\frac{i}{8\pi^2 \epsilon}. \quad (3.5.12)$$

A proper use of this trick allows to significantly simplify the computation of the divergent parts of one-loop diagrams when using DR. If finite parts are needed, one should use the usual route of computing Γ functions. For convenience of the reader, we report here the result for a class of one-loop integrals (α and β not necessarily integers):

$$\int \frac{d^d p_E}{(2\pi)^d} \frac{(p_E^2)^\alpha}{(p_E^2 + \Delta)^\beta} = \frac{\Delta^{\alpha-\beta+d/2}}{(4\pi)^{d/2}} \frac{\Gamma(\alpha + d/2)\Gamma(\beta - \alpha - d/2)}{\Gamma(d/2)\Gamma(\beta)}. \quad (3.5.13)$$

Chapter 4

External Fields and Generating Functionals

It is often useful to study the properties of QFT's coupled with external sources. Among other things, this is a useful way to repackage n -point correlation functions of a given kind, for any n , in generating functionals. It also turns out that this is the easiest way to discuss the consequences of symmetries in a QFT.

The structure of this chapter is as follows. After reviewing how to obtain the one-point irreducible generating functional, we discuss in section 4.1 why it can be seen as an effective action. We then introduce the notion of effective potential and compute its explicit form in a specific model ($\lambda\phi^4$ theory) at one-loop level. A subtlety related to effective potentials is discussed in section 4.3. Finally, we discuss in section 4.4 the Schwinger-Dyson equations and the Ward-Takahashi identities, quantum mechanical analogues of the classical equations of motions and of the conservation of currents, respectively. The section ends with a detailed derivation of Ward-Takahashi identities for QED.

4.1 The 1PI Effective Action

Consider a set of classical sources $J_r(x)$ coupled to some fields $\phi_r(x)$. The fields ϕ_r can be scalars, fermions or vectors, and will correspondingly couple to scalar, fermion or vector currents J_r . In general ϕ_r can also be composite fields, although we take them elementary for simplicity. We can define the J -dependent vacuum functional

$$\langle 0|0\rangle_J \equiv Z[J] = \int \mathcal{D}\phi_r e^{iS(\phi) + i \int d^4x J_r(x) \phi_r(x)} . \quad (4.1.1)$$

All correlation functions of the theory will be given by functional derivatives of $Z[J]$ with respect to J_r , evaluated at $J_r = 0$:

$$\frac{1}{Z} \frac{\delta^n Z[J]}{\delta J_{r_1}(x_1) \dots \delta J_{r_n}(x_n)} \Big|_{J_r=0} = i^n \langle 0 | T [\phi_{r_1}(x_1) \dots \phi_{r_n}(x_n)] | 0 \rangle. \quad (4.1.2)$$

The functional $Z[J]$ is the generator of all amplitudes, connected and disconnected. We can define a similar functional generating connected amplitudes only, $W[J]$. This is simply given by

$$W[J] = -i \log Z[J]. \quad (4.1.3)$$

Connected Green functions are defined similarly to eq.(4.1.2):

$$\frac{\delta^n W[J]}{\delta J_{r_1}(x_1) \dots \delta J_{r_n}(x_n)} \Big|_{J_r=0} = i^{n-1} \langle 0 | T [\phi_{r_1}(x_1) \dots \phi_{r_n}(x_n)] | 0 \rangle_c. \quad (4.1.4)$$

For instance, for a single field ϕ , we have

$$\frac{\delta W[J]}{\delta J(x)} \Big|_{J=0} = (-i) \frac{1}{Z} \frac{\delta Z}{\delta J(x)} \Big|_{J=0} = \langle 0 | \phi(x) | 0 \rangle, \quad (4.1.5)$$

and hence

$$\langle 0 | \phi(x) | 0 \rangle_c = \langle 0 | \phi(x) | 0 \rangle. \quad (4.1.6)$$

Taking one further functional derivative gives

$$\begin{aligned} \frac{\delta^2 W[J]}{\delta J(x) \delta J(y)} \Big|_{J=0} &= (-i) \left(\frac{1}{Z} \frac{\delta^2 Z}{\delta J(x) \delta J(y)} - \frac{1}{Z^2} \frac{\delta Z}{\delta J(x)} \frac{\delta Z}{\delta J(y)} \right) \Big|_{J=0} \\ &= i \left(\langle 0 | T [\phi(x) \phi(y)] | 0 \rangle - \langle 0 | \phi(x) | 0 \rangle \langle 0 | \phi(y) | 0 \rangle \right), \end{aligned} \quad (4.1.7)$$

from which we find the connected time-ordered two-point function

$$\langle 0 | T [\phi(x) \phi(y)] | 0 \rangle_c = \langle 0 | T [\phi(x) \phi(y)] | 0 \rangle - \langle 0 | \phi(x) | 0 \rangle \langle 0 | \phi(y) | 0 \rangle. \quad (4.1.8)$$

Continuing in this way iteratively allows us to find all connected higher point functions.

We can proceed further and define a functional $\Gamma[\Phi]$ that generates 1-particle irreducible amplitudes only.¹ The source $\Phi(x)$ is defined as (omitting for simplicity the index r from now on)

$$\Phi(x) \equiv \frac{\delta W[J]}{\delta J(x)} \quad (4.1.9)$$

and is then a (in general very complicated) functional of the external sources J .² The 1PI generator is defined as

$$\Gamma[\Phi] = W[J] - \int d^4x \Phi(x) J(x). \quad (4.1.10)$$

¹Recall that a 1PI amplitude is defined as the amplitude whose defining graphs cannot be divided in two by cutting one internal line.

²The analysis below follows in part section 16.1 of ref. [60].

It is the Legendre transform of $W[J]$, very much like the Lagrangian is the Legendre transform of the Hamiltonian. In eq. (4.1.10) it is understood that we have inverted eq. (4.1.9) so that $J = J[\Phi]$. We have

$$\frac{\delta\Gamma}{\delta\Phi(x)} = \int d^4y \left(\frac{\delta W}{\delta J(y)} - \Phi(y) \right) \frac{\delta J(y)}{\delta\Phi(x)} - J(x) = -J(x). \quad (4.1.11)$$

When $J = 0$, eq. (4.1.11) can be seen as an equation of motion of Φ with action Γ . Γ is in fact the quantum generalization of the action and is also denoted the quantum effective action. It is an effective action in the sense that all loop amplitudes can be calculated as sum of connected tree diagrams obtained using $\Gamma[\Phi]$ instead of $S[\phi]$.³ Thus, all loop effects are encoded in $\Gamma[\Phi]$. In order to see this, let us define the functional generator $W_\Gamma[J]$ as the generator of connected amplitudes obtained starting from an action given by Γ :

$$e^{iW_\Gamma[J,g]} = \int \mathcal{D}\phi \exp \left(\frac{i}{g} \left(\Gamma[\phi] + \int d^4x J(x)\phi(x) \right) \right). \quad (4.1.12)$$

In (4.1.12) we have introduced a parameter g , that plays the role of $\hbar$ and is a loop-counting parameter. For connected graphs the number of loops L is given by

$$L = I - V + 1, \quad (4.1.13)$$

where I is the total number of propagators and V is the total number of vertices in the graph. Indeed, this coincides with the number of independent virtual momenta entering the graph, the $+1$ coming from the conservation of the overall external momenta. We have already used this relation in eq. (3.1.6). We see from eq. (4.1.12) that the propagator of each field is proportional to g (inverse of the action), while any vertex goes like $1/g$. Thus the L -loop contribution to W_Γ , $W_\Gamma^{(L)}$, goes like $g^{I-V} = g^{L-1}$. In a loop expansion we can write

$$W_\Gamma[J, g] = \sum_{L=0}^{\infty} g^{L-1} W_\Gamma^{(L)}[J]. \quad (4.1.14)$$

Consider now in eq. (4.1.12) the tree-level contribution only, which can be selected by taking the limit $g \rightarrow 0$. By using a saddle point approximation we have

$$\lim_{g \rightarrow 0} e^{iW_\Gamma[J,g]} = e^{\frac{i}{g} W_\Gamma^{(0)}[J]} = \exp \left(\frac{i}{g} \left(\Gamma[\Phi] + \int d^4x J(x)\Phi(x) \right) \right), \quad (4.1.15)$$

where $\Phi(x)$ extremizes the exponential, namely

$$\left. \frac{\delta\Gamma}{\delta\phi(x)} \right|_{\phi=\Phi} + J(x) = 0. \quad (4.1.16)$$

³It is important to keep in mind that Γ , contrary to S , is *not* in general a local functional of the fields: $\Gamma[\Phi] \neq \int d^4x \mathcal{L}_{eff}(\Phi(x))$, see eq.(4.1.19).

Comparing eq. (4.1.10) with eq. (4.1.15) and eq. (4.1.11) with (4.1.16), we conclude that

$$W_{\Gamma}^{(0)}[J] = W[J]. \quad (4.1.17)$$

Stated differently, we can write

$$iW[J] = \int \underset{\substack{\text{Connected} \\ \text{tree graphs}}}{\mathcal{D}\phi} e^{i\Gamma(\phi) + i \int J\phi}. \quad (4.1.18)$$

Since any connected graph is a tree with 1PI vertices, we conclude that $\Gamma(\phi)$ generates all 1PI amplitudes:

$$\Gamma(\phi) = \sum_{n=0}^{\infty} \frac{1}{n!} \int d^4x_1 \dots \int d^4x_n \Gamma^{(n)}(x_1, \dots, x_n) \phi(x_1) \dots \phi(x_n). \quad (4.1.19)$$

The quantum effective action Γ can also be computed directly starting from the action S , without passing through $W[J]$. This is obtained as follows [38]. Starting from

$$e^{iW[J]} = \int \mathcal{D}\phi e^{iS(\phi) + i \int d^4x J(x)\phi(x)}, \quad (4.1.20)$$

define $\phi(x) = \phi_0(x) + \phi_q(x)$, where $\phi_0(x)$ is an arbitrary background configuration and $\phi_q(x)$ is the quantum field to be integrated over in the path integral. We then get

$$e^{iW[J]} = e^{i \int d^4x J(x)\phi_0(x)} \int \mathcal{D}\phi_q e^{iS(\phi_0 + \phi_q) + i \int d^4x J(x)\phi_q(x)}. \quad (4.1.21)$$

Bringing the first factor in the right hand side of eq.(4.1.21) to the left hand side gives

$$e^{i\Gamma[\phi_0]} = \int \mathcal{D}\phi_q e^{iS(\phi_0 + \phi_q) + i \int d^4x J(x)\phi_q(x)}, \quad (4.1.22)$$

where by definition

$$J(x) = -\frac{\delta\Gamma[\phi_0]}{\delta\phi_0(x)}, \quad \phi_0(x) = \langle\phi(x)\rangle_J = \frac{\delta W[J]}{\delta J(x)}, \quad (4.1.23)$$

namely the external current $J(x)$ has to be chosen such that the vacuum expectation value of the field equals $\phi_0(x)$ or equivalently $\langle\phi_q(x)\rangle_J = 0$. At leading order in a loopwise expansion, classically, we have $\Gamma[\phi_0] = S(\phi_0)$. Omitting the obvious x -dependence, we get

$$S(\phi_0 + \phi_q) + \int J\phi_q = S(\phi_0) + \int \frac{\delta S}{\delta\phi_0} \phi_q + \dots + \int \left(-\frac{\delta S}{\delta\phi_0} + \dots \right) \phi_q = S(\phi_0) + \dots$$

We see that the source term $\int J\phi_q$ precisely cancels the linear terms in ϕ_q coming from the expansion of $S(\phi_0 + \phi_q)$. At higher orders the source J should be modified in such a way that $\langle\phi_q(x)\rangle_J = 0$ order by order in a loopwise expansion. In practice we do not

need to actually determine J . A non-vanishing one-point function for ϕ_q can contribute to $\Gamma[\phi_0]$ only in 1-point reducible diagrams, so it is enough to restrict the path integral in eq.(4.1.22) to 1PI graphs and forget about J . In this way one has

$$e^{i\Gamma(\phi_0)} = \int_{1PI} \mathcal{D}\phi_q e^{iS(\phi_0+\phi_q)}, \quad (4.1.24)$$

where 1PI means that we should only consider 1PI graphs in evaluating the path integral. This formula is a direct consequence of the fact that Γ is the quantum version of S . Given a classical configuration ϕ_0 , not necessarily an extremum of S , the quantum action of ϕ_0 is given by integrating out quantum fluctuations ϕ_q around ϕ_0 . We invite the reader to check the validity (and the usefulness) of eq. (4.1.24) by formally computing $\Gamma(\phi)$ at one-loop level using eq. (4.1.24) and directly from the definition (4.1.10).

4.2 The Coleman-Weinberg Effective Potential

The effective potential V_{eff} is defined as the quantum action Γ , evaluated for constant field configurations ϕ_0 .⁴

$$\Gamma(\phi = \phi_0) = -V_4 V_{eff}(\phi_0), \quad (4.2.1)$$

where V_4 is the volume of space-time, coming from $\int d^4x$. It is the quantum analogue of the classical potential V_0 defined as $S(\phi = \phi_0) = -V_4 V_0(\phi_0)$. Let us explicitly compute V_{eff} at one-loop level for the simple scalar theory

$$\mathcal{L}(\phi) = \frac{1}{2}(\partial_\mu \phi_B)(\partial^\mu \phi_B) - \frac{1}{2}m_B^2 \phi_B^2 - \frac{\lambda_B}{4!} \phi_B^4. \quad (4.2.2)$$

First of all, we express the Lagrangian in terms of the physical parameters and fields:

$$\phi_B = \sqrt{Z} \phi, \quad m_B^2 Z = m^2 + \delta m^2, \quad \lambda_B Z^2 = Z_\lambda \lambda, \quad (4.2.3)$$

so that

$$\mathcal{L}(\phi) = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}m^2 \phi^2 - \frac{\lambda}{4!} \phi^4 + \mathcal{L}_{c.t.}(\phi), \quad (4.2.4)$$

where the counter-term Lagrangian reads

$$\mathcal{L}_{c.t.}(\phi) = \frac{1}{2}(Z-1)(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\delta m^2 \phi^2 - \frac{\lambda}{4!}(Z_\lambda-1)\phi^4. \quad (4.2.5)$$

We compute the effective potential starting from eq. (4.1.24) with ϕ_0 constant. We get

$$\begin{aligned} \mathcal{L}(\phi_0 + \phi) = & -\frac{1}{2}m^2 \phi_0^2 - \frac{\lambda}{4!} \phi_0^4 + \phi(-m^2 \phi_0 - \frac{\lambda}{6} \phi_0^3) \\ & + \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\phi^2(m^2 + \frac{\lambda}{2}\phi_0^2) - \frac{\lambda}{6}\phi_0\phi^3 - \frac{\lambda}{4!}\phi^4 + \mathcal{L}_{c.t.}(\phi_0 + \phi). \end{aligned} \quad (4.2.6)$$

⁴The analysis in this section closely follows section 16.2 of ref. [60].

We have ordered the terms according to the powers of quantum fluctuations ϕ . The first three terms in the first row of eq. (4.2.6) reproduce the initial tree-level potential. Terms cubic or quartic in ϕ cannot contribute at one-loop level and start contributing at two-loops. The term linear in ϕ can be neglected altogether: it can only enter in one-particle reducible graphs and never in 1PI amplitudes. The counter-terms are non-vanishing only at loop order, so at one-loop level we should only keep $\mathcal{L}_{c.t.}(\phi_0)$. The relevant terms of the Lagrangian (4.2.6) are then

$$\mathcal{L}(\phi_0 + \phi) \supset -\frac{1}{2}m^2\phi_0^2 - \frac{\lambda}{4!}\phi_0^4 + \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}\mu^2(\phi_0)\phi^2 + \mathcal{L}_{c.t.}(\phi_0), \quad (4.2.7)$$

where we have defined the ϕ_0 -dependent mass

$$M^2(\phi_0) \equiv m^2 + \frac{\lambda}{2}\phi_0^2. \quad (4.2.8)$$

Up to one-loop level we have

$$V_{eff}(\phi_0) = V_0(\phi_0) + V_1(\phi_0) + V_{c.t.}(\phi_0), \quad (4.2.9)$$

with

$$V_{c.t.}(\phi_0) = \frac{1}{2}\delta m^2\phi_0^2 + \frac{\lambda}{4!}(Z_\lambda - 1)\phi_0^4 \quad (4.2.10)$$

and

$$\begin{aligned} e^{-iV_4V_1(\phi_0)} &= \int \mathcal{D}\phi e^{i\int d^4x \left(\frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}M^2(\phi_0)\phi^2 \right)} \\ &= \mathcal{N} \det^{-\frac{1}{2}} \left(\square + M^2(\phi_0) \right) = \mathcal{N} e^{-\frac{1}{2}\text{Tr} \log \left(\square + M^2(\phi_0) \right)}. \end{aligned} \quad (4.2.11)$$

The determinant or the trace of the differential operator $\square + M^2(\phi_0)$ is easily computed in momentum space, where the operator is diagonal. We have

$$\log \left(\square + M^2(\phi_0) \right) (p, q) = \delta(p - q) \log \left(-p^2 + M^2(\phi_0) \right) \quad (4.2.12)$$

and hence

$$-iV_4V_1(\phi_0) = -\frac{1}{2}\delta^{(4)}(0) \int d^4p \log \left(-p^2 + M^2(\phi_0) \right). \quad (4.2.13)$$

The $\delta^{(4)}(0)$ accounts for the space-time volume, $\delta^{(4)}(0) \rightarrow V_4/(2\pi)^4$. Wick rotating $p_0 = ip_4$, we finally get

$$V_1(\phi_0) = \frac{1}{2} \int \frac{d^4p_E}{(2\pi)^4} \log \left(p_E^2 + M^2(\phi_0) \right), \quad (4.2.14)$$

where $p_E^2 = p_1^2 + p_2^2 + p_3^2 + p_4^2$. The integral in eq. (4.2.14) is highly divergent. In terms of the degree of divergence defined in section 3.1 it has $\delta = 4$. We can lower this divergence

by taking derivative with respect to $M^2(\phi_0)$. Since each derivative lowers δ by two, we have to derive three times to get a finite result:

$$\frac{d^3 V_1(\phi_0)}{(dM^2)^3} = \int \frac{d^4 p_E}{(2\pi)^4} \frac{1}{(p_E^2 + M^2(\phi_0))^3} = \int_0^\infty \frac{dp_E}{8\pi^2} \frac{p_E^3}{(p_E^2 + M^2(\phi_0))^3} = \frac{1}{32\pi^2} \frac{1}{M^2(\phi_0)}. \quad (4.2.15)$$

Integrating back three times gives, modulo an irrelevant constant factor, gives $(\phi_0 \rightarrow \phi)$

$$V_1(\phi) = \frac{1}{64\pi^2} M^4(\phi) \log M^2(\phi) + AM^2(\phi) + BM^4(\phi). \quad (4.2.16)$$

In analogy to the case of loop amplitudes considered in section 3.3, the effective potential is the sum of a non-analytic (in ϕ) and finite piece, the first term in eq.(4.2.16), plus local divergent terms. The divergences hidden in the constants A and B are cancelled by a proper choice of the counter-terms δm^2 and Z_λ in eq.(4.2.10). Putting all together, we finally get the total tree+one-loop level potential:

$$V_{eff}(\phi) = \frac{1}{2} m^2 \phi^2 + \frac{\lambda}{4!} \phi^4 + \frac{1}{64\pi^2} M^4(\phi) \log M^2(\phi). \quad (4.2.17)$$

When $m^2 = 0$, the potential (4.2.17) becomes

$$V_{eff}(\phi) = \frac{\lambda}{4!} \phi^4 + \frac{\lambda^2}{256\pi^2} \phi^4 \log \frac{\phi^2}{\phi_0^2}, \quad (4.2.18)$$

where ϕ_0 is an arbitrary fixed energy scale (not to be confused with the ϕ_0 appearing previously from eqs.(4.2.1) to (4.2.15)). Its value depends on the precise form of the counterterms chosen to remove the divergences, and is hence renormalization scheme dependent.

The effective potential, also known as Coleman-Weinberg (CW) potential [17], is a useful tool in QFT. It is in particular crucial to determine the correct vacuum of a theory when the tree level potential does not fix it. A case of this kind is precisely the model considered above, the ϕ^4 theory when the tree level mass vanishes, with effective potential (4.2.18). Classically the vacuum is at the origin in field space, but locally the potential is so flat that small quantum corrections can destabilize it. An arbitrarily small, loop-induced, negative squared mass term would destabilize the classical vacuum at the origin. We will come back to the problem of determining the actual vacuum in the massless ϕ^4 theory in section 5.11, after we have discussed the definition of coupling constants in various schemes and we have introduced the concept of renormalization group.

4.3 A Subtlety about Effective Potentials**

In this section we will discuss a subtlety regarding effective potentials. In eq.(4.1.10) we have defined the 1PI action $\Gamma[\Phi]$ as the Legendre transform of $W[J]$. It can be shown

that the effective potential, defined as in eq.(4.2.1) with $\Gamma[\Phi]$ above, must necessarily be a convex function [54, 37]. This is however puzzling, since the classical potential might clearly be non-convex to start with. If the system is parametrically weakly coupled, it is hard to imagine that quantum corrections could significantly modify the classical potential. The resolution of the puzzle is a bit tricky and is related to the fact that in general the 1PI action defined in eq.(4.1.10) does *not* correspond to that defined in eq.(4.1.24). In particular, the effective potential $V_{1\text{PI}}$ coming from eq.(4.1.24) is the natural extension of the classical potential V_0 and can in general be non-convex, while the effective potential V_{eff} defined as the constant field configurations of eq.(4.1.10) is the one that must necessarily be convex. Starting from eq.(4.1.24), we can define $W[J]$ as the Legendre transform of $\Gamma[\Phi]_{1\text{PI}}$, which is the Γ appearing on the left-hand-side of eq.(4.1.24). Taking now the Legendre transform of the so derived $W[J]$ we get $\Gamma[\Phi]_{eff}$. We then see that V_{eff} is the double Legendre transform of $V_{1\text{PI}}$, also called its convex envelope. *Only* for convex functions we are ensured that $V_{eff} = V_{1\text{PI}}$. Physically, $V_{eff}(\phi_0)$ represents the expectation value of the energy density in the state $|\Psi\rangle$ that minimizes $\langle\Psi|H|\Psi\rangle$ and such that $\langle\Psi|\phi(x)|\Psi\rangle = \phi_0$. This quantity is always real and, as we said, must be convex everywhere. The key point on how $V_{eff}(\phi_0)$ is convex starting from a non-convex classical potential is understood from a purely classical analysis, where $\langle\Psi|\phi(x)|\Psi\rangle$ represents the spatial average of the field $\phi(\vec{x})$. Consider a classical potential V_c like that depicted in the left panel of fig.4.1, with two minima at $\phi(\vec{x}) = \pm\sigma$. For $|\phi_0| \geq \sigma$, the field configuration minimizing the energy is given by $\phi(\vec{x}) = \phi_0$ over the whole space, and $V_{eff}(\phi_0) = V_c(\phi_0)$. For $|\phi_0| < \sigma$ the minimum energy is given by an inhomogeneous state where $\phi(\vec{x}) = \sigma$ in a fraction $(\sigma + \phi_0)/2\sigma$ of space, and $\phi(\vec{x}) = -\sigma$ in the remaining $(\sigma - \phi_0)/2\sigma$ fraction of space, so that the spatial average of the field is

$$\sigma \times \frac{\sigma + \phi_0}{2\sigma} + (-\sigma) \times \frac{\sigma - \phi_0}{2\sigma} = \phi_0, \quad (4.3.1)$$

as it should be. The energy density in such inhomogeneous state is all concentrated in the regions between the two phases. In the infinite volume limit the boundary energy contribution becomes negligible, and hence $V_{eff}(\phi_0) = 0$ for $|\phi_0| < \sigma$. The potential V_{eff} , depicted in the right panel of fig.4.1 is then everywhere convex.

In the quantum case, the inhomogeneous state is replaced by a linear combination of the two vacua at $\pm\sigma$, call them $|\pm\rangle$. For $|\phi_0| < \sigma$ the minimum energy is given by a state of the form $\alpha|+\rangle + \beta|-\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$, and $|\alpha|^2 - |\beta|^2 = \phi_0$. The ending result is the same, giving rise to a flat effective potential V_{eff} for $|\phi_0| < \sigma$. As we will see in more detail in chapter 8 in the context of spontaneous symmetry breaking, inhomogeneous states of the form $\alpha|+\rangle + \beta|-\rangle$ violate cluster decomposition and should not be considered

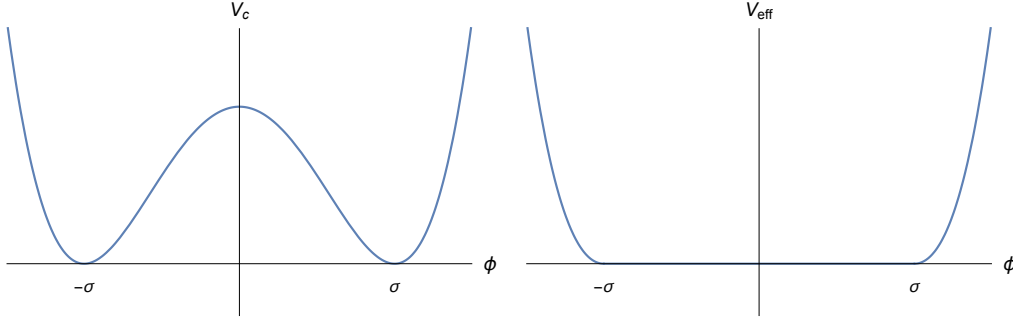


Figure 4.1: Classical (left) and effective (right) potentials V_c and V_{eff} .

viable vacua for a QFT. If the vacua are degenerate, like in fig.4.1, the system will be unstable under any external small perturbation and will fall in either the vacuum $|+\rangle$ or $|-\rangle$, depending on the perturbation. In this case the relevant effective potential is given by V_{1PI} , in agreement with our naive intuition.

4.4 Functional Relations

The formalism based on the generating functions introduced and discussed above is particularly useful for practical applications because it can make the derivation of important relationships rather straightforward. This is particularly true if one proceeds heuristically, as we will do, without taking into account the effects of renormalization. In other words, we will effectively be dealing in this section with bare quantities, without introducing counterterms. This is justified by the fact that eventually the same results apply for the renormalized expressions, but in a much simpler way! See e.g. ref.[19] for a proper derivation where renormalization is taken into account.

4.4.1 Schwinger-Dyson Equations

The Schwinger-Dyson equations are the quantum mechanical analogue of the classical equations of motion. They are derived from the observation (based on the extension of the concept of ordinary differentiation and integration to the case of functional differentiation and integration) that the functional integral of a functional derivative vanishes if one can neglect boundary terms, which is always assumed here. Accordingly, taking the exponential in eq. (4.1.1) as the function to be differentiated, one concludes that

$$0 = \int \mathcal{D}\phi \frac{\delta}{\delta\phi(x)} e^{iS(\phi) + i \int d^4x' J(x')\phi(x')} = \int \mathcal{D}\phi \left[i \frac{\delta S(\phi)}{\delta\phi(x)} + iJ(x) \right] e^{iS(\phi) + i \int d^4x' J(x')\phi(x')}. \quad (4.4.1)$$

Setting $J = 0$ in eq.(4.4.1) gives

$$\left\langle \frac{\delta S}{\delta \phi(x)} \right\rangle = 0, \quad (4.4.2)$$

which is the strict quantum-mechanical analogue of the classical equations of motion. Applying n functional derivatives with respect to the source J to eq.(4.4.1), and then setting $J = 0$, gives the Schwinger-Dyson equations

$$i \left\langle \frac{\delta S}{\delta \phi(x)} \phi(x_1) \dots \phi(x_n) \right\rangle = - \sum_{i=1}^n \delta^{(4)}(x - x_i) \langle \phi(x_1) \dots \hat{\phi}(x_i) \dots \phi(x_n) \rangle, \quad (4.4.3)$$

where the hat means that that operator should be omitted from the correlator. We see that at the quantum level the equations of motion turn into an infinite set of relations among different Green functions. We will make use of eqs.(4.4.3) later on in these lectures.

4.4.2 Symmetries and Ward-Takahashi Identities

Symmetries are important in QFT because they constrain the form of amplitudes, correlation functions, and establish relationships among them. Let us consider a field theory characterized by a Lagrangian density $\mathcal{L}$. An infinitesimal transformation

$$\phi(x) \mapsto \phi'(x) = \phi(x) + \epsilon \Delta \phi(x) \quad (4.4.4)$$

parameterized by a parameter ϵ is a symmetry of the theory if the action $S = \int d^4x \mathcal{L}(\phi, \partial_\mu \phi)$ does not change. For any continuous symmetry transformation, Noether's theorem ensures the existence of a classically conserved current j^μ : $\partial_\mu j^\mu = 0$. If the symmetry transformation is taken to be space-time dependent, namely if $\epsilon = \epsilon(x)$, then the action is in general no longer invariant and we have

$$\delta S = \int d^4x j^\mu(x) \partial_\mu \epsilon(x) \neq 0. \quad (4.4.5)$$

In order to work out the consequences of the symmetry (4.4.4) at the quantum level, consider the generating function $Z[J]$ in eq. (4.1.1) and let us perform a space-time dependent infinitesimal field transformation. We have

$$\begin{aligned} Z[J] &= \int \mathcal{D}\phi' e^{iS(\phi') + i \int d^4x J(x) \phi'(x)} = \int \mathcal{D}\phi' e^{iS(\phi) + i \int d^4x j^\mu \partial_\mu \epsilon(x) + i \int d^4x J(x) [\phi(x) + \epsilon \Delta \phi(x)]} \\ &= \int \mathcal{D}\phi \left(1 + i \int d^4x \epsilon(x) [-\partial_\mu j^\mu(x) + J(x) \Delta \phi(x)] + \mathcal{O}(\epsilon^2) \right) e^{iS(\phi) + i \int d^4x J(x) \phi(x)} \\ &= Z[J] + i \int d^4x \epsilon(x) \left(-\langle \partial_\mu j^\mu(x) \rangle_J + J(x) \langle \Delta \phi(x) \rangle_J \right) + \mathcal{O}(\epsilon^2). \end{aligned} \quad (4.4.6)$$

Note that in general we also have a Jacobian factor arising from the change of variable $\phi' \mapsto \phi$ in the functional integration measure. In writing eq.(4.4.6) we have assumed the

Jacobian is independent of the fields and can be absorbed in the overall normalization of the measure. The previous equation has to be valid for an arbitrary choice of $\epsilon(x)$ and therefore it implies that

$$\langle \partial_\mu j^\mu(x) \rangle_J = J(x) \langle \Delta \phi(x) \rangle_J. \quad (4.4.7)$$

This relation is an example of a Ward-Takahashi (WT) identity associated with the symmetry (4.4.4) and constitutes the equivalent of the Noether theorem in QFT. If we set $J = 0$ in eq.(4.4.7) we recover the classical conservation of the current as a quantum average:

$$\langle \partial_\mu j^\mu(x) \rangle = 0. \quad (4.4.8)$$

By taking functional derivatives of eq.(4.4.7) with respect to the source J , and then setting $J = 0$, gives an infinite set of relationships between correlation functions of the fields ϕ on the r.h.s. and those of the fields with an insertion of the current operator $j_\mu(x)$ on the l.h.s.

While at the classical level a symmetry requires the invariance of the action, at the quantum level we also require that the measure of integration is invariant or independent of the fields, namely no field-dependent Jacobian factor. This is in general a non-trivial constraint, which is however satisfied for the important class of transformations that are *linearly* realized on the fields, meaning that $\Delta \phi$ is linear in ϕ . For global abelian or non-abelian field rotations, for instance, we have

$$\phi_l(x) \mapsto \phi'_l(x) = \phi_l(x) + \epsilon^\alpha (t_\alpha)_l^m \phi_m(x), \quad (4.4.9)$$

where ϕ_l is a field multiplet in some representation of the group, with generators t_α . In this case the associated current is given by $j_{\alpha,\mu} = (\partial_\mu \phi)_l (t_\alpha)_l^m \phi_m$. By differentiating eq.(4.4.7) k times with respect to $J_{n_1}(x_1), J_{n_2}(x_2), \dots, J_{n_k}(x_k)$ and then by setting $J = 0$ we get

$$i \partial_\mu \langle j_\alpha^\mu(x) \phi_{n_1}(x_1) \dots \phi_{n_k}(x_k) \rangle = \sum_{p=1}^k \delta^{(4)}(x - x_p) \langle \phi_{n_1}(x_1) \dots \Delta \phi_{n_p}(x_p) \dots \phi_{n_k}(x_k) \rangle, \quad (4.4.10)$$

where $\Delta \phi_n = (t_\alpha)_n^m \phi_m$. The current $j_\mu(x)$ is a so-called *composite operator* because it involves the product of various quantum fields taken at the same point in space-time. After the renormalization of the theory, correlation functions with the insertion of composite fields still show ultraviolet divergences and need to be additionally renormalized by introducing suitable renormalization constants for these operators. As we shall discuss in more detail in sec. 5.9, the presence of these non-trivial additional renormalization constants implies that the composite operators acquire an *anomalous dimension*. However, conserved currents $j_\mu(x)$ associated to symmetries of the theory do not need to be renormalized and

therefore they do not acquire an anomalous dimension. This can be seen from the example in eq.(4.4.10). Correlation functions with one insertion of j_μ can be expressed in terms of correlation functions without the current. When the latter are properly renormalized by the introduction of suitable counter-terms and renormalization constants, the former will also be finite, because the number of elementary fields in the left and right hand side of eq.(4.4.10) is the same. In other words, a conserved current into a correlation function does not cause additional divergences to appear.

The effective action Γ is not always invariant under the same field transformations for which the classical action S is. The transformations of S and Γ coincide only when they act at most linearly on fields. Let us discuss in more detail this important point. Consider a transformation of the form

$$\phi(x) \mapsto \phi'(x) = \phi(x) + \epsilon F[\phi(x)], \quad (4.4.11)$$

where $F[\phi]$ is a generic functional of the field ϕ , not necessarily linear. We shall encounter non-linear transformations in section 6.3 when we will study the BRST symmetries in the context of non-abelian gauge theories. Assume now that both the action $S(\phi)$ and the integration measure are invariant under the transformation $\phi \rightarrow \phi'$, such that $S(\phi) = S(\phi')$ and $\mathcal{D}\phi' = \mathcal{D}\phi$. In terms of the generating function $Z[J]$ one finds

$$\begin{aligned} Z[J] &= \int \mathcal{D}\phi' e^{iS(\phi') + i \int d^4x J(x) \phi'(x)} = \int \mathcal{D}\phi e^{iS(\phi) + i \int d^4x J(x) \{\phi(x) + \epsilon F[\phi(x)]\}} \\ &= \int \mathcal{D}\phi \left\{ 1 + i\epsilon \int d^4x J(x) F[\phi(x)] + \mathcal{O}(\epsilon^2) \right\} e^{iS(\phi) + i \int d^4x J(x) \phi(x)}, \end{aligned} \quad (4.4.12)$$

which implies

$$\int d^4x \langle F[\phi(x)] \rangle_J J(x) = 0. \quad (4.4.13)$$

Using eq.(4.1.11), we can rewrite eq.(4.4.13) as

$$\int d^4x \langle F[\phi(x)] \rangle_{J(\Phi)} \frac{\delta \Gamma}{\delta \Phi(x)} = 0, \quad (4.4.14)$$

which implies that the 1PI action is invariant for $\Phi \rightarrow \Phi + \epsilon \langle F[\phi(x)] \rangle_{J(\Phi)}$:

$$\Gamma(\Phi + \epsilon \langle F[\phi(x)] \rangle_{J(\Phi)}) = \Gamma(\Phi) + \epsilon \int d^4x \langle F[\phi(x)] \rangle_{J(\Phi)} \frac{\delta \Gamma}{\delta \Phi(x)} = \Gamma(\Phi). \quad (4.4.15)$$

Notice that, in general,

$$\langle F[\phi(x)] \rangle_{J(\Phi)} \neq F(\Phi). \quad (4.4.16)$$

They coincide only for transformations that are at most linear in the fields, namely for⁵

$$F(\phi) = s(x) + \int d^4y t(x, y) \phi(y). \quad (4.4.17)$$

In this case, and only in this case, we have

$$\langle F[\phi(x)] \rangle_{J(\Phi)} = s(x) + \int d^4y t(x, y) \langle \phi(x) \rangle_{J(\Phi)} = s(x) + \int d^4y t(x, y) \Phi(y) = F(\Phi), \quad (4.4.18)$$

where the second identity is a consequence of the definition of Φ .

The general approach described above can be used in order to derive the consequences of the global $U(1)$ symmetry of QED on the correlation functions of the field. This will be the subject of next section.

4.4.3 WT Identities in QED

Gauge invariance in QED requires the addition of a gauge-fixing term

$$\mathcal{L}_{g.f.} = -\frac{1}{2\xi}(\partial_\mu A^\mu)^2, \quad (4.4.19)$$

where ξ is an arbitrary real parameter.⁶ The total Lagrangian density is then

$$\mathcal{L}(A_\mu, \psi, \bar{\psi}) = \mathcal{L}_{QED}(A_\mu, \psi, \bar{\psi}) + \mathcal{L}_{g.f.}(A_\mu), \quad (4.4.20)$$

where $\mathcal{L}_{QED}$ is given in eq.(3.4.16) and the functional integration is done both on the vector field A_μ and the spinors $\bar{\psi}$ and ψ . The source term at the exponential has a density of the form

$$J^\mu(x) A_\mu(x) + \bar{J}(x) \psi(x) + \bar{\psi}(x) J(x), \quad (4.4.21)$$

where J and $\bar{J}$ are Grassmann-valued (and hence anticommuting) functions. If we make the infinitesimal change of variable associated to a gauge transformation,

$$\psi(x) \rightarrow \psi(x) + ie\epsilon(x)\psi(x), \quad \bar{\psi}(x) \rightarrow \bar{\psi}(x) - ie\epsilon(x)\bar{\psi}(x), \quad A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu\epsilon(x), \quad (4.4.22)$$

the measure and $\mathcal{L}_{QED}$ remain invariant, while $\mathcal{L}_{g.f.}$ and the source term will change:

$$\begin{aligned} \mathcal{L}(A_\mu, \psi, \bar{\psi}) &\rightarrow \mathcal{L}(A_\mu, \psi, \bar{\psi}) - \frac{\partial^\mu A_\mu}{\xi} \square \epsilon. \\ J^\mu A_\mu + \bar{J}\psi + \bar{\psi}J &\rightarrow J^\mu A_\mu + \bar{J}\psi + \bar{\psi}J + J^\mu \partial_\mu \epsilon + ie\bar{J}\psi\epsilon - ie\bar{\psi}J\epsilon. \end{aligned} \quad (4.4.23)$$

⁵The usual purely linear transformations of the fields are obtained from eq.(4.4.17) by taking $s(x) = 0$ and $t(x, y) = t\delta(x - y)$.

⁶The derivation of $\mathcal{L}_{g.f.}$ from a functional point of view, for both abelian and non-abelian theories, is discussed in section 6.2.

By bringing down the $\mathcal{O}(\epsilon)$ terms from the exponential and taking a functional derivative with respect to $\epsilon(x)$, we get the identity

$$i\partial_\mu J^\mu(x)Z = e\langle\bar{\psi}(x)\rangle_J J(x) - e\bar{J}(x)\langle\psi(x)\rangle_J - \frac{i}{\xi}\square\partial_\mu\langle A^\mu(x)\rangle_J. \quad (4.4.24)$$

We also have⁷

$$\langle\bar{\psi}(x)\rangle_J = i\frac{\delta Z}{\delta J(x)}, \quad \langle\psi(x)\rangle_J = -i\frac{\delta Z}{\delta\bar{J}(x)}, \quad \langle A^\mu(x)\rangle_J = -i\frac{\delta Z}{\delta J_\mu(x)} \quad (4.4.25)$$

and thus we can rewrite eq.(4.4.24) in terms of $W = -i\log Z$ as

$$i\partial_\mu J^\mu(x) = -e\frac{\delta W}{\delta J(x)}J(x) - e\bar{J}(x)\frac{\delta W}{\delta\bar{J}(x)} - \frac{i}{\xi}\square\partial_\mu\frac{\delta W}{\delta J_\mu(x)}. \quad (4.4.26)$$

We can go further and write the WT Identity (4.4.26) in terms of the 1PI generator Γ defined as

$$\Gamma = W - \int d^4x \left(\mathcal{A}_\mu(x)J^\mu(x) + \bar{J}(x)\Psi(x) + \bar{\Psi}(x)J(x) \right), \quad (4.4.27)$$

where

$$\mathcal{A}_\mu(x) = \frac{\delta W}{\delta J^\mu(x)}, \quad \Psi(x) = \frac{\delta W}{\delta\bar{J}(x)}, \quad \bar{\Psi}(x) = -\frac{\delta W}{\delta J(x)}, \quad (4.4.28)$$

and correspondingly

$$J^\mu(x) = -\frac{\delta\Gamma}{\delta\mathcal{A}_\mu(x)}, \quad J(x) = -\frac{\delta\Gamma}{\delta\bar{\Psi}(x)}, \quad \bar{J}(x) = \frac{\delta\Gamma}{\delta\Psi(x)}. \quad (4.4.29)$$

In terms of Γ , eq.(4.4.26) reads

$$i\partial_\mu\frac{\delta\Gamma}{\delta\mathcal{A}_\mu(x)} = e\bar{\Psi}(x)\frac{\delta\Gamma}{\delta\bar{\Psi}(x)} + e\frac{\delta\Gamma}{\delta\Psi(x)}\Psi(x) + \frac{i}{\xi}\square\partial_\mu\mathcal{A}^\mu(x). \quad (4.4.30)$$

By taking arbitrary functional derivatives of eqs.(4.4.24), (4.4.26) and (4.4.30) with respect to the sources we get an infinite set of WT identities between connected and 1PI amplitudes, respectively.

Let us see more closely the WT identities that will allow us to prove three assertions in section 3.4.1 when studying the renormalizability of this theory:

1. A relation linking the electron two-point function with the vertex, eventually proving eq.(3.4.17).
2. The transversality of the photon propagator, eq.(3.4.7).
3. Decoupling of unphysical photon polarization states, eq.(3.4.11) (and its generalization to any other scattering amplitude involving at least one photon).

⁷Pay attention to the anticommuting nature of the Grassmann fields to get the signs right!

Relation 1 is obtained by taking one functional derivative with respect to Ψ and $\bar{\Psi}$ of eq.(4.4.30). In this way one has

$$i\partial_\mu \frac{\delta^3 \Gamma}{\delta \Psi(x_1) \delta \bar{\Psi}(x_2) \delta \mathcal{A}_\mu(x)} = e\delta(x-x_2) \frac{\delta^2 \Gamma}{\delta \Psi(x_1) \delta \bar{\Psi}(x)} - e\delta(x-x_1) \frac{\delta^2 \Gamma}{\delta \Psi(x) \delta \bar{\Psi}(x_2)}. \quad (4.4.31)$$

In momentum space eq.(4.4.31) becomes

$$q^\mu \Gamma_\mu(p, q) = e\Gamma^{(2)}(q+p) - e\Gamma^{(2)}(p) \quad (4.4.32)$$

where Γ_μ is precisely the function vertex defined in eq.(3.4.3) and $\Gamma^{(2)} = \not{p} - m + \Sigma(p)$ is the inverse of the electron propagator, with Σ defined in eq.(3.4.4). Using eq.(4.4.32) and the above definitions, it is straightforward to derive the identity (3.4.17). We see that a similar identity halso holds for the finite parts:

$$q^\mu \Gamma_\mu^{finite}(p, q) = \Sigma^{finite}(p+q) - \Sigma^{finite}(p). \quad (4.4.33)$$

Relation 2 is obtained by setting $\Psi = \bar{\Psi} = 0$ and taking one derivative with respect to $\mathcal{A}_\mu$ of eq.(4.4.30). This gives, in momentum space

$$q^\nu \Gamma_{\mu\nu}(q) = \frac{1}{\xi} q^2 q_\mu, \quad (4.4.34)$$

where $\Gamma_{\mu\nu}$ is the inverse photon propagator. We decompose $\Gamma_{\mu\nu}(q) = \Gamma_{\mu\nu}^{(0)}(q) + \Pi_{\mu\nu}(q)$ in terms of its tree-level and quantum corrections, where $\Pi_{\mu\nu}(q)$ is the same entering eq.(3.4.7). From the classical action we readily get

$$\Gamma_{\mu\nu}^{(0)}(q) = \eta_{\mu\nu} q^2 - q_\mu q_\nu + \frac{1}{\xi} q_\mu q_\nu. \quad (4.4.35)$$

We notice that $\Gamma_{\mu\nu}^{(0)}$ alone saturates the identity (4.4.34), automatically implying eq.(3.4.7) to all orders in perturbation theory. The fundamental importance of this conclusion relies on the fact that it ensures that the photon does not acquire a mass as a consequence of interactions, being protected by gauge invariance. An analogous conclusion is drawn in chapter 6 for non-abelian gauge theories, where a relation very similar to eq.(4.4.34), eq.(6.3.34), will be proved starting from the so called BRST symmetries.

Let us now consider relation 3. In order to relate the resulting WT identities directly with S -matrix amplitudes, that are associated to amputated connected Green functions, it is better to start from eq.(4.4.26). In absence of fermion fields, we can set $J = \bar{J} = 0$ and take an arbitrary number n of functional derivatives of eq.(4.4.26) with respect to J^μ . In this way the l.h.s. and the first two terms in the r.h.s. of eq.(4.4.26) vanish and we trivially get

$$q^2 q^\mu G_{\mu\alpha_1 \dots \alpha_n}(q, p_1, \dots, p_n) = 0. \quad (4.4.36)$$

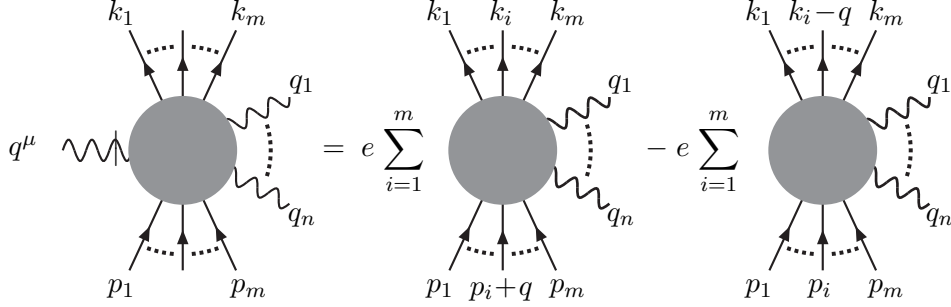


Figure 4.2: Diagrammatic representation of the WT identity for QED, which generalizes eq. (4.4.32) for connected amplitudes involving additional $2m$ fermionic and an arbitrary number n of photon external lines. The extra photon line in the diagram in the l.h.s. is amputated.

for the $n+1$ connected amplitudes involving photons only. The factor of q^2 in eq.(4.4.36) is removed when we amputate the amplitude, see below. In this way we get the generalization of eq.(3.4.11) for any number of external photon fields.

When external fermions are present, the analysis is slightly more complicated. If $2m$ external fermion lines are present, we have to take m functional derivatives of eq.(4.4.26) with respect to J and $\bar{J}$ (in addition to the ones necessary for the photon ones as above). In this way we derive WT identities for connected correlation functions involving an arbitrary number of gauge and fermion fields. The term in the l.h.s. still vanishes, but the first two terms in the r.h.s. are non-zero now. In momentum space we have

$$\frac{i}{\xi} q^2 q^\mu G_\mu^{(2m,n+1)}(p_i, k_i, q_j) = e \sum_{i=1}^m \left(G^{(2m,n)}(p_i + q, k_i, q_j) - G^{(2m,n)}(p_i, k_i - q, q_j) \right), \quad (4.4.37)$$

where $G^{(2m,n)}$ schematically represent the connected Green functions for $2m$ external fermions and n external photons and we have made explicit in $G^{(2m,n+1)}$ only the Lorentz index contracted with q_μ . As we mentioned, physical S -matrix amplitudes are related to amputated connected Green functions. Let us start by amputating the Green function $G^{(2m,n+1)}$ in the l.h.s. of eq.(4.4.37) of its additional photon line. One has

$$G_\mu^{(2m,n+1)} = G_{\mu\nu}^\gamma(q) G_{amp}^{\nu(2m,n+1)}, \quad (4.4.38)$$

where $G_{\mu\nu}^\gamma$ is the (all-orders) photon propagator $G_{\mu\nu}^\gamma \Gamma^{\nu\rho} = -i\delta_\mu^\rho$ and $G_{amp}^{\nu(2m,n+1)}$ refers to

the amputation of this photon line only. Given eq.(4.4.34), one immediately gets

$$q^2 q^\mu G_{\mu\nu}^\gamma(q) = -i\xi q_\nu. \quad (4.4.39)$$

Plugging eq.(4.4.39) in eq.(4.4.37) we get

$$q^\mu G_{\mu,amp}^{(2m,n+1)}(p_i, k_i, q_j) = e \sum_{i=1}^m \left(G^{(2m,n)}(p_i + q, k_i, q_j) - G^{(2m,n)}(p_i, k_i - q, q_j) \right). \quad (4.4.40)$$

This relation is diagrammatically represented in fig. 4.2, where n physical photons and $2m$ fermions are indicated by wiggly and straight lines, respectively, and q^μ denotes the wiggly line of the amputated photon with polarization $\epsilon_\mu(q) = q_\mu$. The grey circle stands for all possible diagrams which can be constructed with the specified external legs. By using the LSZ reduction formula, the correlation functions (4.4.40) are directly related to the physical S -matrix amplitudes. The latter is in particular proportional to the total residue given by the $2m + n$ poles in the correlation function (see eq. (2.3.10)), since we still have to amputate the additional $2m + n$ external states. However, we now notice that on the r.h.s. of eq.(4.4.40) one fermion momentum is always shifted from its on-shell value. If $p_i^2 = k_i^2 = m^2$, generically $(p_i + q)^2, (k_i - q)^2 \neq m^2$. Hence the correlation functions on the r.h.s. can only have $2m + n - 1$ poles and hence eq. (2.3.10) necessarily vanishes. We conclude that any physical amplitude involving an arbitrary number of photon and fermion fields must vanish when any of the external photon polarizations is taken proportional to the photon momentum:

$$\epsilon_{\mu_1}(p_1) \dots p_{\mu_j} \dots \epsilon_{\mu_{n+1}}(p_{n+1}) \mathcal{M}^{\mu_1 \dots \mu_{n+1}} = 0. \quad (4.4.41)$$

Chapter 5

The Renormalization Group

The renormalization group (RG) is a key concept in quantum field theory. It essentially tells us that instead of describing a physical system by some constant parameters in a Lagrangian, it is more convenient to let the parameters vary and keep only some of them, depending on the scale at which we are looking at the system. Historically, the RG was developed by Gell-Mann and Low in ref.[27] as a way to improve the perturbative expansion in QED. Several years after K.G. Wilson formulated the RG evolution in QFT in a much broader sense in terms of effective theories [61, 62]. If we want to know what happens at scales of order L , we do not need to know the details of what happens at distances $\ll L$ to have a good understanding of the process. In general, however, the microscopic short distance behaviour of a system is not completely negligible. When this is the case, we can “integrate out”, rather than simply neglect, all short-distance modes and retain only the effective degrees of freedom of interest.

Effective theories are at the basis of the reductionism in physics. We do not need the SM Lagrangian to study the energy levels of the hydrogen atom! The latter are well described by a much simpler Schrödinger equation, which captures the effective dynamics entering at the eV scale, namely the Coulomb potential between the electron and the proton. We will mostly consider in this chapter weakly coupled QFTs which remain so under RG evolution. In the context of particle physics, this essentially reduces to the original approach of ref.[27]. As mentioned, when heavy particles to be integrated out are present, the RG evolution leads to the idea of effective theories, which will be discussed in chapter 7. The case in which a weakly coupled QFT at high energy becomes strongly coupled at low energy can lead to a reshuffling of degrees of freedom and to the need to use a completely different Lagrangian. A relevant case of this sort is QCD, which we will discuss in chapter 8, see in particular sections 8.6 and 8.7.

The structure of this chapter is as follows. In section 5.1 we introduce Wilson’s idea of RG evolution induced by integrating out degrees of freedom, working out a specific model ($\lambda\phi^4$ theory). In section 5.2 the RG evolution in the same model is seen from a particle physics point of view, the point of view which we will mostly take in what follows. The minimal subtraction scheme and renormalization scheme dependences of coupling constants are discussed in sections 5.3 and 5.4, respectively. The typical evolution of coupling constants and more generally of RG flows in QFT are discussed in section 5.5. The RG flow equations satisfied by correlation functions (so called Callan-Symanzik equations) are described in section 5.6. As application, we compute in section 5.7 the one-loop beta-function and photon anomalous dimension in QED. The meaning of RG flow for dimensionful couplings is explained in sections 5.8 and 5.10, while in section 5.9 we discuss the renormalization of composite operators and how to compute their anomalous dimensions at one-loop level. In section 5.11 we discuss how RG techniques are useful in the analysis of effective potentials.

5.1 Relevant, Marginal and Irrelevant Couplings

In this section we consider the RG in the spirit of Wilson’s original idea, focusing on a particular model, the ϕ^4 theory in four space-time dimensions, Wick rotated in euclidean space.¹ Wilson’s approach has the main advantage of being conceptually very clear. It assumes the presence of a physical cut-off Λ in the theory, above which no mode can be excited. Imagine we are interested in processes occurring at scales of order $b\Lambda$, with $b < 1$. We label all modes $\phi(k)$ as “heavy” and “light”, depending on their momentum k . We write

$$\phi(k) = \phi_L(k)\theta(b\Lambda - |k|) + \phi_H(k)\theta(|k| - b\Lambda), \quad |k| < \Lambda \quad (5.1.1)$$

and we correspondingly decompose the action of the ϕ^4 theory as follows:²

$$S = S_L + S_H^0 + S_{int}, \quad (5.1.2)$$

¹The analysis in this section closely follows section 12.1 of ref. [49].

²Strictly speaking, the actions (5.1.3) should be written in momentum space using eq. (5.1.1) and recalling the overall bound $|k| < \Lambda$. This results in unnecessarily long expressions that are avoided in the rough, but more brief form (5.1.3).

where

$$\begin{aligned}
S_L &= \int d^4x \left(\frac{1}{2} (\partial \phi_L)^2 + \frac{1}{2} m^2 \phi_L^2 + \frac{\lambda}{4!} \phi_L^4 \right), \\
S_H^0 &= \int d^4x \left(\frac{1}{2} (\partial \phi_H)^2 + \frac{1}{2} m^2 \phi_H^2 \right), \\
S_{int} &= \frac{\lambda}{4!} \int d^4x \left(\phi_H^4 + 4 \phi_H \phi_L^3 + 4 \phi_H^3 \phi_L + 6 \phi_H^2 \phi_L^2 \right).
\end{aligned} \tag{5.1.3}$$

It is clear from eq. (5.1.1) that the quadratic terms of the form $\phi_H \phi_L$ vanish. The modes ϕ_H cannot be excited for processes at scales below or of order $b\Lambda$, so at first approximation they can be ignored, in which case we just recover the ϕ^4 theory S_L for the light mode only. Quantum mechanically, however, the modes ϕ_H contribute as virtual particles. Instead of neglecting them, we should more properly integrate them out, getting in this way an effective action for the light modes ϕ_L :

$$e^{-S_{eff}(\phi_L)} = e^{-S_L(\phi_L)} \int \mathcal{D}\phi_H(k) e^{-S_H^0(\phi_H) - S_{int}(\phi_L, \phi_H)}. \tag{5.1.4}$$

The effective action $S_{eff}(\phi_L)$ takes into account of the effects of the heavy fields at the full quantum level. In this sense, it resembles the 1PI action $\Gamma(\phi)$ defined in section 4.1, but only for the heavy fields. This makes a crucial difference: while the functional $\Gamma(\phi)$ is in general non-local and is expanded as in eq.(4.1.19), the action $S_{eff}(\phi_L)$, admits an expansion in terms of a local action, though in general containing an infinite number of terms. In eq. (5.1.4) the modes ϕ_L act like external fields. It is not difficult to see that the tree-level exchange of the ϕ_H modes generates effective ϕ_L couplings of the form $(\lambda^2/m^2)\phi_L^6$ (from two $\phi_H \phi_L^3$ vertices) plus an infinite set of higher derivative couplings involving six ϕ_L fields. Similar considerations can be made for all other couplings. At some order in perturbation theory all possible couplings compatible with the symmetries will be generated. The effective action $S_{eff}(\phi_L)$ reads as

$$S_{eff}(\phi_L) = \int d^4x \left(\frac{1}{2} (1 + \delta Z) (\partial \phi_L)^2 + \frac{1}{2} (m^2 + \delta m^2) \phi_L^2 + \frac{\lambda + \delta \lambda}{4!} \phi_L^4 + \delta Z_1 \phi_L^6 + \delta Z_2 (\partial \phi_L)^4 + \dots \right) \tag{5.1.5}$$

where $\dots$ stands for higher dimensional operators. Recall that ϕ_L contains only momenta $k \leq b\Lambda$. The rescaled momentum $k' = k/b$ satisfies the same constraint as the original system, $k' < \Lambda$. Correspondingly we redefine coordinates as $x' = xb$. We also have to redefine the field ϕ so that it has a canonically normalized kinetic term:

$$\phi_L^{can} = b^{-1} \sqrt{1 + \delta Z} \phi_L. \tag{5.1.6}$$

The action S_{eff} reads now (redefining $x' \rightarrow x$ and $\phi_L^{can} \rightarrow \phi_L$)

$$S_{eff}(\phi_L) = \int d^4x \left(\frac{1}{2}(\partial\phi_L)^2 + \frac{1}{2}m^2(b)\phi_L^2 + \frac{\lambda(b)}{4!}\phi_L^4 + \delta Z_1(b)\phi_L^6 + \delta Z_2(b)(\partial\phi_L)^4 + \dots \right), \quad (5.1.7)$$

where³

$$\begin{aligned} m^2(b) &= \frac{1}{b^2} \frac{m^2 + \delta m^2}{1 + \delta Z}, & \lambda(b) &= \frac{1}{b^0} \frac{\lambda + \delta\lambda}{(1 + \delta Z)^2}, \\ \delta Z_1(b) &= b^2 \frac{\delta Z_1}{(1 + \delta Z)^3}, & \delta Z_2(b) &= b^4 \frac{\delta Z_2}{(1 + \delta Z)^2}. \end{aligned} \quad (5.1.8)$$

The process of integrating out heavy degrees of freedom and rescale the momentum is called “renormalization group”.

The action (5.1.7) is the proper action for describing processes at scales $E \leq b\Lambda$. It is straightforward to see that a coupling $c_{\mathcal{O}}$ of a generic operator $\mathcal{O}$ of dimension Δ (in mass) scales as

$$c_{\mathcal{O}}(b) = b^{\Delta-4} c_{\mathcal{O}}. \quad (5.1.9)$$

At lower and lower energies (smaller and smaller b) among all infinite operators appearing in eq. (5.1.7), only the finite subset of those with $\Delta \leq 4$ do actually matter, all the others being “irrelevant”. This is a very remarkable result, which dramatically simplifies the physical description of a system. We define as

- *Irrelevant* the operators with $\Delta > 4$
- *Relevant* the operators with $\Delta < 4$
- *Marginal* the operators with $\Delta = 4$

Correspondingly, we define as

- *Irrelevant* the couplings with mass dimensions < 0
- *Relevant* the couplings with mass dimensions > 0
- *Marginal* the couplings with mass dimensions $= 0$

Relevant couplings grow in the IR and dominate the physics. In our ϕ^4 example, the only relevant coupling is the mass term. When all relevant couplings vanish, the IR theory is controlled by marginal couplings only, in which case we say that the theory is (classically,

³The couplings δZ , δm^2 , etc. appearing in eq. (5.1.5), after the redefinition $k \rightarrow bk$, can depend on b , as we will explicitly see in what follows. However, in order to avoid confusion with the couplings defined in eqs. (5.1.8), we have omitted to write this b -dependence.

see below) scale invariant, namely the form of the action does not change under the RG flow.

The fate of marginal operators under the RG flow cannot be deduced from classical scaling and requires a quantum computation. In the ϕ^4 example, the only marginal operator is λ and its b -dependence is determined by that of δZ and $\delta\lambda$. It is very useful to determine $\lambda(b)$ at the lowest non-trivial order in perturbation theory, which is one-loop. At one-loop level the exchange of ϕ_H fields modifies the ϕ_L^4 coupling by means of two $\phi_L^2\phi_H^2$ interactions. No correction arises in δZ at one-loop order, so that we can neglect it. By denoting the heavy fields with a wavy line, the relevant Feynman graph to consider is

$$\begin{array}{c} k \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ \diagdown \quad \diagup \\ p+k \end{array} + 2 \text{ perms.} \equiv -\delta\lambda = \frac{3\lambda^2}{2} \int_{b\Lambda \leq k \leq \Lambda} \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 + m^2} \frac{1}{(p+k)^2 + m^2}, \quad (5.1.10)$$

where p is the incoming external momentum and “perms.” refer to two other diagrams obtained by permuting the external lines. When the external momentum and the mass m are much smaller than $b\Lambda$, $\delta\lambda$ is easily computed:

$$\delta\lambda = -\frac{3\lambda^2}{32\pi^4} \int_{b\Lambda \leq k \leq \Lambda} \frac{d^4k}{k^4} = -\frac{3\lambda^2}{16\pi^2} \log \frac{1}{b}. \quad (5.1.11)$$

The effective quartic interaction between the light modes is then

$$\lambda(b) = \lambda + \delta\lambda = \lambda + \frac{3\lambda^2}{16\pi^2} \log b. \quad (5.1.12)$$

As we can see from eq. (5.1.12), $\lambda(b)$ decreases at larger distances (smaller b), so we say that the coupling λ is *marginally irrelevant*, meaning in this way that it decreases in the IR, although not as quickly as a classically irrelevant operator. Similarly, we call *marginally relevant* a classically marginal coupling that increases at larger distance due to quantum effects. Classically marginal couplings that remain marginal at the quantum level are called exactly marginal.

At the quantum level, as we have just seen, all the terms δZ , δm^2 etc., appearing in eqs. (5.1.7) and (5.1.8), can depend on b . This implies that the scaling dimension of an operator in general receives corrections with respect to its classical value determined from the powers of b appearing in eq. (5.1.7). The difference between the quantum contribution to the scaling dimension of an operator from its classical value is called the *anomalous dimension* of the operator. We will come back to it later on in this chapter.

We started our analysis from the UV action (5.1.2) but it is clear that if we had started at the UV with the most general non-renormalizable action of the form (5.1.5), under the

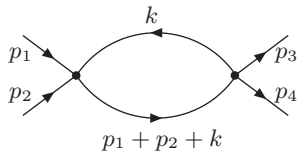
RG flow we would have again ended up in the ϕ^4 theory, plus an infinite number of irrelevant operators. From the Wilsonian RG point of view renormalizable theories can be seen as critical surfaces, in parameter space, where a much larger class of theories flow to. It is important to stress here that our classification of the operators in relevant, irrelevant and marginal is based on the classical dimension of fields, valid at parametrically weak coupling in perturbation theory around free field theories. At strong coupling, it might happen that, say, an operator that is classically irrelevant becomes relevant due to non-perturbative effects.

The Wilsonian picture of the RG flow is very intuitive and physical but, pragmatically speaking, it is not the best way to proceed in high energy physics. Distinguishing light and heavy modes in a single field can give rise to complicated expressions and in Minkowski space is not a Lorentz invariant notion.⁴ Cut-off regularization is often unavailable, like in gauge theories. Moreover, the Wilsonian RG flow requires, as a starting point, some UV-regulated Lagrangian, while in high energy physics we prefer to hide our ignorance about the UV physics in the renormalization of the parameters entering into the Lagrangian. For all these reasons, in the remaining of this chapter we will change perspective and consider the RG flow from an “high energy physics” point of view. As we will see, in a perturbative context, it essentially gives us a way of improving the perturbative expansion, in the spirit of ref.[27].

5.2 The Sliding Scale and the Summation of Leading Logs

The RG flow technique is very useful in standard four dimensional (weakly interacting) theories in high energy physics. In particular, it improves the perturbative expansion, allowing us to sum whole series of higher order effects. An example will clarify the problem and its resolution.

Consider the 1PI 4-point function at one-loop level in our usual ϕ^4 theory in $d = 4$ space-time dimensions. Using a cut-off Λ in the momenta, we get

$$\begin{aligned}
 \text{Diagram} + \text{ perms.} &= \frac{i\lambda^2}{32\pi^2} \int_0^1 dx \left(\log \frac{\Lambda^2}{m^2 - sx(1-x)} - 1 \right) \\
 &+ (s \rightarrow t) + (s \rightarrow u), \tag{5.2.1}
 \end{aligned}$$


where s , t and u are the Mandelstam variables defined in eq.(2.6.15). As we already

⁴On the contrary, when the mass of the field is sufficiently large, *all* frequency modes, i.e. the whole field, can be integrated out. In this case the Wilsonian picture is the way to go and results in a simplification of the physical system. More on this in chapter 7.

discussed at length, we need a renormalization condition that fixes the counter-terms needed to remove the logarithmic divergence in eq. (5.2.1). Let us define the coupling constant λ as the value of $\Gamma^{(4)}(p_i)$ at the (unphysical) symmetric point $s = t = u = 4m^2/3$:

$$i\Gamma^{(4)}(s = t = u = 4m^2/3) \equiv -i\lambda. \quad (5.2.2)$$

In this way, the finite, renormalized 1PI four-point function reads

$$\Gamma^{(4)}(s, t, u) = -\lambda + \frac{\lambda^2}{32\pi^2} \int_0^1 dx \log \frac{m^2(1 - 4x(1 - x)/3)}{m^2 - sx(1 - x)} + (s \rightarrow t) + (s \rightarrow u). \quad (5.2.3)$$

At high energies, when the absolute values of the Mandelstam variables are much greater than m , large logs $\sim \log m^2/E^2$ appear.⁵ If E is sufficiently high, it can happen that the $\log m^2/E^2$ term is so large to compensate for the one-loop suppression given by $\sim \lambda/(16\pi^2)$, breaking down the perturbative expansion. Similarly, at two-loop level, terms of the form $\lambda^3 \log^2 m^2/E^2$ and $\lambda^3 \log m^2/E^2$ appear. In general, at l -loop level, all terms of the form $\lambda^{l+1} \log^l m^2/E^2$, $\lambda^{l+1} \log^{l-1} m^2/E^2$, $\dots$, $\lambda^{l+1} \log m^2/E^2$ can appear. It is clear that if $\lambda \ll 1$, but $\lambda \log E^2/m^2 \sim 1$, the terms of the form $\lambda(\lambda^l \log^l m^2/E^2)$ are all of the same order. Logs of this form are called leading logs (LL) for obvious reasons. The terms of the form $\lambda^2(\lambda^l \log^l m^2/E^2)$ are denoted next-to-leading logs (NLL) and so on. There is actually a simple way to avoid the explicit appearance of large logs in $\Gamma^{(4)}$. The idea is based on the fact that we can choose the renormalization condition for the coupling as we wish. If we define the coupling constant at an energy scale $\mu \sim E$, replacing eq. (5.2.2) by

$$i\Gamma^{(4)}(s = t = u = -\mu^2) \equiv -i\lambda(\mu), \quad (5.2.4)$$

the finite, 1PI four-point function (5.2.3) would now read

$$\Gamma^{(4)}(s, t, u) = -\lambda(\mu) + \frac{\lambda^2(\mu)}{32\pi^2} \int_0^1 dx \log \frac{m^2 + \mu^2 x(1 - x)}{m^2 - sx(1 - x)} + (s \rightarrow t) + (s \rightarrow u), \quad (5.2.5)$$

and no large log term appears anymore. The arbitrary scale μ is denoted the sliding or renormalization scale. The coupling $\lambda(\mu)$ is determined by noting that the physics cannot depend on our arbitrary choice of scale μ . We must require that

$$\mu \frac{d\Gamma^{(4)}}{d\mu} = \left[\mu \frac{\partial}{\partial \mu} + \beta\left(\lambda, \frac{m}{\mu}\right) \frac{\partial}{\partial \lambda} \right] \Gamma^{(4)} = 0, \quad (5.2.6)$$

⁵Since $s + t + u = 4m^2$, the Mandelstam variables will necessarily have, at high energies, different signs, so that the argument of some of the logs in eq. (5.2.3) will be negative with branch cuts singularities. This is of course expected, since by the optical theorem $\Gamma^{(4)}$ should have an imaginary part. In order to avoid branch cut-singularities and imaginary amplitudes, which will not change the discussion that follows, we consider off-shell amplitudes at euclidean values of the momenta, where s, t and u are all negative.

where we have defined the function β as

$$\beta\left(\lambda, \frac{m}{\mu}\right) \equiv \mu \frac{d\lambda}{d\mu}. \quad (5.2.7)$$

The behaviour of λ as a function of μ , as given by the first-order differential equation (5.2.7), is called the renormalization group (RG) flow of λ . It is straightforward to compute β given the explicit form (5.2.5) of $\Gamma^{(4)}$. We get

$$\beta\left(\lambda, \frac{m}{\mu}\right) = \frac{3\lambda^2}{16\pi^2} \int_0^1 dx \frac{\mu^2 x(1-x)}{m^2 + \mu^2 x(1-x)} + \mathcal{O}(\lambda^3). \quad (5.2.8)$$

The coupling $\lambda(\mu)$ is determined by the first-order differential equation (5.2.7). In the high (UV) and low (IR) energy regimes $\mu \gg m$, $\mu \ll m$, eq. (5.2.8) simplifies, giving

$$\begin{aligned} \beta_{UV} &\simeq \frac{3\lambda^2}{16\pi^2}, \\ \beta_{IR} &\simeq \frac{\lambda^2}{32\pi^2} \left(\frac{\mu}{m}\right)^2 \simeq 0, \end{aligned} \quad (5.2.9)$$

whose solutions are

$$\lambda_{UV}(\mu) \simeq \frac{\lambda(\mu_0)}{1 - \frac{3\lambda(\mu_0)}{16\pi^2} \log \frac{\mu}{\mu_0}}, \quad \mu_0, \mu \gg m, \quad (5.2.10)$$

$$\lambda_{IR}(\mu) \simeq \text{constant}, \quad \mu \ll m. \quad (5.2.11)$$

Strictly speaking, eqs. (5.2.10) and (5.2.11) are valid only in the asymptotic UV and IR regimes, but as a first crude approximation we can take $\mu_0 = m$ in eq. (5.2.10) and $\lambda(\mu) \simeq \lambda(m)$ for any $\mu \leq m$ in eq. (5.2.11). We will perform a refined approximation in section 5.5, but for the moment this suffices to understand the main point of the RG analysis. Let us compare the two equivalent 1PI 4-point functions (5.2.3) and (5.2.5) in the euclidean UV point $s = t = u \equiv -E^2 \gg m^2$:

$$\Gamma^{(4)}(E) \simeq -\lambda - \frac{3\lambda^2}{16\pi^2} \log \frac{E}{m} + \mathcal{O}(\lambda^2), \quad (\text{no RG}) \quad (5.2.12)$$

$$\Gamma^{(4)}(E) \simeq -\lambda(E) = -\frac{\lambda}{1 - \frac{3\lambda}{16\pi^2} \log \frac{E}{m}} + \mathcal{O}(\lambda^2), \quad (\text{RG}) \quad (5.2.13)$$

where $\lambda \simeq \lambda(m)$. Expanding in λ , we see that eq. (5.2.13) reproduces the one-loop result in eq. (5.2.12) but, in addition, automatically gives us all the LL logs of the form $\lambda \lambda^l \log^l E/m$ (this will be proved in section 5.6.1). This is the key point of the RG evolution in perturbation theory: a powerful way to improve the perturbative expansion.

Various approximations have been made in comparing eqs. (5.2.3) and (5.2.5), such as $\lambda(m) \simeq \lambda(2m/\sqrt{3})$ and $\mu_0 = m$. All these approximations change $\Gamma(E)$ at $\mathcal{O}(\lambda^2 \log^0 E/M)$

and are hidden in the $\mathcal{O}(\lambda^2)$ terms in eqs. (5.2.12) and (5.2.13). When $E \gg m$, these terms can consistently be neglected, being sub-leading with respect to the $\lambda^2 \log E/M$ terms. They are important if we want to go beyond the LL approximation, in which case they are the first terms in the $\lambda^2(\lambda^{l-1} \log^{l-1} E/M)$ (NLL) series. Resumming these logs require the knowledge of the $\lambda^3 \log E/M$ term, i.e. a two-loop perturbative computation.

The sliding scale μ and the RG evolution of the coupling are also useful in the IR. For instance, when $m = 0$, we immediately see that there is an IR singularity in the 1PI 4-point function (5.2.3), while this is absent in eq. (5.2.5), which is well defined for $m = 0$.

Going back to eqs. (5.2.10) and (5.2.11), the RG evolution implies that the effective coupling constant in a QFT (i.e. the one which does not give rise to large logs in amplitudes) depends on the energy scale, namely it is a “running” coupling constant. In particular eq. (5.2.10) implies that $\lambda(\mu_1) > \lambda(\mu_0)$ for $\mu_1 > \mu_0$ (in agreement with the results derived using the Wilsonian RG flow) and predicts a pole at the scale

$$\Lambda_L = \mu_0 e^{\frac{16\pi^2}{3\lambda(\mu_0)}} \quad (5.2.14)$$

where the coupling diverges (Landau pole).⁶ Equation (5.2.14) indicates that, no matter how small $\lambda(\mu_0)$ is in the IR, at energies of order Λ_L perturbation theory breaks down. On the other hand, perturbativity improves at low energies. When $m = 0$, and eq. (5.2.10) is valid for any range of μ_0 and μ , we see that $\lambda(\mu) \rightarrow 0$ as $\mu \rightarrow 0$. In other words, provided the theory is in a perturbative regime at some scale μ_0 , in the far IR it asymptotes a free field theory. Theories of this sort are denoted IR free. When $m \neq 0$, the running instead essentially stops for $\mu \lesssim m$. This is in agreement with the fact that at low energies no large logs to be resummed appear, see eq. (5.2.3). This is a manifestation of a general principle according to which the effects of massive particles at low energies should be negligible and go to zero when $m \rightarrow \infty$ [4].⁷ This is a key principle in QFT that allows us to reliably describe physical processes at some scale without necessarily knowing the “true” (if any) physical theory underlying all processes. In this sense any QFT should be seen as an “effective” theory. We will come back to this point in chapter 7, where we will study effective field theories.

5.3 Minimal Subtraction

The sliding scale μ does not necessarily appear as the momenta where we renormalize the Green functions. In particular, it arises in dimensional regularization for the need

⁶The scale dependence of Λ_L on μ_0 is only apparent. One can check that $d\Lambda_L/d\mu_0 = 0$.

⁷There can be exceptions when the particle mass arises from a spontaneous breaking of a symmetry, in which case it is governed by a coupling.

to fix the scaling dimension of coupling constants when we depart from $d = 4$ space-time dimensions. Let $g_B(d)$ be a dimensionless bare coupling constant in 4 space-time dimensions. In d dimensions the coupling will acquire a mass dimension

$$\Delta(d) = (4 - d)\rho, \quad (5.3.1)$$

where ρ is a coupling-dependent coefficient. For the quartic coupling in the ϕ^4 theory $\rho = 1$. For gauge couplings in abelian and non-abelian gauge theories, and Yukawa couplings $\rho = 1/2$. In the last cases the actual expansion parameter in perturbation theory is given by their square, so if we identify g above with the effective loop-wise parameter, we can set $\rho = 1$ in all the cases.

Since divergences appear as poles in $(d - 4)$, a dimensionless renormalized coupling constant can be defined as

$$g_B(d)\mu^{-\Delta(d)} \equiv g(\mu, d) + \sum_{n=1}^{\infty} (4 - d)^{-n} b_n(g(\mu, d)). \quad (5.3.2)$$

Let us now take a derivative with respect to $\mu d/d\mu$ of eq. (5.3.2). We get

$$-(4 - d) \left(g + \sum_{n=1}^{\infty} (4 - d)^{-n} b_n(g) \right) = \beta_d(g) + \sum_{n=1}^{\infty} (4 - d)^{-n} \beta_d \frac{\partial b_n}{\partial g}, \quad (5.3.3)$$

where for simplicity we have denoted $g(\mu, d)$ by g . Eq. (5.3.3) should apply to any d , and hence we get an independent relation for any power of $d - 4$. At $\mathcal{O}(4 - d)$ we have

$$\beta_d(g) = -(4 - d)g + \beta(g), \quad (5.3.4)$$

where $\beta(g)$ does not vanish for $d \rightarrow 4$ and is the actual four-dimensional β -function. At $\mathcal{O}(d - 4)^0$ we have

$$\beta(g) = -b_1 + g \frac{\partial b_1}{\partial g}. \quad (5.3.5)$$

We see that the $\beta(g)$ is determined by looking at the simple poles in eq. (5.3.2).

Notice that we have the freedom to replace μ in the left-hand side of eq.(5.3.2) with $c\mu$, where c is an arbitrary constant factor. In so doing nothing changes in the derivation and we would still get eq.(5.3.5). So we can have a one-parameter family of renormalization schemes depending on the choice of c . The choice $c = 1$ corresponds to the Minimal Subtraction (MS) scheme.

In dimensional regularization $d - 4$ poles arise typically in the combination

$$(c\mu)^{\Delta(d)} \Gamma\left(\frac{4 - d}{2}\right) (4\pi)^{\frac{4 - d}{2}} \xrightarrow{d \rightarrow 4} \frac{2}{4 - d} - \gamma_E + \log 4\pi + 2 \log c + \log \mu^2 + \mathcal{O}(4 - d), \quad (5.3.6)$$

where $\gamma_E \simeq 0.577$ is the Euler-Mascheroni constant. It is then quite convenient to choose the factor c in such a way that the finite pieces given by $-\gamma_E + \log 4\pi$ cancel. This is obtained by taking

$$c_{\overline{\text{MS}}} = \frac{1}{\sqrt{4\pi}} e^{\gamma_E/2}. \quad (5.3.7)$$

The associated scheme is called modified Minimal Subtraction and is denoted by $\overline{\text{MS}}$. Note that the counterterms in $\overline{\text{MS}}$ are identical to those in MS, only poles in $1/(4-d)$. As is customary in the literature, we will not write explicitly factors like (5.3.7). So when we will refer to, say, $\overline{\text{MS}}$ and write an expression involving the factor μ^{4-d} , by μ we really mean

$$\mu_{\overline{\text{MS}}} = c_{\overline{\text{MS}}} \mu. \quad (5.3.8)$$

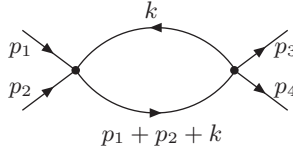
In contrast to the definition of β given in eq. (5.2.8), the MS or $\overline{\text{MS}}$ β -functions do not depend on masses but only on coupling constants. As we will see, a similar result applies to field anomalous dimensions. For this reason, such schemes are called mass-independent. We will explain in next section why and how these two results are not in contradiction with each other.

5.4 Renormalization Scheme Dependence

The detailed form of the β -function of couplings (and anomalous dimensions γ of fields) has not per se an intrinsic physical meaning, since it depends on the renormalization scheme chosen. Only physical quantities are scheme-independent, and hence if distinct schemes give different expressions for physical amplitudes, necessarily the evolution of the couplings should compensate for the difference. For instance, as we have just seen in section 5.3, in MS or $\overline{\text{MS}}$ schemes, β is mass-independent, as opposed to momentum subtraction used in section 5.2 (hereafter denoted by MOM) and more in general to all schemes where β depends on the masses of the particles. Is it then meaningful to talk about running coupling? If so, how do we determine the “correct” running? In order to answer these questions, it is useful to focus on a concrete example and work it out in some detail. As usual, we take the ϕ^4 theory and compare the 1PI 4-point function (5.2.5), evaluated using MOM, with the expression one obtains in $\overline{\text{MS}}$. By denoting $d = 4 - \epsilon$, and recalling eqs. (4.2.3) and (5.3.2), one has

$$\lambda_B(d)\mu^{-\epsilon} = \lambda + \left(\frac{Z_\lambda}{Z^2} - 1\right)\lambda \equiv \lambda + \delta_\lambda. \quad (5.4.1)$$

We compute



$$\begin{aligned}
&= \frac{(-i\lambda\mu^\epsilon)^2}{2} \int \frac{d^d k}{(2\pi)^d} \int_0^1 dx \frac{-i}{[k_E^2 + m^2 - sx(1-x)]^2} \\
&= \frac{i\lambda^2\mu^{2\epsilon}}{2(4\pi)^{d/2}} \Gamma(\epsilon/2) \int_0^1 \frac{dx}{[m^2 - sx(1-x)]^{\epsilon/2}}. \tag{5.4.2}
\end{aligned}$$

The counter-term is easily computed and one finds

$$\delta_\lambda = \frac{3\lambda^2}{16\pi^2\epsilon}. \tag{5.4.3}$$

This is related to the coefficient b_1 in eq.(5.3.2), $\delta_\lambda = b_1/\epsilon$. Summing over the t and u channel contributions, we get the finite, renormalized 4-point function

$$\Gamma_{\overline{\text{MS}}}^{(4)}(s, t, u) = -\lambda(\mu) + \frac{\lambda^2(\mu)}{32\pi^2} \int_0^1 dx \log \frac{\mu^2}{m^2 - sx(1-x)} + (s \rightarrow t) + (s \rightarrow u). \tag{5.4.4}$$

Notice how the numerators of the log term in eqs. (5.2.5) and (5.4.4) differ, the latter being simply μ^2 . Demanding the μ -independence of the amplitude, as in section 4.1, we get

$$\beta_{\overline{\text{MS}}} = \frac{3\lambda^2}{16\pi^2}, \tag{5.4.5}$$

for *any* value of μ , independently of m . Of course, the same result would have been obtained directly using eq. (5.3.5) and b_1 given by eq. (5.4.3). Thus, $\beta_{\overline{\text{MS}}}$ in eq. (5.4.5) differs from β_{MOM} in eq. (5.2.8), although their UV behaviour is identical and given by eq. (5.2.9). The latter also coincides with $bd\lambda(b)/db$ in eq. (5.1.12), as expected. This is a general result: the detailed form of β is scheme-dependent, but its UV behaviour, at one-loop (and sometimes two-loops) level, is scheme-independent, as we will show in section 5.5. As far as the LL summation is concerned, both renormalization schemes are valid. At low energies, however, the two schemes lead to different behaviours: in MOM λ essentially stops running below the scale m , in $\overline{\text{MS}}$ the running never stops and would lead to a free theory when $\mu \rightarrow 0$! In the IR the physical picture is best given by MOM. This can be seen by noticing that eq. (5.2.5) is regular when $\mu \rightarrow 0$, while eq. (5.4.4) seems IR divergent. This “fake” IR divergence and the corresponding large logs associated to it, are easily evaded by noticing that at arbitrarily low energies $s, t, u \ll m^2$, the large logs are avoided by taking $\mu^2 \simeq m^2$. In other words, in the $\overline{\text{MS}}$, the necessity of avoiding spurious large logs forces us to never use the IR evolution of the coupling, being $\mu \simeq m$ the correct sliding scale in the IR. We thus conclude that the “correct” IR running behaviour is the one given by the MOM scheme. However, provided that one keeps in mind that the IR running in mass-independent schemes is fake, the latter can reliably be used. The best

way to automatically get rid of this non-decoupling of heavy particles in mass-independent schemes is provided by using them in an effective field theory approach. One integrates out the heavy particle so that, for $\mu \ll m$, it is not in the spectrum and no longer contributes to the running.

Amplitudes are easier to compute in mass-independent rather than in mass-dependent schemes. However, physical couplings are typically defined by processes occurring at some energy scale and are directly related to the more physical mass-dependent schemes. It is important to understand how to match couplings in mass-independent schemes with the physical couplings. Again, this is best illustrated with the specific instance of the ϕ^4 theory. Since $\Gamma^{(4)}(s, t, u)$ is a physical, scheme-independent, amplitude, we can determine the relation between the MOM and $\overline{\text{MS}}$ couplings by matching eqs.(5.2.5) and (5.4.4). At one-loop level we have

$$\lambda_{\text{MOM}} = \lambda_{\overline{\text{MS}}} + a\lambda_{\overline{\text{MS}}}^2 + \mathcal{O}(\lambda^3), \quad (5.4.6)$$

where a is given by

$$a = \frac{3}{32\pi^2} \int_0^1 dx \log \frac{m^2 + \mu^2 x(1-x)}{\mu^2}. \quad (5.4.7)$$

For $\mu \gtrsim m$ we can neglect the mass term appearing in eq.(5.4.7), in which case we get

$$a = -\frac{1}{16\pi^2}. \quad (5.4.8)$$

We can finally answer the previous questions: is it meaningful to talk about running coupling? If so, how do we determine the “correct” running? There is no notion of “correct” coupling. The running given by any sensible scheme is meaningful, provided we consistently associate it to expressions computed in that scheme. The simplest mass-independent schemes and their associated simple running, can reliably be used in the UV, and then matched, by means of formulae analogous to eq. (5.4.6), to the physically defined coupling constants. In these lecture notes most of the time we will adopt the $\overline{\text{MS}}$ renormalization scheme.

5.4.1 Mass Threshold Effects

Let us find an approximation to the RG evolution of λ in MOM which is more refined than eqs. (5.2.10) and (5.2.11). We proceed as follows. The solution of eq. (5.2.7), for $\mu \gg m$, is given by

$$\lambda^{-1}(\mu) = -\frac{3}{16\pi^2} \log \mu + c, \quad (5.4.9)$$

where c is an integration constant. This is fixed by matching eq. (5.4.9) with the exact flow implied by eq. (5.2.7):

$$\lambda(\mu) = \lambda(\mu_0) + \int_{\mu_0}^{\mu} \frac{\beta}{\mu'} d\mu'. \quad (5.4.10)$$

Let us take $\mu_0 = 0$, $\lambda(0) \equiv \lambda$ and $\mu \gtrsim m$, i.e. slightly larger than μ , so that we can still neglect m with respect to μ , but small enough that no large $\log(\mu/m)$ term appears. In the range $[\mu_0, \mu]$, $\lambda(\mu) \simeq \lambda$, and eq. (5.4.10) is well approximated by

$$\lambda(\mu) \simeq \lambda + \frac{3\lambda^2}{32\pi^2} \int_0^1 dx \log \frac{\mu^2 x(1-x)}{m^2} = \lambda + \frac{3\lambda^2}{16\pi^2} \left(\log \frac{\mu}{m} - 1 \right), \quad (5.4.11)$$

from which $c = \lambda^{-1} + 3/(16\pi^2)(1 + \log m)$ and hence

$$\lambda^{-1}(\mu) = \lambda^{-1} - \frac{3}{16\pi^2} \left(\log \frac{\mu}{m} - 1 \right) \theta(\mu - m) \quad (\text{MOM}). \quad (5.4.12)$$

Notice that $\lambda(\mu)$ in eq. (5.4.12) is discontinuous in m :

$$\lim_{\mu \rightarrow m^-} \lambda^{-1}(\mu) = \lambda^{-1} \neq \lim_{\mu \rightarrow m^+} \lambda^{-1}(\mu) = \lambda^{-1} + 3/(16\pi^2). \quad (5.4.13)$$

This discontinuity essentially takes into account the mass term disturbance to the UV running for $\mu \gtrsim m$. The constant $3/(16\pi^2)$ is often called mass threshold effect. It can be verified that eq. (5.4.12) is an excellent approximation to the exact one-loop running given by eq. (5.2.7) far away from the threshold region $\mu \simeq m$.

In $\overline{\text{MS}}$ mass threshold effects are extracted by carefully matching the couplings in the UV theory ($\mu \gg m$) with that in the effective IR theory ($\mu \ll m$) where the massive particle has been integrated out.

5.5 Asymptotic Behaviours of β -Functions

The β -function is the crucial object to determine the evolution of a coupling constant. In general, in a theory with n couplings g_i , we have to solve a set of coupled differential equations of the kind

$$\mu \frac{dg_i}{d\mu} = \beta_i, i = 1, \dots, n, \quad (5.5.1)$$

where the β_i in eq. (5.5.1) depend on all the other couplings and masses of the theory. We can get rid of the masses by considering mass-independent renormalization schemes, or equivalently by considering the UV behaviour in mass-dependent schemes. In both cases $\beta_i = \beta_i(g_j)$. Yet, it is not possible to describe the main properties of the solutions $g_i = g_i(\mu)$, because the system (5.5.1) is in general too complicated. We can further simplify the situation by considering a single coupling g . In perturbation theory, $\beta(g)$ admits an expansion as follows:

$$\beta(g) = \beta_0 g^2 + \beta_1 g^3 + \mathcal{O}(g^4). \quad (5.5.2)$$

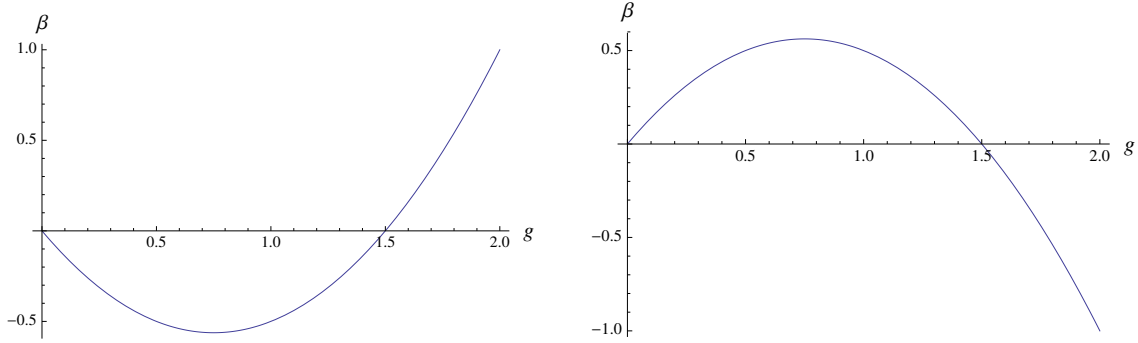


Figure 5.1: (Left panel) Schematic picture of the β -function of an IR stable fixed point (Right panel) Schematic picture of the β -function of an UV stable fixed point. In both cases we have taken $g^* = 3/2$. The units in both axes are arbitrary and irrelevant.

Note that it is always possible, by a proper coupling redefinition, to write the β -function expansion as in eq. (5.5.2). For instance, in QED $\beta(e)$ starts at cubic, rather than quadratic, order in the coupling, but it is enough to consider $\beta = \beta(e^2)$ rather than $\beta(e)$ to put the β in the form (5.5.2). In general eq.(5.5.2) applies when g is the effective loop-wise expansion parameter. At leading order, i.e. neglecting β_1 , eq. (5.5.2) is solved by

$$g(\mu) = \frac{g_0}{1 - g_0\beta_0 \log(\mu/\mu_0)} , \quad (5.5.3)$$

where $g_0 = g(\mu_0)$. The fate of $g(\mu)$ is entirely governed by the sign of β_0 . When $\beta_0 > 0$, like in the ϕ^4 or QED cases, $g(\mu)$ is marginally irrelevant, it increases in the UV and at some high energy scale the theory is no longer perturbative.

When $\beta_0 < 0$, on the other hand, the opposite happens. The coupling is marginally relevant and it decreases in the UV. In the limit of infinite energy, the coupling vanishes. Theories with (all) couplings of this kind are called asymptotically free. The low energy regime of these theories (including the spectrum of particles) is not perturbatively accessible. The coupling formally diverges at the scale⁸

$$\Lambda = \mu e^{\frac{1}{g(\mu)\beta_0}} . \quad (5.5.4)$$

This scale, which is RG invariant, is said to be a dynamically generated scale, since there is no trace of it in the classical theory. It is a purely quantum effect. The most famous theory belonging to this class of theories is QCD.

⁸Equation (5.5.4) applies when the masses of the particles contributing to $\beta(g)$ are smaller than the scale Λ . Otherwise, as we discussed in section 5.4, they should be integrated out, resulting in a change of the coefficient β_0 .

It is useful to consider two other more exotic forms of β -function, which correspond to a perturbative expansion around a non-trivial value of the coupling (see figs. 5.1). Let us assume that there exists a critical value of the coupling, g^* , such that $\beta(g^*) = 0$. If g is sufficiently close to g^* , we can expand β as follows:

$$\beta(g) = \beta(g^*) + \beta^*(g - g^*) + \mathcal{O}(g - g^*)^2 = \beta^*(g - g^*) + \mathcal{O}(g - g^*)^2. \quad (5.5.5)$$

The solution of eq. (5.5.1), with β as in eq. (5.5.5), is

$$g(\mu) = g^* + (g(\mu_0) - g^*) \left(\frac{\mu}{\mu_0} \right)^{\beta^*} + \mathcal{O}(g - g^*)^2. \quad (5.5.6)$$

Independently of $g(\mu_0)$, when $\beta^* < 0$ the coupling approaches g^* in the UV, while for $\beta^* > 0$ it approaches g^* in the IR. In the former case, g^* is called an ultraviolet stable fixed point of the RG flow, in the latter an infrared stable fixed point of the RG flow. When $g = g^*$ and β vanishes, no running occurs. In this case the theory is invariant under scaling transformations. More in general, it can be shown that it is invariant also under conformal transformations. A theory of this sort is called a Conformal Field Theory (CFT). Trivial CFT's are free theories of massless particles, for which $g = 0$ and β vanishes. Non-trivial CFT's are generally strongly coupled, because a non-trivial zero of a β -function requires a cancellation among different orders in the perturbative expansion. But this signals a break down of perturbation theory, where by definition a term of order $n + 1$ should be parametrically smaller than the one of order n . There is however a way to possibly get weakly coupled CFTs. Suppose that in a theory the coefficient β_0 of the one-loop β function is *accidentally* small. If so, setting to zero eq.(5.5.2) we get, aside from $g = 0$, the non-trivial solution

$$g^* = -\frac{\beta_0}{\beta_1}. \quad (5.5.7)$$

This solution (5.5.6) can be trusted only when $\beta_0/\beta_1 \ll 1$, our working hypothesis, otherwise higher order terms would destabilize it and bring g^* at a generically strongly coupled value. The fixed point (5.5.6) is called of the Caswell-Banks-Zaks kind [14, 6]. In non-abelian gauge theories with a large gauge group and a large number of matter fields one can tune β_0 to be parametrically small. With a proper number of matter fields, a Banks-Zaks fixed point can be obtained also in QCD. The clearest example is the one of QCD with 16 fermions, in which case $\beta_0 < 0$ and in modulus is the smallest possible. Lattice simulations show that, contrary to “real world” QCD, at large distances this theory does not confine and approaches a fixed-point, with a β -function qualitatively as the one depicted in the left panel of fig.5.1.

We now show that the coefficients β_0 and β_1 in eq. (5.5.2) do not depend on the mass-independent renormalization scheme chosen, while higher order terms are scheme dependent. The coupling constants in two different schemes, call them schemes g and $\tilde{g}$, are equal at lowest order, but they start to differ at higher orders in the coupling (see eq. (5.4.6) for a concrete relation between coupling constants defined in different renormalization schemes). In general we have

$$\tilde{g}(g) = g + ag^2 + \mathcal{O}(g^3), \quad (5.5.8)$$

where a is a constant. Their associated β -functions are related as

$$\tilde{\beta}(\tilde{g}) = \mu \frac{d\tilde{g}}{d\mu} = \mu \frac{dg}{d\mu} \frac{d\tilde{g}}{dg} = \beta(g) \frac{d\tilde{g}}{dg}. \quad (5.5.9)$$

Taking $\beta(g)$ as in eq. (5.5.2) we have

$$\begin{aligned} \tilde{\beta}(\tilde{g}) &= (\beta_0 g^2 + \beta_1 g^3 + \mathcal{O}(g^4))(1 + 2ag + \mathcal{O}(g^2)) \\ &= (\beta_0(\tilde{g} - a\tilde{g}^2)^2 + \beta_1\tilde{g}^3 + \mathcal{O}(\tilde{g}^4))(1 + 2a\tilde{g} + \mathcal{O}(\tilde{g}^2)) \\ &= \beta_0\tilde{g}^2 + \beta_1\tilde{g}^3 + \mathcal{O}(\tilde{g}^4), \end{aligned} \quad (5.5.10)$$

where the $\mathcal{O}(\tilde{g}^4)$ terms are different from the $\mathcal{O}(g^4)$ terms in $\beta(g)$. Hence we have proved the scheme independence of the coefficients β_0 and β_1 .

The generalization of eqs. (5.5.2) and (5.5.8) for multiple couplings are

$$\begin{aligned} \beta_i(g_j) &= \beta_{ijk}^{(0)} g_j g_k + \beta_{ijkl}^{(1)} g_j g_k g_l + \mathcal{O}(g_i^4), \\ \tilde{g}_i(g_j) &= g_i + a_{ijk} g_j g_k + \mathcal{O}(g^3). \end{aligned} \quad (5.5.11)$$

Proceeding as before, we get

$$\begin{aligned} \tilde{\beta}_{ijk}^{(0)} &= \beta_{ijk}^{(0)}, \\ \tilde{\beta}_{ijkl}^{(1)} &= \beta_{ijkl}^{(1)} + \frac{2}{3} \left(\beta_{sjk}^{(0)} a_{ils} + \beta_{sjl}^{(0)} a_{iks} + \beta_{slk}^{(0)} a_{ijs} - \beta_{ijs}^{(0)} a_{kls} - \beta_{iks}^{(0)} a_{jls} - \beta_{ils}^{(0)} a_{jks} \right). \end{aligned} \quad (5.5.12)$$

The two-loop coefficients of the β -function for multiple couplings are in general scheme-dependent.

5.5.1 CFTs and RG Flows

Roughly speaking CFTs are particular QFTs where the β -functions vanish. They are fixed points of the RG flow and hence can be the end-points of RG trajectories. A QFT can be seen as a deformation of a CFT induced by relevant or marginal operators. RG flows are in one to one correspondence to deformations of the UV CFT. Generally we

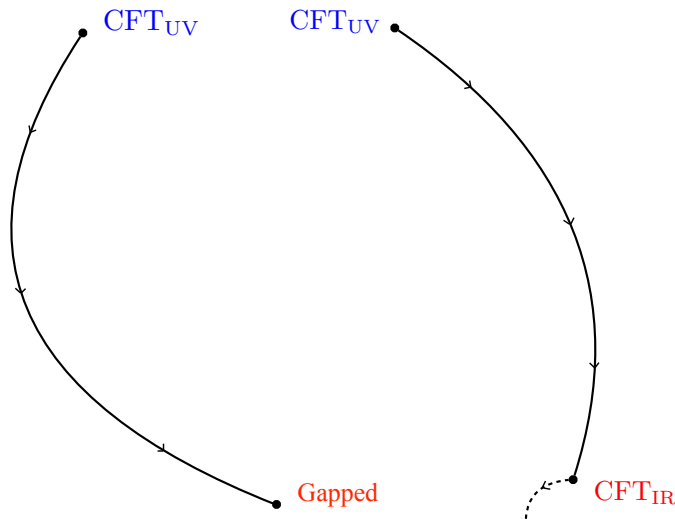


Figure 5.2: Cartoon of two RG flow trajectories from the UV to the IR. The one in the left ends up in a gapped theory, the one in the right to a CFT. The latter can be the starting point of another RG trajectory (dashed line).

describe the UV starting point of an RG flow with a weakly coupled description in terms of a Lagrangian, namely deformations of a free CFT. If the theory remains weakly coupled along the entire RG flow, the UV Lagrangian description remains valid up to the IR. In the most interesting cases, however, the theory becomes strongly coupled along the RG flow. When this happens, the UV description becomes inadequate and some other description is needed. Typically an RG trajectory ends in a vacuum with no degrees of freedom. This happens whenever a QFT in the IR has only massive degrees of freedom and is said to be “gapped”. At energies below the mass of the lightest particle, all excitations are heavy and can be integrated out, resulting only in the vacuum state. Typically mass deformations in the UV give rise to a gapped theory in the IR. Suitable RG trajectories can end in the IR to another CFT, which can be free or interacting. In this case the QFT contains massless excitations and is said to be “gapless”. Note that CFTs, aside from the free case and a few notable cases, do not admit a Lagrangian description and should be studied by other means.

Let us consider a few examples. A massive free particle can be interpreted as a CFT deformed by a relevant operator (corresponding to the mass term). In the UV we have a free CFT and in the IR a gapped theory. The ordinary ϕ^4 theory we have been discussing so far can be seen as a free scalar CFT deformed by a relevant and a classically marginal

operator (ϕ^2 and ϕ^4). Similarly we might refer to QED as the sum of two free CFTs (free photon and electron) deformed by a relevant and a classically marginal interaction ($\bar{\psi}\psi$ and $\bar{\psi}\gamma^\mu\psi A_\mu$). In the last two examples we cannot consider the strict UV limit because of the Landau pole singularity, but we can flow in the IR. If we start at some high energy scale with a weakly coupled description, the theories remain weakly decoupled along the entire RG flow. Both the ϕ^4 theory and QED are gapped or gapless, depending on whether we have or not mass deformations. A more interesting example is given by QCD. In the UV it admits a Lagrangian description in terms of weakly coupled quarks and gluons. In the strict UV limit, $\mu \rightarrow \infty$, the gauge coupling vanishes and quark masses become negligible. QCD in the far UV is a CFT, direct sum of free CFTs, one for each quark and gluon. In the IR QCD becomes strongly coupled and the UV Lagrangian description inadequate. The theory undergoes confinement and chiral symmetry breaking, the asymptotic states are no longer quarks and gluons, but massive excitations given by their bound states: mesons, baryons and glueballs (bound states of gluons only). The theory in the far IR is gapped. If we set the quark masses to zero, QCD in the IR has massless mesons corresponding to the Goldstone bosons of the spontaneously broken chiral symmetry. In the far IR, $\mu \rightarrow 0$, Goldstone bosons become free and we end up with a gapless theory, a free CFT. Consider now non-ordinary QCD with 16 massless quarks. This theory flows in the IR to a gapless theory, but this time this is an interacting CFT. Adding a mass deformation m to all 16 fermions result in a different RG flow. For $\mu \gg m$, the effect of m is negligible and the coupling g approaches its fixed point value approximately given by (5.5.7). When $\mu \ll m$, the fermions can be integrated out and the coupling starts to run again. Eventually we have confinement and chiral symmetry breaking and in the far IR we end up with a gapped theory. Mass deformations do not always give rise to a gapped theory.⁹ For instance, consider the ϕ^4 theory in $d = 3$ dimensions. This theory is super-renormalizable, being the quadratic and quartic deformations both relevant in $d = 3$. In the UV the theory approaches a free CFT. In the IR it is generally gapped. However, by appropriately tuning the mass and the coupling constant we can reach a gapless phase given by an interacting CFT, the renowned Ising model.

The IR CFT, end-point of a given RG flow, can be taken as the starting point of new RG flows induced by its deformations. The IR CFT should be then considered as a UV CFT from the point of view of the new RG flow. Note that the RG flow is irreversible. Given a QFT, we can uniquely compute its RG flow towards the IR, for instance by integrating out UV degrees of freedom à la Wilson. On the other hand, we cannot uniquely reconstruct

⁹For deformations that do not break any symmetry, like scalar masses, the presence or not of a deformation can depend on the renormalization scheme chosen.

the UV theory from where we flowed to, since more UV theories can flow to the same IR theory. This does not imply of course that we can never study the UV properties of a theory. As long as it admits a weakly coupled description in terms of a Lagrangian we can use the perturbative β -functions to infer its UV properties, as we did so far (and continue to do) in these lecture notes. In this spirit, we can sometimes follow the RG flow upward towards the UV. For instance, suppose we deform a free CFT by a marginally irrelevant operator. Typically, as in the ϕ^4 or in the QED case, this would give rise to a QFT with a Landau pole singularity, where the weakly coupled description breaks down and the UV is not under control. However, suppose that the β -function of the coupling looks like the right panel in fig.5.1 and can be controlled in perturbation theory. This can be achieved for instance by considering the Caswell-Banks-Zaks trick and reversing the signs of β_0 and β_1 . In this way the fixed point (5.5.7) will be UV stable and give rise to a non-trivial CFT in the UV.

CFTs do not always appear as isolated points in the space of coupling constants, as depicted in fig.5.2. Sometimes we can have a continuous family of CFTs. This happens when one or more deformations in the theory are exactly (and not only classically) marginal. Perturbing a CFT by one of such deformations does not induce an RG flow, but simply corresponds to move along this space of CFTs. Such space is often denoted a “conformal manifold”.¹⁰ Understanding RG flows, their relation, and the properties of CFTs is crucial in QFT. In particular, CFTs describe second-order phase transitions in critical phenomena, and play an important role in gravity by means of the renowned AdS/CFT correspondence. But these are subjects which are beyond the scope of these lecture notes.

5.6 The Callan-Symanzik RG Equations

We have seen that at the quantum level we are necessarily led to introduce the sliding scale μ . It is often useful to define (renormalize) at the same scale μ not only the coupling constant but also the fields themselves. For instance, in MS or $\overline{\text{MS}}$ the scale μ shows up in a way that affects all Green functions. In particular, the mass renormalized in those schemes does not coincide with the physical mass and the wave function renormalization factor Z is no longer the one appearing in the LSZ reduction formulae. Changing the definition of Z , now a function of μ , $Z = Z(\mu)$, will however make external legs contribute to physical processes, because the Z entering the LSZ reduction formulae is the physical μ -independent one. This is quite clear from eq.(2.3.10). Aside from irrelevant factors, the

¹⁰ In $d = 4$ the existence of conformal manifolds seems to require supersymmetry, but we have no proof at the moment of this statement.

S-matrix reads

$$S_{m \rightarrow n}^C \sim \frac{1}{Z^{\frac{n+m}{2}}} G_B^{(n+m)}(p_i, -q_j) = \frac{1}{Z^{\frac{n+m}{2}}} Z(\mu)^{\frac{n+m}{2}} G^{(n+m)}(p_i, -q_j, \mu), \quad (5.6.1)$$

where $G_B^{(n+m)}$ and $G^{(n+m)}$ are the bare and renormalized connected Green functions, respectively. Since S -matrix elements do not depend on μ , we see that the bare, rather than the renormalized, Green functions are those that are independent of μ . A similar reasoning applies of course also to the 1PI amplitudes. Due to the amputation of the external legs, one has the opposite power of Z 's relating bare and renormalized 1PI amplitudes. Taking again the $\lambda\phi^4$ theory as our working example, we have

$$\Gamma_B^{(n)} = Z^{-\frac{n}{2}}(\mu) \Gamma(\mu), \quad (5.6.2)$$

and hence

$$\mu \frac{d\left(Z^{-\frac{n}{2}} \Gamma^{(n)}\right)}{d\mu} = Z^{-\frac{n}{2}} \left[\mu \frac{\partial}{\partial \mu} + \beta\left(\lambda, \frac{m}{\mu}\right) \frac{\partial}{\partial \lambda} - n\gamma\left(\lambda, \frac{m}{\mu}\right) \right] \Gamma^{(n)} = 0, \quad (5.6.3)$$

where

$$\gamma\left(\lambda, \frac{m}{\mu}\right) \equiv \frac{1}{2} \mu \frac{d \log Z}{d\mu} \quad (5.6.4)$$

is the anomalous dimension of the field ϕ . Eqs.(5.6.3) are called the Callan-Symanzik (CS) equations.¹¹ For simplicity, in eq.(5.6.3) we have assumed that m is the physical μ -independent mass, otherwise we would also get a term proportional to the β -function of the mass. We will consider the RG flow of the mass term, and more in general of dimensionful couplings, in section 5.10. Similar equations hold for the connected Green functions $G^{(n)}$, in which case the exponent of the Z factor in eq.(5.6.2) is $+n/2$ instead of $-n/2$, and correspondingly the last term in eq.(5.6.3) has opposite sign:

$$\mu \frac{d\left(Z^{\frac{n}{2}} G^{(n)}\right)}{d\mu} = Z^{\frac{n}{2}} \left[\mu \frac{\partial}{\partial \mu} + \beta\left(\lambda, \frac{m}{\mu}\right) \frac{\partial}{\partial \lambda} + n\gamma\left(\lambda, \frac{m}{\mu}\right) \right] G^{(n)} = 0. \quad (5.6.5)$$

The CS equations (5.6.3) can be solved as follows in the UV regime where we can neglect the m/μ dependence of β and γ , possibly present only in mass-dependent schemes. As a first step, we can get rid of the last term in eq. (5.6.3) by defining a new Green function $\hat{\Gamma}^{(n)}$:

$$\Gamma^{(n)}(p_i, \lambda, m, \mu) = e^{n \int_{\lambda_0}^{\lambda} d\lambda' \frac{\gamma(\lambda')}{\beta(\lambda')}} \hat{\Gamma}^{(n)}(p_i, \lambda, m, \mu) \quad (5.6.6)$$

¹¹Actually, the original Callan [12] and Symanzik [55] equations were a bit different. There was no sliding scale μ , the physical mass playing essentially the role of μ , and the equations were not homogeneous in $\Gamma^{(n)}$. The more modern version of the Callan Symanzik equations (5.6.3) in terms of μ were developed shortly later, see refs.[59] and [56].

such that

$$\left(\mu \frac{\partial}{\partial \mu} + \beta(\lambda) \frac{\partial}{\partial \lambda} \right) \hat{\Gamma}^{(n)}(p_i, \lambda, m, \mu) = 0. \quad (5.6.7)$$

The above equation is solved by introducing an auxiliary variable t and t -dependent functions $\mu(t)$ and $\lambda(t)$ with

$$\mu(t) = e^t \mu, \quad \lambda(0) = \lambda. \quad (5.6.8)$$

We then demand that $\lambda(t)$ is such that

$$\frac{d}{dt} \hat{\Gamma}^{(n)}(p_i, \lambda(t), m, \mu(t)) = 0. \quad (5.6.9)$$

When eq. (5.6.9) is satisfied, $\hat{\Gamma}^{(n)}(p_i, \lambda(t), m, \mu(t))$ is independent of t . Evaluating it at $t = 0$, we see that it coincides with our original Green function $\hat{\Gamma}^{(n)}(p_i, \lambda, m, \mu)$. Using the chain rule of derivatives, eq. (5.6.9) equals

$$\left(\mu \frac{\partial}{\partial \mu} + \frac{d\lambda}{dt} \frac{\partial}{\partial \lambda} \right) \hat{\Gamma}^{(n)}(p_i, \lambda(t), m, \mu(t)) = 0. \quad (5.6.10)$$

Comparing eqs. (5.6.10) and (5.6.7), we see that the following equation should hold:

$$\frac{d\lambda(t)}{dt} = \beta(\lambda(t)). \quad (5.6.11)$$

For any t , provided that $\lambda(t)$ satisfies eq. (5.6.11), we have

$$\hat{\Gamma}^{(n)}(p_i, \lambda, m, \mu) = \hat{\Gamma}^{(n)}(p_i, \lambda(t), m, e^t \mu). \quad (5.6.12)$$

In terms of the original Green functions $\Gamma^{(n)}$ and changing variable $d\lambda \rightarrow dt$ inside the integral appearing in eq.(5.6.6), eq.(5.6.12) reads

$$e^{-n \int_{t_0}^0 \gamma(t') dt'} \Gamma^{(n)}(p_i, \lambda, m, \mu) = e^{-n \int_{t_0}^t \gamma(t') dt'} \Gamma^{(n)}(p_i, \lambda(t), m, e^t \mu). \quad (5.6.13)$$

Rescaling $p \rightarrow e^t p$, eq.(5.6.13) is rewritten as

$$\Gamma^{(n)}(e^t p_i, \lambda, m, \mu) = e^{-n \int_0^t \gamma(t') dt'} \Gamma^{(n)}(e^t p_i, \lambda(t), m, e^t \mu). \quad (5.6.14)$$

Recall that the classical dimension of $\Gamma^{(n)}$ is $4 - n$, so that

$$\Gamma^{(n)}(e^t p_i, \lambda(t), m, e^t \mu) = e^{t(4-n)} \Gamma^{(n)}(p_i, \lambda(t), e^{-t} m, \mu). \quad (5.6.15)$$

Combining eqs.(5.6.14) and (5.6.15) we finally get

$$\Gamma^{(n)}(e^t p_i, \lambda, m, \mu) = e^{t(4-n)-n \int_0^t \gamma(t') dt'} \Gamma^{(n)}(p_i, \lambda(t), e^{-t} m, \mu). \quad (5.6.16)$$

Equation (5.6.16) tells us that a Green function at some energy scale $e^t p$, modulo an overall factor, equals the same Green function evaluated at the energy scale p , provided

we replace the coupling λ with its running counterpart $\lambda(t)$, solution of eq. (5.6.11). In the high energy regime $t \rightarrow \infty$, mass corrections in the Green functions are negligible, in agreement with our expectations for an (IR) relevant coupling. The anomalous dimension γ affects the effective scaling dimension of the Green function, as expected.

The UV behaviour of $\Gamma^{(n)}$ is particularly simple for theories with an UV stable fixed point λ^* . In this case, $\int \gamma[\lambda(t')] dt' \simeq \int \gamma(\lambda^*) dt' = t\gamma^*$, with $\gamma^* \equiv \gamma(\lambda^*)$. The UV behaviour of $\Gamma^{(n)}$ reduces then to

$$\Gamma^{(n)}(e^t p_i, \lambda, m, \mu) \simeq e^{4t - nt(1+\gamma^*)} \Gamma^{(n)}(p_i, \lambda^*, 0, \mu). \quad (5.6.17)$$

Another interesting situation arises for UV free theories (like QCD) where the (one-loop) running of the coupling is given by eq. (5.5.3) with $\beta_0 < 0$. Assuming that $\gamma(t) = \gamma_0 g$, we have $\int \gamma[g(t')] dt' = \int \gamma(g)/\beta(g') dg' = \gamma_0/\beta_0 \log g(t)/g$ and thus for large t

$$\Gamma^{(n)}(e^t p_i, g, m, \mu) \simeq e^{t(4-n)} \left(\frac{g}{g(t)} \right)^{\frac{n\gamma_0}{\beta_0}} \Gamma^{(n)}(p_i, 0, \mu). \quad (5.6.18)$$

5.6.1 Leading Logs and Callan-Symanzik Equations

In section 5.2 we have seen that the β -function allows us to improve the perturbative expansion by resumming large logs. However, it was not clear from our derivation that the resummed expression (5.2.13) captures exactly all higher-loops leading logs. The Callan-Symanzik equations allow us to fill this gap. Although this result is effectively encoded in eq.(5.6.16), it is more transparent to work out the explicit form of the solution (5.6.16) in perturbation theory, which is what we will do in this section. As usual, for concreteness we consider $\Gamma^{(4)}$ in the $\lambda\phi^4$ theory, though the derivation is more general. At high energies, neglecting masses and taking $s = t = u = -E^2 \gg m^2$, $\Gamma^{(4)}(E)$ is a function of λ and μ/E . In perturbation theory we can then write

$$\Gamma^{(4)}(E) = -\lambda \sum_{l=0}^{\infty} \lambda^l c_l(\mu/E), \quad (5.6.19)$$

where c_l are functions to be determined, and $c_0 = 1$. This expression should satisfy the Callan-Symanzik equation

$$\left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - 4\gamma \right) \Gamma^{(4)}(E) = 0. \quad (5.6.20)$$

In perturbation theory, $\beta = \beta_0 \lambda^2 + \beta_1 \lambda^3 + \mathcal{O}(\lambda^4)$, $\gamma = \gamma_0 \lambda + \gamma_1 \lambda^2 + \mathcal{O}(\lambda^3)$ and eq.(5.6.20) should be satisfied order by order. At the first non-trivial order $\mathcal{O}(\lambda^2)$ we have, recalling that $\gamma_0 = 0$,

$$\mu \partial_\mu c_1 = -\beta_0 \longrightarrow c_1 = -\beta_0 \log \frac{\mu}{E} + C_1, \quad (5.6.21)$$

where C_1 is an integration constant. At $\mathcal{O}(\lambda^3)$ we get

$$\mu \partial_\mu c_2 = -2\beta_0 c_1 - \beta_1 + 4\gamma_1 = 0 \longrightarrow c_2 = \beta_0^2 \log^2 \frac{\mu}{E} + (4\gamma_1 - \beta_1 - 2\beta_0 C_1) \log \frac{\mu}{E} + C_2. \quad (5.6.22)$$

Proceeding in this way allows us to establish an important result. *At loop order l the Green function is a polynomial of degree l in $\log \mu/E$.* At high energies the most relevant terms at loop l are the $\log^l$ terms (the leading logs). These are entirely determined in terms of β_0 :

$$c_l = (-\beta_0)^l \log^l \frac{\mu}{E} + \mathcal{O}\left(\log^{l-1} \frac{\mu}{E}\right). \quad (5.6.23)$$

Resumming these terms give

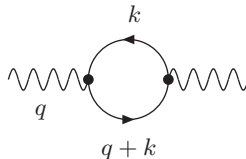
$$\Gamma_{\text{LL}}^{(4)}(E) = -\lambda \sum_{l=0}^{\infty} (-\beta_0 \lambda)^l \log^l(\mu/E) = \frac{-\lambda(\mu)}{1 - \beta_0 \lambda \log \frac{E}{\mu}}, \quad (5.6.24)$$

If we take $\mu = m$ in eq.(5.6.24) and recall that $\beta_0 = 3/(16\pi^2)$ we precisely recover eq.(5.2.13). This proves that there are no extra leading logs contributions that are not captured by the β -function. More in general, the leading log contributions of a correlation function also depends on the anomalous dimension coefficient γ_0 .

This analysis can be repeated to sub-leading orders (next-to leading logs, etc.). For example, by knowing the two-loop coefficients β_1 , γ_1 and the integration constant C_1 allows us to determine the next-to-leading logs, such as the log term in eq.(5.6.22), at any loop order. More in general, the series of logs of the form $\log^{l-n}$ at loop order l can be determined from the knowledge of $\beta_0, \dots, \beta_n$, $\gamma_0, \dots, \gamma_n$, $C_1, \dots, C_n$.

5.7 Anomalous Dimension of the Photon and QED β -function

The WT identities imply that the radiative corrections to the photon propagator are transverse to all orders in perturbation theory, as described by eq. (3.4.7). It is useful to explicitly check this result at one-loop order. In so doing, we will determine the one-loop counter-term Z_3 defined in eq.(3.4.13). Once Z_3 is known, we can compute the photon anomalous dimension and β -function for e . The only one-loop graph contributing to the photon propagator is the electron loop:



$$\equiv i\Pi_{\mu\nu}^{(1)}(q) = (-1)(ie)^2 \mu^\epsilon \text{Tr} \int \frac{d^d k}{(2\pi)^d} \frac{i(\not{k} + m)}{k^2 - m^2} \gamma_\mu \frac{i(\not{q} + \not{k} + m)}{(k+q)^2 - m^2} \gamma_\nu.$$

(5.7.1)

We can bring together the two electron propagators appearing in eq.(5.7.1) by introducing the Feynman parameter x :

$$\frac{1}{ab} = \int_0^1 dx \left[ax + b(1-x) \right]^{-2}. \quad (5.7.2)$$

Performing also the Dirac algebra, we have

$$i\Pi_{\mu\nu}^{(1)} = -de^2\mu^\epsilon \int_0^1 dx \int \frac{d^d k}{(2\pi)^d} \frac{k_\mu(k+q)_\nu + k_\nu(k+q)_\mu + (m^2 - k \cdot (q+k))\eta_{\mu\nu}}{\left[(k^2 - m^2)(1-x) + x(q+k)^2 - xm^2\right]^2}. \quad (5.7.3)$$

Let us now redefine $k \rightarrow k - qx$, so that the odd term $k \cdot q$ in the denominator is cancelled. By symmetry, after the momentum shift, the terms in the numerator odd in k vanish. We have then

$$i\Pi_{\mu\nu}^{(1)} = -de^2\mu^\epsilon \int_0^1 dx \int \frac{d^d k}{(2\pi)^d} \frac{2(k_\mu k_\nu - q_\mu q_\nu x(1-x)) + \eta_{\mu\nu}(m^2 + q^2 x(1-x) - k^2)}{(k^2 - \Delta)^2}. \quad (5.7.4)$$

where $\Delta \equiv m^2 - q^2 x(1-x)$. Recall that $k_\mu k_\nu \not\equiv (k^2/d)\eta_{\mu\nu}$ and perform then the Wick rotation to euclidean signature: $k^0 \rightarrow ik_4$, $k^2 \rightarrow -k^2$. After these steps, eq.(5.7.4) is rewritten as

$$i\Pi_{\mu\nu}^{(1)} = -ide^2\mu^\epsilon \int_0^1 dx \int \frac{d^d k}{(2\pi)^d} \frac{(1 - 2/d)k^2\eta_{\mu\nu} + (m^2 + q^2 x(1-x))\eta_{\mu\nu} - 2q_\mu q_\nu x(1-x)}{(k^2 + \Delta)^2}. \quad (5.7.5)$$

Let us now focus on the divergent part of eq.(5.7.5), i.e. to the $1/\epsilon$ pole. Using the trick described in section 3.5, we immediately get

$$\begin{aligned} i\Pi_{\mu\nu}^{(1)\text{div}} &= -id\frac{e^2}{8\pi^2\epsilon} \int_0^1 dx \left[\left(1 - \frac{2}{d}\right)\eta_{\mu\nu}(-2\Delta) + (m^2 + q^2 x(1-x))\eta_{\mu\nu} - 2q_\mu q_\nu x(1-x) \right] \\ &= i(\eta_{\mu\nu}q^2 - q_\mu q_\nu) \frac{e^2}{8\pi^2} \left(-\frac{4}{3} \right) \frac{1}{\epsilon}. \end{aligned} \quad (5.7.6)$$

The one-loop correction $\Pi_{\mu\nu}^{(1)}$ is in the transverse form expected from the WT identity (3.4.7). The value of the counter-term Z_3 needed to cancel the divergence (5.7.6) in a minimal subtraction scheme is (see fig.3.2)

$$Z_3 = 1 - \frac{e^2}{6\pi^2} \frac{1}{\epsilon}. \quad (5.7.7)$$

The one-loop photon anomalous dimension is

$$\gamma_A = \frac{1}{2} \frac{d \log Z_3}{d \log \mu} = -\frac{e\beta_e^{(d)}}{6\pi^2} \frac{1}{\epsilon}, \quad (5.7.8)$$

where $\beta_e^{(d)}$ is the QED β -function in d (and not 4) dimensions, which can be determined from eq.(5.3.2) by recalling the QED counter-term definitions (3.4.15). By dimensional analysis and using the identity $Z_1 = Z_2$, we have $e_B = Z_3^{-1/2} \mu^{\epsilon/2} e$ and hence

$$0 = \mu \frac{de_B}{d\mu} = Z_3^{-1/2} \left(-e\gamma_A + \frac{1}{2}\epsilon e + \beta_e^{(d)} \right) \implies \beta_e^{(d)} = -\frac{1}{2}\epsilon e + e\gamma_A. \quad (5.7.9)$$

Plugging the tree-level expression of $\beta_e^{(d)}$ in eq.(5.7.9) allows us to determine the actual 1-loop 4D photon anomalous dimension:

$$\gamma_A = \frac{e^2}{12\pi^2}. \quad (5.7.10)$$

Correspondingly, the 4D 1-loop β -function β_e reads

$$\beta_e = e\gamma_A = \frac{e^3}{12\pi^2}. \quad (5.7.11)$$

It is useful to write the RG behavior in terms of $\alpha \equiv e^2/(4\pi)$ and solve for its inverse. One gets, for $\mu_0, \mu \gg m$,

$$\alpha^{-1}(\mu) = \alpha^{-1}(\mu_0) - \frac{2}{3\pi} \log \frac{\mu}{\mu_0}. \quad (5.7.12)$$

For $\mu < m$, instead, like for the coupling λ in the ϕ^4 theory, α does not run and its value is approximately given by $\alpha(m)$. We are assuming here that there is a single charged particle. In presence of more charged particles with different masses, α stops running at the scale given by the lowest charged particle. Notice that in the real world, where the electron is the lightest charged particle, the absence of massless charged particles is crucial to prevent the electric charge (and hence all electrodynamical interactions) to vanish in the far IR.

5.8 “Irrelevant” RG Flow of Dimensionful Couplings

Depending on the energy scale, relevant and irrelevant operators parametrically either dominate the physics or are negligible. For instance, in the IR relevant operators, such as masses, are the dominant effect, while in the UV irrelevant operators make a theory ill-defined. We focus here on a perturbative situation in which the (ir)relevant operators can be seen as a small deformation in the theory. For instance, the effect of a mass term in the UV or the insertion of an irrelevant operator in the IR.

The concept of sliding scale and running coupling can be extended to relevant or irrelevant couplings, but care should be paid in this case to the scheme dependence of the results. In general, in presence of marginal and irrelevant couplings, the β -function

coefficients are scheme-dependent *even in the limit where mass effects are neglected*. If we denote by g_i and $\tilde{g}_i$ two coupling constants with classical mass dimension Δ_i in two different schemes, we have

$$\tilde{g}_i = g_i + a_{ijk}(\mu)g_jg_k + \mathcal{O}(g^3), \quad (5.8.1)$$

where by dimensional analysis

$$a_{ijk}(\mu) = \mu^{\Delta_i - \Delta_j - \Delta_k} c_{ijk}, \quad (5.8.2)$$

and c_{ijk} constant coefficients. The perturbative expansion of the β -functions for the g_i 's read

$$\beta_i = \mu \frac{dg_i}{d\mu} = \beta_0^{ijk}(\mu)g_jg_k + \mathcal{O}(g^3), \quad (5.8.3)$$

where

$$\beta_0^{ijk}(\mu) = \mu^{\Delta_i - \Delta_j - \Delta_k} b_0^{ijk}. \quad (5.8.4)$$

On the other hand, in the other scheme we have

$$\tilde{\beta}_i = \mu \frac{d\tilde{g}_i}{d\mu} = \tilde{\beta}_0^{ijk}(\mu)\tilde{g}_j\tilde{g}_k + \mathcal{O}(\tilde{g}^3), \quad (5.8.5)$$

where

$$\tilde{\beta}_0^{ijk}(\mu) = \mu^{\Delta_i - \Delta_j - \Delta_k} \tilde{b}_0^{ijk}, \quad \tilde{b}_0^{ijk} = b_0^{ijk} + c_{ijk}(\Delta_i - \Delta_j - \Delta_k). \quad (5.8.6)$$

Universal coefficients arise only when $\Delta_i - \Delta_j - \Delta_k = 0$. The relation (5.8.6) can alternatively be derived using dimensionless couplings h_i defined as $g_i = \mu_i^\Delta h_i$, in which case we have

$$\beta_{h_i} = -\Delta_i h_i + b_0^{ijk} h_j h_k + \mathcal{O}(h^3). \quad (5.8.7)$$

Renormalization schemes where classical dimensional analysis is preserved at the quantum level (i.e. where no powers of μ can be generated from quantum corrections), like MS or $\overline{\text{MS}}$ in DR, give automatically $b_0^{ijk} = 0$ when $\Delta_i - \Delta_j - \Delta_k \neq 0$, and only keep the scheme-independent coefficients. This is related to the property of dimensional regularization to set to zero all power-like divergences and keep only the logarithmic ones. Logarithmic divergences are special, because they are the only ones not saturated by UV physics, and sample uniformly all energy scales, up to the IR. Since the IR physics should be insensitive to the details of the different renormalization schemes, it follows that the associated β -function coefficients should be scheme-independent.

Let us now consider theories with only irrelevant couplings, assuming that no relevant or marginal couplings are present. A concrete example of this sort is given by the $\lambda\phi^4$

theory in five space-time dimensions. In 5d the ϕ^4 theory is non-renormalizable and λ is an irrelevant coupling of mass dimension -1 . Let us compute the one-loop RG evolution of λ . In presence of irrelevant couplings, dimensional regularization can no longer be considered a mass-independent scheme, since by dimensional analysis the β -functions can have a dependence on masses. If we neglect mass terms, however, dimensional analysis dictates that the irrelevant coupling with the smallest dimensions, λ itself in this example, has a vanishing β -functions to *all orders in perturbation theory*:

$$\beta_{5d}^{(\lambda)} = 0. \quad (5.8.8)$$

There is no analogue of the log resummation needed in treating marginal couplings and hence no need to improve the perturbative expansion.

Let us see what happens in a mass-dependent scheme. It is convenient to introduce a mass scale M and write the interaction as $\lambda/M\phi^4$, with λ dimensionless and of $\mathcal{O}(1)$. The scale M is the scale below which this theory makes sense as an effective field theory. We assume that the size of all dimensionful operators is governed by this scale, that is any dimensionful coupling of mass dimension $-n$ can be written as a dimensionless coupling of order one times $1/M^n$. Let us focus on the 1PI 4-point function. This is linearly divergent in a cut-off regularization and can be renormalized by momentum subtraction using eq. (5.2.4), with $\lambda \rightarrow \lambda/M$. The finite, 1PI four-point function, the 5d analogue of eq. (5.2.3), is

$$\begin{aligned} \Gamma_{5d}^{(4)}(s, t, u) = & -\frac{\lambda(\mu)}{M} - \frac{\lambda^2(\mu)}{32\pi^2 M^2} \int_0^1 dx \left(\sqrt{m^2 - sx(1-x)} - \sqrt{m^2 + \mu^2 x(1-x)} \right) \\ & + (s \rightarrow t) + (s \rightarrow u). \end{aligned} \quad (5.8.9)$$

The logs are replaced by square roots in 5d. In order to see that the sliding scale is not of great help, fix, say, $\mu = 0$ and consider the high energy regime $-s \gg m^2$. Like the 4d case (5.2.4), the one-loop term grows and can in principle be comparable to the tree-level term. This will occur at energies $\lambda\sqrt{|s|}/(32\pi^2 M) \sim 1$, which implies $|s| > M^2$. But this is beyond the regime in which the effective field theory makes sense, which is $|s| \ll M^2$! There is no energy regime in which the perturbative expansion can be improved by some RG resummation. The theory is simply weakly coupled for energies below M and becomes strongly coupled for energies above M . Nevertheless, there is nothing intrinsically wrong in introducing the sliding scale μ , so we can keep going and compute the β -function for λ by using the CS equations (5.6.3). In this way we get the 5d analogue of eq. (5.2.8):

$$\beta_{5d}^{(\lambda)} = \frac{3\lambda^2}{32\pi^2 M} \int_0^1 dx \frac{\mu^2 x(1-x)}{\sqrt{m^2 + \mu^2 x(1-x)}} + \mathcal{O}(\lambda^3). \quad (5.8.10)$$

For $\mu \gg m$ this simplifies to

$$\beta_{5d}^{(\lambda)} \simeq \frac{3\lambda^2 \sqrt{\mu^2}}{256\pi M}, \quad \mu \gg m. \quad (5.8.11)$$

The approximate solution of the RG flow is

$$\lambda_{UV}(\mu) \simeq \frac{\lambda(\mu_0)}{1 - \frac{3\lambda(\mu_0)}{256\pi M}(\mu - \mu_0)}, \quad \mu_0, \mu \gg m, \quad (5.8.12)$$

$$\lambda_{IR}(\mu) \simeq \text{constant}, \quad \mu \ll m. \quad (5.8.13)$$

Let us finally compare the $\Gamma_{5d}^{(4)}$ one obtains with and without the use of the RG technique at high energies (the analogue of eqs. (5.2.12) and (5.2.13)) for $s = t = u \equiv -E^2 \gg m^2$:

$$\Gamma_{5d}^{(4)}(E) \simeq -\frac{\lambda}{M} - \frac{3\lambda^2 E}{256\pi M^2} + \mathcal{O}(\lambda^2), \quad (\text{no RG}) \quad (5.8.14)$$

$$\Gamma_{5d}^{(4)}(E) \simeq -\frac{\lambda(E)}{M} = \frac{-\lambda/M}{1 - \frac{3\lambda E}{256\pi M}} + \mathcal{O}(\lambda^2). \quad (\text{RG}) \quad (5.8.15)$$

Like in 4d, eq. (5.8.15) reproduces the one-loop result (5.8.14) and, in addition, encodes higher order terms. However, in contrast to the 4d case, the higher order terms are scheme-dependent and are not “special” in the perturbative expansion. They are of order $\lambda^{l+1}(E/M)^l$, which is the order expected for a generic l -loop computation. There is no analogue of the logarithmic enhancement found in the 4d case. Hence eq. (5.8.15) should only be trusted at $\mathcal{O}(\lambda^2)$, in which case it merely reproduces the perturbative result obtained with no RG technique.

The vanishing of $\beta_{5d}^{(\lambda)}$ in DR at one-loop level is immediately seen by noting that the one-loop integral is proportional to $\Gamma(2 - d/2)$. No divergence then occurs because the (analytic continuation of the) Gamma function is well-behaved for negative half-integers values of its argument.

Scheme-independent β -functions can occur in this example for higher dimensional couplings. Consider for instance the dimension 9 operator ϕ^6 , that will be generated at some order in perturbation theory due to the non-renormalizability of the theory. Its coupling g has classical dimension -4 , so by dimensional analysis we expect in a mass-independent scheme

$$\beta_{5d}^{(g)} = c\lambda^4, \quad (5.8.16)$$

where c is a generally non-vanishing constant.¹² Since λ does not run, the solution to eq.(5.8.16) is simply

$$g(\mu) = g(\mu_0) + c\lambda^4 \log\left(\frac{\mu}{\mu_0}\right). \quad (5.8.17)$$

¹²By dimensional analysis other terms could appear in eq.(5.8.16), proportional to couplings associated to operators of dimension 7 and 8. These operators are however redundant and by the equations of motion (or equivalently a field redefinition) can be removed. We will discuss redundant operators in section 7.9.

We see that log terms can appear but in a rather dull way. Moreover, we should keep in mind that log's grow slowly and the range of energies we can explore is limited by the range of validity of the non-renormalizable effective theory.

A similar analysis can also be made for relevant couplings, with the obvious crucial difference that their effect decreases, rather than increases, in the UV. An example is once again provided by the ϕ^4 theory, this time in three dimensions, where the theory is super-renormalizable. The coupling λ has classical mass dimension $+1$ and in mass-independent schemes, and to all orders in perturbation theory, we have

$$\beta_{3d}^{(\lambda)} = 0, \quad (5.8.18)$$

while in mass-dependent schemes $\beta_{3d}^{(\lambda)} \neq 0$.¹³

5.9 Renormalization of Composite Operators**

Composite operators are obtained by taking products of elementary fields (and their derivatives) at the same space-time point. At the quantum level, when two fields approach the same space-time point, UV divergences arise. Naive product of fields does not hold, i.e. $\phi^2(x) \neq (\phi(x))^2$.¹⁴ For this reason, correlations functions involving composite operators are divergent, even in a renormalized theory where the infinities have been already buried in the counter-terms we discussed. Such additional divergences are removed by a renormalization of the composite operators themselves, generalization of the wave-function renormalization of the elementary fields. Composite operators can be relevant, marginal or irrelevant, and are labelled by their quantum numbers, including spin. In addition to scalar operators such as

$$\phi^2, \bar{\psi}\psi, F_{\mu\nu}F^{\mu\nu}, \phi^4, \phi\bar{\psi}\psi, (\partial\phi)^2, \phi^6, (\bar{\psi}\gamma^\mu\psi)(\bar{\psi}\gamma_\mu\psi), (\partial\phi)^4, (F_{\mu\nu}F^{\mu\nu})^2, \dots \quad (5.9.1)$$

we can have fermion, vector and more generally tensor operators

$$\bar{\psi}\gamma^\mu\psi, \phi^2\psi, \phi^2\bar{\psi}, F_{\mu\nu}F^{\nu\rho}, \partial_\mu\phi\partial_\nu\phi, \partial_\mu\bar{\psi}\partial_\nu\psi, \dots \quad (5.9.2)$$

If we denote by $\mathcal{O}$ a generic composite operator, omitting for simplicity of notation all indices labelling its quantum numbers, including its spin, in analogy to the elementary

¹³RG techniques can be used in this theory in another way, by introducing the so called ϵ -expansion [63]. This is an interesting and useful technique which will not be discussed in these notes.

¹⁴In the operatorial formalism, where ϕ is an operator in the Hilbert space, some (in general not all) divergences of composite fields are removed by the normal ordering procedure.



Figure 5.3: Example of a schematic way of depicting two composite operators (ϕ^2 and $\bar{\psi}\psi$) in Feynman diagrams.

field case we write the bare operator $\mathcal{O}^B = \mathcal{O}(\mu)Z^{\mathcal{O}}(\mu)$. Let $G^{(n)}$ be a generic n -point connected Green function of n operators $\mathcal{O}_i$.¹⁵

$$G_B^{(n)}(p_1, \dots, p_n) = \langle \mathcal{O}_1^B(p_1) \dots \mathcal{O}_n^B(p_n) \rangle = \left(\prod_{i=1}^n Z^{\mathcal{O}_i}(\mu) \right) G^{(n)}(p_1, \dots, p_n). \quad (5.9.3)$$

The straightforward generalization of the CS eqs. (5.6.5) for the above connected Green functions involving composite operators are

$$\left(\mu \frac{\partial}{\partial \mu} + \beta_i(\lambda) \frac{\partial}{\partial \lambda_i} + \sum_{i=1}^n \gamma^{\mathcal{O}_i}(\lambda) \right) G^{(n)} = 0, \quad (5.9.4)$$

where

$$\gamma^{\mathcal{O}_i}(\lambda) = \mu \frac{d \log Z^{\mathcal{O}_i}}{d \mu} \quad (5.9.5)$$

is the anomalous dimension of the operator $\mathcal{O}_i$. The anomalous dimension of a composite operator $\mathcal{O}$ is the most interesting quantity characterizing $\mathcal{O}$ and the one we are mostly interested in. It would then be useful to derive a more explicit expression for $\gamma^{\mathcal{O}_i}$.

In terms of Feynman diagrams, a composite operator made out of n elementary fields, can be represented by a single external line (representing a single operator) that splits into n lines representing the fundamental fields. We illustrate in fig.5.3 the case for the operators ϕ^2 and $\bar{\psi}\psi$, where the “external” line is depicted dashed to avoid confusion with the lines representing the elementary constituents. Of course more labels are needed if the operators carry extra structures, such as derivatives, gamma matrices, flavour indices, etc.

The simplest Green function to consider for a composite operator $\mathcal{O}$ made of n fundamental fields Φ_i ($i = 1, \dots, n$), is

$$G^{(n,1)}(p_1, \dots, p_n, q) = \langle \Phi_1(p_1) \dots \Phi_n(p_n) \mathcal{O}_n(q) \rangle, \quad (5.9.6)$$

where Φ_i denote generic scalar, fermion or gauge fields. By Wick-contracting each Φ_i with the corresponding operator contained in $\mathcal{O}_n$ we get a non-vanishing connected term

¹⁵Composite or elementary. In the latter case $Z^\phi(\mu) = \sqrt{Z(\mu)}$.

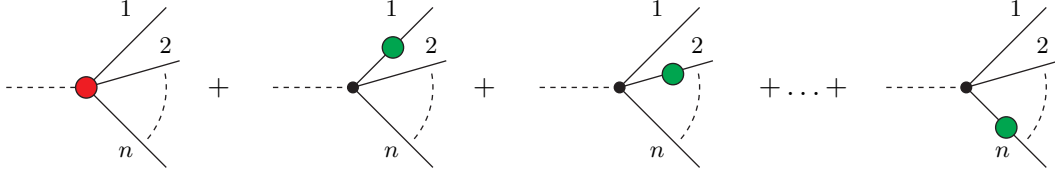


Figure 5.4: Counterterms needed to make a generic correlator $\langle \Phi_1 \Phi_2 \dots \Phi_n \mathcal{O}_n \rangle$ one-loop finite. The Green counterterms coincide with the wave function renormalization of the elementary fields Φ_i .

already at tree-level. For instance, for the two operators in fig.5.3 we have (all momenta incoming)

$$\begin{aligned} \langle \phi(p_1) \phi(p_2) \phi^2(q) \rangle_{\text{tree}} &= (2\pi)^4 \delta(p_1 + p_2 + q) 2 \frac{i}{p_1^2 - m^2} \frac{i}{p_2^2 - m^2}, \\ \langle \psi_a(p_1) \bar{\psi}_b(p_2) \bar{\psi} \psi(q) \rangle_{\text{tree}} &= (2\pi)^4 \delta(p_1 + p_2 + q) \frac{i(\not{p}_1 + m)_{ac}}{p_1^2 - m^2} \frac{i(\not{p}_2 + m)_{cb}}{p_2^2 - m^2}, \end{aligned} \quad (5.9.7)$$

where the factor 2 in the first relation arises from the two possible choices of Wick-contracting ϕ^2 with the external ϕ 's. The two figures in fig.5.3 can be interpreted as representing the tree-level value (5.9.7) of the corresponding Green functions. The tree-level contributions (5.9.7) can be factored out when we consider radiative corrections, relevant for the computation of anomalous dimensions. Since we are considering connected, rather than 1PI, Green functions, the corrections to the external legs should be included. Let us focus for simplicity on one-loop corrections only. In this case we have

$$Z^{\mathcal{O}_n} = 1 + \tilde{\delta}_{\mathcal{O}_n}, \quad Z = 1 + \delta, \quad (5.9.8)$$

where δ is the counterterm for the elementary field Φ (green counterterm in fig.5.4), with $\gamma_\Phi = 1/2d \delta/d \log \mu$ (recall the 1/2 factor in the definition of γ_Φ in eq.(5.6.4)).¹⁶ Let us denote by $\delta_{\mathcal{O}_n}$ the red counterterm that renormalizes the vertex defining the operator $\mathcal{O}_n$ (red counterterm in fig.5.4). It is not difficult to find the relation between $\tilde{\delta}_{\mathcal{O}_n}$, $\delta_{\mathcal{O}_n}$ and δ by looking at some special cases. If $\delta_{\mathcal{O}_n} = 0$, the operator $\mathcal{O}_n$ is made finite by rescaling its constituent fields by n powers of $\sqrt{Z}$ factors, so that we get

$$\tilde{\delta}_{\mathcal{O}_n} = \frac{n}{2} \delta \rightarrow \gamma^{\mathcal{O}_n} = n\gamma \quad (\delta_{\mathcal{O}_n} = 0). \quad (5.9.9)$$

On the other hand, if $\tilde{\delta}_{\mathcal{O}_n} = 0$ the counterterm $\delta_{\mathcal{O}_n}$ should precisely compensate for the external leg contributions to give a vanishing contribution:

$$\delta_{\mathcal{O}_n} = \frac{n}{2} \delta \quad (\tilde{\delta}_{\mathcal{O}_n} = 0). \quad (5.9.10)$$

¹⁶For simplicity we assume that $\mathcal{O}_n$ is made out of n identical elementary operators. The generalization in presence of different kinds of field is straightforward.



Figure 5.5: One-loop correction to the renormalization of the ϕ^2 operator.

We conclude that

$$\gamma^{\mathcal{O}_n} = \mu \frac{d}{d\mu} \tilde{\delta}_{\mathcal{O}_n}, \quad \tilde{\delta}_{\mathcal{O}_n} = -\delta_{\mathcal{O}_n} + \frac{n}{2} \delta, \quad (\text{one loop}). \quad (5.9.11)$$

It is useful to consider some example of computation of one-loop anomalous dimensions. We start with the ϕ^2 operator in the $\lambda\phi^4$ theory. In this theory $\delta = 0$ to one-loop level, so there are no external leg corrections. The only divergence comes from the one-loop diagram in fig.5.5, which is cancelled by the appropriate counterterm δ_{ϕ^2} . Modding out the tree-level contribution (5.9.7) to the scalar Green function $\langle\phi\phi\phi^2\rangle$, setting to zero the external momenta and scalar masses, we have

$$\left. \text{Diagram} \right|_{\text{div}} = -\frac{i\lambda}{2} \int \frac{d^d p}{(2\pi)^d} \frac{i^2}{(p^2)^2} = -\frac{\lambda}{16\pi^2\epsilon} = -\delta_{\phi^2}. \quad (5.9.12)$$

Recalling that $\beta(\lambda) = -\epsilon\lambda + \mathcal{O}(1)$ in d dimensions, one gets¹⁷

$$\gamma^{\phi^2} = \frac{d\tilde{\delta}_{\phi^2}}{d\log\mu} = -\frac{d\delta_{\phi^2}}{d\log\mu} = \frac{\lambda}{16\pi^2}. \quad (5.9.13)$$

Consider now the operator $\bar{\psi}\psi$ in QED. The one-loop diagrams to consider are depicted in fig.5.6 and now include the external leg corrections. We have

$$\text{Diagram} = i\Sigma_{1\text{loop}}(q) = (ie)^2 \mu^\epsilon \int \frac{d^d p}{(2\pi)^d} \frac{-i\eta_{\mu\nu}}{(q-p)^2} \gamma^\mu \frac{i(\not{p} + m)}{p^2 - m^2} \gamma^\nu. \quad (5.9.14)$$

The one-loop wave function renormalization for ψ is obtained by looking at the divergence in $i\Sigma_{1\text{loop}}(q)$ proportional to $\not{q}$. We can neglect the mass term in the numerator in eq.(5.9.14) and use the relation $\gamma^\mu \not{p} \gamma_\mu = (2-d)\not{p}$. Instead of introducing Feynman parameters to combine the fermion and photon propagators in a single denominator, we can use the methods of section 3.5 to directly read the divergent part. Expanding for $p \rightarrow \infty$, we

¹⁷Recall that the limit $d \rightarrow 4$ should always be performed at the end of the computation.

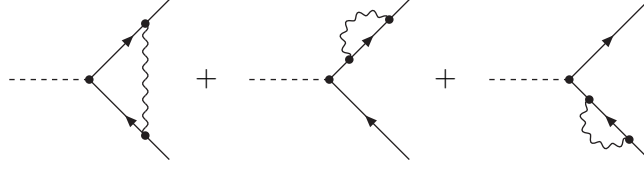


Figure 5.6: One-loop corrections to the renormalization of the $\bar{\psi}\psi$ operator.

have

$$\begin{aligned} i\Sigma_{1\text{loop}}(q)|_{\not{q} \text{ div}} &= -e^2(2-d) \int \frac{d^d p}{(2\pi)^d} \frac{\not{p}}{(p^2)^2} \left(1 + \frac{2q \cdot p}{p^2} + O(1/p^2)\right) \\ &= -e^2(2-d) \frac{2}{d} \frac{i\not{q}}{8\pi^2\epsilon} = i\not{q} \frac{e^2}{8\pi^2\epsilon}, \end{aligned} \quad (5.9.15)$$

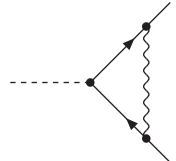
where in the second identity the factor of i arises from Wick rotating to euclidean momenta.

We conclude that

$$\delta_\psi = -\frac{e^2}{8\pi^2\epsilon} \quad (5.9.16)$$

is the wave function renormalization counterterm necessary to cancel the divergence (5.9.15).

Let us now compute the counterterm $\delta_{\bar{\psi}\psi}$ required to remove the divergence coming from the first graph in fig.5.6. Its superficial degree of divergence is zero, so we can safely set the external momenta and fermion masses to zero.



$$\left|_{\text{div}} = (ie)^2 \int \frac{d^d p}{(2\pi)^d} \frac{-i\eta_{\mu\nu}}{p^2} \gamma^\mu \frac{i\not{p}}{p^2} \frac{i\not{p}}{p^2} \gamma^\nu = \frac{4e^2}{8\pi^2\epsilon} = -\delta_{\bar{\psi}\psi}. \quad (5.9.17)$$

Using eq.(5.9.11), we get

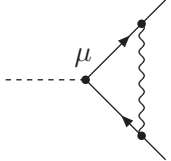
$$\gamma_{\bar{\psi}\psi} = \mu \frac{d}{d\mu} (-\delta_{\bar{\psi}\psi} + \delta_\psi) = \frac{6e\beta_e^{(d)}}{8\pi^2\epsilon}, \quad (5.9.18)$$

Plugging in eq.(5.9.18) finally gives

$$\gamma_{\bar{\psi}\psi} = -\frac{3e^2}{8\pi^2}. \quad (5.9.19)$$

As final example we compute the one-loop anomalous dimension of the vector operator $J^\mu = \bar{\psi}\gamma^\mu\psi$ in QED. We have already anticipated below eq.(4.4.10) that this operator, being conserved, should not give rise to additional divergences when inserted in Green functions. In other words, we expect that $\gamma_{J^\mu} = 0$, to all orders in perturbation theory.

Let us check this statement at the one-loop level. The computation is very similar to the one we performed for $\bar{\psi}\psi$. In particular, we still have the divergent graphs in fig.5.6, the only difference being in the insertion of a gamma matrix γ^μ in the vertex where the dashed line ends. Having already computed δ_ψ , we are only left with the computation of the counterterm δ_{J^μ} coming from the first graph in fig.5.6. We have



$$\begin{aligned} & \left|_{\text{div}} = (ie)^2 \int \frac{d^d p}{(2\pi)^d} \frac{-i\eta_{\alpha\beta}}{p^2} \gamma^\alpha \frac{i\not{p}}{p^2} \gamma^\mu \frac{i\not{p}}{p^2} \gamma^\beta = 2ie^2 \int \frac{d^d p}{(2\pi)^d} \frac{\not{p} \gamma^\mu \not{p}}{(p^2)^3} \\ & = -4ie^2 \frac{1}{4} \gamma^\mu \int \frac{d^d p}{(2\pi)^d} \frac{1}{(p^2)^2} = \gamma^\mu \frac{e^2}{8\pi^2 \epsilon} = -\gamma^\mu \delta_{J^\mu}. \end{aligned} \quad (5.9.20)$$

In eq.(5.9.20) the letter μ above the dashed line indicates the insertion of the matrix γ^μ and has been used to distinguish J^μ from $\bar{\psi}\psi$. We see that $\delta_{J^\mu} = \delta_\psi$ and hence

$$\gamma_{J^\mu} = \mu \frac{d}{d\mu} (-\delta_{J^\mu} + \delta_\psi) = 0, \quad (5.9.21)$$

as expected.

So far we have been assuming that the radiative corrections induced by a composite operator inserted in a given Green function are proportional to its tree-level expression. If this is not the case, we are forced to consider the composite operators that reproduce at tree-level the above radiative corrections. In other words, composite operators can mix under renormalization. On general grounds, mixing is possible for operators sharing the same quantum numbers. The degree of mixing also depends on the renormalization scheme. In particular, operators with different classical dimensions cannot mix in mass-independent schemes in theories with marginal couplings only. In this more complicated setting, the anomalous dimension becomes a matrix of anomalous dimensions:

$$\mathcal{O}_i^B = Z_{ij}^{\mathcal{O}}(\mu) \mathcal{O}_j(\mu), \quad \gamma_{ij}(\lambda) = (Z_{ik}^{\mathcal{O}})^{-1} \mu \frac{dZ_{kj}^{\mathcal{O}}}{d\mu}. \quad (5.9.22)$$

We should now consider a set of Green functions $G^{(i,j)}(p_1, \dots, p_{n_1}, q_j)$ defined as

$$G^{(i,j)}(p_1, \dots, p_{n_1}, q_j) = \left\langle \left(\prod_{k=1}^{n_i} \Phi_k(p_k) \right) \mathcal{O}_j(q) \right\rangle. \quad (5.9.23)$$

where $\prod_{k=1}^{n_i} \Phi_k$ denotes the elementary fields entering the composite operator $\mathcal{O}_i$. For non-scalar operators the product of fundamental fields Φ_k can be taken in different ways, giving rise to multiple tensor structures. In order to simplify the notation and the discussion, we do not keep track of these different structures. Schematically we have

$$G^{(i,j)} = \delta_{i,j} c^{(\text{tree})} + c_{ij}^{(\text{loop})}, \quad (5.9.24)$$

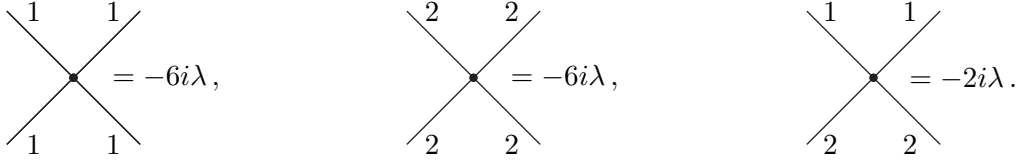


Figure 5.7: Feynman rules for the $O(2)$ ϕ^4 model.

where $c^{(\text{tree})}$ and $c_{ij}^{(\text{loop})}$ represent the classical and radiative contributions to the Green functions, respectively. Note that eq.(5.9.24) applies for a specific operator basis where $\mathcal{O}_i$ and $\prod_{k=1}^{n_j} \Phi_k$ are orthogonal at tree-level. Some of the $\mathcal{O}_i$ could have the same content of fundamental operators (a case of this kind will be considered in section 7.7), in which case they are distinguished by the above mentioned tensor structures. In general the coefficients $c_{ij}^{(\text{loop})}$ are divergent. Denoting by δ_{ij} the counterterms required to absorb the divergences in $c_{ij}^{(\text{loop})}$, in analogy to eq.(5.9.11) the anomalous dimension matrix at one-loop level reads¹⁸

$$\gamma_{ij} = \mu \frac{d}{d\mu} \tilde{\delta}_{ij}, \quad \tilde{\delta}_{ij} = -\delta_{ij} + \frac{1}{2} \delta_{i,j} \sum_{k=1}^{n_i} \delta_k, \quad (\text{one loop}). \quad (5.9.25)$$

A simple example of operator mixing is given by the quadratic scalar operators in the $O(2)$ scalar theory:

$$\mathcal{L} = \frac{1}{2}(\partial\phi_1)^2 + \frac{1}{2}(\partial\phi_2)^2 - \frac{1}{2}m^2(\phi_1^2 + \phi_2^2) - \frac{\lambda}{4}(\phi_1^2 + \phi_2^2)^2. \quad (5.9.26)$$

This is a generalization of the $\lambda\phi^4$ model, with two scalar operators ϕ_1 and ϕ_2 and a $O(2)$ global symmetry under which the fields (ϕ_1, ϕ_2) form a doublet. At one-loop level the anomalous dimensions of the operators $\phi_i\phi_j$ will be given by the insertion of one of the quartic vertices depicted in fig.5.7, with no contribution from the external legs. All the graphs have the topology in fig.5.5. We immediately get

$$\begin{aligned} \text{Diagram 1: } &= -\frac{3\lambda}{8\pi^2\epsilon} = -\delta_{11}, & \text{Diagram 2: } &= -\frac{\lambda}{8\pi^2\epsilon} = -\delta_{12}, \end{aligned} \quad (5.9.27)$$

where it is understood that we look for the divergent part of the diagram and for simplicity of notation we omitted the dashed line indicating the composite operator. Similarly we have $\delta_{22} = \delta_{11}$ and $\delta_{21} = \delta_{12}$. We see that, due the last vertex in fig.5.7, the operators

¹⁸Do not confuse in eq.(5.9.25) the Kronecker delta symbol $\delta_{i,j}$ from the matrix of counterterms δ_{ij} !

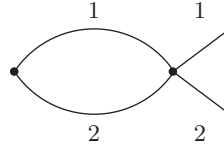
ϕ_1^2 and ϕ_2^2 do not have well-defined anomalous dimensions but mix under renormalization. The one-loop anomalous dimension matrix reads

$$\gamma = \frac{\lambda}{8\pi^2} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \quad (5.9.28)$$

and is diagonalized by the combination

$$\begin{aligned} \phi_n^2 &\equiv \phi_1^2 + \phi_2^2, & \gamma_n &= \frac{8\lambda}{16\pi^2}, \\ \phi_c^2 &\equiv \phi_1^2 - \phi_2^2, & \gamma_c &= \frac{4\lambda}{16\pi^2}. \end{aligned} \quad (5.9.29)$$

The remaining composite operator $\phi_1\phi_2$ does not mix at one-loop level with other operators and renormalizes multiplicatively:



$$= -\frac{2\lambda}{8\pi^2\epsilon} = -\delta, \quad (5.9.30)$$

so that

$$\gamma_{\phi_1\phi_2} = \frac{4\lambda}{16\pi^2}. \quad (5.9.31)$$

The careful reader might have noticed that in the above computation we did not pay attention to the $O(2)$ quantum numbers of the operators. In fact, the above mixing could have been avoided by a more clever choice of basis for the composite operators! The three operators ϕ_1^2 , ϕ_2^2 and $\phi_1\phi_2$ are better rearranged in a $O(2)$ neutral singlet ϕ_n^2 and a pair of charged operators ϕ_c^2 and $2\phi_1\phi_2$, the latter two being equal to the real and imaginary components of the complex field ϕ^2 , where $\phi = \phi_1 + i\phi_2$. The $O(2)$ symmetry explains why the operator combination (5.9.29) diagonalizes the anomalous dimension matrix and why ϕ_c^2 and $\phi_1\phi_2$ have the same anomalous dimension. As we mentioned, if two or more operators share the same quantum numbers, operator mixing will be unavoidable.

5.10 “Relevant” RG Flow of Dimensionful Couplings

We have seen in section 5.8 that the use of RG techniques applied to relevant and irrelevant couplings is not always as useful as in the standard situation where we have marginal couplings only. We have provided an example, the ϕ^4 theory in 5 space-time dimensions, where all couplings are irrelevant. In presence of marginal couplings, the situation changes. In a theory where marginal *and* (ir)relevant couplings are present together, the former can lead to large logs effects on the latter that should be summed. In other words, the quantum

RG flow of (ir)relevant operators, induced by marginal couplings, can and should be taken into account. This is what we mean by “relevant” RG flow in the title of this section.

If we denote by c the coupling constant associated to the classically lowest dimensional irrelevant operator $\mathcal{O}_c$, assuming it to be unique for the moment, dimensional analysis requires that to all orders in perturbation theory¹⁹

$$\beta_c = f(\lambda_i)c, \quad (5.10.1)$$

where we have denoted by λ_i all the marginal couplings in the theory. Note that β_c can have only a linear dependence on the coupling constant c , but an arbitrary dependence on the marginal couplings λ_i , parametrized in eq.(5.10.1) by the function f . We now show that f can be identified with the anomalous dimension of the operator $\mathcal{O}_c$. Let us denote by $G^{(n)}$ a generic n -point Green function obtained from a Lagrangian containing, besides marginal interactions, the term $c\mathcal{O}_c$. We can bring down from the action the term $c\mathcal{O}_c$ so that (in momentum space)

$$\begin{aligned} G^{(n)}(c, p_1, \dots, p_n) &= \langle \phi(p_1) \dots \phi(p_n) \rangle_c = \sum_{k=0}^{\infty} \frac{(ic)^k}{k!} \langle \phi(p_1) \dots \phi(p_n) \mathcal{O}_c^k(0) \rangle_0 \\ &\equiv \sum_{k=0}^{\infty} \frac{(ic)^k}{k!} G^{(n,k)}(0, p_1, \dots, p_n). \end{aligned} \quad (5.10.2)$$

The CS eqs. satisfied by the $G^{(n,k)}$ are

$$\left(\mu \frac{\partial}{\partial \mu} + \beta_i(\lambda) \frac{\partial}{\partial \lambda_i} + n\gamma + k\gamma^{\mathcal{O}_c}(\lambda) \right) G^{(n,k)}(0, p_1, \dots, p_n) = 0. \quad (5.10.3)$$

We can multiply eq. (5.10.3) by $(ic)^k/k!$ and sum over k to write

$$\left(\mu \frac{\partial}{\partial \mu} + \beta_i(\lambda) \frac{\partial}{\partial \lambda_i} + n\gamma + c\gamma^{\mathcal{O}_c}(\lambda) \frac{\partial}{\partial c} \right) G^{(n)}(c, p_1, \dots, p_n) = 0, \quad (5.10.4)$$

where we used that $c\partial G^{(n)}/\partial c = \sum_k k(ic)^k/k! G^{(n,k)}$. The β -function for c can be identified by looking at the term multiplying $\partial/\partial c$. We then have

$$\beta_c = c\gamma^{\mathcal{O}}(\lambda_i), \quad (5.10.5)$$

where the anomalous dimension $\gamma^{\mathcal{O}}(\lambda_i)$ is computed in the theory with $c = 0$. Comparing eq.(5.10.5) with (5.10.1), we see that

$$f = \gamma^{\mathcal{O}}. \quad (5.10.6)$$

¹⁹Again, this is only true in a mass-independent scheme. From now on, the use of a mass-independent scheme will be assumed.

Let us solve the RG flow equation (5.10.5) in the case in which we have only one marginal coupling λ , with $\gamma^{\mathcal{O}}(\lambda) = \gamma_0\lambda + \dots$, $\beta_\lambda(\lambda) = b_0\lambda^2 + \dots$. In this case, using the chain rule, we can rewrite eq.(5.10.5) as

$$\frac{dc}{d\lambda}\beta_\lambda(\lambda) = c\gamma^{\mathcal{O}}(\lambda), \quad (5.10.7)$$

that, at leading order in λ , admits the solution

$$c(\mu) = c(\mu_0) \left(\frac{\lambda(\mu)}{\lambda(\mu_0)} \right)^{\frac{\gamma_0}{b_0}}. \quad (5.10.8)$$

We see how the RG flow of the marginal coupling $\lambda(\mu)$ drives an RG flow for the irrelevant coupling c , with an effect that sensitively depends on the anomalous dimension of the operator $\mathcal{O}_c$ (determined by γ_0) and of course on the running of λ (determined by b_0), that enters both explicitly and implicitly in defining $\lambda(\mu)$ in eq.(5.10.8).

More in general, if the lowest dimensional irrelevant operator is not unique, we can have a set of operators $c_n\mathcal{O}_n$ that mix under renormalization. Repeating the analysis above, we get

$$\beta_{c_n} = \gamma_{nm}(\lambda)c_m. \quad (5.10.9)$$

In many cases eq.(5.10.9) is enough to have a good description of the physics because, by definition, irrelevant couplings are small in the IR and the lowest dimensional ones capture the main effects we want to study.²⁰

Irrelevant operators can also come from an underlying renormalizable theory when some degrees of freedom are integrated out. We will study this topic in chapter 7, where we will also consider an explicit example of RG flow of an irrelevant coupling.

The above analysis can be repeated in the case in which c is a relevant operator. We will not repeat the general analysis for relevant operators, but focus on the particular case of the mass term, which is a relevant operator. The physical mass, defined in a mass-dependent scheme as $\Gamma^{(2)}(p^2 = m^2) = 0$, cannot depend on the energy scale. But non-physical masses in mass-independent schemes do have such a dependence. This is particularly simple in a theory with no irrelevant couplings where, by dimensional analysis and to all orders in perturbation theory, we have

$$\beta_{m^2} = \gamma_m(\lambda)m^2, \quad (5.10.10)$$

²⁰ Unless selection rules require to consider higher dimensional operators. Moreover, this is only true in a perturbative context. At strong coupling several scenarios can possibly happen, e.g. a classically irrelevant operator might turn into a relevant one.

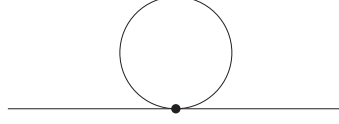


Figure 5.8: One-loop tadpole graph contributing to the renormalization of the mass in the ϕ^4 theory.

where γ_m is the anomalous dimension of the mass operator (5.9.13) computed in section 5.9. Using eq. (5.9.13), we conclude that

$$\beta_{m^2} = m^2 \frac{\lambda}{16\pi^2}. \quad (5.10.11)$$

We can also get eq.(5.10.11) directly, by-passing the computation of anomalous dimensions of composite operators. The one-loop mass correction in the ϕ^4 theory is given by the tadpole graph in fig. 5.8. In DR, the one-loop 1PI 2-point function reads

$$i\Gamma^{(2)}(p) = \frac{-i\mu^\epsilon \lambda}{2} \int \frac{d^d k}{(2\pi)^d} \frac{i}{k^2 - m^2} - i\delta_m = -\frac{i\mu^\epsilon \lambda}{2(4\pi)^{d/2}} \frac{\Gamma(2 - d/2)}{(1 - d/2)} m^{d-2} - i\delta_m. \quad (5.10.12)$$

In the $\overline{\text{MS}}$ scheme the counter-term is

$$\delta_m = \frac{m^2 \lambda}{16\pi^2 \epsilon}, \quad (5.10.13)$$

so that the tree+1-loop 1PI 2-point function reads

$$\Gamma^{(2)}(p) = p^2 - m_{\overline{\text{MS}}}^2 + \frac{m^2 \lambda}{32\pi^2} \left(1 + \log \frac{\mu^2}{m^2}\right). \quad (5.10.14)$$

In the physical scheme we simply have

$$\Gamma^{(2)}(p) = p^2 - m^2, \quad (5.10.15)$$

with no finite one-loop corrections left, where m is the physical, μ -independent, mass. Matching eqs.(5.10.14) and (5.10.15) gives

$$m^2 = m_{\overline{\text{MS}}}^2(\mu) - \frac{m^2 \lambda}{32\pi^2} \left(1 + \log \frac{\mu^2}{m^2}\right), \quad (5.10.16)$$

and hence

$$0 = \mu \frac{dm^2}{d\mu} = \beta_{m^2} - \frac{m^2 \lambda}{16\pi^2} + \mathcal{O}(\lambda^2) \implies \beta_{m^2} = m^2 \frac{\lambda}{16\pi^2}, \quad (5.10.17)$$

which reproduces eq.(5.10.11), as expected.

The use of μ -dependent masses is unavoidable when using mass-independent schemes such as $\overline{\text{MS}}$ or $\overline{\text{MS}}$. There is nothing wrong in computing amplitudes in terms of unphysical masses (like the mass terms in the $\overline{\text{MS}}$ scheme), in which case RG flow techniques should be used to relate mass terms at different energy scales. Eventually one relates these masses to the physical ones, using formulae like eq. (5.10.16). As a matter of fact, $\overline{\text{MS}}$ masses are often used in the literature, being the $\overline{\text{MS}}$ scheme one of the most popular schemes in perturbative computations. For particles like quarks that do not appear in asymptotic states and for which the direct physical mass definition is unavailable, the story is more complicated. The $\overline{\text{MS}}$ quark mass, for example, can be defined to be $m_{\overline{\text{MS}}}^2(M)$, evaluated at some indirectly derived “pole” mass M . For a heavy quark, for which $M \gg \Lambda_{QCD}$, the mass M can roughly be computed as the mass of the meson bound states $\bar{Q}Q$ divided by two. For light quarks, M can be computed using chiral perturbation theory, as explained in section 8.7. For $M \simeq \Lambda_{QCD}$, more complicated procedures are needed.

5.11 RG Improved Effective Potential

The RG technique is also useful in the context of effective potentials. The summation of LL leads to so called RG improved effective potentials. Our favorite ϕ^4 theory is particularly instructive in this case, since it shows how RG improved potentials help us in correcting fake perturbative results. We have seen in section 4.2 that the CW potential for a massless ϕ^4 theory reads (omitting the irrelevant constant term):

$$V_{eff}(\phi) = \frac{\lambda}{4!}\phi^4 + \frac{\lambda^2}{256\pi^2}\phi^4 \log \frac{\phi^2}{\phi_0^2}, \quad (5.11.1)$$

where ϕ_0 is an arbitrary fixed energy scale. Let us look for the extrema of ϕ . The equation

$$\frac{dV_{eff}(\phi)}{d\phi} = \frac{\phi^3\lambda}{384\pi^2}(64\pi^2 + 3\lambda + 6\lambda \log \lambda\phi^2/(2\phi_0^2)) = 0 \quad (5.11.2)$$

admits the three solutions

$$\phi = 0, \quad \phi = \pm\bar{\phi} \equiv \pm\phi_0 \sqrt{\frac{2}{\lambda}} e^{-(\frac{1}{4} + \frac{16\pi^2}{3\lambda})}. \quad (5.11.3)$$

The classical minimum at $\phi = 0$ turns into a maximum and two new symmetric minima arise at $\pm\bar{\phi}$. The one-loop truncation of the potential is reliable provided the tree-level term is greater than the one-loop term, namely for field values such that

$$\frac{3\lambda}{16\pi^2} \left| \log \frac{\phi}{\phi_0} \right| \ll 1. \quad (5.11.4)$$

The condition (5.11.4) is manifestly violated at the minima $|\bar{\phi}|$, so we cannot trust the result we have found for such small values of ϕ . We have here a large log problem, like

the one we found in section 5.2, though here the problem is in the IR (small values of ϕ) and not in the UV, like before. In the same way RG arguments allowed us to go from eq.(5.2.12) to eq.(5.2.13), they will allow us to improve the effective potential and solve the problem of large logs.

First of all, now that we have studied in some detail how coupling constants differ as we vary the renormalization schemes, we can compute once again the one-loop effective potential, paying more attention to the renormalization scheme chosen. Let us consider $\overline{\text{MS}}$. In this scheme eq.(4.2.14) gets modified as

$$V_{1,B}(\phi) = \mu^{4-d} \frac{1}{2} \int \frac{d^d p_E}{(2\pi)^d} \log \left(p_E^2 + M^2(\phi) \right), \quad (5.11.5)$$

where the RG scale μ has been introduced in such a way that V_1 remains of dimension 4 for any d . We can compute V_1 by taking a derivative with respect to $M^2(\phi)$, use eq.(3.5.1), and then integrate back the result over the same variable. In this way we get, modulo an irrelevant constant factor,

$$V_{1,B}(\phi) = \frac{1}{d(4\pi)^{d/2}} \mu^\epsilon (M^2(\phi))^{d/2} \Gamma(1 - d/2). \quad (5.11.6)$$

The divergence is removed in $\overline{\text{MS}}$ by adding to the potential the counterterm

$$\delta V_1(\phi) = \frac{M^4(\phi)}{32\pi^2\epsilon} \quad (5.11.7)$$

so that eventually we get

$$V_{1,\overline{\text{MS}}}(\phi) = V_{1,B}(\phi) + \delta V_1(\phi) = \frac{M^4(\phi)}{64\pi^2} \left(\log \frac{M^2(\phi)}{\mu^2} - \frac{3}{2} \right). \quad (5.11.8)$$

When $m^2 = 0$, the total tree+1-loop potential reads

$$V_{eff}(\phi) = \frac{\lambda(\mu)}{4!} \phi^4 + \frac{\lambda^2(\mu)}{256\pi^2} \phi^4 \left(\log \frac{\lambda\phi^2}{2\mu^2} - \frac{3}{2} \right), \quad (5.11.9)$$

where we omitted for simplicity the subscript $\overline{\text{MS}}$. We see that the one-loop potential consists of two terms: a non-analytic term in ϕ proportional to $\log \phi^2$ and an analytic one. The former is identical to the analogue term in eq.(5.11.1). This is expected, since this term is finite and hence has to be renormalization scheme independent. The factor $-3/2$ is instead scheme-dependent.

We already know how $\lambda(\mu)$ flows with the energy scale, but it is instructive to see how the β -function for λ can be obtained by demanding the μ -invariance of V_{eff} . Recalling that $\gamma_\phi = 0$ up to one-loop level, we get

$$0 = \mu \frac{dV_{eff}}{d\mu} = \beta(\lambda) \frac{\phi^4}{4!} - \frac{\lambda^2 \phi^4}{128\pi^2} + \mathcal{O}(\lambda^3) \rightarrow \beta(\lambda) = \frac{3\lambda^2}{16\pi^2}. \quad (5.11.10)$$

By choosing $\mu^2 = \lambda\phi^2/2$ in eq. (5.11.9), we can get rid of the log term and write

$$V_0^{\text{RG}}(\sqrt{\lambda/2}\phi) = \frac{\lambda(\sqrt{\lambda/2}\phi)}{4!}\phi^4 + \mathcal{O}(\lambda^2) = \frac{\lambda_0}{1 - \frac{3\lambda_0}{32\pi^2} \log \frac{\lambda\phi^2}{2\phi_0^2}} \frac{\phi^4}{4!} + \mathcal{O}(\lambda^2), \quad (5.11.11)$$

where $\lambda_0 = \lambda(\phi_0)$, ϕ_0 is an arbitrary scale and $\mathcal{O}(\lambda^2)$ represent the scheme-dependent term that is not log-enhanced and is then neglected. Expanding V_0^{RG} in powers of λ reproduces the one-loop potential (5.11.1) and, in addition, gives us all the LL logs of the form $\lambda\lambda^l \log^l \phi^2/\phi_0^2$. The potential (5.11.11) increases monotonically when ϕ increases and the minima at $\phi = \pm\bar{\phi}$ have disappeared. The origin of the fake result (5.11.2) should now be clear. The minima at $\phi = \pm\bar{\phi}$ were obtained by forgetting the large log's that appear for so small values of ϕ (small energies). Being the ϕ^4 theory free in the IR, the effective coupling at such small energies becomes smaller and smaller and the actual minimum is in fact the tree-level one. The potential (5.11.11) is called the RG improved tree-level effective potential, because it is obtained by replacing the coupling constant in the tree-level potential with its one-loop running value.

In general we should also take into account of field anomalous dimensions. Recall that the effective potential, modulo a term proportional to the volume of space-time, coincides with the 1PI effective action for constant field configurations ϕ and satisfy a Callan-Symanzik equation of the form²¹ (see eq.(5.6.3))

$$\left[\mu \frac{\partial}{\partial \mu} + \beta(\lambda) \frac{\partial}{\partial \lambda} - \gamma(\lambda) \phi \frac{\partial}{\partial \phi} \right] V_{\text{eff}} = 0. \quad (5.11.12)$$

The solution of eq.(5.11.12) can be read directly from eq.(5.6.16). If at a scale $\mu = \phi_0$ the classical potential is given by $V_0(\phi, \lambda(\phi_0))$, at the scale $\mu = \phi$ its tree level RG improved form reads as $V_0^{\text{RG}}(\xi(\phi)\phi, \lambda(\phi))$, where

$$\xi(\phi) = e^{-\int_{\phi_0}^{\phi} \gamma(\mu) d\mu/\mu} \quad (5.11.13)$$

takes into account of the anomalous dimension of the fields. The potential V_0^{RG} resums the LL logs in the effective potential to all orders in perturbation theory. If we want to go beyond the LL series, the scheme-dependent terms such as the one proportional to $-3/2$ in eq.(5.11.8) should be kept. Similarly to what discussed for amplitudes, the NLL series can be resummed by considering the evolution of the couplings and fields as given by the two-loop β and γ functions in the tree level potential, and take into account the RG evolution of the scheme-dependent terms up to one-loop level. More in general the N^l LL series of logarithms is obtained by evaluating the effective potential up to loop level

²¹In eq.(5.11.12) ϕ and λ represent schematically generic scalar fields and marginal couplings.

l and evolving the fields and couplings using the $l + 1$ -loop β -functions and anomalous dimensions γ . However, we should not forget that field kinetic terms also depend on the RG scale and are determined by the anomalous dimensions of the corresponding fields and the form of the effective potential depends on the particular choice of fields. We will come back to this point at the end of section 11.4, where we will study the abelian Higgs model. In particular, we will show at one-loop level that in the field basis where the scalars are canonically normalized, the ξ term (5.11.13) cancels out and the tree-level RG improved potential reads $V_0^{\text{RG}}(\phi, \lambda(\phi))$ (canonical basis).

Chapter 6

Non-Abelian Gauge Theories

Non-abelian gauge theories are at the base of our current understanding of particle physics. Both the strong and the electroweak interactions are described in terms of them. These theories are based on a generalization of the QED $U(1)$ gauge symmetry, where two transformations do not necessarily commute with each other (hence the name non-abelian).

The structure of this chapter is as follows. In section 6.1 we briefly recap a few notions of Lie algebra and group theory necessary to construct a classical theory based on non-abelian gauge fields. Quantization of gauge theories is discussed in section 6.2. In section 6.3 we introduce the BRST symmetry and discuss how the physical spectrum of non-abelian gauge theories can be understood using this symmetry. In section 6.4 we introduce the background field formalism in non-abelian gauge theories as a useful way to compute the one-loop beta-function in these theories. Finally, in section 6.5 we prove the renormalizability of non-abelian gauge theories, using BRST symmetry and the so called master equation.

6.1 Introduction and Classical Analysis

Before describing non-abelian gauge theories, let us quickly review the role of the $U(1)$ symmetry in QED. We assume an invariance of the Lagrangian under local (i.e. space-time dependent) transformations parametrized by a function $\lambda(x)$, under which any field ψ carrying charge q transforms as

$$\psi(x) \rightarrow e^{iq\lambda(x)}\psi(x). \quad (6.1.1)$$

Due to the space-time dependence of the transformation, the derivative of the field $\partial_\mu\psi$ does not transform covariantly:

$$\partial_\mu\psi(x) \rightarrow e^{iq\lambda(x)}\left(\partial_\mu\psi(x) + iq\psi(x)\partial_\mu\lambda(x)\right). \quad (6.1.2)$$

We then add a gauge field (photon) A_μ that transforms inhomogeneously:

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \lambda(x), \quad (6.1.3)$$

that allows us to define a new (covariant) derivative transforming covariantly under $U(1)$ gauge transformations:

$$D_\mu \psi(x) \equiv \partial_\mu \psi(x) - iq A_\mu(x) \psi(x) \rightarrow e^{i\lambda(x)} D_\mu \psi(x). \quad (6.1.4)$$

Under infinitesimal $U(1)$ transformations parametrized by $\epsilon(x) \ll 1$, we have

$$\delta_\epsilon \psi(x) = iq \epsilon(x) \psi(x), \quad \delta_\epsilon A_\mu(x) = \partial_\mu \epsilon(x). \quad (6.1.5)$$

The QED Lagrangian is constructed by forming gauge invariant combinations of ψ , $D_\mu \psi$ and of the field strength

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (6.1.6)$$

Non-abelian gauge theories are constructed by generalizing the above construction to a set of fields (scalars or fermions) ψ_l , with l labeling the different fields. We assume that under an infinitesimal gauge transformation, the ψ_l are rotated among each other:

$$\delta_\epsilon \psi_l(x) = i\epsilon^\alpha (t_\alpha)_l^m \psi_m(x), \quad (6.1.7)$$

where t_α are a set of constant matrices, labelled by the index α , satisfying the relation

$$[t_\alpha, t_\beta] = iC_{\alpha\beta}^\gamma t_\gamma. \quad (6.1.8)$$

The coefficients $C_{\alpha\beta}^\gamma$ are a set of real parameters denoted the structure constants. They are manifestly antisymmetric in the two lower indices: $C_{\alpha\beta}^\gamma = -C_{\beta\alpha}^\gamma$. We have written the index γ upstairs, because in principle there might be a non-trivial metric $g_{\alpha\beta}$ in group space, to raise and lower the group indices. In the physically most interesting cases, this metric can be chosen to be the identity. From now on, we will assume a trivial group metric. Correspondingly, the position of the indices α , β , etc. will be irrelevant. In this case, it can be shown that the structure constants $C_{\alpha\beta\gamma}$ become antisymmetric in all three indices.

Matrices satisfying the condition (6.1.8) form a so-called Lie algebra, namely they define infinitesimal transformations of a so called Lie group. Differently from the QED case, the group of transformations (6.1.7) is multidimensional and this explains the origin of the index α , that runs from 1 up to $\dim G$, the number of dimensions of the group. The matrices t_α are nothing else than a set of generators of the Lie group, meaning that any group transformation can be written in terms of these matrices, that are linearly

independent from each other. Like any set of matrices, the t_α 's also satisfy the so called Jacobi identity

$$[[t_\alpha, t_\beta], t_\gamma] + [[t_\gamma, t_\alpha], t_\beta] + [[t_\beta, t_\gamma], t_\alpha] = 0. \quad (6.1.9)$$

Using eq. (6.1.8), this identity implies the following constraints among the structure constants:

$$C_{\alpha\beta}^\omega C_{\omega\gamma}^\delta + C_{\gamma\alpha}^\omega C_{\omega\beta}^\delta + C_{\beta\gamma}^\omega C_{\omega\alpha}^\delta = 0. \quad (6.1.10)$$

Depending on how many fields ψ_l we have, the matrices t_α are said to be in different representations of the Lie algebra. In general, there is an infinite set of matrices (of different size), satisfying eqs. (6.1.8) and (6.1.9). Among these, a special role is played by the “adjoint” representation. This is the representation in which the generators t_α^{Adj} are $\dim G \times \dim G$ matrices, i.e. the indices l, m coincide with the indices α, β . An explicit form of this representation is given by

$$(t_\alpha^{Adj})_\gamma^\beta = iC_{\alpha\gamma}^\beta. \quad (6.1.11)$$

Indeed, it is straightforward to check that the Jacobi identity (6.1.10) can be rewritten as

$$[t_\alpha^{Adj}, t_\beta^{Adj}] = iC_{\alpha\beta}^\gamma t_\gamma^{Adj}. \quad (6.1.12)$$

We do not enter here in any detail concerning the definition of Lie algebras, Lie groups, etc., because all this will be extensively treated in the group theory course. However, we need to introduce another relevant representation, the fundamental. It might be useful to consider a simple example of non-abelian Lie group, $G = SU(2)$. $SU(2)$ is defined by the set of 2×2 unitary matrices $U^\dagger = U^{-1}$ with unit determinant (that's why the name S(for special, with unit determinant)U(for unitary)(2)). It is straightforward to check that this set is in fact a group and it is three dimensional. Any $SU(2)$ matrix U can be written as

$$U = e^{i\omega_\alpha t_\alpha}, \quad \alpha = 1, 2, 3, \quad (6.1.13)$$

where $t_\alpha = \sigma_\alpha/2$ and σ_α are the usual Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (6.1.14)$$

The structure constants are

$$C_{\alpha\beta\gamma} = \epsilon_{\alpha\beta\gamma}, \quad (6.1.15)$$

where $\epsilon_{\alpha\beta\gamma}$ is the completely antisymmetric tensor, with $\epsilon_{123} = +1$. The above 2×2 matrices t_α , that enter the definition of the group $SU(2)$, form the fundamental representation.

The 3×3 matrices (6.1.11) define instead the adjoint representation. We can then have field “doublets” D_l ($l = 1, 2$), that transforms as $D_l \rightarrow U_{lm} D_m$, with U as in eq. (6.1.13) or field “triplets” T_l ($l = 1, 2, 3$), transforming as $T_l \rightarrow (U^{Adj})_{lm} D_m$, where

$$U^{Adj} = e^{i\omega_\alpha t_\alpha^{Adj}}, \quad \alpha = 1, 2, 3. \quad (6.1.16)$$

In addition to the fundamental and adjoint representations, there are an infinite number of other irreducible representations, labelled by the angular momentum J , of dimension $2J + 1$, with J any positive integer or semi-integer number.

Similar considerations apply for more general groups, such as $SU(N)$, with $N > 2$, or $SO(N)$, with $N > 3$.¹ Instead of defining $\dim r \times \dim r$ matrices U^r , for each different representation r , it is practically more convenient to write the transformation properties of r in terms of, say, the matrices U in the fundamental representation. All we need to know is how the representation r in question appears in the tensor product of fundamentals. For instance, for $SU(N)$ groups, we have $\mathbf{N} \otimes \bar{\mathbf{N}} = \mathbf{N}^2 - \mathbf{1} \oplus \mathbf{1}$, where $\mathbf{N}$ and $\bar{\mathbf{N}}$ are the fundamental and anti-fundamental (i.e. complex conjugate) representations, and $\mathbf{N}^2 - \mathbf{1}$ is the adjoint one. A field ψ in the adjoint representation of $SU(N)$ can correspondingly be written as an $N \times N$ matrix field ψ_{ij} , transforming as

$$\psi \rightarrow U\psi U^\dagger. \quad (6.1.17)$$

Coming back to physics, given the transformation (6.1.7), we add gauge fields A_μ^α , one for each independent direction in field space, so that we can form a covariant derivative. In analogy to the $U(1)$ case, the transformations of the gauge fields A_μ^α must contain a term of the form $\partial_\mu \epsilon^\alpha$. Contrary to the $U(1)$ case, this cannot be the end of the story, since α is an index in the adjoint representation. The natural guess for the infinitesimal transformation of A_μ^α is then

$$\delta A_\mu^\alpha = \partial_\mu \epsilon^\alpha + i\epsilon^\beta (t_\beta^{Adj})_\gamma^\alpha A_\mu^\gamma = \partial_\mu \epsilon^\alpha + C_{\beta\gamma}^\alpha A_\mu^\beta \epsilon^\gamma. \quad (6.1.18)$$

The covariant derivative is defined as

$$D_\mu \psi_l \equiv \partial_\mu \psi_l - iA_\mu^\alpha (t_\alpha)_l^m \psi_m. \quad (6.1.19)$$

It is straightforward to show that this guess is in fact correct and the covariant derivative transforms as it should:

$$\delta(D_\mu \psi)_l = i\epsilon^\alpha (t_\alpha)_l^m (D_\mu \psi)_m. \quad (6.1.20)$$

¹The group $SO(3)$ is locally isomorphic to $SU(2)$ and is defined by the same 3 generators t_α^{Adj} introduced above for $SU(2)$. The group $SO(2)$ is isomorphic to $U(1)$.

It is often convenient to write the components of the gauge fields A_μ^α in matrix form by defining

$$A_\mu = A_\mu^\alpha t_\alpha. \quad (6.1.21)$$

For simplicity of notation we have omitted, and from now on will be done most of the time, the gauge group indices in eq. (6.1.21). The finite form of eq. (6.1.18) is easily found by demanding that the gauge transformed connection $A \rightarrow A^U$, is such that

$$D_\mu(A)\psi \rightarrow U D_\mu(A)\psi. \quad (6.1.22)$$

for any field in a given representation r , transforming as $\psi \rightarrow \psi^U = U\psi$, where U is defined as

$$U(x) = e^{i\Lambda_\alpha(x)t_\alpha}. \quad (6.1.23)$$

We have

$$(\partial_\mu - iA_\mu)\psi \rightarrow \partial_\mu(U\psi) - iA_\mu^U U\psi = U(\partial_\mu\psi - iA_\mu\psi) + (\partial_\mu U)\psi + (iUA_\mu - iA_\mu^U U)\psi. \quad (6.1.24)$$

Demanding that the last three terms in eq. (6.1.22) vanish uniquely fixes

$$A_\mu^U = UA_\mu U^{-1} - i(\partial_\mu U)U^{-1}. \quad (6.1.25)$$

The first term in eq. (6.1.25) is the one expected from a field in the adjoint representation (see eq. (6.1.17)), while the second is the inhomogeneous one characterizing a gauge connection. For the $U(1)$ case the index α is trivial and eq. (6.1.25) reduces to eq. (6.1.3), with $\Lambda = \lambda$. For infinitesimal transformations, $\Lambda_\alpha = \epsilon_\alpha$, the transformation (6.1.25) correctly reproduces eq. (6.1.18), as it should.

The generalization of the $U(1)$ field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ can be found by recalling that for any field ψ with charge q , the field strength is proportional to the (commutator part of the) action of two covariant derivatives acting on the field itself. Taking $q = 1$, we have

$$D_\mu D_\nu \psi = (\partial_\mu - iA_\mu)(\partial_\nu - iA_\nu)\psi = (\partial_\mu \partial_\nu - i\partial_\mu A_\nu - iA_\nu \partial_\mu - iA_\mu \partial_\nu - A_\mu A_\nu)\psi. \quad (6.1.26)$$

Only the second term in the above equation survives when we take the antisymmetric combination in $\mu \leftrightarrow \nu$:

$$[D_\mu, D_\nu]\psi = -iF_{\mu\nu}\psi. \quad (6.1.27)$$

The field strength in a non-abelian gauge theory can be defined by generalizing eq. (6.1.27). Denoting ψ a field in an arbitrary representation of the gauge group, eq. (6.1.26) still applies, but now the last term does not vanish, since A_μ is a matrix. We have

$$[D_\mu, D_\nu]\psi = -i(\partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu])\psi \equiv -iF_{\mu\nu}\psi. \quad (6.1.28)$$

In components, $F_{\mu\nu} = F_{\mu\nu}^\alpha t_\alpha$, with

$$F_{\mu\nu}^\alpha = \partial_\mu A_\nu^\alpha - \partial_\nu A_\mu^\alpha + C_{\alpha\beta\gamma} A_\mu^\beta A_\nu^\gamma. \quad (6.1.29)$$

Contrary to the abelian case, the field strength $F_{\mu\nu}$ is not gauge invariant. Its transformation properties can easily be found by considering the gauge transform of eq. (6.1.28):

$$[D_\mu, D_\nu]\psi \rightarrow U[D_\mu, D_\nu]\psi = -iUF_{\mu\nu}\psi = -iF_{\mu\nu}^U\psi^U, \quad (6.1.30)$$

from which we immediately get

$$F_{\mu\nu}^U = UF_{\mu\nu}U^{-1}. \quad (6.1.31)$$

In non-abelian gauge theories the gauge field strength transforms in the adjoint representation of the gauge group. The most general gauge-invariant Lagrangian can be written as a Lorentz invariant functional of matter fields and their covariant derivatives, and of the field strengths $F_{\mu\nu}$ and their covariant derivatives. At the level of dimension 4 (or less) operators, we have

$$\mathcal{L} = \mathcal{L}_{YM} + \mathcal{L}_{Matter} \quad (6.1.32)$$

where

$$\mathcal{L}_{YM} = -\frac{1}{2g^2} \text{Tr } F_{\mu\nu} F^{\mu\nu}, \quad (6.1.33)$$

$$\mathcal{L}_{Matter} = \sum_i \bar{\psi}_i (i\not{D} - m_i) \psi_i + \sum_j |D_\mu \phi_j|^2 - V(\psi_i, \phi_j). \quad (6.1.34)$$

In eq. (6.1.33), we have introduced a dimensionless parameter g that is identified as the gauge coupling constant of the non-abelian theory. By rescaling the gauge fields as

$$A_\mu^\alpha \rightarrow gA_\mu^\alpha \quad (6.1.35)$$

we get canonical kinetic terms, having normalized the generators as

$$\text{Tr } t^\alpha t^\beta = \frac{\delta^{\alpha\beta}}{2} \quad (6.1.36)$$

in the fundamental representation. The coupling g now appears in all covariant derivatives in the matter sector and in the self-coupling of the gauge fields in the Yang-Mills (YM) Lagrangian. Gauge invariance requires that the same coupling g governs all these interactions. In eq. (6.1.34) i and j run over all fermions and scalars in the theory, D_μ are the covariant derivatives in the appropriate representations and V encodes the scalar potential of the scalar fields ϕ_j and their Yukawa interactions with the fermions ψ_i . Gauge invariance requires that $\delta_\epsilon V = 0$.

The carriers of the force associated to the non-abelian group, that we generally denote by “gluons”, are themselves subject to the force they carry. The equations of motion (e.o.m.) for A_μ^α deduced from the Lagrangian (6.1.32) are

$$\partial_\mu F_\alpha^{\mu\nu} = -g C_{\alpha\beta\gamma} A_\mu^\beta F_\gamma^{\mu\nu} - \frac{\delta \mathcal{L}_{Matter}}{\delta A_\nu^\alpha} \equiv -\mathcal{J}_\alpha^\nu \quad (6.1.37)$$

where $\mathcal{J}_\nu^\alpha$ are the $\dim G$ conserved currents associated to the symmetry group G :

$$\partial_\mu \mathcal{J}_\alpha^\mu = 0. \quad (6.1.38)$$

The e.o.m. (6.1.37), written in terms of $\mathcal{J}_\mu^\alpha$ are not covariant. It is convenient to shift the gluon contribution to the current, the first term in the second relation in eq. (6.1.37), to the left-hand side of that equation. In doing so, the e.o.m. read

$$D_\nu F_\alpha^{\mu\nu} = -J_\alpha^\mu, \quad (6.1.39)$$

where

$$\begin{aligned} D_\rho F_\alpha^{\mu\nu} &= \partial_\rho F_\alpha^{\mu\nu} - ig(t_\beta^{Adj})_{\alpha\gamma} A_\rho^\beta F_\gamma^{\mu\nu}, \\ J_\alpha^\mu &= \frac{\delta \mathcal{L}_{Matter}}{\delta A_\mu^\alpha} = \frac{\delta \mathcal{L}_{Matter}}{\delta D_\mu \Psi_I} (-igt_\alpha \Psi_I). \end{aligned} \quad (6.1.40)$$

In eq. (6.1.40) we have rewritten the form of the current J_μ^α to make explicit its covariant properties, in contrast to the conserved current $\mathcal{J}_\mu^\alpha$. The field Ψ encodes both fermions and scalars and $I = (i, j)$. The current J_μ^α is covariantly conserved, namely $D_\mu J_\alpha^\mu = 0$. Indeed,

$$D_\mu J_\alpha^\mu = -D_\mu D_\nu F_\alpha^{\mu\nu} = -[D_\mu, D_\nu] F_\alpha^{\mu\nu} = i F_{\mu\nu}^\beta (t_\beta^{Adj})_{\alpha\gamma} F_\gamma^{\mu\nu} = C_{\alpha\beta\gamma} F_{\mu\nu}^\beta F_\gamma^{\mu\nu} = 0. \quad (6.1.41)$$

The field strength $F_{\mu\nu}$ satisfies another important relation, called Bianchi identity. It is a consistency relation, and can be derived starting from the Jacobi identity (6.1.9), applied to covariant derivatives:

$$[D_\mu, [D_\nu, D_\rho]] + [D_\rho, [D_\mu, D_\nu]] + [D_\nu, [D_\rho, D_\mu]] = 0. \quad (6.1.42)$$

Using eq. (6.1.27), we can rewrite the above expression as

$$D_\mu F_{\nu\rho} + D_\rho F_{\mu\nu} + D_\nu F_{\rho\mu} = 0. \quad (6.1.43)$$

It is a straightforward exercise to show that eq. (6.1.43) is identically satisfied.

6.2 Quantum Treatment: the Faddeev-Popov Method

The quantization of gauge theories is non-trivial. The essential point is that Lorentz invariance forces us to describe helicity-one fields in terms of a four-vector field, but the latter has four components, and hence more degrees of freedom than necessary. The time-component of a vector field, in addition, is problematic, because it would lead to a kinetic term with an opposite sign with respect to its spatial components. Non-physical degrees of freedom are then expected. The way in which gauge theories solve the problem is to introduce a redundancy in the theory, gauge invariance, so that we can eliminate these extra unwanted and unphysical degrees of freedom.

Functional methods based on the path integral are by far the best way to quantize gauge theories [24]. Before considering non-abelian theories, it is very useful to recall how QED can be quantized using the path integral. Let's first see why a naive path integral quantization cannot work, by computing the photon propagator

$$\langle A_\mu(x) A_\nu(y) \rangle = \mathcal{N} \int \mathcal{D}A_\mu A_\mu(x) A_\nu(y) e^{iS(A)}, \quad (6.2.1)$$

where $\mathcal{N}$ is a normalization constant and

$$S(A) = -\frac{1}{4} \int d^4x F_{\mu\nu}^2 = \frac{1}{2} \int \frac{d^4p}{(2\pi)^4} A^\mu(-p) D_{\mu\nu}(p) A^\nu(p) \quad (6.2.2)$$

is the usual free action, written both in configuration and momentum space. The tensor $D_{\mu\nu}$ equals

$$D_{\mu\nu}(p) = (-p^2 \eta_{\mu\nu} + p_\mu p_\nu). \quad (6.2.3)$$

The photon propagator is equal to the Fourier transform of the inverse of $D_{\mu\nu}(p)$:

$$\langle A_\mu(x) A_\nu(y) \rangle =? \int \frac{d^4p}{(2\pi)^4} i D_{\mu\nu}^{-1}(p) e^{ip \cdot (x-y)}. \quad (6.2.4)$$

However, $\det D_{\mu\nu}(p) = 0$, no inverse exists and no propagator can be defined. Another way of looking at the problem is obtained by performing a shift of variables in the path integral: $A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \lambda(x)$. We get

$$\langle A_\mu(x) A_\nu(y) \rangle = \langle A_\mu(x) A_\nu(y) \rangle + (\partial_\mu \lambda \partial_\nu \lambda) \mathcal{N} \int \mathcal{D}A_\mu e^{iS(A)}, \quad (6.2.5)$$

whose only solution is $\langle A_\mu(x) A_\nu(y) \rangle = \infty$. The problem arises from the fact that the action is gauge invariant and we are integrating over all possible field configurations, including those that are related by a gauge transformation. All pure gauge configurations, such as $A_\mu(x) = \partial_\mu \lambda$, are not damped by the action and lead to the above divergence. The

problem is solved by restricting the integration to gauge inequivalent configurations only. In other words, we have to implement a gauge-fixing condition of the form $G(A) = 0$, where $G(A)$ is a functional of the gauge field A_μ . This can be imposed inside the path integral by means of a functional generalization of the Dirac delta-function. Care should be paid on possible Jacobian factors. These can arise by recalling the formula

$$\delta[f(x)] = \sum_i \left| \frac{df}{dx} \right|_{x=x_0^{(i)}}^{-1} \delta(x - x_0^{(i)}), \quad (6.2.6)$$

where $x_0^{(i)}$ are the values where the function f vanishes. The n -dimensional integral generalization of eq. (6.2.6) is

$$1 = \int \prod_i dx_i \delta^{(n)}(f_j(x_i)) \left| \det \frac{\partial f_j}{\partial x_i} \right| \quad (6.2.7)$$

where f_j are n functions of the n variables x_i and we have assumed that they all vanish at a single point $x_{0,i}$. The further infinite dimensional generalization of eq. (6.2.7) is

$$1 = \int \mathcal{D}\lambda \delta(G(A^\lambda)) \left| \det \frac{\delta G(A^\lambda)}{\delta \lambda(x)} \right|, \quad (6.2.8)$$

where $A_\mu^\lambda(x) = A_\mu(x) + \partial_\mu \lambda(x)$ and $G(A^\lambda)$ is an arbitrary functional, assumed to vanish only for a specific single function λ_0 . A simple choice for $G(A)$ is

$$G(A) = \partial_\mu A^\mu \implies G(A^\lambda) = \partial_\mu A^\mu + \square \lambda. \quad (6.2.9)$$

Inserting eq. (6.2.8) inside the path integral gives

$$\begin{aligned} \mathcal{N} \int \mathcal{D}A_\mu e^{iS(A)} \int \mathcal{D}\lambda \delta(G(A^\lambda)) \left| \det \frac{\delta G(A^\lambda)}{\delta \lambda(x)} \right| &= \mathcal{N} |\det \square| \int \mathcal{D}A_\mu e^{iS(A)} \int \mathcal{D}\lambda \delta(G(A^\lambda)) \\ &= \mathcal{N}' \int \mathcal{D}\lambda \int \mathcal{D}A_\mu e^{iS(A)} \delta(G(A)) = \mathcal{N}'' \int \mathcal{D}A_\mu e^{iS(A)} \delta(G(A)), \end{aligned} \quad (6.2.10)$$

where we have included into the normalization constants the gauge-field independent factor $\det \square$ and the integration over the gauge parameter λ . The functional delta into the last term of eq. (6.2.10) avoids to integrate over redundant field configurations. With a simple trick, we can also get rid of the functional delta and yet have a well-defined path integral. Instead of taking a single gauge fixing like the one in eq. (6.2.9) we can introduce a family of gauge fixing terms, parametrized by an arbitrary function f :

$$G_f(A) = \partial_\mu A^\mu - f. \quad (6.2.11)$$

Since no physical observable can depend on the gauge fixing, we can average over the different gauge fixings by introducing a phase factor

$$\exp \left(- \frac{i}{2\xi} \int d^4x f^2(x) \right) \quad (6.2.12)$$

and integrating over $f(x)$. This is useful, since in so doing we can get rid in one step of the functional delta and of the integration over $f(x)$. In eq. (6.2.12) ξ is a positive parameter. The path integral becomes now

$$\mathcal{N} \int \mathcal{D}A_\mu e^{iS(A) - \frac{i}{2\xi} \int d^4x (\partial_\mu A^\mu)^2}. \quad (6.2.13)$$

The final outcome of these manipulations is the addition of a new gauge-variant term in the action, called gauge-fixing term. The latter is crucial to make sense of the photon propagator. Equation (6.2.1) is replaced by

$$\langle A_\mu(x) A_\nu(y) \rangle = \mathcal{N} \int \mathcal{D}A_\mu A_\mu(x) A_\nu(y) e^{iS(A) - \frac{i}{2\xi} \int d^4x (\partial_\mu A^\mu)^2}. \quad (6.2.14)$$

The tensor (6.2.3) becomes

$$D_{\mu\nu}(p) = -p^2 \eta_{\mu\nu} + p_\mu p_\nu \frac{(\xi - 1)}{\xi}, \quad (6.2.15)$$

and it admits the inverse

$$D_{\mu\nu}^{-1}(p) = -\frac{1}{p^2} \left(\eta_{\mu\nu} - (1 - \xi) \frac{p_\mu p_\nu}{p^2} \right). \quad (6.2.16)$$

Reintroducing the $i\epsilon$, we finally get

$$\langle A_\mu(x) A_\nu(y) \rangle = \int \frac{d^4p}{(2\pi)^4} \frac{-i}{p^2 + i\epsilon} \left(\eta_{\mu\nu} - (1 - \xi) \frac{p_\mu p_\nu}{p^2} \right) e^{ip \cdot (x-y)}. \quad (6.2.17)$$

The photon propagator is not a direct physical observable and depends on ξ . Common choices for ξ in explicit computations are $\xi = 0$ (Landau gauge), in which case the tree-level propagator is transverse and $\xi = 1$ (Feynman gauge), in which this simplifies considerably.

The quantization of non-abelian gauge theories proceeds along the same way as the abelian case, but it presents additional complications. The naive measure is

$$\mathcal{N} \int \mathcal{D}A_\mu^\alpha e^{iS(A)}. \quad (6.2.18)$$

It is gauge-invariant, but in a less trivial way than in abelian theories, since a gauge transformation rotates the fields. The Jacobian associated to the infinitesimal transformation (6.1.18) is

$$\text{Jac}_{\mu\nu}^{\alpha\beta}(x, y) = \frac{\delta A_\mu'^\alpha(x)}{\delta A_\nu^\beta(y)} = \delta_\mu^\nu \delta(x - y) (\delta_{\alpha\beta} + \epsilon^\gamma C_{\alpha\beta\gamma}). \quad (6.2.19)$$

Since $\text{Det}(1 + \epsilon) = 1 + \text{Tr } \epsilon + \mathcal{O}(\epsilon^2)$, we have

$$\det \text{Jac}_{\mu\nu}^{\alpha\beta}(x, y) = \delta_\mu^\mu \delta(0) (\delta_{\alpha\alpha} + \epsilon^\gamma C_{\alpha\alpha\gamma}) = \delta_\mu^\mu \delta(0) \delta_{\alpha\alpha} \quad (6.2.20)$$

that is the infinite dimensional generalization of the unit matrix. We then conclude that the measure $\mathcal{D}A_\mu^\alpha$ is gauge invariant. Let us now proceed like in the abelian case, introducing a delta-functional gauge-fixing in the path integral, like in eq. (6.2.8). Let us define

$$\Delta_G^{-1}(A) = \int \mathcal{D}U \delta\left(G(A^U(x))\right). \quad (6.2.21)$$

In eq. (6.2.21), $\mathcal{D}U$ is the so-called invariant measure of the group G , parametrized by group elements $U(\Lambda)$. It is not, like in the abelian case, simply the integration over the Lie algebra generators, $\prod_\alpha \mathcal{D}\Lambda^\alpha(x)$, but it includes a non trivial measure $\rho(\Lambda)$. It is called invariant measure because it satisfies the following properties:

$$\int \mathcal{D}U f(U) = \int \mathcal{D}U f(U^{-1}) = \int \mathcal{D}U f(U \cdot U_0) = \int \mathcal{D}U f(U_0 \cdot U) \quad (6.2.22)$$

where U_0 is a constant element of the group and $f(U)$ is an arbitrary function over the group.

** Although we will never need its explicit form in the following, it is worth to spend a few more words on $\rho(\Lambda)$. This can be defined starting from the metric in group space $g_{\alpha\beta}$ as

$$g_{\alpha\beta} = \text{Tr} \left(U^{-1}(\Lambda) \left(\frac{\partial U(\Lambda)}{\partial \Lambda_\alpha} \right) U^{-1}(\Lambda) \left(\frac{\partial U(\Lambda)}{\partial \Lambda_\beta} \right) \right). \quad (6.2.23)$$

The invariant measure is

$$\rho(\Lambda) = \sqrt{\det g_{\alpha\beta}}. \quad (6.2.24)$$

The parameters Λ_α in eq. (6.2.23) are space-time independent coordinates spanning the group G . They do not necessarily correspond to the x -independent version of the generators appearing in eq. (6.1.23). An example will clarify this point. The group $SU(2)$ is defined as the set of 2×2 unitary matrices U of unit determinant. Any matrix U can be written as

$$U = \begin{pmatrix} z_1 & z_2 \\ -z_2^* & z_1^* \end{pmatrix}, \quad (6.2.25)$$

with $|z_1|^2 + |z_2|^2 = 1$, $z_{1,2} \in \mathbb{C}$. The group $SU(2)$ is isomorphic to the three-dimensional sphere S^3 . Instead of using standard coordinates subject to a constraint, we might use radial coordinates. Denoting $z_i = x_i + iy_i$, $i = 1, 2$, we have

$$\begin{aligned} x_1 &= \sin \psi \sin \theta \cos \phi, & y_1 &= \sin \psi \sin \theta \sin \phi \\ x_2 &= \sin \psi \cos \theta, & y_2 &= \cos \psi, \end{aligned} \quad (6.2.26)$$

with $0 \leq \phi < 2\pi$, $0 \leq \theta < 2\pi$, $0 \leq \psi \leq \pi$. The invariant $SU(2)$ measure coincides with

the standard metric of S^3 : $\rho(\psi, \theta, \phi) = \sin^2 \psi \sin \theta$.² Given the explicit form of the metric (6.2.23) it is straightforward to prove the relations (6.2.22).

The functional $\Delta_G(A)$ is gauge-invariant: $\Delta_G(A^U) = \Delta_G(A)$, that immediately follows from eq. (6.2.22). The naive measure (6.2.18) can be rewritten as (omitting Lorentz and color indices in the gauge measure)

$$\mathcal{N} \int \mathcal{D}A \int \mathcal{D}U e^{iS(A)} \Delta_G(A) \delta(G(A^U)). \quad (6.2.27)$$

We can change variables in the path integral by defining $A_\mu = A'^U_{\mu}$. Since $\Delta_G(A)$, the measure and the action are all gauge invariant, we get (redefine $A' \rightarrow A$)

$$\mathcal{N} \int \mathcal{D}A \int \mathcal{D}U e^{iS(A)} \Delta_G(A) \delta(G(A)) = \mathcal{N}' \int \mathcal{D}A e^{iS(A)} \Delta_G(A) \delta(G(A)), \quad (6.2.28)$$

and we can reabsorb the invariant group measure into the overall path integral normalization $\mathcal{N}$. Let us verify that any correlation function of gauge invariant operators does not depend on the gauge-fixing, namely does not depend on the specific choice of the functional $G(A)$. If $O(A)$ schematically represents some product of gauge invariant operators, given two arbitrary gauge fixing functionals G and F , one has

$$\begin{aligned} \langle O(A) \rangle_G &= \mathcal{N} \int \mathcal{D}A e^{iS(A)} \Delta_G(A) \delta(G(A)) O(A) \\ &= \mathcal{N} \int \mathcal{D}A e^{iS(A)} \Delta_G(A) \delta(G(A)) \int \mathcal{D}U \delta(F(A^U)) \Delta_F(A) O(A) \\ &= \mathcal{N} \int \mathcal{D}A \int \mathcal{D}U e^{iS(A)} \Delta_G(A) \delta(G(A^{U^{-1}})) \delta(F(A)) \Delta_F(A) O(A) \\ &= \mathcal{N} \int \mathcal{D}A \int \mathcal{D}U e^{iS(A)} \Delta_G(A) \delta(G(A^U)) \delta(F(A)) \Delta_F(A) O(A) \\ &= \mathcal{N} \int \mathcal{D}A e^{iS(A)} \delta(F(A)) \Delta_F(A) O(A) = \langle O(A) \rangle_F. \end{aligned} \quad (6.2.29)$$

On the contrary, the correlation functions of gauge dependent quantities do depend on the choice of $G(A)$. Similarly to the abelian case, we can also take a family of gauge fixing functionals and integrate over them. The typical choice will be the non-abelian generalization of eq. (6.2.11):

$$G_f(A) = \partial_\mu A^\mu_\alpha - f_\alpha, \quad (6.2.30)$$

weighted by the phase factor

$$\exp \left(- \frac{i}{2\xi} \int d^4x f_\alpha^2(x) \right). \quad (6.2.31)$$

²Notice the similarity of the definition (6.2.24) with the definition of diffeomorphism-invariant measure in general relativity.

In this way, we get

$$\langle O(A) \rangle = \mathcal{N} \int \mathcal{D}A e^{iS(A) - \frac{i}{2\xi} \int d^4x (\partial_\mu A_\alpha^\mu)^2} \Delta(A) O(A), \quad (6.2.32)$$

where

$$\Delta^{-1}(A) = \int \mathcal{D}U \delta\left(\partial_\mu(A^U)_\alpha^\mu - \partial_\mu A_\alpha^\mu\right). \quad (6.2.33)$$

The integrand in eq. (6.2.33) has a non-trivial support for values of U infinitesimally close to the identity: $U = 1 + i\epsilon^\alpha t_\alpha$. We have

$$\delta\left(\partial_\mu(A^U)_\alpha^\mu - \partial_\mu A_\alpha^\mu\right) = \delta\left(\partial_\mu D^\mu \epsilon_\alpha\right) = \frac{\delta(\epsilon_\alpha)}{|\det \partial_\mu D^\mu|}. \quad (6.2.34)$$

Modulo irrelevant constants, as usual absorbable in the path integral normalization $\mathcal{N}$, we get

$$\Delta(A) = |\det \partial_\mu D^\mu|. \quad (6.2.35)$$

This is nothing else than the non-abelian generalization of the $|\det \square|$ term appearing in eq. (6.2.10). The crucial difference with the abelian case is its gauge field dependence by means of the covariant derivative. For this reason it should be kept inside the path integral. Loosely speaking, in perturbation theory

$$|\det \partial_\mu D^\mu| = |\det \square| |\det (1 + \mathcal{O}(g))|. \quad (6.2.36)$$

The first factor is irrelevant, while the second one is manifestly positive in perturbation theory. We can then remove the absolute value, which would complicate the analysis, and write

$$\Delta(A) = \det \partial_\mu D^\mu. \quad (6.2.37)$$

It is not easy to directly compute determinants. It is more convenient to turn a determinant into a local action by inserting additional non-physical degrees of freedom and use the identity, valid for Grassmann variables ω and ω^* ,³

$$\int \mathcal{D}\omega \mathcal{D}\omega^* e^{-i \int d^4x \omega^*(x) F(x) \omega(x)} \propto \det F, \quad (6.2.38)$$

where $F(x)$ is an arbitrary differential operator. Using eq. (6.2.38), we can write, modulo irrelevant constants,

$$\det \partial_\mu D^\mu = \int \mathcal{D}\omega_\alpha \mathcal{D}\omega_\alpha^* e^{iS_{ghost}}, \quad (6.2.39)$$

³Recall that inside the path integral ω and ω^* are two independent variables. In particular, ω^* is not the complex conjugate of ω .

where

$$S_{ghost} = \int d^4x \partial_\mu \omega_\alpha^* D^\mu \omega_\alpha = \int d^4x \partial_\mu \omega_\alpha^* \left(\partial^\mu \omega_\alpha + g C_{\alpha\beta\gamma} A_\beta^\mu \omega_\gamma \right). \quad (6.2.40)$$

ω_α and ω_α^* are dim-G scalar fields with fermion statistic, transforming in the adjoint representation of G . They are not associated to physical propagating particles and for this reason they are denoted ghost fields. They cannot appear as physical external states, but they can and do appear in loops as virtual particles, by means of their interaction with the gauge fields. Ghosts are crucial to restore unitarity in non-abelian gauge theories. Loosely speaking, they compensate for the (also unphysical) contribution of the longitudinal and time component of the gauge fields that, contrary to the abelian case, do not automatically decouple from scattering amplitudes.

Putting all together, the complete non-abelian Lagrangian density at the quantum level is the sum of three terms:

$$\mathcal{L}_{tot} = \mathcal{L} + \mathcal{L}_{g.f.} + \mathcal{L}_{ghost}, \quad (6.2.41)$$

where

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^\alpha F_\alpha^{\mu\nu} + \sum_i \bar{\psi}_i (\not{D} - m_i) \psi_i + \sum_j |D_\mu \phi_j|^2 - V(\psi_i, \phi_j), \quad (6.2.42)$$

$$\mathcal{L}_{g.f.} = -\frac{1}{2\xi} (\partial_\mu A_\alpha^\mu)^2, \quad (6.2.43)$$

$$\mathcal{L}_{ghost} = \partial_\mu \omega_\alpha^* D^\mu \omega_\alpha. \quad (6.2.44)$$

The propagator of gauge and ghost field is readily found by the quadratic term of the above Lagrangian density. We have

$$\begin{aligned} \langle A_\mu^\alpha(x) A_\nu^\beta(y) \rangle &= \delta^{\alpha\beta} \int \frac{d^4p}{(2\pi)^4} \frac{-i}{p^2 + i\epsilon} \left(\eta_{\mu\nu} - (1 - \xi) \frac{p_\mu p_\nu}{p^2} \right) e^{ip \cdot (x-y)}, \\ \langle \omega_\alpha^*(x) \omega_\beta(y) \rangle &= \delta_{\alpha\beta} \int \frac{d^4p}{(2\pi)^4} \frac{i}{p^2 + i\epsilon} e^{ip \cdot (x-y)}. \end{aligned} \quad (6.2.45)$$

The gauge field propagator is the trivial generalization of the photon propagator (6.2.16), while the ghost propagator coincides with that of a complex massless scalar field. Recall that the ghost Lagrangian (6.2.44) depends on the specific form of $\Delta(A)$, which in turn depends on the specific gauge fixing chosen. It is invariant under a $U(1)$ symmetry (ghost number) under which the ghost fields ω_α and ω_α^* have respectively charges +1 and -1. For this reason the latter fields are commonly denoted anti-ghosts and this explains the notation ω_α^* . All the operators appearing in the Lagrangian (6.2.41) have dimensions less or equal to four, like in a renormalizable theory. However, not all possible operators of

dimensions $\Delta \leq 4$ appear in $\mathcal{L}_{tot}$ and hence the renormalizability of these theories is not obvious. The appearance of unphysical fields in the theory complicates also the notion of physical field in non-abelian gauge theories. All these issues are best addressed by introducing the BRST symmetry, subject of the next section.

Before going on we would like to comment on the assumption made below eq.(6.2.8), that was implicit in writing eq.(6.2.21), about the existence, at a given $A_\mu(x)$, of only one gauge element $U_0(x)$ where the functional $G(A^U)$ vanishes. This is equivalent to say that the equation

$$\partial^\mu \left(U^\dagger (\partial_\mu U - i[A_\mu, U]) \right) = 0, \quad (6.2.46)$$

has either no solutions or such solutions should be discarded in the path integral. In the abelian case, eq.(6.2.46) boils down to $\square\lambda = 0$ that clearly admits plane wave solutions. These can be ruled out by demanding proper boundary conditions for A_μ in the path integral. This is particularly clear in the euclidean case, where we demand A_μ to vanish at infinity. The conditions $\lambda \rightarrow 0$ at infinity and $\square\lambda = 0$ in euclidean space imply $\lambda = 0$. In contrast, in the non-abelian case there exist in general non-trivial solutions of the above equation which would force us to consider other gauge field configurations A^U , called Gribov copies [30]. Their possible presence is important in a non-perturbative treatment of gauge theories, but will not affect the perturbative expansion around the free theory (like the absolute value in eq.(6.2.36)). The above assumption can then be justified only in a perturbative set-up.

6.3 BRST Symmetry

The Faddeev-Popov path integral quantization of gauge theories reviewed above requires to fix a gauge and hides the underlying gauge invariance of the theory.⁴ In other words, the Lagrangian $\mathcal{L}_{tot}$ cannot obviously be gauge invariant. On the other hand, it was found by Becchi, Rouet and Stora [7, 8], and independently by Tyutin [58], that $\mathcal{L}_{tot}$ is invariant under an additional symmetry, called BRST symmetry. It is useful to rewrite the gauge fixing term in a different fashion by “integrating in” an auxiliary field $H_\alpha(x)$:

$$e^{-\frac{i}{2\xi} \int d^4x f_\alpha^2(x)} = \int \mathcal{D}H_\alpha e^{\frac{i\xi}{2} \int d^4x H_\alpha^2 - i \int d^4x f_\alpha H_\alpha}, \quad (6.3.1)$$

so that

$$\mathcal{L}_{tot} = \mathcal{L} + \mathcal{L}_{ghost} + \frac{\xi}{2} H_\alpha^2 - f_\alpha H_\alpha. \quad (6.3.2)$$

⁴The analysis in this section follows section 15.7 of ref. [60].

The total Lagrangian (6.3.2) is invariant under infinitesimal transformations parametrized by an anticommuting variable θ :

$$\begin{aligned}
\delta_\theta \Psi &= it_\alpha \theta \omega_\alpha \Psi, \\
\delta_\theta A_\mu^\alpha &= \theta D_\mu \omega_\alpha, \\
\delta_\theta \omega_\alpha &= -\frac{1}{2} \theta C_{\alpha\beta\gamma} \omega_\beta \omega_\gamma, \\
\delta_\theta \omega_\alpha^* &= -\theta H_\alpha, \\
\delta_\theta H_\alpha &= 0.
\end{aligned} \tag{6.3.3}$$

These are the BRST transformations, written in the non-canonical basis for the gauge fields where the gauge coupling does not appear in the interactions. The field Ψ represents any matter field, fermionic or bosonic. It will be useful in what follows to collectively denote all fields with the symbol Φ ($\Phi = A, \Psi, \omega, \omega^*, H$).

Crucially, BRST transformations are nilpotent. If we denote by $F(\Phi)$ a local functional of the fields, we can define a BRST functional operator s as $\delta_\theta F \equiv \theta sF$. Nilpotency implies that

$$s^2 F(\Phi) = 0. \tag{6.3.4}$$

Let us first check that s^2 vanishes when acting on single fields, i.e.

$$s^2 \Phi = 0, \tag{6.3.5}$$

for any field $\Phi = A, \Psi, \omega, \omega^*, H$. To this purpose it is convenient to adopt a compact notation for the BRST transformations (6.3.3). Omitting Lorentz indices, let

$$A = A^\alpha t_\alpha, \quad \omega = \omega_\alpha t^\alpha, \quad \omega^* = \omega_\alpha^* t^\alpha, \tag{6.3.6}$$

be index-free fields. Since (6.3.6) are matrix-valued fields, care has to be paid in the order in which they are written. In terms of these fields, eq.(6.3.3) can be rewritten as

$$\begin{aligned}
s\Psi &= i\omega\Psi, \\
sA &= D\omega = d\omega - i[A, \omega], \\
s\omega &= i\omega^2, \\
s\omega^* &= -H, \\
sH &= 0,
\end{aligned} \tag{6.3.7}$$

where in the transformation of ω we used that

$$\omega^2 = \omega_\beta t^\beta \omega_\gamma t^\gamma = \frac{1}{2} [t_\beta, t_\gamma] \omega_\beta \omega_\gamma = \frac{i}{2} C_{\alpha\beta\gamma} \omega_\beta \omega_\gamma t^\alpha. \tag{6.3.8}$$

We start with Ψ . We have

$$s^2\Psi = s(i\omega\Psi) = i\left[(s\omega)\Psi - \omega(s\Psi)\right] = i(i\omega^2\Psi - i\omega^2\Psi) = 0. \quad (6.3.9)$$

Note that s is a fermionic operator and hence it anti-commutes with ω .

Consider now the BRST transformations of the gauge fields. We have

$$\begin{aligned} s^2A &= s(d\omega - iA\omega + i\omega A) = d(s\omega) - i(sA)\omega - iA(s\omega) + i(s\omega)A - i\omega(sA) \\ &= id\omega^2 - i(d\omega - iA\omega + i\omega A)\omega + A\omega^2 - \omega^2A - i\omega(d\omega - iA\omega + i\omega A) \\ &= 0. \end{aligned} \quad (6.3.10)$$

The nilpotency of the BRST transformations acting on ω^* and H is trivial. We have, from eq. (6.3.7),

$$s^2\omega^* = s(-H) = 0, \quad s^2H = 0. \quad (6.3.11)$$

We have then proved eq.(6.3.5). For products of two elementary fields we have

$$\delta_\theta(\Phi_1\Phi_2) = \theta(s\Phi_1)\Phi_2 + \Phi_1(\theta s\Phi_2) = \theta\left((s\Phi_1)\Phi_2 \pm \Phi_1(s\Phi_2)\right), \quad (6.3.12)$$

with $+$ or $-$ depending whether the field Φ_1 is bosonic or fermionic.⁵ Hence

$$s(\Phi_1\Phi_2) = \left((s\Phi_1)\Phi_2 \pm \Phi_1(s\Phi_2)\right). \quad (6.3.13)$$

Acting again with a BRST transformation gives

$$\begin{aligned} \delta_\theta(s\Phi_1\Phi_2) &= \delta_\theta(s\Phi_1)\Phi_2 + s\Phi_1\delta_\theta\Phi_2 \pm (\delta_\theta\Phi_1)s\Phi_2 \pm \Phi_1\delta_\theta(s\Phi_2) \\ &= \theta\left((s^2\Phi_1)\Phi_2 \mp (s\Phi_1)(s\Phi_2) \pm (s\Phi_1)(s\Phi_2) + \Phi_1(s^2\Phi_2)\right) = 0 \end{aligned} \quad (6.3.14)$$

if eq. (6.3.5) is satisfied, for Φ_1 both bosonic and fermionic. Iterating the argument, we can prove the nilpotency of an arbitrary product of fields, and hence eq.(6.3.4) follows.

We now proceed to prove that the Lagrangian (6.3.2) is BRST-invariant. On the physical fields Ψ and A_μ^α , the BRST transformations (6.3.3) can be seen as infinitesimal gauge transformations with parameter $\epsilon_\alpha(x) = \theta\omega_\alpha(x)$. The BRST-invariance of the gauge and matter Lagrangian term $\mathcal{L}$ immediately follows from the fact that $\mathcal{L}$ is gauge invariant. Let us now turn to the remaining three terms in eq. (6.3.2). We have

$$\delta_\theta f_\alpha = \delta_\theta \partial^\mu A_\mu^\alpha = \theta \partial_\mu D^\mu \omega_\alpha = \theta s f_\alpha. \quad (6.3.15)$$

Modulo total derivatives, we then get

$$\mathcal{L}_{ghost} = -\omega_\alpha^* s f_\alpha. \quad (6.3.16)$$

⁵Namely with Fermi statistics. In particular, ghosts and anti-ghosts are fermionic.

Since $s\omega_\alpha^* = -H_\alpha$, the last three terms in eq. (6.3.2) can be rewritten as

$$\mathcal{L}_{ghost} + \frac{\xi}{2}H_\alpha^2 - f_\alpha H_\alpha = s\left(f_\alpha\omega_\alpha^* - \frac{1}{2}\xi\omega_\alpha^*H_\alpha\right). \quad (6.3.17)$$

In this way, the BRST invariance of these terms is automatically ensured by the fact that $s^2 = 0$ for any functional. We conclude that the whole Lagrangian $\mathcal{L}_{tot}$ is invariant under BRST transformations.

From an operatorial point of view, the BRST transformations (6.3.3) are generated by a Grassmann Hermitian operator Q . For any field Φ , we have

$$\delta_\theta\Phi = i[\theta Q, \Phi] = i\theta[Q, \Phi]_\mp, \quad (6.3.18)$$

where $-$ and $+$ denote commutator and anti-commutator, respectively, depending on whether the field Φ is bosonic or fermionic. We then have

$$[Q, \Phi]_\mp = -is\Phi. \quad (6.3.19)$$

The nilpotency of s , $s^2 = 0$, is equivalent to

$$(-is)^2\Phi = [Q, [Q, \Phi]_\mp]_\pm = [Q^2, \Phi]_- = 0 \implies Q^2 = 0. \quad (6.3.20)$$

The BRST operator Q allows us to make the following partition of the Hilbert space in non-abelian gauge theories. Any state $|\phi\rangle$ in the Hilbert space falls in one of the following three categories:

$$\begin{aligned} Q|\phi_1\rangle &\neq 0 \\ Q|\phi_2\rangle &= 0, \quad \text{with } |\phi_2\rangle = Q|\phi_1\rangle \\ Q|\phi_3\rangle &= 0, \quad \text{but } |\phi_3\rangle \neq Q|\phi_1\rangle. \end{aligned} \quad (6.3.21)$$

The fields in the second class are manifestly unphysical, since they have vanishing norm:

$$||\phi_2\rangle|^2 = \langle\phi_2|\phi_2\rangle = \langle\phi_1|Q^2|\phi_1\rangle = 0. \quad (6.3.22)$$

We now show that gauge invariance implies that physical states should be annihilated by Q . More precisely, matrix elements between physical states $|\alpha\rangle$ and $|\beta\rangle$ should not depend on the choice of gauge fixing term. We have just seen that the total Lagrangian $\mathcal{L}_{tot}$ can be written as a physical gauge invariant term $\mathcal{L}$ plus a BRST variation of some functional F : $\mathcal{L}_{tot} = \mathcal{L} + sF(\Phi)$. The specific form of F depends on the gauge-fixing chosen. If we infinitesimally deform the gauge fixing, the functional F will also be deformed $F \rightarrow F + \delta F$. Demanding that

$$\langle\alpha|\beta\rangle_F = \langle\alpha|\beta\rangle_{F+\delta F} \quad (6.3.23)$$

is equivalent to the condition

$$\langle \alpha | [Q, \delta F]_+ | \beta \rangle_F = 0, \quad \forall \alpha, \beta \in \text{physical} \quad (6.3.24)$$

and for any sensible choice of gauge-fixing functional δF . We then conclude that Q should annihilate physical states. These are then identified with the states in the third category in eq. (6.3.21). States that are annihilated by the operator Q ($|\phi_3\rangle$) are said to be in the kernel of Q . The states $|\phi_2\rangle$ are said to be in the image of Q . The physical states are states in the kernel that are not in the image of Q . Such states are said to be in the cohomology of Q . It is clear that physical states are not uniquely defined. If $|\alpha\rangle$ is a given physical state, then any state of the form

$$|\tilde{\alpha}\rangle = |\alpha\rangle + Q|\phi_1\rangle \quad (6.3.25)$$

defines the same physical state, since

$$\langle \tilde{\alpha} | \beta \rangle = \langle \alpha | \beta \rangle \quad (6.3.26)$$

for any other physical state $|\beta\rangle$. Physical states $|\text{phys}\rangle$ correspond to equivalence classes within the class $|\phi_3\rangle$. This complicated structure of the Hilbert space in non-abelian gauge theories is a consequence of the redundancy introduced by the gauge symmetries. Like ghosts, any unphysical state, even if not present as an external line in a scattering process, can contribute as a virtual particle in loops. On the other hand, the optical theorem relates the imaginary part of a loop diagram with the square of scattering diagrams where the virtual particles become external on-shell states. Unitarity is then not obvious. BRST invariance is of great help to show us that, in fact, no problem arises because physical states are unitary by themselves, namely the contribution of unphysical states in loop diagrams always cancels. Ghosts are crucial for this cancellation to occur. This is best seen by considering

$$\langle \alpha | S^\dagger S | \beta \rangle = \sum_{i=1,2,3} \langle \alpha | S^\dagger | \phi_i \rangle \langle \phi_i | S | \beta \rangle = \langle \alpha | \beta \rangle, \quad (6.3.27)$$

where $|\alpha\rangle$ and $|\beta\rangle$ are two arbitrary physical states. The BRST operator Q commutes with the S matrix and hence

$$\langle \alpha | S^\dagger Q = Q S | \beta \rangle = 0. \quad (6.3.28)$$

This implies that both $S|\alpha\rangle$ and $S|\beta\rangle$ can be expanded as linear combination of states $|\phi_2\rangle$ and $|\phi_3\rangle$. Hence the states $|\phi_1\rangle$ cannot enter eq. (6.3.27). The states $|\phi_2\rangle$ have vanishing inner product with states in both classes $|\phi_2\rangle$ and $|\phi_3\rangle$. We then conclude that effectively

$$\langle \alpha | S^\dagger S | \beta \rangle = \sum_{\text{phys}} \langle \alpha | S^\dagger | \text{phys} \rangle \langle \text{phys} | S | \beta \rangle = \langle \alpha | \beta \rangle, \quad (6.3.29)$$

and hence unitarity is recovered.

The non-abelian generalizations of the QED WT identities, so called Slavnov-Taylor identities, provide constraints on the form of correlation functions. As an example, let us consider the general relation (4.4.13), adapted to a BRST transformation in a non-abelian gauge theory. Let us take $J_\psi = J_\omega = 0$ and keep only the currents for the non-abelian gauge field and the anti-ghosts. Explicitly, eq. (4.4.13) becomes

$$\begin{aligned} \int d^4x \left(\langle sA_\alpha^\mu(x) \rangle_J J_{A,\mu}^\alpha(x) + \langle s\omega_\alpha^*(x) \rangle_J J_{\omega^*}^\alpha(x) \right) = \\ \int d^4x \left(\langle D^\mu \omega_\alpha(x) \rangle_J J_{A,\mu}^\alpha(x) - \frac{1}{\xi} \langle \partial_\mu A_\alpha^\mu(x) \rangle_J J_{\omega^*}^\alpha(x) \right) = 0, \end{aligned} \quad (6.3.30)$$

where in the last equality we have integrated out the auxiliary field $H_\alpha = \partial_\mu A_\alpha^\mu / \xi$. Let us take two functional derivatives of eq. (6.3.30) with respect to the anti-ghost and gauge currents and then set all currents to zero. We get

$$\begin{aligned} \int d^4x \frac{\delta^2}{\delta J_{A,\alpha}^\nu(y) \delta J_{\omega^*}^\beta(z)} \left(\langle D^\mu \omega_\gamma(x) \rangle_J J_{A,\mu}^\gamma(x) - \frac{1}{\xi} \langle \partial_\mu A_\gamma^\mu(x) \rangle_J J_{\omega^*}^\gamma(x) \right) \Big|_{J_A=J_{\omega^*}=0} = \\ i \langle \omega_\beta^*(z) D_\nu \omega_\alpha(y) \rangle - \frac{i}{\xi} \langle A_{\nu,\alpha}(y) \partial_\mu A_\beta^\mu(z) \rangle = 0. \end{aligned} \quad (6.3.31)$$

Multiply now expression (6.3.31) by $\partial/\partial y_\nu$. In so doing the first term becomes trivial, because the operator $\partial^\nu D_\nu$ is, by definition, the inverse of the exact ghost propagator:

$$\langle \omega_\beta^*(z) \partial^\nu D_\nu \omega_\alpha(y) \rangle = -i \delta_{\alpha\beta} \delta^{(4)}(z-y). \quad (6.3.32)$$

Plugging in eq. (6.3.31) gives

$$\langle \partial^\nu A_\nu^\alpha(y) \partial_\mu A_\beta^\mu(z) \rangle = -i \xi \delta_{\alpha\beta} \delta^{(4)}(z-y). \quad (6.3.33)$$

The exact gluon propagator is trivial in the color indices and can be written as

$$G_{\mu\nu}^{\alpha\beta}(y,z) \equiv \langle A_\nu^\alpha(y) A_\mu^\beta(z) \rangle = \delta^{\alpha\beta} \int \frac{d^4p}{(2\pi)^4} G_{\mu\nu}(p) e^{ip(y-z)}, \quad (6.3.34)$$

with

$$p^\mu p^\nu G_{\mu\nu}(p) = -i\xi. \quad (6.3.35)$$

In perturbation theory, we can write $G_{\mu\nu}(p) = G_{\mu\nu}^{(0)}(p) + G_{\mu\nu}^{(q)}(p)$, where $G_{\mu\nu}^{(0)}(p)$ is the tree-level propagator we have computed in eq. (6.2.45) and $G_{\mu\nu}^{(q)}(p)$ encodes the quantum corrections. We notice that the classical propagator alone satisfies eq. (6.3.35), implying that the quantum corrections must be transverse: $p^\mu p^\nu G_{\mu\nu}^{(q)}(p) = 0$. The gluon propagator to all orders in perturbation theory can be written as

$$G_{\mu\nu}(p) = \frac{-i}{p^2 + i\epsilon} \left(\frac{\eta_{\mu\nu} p^2 - p^\mu p^\nu}{p^2(1 - \Pi(p^2))} + \xi \frac{p_\mu p_\nu}{p^2} \right), \quad (6.3.36)$$

where all quantum corrections are encoded in the scalar form factor $\Pi(p^2)$.

BRST invariance is very useful to also establish the renormalization properties of non-abelian gauge theories. Being a symmetry of the total action $S_{tot} = \int d^4x \mathcal{L}_{tot}$, one can show that the total quantum action Γ_{tot} is of the same form of S_{tot} , with renormalized parameters. This derivation is discussed in detail in section 6.5.

6.4 The Background Field Method

The renormalization of non-abelian gauge theories can proceed in a perturbative way as it was explained in chapter 3 — and exemplified in the case of the abelian $U(1)$ gauge theory (QED), by singling out the leading divergences and the class of diagrams which are in need of renormalization. However, the corresponding analysis becomes quite challenging and demanding already for the lowest-order calculations, because of the rather large number of diagrams involved. In particular, in non-abelian gauge theories the identity (3.4.17) no longer holds. This important difference between abelian and non-abelian theories can be traced back to the different nature of the associated conserved currents in the two classes of theories. In abelian gauge theories we have a gauge-invariant divergence-less current J^μ and its associated well-defined conserved charge Q given by the spatial integral of its time component. The notion of an absolute normalization of the charge Q guarantees that $Z_1 = Z_2$ for all charged particles. As we have seen below eq.(6.1.37), in non-abelian gauge theories the currents are either conserved but not covariant, or non-conserved but covariant. Hence there is no a well-defined standard gauge-invariant conserved charge Q as in the QED case and correspondingly $Z_1 \neq Z_2$, though the ratio Z_1/Z_2 is the same for all charged fields. Contrary to QED, then, in addition to the gluon two-point function we should in general compute vertex corrections in order to derive the β -function for the gauge coupling. This can in principle be computed in different ways, using gauge, ghost or matter vertices. Gauge invariance guarantees that the same result is obtained independently of this choice.

As a matter of fact, also in non-abelian gauge theories the β -function can be deduced from the gluon two-point function using an alternative approach, denoted background field method, that will be discussed in this section. This method has a long history and dates back to DeWitt [22]. It is usually discussed in terms of determinants, where applicability is restricted to one-loop processes. We present it here in the formulation given by [1], which has the advantage of being applicable to any loop in perturbation theory. We use the background field method primarily as a tool to compute the (one-loop) β -function in non-abelian gauge theories. It should be clear, however, that this method has a wider

variety of applications.

6.4.1 The method

The basic idea behind the background field method consists in calculating the effective action in the presence of a classical background field gauge A_μ , in total analogy to the scalar analogue (4.1.24):

$$e^{i\Gamma(A_\mu)} = \int_{1PI} \mathcal{D}Q_\mu e^{iS(A_\mu+Q_\mu)} \delta(G(A, Q)) \Delta_G(A, Q) , \quad (6.4.1)$$

where we have denoted by A_μ and Q_μ the background and fluctuating fields, respectively. The key point is to choose a gauge fixing in the path integral such that $\Gamma(A)$ is gauge invariant under transformations of the background:

$$\begin{aligned} \delta_B A_\mu^\alpha &= \partial_\mu \epsilon_\alpha + g C_{\alpha\beta\gamma} A_\mu^\beta \epsilon_\gamma , \\ \delta_B Q_\mu^\alpha &= g C_{\alpha\beta\gamma} Q_\mu^\beta \epsilon_\gamma . \end{aligned} \quad (6.4.2)$$

It is important to distinguish the gauge transformations of the background (6.4.2) from the quantum gauge transformations under which⁶

$$\begin{aligned} \delta A_\mu^\alpha &= 0 , \\ \delta Q_\mu^\alpha &= \partial_\mu \epsilon_\alpha + g C_{\alpha\beta\gamma} (A_\mu^\beta + Q_\mu^\beta) \epsilon_\gamma . \end{aligned} \quad (6.4.3)$$

The gauge fixing where $\Gamma(A)$ is gauge invariant is called background field gauge and is obtained by choosing the functional (6.2.30) as

$$G_f(A, Q) = D_\mu(A) Q_\alpha^\mu - f_\alpha = \partial_\mu Q_\alpha^\mu + g C_{\alpha\beta\gamma} A_\mu^\beta Q_\gamma^\mu - f_\alpha . \quad (6.4.4)$$

The ghost action associated to this gauge fixing is obtained, as usual, by taking a quantum gauge transformation of the gauge fixing functional:

$$\delta G(A, Q) = D_\mu(A) D^\mu(A + Q) \epsilon_\alpha \quad (6.4.5)$$

and hence

$$\mathcal{L}_{ghost} = D_\mu(A) \omega_\alpha^* (D^\mu(A) \omega_\alpha + g C_{\alpha\beta\gamma} Q_\beta^\mu \omega_\gamma) . \quad (6.4.6)$$

Weighting the path integral in eq. (6.4.1) with the phase factor (6.2.31) and integrating over the delta functional gives

$$e^{i\Gamma(A_\mu)} = \int_{1PI} \mathcal{D}Q_\mu \mathcal{D}\omega \mathcal{D}\omega^* e^{iS(A_\mu+Q_\mu) + iS_{ghost} - \frac{i}{2\xi} \int d^4x (D_\mu(A) Q_\alpha^\mu)^2} . \quad (6.4.7)$$

⁶Note that the sum $A+Q$ transforms in the same way (i.e., as in eq. (6.1.18)) under both transformations (6.4.2) or (6.4.3), as expected from the full gauge field. The different splitting distinguishes however the background from the quantum gauge transformations.

Under the background gauge transformations (6.4.2) the ghosts and anti-ghosts transform as fields in the adjoint representations:

$$\begin{aligned}\delta_B \omega_\alpha &= g C_{\alpha\beta\gamma} \epsilon_\beta \omega_\gamma, \\ \delta_B \omega_\alpha^* &= g C_{\alpha\beta\gamma} \epsilon_\beta \omega_\gamma^*.\end{aligned}\tag{6.4.8}$$

Notice that while the ghost and gauge fixing terms are clearly not gauge invariant under the quantum gauge transformations (6.4.3), they are invariant under the background gauge transformations (6.4.2) and (6.4.8). The matter action in $S(A + Q)$, in fact, is manifestly invariant. The gauge kinetic term is

$$\begin{aligned}& -\frac{1}{4} \left(\partial_\mu (A_\nu^\alpha + Q_\nu^\alpha) - \partial_\nu (A_\mu^\alpha + Q_\mu^\alpha) + g C_{\alpha\beta\gamma} (A_\mu^\beta + Q_\mu^\beta) (A_\nu^\gamma + Q_\nu^\gamma) \right)^2 = \\ & -\frac{1}{4} \left(F_{\mu\nu}^\alpha(A) + D_\mu(A) Q_\nu^\alpha - D_\nu(A) Q_\mu^\alpha + g C_{\alpha\beta\gamma} Q_\mu^\beta Q_\nu^\gamma \right)^2,\end{aligned}\tag{6.4.9}$$

which is also invariant. We have then proved that

$$\delta_B \Gamma(A) = 0.\tag{6.4.10}$$

The divergent part of $\Gamma(A)$ has to be a *gauge-invariant* polynomial of dimension 4 and therefore it is necessarily of the form

$$-\frac{Z_A^{-1}}{4} \int d^4x F_{\mu\nu,\alpha}(A_B) F^{\mu\nu,\alpha}(A_B) + \text{finite parts}\tag{6.4.11}$$

where Z_A is a divergent constant and we have redefined the background field as A_B to emphasize that this is a bare quantity. Note that, as usual, $Z_A = 1$ for $g \rightarrow 0$, and therefore $Z_A = 1 + \mathcal{O}(g^2)$. The divergence in eq.(6.4.11) is reabsorbed in a redefinition of the background field: $A_B = Z_A^{1/2} A$. Because of background gauge invariance, the bare gauge coupling g_B entering eq.(6.4.11) has to correspondingly be redefined as

$$g_B = g \mu^{\epsilon/2} Z_A^{-1/2}.\tag{6.4.12}$$

The renormalization of the coupling constant g is then fixed by the renormalization constant of the field A , as it was the case for QED, and the β -function can be calculated as in subsection 5.7.

6.4.2 Two-Point Function of the Background Field: Feynman Rules

The calculation of Z_A requires to take into account for the divergences which arise when the integration over Q , ω , and ω^* is performed in eq. (6.4.7). Given eq.(6.4.11), it is

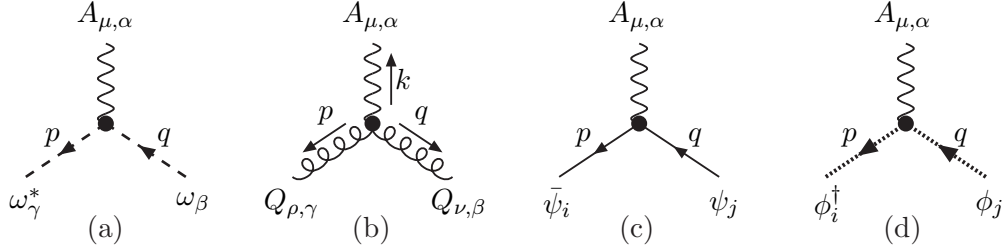


Figure 6.1: Trilinear vertices involving one background field A within the background field method. Wiggly lines indicate the background field A , curly lines the quantum fluctuations Q , dashed directed lines the ghosts, while dotted directed lines the scalars. We list here only the trilinear vertices which, at the lowest order in perturbation theory, give a non-vanishing contribution to the two-point function of the background fields: (a) A - ω^* - ω vertex $V^{A\omega^*\omega}$ in eq. (6.4.13), (b) A - Q - Q vertex V^{AQQ} in eq. (6.4.19), (c) A - $\bar{\psi}$ - ψ vertex $V^{A\bar{\psi}\psi}$ in eq. (6.4.22), and (d) A - $\phi^\dagger$ - ϕ vertex $V^{A\phi^\dagger\phi}$ in eq. (6.4.23).

enough to consider the two-point function of the background field A . This is analogous to a standard propagator if were not for the fact that A is considered as an external field. Let us report below the Feynman rules that are necessary to the computation of Z_A at one-loop level.

From $\mathcal{L}_{ghost}$ in eq. (6.4.6) one gets a A - ω^* - ω -vertex (represented in fig. 6.1(a))

$$V_{\mu\alpha;\beta;\gamma}^{A\omega^*\omega} \equiv \frac{\delta i\mathcal{L}_{ghost}}{\delta A_\alpha^\mu \delta \omega_\beta(q) \delta \omega_\gamma^*(p)} = -gC_{\alpha\beta\gamma}(p+q)_\mu, \quad (6.4.13)$$

where p_μ and q_μ are the momenta of the ghosts. In eq.(6.4.13) we have omitted to write the dependence of the gauge field A_μ on its momentum, since the vertex does not depend on it. Analogue simplifications will be done below. The additional vertices involving A which are generated by this term are of the form A - A - ω^* - ω and A - Q - ω^* - ω . While the latter does not contribute at one-loop to the two-point correlation function of the background field, the former can. We have (see fig. 6.2(a))

$$V_{\mu\alpha;\beta;\gamma;\delta}^{AA\omega^*\omega} \equiv \frac{\delta i\mathcal{L}_{ghost}}{\delta A_\alpha^\mu \delta A_\beta^\nu \delta \omega_\gamma \delta \omega_\delta^*} = ig^2\eta_{\mu\nu}(C_{\alpha\gamma\omega}C_{\beta\delta\omega} + C_{\alpha\delta\omega}C_{\beta\gamma\omega}). \quad (6.4.14)$$

The gauge-fixing term $\mathcal{L}_{g.f.}$ generates a A - Q - Q -vertex

$$\tilde{V}_{\mu\alpha;\nu\beta;\rho\gamma}^{AQQ} \equiv \frac{\delta i\mathcal{L}_{g.f.}}{\delta A_\alpha^\mu \delta Q_\beta^\nu(q) \delta Q_\gamma^\rho(p)} = -\frac{g}{\xi} \left(\eta_{\mu\nu}C_{\alpha\beta\gamma}p_\rho + \eta_{\mu\rho}C_{\alpha\gamma\beta}q_\nu \right), \quad (6.4.15)$$

with the momenta of the fields oriented as in fig. 6.1(b). In addition to this interaction, $\mathcal{L}_{g.f.}$ generates a vertex A - A - Q - Q ,

$$\tilde{V}_{\mu\alpha;\nu\beta;\rho\gamma;\sigma\delta}^{AAQQ} \equiv \frac{\delta i\mathcal{L}_{g.f.}}{\delta A_\alpha^\mu \delta A_\beta^\nu \delta Q_\gamma^\rho \delta Q_\delta^\sigma} = -\frac{ig^2}{\xi} \left(\eta_{\mu\rho}\eta_{\nu\sigma}C_{\alpha\gamma\omega}C_{\beta\delta\omega} + \eta_{\mu\sigma}\eta_{\nu\rho}C_{\beta\gamma\omega}C_{\alpha\delta\omega} \right). \quad (6.4.16)$$

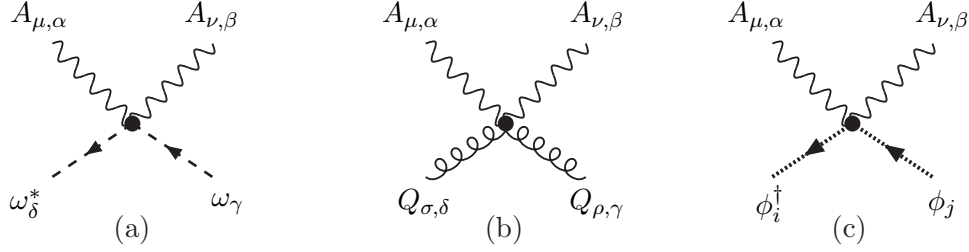


Figure 6.2: Quartic vertices involving two background fields A within the background field method. Wiggly lines indicate the background field A , curly lines the quantum fluctuations Q , dashed directed lines the ghosts, while dotted directed lines the scalars. We list here only the quartic vertices which, at the lowest order in perturbation theory, might give a non-vanishing contribution to the two-point function of the background fields: (a) A - ω^* - ω vertex $V^{AA\omega\omega^*}$ in eq. (6.4.14), (b) A - A - Q - Q vertex V^{AAQQ} in eq. (6.4.20), (c) A - A - $\phi^\dagger$ - ϕ vertex $V^{AA\phi^\dagger\phi}$ in eq. (6.4.24).

The vertex involving three gauge fields A in the standard YM theory generates several vertices: A - A - Q , A - Q - Q , and Q - Q - Q but only the second one is relevant for determining Z_A at the lowest order. We have

$$\hat{V}_{\mu\alpha;\nu\beta;\rho\gamma}^{AQQ} \equiv \frac{\delta i\mathcal{L}}{\delta A_\alpha^\mu(k)\delta Q_\beta^\nu(q)\delta Q_\gamma^\rho(p)} = gC_{\alpha\beta\gamma} \left((k-q)_\rho \eta_{\mu\nu} + (q-p)_\mu \eta_{\nu\rho} + (p-k)_\nu \eta_{\rho\mu} \right), \quad (6.4.17)$$

where all momenta are assumed to be outgoing from the vertex. The expression (6.4.17) coincides with the trilinear vertex computed in YM theories in standard gauges. The vertex involving four gauge fields A in the standard YM theory generates several vertices: A - A - A - A , A - A - A - Q , A - A - Q - Q , A - Q - Q - Q , and Q - Q - Q - Q but only the third one is relevant for determining Z_A at the lowest order. The corresponding expression is again the standard one:

$$\hat{V}_{\mu\alpha;\nu\beta;\rho\gamma;\sigma\delta}^{AAQQ} \equiv \frac{\delta i\mathcal{L}}{\delta A_\alpha^\mu\delta A_\beta^\nu\delta Q_\gamma^\rho\delta Q_\delta^\sigma} = -ig^2 \left(C_{\alpha\beta\omega}C_{\gamma\delta\omega}(\eta^{\mu\rho}\eta^{\nu\sigma} - \eta_{\mu\sigma}\eta_{\nu\rho}) + C_{\alpha\gamma\omega}C_{\beta\delta\omega}(\eta_{\mu\nu}\eta_{\rho\sigma} - \eta_{\mu\sigma}\eta_{\nu\rho}) + C_{\alpha\delta\omega}C_{\beta\gamma\omega}(\eta_{\mu\nu}\eta_{\rho\sigma} - \eta_{\mu\rho}\eta_{\nu\sigma}) \right). \quad (6.4.18)$$

The total vertices A - Q - Q and A - A - Q - Q of this theory are given by the sum of the one stemming from the gauge-fixing term in eqs. (6.4.15) and (6.4.16) and the ones from $\mathcal{L}$ in

eqs. (6.4.17) and (6.4.18), see figs. 6.1(b) and 6.2(b),

$$V_{\mu\alpha;\nu\beta;\rho\gamma}^{AQQ} = gC_{\alpha\beta\gamma} \left(\left(k - q - \frac{1}{\xi}p \right)_\rho \eta_{\mu\nu} + (q - p)_\mu \eta_{\rho\nu} + \left(p - k + \frac{1}{\xi}q \right)_\nu \eta_{\rho\mu} \right), \quad (6.4.19)$$

$$V_{\mu\alpha;\nu\beta;\rho\gamma;\sigma\delta}^{AAQQ} = -ig^2 \left(C_{\alpha\beta\omega} C_{\gamma\delta\omega} (\eta_{\mu\rho} \eta_{\nu\sigma} - \eta_{\mu\sigma} \eta_{\nu\rho}) + C_{\alpha\gamma\omega} C_{\beta\delta\omega} \right. \\ \left. \left(\eta_{\mu\nu} \eta_{\rho\sigma} - \eta_{\mu\sigma} \eta_{\nu\rho} + \frac{1}{\xi} \eta_{\mu\rho} \eta_{\nu\sigma} \right) + C_{\alpha\delta\omega} C_{\beta\gamma\omega} \left(\eta_{\mu\nu} \eta_{\rho\sigma} - \eta_{\mu\rho} \eta_{\nu\sigma} + \frac{1}{\xi} \eta_{\mu\sigma} \eta_{\nu\rho} \right) \right), \quad (6.4.20)$$

where

$$V_{\mu\alpha;\nu\beta;\rho\gamma}^{AQQ} \equiv \tilde{V}_{\mu\alpha;\nu\beta;\rho\gamma}^{AQQ} + \hat{V}_{\mu\alpha;\nu\beta;\rho\gamma}^{AQQ}, \\ V_{\mu\alpha;\nu\beta;\rho\gamma;\sigma\delta}^{AAQQ} \equiv \tilde{V}_{\mu\alpha;\nu\beta;\rho\gamma;\sigma\delta}^{AAQQ} + \hat{V}_{\mu\alpha;\nu\beta;\rho\gamma;\sigma\delta}^{AAQQ}. \quad (6.4.21)$$

The interaction with the matter is ruled by the standard vertices of the YM theory. For fermions we have, see fig. 6.1(c),

$$V_{\mu\alpha;ij}^{A\bar{\psi}\psi} \equiv \frac{\delta i\mathcal{L}}{\delta A_\alpha^\mu \delta \psi_j \delta \bar{\psi}_i} = ig\gamma_\mu t_{ij}^\alpha, \quad (6.4.22)$$

where t^α are the gauge generators in the appropriate representation. For scalars, we can have both the trilinear and quartic vertices, see figs. 6.1(d) and 6.2(c):

$$V_{\mu\alpha;ij}^{A\phi^\dagger\phi} \equiv \frac{\delta i\mathcal{L}}{\delta A_\alpha^\mu \delta \phi_j(q) \delta \phi_i^\dagger(p)} = ig t_{ij}^\alpha (p + q)_\mu, \quad (6.4.23)$$

$$V_{\mu\alpha;\nu\beta;ij}^{AA\phi^\dagger\phi} \equiv \frac{\delta i\mathcal{L}}{\delta A_\alpha^\mu \delta A_\beta^\nu \delta \phi_j \delta \phi_i^\dagger} = ig^2 \eta_{\mu\nu} \{t^\alpha, t^\beta\}_{ij}, \quad (6.4.24)$$

where again t^α are the gauge generators in the appropriate representation.

With all these vertices at hand, the radiative corrections to the two-point function of the background field can be computed. We leave to the reader to perform the computation of Z_A at one-loop level and report here only the final result for the β -function.

In a $SU(N)$ gauge theory with n_f Dirac fermions in representations $r_i^{(F)}$ and n_s scalars in representations $r_i^{(S)}$ of $SU(N)$, one has at one-loop level

$$\beta(g) = -\frac{g^3}{(4\pi)^2} \left(\frac{11N}{3} - \frac{4}{3} \sum_{i=1}^{n_f} T_2(r_i^{(F)}) - \frac{1}{3} \sum_{i=1}^{n_s} T_2(r_i^{(S)}) \right), \quad (6.4.25)$$

where

$$\text{tr}_r t^\alpha t^\beta \equiv \delta^{\alpha\beta} T_2(r), \quad T_2(\text{adj.}) = N. \quad (6.4.26)$$

In particular, for $n_s = 0$ and $r_i^{(F)}$ equal to the fundamental representation for any fermion, we have:

$$\beta(g) = -\frac{g^3}{(4\pi)^2} \left(\frac{11N}{3} - \frac{2}{3} n_f \right), \quad (6.4.27)$$

QCD corresponds to the choice $N = 3$ and $n_f = 6$ and is then an asymptotically free theory with $\beta(g) < 0$, as we discuss in general in sec. 5.5.

6.5 Proof of the Renormalizability of Non-Abelian Gauge Theories**

The BRST symmetry provides a useful set-up to prove the renormalizability of non-abelian gauge theories. Before attaching the problem, we need however to develop a bit on the formalism and derive what is called the master equation.⁷

6.5.1 The Master Equation**

The relation (4.4.14), applied to the BRST transformations (6.3.3), give

$$\int d^4x \langle \Delta^n[\chi(x)] \rangle_{J(\chi)} \frac{\delta_L \Gamma}{\delta \chi^n(x)} = 0, \quad (6.5.1)$$

where χ^n is a compact notation to denote all the fields: $\chi^n = (A_\mu^\alpha, \omega_\alpha, \omega_\alpha^*, \psi, \phi, H_\alpha)$, and we have defined $\delta \chi^n = \theta s \chi^n \equiv \theta \Delta^n(\chi)$, Care has to be taken in taking functional derivatives in presence of Grassmann fields, so we distinguish with a subscript L and R derivatives defined as follows:

$$\frac{\delta_L F}{\delta \phi(x)} = \frac{\overrightarrow{\delta}_L}{\delta \phi(x)} F, \quad \frac{\delta_R F}{\delta \phi(x)} = F \frac{\overleftarrow{\delta}_R}{\delta \phi(x)}. \quad (6.5.2)$$

The BRST transformations are quadratic in the fields and hence $\langle \Delta^n[\chi(x)] \rangle_{J(\chi)} \neq \Delta^n[\chi(x)]$. In order to overcome this problem, we introduce, in addition to the source J_n that couple to χ^n , extra sources $K_n(x)$ that couple to $\Delta^n(\chi)$. Correspondingly we define

$$e^{iW[J,K]} = \int \mathcal{D}\chi^n e^{iS_{tot} + i \int \Delta^n K_n + i \int \chi^n J_n} \quad (6.5.3)$$

and the 1PI generator

$$\Gamma[\chi, K] = W[J, K] - \int d^4x \chi^n(x) J_n(x), \quad (6.5.4)$$

where $J_n = J_n(\chi, K)$, obtained by inverting the relation

$$\frac{\delta_R W[J, K]}{\delta J_n(x)} = \chi^n(x). \quad (6.5.5)$$

Under a BRST transformation $\delta \Delta^n = \theta s^2 \chi^n = 0$ and hence the 1PI action $\Gamma[\chi, K]$ satisfies the analogue of eq.(6.5.1):

$$\int d^4x \langle \Delta^n[\chi(x)] \rangle_{J(\chi, K)} \frac{\delta_L \Gamma[\chi, K]}{\delta \chi^n(x)} = 0, \quad (6.5.6)$$

Taking a functional R-derivation with respect to $K_n(x)$ of eq.(6.5.4) gives

$$\begin{aligned} \frac{\delta_R \Gamma[\chi, K]}{\delta K_n(x)} &= \frac{\delta_R W[J, K]}{\delta K_n(x)} + \int d^4y \frac{\delta_R W[J, K]}{\delta J_m(y)} \frac{\delta_R J_m(y)}{\delta K_n(x)} - \int d^4y \chi^m(y) \frac{\delta_R J_m(y)}{\delta K_n(x)} \\ &= \frac{\delta_R W[J, K]}{\delta K_n(x)} = i \langle \Delta^n[\chi(x)] \rangle_{J(\chi, K)}. \end{aligned} \quad (6.5.7)$$

⁷The analysis in this section closely follows sections 16.4, 17.1 and 17.2 of ref. [60].

We can then rewrite eq.(6.5.6) as

$$\int d^4x \frac{\delta_R \Gamma[\chi, K]}{\delta K_n(x)} \frac{\delta_L \Gamma[\chi, K]}{\delta \chi^n(x)} = 0. \quad (6.5.8)$$

Since the sources K_n have the same commuting properties as Δ^n , which are opposite to those of χ^n , $\delta_R \Gamma / \delta K_n$ and $\delta_L \Gamma / \delta J_n$ anti-commute with each other. We can then rewrite more compactly eq.(6.5.8) as

$$(\Gamma, \Gamma) = 0, \quad (6.5.9)$$

where we have defined, for any two functionals F and G , the operator

$$(F, G) \equiv \int d^4x \left(\frac{\delta_R F}{\delta \chi^n(x)} \frac{\delta_L G}{\delta K_n(x)} - \frac{\delta_R F}{\delta K_n(x)} \frac{\delta_L G}{\delta \chi^n(x)} \right). \quad (6.5.10)$$

Equation (6.5.9) is denoted the master equation and is the key relation that we will use in the next subsection to prove the renormalizability of non-abelian gauge theories.

6.5.2 Structure of Divergences**

In this subsection we will prove that all the divergences occurring at any loop in perturbation theory can be reabsorbed by a renormalization of the couplings appearing in the classical action, namely by a proper choice of counter-terms. We define the classical bare action including the extra K -source term:

$$S_B = S_{tot}(\chi) + \int \Delta^n K_n. \quad (6.5.11)$$

As usual, we rewrite the action $S_B = S + S_\infty$, where S is the action written in terms of physical fields and couplings and S_∞ includes the counter-terms. The 1PI action $\Gamma[\chi, K]$ admits a loop expansion of the form

$$\Gamma[\chi, K] = \sum_{N=0}^{\infty} \Gamma_N[\chi, K] \quad (6.5.12)$$

where $\Gamma_N[\chi, K]$ is the N -loop effective action. Clearly, we have $\Gamma_0 = S$. Plugging eq.(6.5.12) in the master equation (6.5.9) gives an infinite set of equations, valid for any N :

$$\sum_{N'=0}^N (\Gamma_{N'}, \Gamma_{N-N'}) = 0. \quad (6.5.13)$$

We denote by $\Gamma_{N,\infty}$ the divergent part of the N -loop effective action. If the theory is renormalizable, by a proper choice of S_∞ , Γ_N should be finite for each value of N , i.e. $\Gamma_{N,\infty}$ should all vanish. We will prove this statement by induction, assuming that this is

true for $M \leq N - 1$, namely that all the divergences in Γ_M have been reabsorbed in S_∞ . If this is true, divergences in eq.(6.5.13) can only arise from the terms with $N' = 0, N$, that give rise to the equation

$$(S, \Gamma_{N,\infty}) = 0. \quad (6.5.14)$$

Let us see the possible form of $\Gamma_{N,\infty}$. It should be a local functional, invariant under all the linearly realized symmetries of the classical action. By power counting renormalization, all operators should have dimension less or equal to 4. Notice that

$$\begin{aligned} [\Delta_n] &= [\chi_n] + 1, \quad [K_n] = 3 - [\chi_n], \\ \text{gh.}(\chi_n) &\equiv \gamma_n, \quad \text{gh.}(\Delta_n) = \gamma_n + 1, \quad \text{gh.}(K_n) = -\gamma_n - 1, \end{aligned} \quad (6.5.15)$$

where we have denoted by γ_n the ghost number of the field χ^n . Since Δ^H trivially vanishes, $K_H = 0$. We see from eq.(6.5.15) that all K -currents have negative ghost charge, with the exception of K_{ω^*} that has ghost charge zero. Thus the only allowed operator quadratic in the K -currents is $K_{\omega^*}^\alpha K_{\omega^*}^\alpha$. This is already of dimension four, so no extra fields can be added to it. From eq.(6.5.7) we have

$$\frac{\delta_R \Gamma_{N,\infty}}{\delta K_{\omega^*}^\alpha(x)} = i \langle \Delta^{\omega^*}[\chi(x)] \rangle_{J(\chi,K)} = -i H_\alpha(x), \quad (6.5.16)$$

where the last identity applies because the BRST transformation of ω_α^* is linear. The right-hand side of eq.(6.5.16) is fully saturated by the classical action, and we conclude that $\Gamma_{N,\infty}$ does not depend on K_{ω^*} for $N > 0$ and is at most linear in the K 's for any N . We can write

$$\Gamma_{N,\infty}[\chi, K] = \Gamma_{N,\infty}[\chi, 0] + \int d^4x \mathcal{D}_N^n[\chi] K_n(x). \quad (6.5.17)$$

Recall that

$$S[\chi, K] = S[\chi] + \int d^4x \Delta_n[\chi] K_n(x). \quad (6.5.18)$$

Let us plug eqs.(6.5.17) and (6.5.18) in eq.(6.5.14). The terms with no K 's give

$$\int d^4x \left(-\frac{\delta_L S[\chi]}{\delta \chi^n(x)} \mathcal{D}_N^n[\chi(x)] - \Delta_n[\chi(x)] \frac{\delta_L \Gamma_{N,\infty}}{\delta \chi^n(x)} \right) = 0. \quad (6.5.19)$$

In deriving eq.(6.5.19) we used the fact that K_n and χ_n are bosonic (fermionic) and fermionic (bosonic) for any n , hence for any two arbitrary bosonic functionals A and B one has

$$\frac{\delta_R A}{\delta \chi^n} \frac{\delta_L B}{\delta K_n} = -\frac{\delta_L A}{\delta \chi^n} \frac{\delta_R B}{\delta K_n} = -\frac{\delta_L B}{\delta K_n} \frac{\delta_L A}{\delta \chi^n}. \quad (6.5.20)$$

The terms linear in K 's give

$$\int d^4x \left(\mathcal{D}_N^n[\chi(x)] \frac{\delta_L \Delta_m[\chi]}{\delta \chi^n(x)} + \Delta_n[\chi(x)] \frac{\delta_L \mathcal{D}_N^m[\chi]}{\delta \chi^n(x)} \right) = 0. \quad (6.5.21)$$

Equation (6.5.19) can be rewritten as

$$\int d^4x \Delta_N^{(\epsilon)n} \frac{\delta_L \Gamma_N^{(\epsilon)}}{\delta \chi^n} = 0, \quad (6.5.22)$$

where we have defined

$$\begin{aligned} \Gamma_N^{(\epsilon)}[\chi] &= S[\chi] + \epsilon \Gamma_{N,\infty}[\chi, 0], \\ \Delta_N^{(\epsilon)n} &= \Delta^n + \epsilon \mathcal{D}_N^n, \end{aligned} \quad (6.5.23)$$

with ϵ an infinitesimal parameter. The $\mathcal{O}(\epsilon^0)$ term vanishes because S is BRST-invariant, while the $\mathcal{O}(\epsilon)$ terms reproduce eq.(6.5.19). Similarly, eq.(6.5.19) can be rewritten as the $\mathcal{O}(\epsilon)$ terms of the following relation:

$$\Delta_N^{(\epsilon)n} \frac{\delta_L}{\delta \chi^n} \left(\Delta_N^{(\epsilon)n} \frac{\delta_L}{\delta \chi^n} \right) = 0. \quad (6.5.24)$$

We conclude from eqs.(6.5.23) and (6.5.24) that the action $\Gamma_N^{(\epsilon)}$ is invariant under the transformations generated by $\Delta_N^{(\epsilon)n}$, which are nilpotent, like the original BRST transformations Δ^n .

Comparing eqs.(6.5.17) and (6.5.18) we notice that $\mathcal{D}_N^n$ and Δ_n have the same ghost and Lorentz quantum numbers. The most general $\Delta_N^{(\epsilon)n}$ transformations is then parametrized by a simple generalization of the original BRST transformations (6.3.3):

$$\begin{aligned} \delta_\theta^{(\epsilon)N} \Psi &= i\theta Q_\alpha^N \omega_\alpha \Psi, \\ \delta_\theta^{(\epsilon)N} A_{\mu\alpha} &= \theta (B_{\alpha\beta}^N \partial_\mu \omega_\beta + D_{\alpha\beta\gamma}^N A_{\mu\beta} \omega_\gamma), \\ \delta_\theta^{(\epsilon)N} \omega_\alpha &= -\frac{1}{2} \theta E_{\alpha\beta\gamma}^N \omega_\beta \omega_\gamma. \end{aligned} \quad (6.5.25)$$

The transformations of ω_α^* and H_α are like the original BRST transformations (6.3.3) since $\mathcal{D}_N^{\omega^*} = K_H = 0$. For simplicity, from now on we omit the superscript N on $\delta_\theta^{(\epsilon)N}$ and the coefficients Q_α^N , $B_{\alpha\beta}^N$, $D_{\alpha\beta\gamma}^N$ and $E_{\alpha\beta\gamma}^N$. Nilpotency of the ghost transformation in eq.(6.5.25) requires that $E_{\alpha\beta\gamma}$ satisfies a Jacobi identity (compare with eq.(??)). Hence $E_{\alpha\beta\gamma}$ should be the structure constant of a Lie algebra. Since for $\epsilon \rightarrow 0$ we should recover the original BRST transformations, we conclude that $E_{\alpha\beta\gamma}$ must be the structure constant of the original Lie algebra, up to a constant:

$$E_{\alpha\beta\gamma} = Z C_{\alpha\beta\gamma}. \quad (6.5.26)$$

Consider now the gauge fields. We have $s^{(\epsilon)} A_{\mu\alpha} = B_{\alpha\beta} \partial_\mu \omega_\beta + D_{\alpha\beta\gamma} A_{\mu\beta} \omega_\gamma$. Then

$$\begin{aligned} \delta_\theta^{(\epsilon)} (s^{(\epsilon)} A_\mu^\alpha) &= \partial_\mu \left(-\frac{1}{2} B_{\alpha\beta} \theta E_{\beta\gamma\delta} \omega_\gamma \omega_\delta \right) + D_{\alpha\beta\gamma} \theta (B_{\beta\sigma} \partial_\mu \omega_\sigma + D_{\beta\rho\sigma} A_\mu^\rho \omega_\sigma) \omega_\gamma \\ &\quad + D_{\alpha\beta\gamma} A_\mu^\beta \left(-\frac{1}{2} \theta E_{\gamma\rho\sigma} \omega_\rho \omega_\sigma \right). \end{aligned} \quad (6.5.27)$$

Using eq.(6.5.26) and reshuffling indices, the terms proportional to A_μ in eq. (6.5.27) give

$$\delta_\theta^{(\epsilon)}(s^{(\epsilon)}A_\mu^\alpha)|_A = -\frac{1}{2}\theta A_\mu^\beta\omega_\rho\omega_\sigma(D_{\alpha\gamma\rho}D_{\gamma\beta\sigma} - D_{\alpha\gamma\sigma}D_{\gamma\beta\rho} + ZC_{\gamma\rho\sigma}D_{\alpha\beta\gamma}). \quad (6.5.28)$$

This term vanishes, again due to the Jacobi identity, if

$$D_{\alpha\beta\gamma} = ZC_{\alpha\beta\gamma}. \quad (6.5.29)$$

The terms with no gauge fields give

$$\delta_\theta^{(\epsilon)}(s^{(\epsilon)}A_\mu^\alpha)|_{no A} = \theta\partial_\mu\omega_\sigma\omega_\gamma Z(C_{\alpha\beta\gamma}B_{\beta\sigma} - B_{\alpha\beta}C_{\beta\sigma\gamma}), \quad (6.5.30)$$

and vanishes if

$$B_{\alpha\beta} = ZN\delta_{\alpha\beta}. \quad (6.5.31)$$

Finally, let's look at the fermion transformations.

$$\begin{aligned} \delta_\theta^{(\epsilon)}(s^{(\epsilon)}\Psi) &= iQ_\alpha\left((\delta_\theta^{(\epsilon)}\omega_\alpha)\Psi + \omega_\alpha(\delta_\theta^{(\epsilon)}\Psi)\right) = iQ_\alpha\left(-\frac{1}{2}Z\theta C_{\alpha\beta\gamma}\omega_\beta\omega_\gamma\Psi + \omega_\alpha iQ_\beta\theta\omega_\beta\Psi\right) \\ &= \theta\omega_\alpha\omega_\beta\Psi\left(-\frac{i}{2}ZC_{\alpha\beta\gamma}Q_\gamma + \frac{1}{2}[Q_\alpha, Q_\beta]\right). \end{aligned} \quad (6.5.32)$$

This vanishes if

$$Q_\alpha = Zt_\alpha. \quad (6.5.33)$$

Summarizing, we find that the $\Delta_N^{(\epsilon)n}$ transformations (6.5.25) read

$$\begin{aligned} \delta_\theta^{(\epsilon)}\Psi &= i\theta Zt_\alpha\omega_\alpha\Psi, \\ \delta_\theta^{(\epsilon)}A_{\mu\alpha} &= \theta Z(Z'\partial_\mu\omega_\alpha + C_{\alpha\beta\gamma}A_{\mu\beta}\omega_\gamma), \\ \delta_\theta^{(\epsilon)}\omega_\alpha &= -\frac{1}{2}Z\theta C_{\alpha\beta\gamma}\omega_\beta\omega_\gamma. \end{aligned} \quad (6.5.34)$$

Modulo the two (N-dependent) constants Z and Z' , they are like the original BRST transformations (6.3.3). Let us now construct the possible form of the action $\Gamma_N^{(\epsilon)}[\chi]$. It is a local functional, has to contain up to dimension four operators, and has to be invariant under the transformations (6.5.34) and under all the linearly realized symmetries of the original total tree-level action (6.3.2). These are Lorentz invariance, ghost number conservation and the global gauge transformations

$$\begin{aligned} \delta\Psi &= it_\alpha\epsilon_\alpha\Psi, \\ \delta A_\mu &= C_{\alpha\beta\gamma}A_{\mu\beta}\epsilon_\gamma, \\ \delta\omega_\alpha &= C_{\alpha\beta\gamma}\omega_\beta\epsilon_\gamma, \\ \delta\omega_\alpha^* &= C_{\alpha\beta\gamma}\omega_\beta^*\epsilon_\gamma, \\ \delta H_\alpha &= C_{\alpha\beta\gamma}H_\beta\epsilon_\gamma. \end{aligned} \quad (6.5.35)$$

For the gauge fixing choice $f_\alpha = \partial_\mu A_\alpha$, we have an additional symmetry, invariance under anti-ghost shifts:

$$\delta_T \omega_\alpha^* = c_\alpha, \quad (6.5.36)$$

where c_α is an arbitrary Grassmann constant parameter. It is clear that the only dimension four or less terms containing ghosts and compatible with all the above symmetries are $\partial_\mu \omega_\alpha^* \partial^\mu \omega_\alpha$ and $d_{\alpha\beta\gamma} \partial_\mu \omega_\alpha^* A_\gamma^\mu \omega_\beta$. The only terms involving H_α are $H_\alpha H_\alpha$, $H_\alpha \partial_\mu A_\alpha^\mu$ and $e_{\alpha\beta\gamma} H_\alpha A_\beta^\mu A_{\mu\gamma}$. We have $\Gamma_N^{(\epsilon)}[\chi] = \int d^4x \mathcal{L}_N^{(\epsilon)}$, with

$$\begin{aligned} \mathcal{L}_N^{(\epsilon)} = & \mathcal{L}'(\psi, A) + \frac{1}{2} \xi' H_\alpha H^\alpha + c H_\alpha \partial_\mu A_\alpha^\mu - e_{\alpha\beta\gamma} H_\alpha A_\beta^\mu A_{\mu\gamma} \\ & - Z_\omega \partial_\mu \omega_\alpha^* \partial^\mu \omega_\alpha - d_{\alpha\beta\gamma} \partial_\mu \omega_\alpha^* A_\gamma^\mu \omega_\beta. \end{aligned} \quad (6.5.37)$$

In eq.(6.5.37), $\mathcal{L}'(\psi, A)$ is the part of the Lagrangian that depends only on the gauge and matter fields. Let us impose the modified BRST transformations (6.5.34) to $\mathcal{L}_N^{(\epsilon)}$. The terms linear in the H_α field give

$$\begin{aligned} \delta^{(\epsilon)} \mathcal{L}_N^{(\epsilon)}|_{H_\alpha} = & c H_\alpha \theta Z \partial^\mu (Z' \partial_\mu \omega_\alpha + C_{\alpha\beta\gamma} A_{\mu\beta} \omega_\gamma) - 2 e_{\alpha\beta\gamma} H_\alpha A_{\mu\gamma} \theta Z (Z' \partial_\mu \omega_\beta + C_{\beta\rho\sigma} A_{\mu\rho} \omega_\sigma) \\ & + Z_\omega \theta \partial_\mu H_\alpha \partial^\mu \omega_\alpha + d_{\alpha\beta\gamma} \theta \partial_\mu H_\alpha A_\gamma^\mu \omega_\beta \\ = & \theta \partial_\mu H_\alpha \partial^\mu \omega_\alpha (-c Z Z' + Z_\omega) + \theta \partial_\mu H_\alpha A_\gamma^\mu \omega_\beta (c Z C_{\alpha\beta\gamma} + d_{\alpha\beta\gamma}) \\ & - 2 \theta H_\alpha A_\mu^\gamma A_{\mu\rho} \omega_\sigma (e_{\alpha\beta\gamma} Z C_{\beta\rho\sigma}) - 2 \theta H_\alpha A_\mu^\gamma \partial^\mu \omega_\beta (e_{\alpha\beta\gamma} Z Z'). \end{aligned} \quad (6.5.38)$$

This vanishes only if

$$e_{\alpha\beta\gamma} = 0, \quad d_{\alpha\beta\gamma} = -c Z C_{\alpha\beta\gamma}, \quad c = \frac{Z_\omega}{Z Z'}. \quad (6.5.39)$$

The remaining terms are of the form $\partial \omega^* \omega \partial \omega$ and $\partial \omega^* A \omega^2$. They do not give additional constraints. Indeed,

$$\begin{aligned} \delta^{(\epsilon)} \mathcal{L}_N^{(\epsilon)}|_{\partial \omega^* \omega \partial \omega} = & - Z_\omega \partial_\mu \omega_\alpha^* \left(-\frac{1}{2} Z \right) C_{\alpha\beta\gamma} \theta \partial^\mu (\omega_\beta \omega_\gamma) - d_{\alpha\beta\gamma} \partial_\mu \omega_\alpha^* \theta Z Z' \partial^\mu \omega_\gamma \omega_\beta \\ = & \theta \partial_\mu \omega_\alpha^* \partial^\mu \omega_\gamma \omega_\beta Z (Z_\omega C_{\alpha\beta\gamma} + Z' d_{\alpha\beta\gamma}) = 0, \end{aligned} \quad (6.5.40)$$

and

$$\begin{aligned} \delta^{(\epsilon)} \mathcal{L}_N^{(\epsilon)}|_{\partial \omega^* A \omega^2} = & - d_{\alpha\beta\gamma} \partial_\mu \omega_\alpha^* \left(\theta C_{\gamma\rho\sigma} Z A_\rho^\mu \omega_\sigma \omega_\beta + A_\gamma^\mu \left(-\frac{1}{2} \theta Z \right) C_{\beta\rho\sigma} \omega_\rho \omega_\sigma \right) \\ = & \partial_\mu \omega_\alpha^* \theta \omega_\sigma \omega_\beta A_\rho^\mu \left(\frac{Z_\omega Z}{2 Z'} \right) \left(C_{\gamma\rho\sigma} C_{\gamma\alpha\beta} + C_{\gamma\rho\beta} C_{\gamma\sigma\alpha} + C_{\gamma\rho\alpha} C_{\gamma\beta\sigma} \right) = 0. \end{aligned} \quad (6.5.41)$$

Invariance of $\mathcal{L}'(\psi, A)$ under the transformations (6.5.34) is ensured by demanding gauge invariance, with gauge parameter

$$\epsilon_\alpha(x) = Z Z' \theta \omega_\alpha(x), \quad (6.5.42)$$

in terms of rescaled Lie algebra generators and structure constants: $\tilde{t}_\alpha = t_\alpha/Z'$, $\tilde{C}_{\alpha\beta\gamma} = C_{\alpha\beta\gamma}/Z'$. In the canonical basis, such rescalings correspond of course to a renormalization of the gauge coupling constant $\tilde{g} = g/Z'$. Putting all together, we can finally write down the most general form of $\mathcal{L}_N^{(\epsilon)}$:

$$\begin{aligned} \mathcal{L}_N^{(\epsilon)} = & -Z_A \tilde{F}_{\mu\nu}^\alpha \tilde{F}_\alpha^{\mu\nu} - Z_\psi \bar{\psi} \gamma^\mu (\partial_\mu - i \tilde{t}_\alpha A_\mu^\alpha) \psi \\ & + \frac{1}{2} \xi' H_\alpha H_\alpha + \left(\frac{Z_\omega}{Z Z'} \right) H_\alpha \partial^\mu A_\mu^\alpha - Z_\omega (\partial_\mu \omega^\star) (\partial^\mu \omega_\alpha + \tilde{C}_{\alpha\beta\gamma} A_\beta^\mu \omega_\gamma), \end{aligned} \quad (6.5.43)$$

where schematically $\tilde{F} = dA + \tilde{C}A^2$. We notice that this Lagrangian has the same form as the tree level one. Hence, by redefining at level N the fields and couplings in the tree-level action, we can set $\mathcal{L}_N^{(\epsilon)}[\chi, 0] = S[\chi]$, namely $\Gamma_{N,\infty}[\chi, 0] = 0$. This concludes the proof of the renormalizability of non-abelian gauge theories.

Chapter 7

Effective Field Theories*

We have so far mostly considered renormalizable quantum field theories, since these theories have simple renormalization properties. Renormalizable theories are special because they allow us to hide in a redefinition of a finite number of parameters (fixed by experiment) all our ignorance of the UV physics which is beyond the model in consideration. When weakly coupled, these theories also allow us to give very accurate predictions, as in the case of QED. Aside from an improved calculability, however, there is no conceptual reason to focus only on such a restricted class of theories.

Unless one is so ambitious to try to construct the ultimate, possibly, finite theory of everything, any quantum field theory should be seen as an *effective* field theory, namely a theory that is reasonably accurate in a given energy regime and is replaced by some other more complete theory at a given UV scale M . Classical examples are the Fermi theory of electroweak interactions or the pion chiral Lagrangian. These theories are effective, being replaced, at sufficiently high energies, by the SM and QCD, respectively.

The chapter is structured as follows. In section 7.1 we introduce basics of effective field theories and the notion of matching. In section 7.2 we discuss the differences between 1PI and Wilsonian actions. The subsequent sections 7.3, 7.4 and 7.5 provide three simple examples where some aspects of matching between the effective theory and its underlying UV completion are emphasized, notably the sensitivity to the scale M of light scalar masses. In section 7.6 we discuss the notion of naturalness and of hierarchy problem. A significantly more complicated, but more interesting and realistic, example of matching is discussed in section 7.7. In section 7.8 we explain why dimensional regularization with minimal subtraction are the ideal regularization and renormalization scheme to study effective field theories and then in section 7.9 we discuss redundant operators and how to deal with them.

7.1 Matching

Consider a system defined by a UV action S_{UV} and composed by two sets of fields which we schematically denote by ϕ_l and ϕ_h . The fields ϕ_l are by definition light, i.e. their masses are much smaller than a characteristic scale M , while the heavy fields have mass of order M or higher. If we are interested in processes happening at a given scale $E \ll M$ and involving light fields only, the heavy states cannot be produced and can be then integrated out, in the spirit of Wilson's approach to the RG evolution in QFT. In a path integral approach, the complete 1PI generating functional of light fields is given by

$$e^{i\Gamma(\phi_l)} = \int_{1PI} \mathcal{D}\eta_l e^{iS_{\text{IR}}(\phi_l + \eta_l)}, \quad e^{iS_{\text{IR}}(\phi_l)} = \int \mathcal{D}\phi_h e^{iS_{\text{UV}}(\phi_l, \phi_h)}, \quad (7.1.1)$$

where S_{IR} represents the underlying IR effective action. At tree-level, we simply have

$$S_{\text{IR}}^{(0)}(\phi_l) = S_{\text{UV}}(\phi_l, \phi_h(\phi_l)), \quad (7.1.2)$$

where $\phi_h(\phi_l)$ is the classical solution to the heavy field equations of motion. Being the integrated fields heavy, the action admits a well-defined momentum expansion in terms of local operators. We could proceed by computing the one-loop and higher loop effective actions $S_{\text{IR}}^{(l)}$ and reconstruct in this way the whole effective action S_{IR} . The key point of effective field theory is to *replace* this procedure by a simpler one, where we compute radiative effects involving the light fields starting from the action

$$S_{\text{EFT}}(\phi_l) = S_{\text{IR}}^{(0)}(\phi_l) + \Delta S(\phi_l), \quad (7.1.3)$$

where ΔS encodes all higher dimensional local operators up to some order, depending on the accuracy we want to achieve, and compatibly with the global symmetries of the theory. In an effective field theory the actual expansion parameter is E/M , and the higher dimensional an operator is, the smaller is its contribution. Hence, truncating the effective action up to some order in the dimension of the operators corresponds to work up to some order in E/M .

All the unknown couplings multiplying the higher dimensional local operators are fixed by comparing given quantities as computed in the UV theory and in the EFT. In principle, one might want to directly compare S -matrix elements, but it is often easier to compare 1PI functions, generically off-shell. Note that 1PI functions can be compared at an unphysical point, since by analyticity we are ensured that their on-shell properties are equal. However, this is somewhat improper and too conservative, because physical observables correspond to on-shell S matrix elements and not to the whole Green functions, and different off-shell Green functions can lead to the same on-shell results, as we will see in section 7.9. We will

nevertheless require the off-shell equality of 1PI functions in our examples for simplicity. Note also that in the UV theory 1PI refers to the light fields *only*. Graphs that can be splitted in two disconnected ones by cutting a propagator of a heavy field should be included.

The procedure of fixing the parameters of the IR theory using the UV theory is called “matching”. Of course, matching requires to also perform a computation in the full theory but, as we will see, it is typically much easier to do computations in the EFT and then matching, rather than computing everything in the full theory.

7.2 1PI vs Wilsonian Actions*

Before considering various examples of use of effective field theories techniques, we would like here to emphasize the difference between the various effective actions introduced so far.¹ The effective action $\Gamma(\phi)$ defined in eq.(4.1.10) is the generator of all 1PI amplitudes in the theory, namely those involving an arbitrary combination of heavy and light fields. In the context of EFT, we also have $\Gamma(\phi_l)$ appearing in the l.h.s. of the first equation in (7.1.1), which generates the 1PI amplitudes associated to the light fields. Another effective action is S_{IR} , defined in the l.h.s. of the second equation in (7.1.1). The effective actions $\Gamma(\phi)$ and $\Gamma(\phi_l)$ are shortly denoted 1PI actions, while S_{IR} is called a Wilsonian effective action. The key difference between 1PI and Wilsonian actions is that in the latter the virtual particles are always massive. Correspondingly, while 1PI actions are intrinsically non-local, namely they cannot be written as the integral of some effective Lagrangian density (even with an infinite number of terms), the Wilsonian action always admits such rewriting. The key point can be understood at tree-level. When we integrate out a scalar particle of mass M at tree-level, roughly speaking we get an effective action term that reads in momentum space as

$$S_{\text{IR}} \supset \int d^4p \Phi_1(-p) \frac{1}{p^2 - M^2} \Phi_2(p), \quad (7.2.1)$$

where $\Phi_{1,2}$ are local functionals of the remaining light fields and $1/(p^2 - M^2)$ represents the propagator of the massive particle that has been integrated out. The term is manifestly non-local and cannot be written in configuration space as the integral of a Lagrangian density. At low energies, $p^2 \ll M^2$, we can expand the propagator as

$$\frac{1}{p^2 - M^2} = -\frac{1}{M^2} - \frac{1}{M^2} \frac{p^2}{M^2} + \dots \quad (7.2.2)$$

¹See 4.3 for a further subtlety.

Plugging eq.(7.2.2) in eq.(7.2.1), however, gives a (generally infinite) sum of local terms of the form

$$S_{\text{IR}} \supset -\frac{1}{M^2} \int d^4x \Phi_1(x) \left(1 + \frac{\square}{M^2} + \dots\right) \Phi_2(x). \quad (7.2.3)$$

On the contrary, when massless fields are present, the term in eq.(7.2.1) is intrinsically non-local and eq.(7.2.3) does not apply.

7.3 Two Scalars*

Consider

$$\mathcal{L}_{UV} = \frac{1}{2}(\partial H)^2 - \frac{1}{2}M^2 H^2 + \frac{1}{2}(\partial L)^2 - \frac{1}{2}m^2 L^2 - \frac{g}{2}HL^2, \quad (7.3.1)$$

where $m \ll M$ and the dimensionful coupling g is assumed to be of $\mathcal{O}(M)$.² The most general EFT for L reads as follows:

$$\mathcal{L}_{IR} = \frac{1}{2}Z_L(\partial L)^2 - \frac{1}{2}\tilde{m}^2 L^2 - \frac{\lambda}{4!}L^4 + \text{h.d.o.}, \quad (7.3.2)$$

where h.d.o. stands for higher dimensional operators. The tree-level integration of H gives $H = -gL^2/(2M^2) + \mathcal{O}(p^2/M^4)$. Plugged back in eq. (7.3.1) gives

$$\mathcal{L}_{IR}^{(0)} = \frac{1}{2}(\partial L)^2 - \frac{1}{2}m^2 L^2 - \frac{\lambda}{4!}L^4 + \mathcal{O}(g^2 M^{-4}) \quad (7.3.3)$$

so that tree-level matching gives

$$Z_L = 1 + \mathcal{O}(g^2), \quad \tilde{m}^2 = m^2 + \mathcal{O}(g^2), \quad \lambda = -\frac{3g^2}{M^2} + \mathcal{O}(g^4). \quad (7.3.4)$$

It is important to emphasize that the factors Z_L , $\tilde{m}^2$, etc. in eq. (7.3.4) are the analogues of $(1+\delta Z)$, $m^2+\delta m^2$, etc. defined in eq. (5.1.5) in the Wilsonian RG flow. They are finite terms, which at the quantum level indicate how the EFT fields, masses etc. differ from their UV counterparts to compensate for the different UV behaviour of the two theories, so that low-energy observables match in the two descriptions. They should not be confused with the usual counter-terms needed to subtract divergences, which we will not write in this chapter, fixing them by an $\overline{\text{MS}}$ subtraction scheme.

We can directly use $\mathcal{L}_{IR}^{(0)}$ as effective theory to compute Green functions for energies much smaller than M . For instance, the $\overline{\text{MS}}$ renormalized four-point 1PI function reads (see eq. (5.4.4)):

$$\begin{aligned} \Gamma^{(4)}(s, t, u) = & -\lambda(\mu) + \frac{\lambda^2(\mu)}{32\pi^2} \int_0^1 dx \left[\log \frac{\mu^2}{m^2 - sx(1-x)} + (s \rightarrow t) + (s \rightarrow u) \right] \\ & + \mathcal{O}\left(\frac{s}{M^2}, \frac{t}{M^2}, \frac{u}{M^2}\right), \end{aligned} \quad (7.3.5)$$

²This theory is not realistic, since the cubic potential HL^2 is not positive definite, but the considerations we want to make are insensitive to this stability problem.

where we neglect the $\mathcal{O}(E^2/M^2)$ terms coming from dimension six (or higher) operators. Expression (7.3.5) is valid in the limit in which $E/M \ll \lambda \ll 1$. By taking $s \simeq t \simeq u \simeq \mu$ we get $\Gamma^{(4)}(\mu) \simeq -\lambda(\mu)$, where

$$\lambda^{-1}(\mu) = \lambda^{-1}(\mu_0) - \frac{3}{16\pi^2} \log\left(\frac{\mu}{\mu_0}\right) \quad (7.3.6)$$

is the usual one-loop running in the ϕ^4 theory, assuming $\mu, \mu_0 > m$. The natural scale where the effective theory parameter λ should be matched, at one-loop level, with the underlying UV parameters g and M is at the scale $\mu_0 = M$, since this is the energy scale boundary between the two theories. This is understood by comparing the 1PI four point amplitude $\Gamma^{(4)}$. The following diagrams contribute to $\Gamma^{(4)}$ in the UV theory (continuous and dashed lines correspond to light and heavy fields, respectively). At leading order in an $1/M$ expansion and for zero external momentum, we get

$$\begin{aligned} & \text{Diagram 1: } \text{---} \bullet \text{---} \bigcirc \text{---} \bullet \text{---} \quad + \text{ perms.} = \left(\frac{1}{2} \times 3\right) (-ig)^4 \mu^{2\epsilon} \int \frac{d^d k}{(2\pi)^d} \left(\frac{i}{k^2 - m^2}\right)^2 \left(\frac{-i}{M^2}\right)^2 \\ & \quad = \frac{-3ig^4}{32\pi^2 M^4} \log \frac{m^2}{\mu^2}, \\ & \text{Diagram 2: } \text{---} \bullet \text{---} \text{---} \bullet \text{---} \quad + \text{ perms.} = (6) (-ig)^4 \mu^{2\epsilon} \int \frac{d^d k}{(2\pi)^d} \left(\frac{i}{k^2 - m^2}\right)^2 \frac{-i}{M^2} \frac{i}{k^2 - M^2} \\ & \quad = \frac{-3ig^4}{8\pi^2 M^4} \left(1 + \log \frac{m^2}{M^2}\right), \\ & \text{Diagram 3: } \text{---} \bullet \text{---} \text{---} \bullet \text{---} \quad + \text{ perms.} = (6) (-ig)^4 \mu^{2\epsilon} \int \frac{d^d k}{(2\pi)^d} \left(\frac{i}{k^2 - m^2}\right)^2 \left(\frac{i}{k^2 - M^2}\right)^2 \\ & \quad = \frac{-6ig^4}{8\pi^2 M^4} \left(1 + \frac{1}{2} \log \frac{m^2}{M^2}\right), \end{aligned} \quad (7.3.7)$$

where in parenthesis we have denoted the geometrical factor (permutations included) of the graph. Recall that $\Gamma^{(4)}$ is 1PI in the light fields only, and hence the first graph in eq.(7.3.7), reducible when cutting the propagator of the heavy field, should be included in the computation.

Including the tree-level term, we get

$$\Gamma^{(4)}(0) = \frac{3g^2(\mu)}{M^2} - \frac{3g^4(\mu)}{8\pi^2 M^4} \left(3 + 2 \log \frac{m^2}{M^2} + \frac{1}{4} \log \frac{m^2}{\mu^2}\right). \quad (7.3.8)$$

Comparing eq. (7.3.5) (at zero momentum) with eq. (7.3.8) and using the tree-level relation (7.3.4) for λ , we get

$$\lambda(\mu) = -\frac{3g^2(\mu)}{M^2} + \frac{3g^4}{8\pi^2 M^4} \left(3 + 2 \log \frac{\mu^2}{M^2}\right). \quad (7.3.9)$$

Notice how all the (potentially large) logs involving the light mass m^2 cancelled from eq. (7.3.9). This is not a coincidence. All the IR properties of the UV theory (as well as large IR effects) are reliably described by the EFT by construction, so that the matching equations are always regular in the IR. It is quite clear from eq. (7.3.9) that the best scale to match the two theories is $\mu = M$, in which case we have

$$\lambda(M) = -\frac{3g^2(M)}{M^2} + \frac{9g^4}{8\pi^2 M^4}. \quad (7.3.10)$$

Taking $\mu_0 = M$ in eq. (7.3.6) and using eq. (7.3.10), allows us to take into account possible large logs of the form $\log(\mu/M)$. Note that the g^4/M^4 term in eq. (7.3.10) can safely be neglected at one-loop level, starting to be relevant only at two-loop level, if we want to resum NLL terms. In other words, no one-loop computation of the 1PI 4-point function would have been needed in matching the theory at $\mu \simeq M$ and we would have been able to resum the leading $\log(\mu/M)$ terms without performing any radiative computation in the UV theory! In general, an l -loop computation in the effective theory requires an $l - 1$ -loop matching computation in the UV theory. This example shows the usefulness of effective theories even in a simple set-up.

7.4 Yukawa Theory I: Heavy Scalar, Light Fermion*

Consider a light fermion field coupled to a massive scalar by means of a Yukawa-like coupling. The UV and IR Lagrangians are

$$\mathcal{L}_{UV} = \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}M^2\phi^2 + \bar{\psi}(i\not{\partial} - m)\psi - g\phi\bar{\psi}\psi \quad (7.4.1)$$

$$\mathcal{L}_{IR} = Z_\psi\bar{\psi}i\not{\partial}\psi - \tilde{m}\bar{\psi}\psi + \frac{\lambda}{2}\bar{\psi}\psi\bar{\psi}\psi + \text{h.d.o.} \quad (7.4.2)$$

The tree-level integration of ϕ gives $\phi = -g\bar{\psi}\psi/M^2 + \mathcal{O}(p^2/M^4)$. Plugging back in eq. (7.4.1) gives

$$Z_\psi = 1 + \delta Z_\psi, \quad \tilde{m} = m + \delta m, \quad \lambda = \frac{g^2}{M^2}. \quad (7.4.3)$$

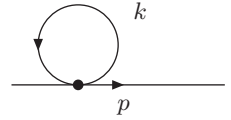
Let us now compute the 1PI 2-point function $\Gamma^{(2)}(p)$ in both the UV and the IR EFT at one-loop level. In the UV theory the relevant one-loop diagram is

$$\begin{aligned} \text{Diagram: } \text{---} \bullet \xrightarrow{p} \bullet \xrightarrow{p+k} \text{---} \quad \text{with a dashed loop from } \bullet \text{ to } \bullet \text{ labeled } k \quad &= (-ig)^2 \mu^\epsilon \int \frac{d^d k}{(2\pi)^d} \frac{i}{k^2 - M^2} \frac{i(\not{p} + \not{k} + m)}{(k+p)^2 - m^2} \\ &= \frac{ig^2}{16\pi^2} \int_0^1 dx \left(\not{p}(1-x) + m \right) \log \left(\frac{\mu^2}{-p^2 x(1-x) + m^2 x + (1-x)M^2} \right). \end{aligned} \quad (7.4.4)$$

Expanding for $M \gg m, p$, we get

$$\begin{aligned}\Gamma^{(2)}(p) &= \not{p} - m - \frac{g^2}{16\pi^2} \left[\not{p} \left(\frac{1}{2} \log \frac{M^2}{\mu^2} - \frac{1}{4} + \frac{m^2}{2M^2} - \frac{p^2}{6M^2} + \mathcal{O}\left(\frac{1}{M^4}\right) \right) \right. \\ &\quad \left. + m \left(\log \frac{M^2}{\mu^2} - 1 - \frac{m^2}{M^2} \log \frac{m^2}{M^2} - \frac{p^2}{2M^2} + \mathcal{O}\left(\frac{1}{M^4}\right) \right) \right].\end{aligned}\quad (7.4.5)$$

In the effective theory, the relevant one-loop diagram is



$$= (-1)(i\lambda\mu^\epsilon) \int \frac{d^d k}{(2\pi)^d} \left(\frac{i(\not{k} + m)}{k^2 - m^2} + \frac{i\text{Tr}(\not{k} + m)}{k^2 - m^2} \right) \quad (7.4.6)$$

$$= -\frac{5i\lambda m^3}{16\pi^2} \left(1 + \log \frac{\mu^2}{m^2} \right), \quad (7.4.7)$$

where the two terms in square brackets indicate the two different contractions among the four fermions in $\bar{\psi}\psi\bar{\psi}\psi$, and we have taken $\text{Tr } I = 4$. The 1PI 2-point function $\Gamma^{(2)}(p)$ in the IR reads then

$$\Gamma^{(2)}(p) = (1 + \delta Z_\psi)\not{p} - (m + \delta_m) - \frac{5\lambda m^3}{16\pi^2} \left(1 + \log \frac{\mu^2}{m^2} \right). \quad (7.4.8)$$

A comparison between eqs.(7.4.5) and (7.4.8) shows that a complete matching cannot be performed, because eq.(7.4.5) contains p^2/M^2 terms which do not appear in eq.(7.4.8). Indeed, a systematic matching of $\Gamma^{(2)}(p)$ in the two theories would require the addition in the IR theory of higher dimensional operators of the form $\bar{\psi}\not{p}\square\psi$, $\bar{\psi}\square\psi$, etc. In order to simplify the analysis, we neglect them, their effect being suppressed by powers of p^2/M^2 times a loop factor. Matching the remaining terms in eqs. (7.4.5) and (7.4.8) gives

$$\begin{aligned}\delta Z_\psi(\mu) &= -\frac{g^2}{16\pi^2} \left(\frac{1}{2} \log \frac{M^2}{\mu^2} - \frac{1}{4} + \frac{m^2}{2M^2} + \mathcal{O}\left(\frac{1}{M^4}\right) \right), \\ \delta_m(\mu) &= -\frac{5\lambda m^3}{16\pi^2} \left(1 + \log \frac{\mu^2}{m^2} \right) + \frac{mg^2}{16\pi^2} \left(\log \frac{M^2}{\mu^2} - 1 - \frac{m^2}{M^2} \log \frac{m^2}{M^2} + \mathcal{O}\left(\frac{1}{M^4}\right) \right) \\ &= -\frac{mg^2}{16\pi^2} \left(1 + \log \frac{\mu^2}{M^2} \right) \left(1 + \frac{m^2}{M^2} \right) - \frac{4m^3 g^2}{16\pi^2 M^2} \left(1 + \log \frac{\mu^2}{m^2} \right),\end{aligned}\quad (7.4.9)$$

where in the last line we have used the tree-level relation (7.4.3) for λ . It is clear from eq. (7.4.9) that the matching is best performed at $\mu = M$. We can now determine the physical fermion mass, defined as

$$\Gamma^{(2)}(\not{p} = m_{\text{phys}}) = 0. \quad (7.4.10)$$

At one-loop level, we get

$$m_{\text{phys}} = m(M) \left[1 - \frac{g^2}{16\pi^2} \left(\frac{5}{4} + \mathcal{O}\left(\frac{m^2}{M^2}\right) \right) \right]. \quad (7.4.11)$$

7.5 Yukawa Theory II: Heavy Fermion, Light Scalar*

Consider now the opposite case with an heavy fermion field coupled to a light scalar:

$$\mathcal{L}_{UV} = \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 + \bar{\psi}(i\not{\partial} - M)\psi - g\phi\bar{\psi}\psi \quad (7.5.1)$$

$$\mathcal{L}_{IR} = \frac{1}{2}Z_\phi(\partial\phi)^2 - \frac{1}{2}\tilde{m}^2\phi^2 + \frac{\lambda_3}{3!}\phi^3 + \frac{\lambda_4}{4!}\phi^4 + \text{h.d.o.} \quad (7.5.2)$$

The tree-level integration of ψ gives $\psi = 0$, meaning that classically a heavy fermion field can simply be neglected. This is a general property of fermions. Lorentz invariance does not allow a fermion line to break, so a heavy fermion either mixes with some light fermion by means of Yukawa interactions, or the heavy fermion line should close in loops. We are clearly here in the second case, having only one fermion species in the spectrum. The parameters of the IR Lagrangian (7.5.2) read then

$$Z_\phi = 1 + \delta Z_\phi, \quad \tilde{m}^2 = m^2 + \delta_{m^2}, \quad \lambda_3 = \mathcal{O}(g^3), \quad \lambda_4 = \mathcal{O}(g^4). \quad (7.5.3)$$

Let us again match the 1PI 2-point function in the UV and in the IR theory. In the UV we have

$$\begin{aligned}
\text{---}\!\!\!\rightarrow \bullet \begin{array}{c} \text{\scriptsize k} \\ \curvearrowright \\ \text{\scriptsize $p+k$} \end{array} \bullet \text{---}\!\!\!\rightarrow &= (-1)(-ig)^2 \mu^\epsilon \text{Tr} \int \frac{d^d k}{(2\pi)^d} \frac{i(\not{k} + M)}{k^2 - M^2} \frac{i(\not{p} + \not{k} + M)}{(k+p)^2 - M^2} \\
&= -\frac{4ig^2}{16\pi^2} \int_0^1 dx \left(M^2 - p^2 x(1-x) \right) \left[3 \log \left(\frac{\mu^2}{-p^2 x(1-x) + M^2} \right) - 5 \right], \quad (7.5.4)
\end{aligned}$$

taking $\text{Tr } I = 4$. Expanding for $M \gg m, p$, we get

$$\begin{aligned} \Gamma^{(2)}(p) &= p^2 - m^2 - \frac{4g^2}{16\pi^2} \left[p^2 \left(-\frac{1}{2} \log \frac{M^2}{\mu^2} + \frac{4}{3} - \frac{p^2}{20M^2} + \mathcal{O}\left(\frac{1}{M^4}\right) \right) \right. \\ &\quad \left. + M^2 \left(-3 \log \frac{M^2}{\mu^2} - 5 + \mathcal{O}\left(\frac{1}{M^4}\right) \right) \right]. \end{aligned} \quad (7.5.5)$$

In the EFT we simply have

$$\Gamma^{(2)}(p) = p^2 - m^2 + p^2 \delta Z_\phi - \delta m^2. \quad (7.5.6)$$

As in section 7.4, a complete matching between eqs.(7.5.5) and (7.5.6) require the addition in the IR theory of higher dimensional operators of the form $(\Box\phi)^2$, etc. In order to simplify the analysis, we neglect these operators, their effect being suppressed by powers of p^2/M^2

times a loop factor. Matching eqs. (7.5.5) and (7.5.6) gives

$$\begin{aligned}\delta Z_\phi(\mu) &= -\frac{4g^2}{16\pi^2} \left(-\frac{1}{2} \log \frac{M^2}{\mu^2} + \frac{4}{3} \right), \\ \delta_{m^2}(\mu) &= -\frac{4g^2 M^2}{16\pi^2} \left(3 \log \frac{M^2}{\mu^2} + 5 \right).\end{aligned}\tag{7.5.7}$$

The physical scalar mass, defined as

$$\Gamma^{(2)}(p^2 = m_{\text{phys}}^2) = 0,\tag{7.5.8}$$

is then given by

$$m_{\text{phys}}^2 = m^2(M) + \frac{g^2}{16\pi^2} M^2 \left(-20 + \mathcal{O}\left(\frac{m^2}{M^2}\right) \right).\tag{7.5.9}$$

7.6 Naturalness and the Hierarchy Problem*

The scalar mass correction in eq. (7.5.9) is of $\mathcal{O}(M^2 g^2 / (16\pi^2))$. For $M \gg m$, the ratio $(M/m)^2$ can overcome the one-loop suppression so that the radiative term is generically much bigger than the tree-level one. If we want to keep the physical scalar mass small, a fine-tuning between the $\overline{\text{MS}}$ mass term $m^2(M)$ and the one-loop correction is needed (of course this readjustment is needed at each order in perturbation theory). This is not a conceptual problem, and there is nothing wrong in principle to do that, but it is nevertheless unpleasant to have in the EFT a parameter that is so sensitive to the UV physics. Even a small change in the UV theory (say, $g \rightarrow g + \delta g$ in our toy example) will give rise to a large radiative correction to m_{phys}^2 , and to a new readjustment. This is the typical example of an “hierarchy problem” or “naturalness problem”, namely the problem of explaining why a parameter in an EFT Lagrangian is much smaller than its expected value at the quantum level.³ Natural parameters are often defined as follows:

dimensionless couplings should be of order one and dimensionful couplings should be of order of the largest mass scale in the theory, to the appropriate power. Exceptions arise if a symmetry is restored when a coupling (dimensionless or not) vanishes, in which case it is natural to have that coupling arbitrarily small.

Let us check, using our results above, that this definition makes sense. In the scalar theory of section 7.3, when $g = \mathcal{O}(M)$ (natural value), we get that the dimensionless

³The hierarchy problem is often defined as due to quadratically (or more generally power-like) divergent graphs in regularizations with some cut-off Λ . Although this argument is essentially correct, it relies on the use of Λ and, taken as it is, does not apply to regularizations like DR, where no Λ is introduced and there are no quadratic divergences. It would then seem that the hierarchy problem is “scheme dependent” and invisible in DR, whereas of course it is not, as we have shown in section 7.5.

coupling λ in the EFT is of $\mathcal{O}(M^2/M^2) = \mathcal{O}(1)$. Contrary to the scalar case, the fermion mass correction in eq. (7.4.11) is proportional to the mass itself. This implies that, for $m \rightarrow 0$, all loop corrections vanish as well. Hence, light fermion masses are natural. According to our general definition of natural parameters, some symmetry should be restored when $m \rightarrow 0$. This is indeed what happens, the symmetry being the chiral symmetry $\psi \rightarrow \gamma_5 \psi$ (combined with the discrete $\mathbf{Z}_2$ inversion symmetry $\phi \rightarrow -\phi$ in the UV theory). No symmetry is instead restored when the scalar mass term goes to zero, so that in this case we have a naturalness problem.

Despite hierarchy problems do not lead to real inconsistencies and we can live with them, they have been the main driving force in going beyond the SM, since the Higgs boson mass term is unnatural. Similarly, one of the main problems in theoretical physics today is provided by another hierarchy problem, which is why the vacuum energy we measure is so smaller than its natural value.

7.7 Non-Leptonic Decays**

As a final example, we consider a commonly used EFT of the weak interactions, obtained by integrating out the massive spin one bosons W and Z . Our aim is to study the QCD corrections to the non-leptonic decays of the K mesons, for which the relevant charged electroweak currents are

$$\mathcal{L}_{SM} \supset \frac{g}{\sqrt{2}} W_\mu^+ \left(V_{ud} \bar{u}_L \gamma^\mu d_L + V_{us} \bar{u}_L \gamma^\mu s_L \right) + h.c. \quad (7.7.1)$$

Integrating out the W gives, among others, the flavour changing four-fermion interaction term

$$- \frac{4G_F}{\sqrt{2}} V_{us} V_{du}^* (\bar{u}_L \gamma^\mu s_L) (\bar{d}_L \gamma_\mu u_L), \quad (7.7.2)$$

with the tree-level matching relation

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}. \quad (7.7.3)$$

On the other hand, the most general EFT dimension six interaction terms that can contribute to K decays are

$$\mathcal{L}_{IR} \supset -c_1 \mathcal{O}_1 - c_2 \mathcal{O}_2 + h.c., \quad (7.7.4)$$

where

$$\mathcal{O}_1 = (\bar{u}_L \gamma^\mu s_L) (\bar{d}_L \gamma_\mu u_L), \quad \mathcal{O}_2 = (\bar{u}_L \gamma^\mu u_L) (\bar{d}_L \gamma_\mu s_L). \quad (7.7.5)$$

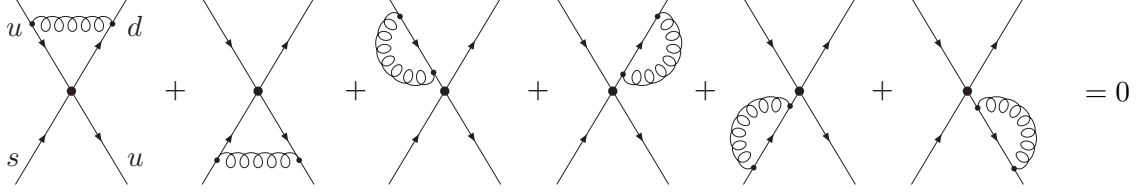


Figure 7.1: Set of one-loop graphs contributing to the renormalization of the four-fermion operator $\mathcal{O}_1$ that we do not need to consider, because of current conservation. The flavor labels are the same for all graphs. The set of one-loop graphs contributing to $\mathcal{O}_2$ is obtained by exchanging the upper d and the lower u indices, where a similar cancellation occurs.

Tree-level matching gives

$$c_1(m_W) = \frac{g^2}{2m_W^2} V_{us} V_{du}^*(m_W), \quad c_2(m_W) = 0. \quad (7.7.6)$$

Although at tree-level only $\mathcal{O}_1$ appears, symmetry arguments do not forbid $\mathcal{O}_2$, which is expected to arise at the quantum level. The natural scale where the operators $\mathcal{O}_{1,2}$ should be renormalized is of order m_K , being m_K the scale of the decay process in the particle rest frame. It is important to compute the RG flow of the couplings $c_i = c_i(\mu)$, $i = 1, 2$, and see its effect. The main contribution is given by the QCD corrections. In what follows we then compute the QCD effects on the RG flow of c_1 and c_2 .

Recall that at the linear level, the RG evolution of the couplings $c_{1,2}$ is fixed by the anomalous dimensions of the corresponding operators (see eq. (5.10.9)):

$$\mu \frac{dc_i}{d\mu} = \beta_i = \gamma_{ij} c_j, \quad (7.7.7)$$

where γ_{ij} is the 2×2 matrix of anomalous dimensions of the composite operators $\mathcal{O}_1$ and $\mathcal{O}_2$ which, as we will see, mix under renormalization. As discussed in section 5.9, we determine γ_{ij} by looking at the Green functions $G_i^{(4,1)} \equiv \langle \bar{u}_L u_L \bar{s}_L d_L \mathcal{O}_i \rangle$. Since the two operators that mix are made of the same constituent fields and differ only by their different contractions, it will be useful to multiply $G_i^{(4,1)}$ by the wave functions of the external fermion fields and denote them by the same symbol of the corresponding fermion field u_L , d_L and s_L . In this way we will more easily disentangle $\mathcal{O}_1$ from $\mathcal{O}_2$. The computation is simplified by

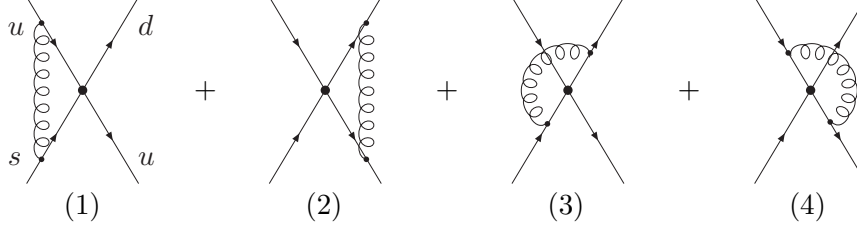


Figure 7.2: Set of one-loop graphs contributing to the renormalization of the four-fermion operator $\mathcal{O}_1$ that should be computed. The flavor labels are the same for all graphs. The set of one-loop graphs contributing to $\mathcal{O}_2$ is obtained by exchanging the upper d and the lower u indices.

setting to zero the external momenta and the quark masses, and choosing the Feynman gauge $\xi = 1$ for the gluon propagator. The relevant one-loop graphs renormalizing $\mathcal{O}_1$ are depicted in figs. 7.1 and 7.2. The order of the fermion arrows is $u - d$ and $s - u$. A similar set of graphs, with $d \leftrightarrow u$ in the external lines, renormalizes $\mathcal{O}_2$.

The leading terms in the EFT corresponds to QCD with $n_f = 3$ quark species, neglecting the other massive quarks, that can also be integrated out. In the limit of vanishing quark masses and electric charge, this theory has an $SU(n_f)$ global symmetry rotating the quark fields, and correspondingly a set of conserved currents given by $J_\mu^\alpha = \bar{q}\gamma_\mu t^\alpha q$, where q is the quark multiplet and t^α are the $SU(n_f)$ generators ($a = 1, \dots, n_f^2 - 1$). All the quark bilinears entering the definition of $\mathcal{O}_{1,2}$ are different components of this set of conserved currents. Since the latter have vanishing anomalous dimensions, the diagrams in fig. 7.1 precisely compensates each other, in analogy to what we have seen in the single fermion case in section 5.9, eq.(5.9.21). We should then only focus on the one-loop diagrams depicted in fig. 7.2. Keeping as usual only the divergent terms, diagram (1) gives

$$\begin{aligned}
 (1) &= (ig_c)^2 \mu^\epsilon \int \frac{d^d k}{(2\pi)^d} \left(\bar{d}_L \gamma^\mu \frac{i \not{k}}{k^2} \gamma^\nu t^a u_L \right) \left(\bar{u}_L \gamma_\mu \frac{-i \not{k}}{k^2} \gamma_\nu t^a s_L \right) \frac{-i}{k^2} \\
 &= -\frac{g_c^2}{32\pi^2 \epsilon} \left(\bar{d}_L \gamma^\mu \gamma^\rho \gamma^\nu t^a u_L \right) \left(\bar{u}_L \gamma_\mu \gamma_\rho \gamma_\nu t^a s_L \right).
 \end{aligned} \tag{7.7.8}$$

Similarly, we get

$$(2) = (1), \quad (3) = (4) = \frac{g_c^2}{32\pi^2 \epsilon} \left(\bar{d}_L \gamma^\mu \gamma^\rho \gamma^\nu t^a u_L \right) \left(\bar{u}_L \gamma_\nu \gamma_\rho \gamma_\mu t^a s_L \right). \tag{7.7.9}$$

Eqs. (7.7.8) and (7.7.9) are best written using certain identities in both color and spinor

indices, identities that are derived in detail in subsection 7.7.1. Using the results in that section, we have

$$t_{ij}^a t_{kl}^a = \frac{1}{2} \left(\delta_{il} \delta_{jk} - \frac{1}{3} \delta_{ij} \delta_{kl} \right), \quad (7.7.10)$$

$$\psi_{2L} \bar{\psi}_{1L} = -\frac{1}{2} (\bar{\psi}_{1L} \gamma^\mu \psi_{2L}) \gamma_\mu P_R. \quad (7.7.11)$$

Starting from eq.(7.7.8), we first use eq.(7.7.11) to write

$$\begin{aligned} K \equiv & \left(\bar{d}_L^i \gamma^\mu \gamma^\rho \gamma^\nu t_{ij}^a u_L^j \right) \left(\bar{u}_L^k \gamma_\mu \gamma_\rho \gamma_\nu t_{kl}^a s_L^l \right) = -\frac{1}{2} \left(\bar{d}_L^i \gamma^\mu \gamma^\rho \gamma^\nu \gamma^\alpha P_R \gamma_\mu \gamma_\rho \gamma_\nu s_L^l \right) t_a^{kl} t_a^{ij} \\ & \times \left(\bar{u}_L^k \gamma_\alpha u_L^j \right) = 16 \left(\bar{u}_L^k \gamma_\alpha u_L^j \right) \left(\bar{d}_L^i \gamma^\alpha s_L^l \right) t_a^{ij} t_a^{kl}, \end{aligned} \quad (7.7.12)$$

where in the second equality we have used the gamma matrix algebra relation

$$\gamma^\mu \gamma^\rho \gamma^\nu \gamma^\alpha \gamma_\mu = -2 \gamma^\alpha \gamma^\nu \gamma^\rho. \quad (7.7.13)$$

We can now apply eq.(7.7.10) to get

$$\begin{aligned} K = & 16 \left(\bar{u}_L^k \gamma^\alpha u_L^j \right) \left(\bar{d}_L^i \gamma_\alpha s_L^l \right) \frac{1}{2} \left(\delta_{il} \delta_{jk} - \frac{1}{3} \delta_{ij} \delta_{kl} \right) = 8 \left(\bar{u}_L \gamma^\alpha u_L \right) \left(\bar{d}_L \gamma_\alpha s_L \right) \\ & - \frac{8}{3} \left(\bar{u}_L^k \gamma^\alpha u_L^i \right) \left(\bar{d}_L^i \gamma_\alpha s_L^k \right). \end{aligned} \quad (7.7.14)$$

Applying again eq.(7.7.11) to the second term in eq.(7.7.14) gives

$$\begin{aligned} K = & 8 \left(\bar{u}_L \gamma^\mu u_L \right) \left(\bar{d}_L \gamma_\mu s_L \right) - \frac{8}{3} \left(-\frac{1}{2} \right) \left(\bar{d}_L \gamma^\mu u_L \right) \left(\bar{u}_L \gamma^\alpha \gamma_\mu P_R \gamma_\alpha s_L \right) \\ = & 8 \left(\bar{u}_L \gamma^\mu u_L \right) \left(\bar{d}_L \gamma_\mu s_L \right) - \frac{8}{3} \left(\bar{d}_L \gamma^\mu u_L \right) \left(\bar{u}_L \gamma_\mu s_L \right), \end{aligned} \quad (7.7.15)$$

where we used the relation

$$\gamma^\alpha \gamma_\mu \gamma_\alpha = -2 \gamma_\mu. \quad (7.7.16)$$

The second row in eq.(7.7.17) is our final expression of K . As can be seen, we have started with the four fermion operator $\mathcal{O}_1$ with the schematic form $(\bar{u}s)(\bar{d}u)$ and ended up with having generated also the form $(\bar{u}u)(\bar{d}s)$, corresponding to the operator $\mathcal{O}_2$. Identical manipulations can be performed starting from the term (3) in eq.(7.7.9) and will not be reported. Summarizing, we have

$$\begin{aligned} \left(\bar{d}_L \gamma^\mu \gamma^\rho \gamma^\nu t^a u_L \right) \left(\bar{u}_L \gamma_\mu \gamma_\rho \gamma_\nu t^a s_L \right) &= 8 (\bar{d}_L \gamma^\mu s_L) (\bar{u}_L \gamma_\mu u_L) - \frac{8}{3} (\bar{d}_L \gamma^\mu u_L) (\bar{u}_L \gamma_\mu s_L), \\ \left(\bar{d}_L \gamma^\mu \gamma^\rho \gamma^\nu t^a u_L \right) \left(\bar{u}_L \gamma_\nu \gamma_\rho \gamma_\mu t^a s_L \right) &= 2 (\bar{d}_L \gamma^\mu s_L) (\bar{u}_L \gamma_\mu u_L) - \frac{2}{3} (\bar{d}_L \gamma^\mu u_L) (\bar{u}_L \gamma_\mu s_L). \end{aligned} \quad (7.7.17)$$

Collecting all terms, the counterterms δ_{11} and δ_{21} needed to remove the above divergences read

$$\delta_{11} = -2 \frac{\alpha_c}{4\pi} \frac{1}{\epsilon}, \quad \delta_{21} = 6 \frac{\alpha_c}{4\pi} \frac{1}{\epsilon}, \quad (7.7.18)$$

where $\alpha_c \equiv g_c^2/(4\pi)$. A similar analysis for the $\mathcal{O}_2$ operator gives $\delta_{22} = \delta_{11}$ and $\delta_{12} = \delta_{21}$. Using eq.(5.9.25), we get

$$\gamma_{ij} = \mu \frac{d}{d\mu} \tilde{\delta}_{ij} = -\mu \frac{d}{d\mu} \delta_{ij} = \frac{\alpha_c}{4\pi} \begin{pmatrix} -2 & 6 \\ 6 & -2 \end{pmatrix}, \quad (7.7.19)$$

where we have used the tree-level d -dimensional QCD β -function $\beta(\alpha_c) = -\epsilon\alpha_c + \mathcal{O}(1)$. The operator basis in which the anomalous dimension matrix is diagonal is given by the operator combinations⁴

$$O_{1/2} = \frac{\mathcal{O}_1 - \mathcal{O}_2}{2}, \quad O_{3/2} = \frac{\mathcal{O}_1 + \mathcal{O}_2}{2}, \quad (7.7.20)$$

with eigenvalues

$$\gamma_{1/2} = -\frac{8\alpha_c}{4\pi}, \quad \gamma_{3/2} = \frac{4\alpha_c}{4\pi}. \quad (7.7.21)$$

We then have $c_i O_i = c_{1/2} O_{1/2} + c_{3/2} O_{3/2}$, with $c_{1/2} = c_1 - c_2$, $c_{3/2} = c_1 + c_2$. In this new basis the RG flow for the couplings reads

$$\mu \frac{dc_i}{d\mu} = \gamma_i c_i, \quad i = 1/2, 3/2. \quad (7.7.22)$$

Having finally found the anomalous dimensions of the four-fermion operators, we can study the RG running of their irrelevant couplings, induced by the QCD coupling, as explained in section 5.10. The equations are easily integrated by writing $\mu dc_i/d\mu = dc_i/d\alpha_c \beta_c$, where β_c is the (one-loop) QCD β -function

$$\beta_c = -\frac{\alpha_c^2}{2\pi} b_0 \quad b_0 = 11 - \frac{2}{3} n_f. \quad (7.7.23)$$

We get

$$c_{1/2}(\mu) = c_{1/2}(\mu_0) \left(\frac{\alpha_c(\mu)}{\alpha_c(\mu_0)} \right)^{\frac{4}{b_0}}, \quad c_{3/2}(\mu) = c_{3/2}(\mu_0) \left(\frac{\alpha_c(\mu)}{\alpha_c(\mu_0)} \right)^{-\frac{2}{b_0}}. \quad (7.7.24)$$

The integration constants $c_i(\mu_0)$ are fixed by taking $\mu_0 = m_W$ and matching by using eq. (7.7.6): $c_{3/2}(m_W) = c_{1/2}(m_W) = g^2 V_{us} V_{du}^* / (2m_W^2) \equiv c_0$. By taking $\mu \sim m_K$, we can estimate the effect of the QCD LL corrections. For $n_f = 4$,⁵ one roughly gets $c_{1/2}(m_K) \simeq 2c_0$, $c_{3/2}(m_K) \simeq 0.7c_0$, with an enhancement by about a factor 3 of the isospin 1/2 operator with respect to the 3/2 one, showing that QCD corrections are far from being negligible.

⁴The label 1/2 and 3/2 refers to the quantum numbers of such operators under the $SU(2)_L$ global symmetry under which u_L and d_L form a doublet and s_L is a singlet. This is an abuse of language, because while $O_{1/2}$ is in fact a component of an operator doublet with isospin 1/2, the operator $O_{3/2}$ is a component of a mixture of operators with isospin 1/2 and 3/2. We follow here the notation of ref.[49].

⁵This is a simplification because we should perform a more refined matching by taking into account the b quark for energies above its mass and integrate out the c quark below its mass.

7.7.1 Useful Color and Spinor identities**

In this subsection we provide the details to derive eqs.(7.7.10) and (7.7.11). Let us first consider the color indices. When the combination of color generators $t_{ij}^a t_{kl}^a$ is summed over all generators, the indices i, j, k, l can only be carried by the group space metric δ_{ij} . Hence we must have

$$\sum_a t_{ij}^a t_{kl}^a = A \delta_{il} \delta_{jk} + B \delta_{ij} \delta_{lk}, \quad (7.7.25)$$

with A and B coefficients to be determined. Multiplying by δ_{ij} (or by δ_{kl}) and by δ_{jk} (or by δ_{il}) we get, for $SU(N)$,

$$\begin{aligned} 0 &= \delta_{lk}(A + NB), \\ C_2(r) \delta_{il} &\equiv \sum_a (t^a t^a)_{il} = \delta_{il}(NA + B), \end{aligned} \quad (7.7.26)$$

where $C_2(r)$ is the quadratic Casimir group invariant, that depends on the representation of the generators t^a . For $SU(N)$ generators in the fundamental representation we have

$$C_2(fund.) = \frac{N^2 - 1}{2N}. \quad (7.7.27)$$

The solution of eqs.(7.7.26) is

$$A = \frac{1}{2}, \quad B = -\frac{1}{2N}. \quad (7.7.28)$$

For $SU(3)$ we have then

$$\sum_a t_{ij}^a t_{kl}^a = \frac{1}{2} \left(\delta_{il} \delta_{jk} - \frac{1}{3} \delta_{ij} \delta_{kl} \right), \quad (7.7.29)$$

that coincides with eq.(7.7.10). Let us now turn to the spinor indices. Here we look for identities relating products of two fermion bilinears expressed in different combinations. Schematically $(\bar{\psi}_1 \Gamma \psi_4)(\bar{\psi}_3 \Gamma \psi_2) \sim (\bar{\psi}_1 \Gamma \psi_2)(\bar{\psi}_3 \Gamma \psi_4)$, where Γ are products of gamma matrices. Relations of this form are called Fierz identities. The latter are best derived by decomposing an open index fermion bilinear $\psi_a \bar{\psi}_b$ in terms of an independent basis in spinor space. A convenient independent basis is provided by the antisymmetric products of gamma matrices:

$$1_{ab}, \quad \gamma_{ab}^\mu, \quad \gamma_{ab}^{\mu\nu}, \quad \gamma_{ab}^{\mu\nu\rho}, \quad \gamma_{ab}^{\mu\nu\rho\sigma}, \quad (7.7.30)$$

where $\gamma^{\mu_1 \dots \mu_n}$ is the completely antisymmetrized product of gamma matrices:

$$\gamma^{\mu_1 \dots \mu_n} \equiv \frac{1}{n!} (\gamma^{\mu_1} \dots \gamma^{\mu_n} \pm perms.). \quad (7.7.31)$$

A spinor basis should consists of $4 \times 4 = 16$ matrices, with complex valued coefficients. The total number of matrices appearing in eq.(7.7.30) is indeed $1 + 4 + 6 + 4 + 1 = 16$. One also notices that the matrices (7.7.30) are orthogonal in spinor space, namely

$$\text{Tr}(\gamma^{\mu_1 \dots \mu_{n_1}} \gamma^{\mu_1 \dots \mu_{n_2}}) = 0, \quad \text{for } n_1 \neq n_2 \quad (7.7.32)$$

and hence they form a basis.⁶ Notice that the matrix $\gamma^{\mu\nu\rho\sigma}$ is equivalent to the chiral γ_5 matrix. Hence, an alternative, but simpler, basis is given by the set

$$1_{ab}, \quad \gamma_{ab}^\mu, \quad \gamma_{ab}^{\mu\nu}, \quad (\gamma^\mu \gamma_5)_{ab}, \quad \gamma_{ab}^5. \quad (7.7.33)$$

A generic fermion bilinear can then be written as follows

$$\psi_{2a} \bar{\psi}_{1b} = c_1 \delta_{ab} + c_2 \gamma_{ab}^\mu + c_3 \gamma_{ab}^{\mu\nu} + c_4 (\gamma^\mu \gamma_5)_{ab} + c_5 (\gamma_5)_{ab}, \quad (7.7.34)$$

where c_i ($i = 1, \dots, 5$) are coefficients to be determined. Multiplying eq.(7.7.34) by δ_{ab} immediately allows us to fix c_1 :

$$c_1 = \frac{\eta}{4} (\bar{\psi}_1 \psi_2), \quad (7.7.35)$$

where $\eta = +1$ for commuting spinors and $\eta = -1$ for anti-commuting spinors. Similarly, multiplying eq.(7.7.34) by the various gamma matrices allows us to fix the remaining coefficients. We get

$$\psi_{2a} \bar{\psi}_{1b} = \frac{\eta}{4} \left((\bar{\psi}_1 \psi_2) \delta_{ab} + (\bar{\psi}_1 \gamma_\mu \psi_2) \gamma_{ab}^\mu + (\bar{\psi}_1 \gamma_{\mu\nu} \psi_2) \gamma_{ab}^{\mu\nu} + (\bar{\psi}_1 \gamma_\mu \gamma_5 \psi_2) (\gamma_5 \gamma^\mu)_{ab} + (\bar{\psi}_1 \gamma_5 \psi_2) (\gamma_5)_{ab} \right). \quad (7.7.36)$$

Using eq.(7.7.36), it is clear how to relate products of bilinears of the form $(\bar{\psi}_1 \Gamma \psi_4)(\bar{\psi}_3 \Gamma \psi_2)$ to $(\bar{\psi}_1 \Gamma \psi_2)(\bar{\psi}_3 \Gamma \psi_4)$. The relation (7.7.36) is the master equation that, applied to different products of bilinears, gives rise to different sets of Fierz identities. In the situation discussed in section 7.7, eq.(7.7.8), we have to consider chiral fermions. This simplifies the analysis, because upon multiplying eq.(7.7.36) by P_L from the left and by P_R from the right, ($P_L = (1 + \gamma_5)/2$, $P_R = (1 - \gamma_5)/2$), three out of five terms in the r.h.s. of that equation vanish. The remaining two terms combine to give

$$\psi_{2L,a} \bar{\psi}_{1L,b} = \frac{\eta}{2} (\bar{\psi}_{1L} \gamma^\mu \psi_{2L}) (\gamma_\mu P_R)_{ab}, \quad (7.7.37)$$

that is eq.(7.7.11) of section 7.7 with $\eta = -1$.

⁶Notice that this true in any number of even dimensions d . In particular, the total number of matrices equal $(1 + 1)^d = 2^d = 2^{d/2} \times 2^{d/2}$, where $2^{d/2}$ is the dimension of the spinor space in d even dimensions.

7.8 (Ir)relevance of Higher Dimensional Operators*

We have so far neglected all higher dimensional operators appearing in the EFT, assuming that their effect is small. This assumption is correct, otherwise EFTs would not be very useful, but it is not as trivial as what naively expected. At the quantum level the insertion of higher dimensional operators in Feynman amplitudes leads to UV divergences which can obscure the irrelevance of these operators at low energies. As a matter of fact, these operators are not always negligible at low-energies, their effect depending on the regularization and renormalization procedure used to deal with the divergences. As an illustrative example, let us go back to the two scalar theory in section 7.3 and let us add the dimension six operator

$$\frac{\tilde{\lambda}}{M^2} L^3 \square L \quad (7.8.1)$$

to the Lagrangian (7.3.2).⁷ The operator (7.8.1) is indeed generated, at tree-level, when we integrate H out. Naively, we expect that (7.8.1) is suppressed at low-energies as E^2/M^2 with respect to the λ coupling in eq. (7.3.2). However, at the quantum level eq. (7.8.1) leads to new UV divergences. For instance, at one-loop level the 1PI 4-point function with one λ and one $\tilde{\lambda}$ vertex is quadratically divergent, so that in cut-off regularization

$$\delta\Gamma^{(4)} \propto \lambda \frac{\tilde{\lambda}}{M^2} \int^\Lambda d^4k \frac{k^2}{(k^2 + p^2)^2} \sim \lambda \frac{\tilde{\lambda}}{M^2} \left(\Lambda^2 + p^2 \log \frac{\Lambda^2}{p^2} + \dots \right), \quad (7.8.2)$$

where p generically denotes external momenta and/or the mass m . Since the cut-off Λ of the EFT is around M , we see that the divergent term can compensate for the manifest $1/M^2$ suppression to give a contribution to $\Gamma^{(4)}$ of the same order of the leading λ^2 term. After the divergence is subtracted, no dependence on the cut-off appears. However, in a generic regularization and renormalization prescription, *power-like* dependence on the sliding scale is induced, which can spoil the naive dimensional analysis. For instance, by renormalizing the operators L^4 and $L^3 \square L$ at a scale μ , one would get

$$\delta\Gamma^{(4)} \sim \lambda \frac{\tilde{\lambda}}{M^2} \left(p^2 \log \frac{\mu^2}{p^2} + \mu^2 \log \frac{\mu^2}{p^2} + \dots \right). \quad (7.8.3)$$

If the subtraction point $\mu \ll M$, we can still neglect higher dimensional operators, but in a sufficiently complicated set-up powers of M might be induced by mixing with various operators. Moreover, we have seen that it is convenient in effective field theory to renormalize at the scale $\mu \sim M$, in which case dimensional analysis, when using eq.(7.8.3), would break down.

⁷As we will see in the next section, this is a so called redundant operator, but this is inessential for the argument we want to make here.

This potential problem is scheme-dependent and is absent in DR with a mass-independent renormalization subtraction scheme, such as $\overline{\text{MS}}$. In DR, divergences arise as poles in $1/\epsilon$ and, when minimally subtracted, leave a renormalized amplitude where the sliding scale μ appears only in logs. Since in the EFT the available energy scales are light masses and low-energy momenta, by dimensional analysis all higher dimensional operators are suppressed. In the example above, for instance, we get, once the poles are subtracted,

$$\delta\Gamma^{(4)} \propto \lambda \frac{\tilde{\lambda}}{M^2} \int d^d k \frac{k^2}{(k^2 + p^2)^2} \sim \lambda \frac{\tilde{\lambda}}{M^2} \left(p^2 \log \frac{\mu^2}{p^2} + \dots \right), \quad (7.8.4)$$

which is sub-leading with respect to the λ^2 term. In other words, mass-independent schemes preserve dimensional analysis, an observation we already made in subsection 5.8 in the context of the RG flow of general dimensionful couplings. We now see that this property has the fundamental implication of validating a naive treatment of higher dimensional operators in the context of EFT.

7.9 Redundant Operators*

We have discussed at length that in EFT one should write down a local effective action for the lights fields including all the higher dimensional operators up to the desired order, the coefficients of which will be then fixed by a matching procedure. It turns out, however, that there is a redundancy in this description, since the physical observables in general depend on only a subset of the EFT coupling constants (called physical in what follows). The remaining couplings are called redundant, because they can be expressed in terms of the physical ones. This redundancy can be traced back to the LSZ reduction formulas (2.3.12), for which on-shell amplitudes are associated to the residue of multiple poles in a Green function, not to the whole off-shell Green function. We can obtain different Green functions by changing, for instance, the asymptotic field ϕ that creates one-particle states. Any other quantum field ϕ' , as long as $\langle 0|\phi'|p \rangle \neq 0$, is an equally good choice. The off-shell Green functions associated to ϕ' would generally differ from those associated to ϕ , but both would lead to identical on-shell properties. In other words, different off-shell Green functions lead to the same on-shell physics.

The same redundancy can similarly be discussed in terms of operators, and of their associated couplings, using the Schwinger-Dyson equations (4.4.3). Viewing $\delta S/\delta\phi$ as a particular composite operator in the theory, we can interpret eqs.(4.4.3) as the statement that such composite operator, when inserted in Green functions, gives only rise to contact terms with Green functions containing a lower number of fields. On-shell, this implies that some pole is lost and these operators should not contribute to physical processes (compare

with the situation discussed in subsection 4.4.3 in the context of the WI in QED). These operators are called redundant [50]. Composite operators made of $\delta S/\delta\phi$ times other operators evaluated at the same space-time point are also redundant. This can be shown as follows [50]. Let us define

$$\theta(x) = F(\phi(x)) \frac{\delta S}{\delta\phi(x)} \quad (7.9.1)$$

a (suitably renormalized) composite operator, where F is a local functional of ϕ and $\delta S/\delta\phi(x)$ are the equations of motion for the field ϕ . We can introduce external sources s and J for ϕ and θ , respectively, and consider the generating functional

$$e^{iW[J,s]} = \int \mathcal{D}\phi e^{i \int d^d x (\mathcal{L} + s\phi + J\theta)}. \quad (7.9.2)$$

We will be interested in correlation functions with at most one insertion of θ , so we can expand in J and keep only the terms up to linear level. Performing the path integral change of variables

$$\phi(x) \rightarrow \phi(x) - J(x)F(\phi(x)), \quad (7.9.3)$$

we then get

$$e^{iW[J,s]} = \int \mathcal{D}\phi \left| \det \text{Jac}(x, y) \right| e^{i \int d^d x (\mathcal{L} + s(\phi - JF))}, \quad (7.9.4)$$

where

$$\text{Jac}(x, y) = \delta^{(d)}(x - y) - J(x) \frac{\delta F((\phi(x)))}{\delta\phi(y)} \quad (7.9.5)$$

is the Jacobian associated to the change of variables. By taking n functional derivatives with respect to $s(x_i)$ and one with respect to $J(x)$, and setting $s = J = 0$, we get the identity

$$\begin{aligned} i \langle \phi(x_1) \dots \phi(x_n) \theta(x) \rangle &= - \sum_{i=1}^n \delta^{(d)}(x - x_i) \langle \phi(x_1) \dots F(\phi(x)) \dots \phi(x_n) \rangle \\ &\quad - \langle \phi(x_1) \dots \phi(x_n) \frac{\delta F((\phi(x)))}{\delta\phi(x)} \rangle. \end{aligned} \quad (7.9.6)$$

The last term in eq.(7.9.6) arises from the determinant of the Jacobian. For any local functional of fields $F(\phi)$, it is UV divergent and proportional to $\delta^{(d)}(0)$, so it vanishes in DR, where such divergences are set to zero. In this way eq.(7.9.6) is the generalization of the Schwinger-Dyson equations (4.4.3), which are reproduced by taking $F = 1$. Equation (7.9.6) shows that any local composite operator θ of the form (7.9.1) does not contribute to physical processes.

In practice redundant operators can be removed in EFT either by using the equations of motion at the Lagrangian level or alternatively by field redefinitions. It is useful to

see how this works in a simple example: an EFT featuring only a real scalar with $\mathbb{Z}_2$ symmetry $\phi \rightarrow -\phi$. Up to dimension six operators, the most general EFT reads as:

$$\mathcal{L}_{EFT}^{(d \leq 6)} = \frac{1}{2}(\partial\phi)^2 - \lambda\phi^4 + \frac{g_1}{M^2}\phi^6 + \frac{g_2}{M^2}(\Box\phi)^2 + \frac{g_3}{M^2}\phi^3\Box\phi + \mathcal{O}\left(\frac{1}{M^4}\right), \quad (7.9.7)$$

where, for further simplification, we have neglected the mass term $m^2\phi^2$. One could write down other dimension six operators, but these are equal, up to total derivatives, to the operators in eq.(7.9.7). For example, integrating by parts, one easily gets

$$\phi^2(\partial\phi)^2 = -\frac{1}{3}\phi^3\Box\phi + \frac{1}{3}\partial_\mu(\phi^3\partial_\mu\phi), \quad (7.9.8)$$

the last operator giving a vanishing contribution in the action. The equation of motion for ϕ is

$$-\Box\phi - 4\lambda\phi^3 + \mathcal{O}\left(\frac{1}{M^2}\right) = 0. \quad (7.9.9)$$

Plugging back eq.(7.9.9) in the dimension six terms in the Lagrangian (7.9.7) gives

$$\tilde{\mathcal{L}}_{EFT}^{(d \leq 6)} = \frac{1}{2}(\partial\phi)^2 - \lambda\phi^4 + \frac{\tilde{g}_1}{M^2}\phi^6 + \mathcal{O}\left(\frac{1}{M^4}\right), \quad (7.9.10)$$

where

$$\tilde{g}_1 = g_1 + 16\lambda^2 g_2 - 4\lambda g_3. \quad (7.9.11)$$

We see that two out of the three dimension-six operators are redundant. This implies that, at this order, physical observables only depend on the two couplings λ and $\tilde{g}_1$.

The same conclusion can be reached by making the most general field redefinition compatible with the symmetries at this order:

$$\phi \rightarrow \phi + \frac{a}{M^2}\Box\phi + \frac{b}{M^2}\phi^3, \quad (7.9.12)$$

where a and b are so far undetermined dimensionless parameters. Under (7.9.12) the dimension four operators in the Lagrangian change as follows:

$$\begin{aligned} (\partial\phi)^2 &\rightarrow (\partial\phi + \frac{a}{M^2}\Box\phi + \frac{3b}{M^2}\phi^2\partial\phi)^2 = (\partial\phi)^2 + 2(\partial\phi)\left(\frac{a}{M^2}\Box\phi + \frac{3b}{M^2}\phi^2\partial\phi\right) + \text{h.d.o.} \\ \phi^4 &\rightarrow (\phi + \frac{a}{M^2}\Box\phi + \frac{b}{M^2}\phi^3)^4 = \phi^4 + 4\phi^3\left(\frac{a}{M^2}\Box\phi + \frac{b}{M^2}\phi^3\right) + \text{h.d.o.} \end{aligned} \quad (7.9.13)$$

In the first line of eq.(7.9.13) we can integrate by parts the first dimension six operator and use eq.(7.9.8) in the second one to express both operators in the basis chosen in eq.(7.9.7). Up to total derivative terms, the Lagrangian (7.9.7) after the field redefinition (7.9.12) reads then⁸

$$\mathcal{L}_{EFT}^{(d \leq 6)} \rightarrow \mathcal{L}_{EFT}^{(d \leq 6)} = \frac{1}{2}(\partial\phi)^2 - \lambda\phi^4 + \frac{\tilde{g}_1}{M^2}\phi^6 + \frac{\tilde{g}_2}{M^2}(\Box\phi)^2 + \frac{\tilde{g}_3}{M^2}\phi^3\Box\phi + \mathcal{O}\left(\frac{1}{M^4}\right), \quad (7.9.14)$$

⁸We have seen that no contribution in DR arises from the Jacobian associated to the change of variables (7.9.12), which can be neglected.

where

$$\tilde{g}_1 = g_1 - 4b\lambda, \quad (7.9.15)$$

$$\tilde{g}_2 = g_2 - a, \quad (7.9.16)$$

$$\tilde{g}_3 = g_3 - 4a\lambda - b. \quad (7.9.17)$$

We can choose a and b such that $\tilde{g}_2 = \tilde{g}_3 = 0$, i.e. $a = g_2$, $b = g_3 - 4\lambda g_2$. Plugging these values back in eq.(7.9.15) reproduces eq.(7.9.11) and gives rise to the same EFT (7.9.10). The presence of two redundant operators in the Lagrangian (7.9.7) can also be seen by taking the following basis for the dimension six operators:

$$\mathcal{L}_{EFT}^{(d \leq 6)} = \frac{1}{2}(\partial\phi)^2 - \lambda\phi^4 + \frac{\hat{g}_1}{M^2}\phi^6 + \frac{\hat{g}_2}{M^2}(\Box\phi + 4\lambda\phi^3)^2 + \frac{\hat{g}_3}{M^2}\phi^3(\Box\phi + 4\lambda\phi^3) + \mathcal{O}\left(\frac{1}{M^4}\right), \quad (7.9.18)$$

where we immediately see that $\hat{g}_2$ and $\hat{g}_3$ are redundant couplings, and the operators $\hat{\mathcal{O}}_2$ and $\hat{\mathcal{O}}_3$ in eq.(7.9.18) are of the form (7.9.1), with $F_2 = \Box\phi + 4\lambda\phi^3$, $F_3 = \phi^3$.

Removing redundant operators can lead to a significant simplification. It should however be noted that if one starts with a Lagrangian where redundant operators have been removed, like eq.(7.9.10), and considers off-shell Green functions, the cancellation of UV divergences will generally require the introduction of counterterms associated to the redundant operators. Such counterterms can then be reexpressed in terms of counterterms for the non-redundant operators using field redefinitions or the equations of motion.

Chapter 8

Spontaneously Broken Symmetries*

Symmetries are a fundamental concept in QFT but they are often broken in Nature. If a physical system described by a Lagrangian $\mathcal{L}(\phi)$ is invariant under some symmetry, a possible breaking term is obtained by adding terms to $\mathcal{L}(\phi)$ which do not respect the symmetry (explicit breaking). If the corresponding breaking operators have dimensions less than four, the UV behaviour of the theory (and its divergences) will not be affected by the symmetry breaking terms. In this case we say that the symmetry is *softly* broken.

Another possibility occurs when $\mathcal{L}(\phi)$ is invariant under the symmetry but the ground state (the vacuum) is not. If G is the operator in the Hilbert space parametrizing the action of the symmetry group, $G|0\rangle \neq |0\rangle$. In this case we have what is called a spontaneous symmetry breaking. Classical prototypical example is the breaking of the $SO(3)$ spatial rotations in a ferromagnet. The laws of Nature are all spatially symmetric, but the vacuum is not.

The structure of this chapter is as follows. We start in section 8.1 by explaining why and how spontaneous symmetry breaking takes place in QFT. An important consequence of the breaking of a global continuous symmetry is the Goldstone theorem, which is introduced and proved in section 8.2. The effect of an explicit symmetry breaking term, on top of a spontaneous breaking, is discussed in section 8.3. The Higgs mechanism in general gauge theories and in arbitrary ξ -gauges is discussed in section 8.4. The equivalence of the high energy scattering of longitudinal gauge bosons with the corresponding eaten Goldstone bosons (so called Goldstone boson equivalence theorem) is introduced in section 8.5. The effective field theories of Goldstone bosons for a large class of theories is discussed in section 8.6. Two notable examples of such theories are discussed next in sections 8.7

and 8.8, where we respectively study mesons in QCD and a hypothetical scenario where the Higgs field is a (pseudo)-Goldstone boson. Finally, in section 8.9 we generalize the formalism introduced in section 8.6 in order to treat more general effective theories of Goldstone bosons.

8.1 Why Spontaneous Symmetry Breaking?*

The simplest example of a spontaneous symmetry breaking mechanism is provided by the $\lambda\phi^4$ theory with a negative squared mass term, which enjoys a $\mathbf{Z}_2$ symmetry under which $\phi \rightarrow -\phi$, with $\langle\phi\rangle = \pm\phi_0 \neq 0$.¹ Clearly, the $\mathbf{Z}_2$ symmetry exchanges the two degenerate vacua, $\mathbf{Z}_2|\pm\phi_0\rangle = |\mp\phi_0\rangle$. The two vacua are physically identical, so that we can choose any of the two and expand for small fluctuations around the selected one. The story is however not so simple. We are tacitly assuming that the vacuum is one of the two vacua $|\pm\phi_0\rangle$, and *not* a linear combination of them. If the vacuum was $|\pm\rangle = 1/\sqrt{2}(|\phi_0\rangle \pm |-\phi_0\rangle)$, then the symmetry would be exact and not spontaneously broken (the vacua being mapped to themselves, up to a possible sign factor). This is in fact what happens in the quantum mechanical analogue of the double well potential: the hamiltonian eigenstates are given by $|\pm\rangle$, since tunneling effects induce non-trivial transitions of the form $\langle\pm\phi_0|\mp\phi_0\rangle$, and no spontaneous $\mathbf{Z}_2$ symmetry breaking occurs. In fact, spontaneous symmetry breaking does not occur in quantum mechanics. More generally, it does not occur in systems with a finite number of degrees of freedom, like in QFT defined on a finite volume. The tunneling effects between different vacua in QFT are exponentially suppressed by the volume of space V . As $V \rightarrow \infty$, such transitions no longer occur and different degenerate vacua are equivalent and completely disconnected from each other: spontaneous symmetry breaking can then occur.

The absence of transitions between different degenerate vacua in QFT can be shown by looking at the equal-time commutators of two generic local operators A and B . Assume that the vacua are discrete and take them orthonormal:

$$\langle u|v\rangle = \delta_{u,v} . \quad (8.1.1)$$

Inserting a complete basis of states we have

$$\langle u|A(\vec{x})B(0)|v\rangle = \sum_w \langle u|A(0)|w\rangle\langle w|B(0)|v\rangle + \int d^3\vec{p} e^{i\vec{p}\cdot\vec{x}} \rho(\vec{p}) , \quad (8.1.2)$$

¹The analysis in this section follows section 19.1 of ref. [60]. I also found useful chapter 2.2 of Parisi's book [48], that contains several interesting observations that are often omitted in QFT textbooks, in particular about the role of cluster decomposition.

where the first term is the sum over all the possible vacua w present in the theory and

$$\rho(\vec{p}) \equiv \sum_{n \geq 1} \langle u|A(0)|n, q_n \rangle \langle n, q_n|B(0)|v \rangle \delta^{(3)}(\vec{p} - \vec{q}_n) \quad (8.1.3)$$

represents the sum over n -particles states with total momentum $\vec{q}_n$. If we assume that $\rho(\vec{p})$ is integrable, the Riemann-Lebesgue lemma applies and we have

$$\lim_{|\vec{x}| \rightarrow \infty} \int d^3\vec{p} e^{i\vec{p} \cdot \vec{x}} \rho(\vec{p}) = 0. \quad (8.1.4)$$

This condition is automatically satisfied if the theory has a mass gap, namely if the one-particle states are all massive, in which case the Fourier transform of $\rho(\vec{p})$ vanishes exponentially in $|\vec{x}|$ at large distances. It ensures that operators inserted far away from each other have no correlation. This is cluster decomposition at the operator level, that will play an important role again in what follows. Since A and B are space-like separated, microcausality requires that

$$[A(\vec{x}), B(0)] = 0. \quad (8.1.5)$$

Eqs.(8.1.2), (8.1.4) and (8.1.5) imply that the matrix elements $\langle u|A(0)|v \rangle$ and $\langle u|B(0)|v \rangle$ should commute for any u and v , and hence can be simultaneously diagonalizable. There exists then a basis of vacua where

$$\langle u|A(0)|v \rangle = a_u \delta_{u,v}. \quad (8.1.6)$$

Since A is an arbitrary local operator, we can take it to be the hamiltonian density operator and thus we have proved that no transitions between degenerate vacua are possible in QFT. Notice how the infinite volume limit entered crucially in eq.(8.1.4).

However, this is not the end of the story, because we have still to rule out the possibility to choose as vacuum some linear combination $\sum_u c_u |u \rangle$, invariant under the symmetry, where $|u \rangle$ is the orthonormal basis defined by eq.(8.1.6), rather than any of the $|u \rangle$. This possibility is ruled out by demanding cluster decomposition of the correlation function:

$$\lim_{|\vec{x}| \rightarrow \infty} \langle A(\vec{x})B(0) \rangle_c = \lim_{|\vec{x}| \rightarrow \infty} \left(\langle A(\vec{x})B(0) \rangle - \langle A(\vec{x}) \rangle \langle B(0) \rangle \right) = 0. \quad (8.1.7)$$

For instance, for two vacua related by a $\mathbf{Z}_2$ symmetry, we can denote $|\pm \rangle$ the vacua in the diagonal basis (8.1.6) and by $|\alpha \rangle$ the linear combination

$$|\alpha \rangle = \cos \alpha |+\rangle + \sin \alpha |-\rangle, \quad (8.1.8)$$

with $0 \leq \alpha \leq \pi/2$. In a α -vacuum we have

$$\begin{aligned} \lim_{|\vec{x}| \rightarrow \infty} \langle \phi(\vec{x})\phi(0) \rangle_\alpha &= \langle \alpha|\phi(0)|+\rangle \langle +|\phi(0)|\alpha \rangle + \langle \alpha|\phi(0)|-\rangle \langle -|\phi(0)|\alpha \rangle \\ &= v^2(\sin^2 \alpha + \cos^2 \alpha) = v^2, \end{aligned} \quad (8.1.9)$$

where $\pm v$ is the VEV of the field ϕ in the two vacua $|\pm\rangle$. On the other hand

$$\langle\phi(0)\rangle_\alpha = \langle\alpha|\phi(0)|\alpha\rangle = v(\cos^2\alpha - \sin^2\alpha) = v\cos 2\alpha, \quad (8.1.10)$$

so that

$$\lim_{|\vec{x}|\rightarrow\infty} \langle\alpha|\phi(\vec{x})\phi(0)|\alpha\rangle_c = v^2(1 - \cos^2 2\alpha) = v^2 \sin^2 2\alpha. \quad (8.1.11)$$

The only α -vacua that satisfy cluster decomposition are those with $\alpha = 0, \pi/2$, i.e. the states $|+\rangle$ and $|-\rangle$. Spontaneous symmetry breaking can then be seen as an obstruction in having a vacuum that at the same time is invariant under a symmetry and satisfies cluster decomposition.

The discussion above can be generalized to continuous symmetries where we have an infinite family of degenerate vacua. For continuous symmetries we also have an important theorem, the Goldstone theorem, relating the symmetry breaking pattern to the spectrum of the theory.

8.2 The Goldstone Theorem*

Let G be a group of global continuous symmetries with generators t^α , $\alpha = 1, \dots, \dim G$, acting on some system. By Noether's theorem, we have $\dim G$ conserved currents $J_\mu^\alpha(x)$: $\partial^\mu J_\mu^\alpha(x) = 0$ and the associated charges $Q^\alpha = \int d^3x J_0^\alpha(x)$. The group G is said to be spontaneously broken if, on the vacuum, the generators splits into two sets, labelled by a and i , such that

$$Q^a|0\rangle \neq 0, \quad Q^i|0\rangle = 0, \quad (8.2.1)$$

with a non-empty set for a . The unbroken generators labelled by t^i form a subgroup of G , denoted by H , so that we have $\dim G - \dim H$ broken generators ($a = 1, \dots, \dim G - \dim H$). Goldstone's theorem states that, independently of the specific pattern of breaking and physical system we are considering, in the spectrum there will appear one massless and spinless particle for each broken generator. The particle will be scalar or pseudoscalar, depending on the parity of the associated broken generator. These particles are called Goldstone or Nambu-Goldstone (NG) bosons. We denote them in what follows as the NG particles.²

Proof of the theorem. Let ϕ_n be the set of fields responsible for the spontaneous symmetry breaking pattern $G \rightarrow H$. This implies that an infinitesimal action of the group

²The Goldstone theorem and the proof that follows is actually due to Goldstone, Salam and Weinberg [29]. Such particles are denoted Nambu-Goldstone bosons because they were found by Nambu [47] and Goldstone [28] in two specific set-ups.

G on the fields ϕ_n do not leave them invariant, namely $\langle \phi'_n(0) \rangle \neq \langle \phi_n(0) \rangle$, where ϕ' are the fields one gets after the infinitesimal action. In other words, we must have

$$\langle \delta \phi_n(0) \rangle = \epsilon^\alpha \langle [Q^\alpha, \phi_n(0)] \rangle \neq 0, \quad (8.2.2)$$

with $\langle [Q^a, \phi_n(0)] \rangle \equiv \delta \phi_n^a \neq 0$, for each a . Consider now the two-point function of a conserved current J_μ^a with the fields ϕ_n . Since $\partial^\mu J_\mu^a(x) = 0$, we have

$$\begin{aligned} \partial_x^\mu \langle 0 | T[J_\mu^a(x) \phi_n(0)] | 0 \rangle &= \partial_x^\mu \left(\theta(x^0) \langle J_\mu^a(x) \phi_n(0) \rangle + \theta(-x^0) \langle \phi_n(0) J_\mu^a(x) \rangle \right) \\ &= \delta(x^0) \langle [J_0^a(x), \phi_n(0)] \rangle. \end{aligned} \quad (8.2.3)$$

Let us denote by $G_{\mu,n}^a$ the Fourier transform of the two-point function:

$$\langle 0 | T[J_\mu^a(x) \phi_n(0)] | 0 \rangle = \int \frac{d^4 p}{(2\pi)^4} G_{\mu,n}^a(p) e^{-ip \cdot x}. \quad (8.2.4)$$

By Lorentz invariance we have

$$G_{\mu,n}^a(p) = i p_\mu H_n^a(p^2). \quad (8.2.5)$$

Integrating eq. (8.2.3) over space-time gives

$$\int d^4 x \partial_x^\mu \langle 0 | T[J_\mu^a(x) \phi_n(0)] | 0 \rangle = \int d^4 p \delta^{(4)}(p) p^2 H_n^a(p^2) = \langle [Q^a, \phi_n(0)] \rangle \neq 0. \quad (8.2.6)$$

The integral of a divergence does not vanish when long-range particles appear, and these are precisely the NG bosons. More precisely, we must have

$$H_n^a(p^2) = \frac{\delta \phi_n^a}{p^2} + \dots \quad (8.2.7)$$

where $\dots$ represent terms that are regular when $p \rightarrow 0$. A pole in the two-point function of J_μ^a with ϕ_n is a signal of the presence of massless 1-particle states, one for each unbroken generator a . Since ϕ_n are spinless (otherwise in the vacuum we will break the Lorentz symmetries as well), such particles are necessarily of spin zero. Their intrinsic parity is then fixed by the parity of the associated current J^a . Q.e.d.

Denoting by NG_a the one-NG particle states, we have

$$\begin{aligned} \langle 0 | J_\mu^a(x) | NG_b(p) \rangle &= \frac{i p_\mu F_{ab} e^{-ip \cdot x}}{\sqrt{2p_0(2\pi)^3}}, \\ \langle NG_b(p) | \phi_n(x) | 0 \rangle &= \frac{Z_n^b e^{ip \cdot x}}{\sqrt{2p_0(2\pi)^3}}, \end{aligned} \quad (8.2.8)$$

where F_{ab} and Z_n^a are matrices (of mass dimension one and zero, respectively) related to $\delta\phi_n^a$ in eq.(8.2.7). This relation is found by noticing that $G_{\mu,n}^a$ above is associated to the spectral density $\rho_{\mu,n}^a$ of the two point function $\langle 0|J_\mu^a(x)\phi_n(0)|0\rangle$ (recall eq.(2.1.16)):

$$G_{\mu,n}^a(p) = \int_0^\infty d\sigma \rho_{\mu,n}^a(p, \sigma) \frac{i}{p^2 - \sigma + i\epsilon}, \quad (8.2.9)$$

where by Lorentz invariance $\rho_{\mu,n}^a(p, \sigma) = p_\mu \rho_n^a(\sigma)$. The contribution of the one-particle NG bosons to ρ_n^a is easily computed from eq.(8.2.8):

$$\rho_{n,NG}^a(\sigma) = iF_{ab}Z_n^b\delta(\sigma). \quad (8.2.10)$$

Matching eqs.(8.2.7), (8.2.9) and (8.2.10) we then have

$$\delta\phi_n^a = iF_{ab}Z_b^n. \quad (8.2.11)$$

We can also define the NG fields $\pi^a(x)$ with canonical 1-point function with the NG particles:

$$\langle NG_b(p)|\pi^a(x)|0\rangle = \frac{\delta_{ab}e^{ip\cdot x}}{\sqrt{2p_0(2\pi)^3}}. \quad (8.2.12)$$

Putting together eqs. (8.2.8) and eq. (8.2.12), we see that

$$\phi_n(x) = Z_n^a\pi^a(x) + \dots = (iF)_{ab}^{-1}\delta\phi_n^b\pi^a(x) + \dots \quad (8.2.13)$$

where $\dots$ denote field components not associated to the NG bosons. For linearly realized symmetries, in a basis where all the field components ϕ_n are real and all the generators t^α are purely imaginary, we simply have

$$\delta\phi_n^a = it_{nm}^a\langle\phi_m\rangle. \quad (8.2.14)$$

The appearance of one NG particle per broken generator is physically explained by noticing that the potential for the ϕ_n is, by assumption, invariant under the symmetry. Hence, around any vacuum $\langle\phi_m\rangle$, the broken generators do not leave the minimum invariant, but necessarily shift it in a new minimum with exactly the same energy. Hence, we have a $(\dim G - \dim H)$ -dimensional space of flat directions in the potential. It costs no energy to fluctuate around a flat direction, and the small fluctuation is identified with the NG boson.

The above proof of the Goldstone's theorem does not rely on perturbation theory, indicating that NG particles are expected to arise whenever a global symmetry group is spontaneously broken, no matter whether the theory undergoing the symmetry breaking is weakly or strongly coupled. The fields ϕ_n responsible for the symmetry breaking are

also not necessarily elementary fields appearing directly in the Lagrangian at some energy scale, but might be composite fields built with different fields. The most relevant example of this kind is the spontaneous breaking of the $SU(2)$ chiral symmetry in QCD, induced by effective scalar fields ϕ_n constructed out of quark bilinears. In this case, the three NG bosons are spin zero mesons that appear as bound states of the original quarks, the pions $\pi^0, \pi^\pm$. We will come back to this relevant case in much more detail in the following.

The Goldstone's theorem applies for internal symmetries only, namely for those symmetries whose generators commute with the ones of the Poincaré group. An example of non-internal symmetry is provided by the conformal group, which in four dimensions is given by $SO(4, 2)$. A CFT is invariant under the conformal group. When the latter is broken spontaneously to the Poincaré group, one gets $15 - 10 = 5$ broken generators (special conformal transformations and dilatations), but actually only one massless NG boson, commonly denoted as the dilaton.

Let us also mention that a spontaneous breaking of a continuous symmetry cannot occur in QFTs defined in $1 + 1$ dimensions. The quantum effects are so strong in $2d$ that we no longer have small fluctuations around an asymmetric ground state, but rather large fluctuations around the flat direction which is in this way wholly sampled and result eventually in a restoration of the global symmetry.³ A spontaneous breaking of a *discrete* symmetry is instead allowed.

As we will shortly see, the Goldstone's theorem does not apply for local (i.e. space-time dependent) symmetries as well.

8.3 Vacuum Alignment and Pseudo-Goldstone Bosons*

It is interesting to analyze theories where a spontaneous symmetry breaking occurs and, in addition, sub-leading effects introduce explicit symmetry breaking terms.⁴ Relevant examples include the above mentioned case of QCD, where the up and down quark masses can be seen as small explicit symmetry breaking terms of the $SU(2)$ chiral symmetry. We will consider in what follows linearly realized symmetries only, for which we know that quantum effects do not spoil the invariance of the action. By considering space-time independent field configurations ϕ_n , the whole effective action boils down to the effective potential $V(\phi)$. The latter, by definition, can be written as

$$V(\phi) = V_0(\phi) + V_1(\phi), \quad (8.3.1)$$

³This phenomenon is denoted as the Coleman-Mermin-Wagner-Hohenberg theorem [16, 36, 46].

⁴The analysis in this section closely follows section 19.3 of ref. [60].

where V_0 is the invariant term, and V_1 is the breaking one. In the field region of interest, by assumption, we have $|V_1| \ll |V_0|$. In the basis (8.2.14), we have

$$\delta\phi_n = i\epsilon^\alpha t_{nm}^\alpha \phi_m. \quad (8.3.2)$$

Invariance of V_0 implies

$$\frac{\partial V_0}{\partial \phi_n} t_{nm}^\alpha \phi_m = 0, \quad \forall \alpha. \quad (8.3.3)$$

Let ϕ_0 be the minimum of V_0 and $\phi = \phi_0 + \phi_1$ the minimum of the whole potential V :

$$\left. \frac{\partial V_0}{\partial \phi_n} \right|_{\phi=\phi_0} = 0, \quad \left. \frac{\partial V}{\partial \phi_n} \right|_{\phi=\phi_0+\phi_1} = 0. \quad (8.3.4)$$

Since $|V_1| \ll |V_0|$, we also have $\phi_1 \ll \phi_0$ and we can consistently expand eq. (8.3.4) for small V_1 and ϕ_1 . At first non-trivial order, we have

$$\left. \frac{\partial^2 V_0}{\partial \phi_n \partial \phi_m} \right|_{\phi_0} \phi_{1,m} + \left. \frac{\partial V_1}{\partial \phi_n} \right|_{\phi_0} = 0. \quad (8.3.5)$$

Taking a derivative with respect to ϕ of eq. (8.3.3), and evaluating at ϕ_0 , we have

$$\left. \frac{\partial^2 V_0}{\partial \phi_n \partial \phi_m} \right|_{\phi_0} t_{mp}^\alpha \phi_{0,p} = 0. \quad (8.3.6)$$

When the index α runs over the unbroken directions i , (8.3.6) trivially vanishes since $t_{mp}^i \phi_{0,p} = 0$. For $\alpha = a$, (8.3.6) is non-trivial and indicates that there are $\dim G - \dim H$ directions in field space with a massless eigenvector, that are the NG particles (this can be in fact seen as an alternative proof of the Goldstone's theorem). Multiplying eq. (8.3.5) by $t_{np}^a \phi_{0,p}$ and using eq. (8.3.6), we finally have

$$\left. \frac{\partial V_1}{\partial \phi_n} \right|_{\phi_0} t_{np}^a \phi_{0,p} = 0. \quad (8.3.7)$$

In presence of a perturbation V_1 , the vacuum ϕ_0 is no longer arbitrary among the would-be degenerate vacua (in absence of the perturbation) but is restricted to satisfy eq. (8.3.7). Equation (8.3.7) is denoted vacuum alignment condition, because it shows that the symmetry breaking terms typically force the vacuum ϕ_0 to be parallel to the direction of the explicit breaking term. This is easily seen in the particularly simple example of an $SO(N) \rightarrow SO(N-1)$ breaking pattern. We choose as explicit breaking term $V_1 = u_n \phi_n$, where u_n is a fixed vector, explicitly breaking $SO(N)$ down to the $SO(N-1)$ subgroup that leaves it fixed. The condition (8.3.7) gives $u_n t_{np}^a \phi_{0,p} = 0$, namely $\phi_{0,n} \propto u_n$ (the matrices t^a being antisymmetric). The addition of the small perturbation to the invariant Lagrangian lifts the vacuum degeneracy and forces ϕ_0 to be parallel to u_n . The final

unbroken group is then $SO(N-1)$ and not $SO(N-2)$, intersection of the two would-be misaligned $SO(N-1)$ subgroups.

The explicit term V_1 is generally responsible for another important effect: they give masses to the NG bosons, which become what are sometimes denoted as pseudo-NG bosons. Their masses can be extracted from the potential by using eq. (8.2.13):

$$M_{ab}^2 = \left. \frac{\partial^2 V}{\partial \pi^a \partial \pi^b} \right|_\phi = (iF_{ac})^{-1} \delta \phi_n^c (iF_{bd})^{-1} \delta \phi_m^d \left. \frac{\partial^2 V}{\partial \phi_n \partial \phi_m} \right|_\phi. \quad (8.3.8)$$

Expanding for small V_1 and ϕ_1 , it is straightforward to see that the pseudo NG boson masses are linearly proportional to the explicit breaking term:

$$M_{ab}^2 \propto V_1. \quad (8.3.9)$$

8.4 Spontaneously Broken Gauge Symmetries: the Higgs Mechanism*

The Goldstone's theorem does not apply in the case in which the broken symmetry is local.⁵ It is impossible for gauge theories to keep at the same time Lorentz invariance and positivity of the Hilbert space, both conditions being important to establish the theorem. For local symmetries, no NG massless particles appear, but rather the would-be NG bosons are “eaten” by gauge fields that become massive.

The situation is easily illustrated by the abelian Higgs model (model that will be extensively analyzed in the final chapter in the special case $m = 0$), whose Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + |D_\mu \Phi|^2 - V(\Phi), \quad V(\Phi) = -m^2 \Phi^\dagger \Phi + \frac{\lambda}{2}(\Phi^\dagger \Phi)^2, \quad (8.4.1)$$

with $D_\mu \Phi = \partial_\mu \Phi - ieA_\mu \Phi$. For $m^2 > 0$, the minimum of the potential is at

$$|\Phi_0| = \sqrt{m^2/\lambda} \equiv \frac{v}{\sqrt{2}}. \quad (8.4.2)$$

It is useful to choose radial coordinates for the field Φ and write

$$\Phi = \frac{v + \rho}{\sqrt{2}} e^{i\theta/v}, \quad (8.4.3)$$

where the factor v in the exponential has been introduced to get canonical fields of dimension one. Expanding around small fluctuations, we get at quadratic order

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}(\partial_\mu \rho)^2 + \frac{1}{2}(\partial_\mu \theta - evA_\mu)^2 - \frac{1}{2}(2m^2)\rho^2 + \dots \quad (8.4.4)$$

⁵This often used terminology is actually misleading. Gauge symmetries, being merely redundancies of the system, cannot be broken. A more proper term would be gauge theories in a non-linearly realized, or Higgs, phase.

The radial field is massive, with $m_\rho = \sqrt{2}m$, while the angular field θ is massless. Actually θ mixes at quadratic level with the gauge field A_μ and it is not an eigenstate of the free Hamiltonian. We can easily get rid of the unwanted mixing term $eA^\mu\partial_\mu\theta$ by noticing that we have still to gauge-fix the $U(1)$ gauge symmetry. Under a $U(1)$ transformation with parameter $\alpha(x)$, $\Phi(x) \rightarrow \exp(i\alpha(x))\Phi(x)$. The radial field $\rho(x)$ is invariant, while the angular field $\theta(x)$ shifts as $\theta(x) \rightarrow \theta(x) + v\alpha(x)$. For any $v \neq 0$, we can take $\alpha = -\theta/v$ and set in this way $\theta(x) = 0$. In this gauge, denoted unitary gauge, the third term in eq. (8.4.4) boils down to a mass term for the $U(1)$ gauge field, $m_A = ev$. We can say that the gauge field has “eaten” the field θ which becomes the longitudinal component of a massive gauge field. This mechanism is commonly denoted Higgs mechanism, though several authors discovered it at the same time [35, 34, 23, 31]. The number of physical degrees of freedom (d.o.f.) in the process is unchanged. We started with 2 d.o.f. from the massless gauge field and 2 d.o.f. from the complex scalar field, for a total of 4, and ended up in 1 d.o.f. from the neutral field ρ and 3 d.o.f. from the massive gauge field, again for a total of 4.

Notice that in the limit $e \rightarrow 0$ the $U(1)$ symmetry becomes global and the gauge field decouples. In this limit the Goldstone’s theorem applies and we recover the massless NG boson θ .

All the above analysis similarly applies to non-abelian symmetries. Let us consider the set of real fields ϕ_n transforming as (8.3.2) under an infinitesimal local transformation of a group G . For simplicity, we assume the fields ϕ_n to be in some irreducible representation of G , although this is not strictly necessary. The Lagrangian of our system will include the terms

$$\mathcal{L} \supset \frac{1}{2} \left(\partial_\mu \phi_n - ig t_{nm}^\alpha A_\mu^\alpha \phi_m \right)^2 - V(\phi), \quad (8.4.5)$$

where we assume that the potential $V(\phi)$ has minima for $\phi_m = v_m$ such that, in the “ungauged” limit $g \rightarrow 0$, the global group G is spontaneously broken to H . Expanding around the non-trivial vacuum, $\phi_m = v_m + \phi'_m$, the above covariant derivative gives rise to the following terms, up to quadratic order in the fluctuations,

$$\mathcal{L} \subset \frac{1}{2} (\partial_\mu \phi'_n)^2 - ig \partial_\mu \phi'_n t_{nm}^\alpha v_m A_\mu^\alpha + \frac{g^2}{2} (it_{nm}^\alpha v_m)(it_{np}^\beta v_p) A_\mu^\alpha A^{\mu,\beta}. \quad (8.4.6)$$

When $\alpha, \beta = i$, the second and third terms in eq. (8.4.6) vanish identically, since $t_{nm}^i v_m = 0$, while they are generically non-vanishing for $\alpha, \beta = a$. The second term in eq. (8.4.6) is the generalization of the $\partial_\mu \theta A^\mu$ mixing term in the $U(1)$ case. Recall that the directions in field space given by $t_{nm}^a v_m$ correspond to the NG boson directions. As we will show below, it is always possible to choose a gauge (denoted again unitary gauge) in which the

fields ϕ_n do not contain NG boson fields, and hence

$$\phi'_n t_{nm}^\alpha v_m = 0. \quad (8.4.7)$$

The third term in eq. (8.4.6) is a mass term for the gauge fields in the broken directions. The mass matrix

$$\mu_{\alpha\beta}^2 = g^2 (it_{nm}^\alpha v_m) (it_{nl}^\beta v_l) \quad (8.4.8)$$

is symmetric, real and positive. Each gauge field A_μ^a along the broken directions gets a non-vanishing mass, by eating the corresponding would-be NG boson. In the basis in which μ^2 is diagonal, the propagators $G_{\mu\nu}^a$ for the massive gauge fields read

$$G_{\mu\nu}^{a(\text{UG})}(p) = \frac{-i}{p^2 - \mu_a^2} \left(\eta_{\mu\nu} - \frac{p_\mu p_\nu}{\mu_a^2} \right), \quad (8.4.9)$$

where UG stands for unitary gauge and μ_a^2 are the mass square eigenvalues. Notice that for large values of the momentum p , the gauge field propagator goes to a constant. The renormalizability properties of spontaneously broken gauge theories are correspondingly unclear. In order to fix this problem, it is sometimes more useful to use a more general class of gauge-fixing terms, denoted by ξ -gauges. The gauge fixing term $\mathcal{L}_{g.f.}$ for ξ -gauges is a generalization of the usual $1/(2\xi)(\partial_\mu A^\mu)^2$ term. It reads

$$\mathcal{L}_{g.f.} = -\frac{1}{2\xi} f_\alpha f_\alpha, \quad f_\alpha = \partial_\mu A_\alpha^\mu + i\xi g \phi'_n t_{nm}^\alpha v_m. \quad (8.4.10)$$

The associated ghost Lagrangian is obtained by taking an infinitesimal gauge variation of f_α , with parameter ω^α , $\mathcal{L}_{gh.} = -g\omega_\alpha^* \delta_\omega f_\alpha$. We get, modulo total derivatives,⁶

$$\mathcal{L}_{gh.} = (\partial^\mu \omega_\alpha^*) (D_\mu \omega_\alpha) - \xi g^2 (it_{nm}^\alpha v_m) (it_{nl}^\beta \phi_l) \omega_\alpha^* \omega_\beta. \quad (8.4.11)$$

From eq. (8.4.11) we see that the ghost fields along the broken directions have a mass square that equals the gauge boson mass matrix times ξ . For any ξ , the term $ig\xi(\phi' tv)_\alpha \partial_\mu A_\alpha^\mu$ appearing in eq. (8.4.10), upon an integration by parts, cancels the second unwanted term in eq. (8.4.6). The unitary gauge (8.4.7) corresponds to the limit $\xi \rightarrow \infty$ of the above class of gauges. In this limit, the term proportional to ξ in $\mathcal{L}_{g.f.}$ oscillates very rapidly and averages to zero any field configuration for which this term is non-zero. We then effectively recover the unitary gauge (8.4.7). We see from eq. (8.4.11) that in the unitary gauge the ghost fields become infinitely massive and can be neglected.

Let us derive the propagators for gauge fields, ghosts and scalar (NG and not) fields in an arbitrary ξ -gauge. In momentum space, setting μ^2 to diagonal form, the quadratic

⁶We take $\delta v_m = 0$, so that $\delta \phi'_m = \omega^\alpha i t_{mn}^\alpha \phi_n$.

terms of the gauge bosons in the Lagrangian add to

$$\mathcal{L}_{quad.}(A) = \frac{1}{2} A_\mu^\alpha(-p) A_\nu^\beta(p) \delta_{\alpha\beta} \left(-p^2 \eta^{\mu\nu} + p^\mu p^\nu \left(1 - \frac{1}{\xi} \right) + \eta^{\mu\nu} \mu_\alpha^2 \right), \quad (8.4.12)$$

valid for both broken and unbroken generators (in the latter case $\mu_i^2 = 0$). By inverting the above quadratic terms we get

$$G_{\mu\nu}^{\alpha\beta(\xi)}(p) = \frac{-i\delta_{\alpha\beta}}{p^2 - \mu_a^2} \left(\eta_{\mu\nu} - \frac{(1 - \xi)p_\mu p_\nu}{p^2 - \xi\mu_\alpha^2} \right). \quad (8.4.13)$$

The ghost propagator $G_{\alpha\beta}^{\omega(\xi)}(p)$ is easily obtained from eq. (8.4.11):

$$G_{\alpha\beta}^{\omega(\xi)}(p) = \frac{i\delta_{\alpha\beta}}{p^2 - \xi\mu_\alpha^2}. \quad (8.4.14)$$

The scalar propagators are different for NG and non-NG bosons. The total mass matrix in the Lagrangian reads

$$M_{mn}^2 = \frac{\partial^2 V}{\partial \phi_m \partial \phi_n} \Big|_{\phi=v} + \xi g^2 (it_{mp}^a v_p)(it_{nq}^a v_q). \quad (8.4.15)$$

The mass matrix (8.4.15), acting on the NG boson directions $(t_{ns}^b v_s)$, gives

$$M_{mn}^2(t_{ns}^b v_s) = \xi g^2 (it_{mp}^a v_p)(it_{nq}^a v_q)(t_{ns}^b v_s) = \xi \delta_{ab} \mu_a^2 (t_{mp}^a v_p), \quad (8.4.16)$$

where the first term in eq. (8.4.15) vanishes thanks to the relation (8.3.6). In a generic ξ -gauge, the would-be NG bosons have a mass which is $\sqrt{\xi}$ times the gauge boson mass. On the other hand, for the non-NG boson directions, defined by eq. (8.4.7), it is the second term in eq. (8.4.15) that vanishes, giving as masses the eigenvalues m_i^2 of $\partial^2 V / \partial \phi_m \partial \phi_n$. We then get as scalar propagators

$$\begin{aligned} G_{ij}^S(p) &= \frac{i\delta_{ij}}{p^2 - m_i^2}, \quad (\text{no NG bosons}) \\ G_{ab}^{S(\xi)}(p) &= \frac{i\delta_{ab}}{p^2 - \xi\mu_a^2}, \quad (\text{NG bosons}) \end{aligned} \quad (8.4.17)$$

where the index i runs over all, but the NG bosons, scalar directions.

In the unitary gauge $\xi \rightarrow \infty$, the NG bosons decouple and can be seen as eaten by the gauge fields. In any other gauge they should be kept. Notice how the gauge boson propagator (8.4.13) goes like $1/p^2$ for large momenta for any finite values of ξ and it is only in the unitary gauge that it behaves as p^0 . This makes any finite ξ -gauge more suitable than the unitary gauge to address renormalizability properties of spontaneously broken gauge theories. Common choices of ξ -gauges are the Landau gauge $\xi = 0$, in which ghosts and NG bosons are massless, and the Feynman gauge $\xi = 1$, in which the gauge propagator simplifies considerably.

8.5 The Goldstone Boson Equivalence Theorem**

The Goldstone boson equivalence theorem, for short denoted equivalence theorem (ET) in what follows, is a theorem about the behaviour of the scattering amplitude of longitudinally polarized gauge bosons in a spontaneously broken gauge theory.⁷ The theorem states that at high enough energies the scattering amplitude of longitudinally polarized gauge bosons is the same as the one obtained by replacing the gauge boson with its corresponding “eaten” Goldstone boson, considered as a physical state. The relevance of the ET relies in the fact that the scattering amplitude of massive gauge bosons is indeed dominated by the longitudinal polarizations at high energies.

More precisely, the ET states that for $E \gg m_A$

$$\epsilon_{L,\alpha_1}^{\mu_1}(p_1) \dots \epsilon_{L,\alpha_n}^{\mu_n}(p_n) S_{\mu_1 \dots \mu_n \dots}^{\alpha_1 \dots \alpha_n \dots}(p_1, \dots, p_n, \dots) = S^{\alpha_1 \dots \alpha_n \dots}(p_1, \dots, p_n, \dots) + \mathcal{O}\left(\frac{m_A}{E}\right). \quad (8.5.1)$$

In eq. (8.5.1), $S_{\mu_1 \dots \mu_n \dots}^{\alpha_1 \dots \alpha_n \dots}$ is the S -matrix element for the scattering of n longitudinally polarized vector bosons with mass m_A (taken equal for simplicity) with possibly other physical states, unspecified in eq. (8.5.1) and encoded in the second $\dots$ in $S^{\mu_1 \dots \mu_n \dots}$, while $S^{\alpha_1 \dots \alpha_n \dots}$ is the scattering of their corresponding n would-be Goldstone bosons ϕ (with the same physical states), treated as physical particles. In order to prove the ET we start by the generalization of eq. (6.3.24), which still applies in spontaneously broken gauge theories:

$$\langle \alpha | [Q, \mathcal{O}] | \beta \rangle = 0, \quad (8.5.2)$$

where $|\alpha\rangle$ and $|\beta\rangle$ are arbitrary physical states and $\mathcal{O}$ is any operator (not necessarily gauge invariant) of the theory. If we take $\mathcal{O} = \omega_\alpha^*$ and work with the auxiliary field H_α (recall section 6.3), we get

$$\langle \alpha | H_\alpha | \beta \rangle = 0. \quad (8.5.3)$$

In the R_ξ -gauges defined in eq. (8.4.10), solving for H_α we have

$$H_\alpha = \frac{1}{\xi} (\partial_\mu A_\alpha^\mu + i\xi g \phi_n' t_{nm}^\alpha v_m). \quad (8.5.4)$$

For simplicity, let us consider the situation where all massive gauge fields have a common mass m_A and let us define the would-be GB fields

$$m_A \pi^\alpha \equiv g \phi_n' t_{nm}^\alpha v_m. \quad (8.5.5)$$

⁷The ET implicitly appeared for the first time in [20] and was explicitly stated in [41]. A more detailed and careful derivation (although with a quite cumbersome notation and some unclear passages), including the proof in a generic R_ξ -gauge, has been given in [15]. For its clear presentation, that includes also further developments on the ET with respect to the original references, we recommend [5].

The relation (8.5.3) implies that any connected Green function of physical states involving H_α must vanish. Denoting collectively by Φ any physical field (transverse or longitudinal gauge field, physical scalar, matter field), we have

$$\langle \Phi(q_1) \dots \Phi(q_n) H_\alpha(p) \rangle = 0. \quad (8.5.6)$$

S-matrix elements are given by the amputated Green-functions. In terms of these, eq. (8.5.6) becomes

$$p^\mu G_{\mu\nu}^{(\xi)}(p) \langle \Phi(q_1) \dots \Phi(q_n) A^\nu(p) \rangle_{amp} = -\xi m_A G^{S(\xi)}(p) \langle \Phi(q_1) \dots \Phi(q_n) \pi(p) \rangle_{amp}, \quad (8.5.7)$$

where the propagators are those defined in eqs. (8.4.13) and (8.4.17). We have

$$p^\mu G_{\mu\nu}^{(\xi)}(p) = \frac{-ip_\nu(p^2 - \xi m_A^2 - (1 - \xi)p^2)}{(p^2 - m_A^2)(p^2 - \xi m_A^2)} = \frac{-ip_\nu \xi}{(p^2 - \xi m_A^2)} = -p_\nu \xi G^{S(\xi)}(p). \quad (8.5.8)$$

Plugging eq.(8.5.8) in eq.(8.5.7) gives

$$\frac{p^\mu}{m_A} \langle \Phi(q_1) \dots \Phi(q_n) A_\mu(p) \rangle_{amp} = \langle \Phi(q_1) \dots \Phi(q_n) \pi(p) \rangle_{amp}. \quad (8.5.9)$$

Notice that the ξ -dependence completely dropped. The longitudinal polarization for a vector with 4-momentum $p_\mu = (E, \vec{p})$ is given by

$$\epsilon_\mu^L(p) = \frac{1}{m_A} \left(|\vec{p}|, \frac{\vec{p}}{|\vec{p}|} E \right). \quad (8.5.10)$$

For $E \gg m_A$, $|\vec{p}| = E(1 + \mathcal{O}(m_A^2/E^2))$ and hence

$$\epsilon_\mu^L(p) = \frac{p_\mu}{m_A} + \mathcal{O}\left(\frac{m_A}{E}\right). \quad (8.5.11)$$

We finally get

$$\epsilon_{L,\alpha}^\mu(p) \langle \Phi(q_1) \dots \Phi(q_n) A_\mu(p) \rangle_{amp} = \langle \Phi(q_1) \dots \Phi(q_n) \pi_\alpha(p) \rangle_{amp} + \mathcal{O}\left(\frac{m_A}{E}\right). \quad (8.5.12)$$

The above derivation is easily generalized to multiple external longitudinal vector fields. If we take $\mathcal{O} = \omega_\alpha^* \prod_{i=1}^{n-1} H_{\alpha_i}$ in eq. (8.5.2), we immediately get that the Green functions with an arbitrary number of insertions of H_α must vanish. Proceeding as above leads to

$$\left(\prod_{i=1}^n \epsilon_{L,\mu_i}^{\alpha_i}(p_i) \right) \langle A_{\mu_1}^{\alpha_1}(p_1) \dots A_{\mu_n}^{\alpha_n}(p_n) \dots \rangle_{amp} = \langle \pi_{\alpha_1}(p_1) \dots \pi_{\alpha_n}(p_n) \dots \rangle_{amp} + \mathcal{O}\left(\frac{m_A}{E}\right), \quad (8.5.13)$$

that proves the relation (8.5.1). We conclude by noticing that the proof above applies only to tree-level amplitudes. Indeed, while the relation (8.5.2) and its generalizations are exact to all orders in perturbation theory, the form of the amputated Green functions crucially depends on the form of the propagator. In going from eq. (8.5.7) to eq. (8.5.9) we have used the tree-level form of the propagators. At the quantum level the relation (8.5.13) gets corrections, that however can be reabsorbed with a proper choice of the gauge fixing functional. We do not discuss these subtleties further.

8.6 Effective Field Theories for Broken Symmetries*

Massless NG bosons are a generic prediction of the Goldstone's theorem, no matter whether the underlying theory is weakly or strongly coupled. At sufficiently low energies we can then integrate out massive degrees of freedom and write down an EFT for the NG bosons, possibly interacting with other light fields, such as gauge fields. We will see in this section that, surprisingly enough, a lot can be said about this EFT without even knowing the underlying UV Lagrangian (both spectrum and interactions). All we need to know is the symmetry breaking pattern $G \rightarrow H$, namely the starting and final global groups. The key point is that the EFT should be not only invariant under the subgroup H , but under the entire group G (this is precisely what we mean by spontaneous symmetry breaking!). However, while the symmetries for H are manifest and linearly realized on the NG bosons, the ones for G/H are non-linearly realized and in general complicated. Group-theoretical considerations will allow us to find the proper way to repackage the NG bosons into fields that have relatively simple transformation properties so that a general EFT can be found [18, 13].

Let us start by writing down the general commutation relations among the generators $t_i \in \mathcal{H}$ and $t_a \in \mathcal{G}/\mathcal{H}$, $\mathcal{G}$ and $\mathcal{H}$ being the Lie algebras of the corresponding groups. We have

$$[t_i, t_j] = iC_{ijk}t_k, \quad [t_i, t_a] = iC_{iab}t_b, \quad [t_a, t_b] = iC_{abc}t_c + iC_{abi}t_i. \quad (8.6.1)$$

We take here the structure constants completely antisymmetric in its indices. Since H is a subgroup of G , $C_{ija} = -C_{iaj} = 0$. Let ϕ_m be the fields responsible for the $G \rightarrow H$ spontaneous symmetry breaking pattern (again in a basis where all the fields are real). Since we want to write down an EFT for the NG boson fields, integrating out the other model-dependent degrees of freedom present in the theory, we have to find a way to disentangle the NG bosons from the rest. In other words, we have to find a way to generalize for an arbitrary symmetry breaking pattern $G \rightarrow H$, the radial coordinates decomposition (8.4.3) that in the $U(1) \rightarrow \emptyset$ allows to disentangle the NG boson (θ) from the massive excitations (ρ). For this purpose let us define fields $\tilde{\phi}$ that do not contain NG boson field directions:

$$\phi(x) = \gamma(x)\tilde{\phi}(x), \quad \gamma(x) \in G, \quad (8.6.2)$$

namely such that

$$\tilde{\phi}^t(x)t_\alpha v = 0, \quad \forall \alpha. \quad (8.6.3)$$

In eq.(8.6.2) and in what follows we omit for simplicity the matrix indices m, n . In eq.(8.6.3) we clearly have $t_i v = 0$ and $t_a v \neq 0$, so this equation is non trivial only along

the broken directions. In this basis the NG bosons all sit in the matrix field $\gamma(x)$. The definition of $\tilde{\phi}$ is not unique. If $\tilde{\phi}$ satisfies (8.6.3), so does $\tilde{\phi}' = \tilde{\phi} + c_i t_i \tilde{\phi}$, for arbitrary coefficients c_i . Indeed,

$$\begin{aligned}\tilde{\phi}^t t_\alpha v &= (\tilde{\phi}^t + c_i \tilde{\phi}^t t_i^t) t_\alpha v = (\tilde{\phi}^t - c_i \tilde{\phi}^t t_i) t_\alpha v \\ &= -c_i \tilde{\phi}^t t_i t_\alpha v = -c_i \tilde{\phi}^t [t_i, t_\alpha] v = -i c_i C_{iab} \tilde{\phi}^t t_b v = 0.\end{aligned}\tag{8.6.4}$$

The field matrix $\gamma(x)$ is therefore defined only modulo x -dependent transformations under H : $\gamma(x) \sim \gamma(x)h(x)$, with $h(x) \in H$. Given this equivalence class, we can always choose as representative a $\gamma(x)$ of the form

$$\gamma(x) = e^{i\xi_a(x)t_a}, \tag{8.6.5}$$

where ξ^a are essentially the NG boson fields. Under a global G transformation, we have $\phi \rightarrow \phi' = g\phi = g\gamma(x)\tilde{\phi}(x)$. The field ϕ' can also be decomposed as in eq. (8.6.2). In general, we have

$$\phi'(x) = \gamma(\xi'_a(x))h(\xi_a(x), g)\tilde{\phi}(x), \tag{8.6.6}$$

for some $h(\xi_a(x), g) \in H$, so that $\phi' = \gamma(\xi'_a)\tilde{\phi}'$,⁸ with

$$\tilde{\phi}' = h(\xi_a, g)\tilde{\phi}, \tag{8.6.7}$$

$$\gamma(\xi'_a) = g\gamma(\xi_a)h^{-1}(\xi_a, g). \tag{8.6.8}$$

Notice that the field $\tilde{\phi}(x)$, under global G transformations, transforms as effectively *local* transformations under H . The transformations of ξ_a and $\tilde{\phi}$ simplify considerably when $g = h \in H$. The second commutation relation in eq. (8.6.1) shows that the broken generators t_a transform linearly under transformations of h , namely

$$ht_a h^{-1} = t_b R_{ba}(h), \tag{8.6.9}$$

where R is some representation (in general reducible) of H . The same commutation relation shows that when $g = h$, the factor $h(\xi(x), g)$ defined in eq. (8.6.6) reduces to the global x -independent element h : $h(\xi(x), g) = h$. We then have, using also eq. (8.6.9)

$$\gamma' \equiv \gamma(\xi') = h\gamma(\xi)h^{-1} = \gamma(R(h)\xi). \tag{8.6.10}$$

Under H , then, the transformations of both ξ and $\tilde{\phi}$ are simple and linear:

$$\begin{aligned}\xi'_a(x) &= R_{ab}(h)\xi_b(x), \\ \tilde{\phi}'(x) &= h\tilde{\phi}(x).\end{aligned}\tag{8.6.11}$$

⁸From now on we omit, for simplicity, the partial or total dependence on x , $\xi_a(x)$, etc. of the various fields.

On the other hand, for infinitesimal transformations along G/H , the NG fields ξ_a *transform as a shift*, at leading order in the field fluctuations. This is immediately visible from eq.(8.6.8). For $g = 1 + i\epsilon_a t_a$, indeed, we have

$$\delta\xi_a = \epsilon_a + \mathcal{O}(\xi^2, \epsilon\xi). \quad (8.6.12)$$

We now assume that the structure constants C_{abc} in the third commutator in eq. (8.6.1) vanish, in which case the coset G/H is called symmetric and the formalism describing the interactions of NG bosons can be simplified. We refer the reader to subsection 8.9 where a more general formalism allows us to also consider non-symmetric cosets. With $C_{abc} = 0$, the Lie algebra defined by the relations (8.6.1) is invariant under the $\mathbf{Z}_2$ symmetry R under which the broken generators change sign:

$$R(t_i) = t_i, \quad R(t_a) = -t_a. \quad (8.6.13)$$

We clearly have $R(\gamma) = \gamma^{-1}$. Under G transformations

$$\gamma' = g\gamma h^{-1}. \quad (8.6.14)$$

On the other hand

$$\begin{aligned} \gamma'^{-1} &= R(\gamma') = R(g)R(\gamma)R(h^{-1}) = R(g)\gamma^{-1}h^{-1} \\ &\implies \gamma' = h\gamma R(g)^{-1}. \end{aligned} \quad (8.6.15)$$

Using both eqs. (8.6.14) and (8.6.15), we see that

$$U' \equiv \gamma'^2 = g\gamma h^{-1}h\gamma R(g)^{-1} = gUR(g)^{-1}. \quad (8.6.16)$$

In other words, while γ does not transform linearly under transformations of G , its square $U = \gamma^2$ does. This is a significant simplification, because it allows us to no longer worry of the complicated factor $h(x)$.

We have found that an EFT for the NG bosons, invariant under G , is best written when the NG bosons are repackaged in the matrix U . Let us see which kind of invariant terms we can write using U and $U^\dagger$. Let us first focus on potential terms, i.e. no derivatives. Since $R(g) \neq g$, terms of the form U^n or $(U^\dagger)^n$ are forbidden, the only allowed invariant terms being trace of operators proportional to $U^\dagger U$. But $U^\dagger U = 1$, so no potential term is indeed allowed! This should not come as a surprise. We know that NG bosons must be massless. If a potential term would have been allowed, generally it would give a mass to them. The absence of any potential could have been anticipated by the shift symmetry (8.6.12), which is broken by any potential term. Non-trivial invariant operators can instead

be written using derivatives. The leading, two-derivative, term involving the NG bosons and invariant under the symmetry (8.6.16) is

$$\mathcal{L} = \frac{f_\pi^2}{8} \text{tr} \left(\partial_\mu U \partial^\mu U^\dagger \right) + \dots, \quad (8.6.17)$$

where f_π is a mass scale, related to the scale of spontaneous symmetry breaking, introduced for dimensional reasons, and $\dots$ stands for higher dimensional operators involving four or higher number of derivatives. The expansion of the the matrices U and $U^\dagger$ in powers of ξ^a in eq. (8.6.17) allows us to find the explicit form for an infinite number of leading interactions among the NG bosons, all fixed by a single term in the Lagrangian! The relation between the fields ξ_a and the canonical NG fields π_a defined in eq.(8.2.12) is obtained by computing the form of the broken currents J_μ^a from eq.(8.6.17). Using the trick of promoting a global symmetry to a local one, so that $\delta_\epsilon \mathcal{L} = -\partial_\mu \epsilon J^\mu$, we have,

$$\delta_\epsilon \mathcal{L} = f_\pi^2 \partial_\mu \epsilon_a \partial^\mu \xi_b \text{Tr} t^a t^b + \dots = f_\pi^2 \partial_\mu \epsilon^a \partial^\mu \xi_a + \dots, \Rightarrow J_\mu^a = -f_\pi^2 \partial_\mu \xi^a, \quad (8.6.18)$$

where we have used the leading order transformation (8.6.12) for the ξ 's and taken $\text{Tr} t_a t_b = \delta_{ab}$. Matching eqs.(8.2.8) and (8.2.12) with eq.(8.6.18) and demanding that the kinetic term (8.6.17) is canonical in terms of the π_a fields gives

$$\xi_a = \frac{F_{ab}}{f_\pi^2} \pi_b. \quad (8.6.19)$$

Since the NG fields transform linearly under H transformations, the decay constant matrix F_{ab} should correspondingly be invariant under such transformations. This is in general a strong constraint on the form of F_{ab} .

As we mentioned, all the interactions among the NG bosons, including their kinetic terms, are efficiently encoded in the single term (8.6.17). This is the first of an infinite series of higher derivative operators involving the matrices U . By expanding U in terms of field fluctuations, we see that the effective coupling constant of all the interaction terms that can arise from eq. (8.6.17) is E^2/f_π^2 . For sufficiently low-energies $E \ll f_\pi$, this term is the most important one, all the higher derivative interactions being suppressed by additional powers of E/f_π . The EFT Lagrangian (8.6.17) is of course non-renormalizable and becomes unreliable when $E \sim f_\pi$, in which case an alternative description is needed. More precisely, the energy scale where we expect our effective theory to break down is $E = \Lambda \simeq 4\pi f_\pi$, including the 4π factor coming from loops (analogue of the effective QED expansion parameter $e^2/(16\pi^2)$). This is the energy scale where the non-NG degrees of freedom that we have (implicitly) integrated out start to matter and must be included to possibly embed the theory in a consistent UV complete model.

8.6.1 Adding Gauge Fields*

Other light fields might appear in the EFT for the NG bosons, most notably gauge fields. Like the NG bosons, gauge particles can be naturally light or massless and their interactions are governed by gauge invariance. In fact, gauge interactions in this formalism do not pose any difficulty. Let us suppose that a subgroup H_g of G , distinct from H , is gauged. A convenient way to proceed is to pretend that the whole group is gauged, namely take $H_g = G$. Since U transforms linearly under G , the local symmetry is implemented in the usual way discussed in chapter 6 for matter fields; one introduces a gauge field A_μ and replace ordinary derivatives of U with covariant ones. Under a gauge transformation g we have

$$A_\mu^g = g A_\mu g^{-1} - i(\partial_\mu g)g^{-1}, \quad (8.6.20)$$

where $A_\mu = A_\mu^i t_i + A_\mu^a t_a$ is the full (non-canonically normalized) gauge field. Given the transformation (8.6.16) and the above relation, the covariant derivative

$$D_\mu U = \partial_\mu U - i A_\mu U + i U A_\mu^R, \quad (8.6.21)$$

where $A_\mu^R = A_\mu^i t_i - A_\mu^a t_a$, transforms as $D_\mu U^g = g(D_\mu U)R(g)^{-1}$. Since the theory is now invariant under local transformations of G , by a proper choice of g we might completely remove the NG bosons, i.e. we can set $U = 1$. This is nothing else than the unitary gauge in a non-linear realization of the symmetry. In this gauge the covariantization of eq.(8.6.17) leads to mass terms for the gauge fields $A_\mu - A_\mu^R = 2A_\mu^a t_a$, namely to those along the broken direction, as expected from the Higgs mechanism.

The generic case $H_g \subset G$ can be deduced from the case $H_g = G$ by simply switching off the gauge fields that are not in the direction of H_g . Denoting by $t^{\hat{\alpha}}$ the generators of G that are gauged, in general we have

$$t^{\hat{\alpha}} = g_{\hat{\alpha}\alpha} t^\alpha \quad (8.6.22)$$

where $g_{\hat{\alpha}\alpha}$ are the gauge coupling constants associated to the gauging, all different in the most general case. We now set

$$A_\mu^\alpha = g_{\hat{\alpha}\alpha} A_\mu^{\hat{\alpha}}, \quad (8.6.23)$$

where $A_\mu^{\hat{\alpha}}$ are the actual dynamical gauge fields. In the unitary gauge, as expected, only the gauge fields along the broken directions get a mass with the corresponding NG bosons being eaten by them. Notice that gauging a subgroup of H_g explicitly breaks the global symmetry G , since it determines a specific direction in field space. Taking into account the normalization (8.6.19) of the broken generators, the gauge field mass matrix reads

$$\mu_{\hat{\alpha}\hat{\beta}}^2 = F_{ac} F_{ad} g_{\hat{\alpha}c} g_{\hat{\beta}d}. \quad (8.6.24)$$

This relation is the generalization of eq. (8.4.8). The latter only applies for weakly coupled descriptions of the spontaneous symmetry breaking in terms of free fields.

8.7 $SU(3)_V \times SU(3)_A \rightarrow SU(3)_V$: Mesons in QCD*

Consider QCD with three active quarks (say, the up, down and strange quark) in the limit in which they are all massless:

$$\mathcal{L}_{QCD} = -\frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu} + \bar{Q} i \not{D} Q, \quad (8.7.1)$$

where $Q = (u \ d \ s)^t$.⁹ In addition to the $SU(3)_c$ gauge (color) symmetry, the Lagrangian (8.7.1) is invariant under an additional $SU(3)_V \times SU(3)_A$ global (flavour) symmetry acting on Q as

$$Q \rightarrow \exp(i\theta_a^V \lambda_a + i\theta_a^A \lambda_a \gamma_5) Q, \quad (8.7.2)$$

where λ_a are the $SU(3)$ Gell-Mann matrices, normalized such that $\text{tr } \lambda_a \lambda_b = \delta_{ab}/2$ in the fundamental representation. At sufficiently low energies, when the QCD coupling constant becomes strong, a quark bilinear gets a non-vanishing vacuum expectation value:

$$\langle \bar{Q}_i Q_j \rangle = \hat{\Lambda}^3 \delta_{ij}, \quad (8.7.3)$$

that breaks $G = SU(3)_V \times SU(3)_A$ down to $H = SU(3)_V$. Here $\hat{\Lambda}$ is the scale where the chiral symmetry breakdown occurs. The eight NG bosons are encoded in the matrix field

$$\gamma(x) = e^{i\lambda^a \xi^a(x)}. \quad (8.7.4)$$

The $SU(3)_V$ unbroken symmetry fixes the parameters F_{ab} to be proportional to the identity. We take $F_{ab} \equiv \delta_{ab} f_\pi$ in eq.(8.6.19) and write the 8 NG bosons as

$$\lambda_a \xi_a = \frac{1}{f_\pi} \begin{pmatrix} \frac{1}{\sqrt{2}}\pi_0 + \frac{1}{\sqrt{6}}\eta^0 & \pi^+ & K^+ \\ \pi^- & -\frac{1}{\sqrt{2}}\pi_0 + \frac{1}{\sqrt{6}}\eta^0 & K^0 \\ \bar{K}^- & \bar{K}^0 & -\sqrt{\frac{2}{3}}\eta^0 \end{pmatrix}, \quad (8.7.5)$$

where f_π is identified with the pion decay constant: $f_\pi \simeq 92$ MeV. With this normalization the 8 spin zero mesons $\pi^\pm$, π^0 , etc. are all canonically normalized.

The commutator of two vector or two axial transformations is a vector transformation, while the commutator of a vector and axial transformation is an axial transformation. The schematic commutation relations for $SU(3)_V \times SU(3)_A$ are then $[V, V] = V$, $[V, A] = A$, $[A, A] = V$, where V and A schematically represent the $SU(3)_V$ and $SU(3)_A$ generators.

⁹The analysis in this section closely follows section 19.7 of ref. [60].

This is a symmetric coset space, being this algebra invariant under the automorphism action $V \rightarrow V$, $A \rightarrow -A$. We can also construct L and R transformations defined as $V = (L + R)/2$, $A = (L - R)/2$, under which the above automorphism exchanges L and R : $L \leftrightarrow R$. According to the general results of section 8.6, the matrix $U = \gamma^2$ transforms homogeneously under g . We have¹⁰

$$U \rightarrow RUL^\dagger, \quad (8.7.6)$$

with obvious notation. In other words, U transforms as a $(\mathbf{3}, \bar{\mathbf{3}})$ representation of $SU(3)_R \times SU(3)_L$. Correspondingly, the kinetic terms and leading interactions among the 8 mesons π , K and η are collected in the single term (8.6.17). As we have seen in the previous section, the effective field theory is expected to break down at $E = \Lambda \simeq 4\pi f_\pi \sim 1$ GeV. It is natural to assume that $\Lambda \simeq \hat{\Lambda}$, the scale where chiral symmetry breaking occurs. This scale is related to the scale Λ_{QCD} where quarks confine (confinement scale), but it is *not* the same scale. They are conceptually different and there exist vacua in gauge (supersymmetric) theories where it is believed that quarks confine but no chiral symmetry breaking occurs.

In the real world, of course, the up, down and strange quarks are not massless and the spin zero mesons π , K and η are not exact NG bosons. The two things are related. The actual up, down and strange masses explicitly break the $SU(3)_V \times SU(3)_A$ symmetry, but are small enough, compared to Λ , to be considered as a perturbation of the QCD Lagrangian (8.7.1) (this would not be the case for the charm, bottom and top, that are all heavier than Λ). We can get the relation between the quark and meson masses by using the following trick. We formally promote the quark mass term

$$M = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{pmatrix}, \quad (8.7.7)$$

to be an external source that transforms under $SU(3)_L \times SU(3)_R$ as $M \rightarrow LMR^\dagger$, so that the term $\bar{Q}_L M Q_R + h.c.$ is made invariant. We have morally promoted the mass term to a so called *spurion field*, whose vacuum expectation value coincides with the mass matrix M and whose dynamics is frozen. Spurions are often used in QFT to formally restore explicitly broken symmetries. In so doing, one proceeds in the computation demanding the full symmetry, and only at the end of the computation one plugs back the original form

¹⁰There is actually no way to distinguish an $SU(3)_L$ from an $SU(3)_R$ transformation when acting on a 3×3 matrix such as U . Strictly speaking, we should write U as a 6×6 matrix, containing U and $U^\dagger$, to properly distinguish the two $SU(3)$'s. In doing that, one gets eq. (8.7.6).

of the breaking term into the spurion field. This is a very efficient way to make manifest how breaking terms affect physical quantities. Thanks to this trick, at low-energies we can construct $SU(3)_L \times SU(3)_R$ invariant terms using both U and M . If M is small, we can consider terms linear in M only. We have

$$\mathcal{L}_m = cf_\pi^3 \text{tr} UM + h.c. , \quad (8.7.8)$$

where c is an undetermined dimensionless coefficient. By plugging in eq. (8.7.8) the actual form of the NG fields (8.7.5) and of the mass term (8.7.7) we get, after some algebra:

$$\begin{aligned} m_{\pi^\pm}^2 &= C(m_u + m_d) , \\ m_{K^\pm}^2 &= C(m_u + m_s) , \\ m_{K^0}^2 &= C(m_d + m_s) , \end{aligned} \quad (8.7.9)$$

where $C = 2cf_\pi$. The π^0 and η^0 mix with each other. Expanding for $m_{u,d} \ll m_s$ we get

$$\begin{aligned} m_{\pi^0}^2 &\simeq C(m_u + m_d) , \\ m_\eta^2 &\simeq C \frac{4m_s + m_u + m_d}{3} . \end{aligned} \quad (8.7.10)$$

It is surprising that using symmetries only we can fix the meson masses in terms of the quark masses and one unknown coefficient C . The above relations are a particular example of eq. (8.3.9), and show how the masses of the pseudo NG bosons are linearly dependent on the source of the symmetry breaking violating term in the Lagrangian. In deriving eqs. (8.7.9) and (8.7.10) we have actually neglected another relevant source of explicit breaking of the $SU(3)_L \times SU(3)_R$ symmetry, given by the electric charge. As explained in section 8.6, in this EFT framework electromagnetic interactions are introduced by identifying a proper $U(1)$ subgroup and gauge it. This is a subgroup of the unbroken $SU(3)_V$, and its explicit form is found by looking at the infinitesimal transformations of the quark fields Q . We have $\delta_{\epsilon, EM} Q = i\epsilon Q_{el} Q$, where

$$Q_{el} = \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & -\frac{1}{3} \end{pmatrix} , \quad (8.7.11)$$

according to the actual electric charges of the up, down and strange quarks. We therefore get the leading order Lagrangian

$$\mathcal{L}_{Kin+\gamma} = \frac{f_\pi^2}{4} \text{tr} \left(D_\mu U D^\mu U^\dagger \right) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (cf_\pi^3 \text{tr} UM + h.c.) , \quad (8.7.12)$$

with

$$D_\mu U = \partial_\mu U + ieQ_{el} U A_\mu - ieU Q_{el} A_\mu . \quad (8.7.13)$$

It is straightforward to check that eq. (8.7.12) contains the expected interactions of the QED form for the charged mesons $\pi^\pm$ and $K^\pm$. These interactions, at the quantum level, contribute to the mass of the charged mesons by an amount Δ_{el} , which is the same, at leading order, for $\pi^\pm$ and $K^\pm$, thanks to the $SU(3)_V$ global symmetry. Eventually, the effect of the QED interactions is to shift the charged meson masses in eq. (8.7.10) by this amount Δ_{el} :

$$\begin{aligned} m_{\pi^\pm}^2 &\rightarrow m_{\pi^\pm}^2 + \Delta_{el}, \\ m_{K^\pm}^2 &\rightarrow m_{K^\pm}^2 + \Delta_{el}. \end{aligned} \quad (8.7.14)$$

The neutral meson masses, at this order, are unaffected by QED interactions. There are various ways in which the mass formulas (8.7.9), (8.7.10) and (8.7.14) can be used. For instance, the following relation holds, independently of C , Δ_{el} and the quark masses:

$$3m_\eta^2 + 2m_{\pi^\pm}^2 - m_{\pi^0}^2 = 2m_{K^\pm}^2 + 2m_{K^0}^2 \quad (8.7.15)$$

and it is experimentally very well reproduced. Alternatively, we might use the meson masses as experimental input parameters to compute the quark masses or, better, their ratios. One has

$$\begin{aligned} \frac{m_u}{m_s} &= \frac{2m_{\pi^0}^2 - m_{\pi^\pm}^2 + m_{K^\pm}^2 - m_{K^0}^2}{m_{K^\pm}^2 + m_{K^0}^2 - m_{\pi^\pm}^2} \simeq 0.027, \\ \frac{m_d}{m_s} &= \frac{m_{\pi^\pm}^2 + m_{K^0}^2 - m_{K^\pm}^2}{m_{K^\pm}^2 + m_{K^0}^2 - m_{\pi^\pm}^2} \simeq 0.050, \end{aligned} \quad (8.7.16)$$

where we have inserted in the last relation the approximate masses for the spin zero mesons.

The actual global symmetry of the QCD Lagrangian (8.7.1) is $U(3)_V \times U(3)_A$, rather than $SU(3)_V \times SU(3)_A$, since the $U(1)_V \times U(1)_A$ transformations

$$Q \rightarrow \exp(i\theta^V + i\theta^A \gamma_5) Q \quad (8.7.17)$$

are a symmetry of (8.7.1). The $U(1)_V$ symmetry is nothing else than the $U(1)$ baryon number and is unbroken, while $U(1)_A$ is manifestly broken by the quark condensate in eq. (8.7.3). Correspondingly, we should have a ninth pseudo NG boson, called η' . It can be shown, by adding the η' to the matrix field (8.7.4), that its mass $m_{\eta'}$ is bounded to be $m_{\eta'} \leq \sqrt{3}m_\pi$. But no such particle has been observed within this range of energies. This puzzle, denoted the $U(1)_A$ problem, has been solved by noticing that the $U(1)_A$ symmetry is explicitly broken. The classical symmetry does not survive quantum corrections and is said to be anomalous. Anomalies will be the subject of chapter 9. The size of the explicit

breaking is larger than the one given by the quark masses and is governed by instantons, non-perturbative effects that are outside the content of these lectures. As a consequence, the actual η' has a much larger mass, $m_{\eta'} \simeq 960$ MeV.

8.8 $SO(5) \rightarrow SO(4)$: A Composite Higgs?^{**}

The example that follows is a bit more exotic and refers to the electroweak sector of the Standard Model (SM). We have briefly seen in section 7.6 that scalar masses are unnatural, in the sense that they are quadratically sensitive to UV physics that push their values to that scale. This problem clearly applies to *elementary* scalars, including the Higgs boson in the SM. But what if the Higgs is not elementary, but rather a composite particle, very much like the pion? We do not have a hierarchy problem for the pion mass, since we know that this particle is composite and at some energies of order Λ_{QCD} , “diffuse” in its quark constituents. Moreover, pions are actually naturally lighter than Λ_{QCD} itself because, as we have extensively seen in the previous subsection, are pseudo NG bosons of an approximate global symmetry.

Before developing on this idea of considering the Higgs particle as a pseudo NG boson of some spontaneously broken global symmetry, it is useful to recall the symmetries of the standard, elementary Higgs Lagrangian, in the SM. In the global limit in which we switch off the $SU(2)_L \times U(1)_Y$ gauge couplings, the SM Higgs sector is

$$\mathcal{L}_H^{SM} = (\partial_\mu H)^\dagger (\partial^\mu H) + m^2 H^\dagger H - \frac{\lambda}{2} (H^\dagger H)^2, \quad (8.8.1)$$

where H is an $SU(2)_L$ doublet of the form

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} h_2 + ih_1 \\ h_4 - ih_3 \end{pmatrix}. \quad (8.8.2)$$

A closer look at eq. (8.8.1) reveals that the Higgs Lagrangian is invariant under an $SO(4)$, rather than $SU(2)_L$, global symmetry. This is manifest if we recast the four real Higgs components h_i ($i = 1, 2, 3, 4$) in a 4-plet and notice that eq. (8.8.1) is invariant under the transformations $h_i \rightarrow O_{ij} h_j$, with $O \in SO(4)$. It is actually more convenient to exploit the local isomorphism of $SO(4)$ with $SU(2) \times SU(2)$. The fundamental **4** of $SO(4)$ becomes a bidoublet **(2, 2)** of $SU(2) \times SU(2)$. One of the two $SU(2)$ is identified with the original $SU(2)_L$, and we denote by $SU(2)_R$ the other one. We can easily write the $SO(4)$ generators in the $SU(2)_L \times SU(2)_R$ basis. The 6 anti-symmetric hermitian generators of $SO(4)$ are proportional to

$$t_{ij}^{ab} = -t_{ij}^{ba} = \delta_i^a \delta_j^b - \delta_i^b \delta_j^a, \quad (8.8.3)$$

where $a, b = 1, \dots, 4$ label the generators and i, j their matrix components. A simple check of the algebra reveals that the combinations

$$\begin{aligned} t_L^1 &= -\frac{i}{2}(t^{23} + t^{14}), & t_L^2 &= -\frac{i}{2}(t^{31} + t^{24}), & t_L^3 &= -\frac{i}{2}(t^{12} + t^{34}), \\ t_R^1 &= -\frac{i}{2}(t^{23} - t^{14}), & t_R^2 &= -\frac{i}{2}(t^{31} - t^{24}), & t_R^3 &= -\frac{i}{2}(t^{12} - t^{34}), \end{aligned} \quad (8.8.4)$$

satisfy the commutation relations of the $SU(2)_L$ and $SU(2)_R$ algebras, respectively. The action of $SU(2)_L \times SU(2)_R$ is best seen by writing explicitly the Higgs field as a bidoublet using the 2×2 matrices $\sigma^\mu = (1, i\sigma_k)$ ($k = 1, 2, 3$):

$$H_{BD} = \frac{h_4 1_2 + ih_k \sigma_k}{2} = \frac{1}{2} \begin{pmatrix} h_4 + ih_3 & h_2 + ih_1 \\ -h_2 + ih_1 & h_4 - ih_3 \end{pmatrix}. \quad (8.8.5)$$

Under $SU(2)_L \times SU(2)_R$, the bidoublet (8.8.5) transforms as

$$H_{BD} \rightarrow g_L H_{BD} g_R^\dagger, \quad (8.8.6)$$

with $g_L \in SU(2)_L$, $g_R \in SU(2)_R$. It is manifest from eq. (8.8.6) that when the Higgs field develops a non-vanishing VEV, say $\langle h_4 \rangle \equiv v \neq 0$, i.e. $\langle H_{BD} \rangle \propto 1_2$, the $SU(2)_L \times SU(2)_R$ global symmetry is broken to the diagonal $SU(2)_c$ symmetry, with $g_L = g_R$. This unbroken $SU(2)_c$ global symmetry is called custodial symmetry and plays an important role in establishing that the W and Z boson mass ratio at tree-level is

$$\frac{m_Z^2}{m_W^2} = \frac{g^2 + g'^2}{g^2}, \quad (8.8.7)$$

where g and g' are the $SU(2)_L$ and $U(1)_Y$ gauge coupling constants. In this formalism, the gauge fields are introduced by gauging the whole $SU(2)_L$ global group and a subgroup of $SU(2)_R$, along the σ_3 generator, which is identified with the $U(1)_Y$ symmetry. The full, gauged, Higgs Lagrangian reads

$$\mathcal{L}_H^{SM} = \text{tr} (D_\mu H_{BD})^\dagger (D^\mu H_{BD}) + m^2 \text{tr} H_{BD}^\dagger H_{BD} - \frac{\lambda}{2} \text{tr} (H_{BD}^\dagger H_{BD})^2, \quad (8.8.8)$$

where

$$D_\mu H_{BD} = \partial_\mu H_{BD} - ig W_\mu^L H_{BD} + ig' H_{BD} W_\mu^R, \quad (8.8.9)$$

and

$$W_\mu^L = \frac{1}{2} \sigma_i W_\mu^i, \quad W_\mu^R = \frac{1}{2} \sigma_3 B_\mu. \quad (8.8.10)$$

The custodial $SU(2)_c$ symmetry is broken by the hypercharge coupling g' only.¹¹ If we set $g' = 0$, the Lagrangian (8.8.8) is invariant under $SU(2)_L \times SU(2)_R$ global transformations, provided we rotate $W_\mu^L \rightarrow g_L W_\mu^L g_L^\dagger$. The 3 would-be NG bosons associated to the

¹¹The fermion Yukawa couplings also break $SU(2)_c$, but here for simplicity we are focusing on the bosonic sector of the SM, neglecting fermions altogether.

$SU(2)_L \times SU(2)_R \rightarrow SU(2)_c$ breaking are all eaten by the W^L 's, and $SU(2)_c$ ensures that the 3 vector bosons all have equal masses $m_W = gv/2$. When $g' \neq 0$, the custodial symmetry is broken and the 3 massive gauge fields no longer have equal masses. However, the breaking pattern fixes their tree-level masses. From eq. (8.8.9) we immediately see that the gauge bosons along the off-diagonal components of $SU(2)_L$ are unaffected by g' and retain their masses $m_W = gv/2$. On the other hand, along the $U(1)_L \times U(1)_R$ gauged subgroups in the σ_3 directions, the gauge field along the unbroken symmetry remains massless (the photon) and the orthogonal direction (the Z) gets a mass equal to $m_Z = \sqrt{g^2 + g'^2}v/2$. The importance of emphasizing the symmetries of the Higgs Lagrangian becomes clear when we replace the Higgs field by an unspecified sector responsible for the electroweak symmetry breaking. In this more general context, we would conclude that the W and Z boson masses are given by eq. (8.6.24), with $\hat{\alpha} = 1L, 2L, 3L, 3R$. More explicitly, we have

$$g_{aLb} = g\delta_{ab}, \quad g_{aRb} = g'\delta_{ab}\delta_{a3}, \quad (8.8.11)$$

where $a, b = 1, 2, 3$ run over the broken $SU(2)$ generators. The crucial point is played by the custodial $SU(2)_c$ symmetry, that forces the F_{ab} terms to be proportional to the identity (like $SU(3)_V$ in the QCD example before): $F_{ab} = \delta_{ab}F$. Putting all together, the gauge boson mass matrix is of the form

$$\mu^2 = F^2 \begin{pmatrix} g^2 & 0 & 0 & 0 \\ 0 & g^2 & 0 & 0 \\ 0 & 0 & g^2 & gg' \\ 0 & 0 & gg' & g'^2 \end{pmatrix}. \quad (8.8.12)$$

We automatically recover the SM gauge boson masses, in particular the relation (8.8.7), with the identification $F = v/2$. We conclude by stressing that any sector replacing the usual Higgs in the SM Lagrangian will give the correct leading order electroweak gauge boson masses, provided it includes an $SU(2)_c$ global symmetry.

After this long, but necessary, digression on the SM, let us come back to our idea of the Higgs field as a pseudo NG boson. Assuming this idea, the Higgs should be a composite of certain constituents, analogues of quarks and gluons in QCD, that appears at strong coupling at some energy scale. Being the analogue of the pion in QCD, the Higgs particle is expected to be the lightest resonance of the strongly coupled theory, that might include additional heavier resonances (in analogy with the hadron spectroscopy in QCD). Of course, in contrast to the QCD case, we do not know here what is the UV theory that becomes strongly coupled. But we have by now learned that a lot can be said about the dynamics of pseudo NG bosons, by only specifying the group theoretical structure

of the spontaneous symmetry breaking pattern. As far as we are concerned, we have to assume that the UV theory, no matter what it is, has an approximate global symmetry group G , spontaneously broken to H , such that the NG bosons along the G/H directions have the quantum numbers of the SM Higgs. Let us denote by f_H the scale where this breaking occurs. The unbroken group H should be large enough to accommodate an $SU(2)_L \times U(1)_Y$ subgroup, that we will gauge and identify with the SM electroweak gauge group. Finally, we might also demand that H includes the custodial symmetry $SU(2)_c$ that ensures the correct tree-level mass ratio (8.8.7) between the SM gauge bosons. The minimal groups that give rise to the 4 Higgs NG bosons and nothing else are

$$\begin{aligned} G = SU(3) &\rightarrow H = SU(2) \times U(1), \\ G = SO(5) &\rightarrow H = SO(4) \cong SU(2)_L \times SU(2)_R. \end{aligned} \quad (8.8.13)$$

This is understood by looking at how the adjoint representations of $SU(3)$ and $SO(5)$ decompose under the breaking pattern above. We have

$$\begin{aligned} \mathbf{8} &\rightarrow \mathbf{3}_0 \oplus \mathbf{1}_0 \oplus \mathbf{2}_{1/2} \oplus \bar{\mathbf{2}}_{-1/2}, \\ \mathbf{10} &\rightarrow \mathbf{6} \oplus \mathbf{4} = (\mathbf{1}, \mathbf{3}) \oplus (\mathbf{3}, \mathbf{1}) \oplus (\mathbf{2}, \mathbf{2}). \end{aligned} \quad (8.8.14)$$

The subscript in the first line of eq. (8.8.14) refers to the $U(1)$ charge and in the second line we have reported the decomposition in terms of $SO(4)$ and $SU(2)_L \times SU(2)_R$ representations, respectively. The broken generators in the $SU(3)$ case transform as the last two terms in the first line of eq. (8.8.14), and the corresponding NG bosons form a complex doublet of $SU(2)$. The broken generators in the $SO(5)$ case transform as the last term in the second line of eq. (8.8.14), and the corresponding NG bosons form a 4-plet of $SO(4)$ or a bidoublet of $SU(2)_L \times SU(2)_R$. Both NG bosons have the quantum numbers of the ordinary Higgs field. However, the $SU(3) \rightarrow SU(2)$ case does not include the $SU(2)_c$ custodial symmetry, since the only unbroken $SU(2)$ must be the $SU(2)_L$ group. We then focus on the $SO(5) \rightarrow SO(4)$ case in the following.

The $SO(5)$ generators are as in eq. (8.8.3), with the indices a, b, i, j running now from 1 to 5, rather than from 1 to 4. The $SO(4)$ subgroup can be taken to be the one generated by the matrices (8.8.4), with the understanding that they are now 5×5 matrices with an additional row and column of zeros. The remaining $SO(5)/SO(4)$ broken generators are given by

$$t^a = -\frac{i}{\sqrt{2}} t^{a5}, \quad a = 1, 2, 3, 4. \quad (8.8.15)$$

The 4 NG bosons h_a are encoded as usual in the matrix

$$\gamma = e^{i\sqrt{2}\frac{h_a t_a}{f_H}}, \quad (8.8.16)$$

where the $\sqrt{2}$ factor arises from our choice of normalizations of the $SO(5)$ generators. It is straightforward to verify that G/H is a symmetric coset space, invariant under the automorphism (8.6.13), with $t_i \in SO(4)$ ($i = 1, \dots, 6$) and $t_a \in SO(5)/SO(4)$ ($a = 1, \dots, 4$). The matrix $U = \gamma^2$ transforms homogeneously under $SO(5)$ transformations, as in eq. (8.6.16). For a generic gauging, the covariant derivative for U is given in eq. (8.6.21). In our case, the gauging is all within H , so that A_μ^a all vanish and $A_R = A$. The covariant derivative (8.6.21) reduces to (now for canonically normalized gauge fields, including gauge coupling constant factors)

$$D_\mu U = \partial_\mu U + i(gW_\mu^L + g'W_\mu^R)U - iU(gW_\mu^L + g'W_\mu^R), \quad (8.8.17)$$

with

$$W_\mu^L = \sum_{\alpha=1}^3 t_L^\alpha W_\mu^\alpha, \quad W_\mu^R = t_R^3 B_\mu. \quad (8.8.18)$$

The Higgs Lagrangian associated to this “composite Higgs” scenario is

$$\mathcal{L}_H^{CH} = \frac{f_H^2}{16} \text{tr}(D^\mu U)^\dagger D_\mu U. \quad (8.8.19)$$

The NG nature of the Higgs forbids a potential for H , in the limit in which the $SO(5)$ global symmetry is exact and only spontaneously broken to $SO(4)$. But like $SU(3)_V$ in the QCD case is broken by $U(1)_{EM}$, the $SO(5)$ symmetry is explicitly broken by the $SU(2)_L \times U(1)_Y$ gauging¹². This implies that a potential for the Higgs field, even if not included at tree-level, will be generated by radiative effects. We will not elaborate more on this idea, but it suffices to say here that when fermions are also included and coupled to the Higgs matrix field U , electroweak symmetry breaking can in fact be induced at the quantum level. When $h_4 \neq 0$, a straightforward computation shows that the Lagrangian (8.8.19) gives rise to the following mass terms for the SM gauge fields:

$$m_W^2 = \frac{1}{4}g^2 f_H^2 \sin^2\left(\frac{\langle h_4 \rangle}{f_H}\right), \quad m_Z^2 = \frac{1}{4}(g^2 + g'^2) f_H^2 \sin^2\left(\frac{\langle h_4 \rangle}{f_H}\right), \quad m_\gamma^2 = 0. \quad (8.8.20)$$

As expected, thanks to the $SU(2)_c$ custodial symmetry, the mass ratio (8.8.7) is reproduced. We see that for $\langle h_4 \rangle/f_H \rightarrow 0$ we can expand the sin factor and recover the usual SM formula for the W 's and Z , where we identify

$$\langle h_4 \rangle = v \equiv \sqrt{\frac{\sqrt{2}}{2G_F}} \simeq 246 \text{ GeV}. \quad (8.8.21)$$

¹² $SO(5)$ might also be broken by additional terms, such as the analogues of the quarks masses in QCD. Without knowing the UV theory, we might for simplicity assume that there are no such terms.

This is the limit in which we push to very high energies the $SO(5) \rightarrow SO(4)$ breaking pattern, but in so doing we recover the hierarchy problem. On the other hand, various phenomenological bounds constrain the more natural limit $\langle h_4 \rangle \simeq f_H$, so that a little tuning is needed to achieve a mild separation between $\langle h_4 \rangle$ and f_H . Notice that away from the SM limit $f_H \rightarrow \infty$, $\langle h_4 \rangle$ does not coincide with v , as defined in eq. (8.8.21). From eq. (8.8.20), for finite f_H , we get

$$v = f_H \sin \left(\frac{\langle h_4 \rangle}{f_H} \right). \quad (8.8.22)$$

8.9 Effective Field Theories for Broken Symmetries: Generale Case**

In this section we consider the general construction of the effective field theory of NG bosons, discussed for the first time by Callan, Coleman, Wess and Zumino [18, 13]. This slightly more elaborated formalism allows to consider arbitrary symmetry breaking patterns.

The analysis made in section 8.6 continues to be valid up to eq.(8.6.12) included. If the coset is not symmetric, it is not particularly useful to consider the matrix $U = \gamma^2$ and a more elaborate analysis is needed. It is useful in this case to study the transformation properties of the field combination $\gamma^{-1} \partial_\mu \gamma$. By recalling the formula

$$e^X Y e^{-X} = Y + [X, Y] + \frac{1}{2!} [X, [X, Y]] + \frac{1}{3!} [X, [X, [X, Y]]] + \dots \quad (8.9.1)$$

valid for arbitrary matrices X and Y , we see that the combination $\gamma^{-1} \partial_\mu \gamma$ is a field defined in the algebra $\mathcal{G}$ of the group G . As such, it can be decomposed as

$$\gamma^{-1} \partial_\mu \gamma = it_a D_\mu^a(x) + it_i E_\mu^i(x), \quad (8.9.2)$$

where necessarily $D_\mu^a(x) = D^{ab}(\xi(x)) \partial_\mu \xi_b(x)$, $E_\mu^i(x) = E^{ia}(\xi(x)) \partial_\mu \xi_a(x)$, for some fields D^{ab} and E^{ia} . Under a global transformation $g \in G$ we clearly have

$$(g\gamma)^{-1} \partial_\mu (g\gamma) = \gamma^{-1} \partial_\mu \gamma. \quad (8.9.3)$$

On the other hand $g\gamma = \gamma' h$ and hence, after multiplication by h^{-1} to the right, we get

$$\gamma^{-1} \partial_\mu \gamma' = h(\gamma^{-1} \partial_\mu \gamma) h^{-1} - (\partial_\mu h) h^{-1}. \quad (8.9.4)$$

Using the decomposition (8.9.2) for γ' and identifying both sides, we get the transformations of D_μ^a and E_μ^i under global G transformations:

$$\begin{aligned} t_a D_\mu^a(\xi') &= h t_a h^{-1} D_\mu^a(\xi) = t_a R_{ab}(h) D_\mu^b(\xi), \\ t_i E_\mu^i(\xi') &= h t_i h^{-1} E_\mu^i(\xi) + i(\partial_\mu h) h^{-1}. \end{aligned} \quad (8.9.5)$$

Since H is a subgroup of G , we have

$$ht_i h^{-1} = R_{ij}(h)t_j, \quad (\partial_\mu h)h^{-1} = it_i R_{ia}(h)\partial_\mu \xi_a(x), \quad (8.9.6)$$

and thus

$$\begin{aligned} D_\mu^a(\xi') &= R_{ab}(h)D_\mu^b(\xi), \\ E_\mu^i(\xi') &= R_{ij}(h)E_\mu^j(\xi) - R_{ia}(h)\partial_\mu \xi_a(x). \end{aligned} \quad (8.9.7)$$

Under global G transformations, the factors D_μ^a transform linearly, while the factors E_μ^i transform non-linearly like a gauge field. We can in fact use E_μ^i to build covariant derivatives for $\tilde{\phi}$. From eqs. (8.6.7) and (8.6.8), we get that under G transformations

$$\partial_\mu \tilde{\phi}' = (\partial_\mu h)\tilde{\phi} + h(\partial_\mu \tilde{\phi}) = h\left(h^{-1}(\partial_\mu h)\tilde{\phi} + \partial_\mu \tilde{\phi}\right). \quad (8.9.8)$$

Defining

$$D_\mu \tilde{\phi} \equiv (\partial_\mu \tilde{\phi} + it_i E_\mu^i)\tilde{\phi}, \quad (8.9.9)$$

we have

$$D_\mu \tilde{\phi}' = h D_\mu \tilde{\phi}. \quad (8.9.10)$$

The symmetries constrain the NG bosons to appear in the low-energy Lagrangian only through the combinations D_μ^a and E_μ^i in derivative interactions. We conclude that any Lagrangian invariant under H transformations and constructed with D_μ^a , $\tilde{\phi}$ and their covariant derivatives, will automatically be invariant under the whole group G . Invariant terms can also be constructed using the “field strength”

$$F_{\mu\nu}(E) = \partial_\mu E_\nu - \partial_\nu E_\mu + [E_\mu, E_\nu] \quad (8.9.11)$$

and derivatives thereof, keeping in mind that not all invariants constructed in this way are independent of each other.

The leading order term encoding the NG boson kinetic terms is

$$\mathcal{L} \supset \frac{1}{2} F_{ab}^2 D_\mu^a D_\mu^b, \quad (8.9.12)$$

where $D_\mu^a = \partial_\mu \xi^a + \dots$. As we have seen in section 8.6, the linearly realized symmetry H constrains the coefficients F_{ab}^2 .

The above formalism can easily be extended when a subgroup H_g is gauged. The physical considerations are clearly the same as discussed in section 8.6. Let us pretend that the whole group G is gauged and only at the end recover the original theory by setting to zero the gauge fields not in H_g . One can check that all our previous results,

until eq. (8.9.1) included, are formally still valid, provided the obvious understanding that now the G transformations are x -dependent. The G invariant field combination to consider now, analogue of $\gamma^{-1}\partial_\mu\gamma$, is¹³

$$\gamma^{-1}D_\mu\gamma \equiv \gamma^{-1}(\partial_\mu - it^\alpha A_\mu^\alpha)\gamma, \quad (8.9.13)$$

so that the inhomogeneous term in the gauge field transformation cancels the analogue term coming from $\partial_\mu g$. The quantity $\gamma^{-1}D_\mu\gamma$ is also defined in the algebra $\mathcal{G}$, so we have

$$\gamma^{-1}D_\mu\gamma \equiv it_a \hat{D}_\mu^a(x) + it_i \hat{E}_\mu^i(x), \quad (8.9.14)$$

where $\hat{D}_\mu^a$ and $\hat{E}_\mu^i$ are the gauged versions of the fields D_μ^a and E_μ^i defined in eq. (8.9.2), which now include the gauge fields A_μ^α . Following the same steps as above, we immediately get

$$(\gamma^{-1}D_\mu\gamma)' = h(x)(\gamma^{-1}D_\mu\gamma)h^{-1}(x) - \partial_\mu h(x)h^{-1}(x). \quad (8.9.15)$$

Correspondingly, the relations (8.9.5) are valid for $\hat{D}_\mu^a$ and $\hat{E}_\mu^i$ as well. As before, any Lagrangian invariant under H transformations and constructed with $\hat{D}_\mu^a$, $\tilde{\phi}$ and their covariant derivatives, will automatically be invariant under the whole gauge group G . In order to also construct H invariant operators involving the field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$, that transforms linearly under G as $F_{\mu\nu} \rightarrow gF_{\mu\nu}g^{-1}$, we can define NG boson-dependent field strengths

$$f_{\mu\nu} = \gamma^{-1}F_{\mu\nu}\gamma. \quad (8.9.16)$$

The latter transform as $f_{\mu\nu} \rightarrow h\gamma^{-1}g^{-1}gF_{\mu\nu}g^{-1}g\gamma h^{-1} = hf_{\mu\nu}h^{-1}$ under G and can be used together with $\hat{D}_\mu^a$, $\tilde{\phi}$ and their associated operators constructed acting with covariant derivatives. We can now set to zero the “spurionic” gauge field components that do not belong to H_g . The leading order term (8.9.12) in the low-energy effective Lagrangian becomes

$$\mathcal{L} \supset \frac{1}{2}F_{ab}^2 \hat{D}_\mu^a \hat{D}_\mu^b, \quad (8.9.17)$$

where $\hat{D}_\mu^a = \partial_\mu \xi^a - g_{\hat{\alpha}a} A_\mu^{\hat{\alpha}}$, with $g_{\hat{\alpha}a}$ as defined in eq.(8.6.22). As expected, the NG bosons along the gauged directions are eaten by the gauge fields and their mass is given by eq.(8.6.24).

¹³In order not to confuse the group element g with the gauge coupling constant g , we adopt here a non-canonical normalization for the gauge fields A_μ^α .

Chapter 9

Anomalies*

There are different ways of regularizing a QFT. The best choice of regulator is the one which keeps the maximum number of symmetries of the classical action unbroken. Cut-off regularization, for instance, breaks gauge invariance and this is one of the reasons we prefer to work in the somewhat more exotic DR, where instead gauge invariance is unbroken. It is also possible that there exists *no* regulator that preserves a given classical symmetry. In this case we say that the symmetry is anomalous, namely the quantum theory necessarily breaks it, independently of the choice of regulator.

Roughly speaking, anomalies can affect global or local symmetries. The latter case is particularly important, because local symmetries are needed to decouple unphysical fields. Indeed, anomalies in linearly realized local gauge symmetries lead to inconsistent theories. Theories with anomalous global symmetries are instead consistent, yet the effect of the anomaly can have important effects on the theories. We have already seen an example of anomaly. In classically scale invariant theories, a scale dependence is typically generated by quantum effects by means of a non-vanishing β -function. We will however mostly focus on global anomalies associated to chiral currents and their related anomalies in local symmetries. Historically the first anomaly of this kind has been discovered by Adler, Bell, and Jackiw [3, 9], and it is associated to the non-conservation of the axial current in QCD. Among other things, the axial anomaly resolved a puzzle related to the $\pi^0 \rightarrow 2\gamma$ decay rate, predicted by effective Lagrangian considerations to be about three orders of magnitude smaller than the observed one.

The structure of this chapter is as follows. In section 9.1 we start by studying the basic anomaly associated to a global $U(1)$ chiral transformation using Feynman diagrams and its generalization to non-abelian groups. Anomalies in local symmetries are discussed in section 9.2 and their cancellation in the Standard Model worked out in section 9.3. An

alternative derivation of chiral anomalies by means of a path integral is discussed in section 9.4. Anomaly terms are constrained by the so called Wess-Zumino consistency conditions, discussed in section 9.5. Anomalies are invariant under RG flow and should match between a UV and an IR theory. This is the essence of the 't Hooft anomaly matching conditions, introduced in section 9.6. Section 9.7 is dedicated to the study of the anomalous breaking of scale invariance and its relation with beta-functions. In section 9.8 we finally discuss the strong CP problem of QCD and its possible solutions in terms of axions.

9.1 The $U(1)_A$ Chiral Anomaly from One-Loop Graphs*

The Lagrangian for a free Dirac fermion ψ is

$$\mathcal{L} = i\bar{\psi}\not{\partial}\psi - m\bar{\psi}\psi, \quad (9.1.1)$$

and is invariant for vectorial $U(1)_V$ transformations under which $\psi \rightarrow e^{i\alpha}\psi$, $\bar{\psi} \rightarrow \bar{\psi}e^{-i\alpha}$. The associated conserved $U(1)_V$ current is

$$J^\mu = \bar{\psi}\gamma^\mu\psi. \quad (9.1.2)$$

If $m = 0$, as we assume from now on, the Lagrangian (9.1.1) is also invariant under the axial $U(1)_A$ symmetry under which

$$\psi \rightarrow e^{i\alpha\gamma_5}\psi, \quad \bar{\psi} \rightarrow \bar{\psi}e^{i\alpha\gamma_5}. \quad (9.1.3)$$

The associated conserved $U(1)_A$ current is

$$J_5^\mu = \bar{\psi}\gamma^\mu\gamma_5\psi. \quad (9.1.4)$$

While the classical $U(1)_V$ symmetry is quantum mechanically preserved and leads to Ward identities as explained in section 4.4.2, it turns out that there is no way to keep the $U(1)_A$ chiral symmetry at the quantum level. This is best seen from a functional integral point of view. An invariant theory not only requires an invariant action, but also an invariant *measure*. And the functional measure is not invariant under the chiral transformation (9.1.3) (see section 9.4 below for a detailed analysis). The $U(1)_A$ current is quantum mechanically not conserved,

$$\langle\partial_\mu J_5^\mu\rangle = \mathcal{A}(x) \neq 0, \quad (9.1.5)$$

and we say that there is a $U(1)_A$ chiral anomaly given by $\mathcal{A}(x)$. We can probe the anomaly by adding to the free fermion theory external spin one sources K^μ and K_5^μ that couple to J^μ and J_5^μ , respectively:

$$Z[K^\mu, K_5^\mu] = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{i\int d^4x [\mathcal{L} + K_\mu J^\mu + K_{\mu 5} J_5^\mu]}. \quad (9.1.6)$$

Anomalies were originally discovered by evaluating three-point functions between external sources at one-loop level. They arise from loops of the internal fermion lines from current correlations that involve the chiral matrix γ_5 . Despite these one-loop graphs are divergent and would require a choice of regularization, their divergence (i.e. the anomaly) is finite. Since we are only interested to the anomaly, we can proceed with the computation without the need of introducing a regulator.

We are then led to compute the three point functions between two vector currents J^μ and one axial current J_5^μ :

$$\Gamma_{\mu\nu\rho}(x, y, z) = \frac{i\delta^3 Z}{\delta K_5^\mu(x)\delta K^\nu(y)\delta K^\rho(z)} \Big|_{K=K_5=0} = \langle J_{\mu,5}(x)J_\nu(y)J_\rho(z) \rangle, \quad (9.1.7)$$

which in momentum space reads as

$$\Gamma_{\mu\nu\rho}(k_1, k_2) \equiv \langle J_{\mu,5}(-k_1 - k_2)J_\nu(k_1)J_\rho(k_2) \rangle \quad (9.1.8)$$

where, with some abuse of notation, we denote the Fourier transform with the same symbol of the corresponding term. The would-be conservation of the two currents, $\partial_\mu J^\mu = \partial_\mu J_5^\mu = 0$, would imply at the quantum level the relations

$$(k_1 + k_2)^\mu \Gamma_{\mu\nu\rho} = k_1^\nu \Gamma_{\mu\nu\rho} = k_2^\rho \Gamma_{\mu\nu\rho} = 0. \quad (9.1.9)$$

No contact terms arise in eqs.(9.1.9), because the axial and vector currents are classically invariant under the $U(1)_V \times U(1)_A$ transformations. As we will see, due to the anomaly, it turns out to be impossible to impose all three conditions (9.1.9) simultaneously.

The two diagrams in fig. 9.1 contribute at one-loop level to $\Gamma_{\mu\nu\rho}$. They give

$$\Gamma_{\mu\nu\rho}(k_1, k_2) = - \int \frac{d^4 p}{(2\pi)^4} \text{tr} \left(S(p+a)\gamma_\mu\gamma_5 S(p+a+k)\gamma_\rho S(p+a+k_1)\gamma_\nu + \right. \\ \left. S(p+b)\gamma_\mu\gamma_5 S(p+b+k)\gamma_\nu S(p+b+k_2)\gamma_\rho \right), \quad (9.1.10)$$

where $k = k_1 + k_2$ and $S(p) = \not{p}/p^2$ is the free fermion propagator for a massless fermion. In eq. (9.1.10) a and b are arbitrary constant vectors, whose relevance will be clear in the following. Let us first compute the divergence of the axial current, $k^\mu \Gamma_{\mu\nu\rho}$. It is convenient to write

$$S(p+a)\not{k}S(p+a+k) = S(p+a)(\not{p}+\not{a}+\not{k}-\not{p}-\not{a})S(p+a+k) = S(p+a)-S(p+a+k), \quad (9.1.11)$$

so that we can rewrite

$$k^\mu \Gamma_{\mu\nu\rho}(k_1, k_2) = \int \frac{d^4 p}{(2\pi)^4} \left(f_{\rho\nu}(p, a+k, a+k_1) - f_{\rho\nu}(p, a, a+k_1) + (\rho \leftrightarrow \nu, a \leftrightarrow b, k_1 \leftrightarrow k_2) \right), \quad (9.1.12)$$

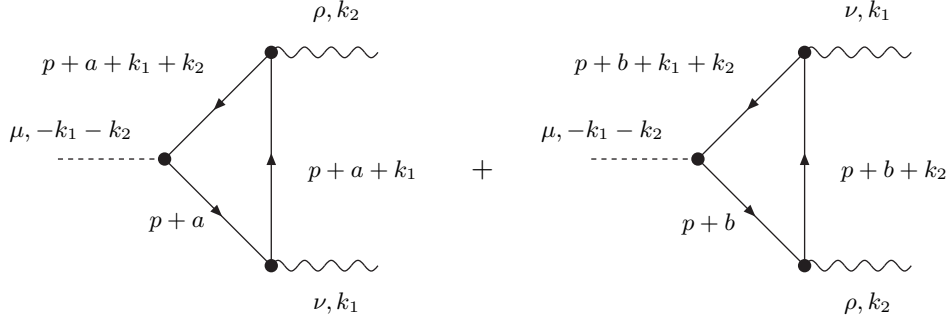


Figure 9.1: One-loop graphs contributing to the anomaly. All external momenta are incoming. The wavy and dashed lines represent the external vector and axial sources, respectively.

where we have defined

$$f_{\rho\nu}(p, a, b) \equiv \text{tr} \left(S(p+a) \gamma_\rho S(p+b) \gamma_\nu \gamma_5 \right). \quad (9.1.13)$$

The explicit form of $f_{\rho\nu}$ is readily found by recalling the definition of the γ_5 matrix: $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, and the property

$$\text{Tr} \left(\gamma_\gamma \gamma_\rho \gamma_\omega \gamma_\nu \gamma_5 \right) = 4i\epsilon_{\gamma\rho\omega\nu}, \quad (9.1.14)$$

with $\epsilon_{0123} = +1$. We then have

$$f_{\rho\nu}(p, a, b) = \frac{(p+a)^\gamma (p+b)^\omega}{(p+a)^2 (p+b)^2} 4i\epsilon_{\gamma\rho\omega\nu} = 4i\epsilon_{\gamma\rho\omega\nu} \frac{p_\omega (a-b)_\gamma + a_\gamma b_\omega}{(p+a)^2 (p+b)^2}. \quad (9.1.15)$$

Using the symmetry ($\rho \leftrightarrow \nu, a \leftrightarrow b, k_1 \leftrightarrow k_2$), we can replace $f_{\rho\nu}(p, a+k, a+k_1)$ in eq.(9.1.12) with $f_{\rho\nu}(p, b+k_2, b+k)$ so that

$$k^\mu \Gamma_{\mu\nu\rho}(k_1, k_2) = \int \frac{d^4 p}{(2\pi)^4} \left(f_{\rho\nu}(p, b+k_2, b+k) - f_{\rho\nu}(p, a, a+k_1) + (\rho \leftrightarrow \nu, a \leftrightarrow b, k_1 \leftrightarrow k_2) \right). \quad (9.1.16)$$

Note that there is a redundancy in the definition of $f_{\rho\nu}$. For any four-vector c we have

$$f_{\rho\nu}(p+c, a, b) = f_{\rho\nu}(p, a+c, b+c). \quad (9.1.17)$$

We can then use eq.(9.1.17) with $c = b-a+k_2$ in the second term in eq. (9.1.16) to write

$$\begin{aligned} k^\mu \Gamma_{\mu\nu\rho}(k_1, k_2) &= \int \frac{d^4 p}{(2\pi)^4} \left(f_{\rho\nu}(p, b+k_2, b+k) - f_{\rho\nu}(p+c, b+k_2, b+k) \right. \\ &\quad \left. + (\rho \leftrightarrow \nu, a \leftrightarrow b, k_1 \leftrightarrow k_2) \right). \end{aligned} \quad (9.1.18)$$

Shifting the virtual momentum p in the second term in eq. (9.1.18), $p + c \rightarrow p$, would naively give $k^\mu \Gamma_{\mu\nu\rho} = 0$. However, we should recall that these expressions are divergent and momentum shift is *not* allowed. Indeed, for a generic function $f(x)$, we have

$$\begin{aligned} \int_{-\infty}^{\infty} dx \left(f(x+c) - f(x) \right) &= \int_{-\infty}^{\infty} dx \left(c \frac{df(x)}{dx} + \frac{c^2}{2} \frac{d^2 f(x)}{dx^2} + \dots \right) = \\ &c \left(f(\infty) - f(-\infty) \right) + \frac{c^2}{2} \left(f'(\infty) - f'(-\infty) \right) + \dots \end{aligned} \quad (9.1.19)$$

Equation (9.1.19) vanishes if $f(\pm\infty) = f'(\pm\infty) = \dots = 0$, conditions automatically satisfied for convergent integrals, but not for divergent ones. So we have to proceed more carefully. The basic expression we want to evaluate is

$$I_{\mu\nu}(c, d, e) \equiv \int \frac{d^4 p}{(2\pi)^4} \left(f_{\mu\nu}(p+c, d, e) - f_{\mu\nu}(p, d, e) \right). \quad (9.1.20)$$

It is useful to rotate the virtual momentum p to euclidean values p_E and expand in Taylor series the integrand:

$$I_{\mu\nu}(c, d, e) = i \int \frac{d^4 p_E}{(2\pi)^4} \sum_{n=1}^{\infty} \frac{1}{n!} c^{\rho_1} \dots c^{\rho_n} \frac{\partial^n f_{\mu\nu}(p, d, e)}{\partial p_E^{\rho_1} \dots \partial p_E^{\rho_n}}. \quad (9.1.21)$$

Like in the one-dimensional case in eq.(9.1.19), the integrals are total derivatives and can be recast in boundary integrals using the divergence theorem (particular case of the more general Stokes theorem) for which

$$\int_V \partial_\rho f = \int_S n_\rho f, \quad (9.1.22)$$

where V is a volume, S its boundary and n_μ is a unit vector orthogonal to the boundary. In the euclidean integral (9.1.21) the measure of integration (V) is a large four-dimensional ball and its boundary (S) is a 3-sphere S^3 , and $n^\rho = p_E^\rho / |p_E|$. Since $f_{\mu\nu} \propto 1/p_E^3$ for $p_E \rightarrow \infty$, we see that only the $n = 1$ term in the Taylor expansion (9.1.21) can give a non-vanishing contribution. We then have

$$I_{\mu\nu}(c, d, e) = \frac{i}{(2\pi)^4} \lim_{p_E \rightarrow \infty} c^\rho \int_{S^3} p_{E,\rho} p_E^2 f_{\mu\nu}(p_E, d, e) d\Omega_3. \quad (9.1.23)$$

By $SO(4)$ invariance

$$\int p_E^\mu p_E^\nu f(p_E^2) d\Omega_3 = \frac{\delta^{\mu\nu}}{4} \Omega_3 p_E^2 f(p_E^2), \quad (9.1.24)$$

where $\Omega_3 = 2\pi^2$ is the volume of the unit three-sphere. Plugging eq.(9.1.15) in eq.(9.1.23) and using eq.(9.1.24) we get

$$I_{\mu\nu}(c, d, e) = \frac{1}{8\pi^2} c^\gamma (d-e)^\omega \epsilon_{\gamma\mu\omega\nu}. \quad (9.1.25)$$

We can now go back to eq.(9.1.18) and compute it using eq.(9.1.25). The derivative of the axial current is finally found to be

$$\begin{aligned} k^\mu \Gamma_{\mu\nu\rho}(k_1, k_2) &= \frac{1}{8\pi^2} \epsilon_{\gamma\rho\omega\nu} \left[(k_2 + b - a)^\omega (-k_1)^\gamma + (k_1 + a - b)^\omega k_2^\gamma \right] \\ &= \frac{1}{8\pi^2} \epsilon_{\gamma\rho\omega\nu} (k_1 - k_2 + \Delta)^\omega (k_1 + k_2)^\gamma, \end{aligned} \quad (9.1.26)$$

with $\Delta \equiv a - b$. It is now clear why we have introduced the otherwise redundant parameters a and b . In a convergent expression, they would trivially be reabsorbed in a shift of the virtual momentum p , but in the case at hand the final expression turns out to depend on their difference Δ . Being Δ arbitrary, one could choose $\Delta = k_2 - k_1$, so that the “anomalous” term (9.1.26) would cancel. However, care has to be paid on the divergence of the two vector currents J_ν and J_ρ . Proceeding exactly as before, we can compute

$$\begin{aligned} k_1^\nu \Gamma_{\mu\nu\rho}(k_1, k_2) &= \frac{1}{8\pi^2} \epsilon_{\gamma\mu\omega\rho} \left[(k_1 + k_2)^\gamma (b - a)^\omega - (b - a - k_1)^\omega k_2^\gamma \right] \\ &= \frac{1}{8\pi^2} \epsilon_{\gamma\mu\omega\rho} k_1^\omega (\Delta + k_2)^\gamma, \end{aligned} \quad (9.1.27)$$

$$\begin{aligned} k_2^\rho \Gamma_{\mu\nu\rho}(k_1, k_2) &= \frac{1}{8\pi^2} \epsilon_{\gamma\mu\omega\nu} \left[(k_1 + k_2)^\gamma (a - b)^\omega - (a - b - k_2)^\omega k_1^\gamma \right] \\ &= \frac{1}{8\pi^2} \epsilon_{\gamma\mu\omega\nu} k_2^\gamma (\Delta - k_1)^\omega. \end{aligned} \quad (9.1.28)$$

The choice $\Delta = k_2 - k_1$ would then lead to the non-conservation of the two vector currents J_ν and J_ρ . Since three-point functions between 3 vector currents do not lead to any anomaly, we insist in having vanishing divergence for the vector currents. This uniquely fixes $\Delta = k_1 - k_2$, opposite to the choice leading to the conservation of the axial current. Plugging $\Delta = k_1 - k_2$ in eq. (9.1.26) gives

$$k^\mu \Gamma_{\mu\nu\rho}(k_1, k_2) = \frac{1}{2\pi^2} \epsilon_{\gamma\rho\omega\nu} k_1^\omega k_2^\gamma. \quad (9.1.29)$$

This is the final form of the chiral anomaly in momentum space. The arbitrariness to shift the anomaly from one current to another by changing Δ corresponds to the possibility to add in the action a local term proportional to $\epsilon^{\mu\nu\rho\sigma} K_{5,\mu} \partial_\sigma K_\nu K_\sigma$. This term gives rise to a contact term interaction to $\Gamma_{\mu\nu\rho}(x, y, z)$, see eq.(9.1.7), that changes the divergence of the currents. Crucially, there is *no* choice of such counterterm (and correspondingly of Δ , as we have seen) for which all three currents are conserved.

When K_μ are external sources, the anomaly term in eq.(9.1.29) does not lead to any non-conservation of currents in a quantum theory, because external sources are supposed to be set to zero eventually. These are called ’t Hooft anomalies (more on this in section 9.6). However, we can equally couple the free fermion theory to dynamical gauge fields, gauging

the $U(1)_V$ symmetry, in which case the vector currents K_μ are nothing else than the $U(1)_V$ gauge fields A_μ (modulo a coupling constant factor, depending on the normalization chosen). The gauging of $U(1)_V$ does not break the $U(1)_A$ symmetry, which still remains at the classical level. In this case the wavy lines depicted in fig.9.1 are actual photons and it is not difficult to reconstruct the local anomaly functional $\mathcal{A}(x)$ entering in eq.(9.1.5). This should be such that, when taking two functional derivatives with respect to the gauge fields A_μ gives, in momentum space, the term in the right hand side of eq.(9.1.29). For canonically normalized gauge fields we get, in configuration space,

$$\langle \partial_\mu J_5^\mu \rangle = -\frac{g^2}{16\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}. \quad (9.1.30)$$

This is the standard form of the chiral anomaly. We see that when gauge fields are coupled to the fermion field the chiral current is not conserved.

It is not difficult to generalize the above analysis to non-abelian global symmetries. The Lagrangian for N massless free Dirac fields ψ_i ($i = 1, \dots, N$) has an $SU(N)_V \times SU(N)_A \times U(1)_A \times U(1)_V$ global symmetry. The vector and axial currents read

$$J_\mu^\alpha = \bar{\psi} \gamma_\mu T^\alpha \psi, \quad J_{\mu 5}^\alpha = \bar{\psi} \gamma_\mu \gamma_5 T^\alpha \psi, \quad (9.1.31)$$

where T^α the $U(N)$ generators, including the $U(1)$ factor. We then evaluate the 3-point function at one-loop level between two vector and 1 axial currents:

$$\Gamma_{\mu\nu\rho}^{\alpha\beta\gamma}(k_1, k_2) \equiv \langle J_{\mu 5}^\alpha(-k_1 - k_2) J_\nu^\beta(k_1) J_\rho^\gamma(k_2) \rangle. \quad (9.1.32)$$

Similarly to eq. (9.1.10), we have

$$\begin{aligned} \Gamma_{\mu\nu\rho}^{\alpha\beta\gamma}(k_1, k_2) = & -\int \frac{d^4 p}{(2\pi)^4} \left[\text{tr}(T^\alpha T^\gamma T^\beta) \text{tr} \left(S(p+a) \gamma_\mu \gamma_5 S(p+a+k) \gamma_\rho S(p+a+k_1) \gamma_\nu \right) \right. \\ & \left. + \text{tr}(T^\alpha T^\beta T^\gamma) \text{tr} \left(S(p+b) \gamma_\mu \gamma_5 S(p+b+k) \gamma_\nu S(p+b+k_2) \gamma_\rho \right) \right]. \end{aligned} \quad (9.1.33)$$

Since the currents transform under a symmetry transformation, their classical conservation turns in non-trivial WT identities, see eq.(4.4.10):

$$i\partial_{x_1}^\mu \langle J_{\mu 5}^\alpha(x_1) J_\nu^\beta(x_2) J_\rho^\gamma(x_3) \rangle = \delta(x_1 - x_2) \langle \delta_\alpha J_\nu^\beta(x_2) J_\rho^\gamma(x_3) \rangle + \delta(x_1 - x_3) \langle J_\nu^\beta(x_2) \delta_\alpha J_\rho^\gamma(x_3) \rangle, \quad (9.1.34)$$

where, under an infinitesimal chiral transformation, $\delta_\alpha J_\mu^\beta = C_{\beta\alpha\gamma} J_{\mu 5}^\gamma$. By parity invariance the two-point function between a vector and an axial current vanishes, the contact terms in eq.(9.1.34) then vanish and in momentum space we simply have¹

$$(k_1 + k_2)^\mu \Gamma_{\mu\nu\rho}^{\alpha\beta\gamma} = 0. \quad (9.1.35)$$

¹Contact terms will not vanish when taking the divergence of currents in other correlators, such as three vector currents. These terms are always proportional to the structure constants $C_{\beta\alpha\gamma}$ and can be shown to not give rise to anomalies.

It is convenient to write

$$\text{tr}(T^\alpha T^\beta T^\gamma) = \frac{1}{2}\text{tr}(\{T^\alpha, T^\beta\}T^\gamma) + \frac{1}{2}\text{tr}([T^\alpha, T^\beta]T^\gamma) = D^{\alpha\beta\gamma} + \frac{i}{2}C^{\alpha\beta\gamma}, \quad (9.1.36)$$

where in eq.(9.1.36) we have introduced the factor

$$D^{\alpha\beta\gamma} \equiv \frac{1}{2}\text{tr}(\{T^\alpha, T^\beta\}T^\gamma), \quad (9.1.37)$$

symmetric under permutation of its indices. It can be shown that the terms proportional to $C^{\alpha\beta\gamma}$ do not give rise to anomalies, which are found in the terms proportional to $D^{\alpha\beta\gamma}$. These can be manipulated exactly as before and eventually lead to the same form of the anomaly:

$$\Gamma_{\mu\nu\rho}^{\alpha\beta\gamma}(k_1, k_2)|_{ano.} = D^{\alpha\beta\gamma}\Gamma_{\mu\nu\rho}(k_1, k_2)|_{ano.}. \quad (9.1.38)$$

As before, a non-conservation of currents in the quantum theory arises only when we gauge some of the global symmetries. With non-abelian symmetries we have to pay attention in the gauging procedure, because it can lead to an explicit violation of some of the global symmetries. For example, gauging the whole $U(N)_V$ factor would break $SU(N)_A$, because the gauge coupling interaction is not $SU(N)_A$ invariant. The $U(1)_A$ symmetry is however classical preserved, so we might consider the case in which the axial current is given by $U(1)_A$, while the vector currents are gauged and correspond to gauge fields. In this case T^α can be taken equal to unity and we get

$$\langle\partial_\mu J_5^\mu\rangle = -\frac{g^2}{4\pi^2}\epsilon^{\mu\nu\rho\sigma}\partial_\mu A_\nu^\beta\partial_\rho A_\sigma^\gamma\text{tr}(T^\beta T^\gamma). \quad (9.1.39)$$

where $D_{\alpha\beta\gamma} \rightarrow \text{tr}(T^\beta T^\gamma)$. In the non-abelian case, additional one-loop (square and pentagon) diagrams contribute to the anomaly and should be considered. When they are summed to eq. (9.1.39), the whole non-abelian form of the field strength is reconstructed and we get the final result

$$\langle\partial_\mu J_5^\mu\rangle = -\frac{g^2}{16\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}^\beta F_{\rho\sigma}^\gamma\text{tr}(T^\beta T^\gamma). \quad (9.1.40)$$

Eq.(9.1.40) is the generalization of eq.(9.1.30) for non-abelian gauge fields and, as expected, boils down to eq.(9.1.30) in the abelian case with $T^\beta = T^\gamma = 1$.

In any given regularization, where all amplitudes are made finite and the shift in the virtual momentum allowed, the anomaly would appear differently. For instance, in DR, the anomaly arises from subtleties related to the definition of γ_5 in $d \neq 4$ dimensions (see, e.g., ref. [49] for a computation of the anomaly in DR). In Pauli-Villars regularization, where a heavy fermion is added, the anomaly is related to the explicit breaking of the axial symmetry given by the heavy fermion mass term. There is no regulator that preserves at the same time the vector and axial symmetries and as a result an anomaly appears.

9.2 Gauge Anomalies*

As we have already mentioned, anomalies can also affect local symmetries, in which case we refer to them as “gauge anomalies”. The latter can arise in chiral, parity non-invariant, gauge theories, where left and right-moving fermions transform in representations of the gauge group that are not complex conjugate with each other. The simplest gauge anomaly can be described by starting from the Lagrangian for a free chiral fermion ψ_L :

$$\mathcal{L} = i\bar{\psi}_L \not{\partial} \psi_L, \quad (9.2.1)$$

invariant under $U(1)_L$ transformations $\psi_L \rightarrow e^{i\alpha}\psi_L$, $\bar{\psi}_L \rightarrow \bar{\psi}_L e^{-i\alpha}$, with $U(1)_L$ current $J^\mu = \bar{\psi}_L \gamma^\mu \psi_L = \bar{\psi} \gamma^\mu P_L \psi$, with $P_L = (1 + \gamma_5)/2$ the chirality projector. The 3-point function at one-loop level between three currents is

$$\begin{aligned} \Gamma_{\mu\nu\rho}(k_1, k_2) = - \int \frac{d^4 p}{(2\pi)^4} \left[\text{tr} \left(\gamma_\mu P_L S(p+a+k) \gamma_\rho P_L S(p+a+k_1) \gamma_\nu P_L S(p+a) \right. \right. \\ \left. \left. + \gamma_\mu P_L S(p+b+k) \gamma_\nu P_L S(p+b+k_2) \gamma_\rho P_L S(p+b) \right) \right]. \end{aligned} \quad (9.2.2)$$

where $\Gamma_{\mu\nu\rho}(k_1, k_2) \equiv \langle J_\mu(-k_1 - k_2) J_\nu(k_1) J_\rho(k_2) \rangle$. Since $P_L^2 = P_L$, eq.(9.2.2) turns into a sum of a non-anomalous 3-point correlators with no γ_5 factors and an anomalous one with one insertion of γ_5 . The last term, modulo a factor 1/2 coming from P_L , equals eq.(9.1.10). We then get

$$k^\mu \Gamma_{\mu\nu\rho}(k_1, k_2) = \frac{1}{2} \frac{1}{8\pi^2} \epsilon_{\gamma\rho\omega\nu} (k_1 - k_2 + \Delta)^\omega (k_1 + k_2)^\gamma, \quad (9.2.3)$$

and similarly 1/2 of the divergences (9.1.27) and (9.1.28) for the $\mathcal{J}^\nu$ and $\mathcal{J}^\rho$ currents. Gauging the chiral theory (9.2.1) amounts to couple the above three currents to a $U(1)$ gauge field. Since the theory is chiral, there are no vector and axial currents, but only left currents to couple to the gauge field. As a result, there is no choice of the factor Δ that allows us to conserve all the three vector currents at the same time. Because of permutation symmetry of the three identical external currents, we should choose Δ so that the anomaly is symmetric under the exchange of any of the three currents, as required by Bose symmetry. This is achieved by taking

$$\Delta = -\frac{1}{3}(k_1 - k_2), \quad (9.2.4)$$

which gives an anomaly one-third smaller than the anomaly (9.1.30):

$$\langle \partial_\mu J^\mu \rangle = -\frac{1}{2} \frac{g^2}{48\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}. \quad (9.2.5)$$

This is the simplest form of a gauge anomaly. Such an anomaly is deadly for a theory. The WI derived in section 4.4.3 no longer hold. In particular, the violation of eq.(4.4.41) does not allow the decoupling of unphysical states. In scattering amplitudes the factor $\Gamma_{\mu\nu\rho}(k_1, k_2)$ is multiplied either by polarizations of the external photons or by internal photon propagators when the triangular graph is embedded within a more complicated Feynman graph. For external photons the computation of the divergence of one of the gauge currents is equivalent to replace the photon polarization vector by its momentum, i.e. to consider longitudinal photons. A non-vanishing result implies that these states do not decouple. If the line is internal, the longitudinal photon runs into the loop, yet we have a problem, because unitarity (the optical theorem) would imply their non-decoupling and a violation of unitarity (at any scale). In other words, minimally coupling a $U(1)$ gauge field to a chiral fermion is quantum mechanically inconsistent.

Similar considerations can be generalized to multiple fermions and to non-abelian gauge theories. We will not enter into the details of non-abelian gauge anomalies. It is sufficient here to say that the key coefficient governing the anomaly is the factor $D^{\alpha\beta\gamma}$ defined in eq.(9.1.37). When we have several fermions transforming in different representations r of a gauge group G , the trace in $D^{\alpha\beta\gamma}$ should be taken over all fermions, each in the corresponding representation r . Crucially, left-handed and right-handed fermions contribute with an opposite sign to $D^{\alpha\beta\gamma}$, due to the chiral nature of the gauge anomaly. A gauge theory is non-anomalous if and only if in total $D^{\alpha\beta\gamma} = 0$. The explicit form of the non-abelian anomaly can be significantly constrained by demanding consistency with the structure of the gauge group (see section 9.5 for some other detail) and cannot be written in a covariant way in terms of field strengths only.

So far, we have only considered massless fermions. Fermion mass terms explicitly break the axial symmetry (9.1.3), but it can easily be shown that they do not change the form of the axial anomaly (9.1.39). In fact, left-handed and right-handed fermions contribute the same to the axial anomaly, because the sign change due to the different chirality of the fermions compensate their opposite charge with respect to the axial symmetry. This explains why we got a non-vanishing axial anomaly starting from vector-like Dirac fermions. The situation is different for gauge anomalies, that can receive a non-vanishing contribution only from massless fermions. This is proved as follows. It is convenient to consider all fermions as left-handed by defining, for each right-handed component ψ_R its left-handed counterpart

$$\psi_L^c = C\psi_R^*, \quad (9.2.6)$$

C being the matrix of charge conjugation in spinor space. If ψ_R transforms under a representation r_R of a gauge group G , ψ_L^c will transform as r_R^* . In the basis of left-handed

two-components spinors χ_i , a generic mass term reads

$$\mathcal{L}_m = \chi_i^t \sigma^2 \hat{M}_{ij} \chi_j + h.c. \quad (9.2.7)$$

where i run over all the fermions in the theory and $\hat{M}$ is a symmetric mass matrix. Let us focus on a subset of terms in eq. (9.2.7) coupling two left-handed fermion multiplets χ_1 and χ_2 in irreducible representations r_1 and r_2 of the group G , with $\dim r_1 = \dim r_2$:

$$\mathcal{L}_m \supset \chi_1^t \sigma^2 M \chi_2 + h.c. \quad (9.2.8)$$

where M is a non-singular mass matrix. The term (9.2.8) is gauge-invariant if

$$-T_1^t M = M T_2, \quad (9.2.9)$$

where T_1 and T_2 are the generators in the representations r_1 and r_2 . Eq. (9.2.9) implies that $-T_1^t$ and T_2 are related by a similarity transformation, since $-T_1^t = M T_2 M^{-1}$. Then

$$D_{\alpha\beta\gamma}^1 = \frac{1}{2} \text{tr}\{T_{1\alpha}, T_{1\beta}\} T_{1\gamma} = (-1)^3 \frac{1}{2} \text{tr}\{T_{2\alpha}^t, T_{2\beta}^t\} T_{2\gamma}^t = -D_{\alpha\beta\gamma}^2. \quad (9.2.10)$$

The contribution to the gauge anomaly given by χ_1 is then exactly cancelled by that of χ_2 . Let us clarify this result with a couple of simple examples. Consider a $U(1)$ gauge theory with a Dirac fermion with charge q . In terms of left-handed fields, the Dirac fermion consists of one left-handed fermion χ_1 with charge q and its conjugate χ_2 with charge $-q$ and admits the mass term $m\chi_1\sigma^2\chi_2 + h.c.$. The total gauge anomaly is proportional to

$$+q^3 + (-q)^3 = 0. \quad (9.2.11)$$

Similarly, Dirac fermions in any representation T_r of a gauge group consist of one left-handed fermion multiplet in the representation T_r and its conjugate in the complex conjugate representation $T_{\bar{r}} = -T_r^t$ and thus do not lead to anomalies. In parity-invariant theories the absence of anomalies is obvious since all currents are manifestly vector-like (i.e no γ_5 appears in a 4-component Dirac notation). Non-trivial anomalies can only arise in non-parity-invariant theories, namely in so called chiral gauge theories, where a fermion and its complex conjugate transform in representations of the gauge group that are not complex conjugate of each other. The absence of anomalies in chiral gauge theories requires a non-trivial cancellation between different fermion multiplets. The most important example of a theory of this sort is the SM, where gauge anomalies cancel between quarks and leptons, as we will see in the next section.

We can now come back to the fate of the $U(1)_A$ symmetry in QCD, mentioned at the end of section 8.7. QCD with n_f massless quarks has actually an $SU(n_f)_V \times SU(n_f)_A \times$

$U(1)_V \times U(1)_A$ global symmetry, spontaneously broken to $SU(n_f)_V \times U(1)_V$ by the quark condensate. Let us compute the possible anomalies that we can have by looking at the $D_{\alpha\beta\gamma}$ coefficients associated to the various groups. Since quarks are vector-like, the gauge $SU(3)^3$ anomalies all cancel. Anomalies with one $SU(N)$ non-abelian current of any kind also manifestly vanish, because $\text{tr } T = 0$. The only possible non-vanishing anomalies are of the form $SU(3)^2 \times U(1)_V$ and $SU(3)^2 \times U(1)_A$. In terms of left-handed quarks, a quark and its conjugate have opposite charges with respect to $U(1)_V$ and equal charges with respect to $U(1)_A$, thus the $SU(3)^2 \times U(1)_V$ anomaly vanishes while $SU(3)^2 \times U(1)_A$ does not. This is nothing else than the anomaly considered before in eq.(9.1.40), with $F_{\mu\nu}^\alpha$ identified with the gluon field strengths. In QCD, then, the $U(1)_A$ symmetry, in addition of being spontaneously broken by the condensate, is also explicitly broken by the anomaly. The latter breaking is large when the theory becomes strongly coupled (due to instantons) and is responsible for the absence of a light η' in the QCD scalar meson spectrum.

Summarizing, we can have three qualitatively different scenarios:

- An anomaly involving three global currents $\rightarrow$ The theory is consistent, the global symmetry is preserved but there is an obstruction to gauging ('t Hooft anomaly).
- An anomaly involving one global and two local currents $\rightarrow$ The theory is consistent, the global symmetry is broken.
- An anomaly involving three local currents $\rightarrow$ The theory is inconsistent.

9.3 A Relevant Example: Cancellation of Gauge Anomalies in the SM*

The SM gauge group is $G_{SM} = SU(3) \times SU(2) \times U(1)$. Its fermion content, in terms of left-handed fields, is composed by three copies (generations) of the following representations:

$$(\mathbf{3}, \mathbf{2})_{\frac{1}{6}} + (\bar{\mathbf{3}}, \mathbf{1})_{-\frac{2}{3}} + (\bar{\mathbf{3}}, \mathbf{1})_{\frac{1}{3}} + (\mathbf{1}, \mathbf{2})_{-\frac{1}{2}} + (\mathbf{1}, \mathbf{1})_1 \quad (9.3.1)$$

corresponding to the quark doublet, up quark singlet, down quark singlet, lepton doublet and charged lepton singlet, respectively. In principle there can be ten possible kinds of gauge anomalies, associated to all possible combinations of $SU(3)$, $SU(2)$ and $U(1)$ currents in the triangular graph. Five of them, where a non-abelian group factor ($SU(3)$ or $SU(2)$) appears only once, $SU(3)^2 \times SU(2)$, $SU(3) \times SU(2)^2$, $SU(3) \times SU(2) \times U(1)$, $SU(3) \times U(1)^2$ and $SU(2) \times U(1)^2$ are trivially vanishing, since for $SU(n)$ groups the generators are traceless: $\text{tr } T = 0$. The $SU(2)^3$ anomaly is also manifestly vanishing because for $SU(2)$ the symmetric factor $D^{\alpha\beta\gamma}$ vanishes. In the case at hand, with doublets only, this is easily

seen: $D^{\alpha\beta\gamma} = 1/2 \text{tr}\{t^\alpha, t^\beta\}t^\gamma = \delta^{\alpha\beta} \text{tr} t^\gamma = 0$. In the general case, $D^{\alpha\beta\gamma}$ vanishes because all $SU(2)$ representations are equivalent to their complex conjugates, namely there exists a matrix A such that $T^t = T^* = -ATA^{-1}$. Using this relation, one immediately sees that $D^{\alpha\beta\gamma} = 0$ for any representation.

The remaining combinations $SU(3)^3$, $SU(3)^2 \times U(1)$, $SU(2)^2 \times U(1)$ and $U(1)^3$ have to be checked. Let us then compute the values of the symmetric coefficients $D^{\alpha\beta\gamma}$ in each of the above 4 cases and show that they always vanish. In order to distinguish the different coefficients, we denote by a subscript c , w and Y the $SU(3)$, $SU(2)$ and $U(1)$ factors, respectively. It is enough to consider a single generation of quarks and doublets, because the cancellation occurs generation per generation. Let us start with the $SU(3)^3$ anomaly. Only quarks contribute to it. We get

$$D_{ccc}^{\alpha\beta\gamma} = 2D_{\mathbf{3}}^{\alpha\beta\gamma} + D_{\mathbf{\bar{3}}}^{\alpha\beta\gamma} + D_{\mathbf{3}}^{\alpha\beta\gamma} = 2D_{\mathbf{3}}^{\alpha\beta\gamma} - D_{\mathbf{3}}^{\alpha\beta\gamma} - D_{\mathbf{3}}^{\alpha\beta\gamma} = 0, \quad (9.3.2)$$

using that $D_{\mathbf{\bar{3}}}^{\alpha\beta\gamma} = -D_{\mathbf{3}}^{\alpha\beta\gamma}$. For $SU(3)^2 \times U(1)$ we have

$$D_{ccY}^{\alpha\beta} = 2\text{tr}_{\mathbf{3}} t^\alpha t^\beta \times \frac{1}{6} + \text{tr}_{\mathbf{\bar{3}}} t^\alpha t^\beta \times \left(-\frac{2}{3}\right) + \text{tr}_{\mathbf{\bar{3}}} t^\alpha t^\beta \times \frac{1}{3} = \text{tr}_{\mathbf{3}} t^\alpha t^\beta \left(\frac{1}{3} - \frac{2}{3} + \frac{1}{3}\right) = 0, \quad (9.3.3)$$

with $\text{tr}_{\mathbf{3}} t^\alpha t^\beta = \text{tr}_{\mathbf{\bar{3}}} t^\alpha t^\beta$. For $SU(2)^2 \times U(1)$, only doublets contribute. We get

$$D_{wwY}^{\alpha\beta} = 3\text{tr}_{\mathbf{2}} t^\alpha t^\beta \times \frac{1}{6} + \text{tr}_{\mathbf{2}} t^\alpha t^\beta \times \left(-\frac{1}{2}\right) = \text{tr}_{\mathbf{2}} t^\alpha t^\beta \left(\frac{1}{2} - \frac{1}{2}\right) = 0. \quad (9.3.4)$$

For $U(1)^3$ anomalies all quarks and leptons contribute and one gets

$$D_{YYY} = 3 \times 2 \times \left(\frac{1}{6}\right)^3 + 3 \times \left(-\frac{2}{3}\right)^3 + 3 \times \left(\frac{1}{3}\right)^3 + 2 \times \left(-\frac{1}{2}\right)^3 + (1)^3 = 0. \quad (9.3.5)$$

There is actually a fifth non-trivial anomaly to check, that arises when we couple the SM to gravity. It is an anomaly involving a $U(1)$ current and two energy-momentum tensors, and it is called a mixed $U(1)$ -gravitational anomaly.² This anomaly is proportional to $\sum_n q_n$, where n runs over all fermions with charges q_n . In the SM, the mixed $U(1)$ -gravitational anomaly is proportional to

$$3 \times 2 \times \frac{1}{6} + 3 \times \left(-\frac{2}{3}\right) + 3 \times \frac{1}{3} + 2 \times \left(-\frac{1}{2}\right) + 1 = 0. \quad (9.3.6)$$

Notice how the fermion charges nicely combine to give a vanishing result for all the anomalies, in particular the $U(1)^3$ and the mixed $U(1)$ -gravitational ones.

²There also exist pure gravitational anomalies. These are vanishing in 4 space-time dimensions, but can occur in $4n + 2$ dimensions (n a non-negative integer). We will not discuss these anomalies, that are beyond our course.

The SM is then a chiral, anomaly-free gauge theory. More precisely, we have shown that the SM is anomaly-free in the unbroken phase where all the gauge group is linearly realized (i.e. no Higgs mechanism is at work). Since the anomaly does not depend on the Higgs field, our results automatically imply that the SM is anomaly-free in the broken phase as well. In particular, the unbroken gauge group $SU(3) \times U(1)_{EM}$ is manifestly anomaly-free being all currents vector-like.

9.4 Path Integral Derivation of the Chiral Anomaly**

In the path-integral formulation of field theory, anomalies arise from the transformation of the measure used to define the fermion path integral. We will consider in what follows path-integral in euclidean space-time.

Let $\psi_A(x)$ be a massless Dirac fermion defined on R^4 , in an arbitrary representation $\mathcal{R}$ of a gauge group G ($A = 1, \dots, \dim \mathcal{R}$). The minimal coupling of the fermion to the gauge fields is described by the Lagrangian

$$\mathcal{L} = \bar{\psi}(x)_A i\gamma^\mu D_\mu^A \psi^B(x). \quad (9.4.1)$$

The covariant derivative is given by

$$D_\mu^A = \partial_\mu \delta_B^A + A_\mu^\alpha T_{\alpha B}^A, \quad (9.4.2)$$

in terms of the gauge connection A_μ^α and the anti-Hermitian generators $T_{\alpha B}^A$ of the group G in the representation $\mathcal{R}$ ($\alpha = 1, \dots, \dim G$), with γ^μ satisfying the anticommutation relations $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$ ($\mu, \nu = 1, 2, 3, 4$).

The classical Lagrangian (9.4.1) is invariant under the global chiral transformation

$$\psi \rightarrow e^{i\alpha\gamma_5}\psi, \quad (9.4.3)$$

where $\gamma_5 = \prod_{\mu=1}^4 \gamma_\mu$, normalized so that $\gamma_5^2 = I$, and α is a constant parameter. The chiral current

$$J_5^\mu = \bar{\psi}_A \gamma^\mu \gamma_5 \psi^A \quad (9.4.4)$$

is classically conserved. At the quantum level, however, this conservation law can be violated and turns into an anomalous WT identity. To derive it, we consider the quantum effective action Γ defined by

$$e^{-\Gamma(A)} = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{-\int d^4x \mathcal{L}}, \quad (9.4.5)$$

and study its behavior under an infinitesimal chiral transformation of the fermions, with a space-time-dependent parameter $\alpha(x)$, given by³

$$\delta_\alpha \psi = i\alpha \gamma_5 \psi, \quad \delta_\alpha \bar{\psi} = i\alpha \bar{\psi} \gamma_5. \quad (9.4.6)$$

Since the external gauge fields A are inert, the transformation (9.4.6) represents a redefinition of dummy integration variables, and should not affect the effective action: $\delta_\alpha \Gamma = 0$. This statement carries however a non-trivial piece of information, since neither the classical action nor the integration measure is invariant under (9.4.6). The variation of the classical action under (9.4.6) is non-vanishing only for non-constant α , and has the form $\delta_\alpha \int \mathcal{L} = \int J_5^\mu \partial_\mu \alpha$. The variation of the measure is instead always non-vanishing, because the transformation (9.4.6) leads to a non-trivial Jacobian factor, which has the form $\delta_\alpha [\mathcal{D}\psi \mathcal{D}\bar{\psi}] = \exp\{i \int \alpha \mathcal{A}\} \mathcal{D}\psi \mathcal{D}\bar{\psi}$, as we will see below. In total, the effective action therefore transforms as

$$\delta_\alpha \Gamma = \int d^4x \alpha(x) \left[i\mathcal{A}(x) + \langle \partial_\mu J_5^\mu(x) \rangle \right]. \quad (9.4.7)$$

The condition $\delta_\alpha \Gamma = 0$ then implies the anomalous WT identity:

$$\langle \partial_\mu J_5^\mu \rangle = -i\mathcal{A}. \quad (9.4.8)$$

In order to compute the anomaly $\mathcal{A}$, we need to define the integration measure more precisely. This is best done by considering the eigenfunctions of the Dirac operator $\mathcal{D} \equiv i\gamma^\mu D_\mu$. Since the latter is Hermitian, the set of its eigenfunctions $\psi_k(x)$ with eigenvalues λ_k , defined by $\mathcal{D}\psi_k = \lambda_k \psi_k$, form an orthonormal and complete basis of spinor modes:

$$\int d^4x \psi_k^{\dagger a}(x) \psi_l^a(x) = \delta_{k,l}, \quad \sum_k \psi_k^{\dagger a}(x) \psi_k^b(y) = \delta^{ab} \delta^{(4)}(x-y), \quad (9.4.9)$$

where we have made explicit the spinor indices a and b . The fermion fields ψ and $\bar{\psi}$, which are independent from each other in Euclidean space, can be decomposed as

$$\psi = \sum_k a_k \psi_k, \quad \bar{\psi} = \sum_k \bar{b}_k \psi_k^\dagger, \quad (9.4.10)$$

so that the measure becomes

$$\mathcal{D}\psi \mathcal{D}\bar{\psi} = \prod_{k,l} da_k d\bar{b}_l. \quad (9.4.11)$$

Under the chiral transformation (9.4.6), we have

$$\delta_\alpha a_k = i \int d^4x \sum_l \psi_k^\dagger \alpha \gamma_5 \psi_l a_l, \quad \delta_\alpha \bar{b}_k = i \int d^4x \sum_l \bar{b}_l \psi_l^\dagger \alpha \gamma_5 \psi_k, \quad (9.4.12)$$

³For simplicity of the notation, we omit the gauge index A in the following equations. It will be reintroduced later on in this section.

and the measure (9.4.11) transforms as

$$\begin{aligned}\delta_\alpha [\mathcal{D}\psi \mathcal{D}\bar{\psi}] &= \mathcal{D}\psi \mathcal{D}\bar{\psi} \det(\delta_{kl} + i \int d^4x \psi_k^\dagger \alpha \gamma_5 \psi_l)^{-2} \\ &= \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp \left\{ -2i \sum_k \int d^4x \psi_k^\dagger \alpha \gamma_5 \psi_k + \mathcal{O}(\alpha^2) \right\}.\end{aligned}\quad (9.4.13)$$

For simplicity we can take α to be constant. This formal expression is ill-defined as it stands, since it decomposes into a vanishing trace over spinor indices ($\text{tr } \gamma_5 = 0$) times an infinite sum over the modes ($\sum_k 1 = \infty$). A convenient way of regularizing it is to introduce a gauge-invariant Gaussian cut-off. The anomaly $\mathcal{A}$ can then be defined as

$$\mathcal{A} = -2 \lim_{\beta \rightarrow 0} \sum_k \psi_k^\dagger \gamma_5 e^{-\beta(\not{D})^2} \psi_k. \quad (9.4.14)$$

Using the completeness relation in eq. (9.4.9), we can write

$$\mathcal{A} = -2 \lim_{\beta \rightarrow 0} \lim_{y \rightarrow x} \text{Tr} \left[\gamma_5 e^{-\beta(\not{D})^2} \right] \delta^{(4)}(x - y) = -2 \lim_{\beta \rightarrow 0} \int \frac{d^4k}{(2\pi)^4} \text{Tr} \left[\gamma_5 e^{-\beta(\not{k} + \not{D})^2} \right], \quad (9.4.15)$$

where the trace is over the spinor and the gauge indices. By using the commutator properties of the γ matrices, we can rewrite

$$(\not{k} + \not{D})^2 = (k + iD)^2 - \frac{1}{4} F^{\mu\nu} [\gamma_\mu, \gamma_\nu]. \quad (9.4.16)$$

Rescaling the momentum $k \rightarrow k/\sqrt{\beta}$, we get

$$\mathcal{A} = -2 \lim_{\beta \rightarrow 0} \frac{1}{\beta^2} \int \frac{d^4k}{(2\pi)^4} \text{Tr} \gamma_5 e^{-(k^2 + 2i\sqrt{\beta} D \cdot k - \beta D^2 - \beta F)}, \quad (9.4.17)$$

where $F \equiv F^{\mu\nu} [\gamma_\mu, \gamma_\nu]/4$. The trace over spinor indices is vanishing unless at least two factors of F (i.e. 4 γ 's) are included in the trace. In this way, we get two powers of β that compensate for the overall factor $1/\beta^2$ in eq. (9.4.17). Hence, in the limit $\beta \rightarrow 0$, we can safely neglect the terms proportional to $D \cdot k$ and D^2 in the exponential. In this way, we finally get

$$\mathcal{A} = -\frac{1}{16\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^\alpha F_{\rho\sigma}^\beta \text{tr}(T_\alpha T_\beta). \quad (9.4.18)$$

In Minkowski space, with $\partial_4 = -i\partial_0$, and in terms of hermitian (rather than anti-hermitian) generators and canonically normalized gauge fields ($AT \rightarrow -igAT$) the non-conservation of the axial current takes the form

$$\langle \partial_\mu J_5^\mu \rangle = -\frac{g^2}{16\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^\alpha F_{\rho\sigma}^\beta \text{tr}(T_\alpha T_\beta). \quad (9.4.19)$$

Anomalies in the chiral transformation (9.4.3) occur in any number of even space-time dimensions (in odd-dimensional space-times, chirality is not defined, since there is

no analogue of the matrix γ_5). The derivation of the anomaly reviewed here, due to Fujikawa, is easily generalized to any number of dimensions. In $2n$ dimensions, we define the chiral matrix γ_{2n+1} as $\gamma_{2n+1} = i^n \prod_{\mu=1}^{2n} \gamma_\mu$ so that $\gamma_{2n+1}^2 = I$ for any n . All the steps are essentially identical to the 4d case and we get the generalization of eq. (9.4.17):

$$\mathcal{A} = 2 \lim_{\beta \rightarrow 0} \frac{1}{\beta^n} \int \frac{d^{2n}k}{(2\pi)^{2n}} \text{Tr} \gamma_{2n+1} e^{-(k^2 + 2i\sqrt{\beta} D \cdot k - \beta D^2 - i\beta F)}. \quad (9.4.20)$$

In $2n$ dimensions, spinors have 2^n components, so we get for the anomaly

$$\mathcal{A} = -\frac{2}{(4\pi)^n n!} \epsilon^{\mu_1 \dots \mu_{2n}} F_{\mu_1 \mu_2}^{\alpha_1} \dots F_{\mu_{2n-1} \mu_{2n}}^{\alpha_n} \text{tr}(T_{\alpha_1} \dots T_{\alpha_n}). \quad (9.4.21)$$

The anomaly (9.4.21) can more elegantly be rewritten in a compact form in terms of differential two-forms. Let us introduce auxiliary anticommuting variables $\phi^1, \dots, \phi^{2n}$ and define

$$F = \frac{1}{2} F_{\mu\nu} \phi^\mu \wedge \phi^\nu. \quad (9.4.22)$$

Then eq. (9.4.21) becomes simply

$$\mathcal{A} = -2 \int d^{2n} \phi \text{tr} e^{F/(2\pi)}, \quad (9.4.23)$$

where the integration over the auxiliary fermion variables ϕ^μ automatically selects the correct number of field strengths F . In evaluating the integrated form of the anomaly, $\int d^{2n} x \mathcal{A}$, we can replace the auxiliary fermion variables ϕ by the differentials dx^μ , so that we have

$$\int d^{2n} x \mathcal{A} = -2 \int d^{2n} x \text{ch}(F), \quad (9.4.24)$$

where

$$\text{ch}(F) = \text{tr} e^{F/(2\pi)} \quad (9.4.25)$$

is the so-called Chern character of the gauge connection. There is a deep connection between anomalies and certain mathematical results that will not be discussed here.

9.5 The Wess-Zumino Consistency Conditions**

In this section we will consider some more formal developments about anomalies. It is convenient in this context to consider the effective action $\Gamma(A)$ arising upon integration of the fermions:

$$e^{-\Gamma(A)} = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{-S(\psi, \bar{\psi}, A)}. \quad (9.5.1)$$

In presence of an anomaly, $\Gamma(A)$ is not gauge invariant. Indeed, under an infinitesimal gauge transformation $A \rightarrow A + D\epsilon_1$

$$\delta_{\epsilon_1}\Gamma(A) = \int d^4x \frac{\delta\Gamma(A)}{\delta A_\mu^\alpha} D_\mu \epsilon_1^\alpha = - \int d^4x D_\mu J_\alpha^\mu \epsilon_1^\alpha \equiv - \int d^4x \mathcal{A}_\alpha \epsilon_1^\alpha = - \int d^4x \epsilon_1^\alpha \mathcal{F}_\alpha \Gamma(A), \quad (9.5.2)$$

where the third equality defines the anomaly term $\mathcal{A}_\alpha$ and we have defined the functional operator

$$\mathcal{F}_\alpha = \partial_\mu \frac{\delta}{\delta A_\mu^\alpha} + C_{\alpha\beta\gamma} A_\mu^\beta \frac{\delta}{\delta A_\mu^\gamma}. \quad (9.5.3)$$

Under a further infinitesimal transformations, we have

$$\delta_{\epsilon_2} \delta_{\epsilon_1} \Gamma(A) = \int d^4x \int d^4y \epsilon_2^\beta(y) \epsilon_1^\alpha(x) \mathcal{F}_\beta(y) \mathcal{F}_\alpha(x) \Gamma(A). \quad (9.5.4)$$

Due to the group structure, performing the commutators of the two infinitesimal transformations parametrized by ϵ_1 and ϵ_2 should be equivalent to perform a single infinitesimal transformation with parameter $[\epsilon_1, \epsilon_2]$, where $\epsilon_{1,2} = \epsilon_{1,2}^\alpha T_\alpha$. In other words, we should have

$$[\delta_{\epsilon_1}, \delta_{\epsilon_2}] \Gamma(A) = \delta_{[\epsilon_1, \epsilon_2]} \Gamma(A), \quad (9.5.5)$$

identity that can easily be verified to hold in general. The relation (9.5.5) implies a non-trivial condition for the anomaly, known as the Wess-Zumino consistency conditions:

$$\mathcal{F}_\alpha(x) \mathcal{A}_\beta(y) - \mathcal{F}_\beta(y) \mathcal{A}_\alpha(x) = -C_{\alpha\beta\gamma} \mathcal{A}_\gamma(x) \delta(x-y). \quad (9.5.6)$$

These conditions can be more conveniently expressed in terms of BRST transformations by defining

$$G(\omega, A) \equiv \int d^4x \omega_\alpha(x) \mathcal{A}_\alpha(x). \quad (9.5.7)$$

Equations (9.5.6) imply that G is BRST-invariant, namely

$$sG(\omega, A) = 0. \quad (9.5.8)$$

This is shown by recalling the BRST transformations (6.3.3) of ω and A , according to which

$$\begin{aligned} sG(\omega, A) &= \int d^4x \left(-\frac{1}{2} C_{\alpha\beta\gamma} \omega_\beta(x) \omega_\gamma(x) \mathcal{A}_\alpha(x) \right. \\ &\quad \left. - \omega_\alpha(x) \int d^4y \frac{\delta \mathcal{A}_\alpha(x)}{\delta A_\mu^\beta(y)} (\partial_\mu \omega_\beta(y) + C_{\beta\gamma\delta} A_\mu^\gamma(y) \omega_\delta(y)) \right) \\ &= \int d^4x d^4y \left(-\frac{1}{2} \omega_\alpha(x) \omega_\beta(y) \right) \left[\delta(x-y) C_{\alpha\beta\gamma} \mathcal{A}_\gamma(x) + \mathcal{F}_\alpha(x) \mathcal{A}_\beta(y) - \mathcal{F}_\beta(x) \mathcal{A}_\alpha(y) \right], \end{aligned} \quad (9.5.9)$$

which vanishes if eq. (9.5.6) is satisfied.

We have seen in the previous sections that there is some arbitrariness in computing anomalies, related to the fact that we can shift the latter from one current to another. From an effective action point of view, this shift corresponds to the possibility of adding local, non-gauge invariant, counter-terms to $\Gamma(A)$ that change $\delta_\epsilon \Gamma(A)$ and hence the anomaly. The impossibility of keeping all currents conserved corresponds to the impossibility of finding a local⁴ counter-term such that $\delta_\epsilon \Gamma(A) = 0$. We can now be more precise exploiting eq. (9.5.8). If there existed a local functional of the gauge fields $F(A)$, such that $G(\omega, A) = sF(A)$, namely suppose that $G(\omega, A)$ would be BRST-exact, then the anomaly would be cancelled by adding to $\Gamma(A)$ the counter-term $-F(A)$. Anomalies then form an equivalence class. Two anomalies related by the addition of a local functional $F(A)$ are equivalent and belong to the same cohomology of the BRST operator s . Thanks to the Wess-Zumino consistency conditions, it is possible to reconstruct the whole form of the non-abelian anomaly by only knowing the terms quadratic in the gauge fields. The procedure does not uniquely fix the anomaly since, as we have just said, the latter can be changed by adding local counter-terms to the action.

We finally mention about the existence of an elegant formalism, known as the Stora-Zumino descent equations, that allows, in any number of even dimensions, to get anomaly functionals $\mathcal{A}_\alpha$ that automatically satisfy the conditions (9.5.6). The descent equations make also manifest the close relationship between chiral anomalies in $2n + 2$ dimensions and gauge anomalies in $2n$ dimensions.

9.6 't Hooft Anomaly Matching and the Wess-Zumino-Witten Term*

Asymptotically free gauge theories are strongly coupled in the IR and can give rise to confinement of its fermion constituents, like quarks in QCD. At low energies the propagating degrees of freedom are bound states of the elementary high energy (UV) states, such as mesons and hadrons in QCD. More generally, let us consider a gauge theory coupled to fermions with local symmetry G_s and global symmetry G_f and denote schematically by D_{abc} the associated anomaly coefficients, where $a, b, c = s, f$. We assume that the theory in the IR becomes strongly coupled. For the theory to be consistent $D_{sss} = 0$ so that no gauge anomalies arise. We also have $D_{fss} = 0$, which means that we keep in G_f only the

⁴Locality is crucial. By using non-local functionals, any anomaly can be cancelled. For instance, a gauge $U(1)^3$ anomaly would be cancelled by adding to $\Gamma(A)$ the non-local functional

$$-F(A) = \frac{g^2}{96\pi^2} \frac{1}{\square} \partial^\mu A_\mu \epsilon^{\alpha\beta\gamma\delta} F_{\alpha\beta} F_{\gamma\delta}. \quad (9.5.10)$$

quantum global symmetries of the theory, not the classical ones. For instance, in QCD we should not include $U(1)_A$ in G_f . We can however have 't Hooft anomalies $D_{fff} \neq 0$. Although these anomalies do not lead to any real effect, 't Hooft has shown that they are invariant under RG flow. They might then be useful to put constraints on the IR theory.

For example, if the anomaly is induced in the UV by elementary fermion fields, in the IR massless fermion bound states with suitable quantum numbers should appear to give rise to the same anomaly coefficients D_{fff} . If G_f is spontaneously broken (like chiral symmetries in QCD), the NG bosons can replace the role of the massless fermion bound states, no longer required. In this situation the low energy Lagrangian of the NG bosons should contain certain terms that reproduce the anomaly.

The argument to prove the RG invariance of the anomalies is simple. Let us assume of (weakly) gauging the group G_f (or a subgroup of it) of the UV theory. In this way the innocuous global anomalies turn into deadly gauge anomalies making the theory inconsistent. We might cancel the gauge anomaly D_{fff} by adding suitable additional massless “spectator” fermions that give a contribution $D_{fff}^{\text{spect.}} = -D_{fff}$ to the anomaly. Crucially, these spectators can be taken neutral under G_s and charged only under the weakly gauged symmetries G_f . At low energies, when the strong gauge group G_s confines, the spectrum of the theory will include IR bound states plus the “spectator” fermions that, being neutral under G_s , are unaffected by the condensation of G_s . The UV theory with the spectators is, by construction, consistent and it has to remain so for all values of the gauge coupling constant. Hence, at low energies, there must be an anomaly contribution equal to $-D_{fff}^{\text{spect.}} = D_{fff}$, canceling that of the fermion spectators. The argument is valid for an arbitrarily weak gauging and thus has to also apply in the original theory where G_f is a global symmetry.

Given the quantum numbers of the elementary constituents, the possible quantum numbers of the massless fermion bound states can be argued. It might happen that there is no choice of quantum numbers for the bound states to reproduce the anomaly of the constituents. In this case, the 't Hooft anomaly matching conditions can be used to prove that our original assumption of unbroken global symmetries is necessarily violated, and some global symmetries are spontaneously broken. Notably, one can show in this way that in QCD the $SU(3)_V \times SU(3)_A \times U(1)_V$ global symmetry must be spontaneously broken, see e.g. section 22.5 of ref.[60].⁵

As we mentioned, when G_f is spontaneously broken the anomaly of the UV fermions can be reproduced by the effective action of the (pseudo) NG bosons. This is the situation

⁵The argument requires $n_f = 3$. Solutions for the bound states quantum numbers can instead be found when $n_f = 2$.

actually realized in QCD, in which case the weak gauging used in 't Hooft argument can be identified with the electromagnetic interactions. Consider for simplicity $n_f = 2$ (up and down quarks only). The quantum global symmetry group is $G_f = SU(2)_V \times SU(2)_A \times U(1)_V$. This is explicitly broken by the electromagnetic interactions to $H_f = U(1)_V \times U(1)_{EM} \times \tilde{U}(1)_A$, where $\tilde{U}(1)_A \subset SU_A(2)$ is the abelian subgroup generated by t^3 . Consider now the $U(1)_{EM}^2 \times \tilde{U}(1)_A$ axial anomaly and denote by $\tilde{J}_A^\mu$ the $\tilde{U}(1)_A$ axial current. We get

$$\begin{aligned}\partial_\mu \tilde{J}_A^\mu &= -\frac{N_c}{16\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} \left[\left(\frac{2e}{3}\right)^2 \times 1 + \left(\frac{-e}{3}\right)^2 \times (-1) \right] \\ &= -\frac{N_c e^2}{48\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} .\end{aligned}\tag{9.6.1}$$

Anomaly matching requires that a similar anomaly should show up in the IR. Since no massless fermion bound states occur in Nature, this should appear in the low-energy effective chiral Lagrangian $\mathcal{L}_\pi$ describing the dynamics of the π mesons interacting with photons. Under the $\tilde{U}(1)_A$ chiral transformation above, $\delta_\epsilon \mathcal{L}_\pi$ should then not vanish, but rather reproduce the axial anomaly (9.6.1). Considering that $\delta_\epsilon \pi^0 = \epsilon f_\pi$, $\mathcal{L}_\pi$ should include the coupling

$$\mathcal{L}_\pi \supset -\frac{N_c e^2}{48\pi^2 f_\pi} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} \pi^0 .\tag{9.6.2}$$

The axial anomaly (9.6.1) has allowed to resolve the puzzle of the $\pi^0 \rightarrow 2\gamma$ decay. In absence of any anomaly, the term (9.6.2) would still appear in the chiral Lagrangian $\mathcal{L}_\pi$ but with a much more suppressed coupling. On the contrary, anomaly considerations uniquely fix its coefficient and it turns out that the experimental rate $\Gamma(\pi^0 \rightarrow 2\gamma)$ is successfully reproduced with $N_c = 3$. We have seen in section 8.7 that the chiral Lagrangian $\mathcal{L}_\pi$ should be described in terms of the matrix of fields $U = \exp(2i\pi^a t^a / f_\pi)$, rather than by the π 's mesons themselves. The anomalous term (9.6.2) should then be rewritten in terms of the U 's. This rewriting is not totally straightforward and will not be done here. The ending result goes under the name of the gauged version of the Wess-Zumino-Witten term. The latter includes many other couplings, including the term (9.6.2).

9.7 Anomalous Breaking of Scale Invariance^{**}

As we have mentioned at the beginning of this chapter, in most QFT's the symmetry under scale transformations, if present at tree level, is broken by quantum effects due to the energy dependence of the couplings.⁶ Let us formalize a bit better this observation.

⁶The analysis in this section closely follows section 19.5 of ref. [49].

We know that the energy-momentum tensor is the conserved tensor associated to the translational symmetries. In general, under an infinitesimal symmetry transformation of the action, the Lagrangian density is not necessarily invariant, but can change by a total derivative: $\mathcal{L} \rightarrow \mathcal{L} + \epsilon \partial_\mu \mathcal{J}^\mu$, with $\mathcal{J}^\mu$ a given tensor.⁷ The associated conserved Noether current in this case is given by

$$J^\mu = \frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \delta \phi - \mathcal{J}^\mu. \quad (9.7.1)$$

Under an infinitesimal translation $x^\mu \rightarrow x^\mu + \epsilon^\mu$, the Lagrangian, being a scalar, transforms as $\mathcal{L} \rightarrow \mathcal{L} + \epsilon^\mu \partial_\mu \mathcal{L}$. The Noether energy-momentum tensor reads

$$T^{\mu\nu} = \frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \partial^\nu \phi - \eta^{\mu\nu} \mathcal{L}, \quad (9.7.2)$$

and is conserved: $\partial_\mu T^{\mu\nu} = 0$. In general, the Noether energy-momentum tensor $T^{\mu\nu}$ is not symmetric, $T^{\mu\nu} \neq T^{\nu\mu}$. A new, symmetric, energy-momentum tensor $\theta^{\mu\nu}$ can be defined by adding a trivially conserved term to $T^{\mu\nu}$:

$$\theta^{\mu\nu} = T^{\mu\nu} + \partial_\rho M^{\mu\nu\rho}, \quad (9.7.3)$$

where $M^{\mu\nu\rho}$ is a tensor antisymmetric in μ and ρ , so that $\partial_\mu \partial_\rho M^{\mu\nu\rho}$ automatically vanishes. The symmetric energy-momentum tensor $\theta^{\mu\nu}$ can directly be derived by coupling the QFT to gravity and noticing that $\theta^{\mu\nu}$ minimally couples to gravity. We have

$$\theta^{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g_{\mu\nu}} \Big|_{g_{\mu\nu} = \eta_{\mu\nu}}, \quad (9.7.4)$$

where $g_{\mu\nu}$ is the metric and $g = \det g_{\mu\nu}$.

Let us see how this works for the massless ϕ^4 theory. In this case we have

$$T^{\mu\nu} = \theta^{\mu\nu} = \partial^\mu \phi \partial^\nu \phi - \eta^{\mu\nu} \mathcal{L}. \quad (9.7.5)$$

Coupling to gravity we have

$$S = \int d^4x \sqrt{-g} \left(\frac{1}{2} g_{\mu\nu} \partial^\mu \phi \partial^\nu \phi - \frac{\lambda}{4!} \phi^4 \right). \quad (9.7.6)$$

Noticing that

$$\frac{\delta \sqrt{-g}}{\delta g_{\mu\nu}} = \frac{\delta}{\delta g_{\mu\nu}} e^{1/2 \text{Tr} \log(-g)} = -\frac{1}{2} g^{\mu\nu} \sqrt{-g}, \quad (9.7.7)$$

we immediately recover eq. (9.7.5).

⁷In fact, this is not the most general possibility. A counterexample is provided by the transformation of the Lagrangian density under scale transformations considered below.

A QFT is (classically) scale invariant if its Lagrangian density $\mathcal{L}(x) \rightarrow e^{-4\sigma}\mathcal{L}(x)$ when $x^\mu \rightarrow e^\sigma x^\mu$, with σ a real constant parameter, so that the action $S = \int d^4x \mathcal{L}$ remains invariant. The transformations of fields are dictated by their classical dimension Δ :

$$\phi(x) \rightarrow e^{-\sigma\Delta}\phi(x). \quad (9.7.8)$$

The classical scaling dimension Δ coincides with the power counting dimension of a field, $\Delta = +1$ for scalars and vectors, and $\Delta = 3/2$ for fermions in four space-time dimensions. Given the action on the coordinates, each derivative acting on a field increases by one the scaling dimension of the operator:

$$\partial_{\mu_1} \dots \partial_{\mu_n} \phi(x) \rightarrow e^{-\sigma(\Delta+n)} \partial_{\mu_1} \dots \partial_{\mu_n} \phi(x). \quad (9.7.9)$$

A QFT is classically scale invariant if only marginal operators appear in its Lagrangian density. In particular, any mass term explicitly breaks the symmetry. The current associated to scaling transformations is called the dilatation current D_μ . Finding its explicit form with the Noether method is non-trivial, because the Lagrangian density is not invariant, even up to total derivatives. On the other hand, it is easy to find D_μ by using the same trick used before of coupling the QFT to gravity. In this way, we can reinterpret the scaling transformation $x^\mu \rightarrow e^\sigma x^\mu$ as a rescaling of the metric $g_{\mu\nu} \rightarrow e^{-2\sigma} g_{\mu\nu}$. Multiplying eq. (9.7.4) by $\delta g_{\mu\nu} = -2\epsilon g_{\mu\nu}$, we get that the action variation vanishes if and only if

$$\theta_\mu^\mu = 0. \quad (9.7.10)$$

The dilatation current is

$$D^\mu = x_\nu \theta^{\mu\nu}, \quad (9.7.11)$$

and is in fact conserved if $\theta^{\mu\nu}$ is traceless. Notice that $\theta^{\mu\nu}$ entering eq. (9.7.11) does not in general coincide with the $\theta^{\mu\nu}$ in eq. (9.7.4), but is related to it by a total derivative of the form (9.7.3). This is the case even in the simplest theories. For instance, in the massless ϕ^4 theory,

$$\theta_\mu^\mu = (\partial_\mu \phi)^2 - 4\mathcal{L} = -(\partial_\mu \phi)^2 + \frac{\lambda}{3!} \phi^4 \neq 0. \quad (9.7.12)$$

We can fix this problem by redefining the energy-momentum tensor. By adding a manifestly conserved tensor, we have for the $\lambda\phi^4$ theory

$$\theta^{\mu\nu} \rightarrow \theta^{\mu\nu} - \frac{1}{6}(\partial^\mu \partial^\nu - \eta^{\mu\nu} \square)\phi^2. \quad (9.7.13)$$

The trace of the redefined energy-momentum tensor reads

$$\theta_\mu^\mu = -(\partial_\mu \phi)^2 + \frac{\lambda}{3!} \phi^4 + (\partial_\mu \phi)^2 + \phi \square \phi = \phi \left(\square \phi + \frac{\lambda}{3!} \phi^3 \right) = 0, \quad (9.7.14)$$

where in the last equality we used the equation of motion of ϕ . One can easily show that by adding a mass term to the ϕ^4 theory, one gets

$$\theta_\mu^\mu = m^2 \phi^2. \quad (9.7.15)$$

No redefinition of θ_ν^μ can get rid of the mass term, which is in fact a genuine source of explicit violation of the scale symmetry. Alternatively, we can keep the energy-momentum tensor as defined in eq. (9.7.4) and change the definition of the dilatation current as

$$D^\mu = x_\nu \theta^{\mu\nu} + V^\mu. \quad (9.7.16)$$

The field V^μ is denoted the field virial. For the ϕ^4 theory $V^\mu = \phi \partial^\mu \phi$.

At the quantum level, scale invariance is typically broken by quantum effects. The “would-be” WT identities implied by scale invariance are

$$\langle \theta_\mu^\mu(q) \phi_1(p_1) \dots \phi_n(p_n) \rangle = - \sum_{r=1}^n \Delta_r \langle \phi_1(p_1) \dots \phi_r(p_r + q) \dots \phi_n(p_n) \rangle, \quad (9.7.17)$$

where Δ_r is the scaling dimension of the field ϕ_r . Quantum corrections responsible of the anomalous dimensions of the field do not violate the WT identities (9.7.17), that continue to hold with a redefinition of Δ_r that includes the anomalous dimension contribution. The actual breaking arises from the energy-dependence of couplings, namely from their β -function. Invariance of the *quantum* action under scaling transformations would imply that

$$\begin{aligned} \int d^4x \mathcal{L}(\phi(x), \partial_\mu \phi(x), \lambda(x)) &= \int d^4x \mathcal{L}(e^{-\sigma\Delta} \phi(e^{-\sigma}x), e^{-\sigma(\Delta+1)} \partial_\mu \phi(e^{-\sigma}x), \lambda(x)) \\ &= \int d^4x e^{4\sigma} \mathcal{L}(e^{-\sigma\Delta} \phi(x), e^{-\sigma(\Delta+1)} \partial_\mu \phi(x), \lambda(e^\sigma x)), \end{aligned} \quad (9.7.18)$$

where λ denotes a coupling constant and in the last equality we have redefined the coordinates $x^\mu \rightarrow e^\sigma x^\mu$. For a classically scale invariant theory $\mathcal{L}(e^{-\sigma\Delta} \phi(x), e^{-\sigma(\Delta+1)} \partial_\mu \phi(x), \lambda) = e^{-4\sigma} \mathcal{L}(\phi(x), \partial_\mu \phi(x), \lambda)$. Under an infinitesimal transformation,

$$\lambda(x + \epsilon x) = \lambda(x) + \epsilon \beta(\lambda), \quad (9.7.19)$$

and the violation of the scaling symmetry is given by

$$\delta S = \int d^4x \left(\mathcal{L}(\phi(x), \partial_\mu \phi(x), \lambda(x + \epsilon x)) - \mathcal{L}(\phi(x), \partial_\mu \phi(x), \lambda(x)) \right) = \epsilon \int d^4x \frac{\partial \mathcal{L}}{\partial \lambda} \beta(\lambda). \quad (9.7.20)$$

Correspondingly, the divergence of the dilatation current is anomalous, with the anomaly given by β :

$$\partial_\mu D^\mu = \beta(\lambda) \frac{\partial \mathcal{L}}{\partial \lambda}. \quad (9.7.21)$$

Let us see how this explicitly works for the usual ϕ^4 theory. We work in DR with $d = 4 - \epsilon$. First of all, since we are no longer in 4 dimensions, we have to reconsider how to redefine $\theta^{\mu\nu}$, like in eq. (9.7.13). Let

$$\theta_B^{\mu\nu} \rightarrow \theta_B^{\mu\nu} + \alpha(\partial^\mu \partial^\nu - \eta^{\mu\nu} \square) \phi_B^2, \quad (9.7.22)$$

where the subscript B stands for bare quantities. On-shell, the trace of the redefined energy-momentum tensor reads

$$\theta_{B\mu}^\mu = \left(1 - \frac{d}{2} + 2\alpha(1-d)\right)(\partial_\mu \phi_B)^2 - \left(d + 8\alpha(1-d)\right) \frac{\lambda_B}{4!} \phi_B^4. \quad (9.7.23)$$

Let us choose α so that the term proportional to $(\partial_\mu \phi_B)^2$ vanishes:

$$\alpha = \frac{(1 - \frac{d}{2})}{2(d-1)}. \quad (9.7.24)$$

We get

$$\theta_{B\mu}^\mu = (d-4) \frac{\lambda_B}{4!} \phi_B^4. \quad (9.7.25)$$

We now have

$$\lambda_B \phi_B^4 = Z_\lambda \lambda \phi^4, \quad (9.7.26)$$

with

$$Z_\lambda = 1 + \frac{3\lambda}{16\pi^2} \frac{1}{\epsilon} \quad (9.7.27)$$

at one-loop level. Substituting in eq. (9.7.25) gives

$$\theta_\mu^\mu = \theta_{B\mu}^\mu = -\frac{3\lambda^2}{16\pi^2} \frac{1}{4!} \phi^4 = \beta(\lambda) \frac{\partial \mathcal{L}}{\partial \lambda}, \quad (9.7.28)$$

which is the expected result. Notice that θ_μ^μ eventually is finite with no need of any wave function renormalization, in agreement to the general result that conserved currents do not renormalize, $Z_J = 1$, and correspondingly have vanishing anomalous dimensions.

9.8 The Strong CP Problem and a Possible Solution: Axions*

Strong interactions seem to respect to a high degree of accuracy invariance under parity P, charge conjugation C and time reversal T. On the other hand, the QCD Lagrangian might contain a gauge and Lorentz invariant dimension four operator of the form

$$\frac{\theta}{64\pi^2} \int d^4x \epsilon_{\mu\nu\rho\sigma} G_a^{\mu\nu}(x) G_a^{\rho\sigma}(x) \quad (9.8.1)$$

where $G_a^{\mu\nu}$ are the (non-canonically normalized) gluon field-strengths and θ is a real parameter. The parameter θ in eq.(9.8.1) is actually not physical. Under a chiral rotation of the quark fields

$$Q_f \rightarrow e^{i\alpha_f \gamma_5} Q_f, \quad (9.8.2)$$

where f is a flavour index, the QCD action is not invariant due to the global anomaly (9.1.40). The non-invariance of the action corresponds to a redefinition of the parameter θ :

$$\theta \rightarrow \theta + 2 \sum_f \alpha_f. \quad (9.8.3)$$

The chiral transformation (9.8.2) affects also the quark masses:

$$m_f \rightarrow e^{2i\alpha_f} m_f. \quad (9.8.4)$$

Hence the invariant parameter $\bar{\theta}$ is the combination

$$\bar{\theta} = \theta + i \log \prod_f m_f. \quad (9.8.5)$$

If any of the quark masses were zero, θ would not be observable, since we can set it to zero by rotating the massless quark. If $m_f \neq 0$, as it appears to be in the real world, $\bar{\theta}$ is a genuine physical source of parity and time reversal violation (and hence also of CP by the CPT theorem) in the strong interactions.

The term (9.8.1) can be shown to be a total derivative and is completely irrelevant at the perturbative level.⁸ Yet, at low energies, when QCD confines, it is expected to give rise to CP violating operators. A particularly relevant CP odd operators is the neutron electric dipole moment operator,

$$i d_N \bar{N} \sigma^{\mu\nu} \gamma_5 N F_{\mu\nu}, \quad (9.8.6)$$

where N is the neutron fermion field. Needless to say, computing d_N from first principles is an hard task, considering also that the neutron has a mass of order Λ_{QCD} and cannot be straightforwardly included in the meson low energy action described in section 8.7. A rough simple estimate of the order of magnitude of d_N is however possible. Since d_N has to vanish if any quark mass is zero, eq.(8.7.9) and dimensional analysis suggest that

$$|d_N| \sim e |\bar{\theta}| \frac{m_\pi^2}{m_N^3} \quad (9.8.7)$$

⁸This term plays an important role when studying non-perturbative configurations in QCD, instantons. We will not consider such effects in these lectures.

The current experimental upper bounds on the neutron electric dipole moment is $|d_N| \leq 2.9 \times 10^{-26}$ e cm. Using the theoretical guess (9.8.7) gives the bound

$$|\bar{\theta}| \lesssim 10^{-10}. \quad (9.8.8)$$

More refined estimates confirm the order of magnitude bound given in eq.(9.8.8).

According to the naturalness criterion reviewed in section 7.6, a dimensionless coupling can be very small if a symmetry is restored in the limit it vanishes. This would indeed be the case for $\bar{\theta}$ in QCD, in isolation. In the real world, however, the electroweak interactions break both P and CP and the naturalness of a small $\bar{\theta}$ is not guaranteed. We are left with the question: why is $\bar{\theta}$ so small?⁹ This naturalness problem is often denoted strong CP problem.

An elegant solution is achieved by introducing a new field $a(x)$, called the axion, that is supposed to be the NG boson of a spontaneously broken $U(1)$ symmetry (at energies much larger than Λ_{QCD}). The original formulation of this idea was due to Peccei and Quinn and for this reason this $U(1)$ symmetry is often denoted Peccei-Quinn (PQ) symmetry. Under a $U(1)_{PQ}$ transformation parametrized by ω , the axion shifts

$$a(x) \rightarrow a(x) + f_a \omega, \quad (9.8.9)$$

where f_a is the axion decay constant. According to the analysis made in sections 8.6 and 8.9, the non-linear realization of the $U(1)_{PQ}$ symmetry requires the axion to have only derivative interactions. Suppose now that the $U(1)_{PQ}$ symmetry, in addition to be spontaneously broken at the scale f_a , is also explicitly broken by (and only by) the strong interactions by means of $SU(3)_c^2 U(1)_{PQ}$ anomalies. Then, in analogy to eq.(9.6.2), the axion effective theory should contain an interaction of the form

$$\frac{1}{64\pi^2 f_a} \int d^4x a(x) \epsilon_{\mu\nu\rho\sigma} G_a^{\mu\nu}(x) G_a^{\rho\sigma}(x). \quad (9.8.10)$$

This is the only allowed non-derivative term in the axion effective action.¹⁰ In presence of this interaction, the θ term (9.8.1) can trivially be reabsorbed in eq.(9.8.10) by the $U(1)_{PQ}$ symmetry (9.8.9). We have essentially promoted θ to be a field, so the problem of the

⁹Notice that this naturalness problem is qualitatively different from the one affecting the Higgs mass (or more generally, relevant operators). There, a delicate cancellation (or fine-tuning) has to occur to get a small number out of big ones, induced by quantum corrections. Here $\bar{\theta}$ is radiatively stable and we have “simply” to understand the smallness of its classical value.

¹⁰Note that $U(1)_{PQ}$ is broken only at the non-perturbative level in α_c , since under the shift (9.8.9) the term (9.8.10) changes by a total derivative. This is non-vanishing only around certain semi-classical configurations denoted instantons in euclidean space. We will not need to discuss instantons in what follows and instead proceed with a much simpler analysis based on effective field theory.

value of θ becomes now a dynamical one. We have to compute the low-energy effective potential for the axion $a(x)$ and find its minima. The best way to study the axion potential is by exploiting the QCD NG boson Lagrangian already introduced in section 8.7. For simplicity consider $n_f = 2$, keeping only the up and down quarks. By a chiral rotation, we can eliminate the VEV of $a(x)$ from eq.(9.8.10) and reabsorb it in, say, the up quark mass term. As we have seen, quark mass terms are sources of explicit violation of the chiral symmetry, responsible for the pion masses and, more generally, for a pion potential. The leading terms are given in eq.(8.7.8). When the axion VEV is reabsorbed into M , this term can be seen as the leading axion-pion effective potential term:

$$V(a, \pi) = -cf_\pi^3 \text{tr}(UM + M^\dagger U^\dagger) \quad (9.8.11)$$

where $U = \exp(i\sigma^a \pi^a / f_\pi)$ and

$$M = \begin{pmatrix} m_u e^{-i\frac{a}{f_a}} & 0 \\ 0 & m_d \end{pmatrix}. \quad (9.8.12)$$

In eq.(9.8.12), a is the axion VEV. By setting $\pi^\pm = 0$, the explicit form of V reads

$$V(a, \pi_0) = -2cf_\pi^3 \left(m_d \cos\left(\frac{\pi_0}{f_\pi}\right) + m_u \cos\left(\frac{a}{f_a} - \frac{\pi_0}{f_\pi}\right) \right). \quad (9.8.13)$$

The extrema of V are found at $a = 0 \bmod \pi f_a$ and $\pi_0 = 0 \bmod \pi f_\pi$. A straightforward computation shows that the minimum is at

$$a = \pi_0 = 0. \quad (9.8.14)$$

Hence the θ angle dynamically vanishes, providing a solution to the strong CP problem. From the potential (9.8.13) we can easily compute the axion mass. Expanding for $f_a \gg f_\pi, m_u, m_d$ we get (assuming a canonical kinetic term for a)

$$m_a^2 = \frac{f_\pi^2}{f_a^2} \frac{m_u m_d}{(m_u + m_d)^2} m_\pi^2. \quad (9.8.15)$$

Once again, we see the power of an effective description of NG bosons, which has allowed us to compute the axion mass in a few simple steps. The axion decay constant f_a governs all the derivative interactions that the axion can possibly have in the UV theory with quarks and gauge fields, and with mesons, baryons and photons in the IR. Astrophysical constraints coming from red giant cooling puts a lower bound on f_a :

$$f_a > 10^9 \text{ GeV}. \quad (9.8.16)$$

Given eq.(9.8.15) and the above bound, we conclude that the axion must be an extremely light particle:

$$m_a \lesssim 10^{-3} \text{ eV} . \tag{9.8.17}$$

The axion has the added virtue of being also a viable dark matter candidate. Upper bounds, coming mainly from cosmological considerations, also exist. They depend on some details about the early evolution of the universe, and will not be discussed here.

Chapter 10

Some Formal Developments[★]

Perturbation theory in QFT is the most important tool at our disposal for analytic considerations and has been a recurrent theme in these notes. It is now time to discuss the nature of the perturbative series from a mathematical point of view. We will see that this series is not convergent, yet it is useful when properly interpreted. In the second part of this chapter we will derive the effective action of the photon when massive charged matter fields are integrated out. This will be done by mapping the problem to that of a partition function of a quantum mechanical problem. The result will be exact for constant field strength configurations and would allow us to study the non-perturbative decay of the vacuum in presence of a constant electric field.

The structure of this chapter is as follows. In section 10.1 we show the divergent nature of the perturbative series, the notion of optimal truncation, and briefly discusses Borel resummation to resum the perturbative series. In section 10.2 we derive the photon effective theory (Euler-Heisenberg action) for both scalar and fermion matter, and then discuss the vacuum instability in presence of a constant electric field.

10.1 Asymptotic Nature of Perturbation Theory[★]

Most of the considerations made in these notes are based on perturbation theory. It is natural at this stage to ask what are the convergent properties of the associated series. Perturbative expansions in QFT give rise to power series in some coupling constant λ or to $\hbar$, if we consider the analogous loopwise expansion. It is useful to recall here some basic mathematical properties of power series of holomorphic functions. If $f(\lambda)$ is analytic at a point λ_0 , in a small disc around λ_0 the function is given by the power series

$$f(\lambda) = \sum_{n=0}^{\infty} f_n(\lambda), \quad f_n(\lambda) = c_n(\lambda - \lambda_0)^n. \quad (10.1.1)$$

The radius of convergence R of the series (10.1.1) is given by

$$R^{-1} = \limsup_{n \rightarrow \infty} |c_n|^{\frac{1}{n}}. \quad (10.1.2)$$

If λ_0 is a regular point of $f(\lambda)$, R is non-vanishing and is given by the distance of λ_0 from the closest singularity of $f(\lambda)$. Viceversa, if R is non-vanishing, necessarily λ_0 is a regular point of the function f . The power series (10.1.1) is uniformly convergent for any $|\lambda| < R$ and divergent for $|\lambda| > R$. The convergence at $|\lambda| = R$ depends on the particular cases, but necessarily there is at least a point where the series diverges, corresponding to the singular point of $f(\lambda)$ closest to λ_0 .

F. Dyson in 1952 presented an argument that led to the conclusion that the perturbative series expansion in QED has zero radius of convergence. It is worth to sketch here Dyson's original simple and brilliant argument. Recall that the QED expansion is a power series in $\alpha \propto e^2$ and not in the charge e . Suppose now that the QED expansion has a non-vanishing radius of convergence. This would imply that the point $\alpha = 0$ should be a regular point of a physical observable $O(\alpha)$, seen as an analytic function of α . As we recalled, this would imply that $O(\alpha)$ has a finite radius of convergence where the function is analytic, including regions where $\alpha < 0$. A world with $\alpha < 0$ is however unstable, because electrons and positrons would repel each other destroying any vacuum in an uncontrolled production of electron-positron pairs.¹ We conclude that sensible physical observables cannot be analytic for $\alpha < 0$ and thus the point $\alpha = 0$ itself cannot be analytic. In turn, this implies that the series around $\alpha = 0$ has zero radius of convergence.

A modern and more general version of Dyson's argument can be obtained by considering the euclidean path integral formulation in QFT. Consider a generic n -point function

$$G^{(n)}(x_1, x_2, \dots, x_n; \hbar) = \int \mathcal{D}\phi \phi(x_1) \phi(x_2) \dots \phi(x_n) e^{-S(\phi)/\hbar}, \quad (10.1.3)$$

where we denote collectively all the fields in the action by ϕ .² In euclidean space the action is positive definite and the path integral converges.³ Consider now the loopwise expansion, that is the expansion of $G^{(n)}$ in powers of $\hbar$. The point $\hbar = 0$ is necessarily non-analytic because for any value of $\hbar < 0$ the Green functions $G^{(n)}$ blow up. We conclude that loopwise perturbative expansions in QFT have generically zero radius of convergence and

¹A similar event occurs for $\alpha > 0$ in presence of a constant electric field. We will study in some detail this phenomenon in the subsequent sections of this chapter.

²This argument works also in presence of fermions. We integrate them out and identify the resulting effective action with S in eq.(10.1.3).

³With an appropriate measure and upon renormalization. Strictly speaking a full non-perturbative definition of the path integral requires a lattice discretization of space-time.

are divergent. A similar conclusion applies for coupling constant expansions that upon rescaling of the fields are equivalent to a loopwise expansion. Considering for simplicity a single coupling constant λ , we write the action as $S = S_0 + \lambda \Delta S$, where S_0 is the free theory and ΔS the interaction term. Simplifying the notation, we have

$$\begin{aligned} G^{(n)} &= \int \mathcal{D}\phi \phi_1 \dots \phi_n \sum_{p=0}^{\infty} \frac{(-\Delta S)^p \lambda^p}{p!} e^{-S_0} \stackrel{?}{=} \sum_{p=0}^{\infty} \lambda^p \int \mathcal{D}\phi \phi_1 \dots \phi_n \frac{(-\Delta S)^p}{p!} e^{-S_0} \\ &= \sum_{p=0}^{\infty} \lambda^p G_p^{(n)}. \end{aligned} \quad (10.1.4)$$

What we typically compute in perturbation theory are the correction terms $G_p^{(n)}$ to the exact Green functions $G^{(n)}$. But since the power series in λ is never uniformly convergent, having zero radius of convergence, we are *not* allowed to exchange the order of sum and integration in the second identity above. Despite we know this is an improper step, this is often the only thing we can do to make concrete computations and is indeed what we have tacitly done all the times so far. The divergent power series appearing in the second row of eq.(10.1.4) is generally an asymptotic series. For sufficiently small values of the coupling such series reproduces quite accurately the exact result $G^{(n)}$ and this explains why perturbation theory in QFT is useful, despite being divergent.

10.1.1 Asymptotic Series and Optimal Truncation*

A series expansion associated to a function $Z(\lambda)$ is asymptotic if, at any fixed order N ,

$$Z(\lambda) - \sum_{n=0}^{N-1} Z_n \lambda^n = \mathcal{O}(\lambda^N), \quad \text{as } \lambda \rightarrow 0. \quad (10.1.5)$$

Notice the crucial difference with respect to convergent series where for $N \rightarrow \infty$ the sum $\sum_{n=0}^N Z_n \lambda^n$ approaches $Z(\lambda)$ for *any* λ within the domain of convergence. For convergent series different functions lead to different series. This is not the case for asymptotic expansions, where different functions can have the same asymptotic expansion. Indeed, if $Z(\lambda)$ has the asymptotic expansion (10.1.5), any other function of the form, say,

$$\tilde{Z}(\lambda) = Z(\lambda) + e^{-1/\lambda} a(\lambda), \quad (10.1.6)$$

with $a(\lambda)$ sufficiently regular, will have exactly the same expansion as $Z(\lambda)$. Asymptotic series are anyhow useful because they approximate the true result with an accuracy that depends on the value of the coupling λ and on the behaviour of the series coefficients Z_n for $n \gg 1$. Contrary to convergent series, where the more terms are added in the series

and the more accurate is the result, in asymptotic series there is an optimal number of terms one should keep, after which adding more terms results in worse and worse accuracy. This is called optimal truncation. Suppose that for $n \gg 1$

$$Z_n \sim n! a^n, \quad (10.1.7)$$

for some real parameter a .⁴ The best accuracy for $Z(\lambda)$ is obtained by finding the value $N = N_{\text{Best}}$ that minimizes the error, estimated as the value of the last term not included in the expansion (10.1.5):

$$\Delta_Z \sim Z_N \lambda^N. \quad (10.1.8)$$

Recalling Stirling approximation

$$n! = \sqrt{2\pi n} n^n e^{-n} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right), \quad (10.1.9)$$

one has

$$Z_N \lambda^N \sim e^{f(N)}, \quad f(N) = -N + N \log(a\lambda N) + \frac{1}{2} \log N. \quad (10.1.10)$$

Since by assumption $N \gg 1$, the minimal error will be given by the minimum of the function $f(N)$:

$$\partial_N f(N) = \log(a\lambda N) + \frac{1}{2N} = 0 \implies N_{\text{Best}} = \frac{1}{a\lambda} \left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right). \quad (10.1.11)$$

Plugging back in eq.(10.1.8) we find at leading order

$$\Delta_Z \sim e^{-\frac{1}{a\lambda}}. \quad (10.1.12)$$

We see that the smaller the coupling is, the smaller (exponentially) the error is, but no matter how many terms we compute in perturbation theory, asymptotic series fail to reproduce the exact function by (at best) exponentially suppressed terms. This is consistent with the intrinsic ambiguity related to asymptotic series shown in eq.(10.1.6). Keeping more than N_{Best} terms in the asymptotic series would lead to an increase in the error.

It has been shown that generally the coefficients of the perturbative expansion in QFT behave at parametrically large order as in eq.(10.1.7), due to the exponentially growing number of Feynman diagrams as the number of loops increases.

⁴The analysis that follows can be generalized for large-order behaviours of the kind $Z_n \sim (n!)^k a^n$.

10.1.2 Borel Summation*

Borel summation is a summation method for asymptotic series. Suppose that a function $Z(\lambda)$ admits an asymptotic expansion of the form (10.1.5) with coefficients Z_n that at large order goes like in eq.(10.1.7). We define the Borel transform as the function obtained by dividing the original series by a factorially growing factor:

$$\mathcal{B}Z(t) \equiv \sum_{n=0}^{\infty} \frac{Z_n}{n!} t^n. \quad (10.1.13)$$

Thanks to the division by $n!$, the series in eq.(10.1.13) has a non-zero radius of convergence where it defines an analytic function $\mathcal{B}Z(t)$. If the analytic continuation of $\mathcal{B}Z(t)$ over the complex t plane is free of singularities for $t > 0$, the integral

$$Z_B(\lambda) = \int_0^{\infty} dt e^{-t} \mathcal{B}Z(t\lambda) \quad (10.1.14)$$

defines a function of λ that is said to be the Borel resum of the original asymptotic series. Recalling the definition of Gamma function

$$\Gamma(x) = \int_0^{\infty} dt e^{-t} t^{x-1}, \quad (10.1.15)$$

it is immediate to verify that $Z_B(\lambda)$ has the same asymptotic expansion as $Z(\lambda)$. On the other hand, it is also clear that if we approximate the series defining $\mathcal{B}Z(t)$ in eq.(10.1.13) with its truncated version, $Z_B(\lambda)$ boils down to the original asymptotic expansion of $Z(\lambda)$ and no progress is achieved. The Borel resummation method works when we can resum the whole series (10.1.13) or estimate the function $\mathcal{B}Z(t)$ by some other means. In general the Borel resummed function $Z_B(\lambda)$ is *not* guaranteed to coincide with the original function $Z(\lambda)$. Indeed, if different functions can admit the same asymptotic series, it is clear that manipulating the latter cannot be enough to uniquely fix $Z(\lambda)$. The uniqueness (and hence the condition $Z_B(\lambda) = Z(\lambda)$) is ensured by a theorem provided that certain analyticity properties of $Z(\lambda)$ near the origin are assumed. When $\mathcal{B}Z(t)$ is free of singularities and gives rise to a well-defined $Z_B(\lambda)$ we say that the associated asymptotic series is Borel resummable. When $Z_B(\lambda) = Z(\lambda)$ we say that the asymptotic series is Borel resummable to the exact result.

We can get some intuition on Borel functions by working out the Borel transform of an asymptotic series with coefficients as in eq.(10.1.7). The Borel series in this case collapses to a simple power series and gives

$$\mathcal{B}Z(t) = \frac{1}{1 - at}, \quad (10.1.16)$$

that has a simple pole at $t = 1/a$. The radius of convergence of the Borel series is given by $R = 1/|a|$, but the function can be analytically continued over the whole t -plane. Borel summability depends on the sign of a . If $a > 0$ (same sign series) the singularity is on the positive real t axis, the integral (10.1.14) is divergent and the series is not Borel resummable. If $a < 0$ (alternating series), the singularity is over the negative real t axis, the integral (10.1.14) is finite and the Borel resummed series is given by $Z_B(\lambda)$. More in general, eq.(10.1.7) gives only the asymptotic form of the coefficients of the series, while the precise form of the latter might be unavailable. When the exact Borel function $\mathcal{B}Z(t)$ is not known, some information on its analytic structure can still be deduced, because the large order behaviour of the asymptotic series determines the position of the singularity closest to the origin. If $a < 0$ the series is not guaranteed to be Borel resummable, because further singularities on the positive real axis might occur, depending on the next to leading large order behaviour of the series coefficients. On the other hand, for $a > 0$ the series is certainly not Borel resummable.

The order of sum and integration in the Borel function cannot be inverted because the Borel series expansion has generally a finite radius of convergence while the integral is taken over the whole positive t axis. Indeed, if we erroneously interchange the two operations, we get back the asymptotic divergent series we started with:

$$Z_B(\lambda) = \int_0^\infty dt e^{-t} \mathcal{B}Z(\lambda t) \neq \sum_{n=0}^\infty \frac{Z_n}{n!} \lambda^n \int_0^\infty dt t^n e^{-t} = \sum_{n=0}^\infty Z_n \lambda^n. \quad (10.1.17)$$

Unfortunately the asymptotic series in the coupling in four dimensional gauge theories are generally not Borel resummable. Examples of Borel resummable series include the $\lambda\phi^4$ theories in two and three space-time dimensions, important in the study of critical phenomena in statistical physics.

10.2 Euler-Heisenberg Effective Action and Vacuum Decay*

The far IR behaviour of a $U(1)$ gauge theory with massive charged particles is obtained by integrating out all the matter fields. By gauge invariance the effective action for the photon field is expected to be a (complicated) functional of the field strength $F_{\mu\nu}$ and its derivatives. Notably all the higher dimensional operators that do not involve derivatives of the field strength can be exactly resummed. This remarkable result was found by Euler and Heisenberg in 1936 [32] and still represents nowadays one of the few effective theories where an exact resummation of an infinite series of operators is doable. The Euler-Heisenberg effective action allows us to describe the leading interactions in light-light scattering. If we consider $F_{\mu\nu}$ as an external field, it can be interpreted as the exact quantum vacuum

energy in presence of a constant electric and magnetic field. In this case we will be able to compute the asymptotic nature of the coupling constant expansion. Interestingly, we will see that in presence of an electric field the ground state energy develops an imaginary part that leads to its decay through particle production (electron-positron pairs in QED).⁵ However, before discussing in detail all these issues, we have to develop a formalism that will allow us to straightforwardly compute the Euler-Heisenberg action. In order to do that, it is useful to consider first a seemingly unrelated problem, the derivation of the Landau level in quantum mechanics using a path integral formulation.

10.2.1 Landau Levels by Path Integral*

Consider a charged particle of mass M in the presence of a magnetic field. Its classical Lagrangian is

$$L = \frac{M}{2} \dot{\vec{q}}^2 + e \dot{\vec{q}} \cdot \vec{A}(\vec{q}), \quad (10.2.1)$$

where $\vec{q}(t)$ is the particle coordinate and $\dot{\vec{q}} = d\vec{q}/dt$. The corresponding Hamiltonian H is given by

$$H \equiv \vec{p} \cdot \dot{\vec{q}} - L = \frac{(\vec{p} - e\vec{A}(\vec{q}))^2}{2M}, \quad (10.2.2)$$

where H has been expressed in terms of the momentum $\vec{p} \equiv \delta L / \delta \dot{\vec{q}} = M\dot{\vec{q}} + e\vec{A}(\vec{q})$.

A classical problem in quantum mechanics is the determination of the energy levels of a quantum particle when it is confined in the x - y -plane while being subject to a constant magnetic field $\vec{B} = B\vec{e}_z$ (where $\vec{e}_i$ is the unit vector along direction i). Given that $\vec{B} = \vec{\nabla} \times \vec{A}$, one can choose $\vec{A} = Bx\vec{e}_y$, such that eq. (10.2.1) in terms of operators becomes

$$\hat{H} = \frac{\hat{p}_x^2}{2M} + \frac{(\hat{p}_y - eB\hat{x})^2}{2M}. \quad (10.2.3)$$

Translation invariance of $\hat{H}$ along $\vec{e}_y$ implies $[\hat{H}, \hat{p}_y] = 0$ and therefore one can look for the eigenstates of $\hat{H}$ in the form of plane waves with definite momentum $p_y = \hbar k_y$, for which

$$\hat{H} = \frac{\hat{p}_x^2}{2M} + \frac{(\hbar k_y - eB\hat{x})^2}{2M} = \frac{\hat{p}_x^2}{2M} + \frac{M\omega_c^2}{2}(\hat{x} - x_0)^2, \quad (10.2.4)$$

where

$$\omega_c = \frac{eB}{M} \quad (10.2.5)$$

is the cyclotron frequency and the harmonic oscillator is centered around $x_0 = \hbar k_y / (eB)$. For a given k_y , the energy levels of this harmonic oscillator are given by $E_n = \hbar\omega_c(n + 1/2)$ with $n = 0, 1, \dots$ (Landau levels). Each of these levels has a degeneracy g determined by the corresponding

⁵The vacuum decay process has been discussed in the context of QED later on by Schwinger [53] and for this reason it is sometimes denoted Schwinger effect or Schwinger pair production. However it should be pointed out that an estimate of the pair production rate was already present in ref.[32].

possible values of k_y . In order to determine g consider a two-dimensional rectangular area with edges of length L_x and L_y . Assuming periodic boundary conditions, one finds that $k_y = 2\pi n_y/L_y$, with $n_y = 0, 1, \dots$, each corresponding to an harmonic oscillator centered at $x_0 = \hbar 2\pi n_y/(eBL_y)$. Requiring that x_0 is within the rectangle $0 \leq x_0 \leq L_x$, one finds that $n_y \leq eBL_x L_y/(2\pi\hbar)$ and therefore $g = eBA/(2\pi\hbar)$, where $A = L_x L_y$ is the area of the surface across which $\vec{B}$ flows. The fact that g is an integer implies a magnetic flux quantization in multiple integers of the quantum of flux $2\pi\hbar/e$. Below we recover this result via a path-integral approach.

Consider a gas of non-interacting particles with Hamiltonian given by eq. (10.2.2) and at temperature β^{-1} . The corresponding single-particle partition function is given by

$$Z(\beta) = \text{Tr} e^{-\beta H} = \int d\vec{q}_a \langle \vec{q}_a | e^{-\beta H} | \vec{q}_a \rangle. \quad (10.2.6)$$

The matrix element $\langle \vec{q}_a | e^{-\beta H} | \vec{q}_a \rangle$ can be obtained via a Wick's rotation of the standard propagator of quantum mechanics $\langle \vec{q}_b | e^{-iH(t_b-t_a)} | \vec{q}_a \rangle$ which quantifies the probability amplitude for a particle with Hamiltonian H to propagate from the position $\vec{q}_a$ at time t_a to the position $\vec{q}_b$ at time t_b . This propagator can be represented as a Feynman path-integral with an exponential weight determined by the action associated to Lagrangian L in eq. (10.2.1):

$$\langle \vec{q}_b | e^{-iH(t_b-t_a)} | \vec{q}_a \rangle = \mathcal{N} \int_{\substack{\vec{q}(t_a) = \vec{q}_a \\ \vec{q}(t_b) = \vec{q}_b}} \mathcal{D}\vec{q}(t) e^{i \int_{t_a}^{t_b} dt L}. \quad (10.2.7)$$

By setting $t_a = 0$ and by continuing $t_b \rightarrow -i\tau$, with $\dot{\vec{q}} \rightarrow id\vec{q}/d\tau$ one gets

$$Z = \text{Tr} e^{-\beta H} = \mathcal{N} \int d\vec{q}_a \int_{\vec{q}(\beta) = \vec{q}(0) = \vec{q}_a} \mathcal{D}\vec{q}(\tau) e^{-\int_0^\beta d\tau L_E}, \quad (10.2.8)$$

where the Euclidean Lagrangian L_E is given by

$$L_E = -L|_{\dot{\vec{q}} \rightarrow id\vec{q}/d\tau} = \frac{M}{2} \left(\frac{d\vec{q}}{d\tau} \right)^2 - ie \frac{d\vec{q}}{d\tau} \cdot \vec{A}(\vec{q}(\tau)). \quad (10.2.9)$$

The path integral in eq. (10.2.8) is calculated by summing over all closed paths originating in a certain point $\vec{q}_a = \vec{q}(0) = \vec{q}(\beta)$, over which one eventually integrates. These paths are in a one-to-one correspondence with periodic functions of period β and therefore can be expressed in terms of their Fourier transforms:

$$\vec{q}(\tau) = \vec{q}_0 + \sum_{n=1}^{\infty} \sqrt{\frac{2}{\beta}} \left(\vec{c}_n \cos \frac{2\pi n\tau}{\beta} + \vec{s}_n \sin \frac{2\pi n\tau}{\beta} \right), \quad (10.2.10)$$

where $\vec{s}_n = (s_{x,n}, s_{y,n}, s_{z,n})$ and $\vec{c}_n = (c_{x,n}, c_{y,n}, c_{z,n})$ are the vector coefficients of the transform which can be obtained by inverting it, i.e.,

$$\vec{q}_0 = \frac{1}{\beta} \int_0^\beta d\tau \vec{q}(\tau), \quad \left. \begin{matrix} \vec{c}_n \\ \vec{s}_n \end{matrix} \right\} = \int_0^\beta d\tau \vec{q}(\tau) \sqrt{\frac{2}{\beta}} \begin{Bmatrix} \cos(2\pi n\tau/\beta) \\ \sin(2\pi n\tau/\beta) \end{Bmatrix}. \quad (10.2.11)$$

Note that because the Fourier transform is a unitary transformation, one has

$$\int_{\vec{q}(0)=\vec{q}(\beta)} \mathcal{D}\vec{q}(\tau) = \int_{-\infty}^{+\infty} d\vec{q}_0 \prod_{n=1}^{\infty} d\vec{c}_n d\vec{s}_n. \quad (10.2.12)$$

Assume now that this gas of particles is constrained to move on the x - y -plane with $q_z(\tau) = 0$ (and therefore $c_{z,n} = s_{z,n} = 0$) subject to a constant magnetic field $\vec{B} = B\vec{e}_z$, as in the case of the Landau levels discussed above. As before, we can choose $\vec{A} = Bx\vec{e}_y$ and, by representing the trajectories as in eq. (10.2.10), one can easily express the (Euclidean) action $S_E = \int_0^\beta d\tau L_E$ associated with the (Euclidean) Lagrangian in eq. (10.2.8) in terms of the coefficients $\vec{c}_n$ and $\vec{s}_n$. In particular, it is useful to note that

$$\int_0^\beta d\tau \frac{d\vec{q}}{d\tau} \cdot \vec{A}(\vec{q}) = B \int_0^\beta d\tau \frac{dq_y}{d\tau} q_x = B \sum_{n=1}^{\infty} \frac{2\pi n}{\beta} (-c_{x,n} s_{y,n} + s_{x,n} c_{y,n}), \quad (10.2.13)$$

$$\int_0^\beta d\tau \left(\frac{d\vec{q}}{d\tau} \right)^2 = \sum_{n=1}^{\infty} \left(\frac{2\pi n}{\beta} \right)^2 (c_{x,n}^2 + c_{y,n}^2 + s_{x,n}^2 + s_{y,n}^2), \quad (10.2.14)$$

and therefore

$$\begin{aligned} S_E &= \sum_{n=1}^{\infty} \left\{ \frac{M}{2} \left(\frac{2\pi n}{\beta} \right)^2 (c_{x,n}^2 + c_{y,n}^2 + s_{x,n}^2 + s_{y,n}^2) - ieB \frac{2\pi n}{\beta} (s_{x,n} c_{y,n} - c_{x,n} s_{y,n}) \right\} \\ &= \sum_{n=1}^{\infty} \left\{ \frac{M}{2} \left(\frac{2\pi n}{\beta} \right)^2 (c_{y,n}^2 + s_{x,n}^2) - ieB \frac{2\pi n}{\beta} s_{x,n} c_{y,n} \right\} \\ &\quad + \sum_{n=1}^{\infty} \left\{ \frac{M}{2} \left(\frac{2\pi n}{\beta} \right)^2 (c_{x,n}^2 + s_{y,n}^2) + ieB \frac{2\pi n}{\beta} c_{x,n} s_{y,n} \right\} \\ &= \sum_{n=1}^{\infty} \frac{1}{2} (c_{y,n}, s_{x,n}) A_n^* \begin{pmatrix} c_{y,n} \\ s_{x,n} \end{pmatrix} + \sum_{n=1}^{\infty} \frac{1}{2} (c_{x,n}, s_{y,n}) A_n \begin{pmatrix} c_{x,n} \\ s_{y,n} \end{pmatrix}, \end{aligned} \quad (10.2.15)$$

where

$$A_n \equiv \begin{pmatrix} M(2\pi n/\beta)^2 & ieB(2\pi n/\beta) \\ ieB(2\pi n/\beta) & M(2\pi n/\beta)^2 \end{pmatrix}. \quad (10.2.16)$$

As expected, S_E is independent of $\vec{q}_0$. Accordingly, eq. (10.2.8) becomes

$$\begin{aligned} Z &= \mathcal{N} \int_{-\infty}^{+\infty} dq_{x,0} dq_{y,0} \prod_{n=1}^{\infty} dc_{x,n} dc_{y,n} ds_{x,n} ds_{y,n} e^{-S_E} \\ &= \mathcal{N} \mathcal{A} \prod_{n=1}^{\infty} \frac{\pi}{\sqrt{\det A_n}} \frac{\pi}{\sqrt{\det A_n^*}} = \mathcal{N}' \mathcal{A} \prod_{n=1}^{\infty} \left[1 + \left(\frac{\beta \omega_c}{2\pi n} \right)^2 \right]^{-1}, \end{aligned} \quad (10.2.17)$$

where ω_c is defined in eq.(10.2.5), we used the fact that the integral over $\vec{q}_0$ renders the area $\mathcal{A}$, and

$$\det M_n = M^2 \left(\frac{2\pi n}{\beta} \right)^4 \left[1 + \left(\frac{\beta\omega_c}{2\pi n} \right)^2 \right]. \quad (10.2.18)$$

All the factors which do not depend on B have been reabsorbed in the overall normalization $\mathcal{N}'$ of the path integral. In turn, this constant can be easily calculated by considering the same problem for $B = 0$, i.e., by calculating the partition function $Z_0(\beta)$ of a free quantum particle with Hamiltonian $H_0 = \vec{p}^2/(2M)$ which is confined in a d -dimensional space with volume $\mathcal{A}$:

$$Z_0(\beta) = \int \frac{d\vec{q}d\vec{p}}{(2\pi\hbar)^d} e^{-\beta H_0} = \mathcal{A} \left(\frac{M}{2\pi\hbar\beta} \right)^{d/2}. \quad (10.2.19)$$

By requiring that Z in eq. (10.2.5) reduces to Z_0 for $\omega_c = 0$ (with $d = 2$), one fixes the value of $\mathcal{N}'$ and finds that

$$Z(\beta) = \mathcal{A} \frac{M}{2\pi\hbar\beta} \frac{\beta\omega_c/2}{\sinh(\beta\omega_c/2)} = \frac{\mathcal{A}eB}{2\pi\hbar} \sum_{n=0}^{\infty} e^{-\beta\omega_c(n+1/2)}, \quad (10.2.20)$$

where we used the identity

$$\prod_{n=1}^{\infty} \left(1 + \frac{x^2}{\pi^2 n^2} \right) = \frac{\sinh x}{x}, \quad (10.2.21)$$

and the expansion of $\sinh$ in terms of exponentials. Recalling the definition of $Z(\beta)$ in eq. (10.2.6) one can alternatively calculate it in the basis of the eigenstate of the energy H (with eigenvalues E_n and degeneracy g_n), finding that $Z(\beta) = \sum_n g_n e^{-\beta E_n}$. By comparing this general expansion with eq. (10.2.20) it is possible to identify $E_n = \hbar\omega_c(n + 1/2)$ and $g_n = \mathcal{A}eB/(2\pi\hbar)$, which are readily recognized to be the Landau levels that we discussed before. Note that, due to the fact that g_n has to be an integer, one concludes that the flux $\mathcal{A}B$ of a magnetic field through a certain surface of area $\mathcal{A}$ has to be a multiple integer of the elementary flux $2\pi/e$. This is nothing but the Dirac quantization condition for a flux of B through a (large) close surface enclosing a magnetic monopole.

10.2.2 Effective Action for Scalar QED*

We consider first the case of a scalar field, i.e., the scalar QED with Lagrangian

$$L = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + |D_\mu\phi|^2 - m^2|\phi|^2, \quad (10.2.22)$$

where $D_\mu = \partial_\mu - ieA_\mu$ is the covariant derivative and the field A_μ is assumed to be assigned from the outset. We neglect the self-interaction of the scalar field, assuming it

provides a negligible correction to the effective action. At energy scales much smaller than the scalar mass m we can integrate out the scalar field exactly, since it enters the action quadratically. In the language of chapter 7, the photon and the charged scalar represent the light and the heavy fields ϕ_l and ϕ_h , respectively, while the Euler-Heisenberg action $S_{\text{EH}}^{(S)}$ represents part of the IR effective action S_{IR} defined in eq.(7.1.1):

$$iS_{\text{IR}}(A) = -\frac{i}{4} \int d^4x F_{\mu\nu} F^{\mu\nu} + i\Delta S_{\text{IR}}(A), \quad (10.2.23)$$

where

$$i\Delta S_{\text{IR}}(A) \equiv \ln \left\{ \mathcal{N} \int \mathcal{D}\phi \mathcal{D}\phi^* e^{i \int d^4x \phi^* (-D^2 - m^2) \phi} \right\}. \quad (10.2.24)$$

It is convenient to perform a Wick's rotation in the Gaussian path integral. In order to keep the form of the covariant derivative unchanged, we rotate $x_0 = -ix_4$ and define $A_0 = iA_{E,4}$, so that $D_0 = \partial_0 - ieA_0 = i(\partial_4 - ieA_{E,4}) = iD_4$. We then have

$$i \int d^4x \phi^* (-D^2 - m^2) \phi = - \int d^4x_E \phi^* (-D_E^2 + m^2) \phi, \quad (10.2.25)$$

where $D_{E,\mu} = \partial_\mu - ieA_{E,\mu}$ with $\mu = 1, 2, 3, 4$, $A_{E,i} = A_i$ ($i = 1, 2, 3$) and $A_{E,4} = -iA_0$. The Gaussian integral can be easily calculated, leading to

$$i\Delta S_{\text{IR}}(A) = -\text{Tr} \ln \frac{-D_E^2 + m^2}{\mu^2} + \dots, \quad (10.2.26)$$

where henceforth we neglect irrelevant constants and we introduce a convenient momentum scale μ . We can get rid of the log appearing in eq.(10.2.26) by using the relation

$$\ln \frac{a}{b} = - \int_0^\infty \frac{d\beta}{\beta} (e^{-a\beta} - e^{-b\beta}), \quad (10.2.27)$$

which allows us to rewrite eq.(10.2.26) as follows:

$$i\Delta S_{\text{IR}}(A) = \int_0^\infty \frac{d\beta}{\beta} e^{-\beta m^2} \text{Tr} e^{-\beta(-D_E^2)} + \text{const.}, \quad (10.2.28)$$

where the constant term is independent of A and will be neglected in what follows. The connection with the problem discussed in Sec. 10.2.1 becomes clear by noticing that

$$-D_E^2 = (-i\partial_\mu - eA_{E,\mu})^2 \quad (10.2.29)$$

formally coincides with the Hamiltonian reported in eq. (10.2.2) of a quantum particle moving in 4 spatial dimensions, with momentum $p_\mu = -i\partial_\mu$, mass $M = 1/2$, in the presence of an external field $A_{E,\mu}$. Accordingly, $\text{Tr} e^{-\beta(-D_E^2)}$ can be identified with the partition function $Z(\beta)$ of this particle, which can be expressed as a path integral like in

eq. (10.2.8). Similarly to the study of the Landau levels in sec. 10.2.1, the computation can be performed in closed form if we consider a linear vector potential A_E , corresponding to a constant field strength $F_{E,\mu\nu} = \partial_\mu A_{E,\nu} - \partial_\nu A_{E,\mu}$. A possible choice of A_E is

$$A_{E,\mu}(\vec{q}) = \frac{1}{2} F_E^{\mu\nu} q_\nu, \quad (10.2.30)$$

such that the Euclidean Lagrangian of this particle becomes

$$L_E = \frac{M}{2} \left(\frac{dq_\mu}{d\tau} \right)^2 - \frac{ie}{2} \frac{dq_\mu}{d\tau} F_E^{\mu\nu} q_\nu, \quad (10.2.31)$$

The Lagrangian (10.2.31) is a straightforward generalization of the Lagrangian (10.2.9) considered before. We can then compute the partition function by decomposing q_μ as in eq. (10.2.10) and by expressing the corresponding Euclidean action S_E in terms of the coefficients $s_{\mu,n}$ and $c_{\mu,n}$, taking into account the obvious generalizations of eqs. (10.2.13) and (10.2.14):

$$\int_0^\beta d\tau \frac{dq_\mu}{d\tau} q_\nu(\tau) = \sum_{n=1}^\infty \frac{2\pi n}{\beta} (-c_{\mu,n} s_{\nu,n} + s_{\mu,n} c_{\nu,n}), \quad (10.2.32)$$

$$\int_0^\beta d\tau \left(\frac{dq_\mu}{d\tau} \right)^2 = \sum_{n=1}^\infty \left(\frac{2\pi n}{\beta} \right)^2 (c_{\mu,n}^2 + s_{\mu,n}^2). \quad (10.2.33)$$

As the term on the l.h.s. of the first equation is contracted in S_E with the antisymmetric matrix F_E , the two terms on the r.h.s. eventually give the same contribution and therefore

$$S_E = \sum_{n=1}^\infty \left\{ \frac{M}{2} \left(\frac{2\pi n}{\beta} \right)^2 \sum_{\mu=1}^4 (c_{\mu,n}^2 + s_{\mu,n}^2) - ie \left(\frac{2\pi n}{\beta} \right) \sum_{\mu,\nu=1}^4 s_{\mu,n} F_{E,\mu\nu} c_{\nu,n} \right\}. \quad (10.2.34)$$

Accordingly, the partition function can be written as

$$Z = \mathcal{N} \int \prod_{\mu=1}^4 \left(dq_{\mu,0} \prod_{n=1}^\infty dc_{\mu,n} ds_{\mu,n} \right) e^{-S_E} \quad (10.2.35)$$

where the integral over $q_{\mu,0}$ renders the Euclidean space-time volume iTV , while the integral over each single $s_{\mu,n}$ is a Gaussian integral in the presence of a linear term of the form

$$\int ds_{\mu,n} e^{-\alpha s_{\mu,n}^2 + i\gamma s_{\mu,n}} = \sqrt{\frac{\pi}{\alpha}} e^{-\gamma^2/(4\alpha)}, \quad (10.2.36)$$

with $\alpha = (\pi n/\beta)^2$ (with $M = 1/2$), and $\gamma = e(2\pi n/\beta) F_{E,\nu\mu} c_{\mu,n}$. After this first integration, one finds (up to irrelevant constants which are absorbed in the definition of $\mathcal{N}'$ and

then $\mathcal{N}''')$

$$\begin{aligned} Z &= \mathcal{N}' iTV \int \prod_{n=1}^{\infty} \left(\prod_{\mu=1}^4 dc_{\mu,n} \exp \left\{ -(\pi n/\beta)^2 c_{\mu,n} M_{\mu\nu} c_{\nu,n} \right\} \right) \\ &= \mathcal{N}'' iTV \prod_{n=1}^{\infty} (\det M_n)^{-1/2} \end{aligned} \quad (10.2.37)$$

where we introduced the 4×4 matrix

$$M_{n,\mu\nu} \equiv \delta_{\mu\nu} - \left(\frac{e\beta}{\pi n} \right)^2 (F_E^2)_{\mu\nu}. \quad (10.2.38)$$

As in the case discussed in Sec. 10.2.1, the normalization constant $\mathcal{N}'''$ can be fixed by comparing this result with the one for $e = 0$, which reduces to the partition function of a free particle (with mass $M = 1/2$) calculated in eq. (10.2.19) for generic dimension d , with an area $\mathcal{A}$ given by the Euclidean volume iVT . Accordingly,

$$Z(\beta) = iTV (4\pi\beta)^{-2} \prod_{n=1}^{\infty} (\det M_n)^{-1/2}. \quad (10.2.39)$$

It is useful now to express the components of F_E in terms of the fields $\vec{E}$ and $\vec{B}$. Denoting by F the usual Lorentzian field strength, we have $F_{0j} = E_j$ and $F_{jk} = \epsilon_{jkl} B_l$, i, j, k being spatial indices running from 1 to 3. Accordingly, the components of the Euclidean tensor are given by $F_{E,4j} = -iF_{0j} = -iE_j$ and $F_{E,jk} = F_{jk} = \epsilon_{jkl} B_l$. Using rotations, we can choose a reference frame where the electric field $\vec{E} = E\vec{e}_x$ is directed along the x -axis and the magnetic field $\vec{B} = B_x\vec{e}_x + B_y\vec{e}_y$ is on the x - y -plane, which gives

$$F_E = \begin{pmatrix} 0 & 0 & -B_y & iE \\ 0 & 0 & B_x & 0 \\ B_y & -B_x & 0 & 0 \\ -iE & 0 & 0 & 0 \end{pmatrix}, \quad (10.2.40)$$

and

$$F_E^2 = \begin{pmatrix} -B_y^2 + E^2 & B_x B_y & 0 & 0 \\ B_x B_y & -B_x^2 & 0 & 0 \\ 0 & 0 & -B_y^2 - B_x^2 & iB_y E \\ 0 & 0 & iB_y E & E^2 \end{pmatrix}. \quad (10.2.41)$$

Due to the block-diagonal structure of F_E^2 , the 4 eigenvalues of this matrix are given by the 2 eigenvalues of the upper left 2×2 matrix and by those of the lower right 2×2 matrix.

However it is easy to see that these two 2×2 matrices have the same eigenvalues, which therefore have multiplicity two and which are solutions of the equation

$$\lambda^2 + (\vec{B}^2 - \vec{E}^2)\lambda - (\vec{B} \cdot \vec{E})^2 = 0, \quad (10.2.42)$$

i.e.,

$$\lambda_{\pm} = \frac{1}{2} \left[\vec{E}^2 - \vec{B}^2 \pm \sqrt{(\vec{B}^2 - \vec{E}^2)^2 + 4(\vec{B} \cdot \vec{E})^2} \right] \equiv \pm a_{\pm}^2. \quad (10.2.43)$$

Accordingly

$$\det M_n = \left(1 - \frac{e^2 \beta^2}{\pi^2 n^2} a_+^2 \right)^2 \left(1 + \frac{e^2 \beta^2}{\pi^2 n^2} a_-^2 \right)^2, \quad (10.2.44)$$

and Z in eq. (10.2.39) can be calculated by using eq. (10.2.21):

$$Z(\beta) = \text{Tr} e^{-\beta(-D_E^2)} = \frac{iTV}{(4\pi\beta)^2} \frac{e\beta a_+}{\sin(e\beta a_+)} \frac{e\beta a_-}{\sinh(e\beta a_-)}. \quad (10.2.45)$$

Modulo a constant factor, we then get the Euler-Heisenberg effective Lagrangian⁶

$$\mathcal{L}_{\text{EH}}^{(S)} = -\frac{1}{4} Z_A^{(S)} F_{\mu\nu} F^{\mu\nu} + \int_0^\infty \frac{d\beta}{\beta} e^{-\beta m^2} (4\pi\beta)^{-2} \frac{e\beta a_+}{\sin(e\beta a_+)} \frac{e\beta a_-}{\sinh(e\beta a_-)}, \quad (10.2.46)$$

where, according to eq.(10.2.23), we have reinserted the tree-level contribution, the first term in eq.(10.2.46). The above relation is expressed in terms of renormalized quantities. Thanks to the QED Ward identities the combination $ea_{\pm}$ is the same for bare and renormalized quantities, while $A_B = \sqrt{Z_A^{(S)}} A$, so we get an explicit counterterm dependence in eq.(10.2.46) only from the tree-level kinetic term. Note that the combinations $a_+^2 - a_-^2$ and $a_+ a_-$ can be more conveniently written as

$$\begin{aligned} a_+^2 - a_-^2 &= \vec{E}^2 - \vec{B}^2 = -\frac{1}{2} F_{\mu\nu} F^{\mu\nu}, \\ a_+ a_- &= 2\vec{E} \cdot \vec{B} = -\frac{1}{2} F_{\mu\nu} \tilde{F}^{\mu\nu}, \end{aligned} \quad (10.2.47)$$

where

$$\tilde{F}^{\mu\nu} \equiv \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma} \quad (10.2.48)$$

is the dual field strength and $\epsilon^{\mu\nu\rho\sigma}$ is the completely antisymmetric four-tensor with $\epsilon^{0123} = -\epsilon_{0123} = +1$. The integral in β above is not straightforward and in fact is not always well defined. However, by expanding the sin and sinh factors in eq.(10.2.46) and using eq.(10.2.47) the term by term integral can be done and allows us to determine all higher dimensional operators that do not involve derivatives of the field strengths.

⁶Strictly speaking, Euler and Heisenberg considered fermion QED, but the idea is the same. The fermion QED case will be discussed in section 10.2.3.

We can interpret the above computation as a resummation of an infinite class of one-loop diagrams in a specific background field gauge where $F_{\mu\nu}$ is constant. The zero temperature limit $\beta \rightarrow \infty$ of the partition function is determined by the vacuum of the theory, an IR property, while the infinite temperature limit $\beta \rightarrow 0$ is sensitive to all frequency modes, so the two limits correspond to the IR and UV of the theory, respectively. Thanks to the mass term, no IR divergencies can arise. Since scalar QED is a renormalizable theory, we expect UV divergences only for the first two terms in an expansion in $a_{\pm}$ of the integrand in eq.(10.2.46). These correspond to the renormalization of the constant and of the term proportional to $F_{\mu\nu}F^{\mu\nu}$. Indeed, we have

$$Z(\beta \rightarrow 0) = \frac{iTV}{(4\pi\beta)^2} \left(1 - \frac{e^2\beta^2}{12} F_{\mu\nu}F^{\mu\nu} + \text{finite} \right), \quad (10.2.49)$$

where the finite terms start at $\mathcal{O}(\beta^4)$. The leading divergence $\sim \beta^{-2}$ is field independent and can be fixed by appropriately adjusting the constant term we have neglected so far. The logarithmic divergence $\sim \beta^0$ corresponds instead to the one-loop contribution of a scalar field to the wave function renormalization of the photon. We can regularize this divergence using dimensional regularization. Although we need a prescription to generalize in arbitrary d dimensions the matrix (10.2.38), a universal effect of having $d \neq 4$ is to change the determination of the normalization constant of the path integral. By comparison with the partition function of a free particle in d dimensions, $Z_0(\beta)$ in eq.(10.2.19), we see that the factor $(4\pi\beta)^{-2}$ in eq.(10.2.46) turns into $(4\pi\beta)^{-d/2}$. The divergent term of the Lagrangian reads then

$$\mathcal{L}_{\text{EH}}^{(S)}|_{\text{div}} = \left(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \right) \frac{e^2}{3(4\pi)^{d/2}} \int_0^\infty \frac{d\beta}{\beta} e^{-\beta m^2} \beta^{-d/2} \beta^2. \quad (10.2.50)$$

The UV divergence comes from $\beta \rightarrow 0$, so we can change the upper limit of the integration over β from ∞ to any other finite value. We can also neglect the exponential factor which does not play any role in the limit. It is convenient to set the upper limit of the integral to 1 and use the following identity:

$$\int_0^1 d\beta \beta^\gamma = \frac{1}{\gamma + 1}. \quad (10.2.51)$$

The integral above is valid for $\gamma > -1$, but it can be analytically continued in the complex γ plane for any other value. Using eq.(10.2.51) in the Lagrangian (10.2.50) with $\gamma = 1 - d/2$, we get a pole in $1/\epsilon$ given by

$$\mathcal{L}_{\text{EH}}^{(S)}|_{\text{div}} = \frac{2e^2}{3(4\pi)^2} \frac{1}{\epsilon} \left(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \right). \quad (10.2.52)$$

Note that the divergence (10.2.52) is independent of the way in which we define the matrix $M_{n,\mu\nu}$ in $d = 4 - \epsilon$, because $\det M_n^{(d)} = \det M_n^{(4)} + \mathcal{O}(\epsilon)$. Finite terms will then be renormalization scheme-dependent, but the $1/\epsilon$ pole is universal. Matching with eq.(10.2.46) we get

$$Z_A^{(S)} = 1 - \frac{2e^2}{3(4\pi)^2} \frac{1}{\epsilon} + \delta Z_A^{(S)}, \quad (10.2.53)$$

where $\delta Z_A^{(S)}$ is a finite, scheme-dependent, renormalization quantity. It is straightforward to check that $Z_A^{(S)}$ gives rise to the correct one-loop photon anomalous dimension and β -function, as reported i.e. in eq.(11.4.15). Needless to say, $Z_A^{(S)}$ receives higher order contributions from loops containing virtual photons, as well as virtual scalars, which cannot be captured from the above computation. We can rewrite eq.(10.2.46) as

$$\mathcal{L}_{\text{EH}}^{(S)} = -\frac{1}{4}(1 + \delta Z_A^{(S)})F_{\mu\nu}F^{\mu\nu} + \frac{m^4}{16\pi^2} \int_0^\infty dt e^{-t} \frac{1}{t^3} \left[\frac{x_+ t}{\sin(x_+ t)} \frac{x_- t}{\sinh(x_- t)} - 1 - \frac{t^2}{6}(x_+^2 - x_-^2) \right], \quad (10.2.54)$$

where

$$x_\pm \equiv \frac{ea_\pm}{m^2}. \quad (10.2.55)$$

The Euler-Heisenberg Lagrangian can be interpreted as the exact vacuum energy in presence of an external constant electro-magnetic field. It is then interesting to expand the exact result in powers of the electric charge in order to see which kind of series we get. For simplicity, let us separately consider the purely magnetic ($a_+ = 0, a_- = |\vec{B}| \equiv B$) and purely electric ($a_- = 0, a_+ = |\vec{E}| \equiv E$) cases. The relevant expansions are

$$\frac{z}{\sinh z} = \sum_{k=0}^{\infty} \frac{(2 - 2^{2k})B_{2k}}{(2k)!} z^{2k}, \quad \frac{z}{\sin z} = \sum_{k=0}^{\infty} (-1)^k \frac{(2 - 2^{2k})B_{2k}}{(2k)!} z^{2k}. \quad (10.2.56)$$

The coefficients B_n are the Bernoulli numbers, defined as

$$B_n = \frac{n!}{2i\pi} \oint_{\mathcal{C}_0} \frac{z}{e^z - 1} \frac{dz}{z^{n+1}}, \quad (10.2.57)$$

with $\mathcal{C}_0$ a counterclockwise contour around the origin with radius $r < 2\pi$ (to avoid poles coming from $e^z - 1$). Plugging the expansions (10.2.56) in the Lagrangian (10.2.54) and integrating in t term by term we get in the two cases series expansions of the form

$$\begin{aligned} & \sum_k \left(\frac{eB}{m^2} \right)^{2k} \frac{\Gamma(2k-2)(2 - 2^{2k})B_{2k}}{(2k)!} \quad (\text{magnetic}), \\ & \sum_k \left(\frac{eE}{m^2} \right)^{2k} (-1)^k \frac{\Gamma(2k-2)(2 - 2^{2k})B_{2k}}{(2k)!} \quad (\text{electric}). \end{aligned}$$

For large values of k , the Bernoulli numbers can be approximated as

$$B_{2k} \approx (-1)^{k+1} 4\sqrt{\pi k} \left(\frac{k}{\pi e}\right)^{2k}, \quad k \gg 1. \quad (10.2.58)$$

So both series are divergent asymptotic with a large order behaviour of the form

$$\begin{aligned} \sum_k \left(\frac{eB}{m^2}\right)^{2k} F_{2k}, \quad F_{2k} \sim (2k)!(-1)^k & \quad (\text{magnetic}), \\ \sum_k \left(\frac{eE}{m^2}\right)^{2k} F_{2k}, \quad F_{2k} \sim (2k)! & \quad (\text{electric}), \end{aligned} \quad (10.2.59)$$

where for simplicity we have written only the leading factorial growth. The first series coefficients alternate in sign, the second ones have always the same sign. According to the discussion in section 10.1.2, this is an indication that the first series might be Borel resummable, while the second will not be so. As a matter of fact, we do not need to look at the large order behaviour of the series, because we have the luxury of having the closed form expression of the Borel function. Indeed, the integral in t in eq.(10.2.54) can be interpreted as the Laplace transform needed to recover an observable from its Borel function, namely we can identify the factor in the square bracket, together with the factor $1/t^3$ as the Borel function associated to the Euler-Heisenberg Lagrangian. In the purely magnetic case all the singularities of the function occur in the pure imaginary t -axis, where the $\sinh$ develops simple poles, in the electric case these occur in the real axis, where the $\sin$ has simple poles. We can thus unambiguously establish (non-)Borel summability of the asymptotic series in the (electric) magnetic case.

As we will see in section 10.2.4, the ambiguity associated to the integration over t in the electric case has a profound implication: the Euler-Heisenberg Lagrangian develops an imaginary part that leads to the decay of the vacuum.

10.2.3 Effective Action for Fermion QED**

In this subsection we consider the actual Lagrangian Euler and Heisenberg discovered, namely that associated to fermion QED. Most of the considerations made in section 10.2.2 apply so the discussion will be brief, focusing on the differences with respect to the scalar QED case. The effective action ΔS_{IR} reads now

$$i\Delta S_{\text{IR}}(A) \equiv \ln \left\{ \mathcal{N} \int \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp \left(i \int d^4x \bar{\psi} (\not{D} - m) \psi \right) \right\} = \text{Tr} \ln \left(\frac{\not{D} - m}{\mu} \right) + \dots \quad (10.2.60)$$

The trace in eq.(10.2.60) includes the trace Tr_s over the spinor indices. This can be simplified by noting that

$$\text{Tr}_s (\not{D})^n = \text{Tr}_s \gamma_5 (\not{D})^n \gamma_5 = \text{Tr}_s (-\not{D})^n, \quad (10.2.61)$$

since $(\gamma_5)^2 = 1$ and $\{\gamma_\mu, \gamma_5\} = 0$. Therefore for a scalar function f , $\text{Tr}_s f(\not{D}) = \text{Tr}_s f(-\not{D})$ and accordingly

$$\text{Tr} \ln \left(\frac{\not{D} - m}{\mu} \right) = \frac{1}{2} \text{Tr} \ln \left(\frac{\not{D} - m}{\mu} \frac{-\not{D} - m}{\mu} \right) = \frac{1}{2} \text{Tr} \ln \left(\frac{\not{D}^2 + m^2}{\mu^2} \right). \quad (10.2.62)$$

Furthermore, we can write

$$\not{D}^2 = \frac{1}{2}(\{\gamma_\mu, \gamma_\nu\} + [\gamma_\mu, \gamma_\nu])D^\mu D^\nu = D^2 + \frac{e}{2}\sigma_{\mu\nu}F^{\mu\nu}, \quad (10.2.63)$$

where $[\gamma_\mu, \gamma_\nu] \equiv 2i\sigma_{\mu\nu}$ and we used the relation $[D_\mu, D_\nu] = -ieF_{\mu\nu}$. As in the QED scalar case it is convenient to perform a Wick's rotation. In addition to the Euclidean coordinates and fields discussed right before eq. (10.2.26), we introduce a set of Euclidean γ matrices γ_E such that $\gamma_i = \gamma_{E,i}$ ($i = 1, 2, 3$), $\gamma_0 = i\gamma_{E,4}$. As a result,

$$\not{D} = \gamma_0 \partial_0 - \vec{\gamma} \cdot \vec{\partial} = -(\gamma_{E,4} \partial_4 + \vec{\gamma} \cdot \vec{\partial}) \equiv -\not{D}_E, \quad (10.2.64)$$

with Euclidean metric. In addition, we introduce

$$[\gamma_{E,\mu}, \gamma_{E,\nu}] \equiv -2i\sigma_{E,\mu\nu}, \quad \sigma_{E,4k} = -i\sigma_{0k}, \quad \sigma_{E,lk} = \sigma_{lk}, \quad (10.2.65)$$

with $l, k \in \{1, 2, 3\}$, such that $F_{\mu\nu}\sigma^{\mu\nu} = F_{E,\mu\nu}\sigma_{E,\mu\nu}$. Using the chiral representation for the γ matrices one has, in terms of Pauli matrices $\{\sigma_l\}_{l=1,2,3}$,

$$\gamma_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma_l = \begin{pmatrix} 0 & \sigma_l \\ -\sigma_l & 0 \end{pmatrix}, \quad \sigma_{0l} = i \begin{pmatrix} -\sigma_l & 0 \\ 0 & \sigma_l \end{pmatrix}, \quad \sigma_{ij} = \epsilon_{ijk} \begin{pmatrix} \sigma_k & 0 \\ 0 & \sigma_k \end{pmatrix}, \quad (10.2.66)$$

and therefore

$$F_{E,\mu\nu}\sigma_{E,\mu\nu} = 2 \begin{pmatrix} \vec{\sigma} \cdot (-i\vec{E} + \vec{B}) & 0 \\ 0 & \vec{\sigma} \cdot (i\vec{E} + \vec{B}) \end{pmatrix}. \quad (10.2.67)$$

Collecting the above results we can then write, modulo irrelevant constant factors,

$$\begin{aligned} i\Delta S_{\text{IR}}(A) &= \frac{1}{2} \text{Tr} \ln \left(\frac{-D_E^2 + (e/2)F_E\sigma_E + m^2}{\mu^2} \right) \\ &= -\frac{1}{2} \int_0^\infty \frac{d\beta}{\beta} e^{-\beta m^2} \text{Tr} \left[e^{-\beta(-D_E^2)} e^{-(\beta e/2)F_E\sigma_E} \right]. \end{aligned} \quad (10.2.68)$$

Interestingly enough, the trace above splits into a direct product of traces, one involving the coordinate space and the differential operator $e^{-\beta(-D_E^2)}$, which is the same as the one discussed in the scalar QED case, and the other involving the spinor structure and the operator $e^{-(\beta e/2)F_E\sigma_E}$. So, the only new element in the analysis compared to the case of scalar QED is the calculation of $\text{Tr}_s e^{-(\beta e/2)F_E\sigma_E}$. Given that

$$\text{Tr} \exp \begin{pmatrix} \vec{\sigma} \cdot \vec{x} & 0 \\ 0 & \vec{\sigma} \cdot \vec{y} \end{pmatrix} = \text{Tr} \begin{pmatrix} e^{\vec{\sigma} \cdot \vec{x}} & 0 \\ 0 & e^{\vec{\sigma} \cdot \vec{y}} \end{pmatrix} = 2(\cosh |\vec{x}| + \cosh |\vec{y}|), \quad (10.2.69)$$

where $|\vec{x}| = \sqrt{\vec{x} \cdot \vec{x}}$, we get from eq. (10.2.67)

$$\text{Tr}_s e^{-(\beta e/2)F_E \sigma_E} = 2 \left[\cosh(\beta e| -i\vec{E} + \vec{B}|) + \cosh(\beta e|i\vec{E} + \vec{B}|) \right]. \quad (10.2.70)$$

We can now use the trigonometric identity

$$\cosh x + \cosh y = 2 \cosh \left(\frac{x+y}{2} \right) \cosh \left(\frac{x-y}{2} \right) \quad (10.2.71)$$

to write

$$\text{Tr}_s e^{-(\beta e/2)F_E \sigma_E} = 4 \cosh(\beta e q_+) \cosh(\beta e q_-), \quad (10.2.72)$$

where

$$q_{\pm} = \frac{1}{2} \left(| -i\vec{E} + \vec{B}| \pm |i\vec{E} + \vec{B}| \right). \quad (10.2.73)$$

It is now straightforward to see that

$$q_+^2 = -\lambda_- = a_-^2, \quad q_-^2 = -\lambda_+ = -a_+^2 \longrightarrow q_+ = \pm a_-, \quad q_- = \pm i a_+, \quad (10.2.74)$$

in terms of $\lambda_{\pm}$ and $a_{\pm}$ defined in eq.(10.2.43). Modulo a constant factor, the Euler-Heisenberg Lagrangian for fermion QED reads

$$\mathcal{L}_{\text{EH}}^{(F)} = -\frac{1}{4} Z_A^{(F)} F_{\mu\nu} F^{\mu\nu} - 2 \int_0^\infty \frac{d\beta}{\beta} e^{-\beta m^2} (4\pi\beta)^{-2} \frac{e\beta a_+}{\tan(e\beta a_+)} \frac{e\beta a_-}{\tanh(e\beta a_-)}. \quad (10.2.75)$$

In complete analogy to the scalar QED case, the UV divergences appearing in the integral above correspond to the renormalization of the constant and of the gauge kinetic term.

We now have

$$Z_A^{(F)} = 1 - \frac{8e^2}{3(4\pi)^2} \frac{1}{\epsilon} + \delta Z_A^{(S)}, \quad (10.2.76)$$

in agreement with the divergence found in eq.(5.7.7), where the photon wave function renormalization constant was denoted by Z_3 . We finally get

$$\mathcal{L}_{\text{EH}}^{(F)} = -\frac{1}{4} (1 + \delta Z_A^{(F)}) F_{\mu\nu} F^{\mu\nu} - \frac{m^4}{8\pi^2} \int_0^\infty dt e^{-t} \frac{1}{t^3} \left[\frac{x_+ t}{\tan(x_+ t)} \frac{x_- t}{\tanh(x_- t)} - 1 + \frac{t^2}{3} (x_+^2 - x_-^2) \right], \quad (10.2.77)$$

in terms of the variables defined in eq.(10.2.55). All the considerations made below eq.(10.2.55) apply to the fermion case. In particular, given the expansions

$$\frac{z}{\tanh z} = \sum_{k=0}^{\infty} \frac{2^{2k} B_{2k}}{(2k)!} x^{2k}, \quad \frac{z}{\tan z} = \sum_{k=0}^{\infty} (-1)^k \frac{2^{2k} B_{2k}}{(2k)!} x^{2k}, \quad (10.2.78)$$

we see that the large order behaviour of the series in the coupling of the vacuum energy is analogous to the scalar QED case and again it develops an imaginary part in the electric case, leading to its decay.

10.2.4 Vacuum Instability for a Constant Electric Field*

Before showing how an imaginary part arises from the Euler-Heisenberg Lagrangian when $E \neq 0$, let us derive the relation between the vacuum time decay and the imaginary part of the action. Recall that the path integral with no operator insertion can be interpreted as the vacuum to vacuum transition $\langle 0|0\rangle$. We typically assume that the vacuum around which we build our Hilbert space is stable, which implies $|\langle 0|0\rangle| = 1$ in absence of external fields. For a system at finite volume V , in presence of an external field we have

$${}_{out}\langle 0|0\rangle_{in} = {}_{in}\langle 0|e^{-iHT}|0\rangle_{in} = e^{iS_{\text{eff}}(\phi_0)} = \int \mathcal{D}\phi e^{iS(\phi, \phi_0)}, \quad (10.2.79)$$

where ϕ_0 represent the extrnal fields and ϕ the quantum fluctuations which are integrated out. As long as the vacuum is stable, the probability amplitude $\langle 0|e^{-iHT}|0\rangle$ has unit modulus and the effective action S_{eff} is real. On the other hand, if the latter acquires an imaginary part, the vacuum decays with a rate per unit volume given by

$$\Gamma \equiv \frac{2\text{Im } S_{\text{eff}}}{VT}. \quad (10.2.80)$$

Let us now come back to the scalar QED case of the Euler-Heisenberg Lagrangian given by eq.(10.2.54). When $x_+ \neq 0$, the integral in t is not well defined because of the presence of simple poles located at $x_+ = \pi n$, with $n > 0$. We can slightly shift the poles off the real axis to avoid the problem, but then the result depends on wether we move the poles below or above the real axis, and an ambiguity arises. Luckily enough, we should remember about the Feynman $i\epsilon$ prescription (which we left implicit in the path integral), which effectively amounts at substituting $m^2 \rightarrow m^2 - i\epsilon$ with $\epsilon \rightarrow 0^+$. With this prescription we know how the singularities should move and the integral can be safely calculated. Given the relation (10.2.55), $m^2 \rightarrow m^2 - i\epsilon$ implies $x_+ \rightarrow x_+ + i\epsilon$. Recalling eq. (2.1.29), we have

$$\lim_{x \rightarrow \pi n} \frac{1}{\sin(x + i\epsilon)} = P\left(\frac{1}{\sin x}\right) - \pi(-1)^n \delta(x - \pi n). \quad (10.2.81)$$

The integrand in the Lagrangian (10.2.54) develops a total imaginary part which is given by the sum over all imaginary parts for the simple poles:

$$\text{Im} \frac{1}{\sin(x_+ t + i\epsilon)} = -\pi \sum_{n=1}^{\infty} (-1)^n \delta(x_+ t - \pi n). \quad (10.2.82)$$

Accordingly, we have

$$\text{Im } \mathcal{L}_{\text{EH}}^{(S)} = \frac{e^2 a_+ a_-}{(4\pi)^2} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \frac{e^{-n\pi m^2/(ea_+)}}{\sinh(n\pi a_-/a_+)}. \quad (10.2.83)$$

Note that this expression is non-perturbative in the coupling constant e , due to the essential singularity associated with the exponential factor. Consider now the two relevant cases in which either field $\vec{E}$ or $\vec{B}$ are present. In the latter case we have $a_+ = 0$, and hence the decay rate Γ of the vacuum vanishes. In the former case $a_+ = E$, $a_- = 0$, and eqs.(10.2.80) and (10.2.83) give

$$\Gamma^{(S)} = \frac{e^2 E^2}{8\pi^3} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} e^{-n\pi m^2/(eE)}. \quad (10.2.84)$$

A similar computation gives the decay rate of the vacuum in fermion QED:

$$\Gamma^{(F)} = \frac{e^2 E^2}{4\pi^3} \sum_{n=1}^{\infty} \frac{1}{n^2} e^{-n\pi m^2/(eE)}. \quad (10.2.85)$$

In both cases the exponential factor governing the decay is proportional to E_{cr}/E , where $E_{\text{cr}} = 2m^2/e$ is a critical value of the electric field. It is useful to understand the appearance of E_{cr} from simple physical arguments. The decay occurs via emission of particle-anti particle pairs created out from the vacuum. A constant electric field E polarizes the virtual particle-anti particle pairs populating the vacuum. If $E \geq E_{\text{cr}}$ the electrostatic energy density is so high that it is energetically favourable to create real pairs to “discharge” the vacuum. This occurs when the electrostatic energy in a region of Compton size $\lambda \sim 1/m$ exceeds the particle rest frame mass $2m$, namely when $eE \times \lambda \gtrsim 2m$, i.e. $E \gtrsim 2m^2/e = E_{\text{cr}}$. For electrons $E_{\text{cr}} \sim 10^{16}$ V/cm, a value that is way too strong to be achieved in any laboratory with the current technologies. On the other hand, no decay occurs when only a magnetic field is present, because magnetic fields do not provide work on charged particles and therefore cannot provide the necessary energy to produce particle-anti particle pairs from the vacuum.

Chapter 11

Final Project: The Abelian Higgs Model^{★★}

In this last chapter we study the abelian Higgs model, along the lines of ref. [17]. This chapter can be seen as an exercise in which many of the notions and techniques introduced in these notes are used at the same time, in particular the effective potential, the background field method, gauge-fixing dependence, Callan-Symanzik equations, β -functions and anomalous dimensions.

The structure of this chapter is as follows. The model and the gauge-fixing used are defined in section 11.1. The one-loop effective potential in a generic ξ -gauge is computed in section 11.2. In section 11.3 we go beyond the effective potential and compute the one-loop effective action, namely the corrections to the gauge and scalar kinetic terms. Using the quantum effective action and a functional form of the Callan-Symanzik equations, we determine in section 11.4 the one-loop beta-functions for the $U(1)$ gauge and quartic scalar couplings, the photon and scalar field anomalous dimensions, and we discuss the phenomenon of radiative symmetry breaking.

11.1 The Abelian Higgs Model in a generic ξ -gauge

The Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + |D_\mu\Phi|^2 - \frac{\lambda}{6}(\Phi^\dagger\Phi)^2, \quad (11.1.1)$$

where $D_\mu\Phi = \partial_\mu\Phi - ieA_\mu\Phi$. Our aim will be to understand the vacuum of this theory, namely whether the $U(1)$ gauge symmetry is broken or not, and the RG flows of the two couplings e and λ . First of all, we have to gauge fix the theory. Since we want to study the effective potential as a function of the VEV of Φ , it is convenient to use a generalized

ξ -gauge which is valid for any value of $\langle\Phi\rangle$:

$$\mathcal{L}_{g.f.} = -\frac{1}{2\xi} \left[\partial_\mu A^\mu + ie\xi(\phi^\dagger\phi_0 - \phi\phi_0^\dagger) \right]^2, \quad (11.1.2)$$

where $\Phi = \phi_0 + \phi$, with ϕ_0 the VEV of Φ and ϕ its quantum fluctuation. It is straightforward to verify that the quadratic mixing terms between ϕ and A_μ vanish when $\mathcal{L}_{g.f.}$ is added to the Lagrangian (11.1.1). Even if the local symmetry is abelian, the ghosts do not decouple in the ξ -gauge we have chosen. The ghost Lagrangian, as usual, is derived by taking the infinitesimal variation of $\mathcal{L}_{g.f.}$ with respect to a $U(1)$ transformation. One gets

$$\mathcal{L}_{ghosts} = \partial_\mu \omega^* \partial^\mu \omega - e^2 \xi \omega^* \omega \left[2|\phi_0|^2 + (\phi^\dagger\phi_0 + \phi_0^\dagger\phi) \right]. \quad (11.1.3)$$

In the Landau gauge $\partial_\mu A^\mu = 0$, reached for $\xi \rightarrow 0$, the ghosts are free and decouple, whereas in the unitary gauge $\xi \rightarrow \infty$ they are infinitely massive and decouple again. We will not fix in the following a specific value of ξ , so that the ghosts should be taken into account. The total Lagrangian is

$$\mathcal{L}_{tot} = \mathcal{L} + \mathcal{L}_{g.f.} + \mathcal{L}_{ghosts}. \quad (11.1.4)$$

11.2 One-loop Effective Potential^{**}

The 1-loop effective potential is completely determined by the terms in $\mathcal{L}_{tot}$ quadratic in the field fluctuations. In momentum space, we have

$$\mathcal{L}_{tot,quad}(p) = -\frac{1}{2} A^\mu(-p) A^\nu(p) \mathcal{L}_{\mu\nu}^{(A)}(p) + \omega^*(-p) \mathcal{L}^{(\omega)}(p) \omega(p) + \frac{1}{2} \phi_i(-p) \mathcal{L}_{ij}^{(\phi)}(p) \phi_j(p), \quad (11.2.1)$$

where $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$, and

$$\begin{aligned} \mathcal{L}_{\mu\nu}^{(A)}(p) &= \eta_{\mu\nu}(p^2 - 2e^2|\phi_0|^2) - \left(1 - \frac{1}{\xi}\right) p_\mu p_\nu, \\ \mathcal{L}^{(\omega)}(p) &= p^2 - 2\xi e^2|\phi_0|^2, \\ \mathcal{L}_{ij}^{(\phi)}(p) &= \begin{pmatrix} p^2 - |\phi_0|^2(e^2\xi + \frac{2\lambda}{3}) + (\phi_0^2 + \phi_0^{*2})(\frac{e^2\xi}{2} - \frac{\lambda}{6}) & i(\phi_0^2 - \phi_0^{*2})(\frac{e^2\xi}{2} - \frac{\lambda}{6}) \\ i(\phi_0^2 - \phi_0^{*2})(\frac{e^2\xi}{2} - \frac{\lambda}{6}) & p^2 - |\phi_0|^2(e^2\xi + \frac{2\lambda}{3}) - (\phi_0^2 + \phi_0^{*2})(\frac{e^2\xi}{2} - \frac{\lambda}{6}) \end{pmatrix}. \end{aligned} \quad (11.2.2)$$

Modulo irrelevant factors,

$$\begin{aligned} \det \mathcal{L}_{\mu\nu}^{(A)}(p) &= (p^2 - 2e^2|\phi_0|^2)^3 (p^2 - 2e^2\xi|\phi_0|^2), \\ \det \mathcal{L}_{ij}^{(\phi)}(p) &= (p^2 - \lambda|\phi_0|^2) \left[p^2 - |\phi_0|^2(2e^2\xi + \frac{\lambda}{3}) \right]. \end{aligned} \quad (11.2.3)$$

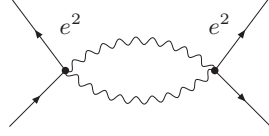


Figure 11.1: One-loop graph leading to a $\lambda\rho^4$ coupling.

Summing over all contributions (gauge, ghosts and scalar fields), we get

$$V_{1-loop}(\rho) = \frac{1}{2} \int \frac{d^4 p_E}{(2\pi)^4} \left[3 \log(p_E^2 + e^2 \rho^2) - \log(p_E^2 + \xi e^2 \rho^2) \right. \\ \left. + \log(p_E^2 + \frac{\lambda}{2} \rho^2) + \log\left(p_E^2 + \left(\xi e^2 + \frac{\lambda}{6}\right) \rho^2\right) \right], \quad (11.2.4)$$

where $\rho^2 = 2|\phi_0|^2$. We renormalize V_{1-loop} by demanding that

$$\left. \frac{d^2 V_{1-loop}}{d\rho^2} \right|_{\rho=0} = 0, \quad \left. \frac{d^4 V_{1-loop}}{d\rho^4} \right|_{\rho=\mu} = 0. \quad (11.2.5)$$

After some simple algebra, we obtain

$$V_{eff}(\rho) = V_{tree}(\rho) + V_{1-loop}(\rho) = \frac{\lambda}{4!} \rho^4 + \frac{\rho^4}{64\pi^2} \left(3e^4 + \frac{5}{18} \lambda^2 + \frac{1}{3} \xi \lambda e^2 \right) \left(\log \frac{\rho^2}{\mu^2} - \frac{25}{6} \right). \quad (11.2.6)$$

Let us study the minima of V_{eff} , assuming that $\lambda \sim e^4$, so that at leading order we can neglect the λ^2 and $\xi \lambda e^2$ terms in eq. (11.2.6). It is important to emphasize here that at tree-level (or, alternatively, at any given energy scale) we can assume any relation we like of the form $\lambda \sim e^n$ for any n , but such relations at the quantum level cannot generally hold at any energy scale for any n . Indeed, radiative corrections will nevertheless generate the $\lambda\rho^4$ coupling in the theory. The leading one-loop correction arises from virtual photons as illustrated in fig. 11.1. Being this correction of $\mathcal{O}(e^4)$, we see that $\lambda \sim e^n$, with $n \leq 4$, are the only radiatively stable assumptions we can make. The extrema of V_{eff} are

$$\frac{dV_{eff}}{d\rho} = \rho^3 \left(\frac{\lambda}{6} + \frac{e^4}{16\pi^2} \left(3 \log \frac{\rho^2}{\mu^2} - 11 \right) \right) = 0. \quad (11.2.7)$$

Taking $\mu = \langle \rho \rangle$ in eq. (11.2.7), we get $\rho = 0$ and

$$\lambda(\langle \rho \rangle) = \frac{33}{8\pi^2} e^4(\langle \rho \rangle). \quad (11.2.8)$$

Equation (11.2.8) is an instance of dimensional transmutation: we have traded the VEV of ρ for the coupling λ . Plugging back in V_{eff} gives

$$V_{eff} \simeq \frac{3e^4}{64\pi^2} \rho^4 \left(\log \frac{\rho^2}{\langle \rho \rangle^2} - \frac{1}{2} \right). \quad (11.2.9)$$

The extremum (11.2.8) is a minimum. The photon and scalar masses are

$$m_\gamma^2 = e^2 \langle \rho \rangle^2, \quad m_\rho^2 = \frac{3e^4}{8\pi^2} \langle \rho \rangle^4. \quad (11.2.10)$$

We conclude that in this theory a dynamical spontaneous symmetry breaking of the $U(1)$ gauge symmetry can occur. In order to firmly establish that, we have to compute the RG evolutions of λ and of the charge e to check the existence of an energy scale $\langle \rho \rangle$ where eq. (11.2.8) is valid. This computation will be the subjects of the following sections.

11.3 The Quantum Effective Action**

An instructive way of computing the β -functions of λ and e , as well as the anomalous dimensions of A_μ and ϕ , makes use of a functional form of the CS equations (5.6.3), which reads:

$$\mu \frac{d\Gamma}{d\mu} = \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - \gamma \int d^4x \phi(x) \frac{\delta}{\delta \phi(x)} \right) \Gamma(\phi) = 0, \quad (11.3.1)$$

where $\Gamma(\phi)$ is the quantum effective action. For a single real scalar field, keeping up to two derivative terms, the latter reads

$$\Gamma(\phi) = \int d^4x \left(\frac{1}{2} Z(\phi) (\partial_\mu \phi)^2 - V_{eff}(\phi) \right), \quad (11.3.2)$$

where V_{eff} is the Coleman-Weinberg potential and Z is the radiative correction to the kinetic term. Once Γ is known, eq. (11.3.1) gives us β and γ . It is straightforward to see that eq. (11.3.1) encodes all eqs. (5.6.3) for any n , by recalling that

$$\Gamma(\phi) = \sum_{n=0}^{\infty} \frac{1}{n!} \int d^4x_1 \dots d^4x_n \Gamma^{(n)}(x_1, \dots, x_n) \phi(x_1) \dots \phi(x_n). \quad (11.3.3)$$

In the case at hand, with two fields and two couplings, eq. (11.3.1) generalizes to

$$\left(\mu \frac{\partial}{\partial \mu} + \beta_\lambda \frac{\partial}{\partial \lambda} + \beta_e \frac{\partial}{\partial e} - \gamma_A \int d^4x A_{\mu,0}(x) \frac{\delta}{\delta A_{\mu,0}(x)} - \gamma_\phi \int d^4x \left(\phi_0(x) \frac{\delta}{\delta \phi_0(x)} + \phi_0^\dagger(x) \frac{\delta}{\delta \phi_0^\dagger(x)} \right) \right) \Gamma(\phi_0, A_{\mu,0}) = 0, \quad (11.3.4)$$

where ϕ_0 , $\phi_0^\dagger$ and $A_{\mu,0}$ are the background field configurations. Invariance under the background $U(1)$ gauge invariance implies that the lowest dimensional operators appearing in Γ are of the form

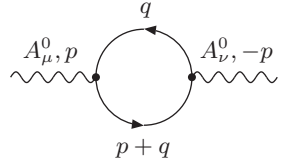
$$\Gamma(\phi_0, A_{\mu,0}) = \int d^4x \left[-\frac{1}{4}H(\rho)F_{\mu\nu,0}^2 + Z(\rho)|D_\mu\phi_0|^2 - V_{eff}(\rho) \right]. \quad (11.3.5)$$

The effective potential V_{eff} has been already computed and is given by eq. (11.2.6). We have then to determine H and Z only. This can be done by decomposing the gauge field as well in terms of background and fluctuation fields: $A_\mu \rightarrow A_\mu^0 + A_\mu$.

Let us start by computing H , in which case we can take $\phi_0 = \text{constant}$. The relevant interaction terms are

$$\mathcal{L} \supset ieA_0^\mu(\phi^\dagger\partial_\mu\phi - \partial_\mu\phi^\dagger\phi), \quad (11.3.6)$$

and H is determined by the contraction of the two scalar currents. The contractions of the form $\langle\phi\phi\rangle$ or $\langle\phi^\dagger\phi^\dagger\rangle$, although non-vanishing for $\phi_0 \neq 0$ due to mass terms of the form ϕ^2 and $(\phi^\dagger)^2$, give only rise to irrelevant divergent contact terms. In DR, for instance, they all trivially vanish. The only relevant contraction is the usual one, of the form $\langle\phi^\dagger\phi\rangle$. The computation essentially boils down to the one-loop photon vacuum polarization in scalar QED. Two diagrams contribute. The first reads



$$= i\Pi_{\mu\nu}^1(p) = (ie)^2\mu^\epsilon \int \frac{d^dq}{(2\pi)^d} \frac{i^2(p+2q)_\mu(p+2q)_\nu}{(q^2 - m^2)[(p+q)^2 - m^2]}, \quad (11.3.7)$$

where, according to eq. (11.2.2),

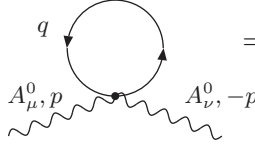
$$m^2 = m^2(\rho) = \rho^2 \left(\frac{2\lambda}{3} + \frac{e^2\xi}{2} \right). \quad (11.3.8)$$

By performing the usual manipulations (introduce the Feynman parameter x , shift $q \rightarrow q - xp$ and Wick rotate to euclidean momentum), we get

$$\begin{aligned} i\Pi_{\mu\nu}^1(p) &= (ie)^2\mu^\epsilon \int \frac{d^dq}{(2\pi)^d} \frac{i^2(p+2q)_\mu(p+2q)_\nu}{(q^2 - m^2)[(p+q)^2 - m^2]} = \frac{ie^2\mu^\epsilon}{(4\pi)^{d/2}} \int_0^1 dx [m^2 - p^2x(1-x)]^{\frac{d-4}{2}} \\ &\quad \times \Gamma\left(\frac{4-d}{2}\right) \left(p_\mu p_\nu (1-2x)^2 - \frac{4}{2-d} \eta_{\mu\nu} [m^2 - p^2x(1-x)] \right). \end{aligned} \quad (11.3.9)$$

The second diagram is a tadpole, that does not depend on the external momentum p . It can be cast in a form close to eq. (11.3.9) by multiplying and dividing it by $(p+q)^2 - m^2$.

Using the same manipulations as before, we get



$$= i\Pi_{\mu\nu}^2(p) = 2ie^2\mu^\epsilon\eta_{\mu\nu}\int\frac{d^dq}{(2\pi)^d}\frac{i}{q^2-m^2}\frac{(p+q)^2-m^2}{(p+q)^2-m^2} = -\frac{2ie^2\mu^\epsilon}{(4\pi)^{d/2}}\eta_{\mu\nu} \quad (11.3.10)$$

$$\times \int_0^1 dx [m^2 - p^2x(1-x)]^{\frac{d-4}{2}} \Gamma\left(\frac{4-d}{2}\right) \left((1-x)^2p^2 - m^2 - \frac{d}{2-d}[m^2 - p^2x(1-x)]\right).$$

Summing the two contributions, we have

$$i\Pi_{\mu\nu} = i\Pi_{\mu\nu}^1 + i\Pi_{\mu\nu}^2 = \frac{ie^2}{24\pi^2}(p_\mu p_\nu - \eta_{\mu\nu}p^2)\left(\frac{1}{\epsilon} + \text{const.}\right) - \frac{ie^2}{16\pi^2}\int_0^1 dx \log \frac{m^2 - p^2x(1-x)}{\mu^2} \\ \times \left(p_\mu p_\nu(1-2x)^2 - \eta_{\mu\nu}p^2(4x^2 - 6x + 2)\right). \quad (11.3.11)$$

We are interested in computing H , which is the coefficient of the F^2 term, quadratic in the external momentum p . The last term in eq. (11.3.11) is already quadratic in p , so we can safely neglect the p^2 term inside the log and keep only the mass term m^2 . In this way we get

$$i\Pi_{\mu\nu} = \frac{ie^2}{24\pi^2}(p_\mu p_\nu - \eta_{\mu\nu}p^2)\left(\frac{1}{\epsilon} + \text{const.} - \log(\rho/\mu)\right), \quad (11.3.12)$$

where we have reabsorbed in the arbitrary constant the ρ -independent factors coming from eq. (11.3.8). As expected, $\Pi_{\mu\nu}$ is transverse. We now perform a non-minimal subtraction, i.e. we add the counter-term¹ $i(Z-1)(p_\mu p_\nu - \eta_{\mu\nu}p^2)$ and require that

$$i\Pi_{\mu\nu} + i(Z-1)(p_\mu p_\nu - \eta_{\mu\nu}p^2) = 0, \text{ at } \rho = \mu. \quad (11.3.13)$$

The renormalized one-loop photon vacuum polarization is then

$$i\Pi_{\mu\nu}^R = \frac{ie^2}{24\pi^2}(\eta_{\mu\nu}p^2 - p_\mu p_\nu) \log \frac{\rho}{\mu} \equiv iAp^2\mathcal{P}_{\mu\nu}, \quad (11.3.14)$$

where $A = e^2/(24\pi^2) \log \phi/\mu$ and we have defined the projector

$$\mathcal{P}_{\mu\nu} = \eta_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}. \quad (11.3.15)$$

The tree-level photon propagator is

$$G_{\mu\nu}(p) = -\frac{i}{p^2}\left(\mathcal{P}_{\mu\nu} + \xi \frac{p_\mu p_\nu}{p^2}\right). \quad (11.3.16)$$

¹Of course, the counter-term Z should not be confused with the finite $Z(\rho)$ appearing in the effective action (11.3.5).

Iterating the one-loop correction (11.3.14) in the tree-level expression (11.3.16) we get

$$G_{\mu\nu}(p) \rightarrow \frac{-i\mathcal{P}_{\mu\nu}}{p^2(1-A)} - \frac{i}{p^2} \frac{\xi p_\mu p_\nu}{p^2} \quad (11.3.17)$$

from which we see that

$$-\frac{1}{4}F_{\mu\nu,0}^2 \rightarrow -\frac{1}{4}(1-A)F_{\mu\nu,0}^2 \quad (11.3.18)$$

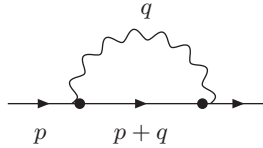
finally giving the desired $H(\rho)$:

$$H(\rho) = 1 - A(\rho) = 1 - \frac{e^2}{24\pi^2} \log \frac{\rho}{\mu}. \quad (11.3.19)$$

Let us now determine Z , the wave-function renormalization of ϕ_0 . In this case we can set $A_{\mu,0} = 0$ but of course ϕ_0 can no longer be taken constant. The tadpole graph given by the quartic scalar interaction induces only a mass renormalization, so we need only to consider the contribution induced by the gauge interactions. These are of the form $A_\mu \phi \phi_0$, given by

$$\mathcal{L} \supset 2ieA^\mu (\phi^\dagger \partial_\mu \phi_0 - \partial_\mu \phi_0^\dagger \phi). \quad (11.3.20)$$

We have



$$= i\Sigma(p) = (2ie)^2 \mu^\epsilon p^\mu p^\nu \int \frac{d^d q}{(2\pi)^d} G_{\mu\nu}(q) \frac{i}{(p+q)^2 - m^2}, \quad (11.3.21)$$

where $G_{\mu\nu}(q)$ is the photon propagator in the generically broken phase, as given by in eq.(8.4.13), with $\mu_\gamma^2 = e^2 \rho^2$.² For $\xi \neq 1$, up to three denominators appear in eq. (11.3.21) and the computation is a bit involved. We can greatly simplify it by noticing that i) the $\log \mu$ (and hence the $\log \rho$) terms we are looking for can be detected by finding the residues of the appropriate $1/q^4$ poles in the integrand, as explained in section 3.5 and ii) we can set $p = 0$ in the scalar propagator since we already have two powers of the external momentum coming from the vertices. After Wick rotating to euclidean momenta and using $SO(d)$ invariance of the integrand, we get

$$\begin{aligned} i\Sigma(p) &= -4ie^2 \mu^\epsilon p^2 \int \frac{d^d q_E}{(2\pi)^d} \left(\frac{1}{q_E^2 + \mu_\gamma^2} - \frac{(1-\xi)q_E^2}{d(q_E^2 + \xi\mu_\gamma^2)(q_E^2 + \mu_\gamma^2)} \right) \frac{1}{q_E^2 + m^2} \\ &= (-4ie^2) p^2 \frac{1}{8\pi^2} \left(\frac{1}{\epsilon} + \log \mu + \dots \right) \frac{\xi + 3}{4}, \end{aligned} \quad (11.3.22)$$

²Pay attention in not confusing the photon mass μ_γ with the sliding scale μ !

where the ... include the ρ -dependent terms we are looking for plus additional finite pieces. The renormalized scalar two-point function reads then

$$i\Sigma^R(p) = \frac{ie^2}{8\pi^2}(3 + \xi)p^2 \log \frac{\rho}{\mu} \equiv ip^2 B, \quad (11.3.23)$$

where, as before, we demand that $\Sigma^R(p) = 0$ when $\rho = \mu$. Iterating the one-loop correction in the tree-level scalar propagator, as in the photon case, gives

$$\frac{i}{p^2} \rightarrow \frac{i}{p^2(1 + B)}, \quad (11.3.24)$$

from which we extract $Z(\rho)$:

$$Z(\rho) = 1 + B(\rho) = 1 + \frac{e^2}{8\pi^2}(3 + \xi) \log \frac{\rho}{\mu}. \quad (11.3.25)$$

The form of the effective action (11.3.5) is finally determined and we can proceed to use it to compute the RG evolution of the couplings e and λ . It is worth to emphasize that the effective action (11.3.5) does *not* appear to be gauge invariant, since ξ enters in both $Z(\rho)$ and in $V_{eff}(\rho)$. This does not signal an inconsistency of the computation, because the effective action (11.3.5) is not a physical observable per se. On the other hand, by a field rescaling the ξ -dependence can be drastically decreased. In fact, the kinetic term of ϕ_0 is not canonical and it is convenient to rescale the field as

$$\phi_0 \rightarrow \frac{1}{\sqrt{Z(\rho)}} \phi_0 \quad (11.3.26)$$

to canonically normalize it. The potential (11.2.6) is then rescaled by a factor $1/Z(\rho)^2$ and we have

$$\begin{aligned} \frac{1}{Z(\rho)^2} V_{eff}(\rho) &= \frac{\lambda}{4!} \rho^4 \left(1 - \frac{e^2}{4\pi^2} (3 + \xi) \log \frac{\rho}{\mu} \right) + \frac{\rho^4}{32\pi^2} \left(3e^4 + \frac{5}{18} \lambda^2 + \frac{1}{3} \xi \lambda e^2 \right) \log \frac{\rho}{\mu} \\ &= \frac{\lambda}{4!} \rho^4 + \frac{\rho^4}{32\pi^2} \left(3e^4 + \frac{5}{18} \lambda^2 - e^2 \lambda \right) \log \frac{\rho}{\mu}. \end{aligned} \quad (11.3.27)$$

In the new basis the ξ -dependence in the effective action has been completely canceled. In eq. (11.3.27) we have focused on the log terms only, since the constant term is scheme-dependent and we have not been careful in systematically using a given scheme.

11.4 RG equations and Their Solutions**

For a single real scalar field, the functional RG equation (11.3.1) splits into two ordinary differential equations for Z and V :

$$\begin{aligned} \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - \gamma \phi \frac{\partial}{\partial \phi} - 2\gamma \right) Z &= 0, \\ \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - \gamma \phi \frac{\partial}{\partial \phi} \right) V_{eff} &= 0. \end{aligned} \quad (11.4.1)$$

In dealing with quartic potentials, like in our case, it is actually convenient to write an RG equation for $V^{(4)} \equiv \partial^4 V_{eff}/\partial\phi^4$ rather than V_{eff} itself. Given that $\partial_\phi^4(\gamma\phi\partial_\phi) = 4\gamma\partial_\phi^4 + \gamma\phi\partial_\phi\partial_\phi^4$, we have

$$\left(\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial\lambda} - 4\gamma - \gamma\phi\frac{\partial}{\partial\phi}\right)V^{(4)} = 0. \quad (11.4.2)$$

The RG equation (11.4.2) can further be simplified by noting that $V^{(4)}$ depends on ϕ only through the dimensionless combination ϕ/μ . Then $\phi\partial_\phi V^{(4)} = -\mu\partial_\mu V^{(4)} = \partial_t V^{(4)}$, where $t = \log \phi/\mu$ and thus we get

$$\left(-\frac{\partial}{\partial t} + \bar{\beta}\frac{\partial}{\partial\lambda} - 4\bar{\gamma}\right)V^{(4)} = 0. \quad (11.4.3)$$

where

$$\bar{\beta} = \frac{\beta}{1+\gamma}, \quad \bar{\gamma} = \frac{\gamma}{1+\gamma}. \quad (11.4.4)$$

Using the same manipulations, the RG equation for Z becomes

$$\left(-\frac{\partial}{\partial t} + \bar{\beta}\frac{\partial}{\partial\lambda} - 2\bar{\gamma}\right)Z = 0. \quad (11.4.5)$$

In the situation at hand, the functional RG equation is given by eq. (11.3.4). It is clear that there can be no cancellations between the three terms (11.3.5) appearing in Γ , so eq. (11.3.4) splits into three independent equations. Let us first focus on the scalar kinetic term $|D\phi_0|^2$:

$$\begin{aligned} &|D_\mu\phi_0|^2\left(-\frac{\partial}{\partial t} + \beta_e\frac{\partial}{\partial e} - \gamma_\phi\rho\partial_\rho\right)Z + Z\left(\beta_e\frac{\partial}{\partial e}|D_\mu\phi_0|^2 - \gamma_A\int d^4xA_{\nu,0}(x)\frac{\delta|D_\mu\phi_0|^2}{\delta A_{\nu,0}(x)}\right. \\ &\left.-\gamma_\phi\int d^4x\left(\phi_0(x)\frac{\delta}{\delta\phi_0(x)} + \phi_0^\dagger(x)\frac{\delta}{\delta\phi_0^\dagger(x)}\right)\right)|D_\mu\phi_0|^2 = 0, \end{aligned} \quad (11.4.6)$$

where we have used the fact that Z depends on constant ρ only, with $\phi\partial_\phi + \phi^\dagger\partial_{\phi^\dagger} = \rho\partial_\rho$. The term in the second line of eq. (11.4.6) equals $-2\gamma_\phi|D_\mu\phi_0|^2$, while the last two terms in the first line of eq. (11.4.6) gives rise to a different operator. As such, two independent equations arise from eq. (11.4.6). Requiring the vanishing of the coefficient proportional to $|D_\mu\phi_0|^2$ gives

$$\left(-\frac{\partial}{\partial t} + \bar{\beta}_e\frac{\partial}{\partial e} - 2\bar{\gamma}_\phi\right)Z = 0, \quad (11.4.7)$$

with

$$\bar{\beta}_e = \frac{\beta_e}{1+\gamma_\phi}, \quad \bar{\gamma}_\phi = \frac{\gamma_\phi}{1+\gamma_\phi}. \quad (11.4.8)$$

It is straightforward to see that

$$\int d^4x A_{\nu,0}(x) \frac{\delta |D_\mu \phi_0|^2}{\delta A_{\nu,0}(x)} = e \frac{\partial}{\partial e} |D_\mu \phi_0|^2 \quad (11.4.9)$$

and hence the vanishing of the coefficient multiplying $\partial |D_\mu \phi_0|^2 / \partial e$ gives

$$\beta_e = e \gamma_A. \quad (11.4.10)$$

Consider now the gauge kinetic term $F_{\mu\nu,0}^2$. We get

$$-\frac{1}{4} F_{\mu\nu,0}^2 \left(-\frac{\partial}{\partial t} + \beta_e \frac{\partial}{\partial e} - \gamma_\phi \rho \partial_\rho \right) H - 2\gamma_A H \left(-\frac{1}{4} F_{\mu\nu,0}^2 \right) = 0, \quad (11.4.11)$$

that gives rise to

$$\left(-\frac{\partial}{\partial t} + \bar{\beta}_e \frac{\partial}{\partial e} - 2\bar{\gamma}_A \right) H = 0, \quad \bar{\gamma}_A = \frac{\gamma_A}{1 + \gamma_\phi}. \quad (11.4.12)$$

Finally we have the potential term. As explained before, we write an RG equation for $V^{(4)}$ rather than V_{eff} , which is the obvious generalization of eq. (11.4.3):

$$\left(-\frac{\partial}{\partial t} + \bar{\beta}_\lambda \frac{\partial}{\partial \lambda} + \bar{\beta}_e \frac{\partial}{\partial e} - 4\bar{\gamma}_\phi \right) V^{(4)} = 0. \quad (11.4.13)$$

Equations (11.4.7), (11.4.10), (11.4.12) and (11.4.13) are enough to determine β_e , β_λ , γ_A and γ_ϕ . Let us recall below the explicit form of Z , H and $V^{(4)}$, the latter computed from eq. (11.2.6):

$$\begin{aligned} Z(t) &= 1 + \frac{e^2}{8\pi^2} (\xi + 3)t \\ H(t) &= 1 - \frac{e^2}{24\pi^2} t \\ V^{(4)}(t) &= \lambda + \frac{1}{4\pi^2} \left(9e^4 + \frac{5}{6} \lambda^2 + \xi \lambda e^2 \right) t \end{aligned} \quad (11.4.14)$$

One immediately gets from eqs. (11.4.7), (11.4.10) and (11.4.12)

$$\begin{aligned} \gamma_\phi &= \bar{\gamma}_\phi + \mathcal{O}(e^4) = -\frac{e^2}{16\pi^2} (\xi + 3) + \mathcal{O}(e^4), \\ \gamma_A &= \bar{\gamma}_A + \mathcal{O}(e^4) = \frac{e^2}{48\pi^2} + \mathcal{O}(e^4), \\ \beta_e &= \bar{\beta}_e + \mathcal{O}(e^5) = \frac{e^3}{48\pi^2} + \mathcal{O}(e^5), \end{aligned} \quad (11.4.15)$$

Plugging the values (11.4.15) in eq. (11.4.13) allows us to determine β_λ :

$$\beta_\lambda = \bar{\beta}_\lambda + \mathcal{O}(e^6, e^4 \lambda, e^2 \lambda^2) = \frac{1}{4\pi^2} \left(9e^4 + \frac{5}{6} \lambda^2 - 3\lambda e^2 \right) + \mathcal{O}(e^6, e^4 \lambda, e^2 \lambda^2). \quad (11.4.16)$$

Notice how all ξ -dependent factors have cancelled in β_λ as it should be, being the latter gauge invariant, like β_e (and γ_A). The scalar field anomalous dimension γ_ϕ , instead, does depend on ξ . This is expected since ϕ changes by a phase under a gauge transformation and at the quantum level there is no gauge invariant notion of γ_ϕ .

The RG flow of e is easily computed from β_e . We get

$$e^2(t) = \frac{e_0^2}{1 - \frac{e_0^2}{24\pi^2}t}. \quad (11.4.17)$$

The RG flow of λ requires some more work. It is convenient to define $R(t) = \lambda(t)/e^2(t)$ and write an RG equation for R . One gets

$$e^2(t)\dot{R}(t) = \frac{e^4(t)}{4\pi^2} \left(\frac{5}{6}R^2(t) - \frac{19}{6}R(t) + 9 \right), \quad (11.4.18)$$

which is further simplified by considering $R = R(e^2)$, so that

$$\dot{R} = \frac{dR}{de^2} 2e\dot{e} = \frac{e^4}{24\pi^2} \frac{dR}{de^2}. \quad (11.4.19)$$

Our desired final equation reads

$$e^2 \frac{dR(e^2)}{de^2} = 5R^2(e^2) - 19R(e^2) + 54, \quad (11.4.20)$$

whose solution is

$$R(e^2) = \frac{1}{10} \left(19 + \sqrt{719} \tan \left(\frac{1}{2} \sqrt{719} \log e^2 + \theta \right) \right) \quad (11.4.21)$$

giving

$$\lambda(t) = \frac{e^2(t)}{10} \left(19 + \sqrt{719} \tan \left(\frac{1}{2} \sqrt{719} \log e^2(t) + \theta \right) \right), \quad (11.4.22)$$

where θ is an integration constant. Both e and λ grow in the UV, with the quartic coupling λ varying significantly over a range in which the electric charge remains essentially constant. We conclude that for a wide range of initial conditions for e and λ there exists an energy scale where $\lambda \sim e^4$, and in particular eq. (11.2.8) is valid.

Dynamical transmutation occurs when in the IR an instability of the vacuum $\langle \rho \rangle = 0$ drives a transition to $\langle \rho \rangle \neq 0$. In terms of the RG improved effective potential, this instability arises because at some energy scale Λ_{IR} the running quartic coupling vanishes. Below Λ_{IR} the quartic is negative definite, and a near-by minimum appears. Neglecting scheme-dependent $\mathcal{O}(1)$ coefficients, the canonically normalized tree-level RG improved version of the effective potential (11.3.27) is

$$\frac{1}{Z(\rho)^2} V_{\text{eff}}^{\text{RG}}(\rho) = \frac{\lambda(t)}{4!} \rho^4. \quad (11.4.23)$$

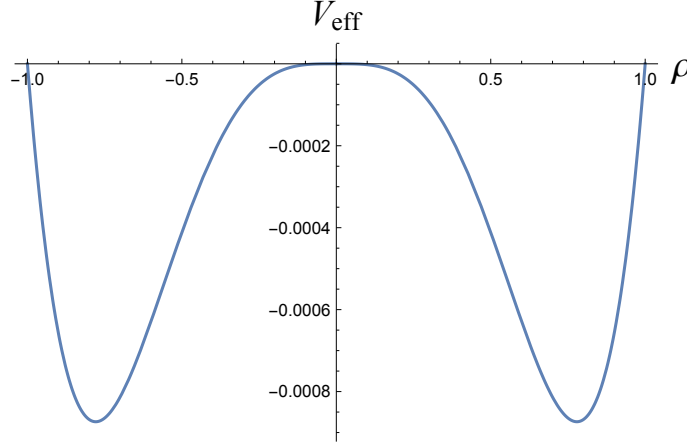


Figure 11.2: The RG improved effective potential in terms of canonically normalized fields (denoted for simplicity just V_{eff} in the figure) in the IR as a function of the field VEV ρ , for field values close to the scale Λ_{IR} where the quartic coupling vanishes, in units of Λ_{IR} .

Let us denote by t_{IR} the scale where $\lambda = 0$. For $t \sim t_{\text{IR}}$, no large logs can develop, and we can expand $\lambda(t)$ in eq.(11.4.23):

$$\frac{1}{Z(\rho)^2} V_{eff}^{\text{RG}}(\rho) \approx \frac{\lambda(t_{\text{IR}}) + (t - t_{\text{IR}})\beta_\lambda(t_{\text{IR}})}{4!} \rho^4 = \frac{9 e_{\text{IR}}^4}{192\pi^2} \rho^4 \log\left(\frac{\rho^2}{\Lambda_{\text{IR}}}\right). \quad (11.4.24)$$

Independently of e_{IR} , two minima develop at $\rho = \pm e^{-1/4} \Lambda_{\text{IR}}$, see fig.11.2 for an illustration. Since ρ is positive definite, the actual minimum is at $\rho = e^{-1/4} \Lambda_{\text{IR}}$. The expression (11.4.23) is an example of the phenomenon discussed at the end of section 5.11. The tree-level RG improved potential at face value would depend on the field anomalous dimension, which in turn would give rise to a gauge dependence of the potential, as shown in the one-loop potential (11.2.6). On the other hand, in the canonical basis and neglecting scheme-dependent terms like the factor proportional to 25/6 in eq.(11.2.6), the potential is just given by eq.(11.4.23) and is manifestly gauge invariant.

Bibliography

- [1] L. F. Abbott. The Background Field Method Beyond One Loop. *Nucl. Phys.*, B185:189–203, 1981.
- [2] Allan Adams, Nima Arkani-Hamed, Sergei Dubovsky, Alberto Nicolis, and Riccardo Rattazzi. Causality, analyticity and an IR obstruction to UV completion. *JHEP*, 10:014, 2006.
- [3] Stephen L. Adler. Axial vector vertex in spinor electrodynamics. *Phys. Rev.*, 177:2426–2438, 1969.
- [4] Thomas Appelquist and J. Carazzone. Infrared Singularities and Massive Fields. *Phys. Rev.*, D11:2856, 1975.
- [5] Jonathan Bagger and Carl Schmidt. Equivalence Theorem Redux. *Phys. Rev.*, D41:264, 1990.
- [6] Tom Banks and A. Zaks. On the Phase Structure of Vector-Like Gauge Theories with Massless Fermions. *Nucl. Phys.*, B196:189–204, 1982.
- [7] C. Becchi, A. Rouet, and R. Stora. Renormalization of the Abelian Higgs-Kibble Model. *Commun. Math. Phys.*, 42:127–162, 1975.
- [8] C. Becchi, A. Rouet, and R. Stora. Renormalization of Gauge Theories. *Annals Phys.*, 98:287–321, 1976.
- [9] J. S. Bell and R. Jackiw. A PCAC puzzle: $\pi^0 \rightarrow \gamma\gamma$ in the σ model. *Nuovo Cim.*, A60:47–61, 1969.
- [10] N. N. Bogoliubov and O. S. Parasiuk. On the Multiplication of the causal function in the quantum theory of fields. *Acta Math.*, 97:227–266, 1957. (in German).
- [11] J. Bros, H. Epstein, and V. Glaser. A proof of the crossing property for two-particle amplitudes in general quantum field theory. *Commun. Math. Phys.*, 1(3):240–264, 1965.
- [12] Curtis G. Callan, Jr. Broken scale invariance in scalar field theory. *Phys. Rev.*, D2:1541–1547, 1970.
- [13] Curtis G. Callan, Jr., Sidney R. Coleman, J. Wess, and Bruno Zumino. Structure of phenomenological Lagrangians. 2. *Phys. Rev.*, 177:2247–2250, 1969.
- [14] William E. Caswell. Asymptotic Behavior of Nonabelian Gauge Theories to Two Loop Order. *Phys. Rev. Lett.*, 33:244, 1974.

- [15] Michael S. Chanowitz and Mary K. Gaillard. The TeV Physics of Strongly Interacting W's and Z's. *Nucl. Phys.*, B261:379–431, 1985.
- [16] Sidney R. Coleman. There are no Goldstone bosons in two-dimensions. *Commun. Math. Phys.*, 31:259–264, 1973.
- [17] Sidney R. Coleman and Erick J. Weinberg. Radiative Corrections as the Origin of Spontaneous Symmetry Breaking. *Phys. Rev.*, D7:1888–1910, 1973.
- [18] Sidney R. Coleman, J. Wess, and Bruno Zumino. Structure of phenomenological Lagrangians. 1. *Phys. Rev.*, 177:2239–2247, 1969.
- [19] John C. Collins. *Renormalization*, volume 26 of *Cambridge Monographs on Mathematical Physics*. Cambridge University Press, Cambridge, 1986.
- [20] John M. Cornwall, David N. Levin, and George Tiktopoulos. Derivation of Gauge Invariance from High-Energy Unitarity Bounds on the s Matrix. *Phys. Rev.*, D10:1145, 1974. [Erratum: *Phys. Rev.*D11,972(1975)].
- [21] R. E. Cutkosky. Singularities and discontinuities of Feynman amplitudes. *J. Math. Phys.*, 1:429–433, 1960.
- [22] Bryce S. DeWitt. Quantum Theory of Gravity. 2. The Manifestly Covariant Theory. *Phys. Rev.*, 162:1195–1239, 1967. [,298(1967)].
- [23] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.
- [24] L. D. Faddeev and V. N. Popov. Feynman Diagrams for the Yang-Mills Field. *Phys. Lett.*, 25B:29–30, 1967.
- [25] Marcel Froissart. Asymptotic behavior and subtractions in the Mandelstam representation. *Phys. Rev.*, 123:1053–1057, 1961.
- [26] Murray Gell-Mann, M. L. Goldberger, and Walter E. Thirring. Use of causality conditions in quantum theory. *Phys. Rev.*, 95:1612–1627, 1954.
- [27] Murray Gell-Mann and F. E. Low. Quantum electrodynamics at small distances. *Phys. Rev.*, 95:1300–1312, 1954.
- [28] J. Goldstone. Field Theories with Superconductor Solutions. *Nuovo Cim.*, 19:154–164, 1961.
- [29] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken Symmetries. *Phys. Rev.*, 127:965–970, 1962.
- [30] V. N. Gribov. Quantization of Nonabelian Gauge Theories. *Nucl. Phys.*, B139:1, 1978.
- [31] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964.
- [32] W. Heisenberg and H. Euler. Consequences of Dirac's theory of positrons. *Z. Phys.*, 98(11-12):714–732, 1936.

- [33] Klaus Hepp. Proof of the Bogolyubov-Parasiuk theorem on renormalization. *Commun. Math. Phys.*, 2:301–326, 1966.
- [34] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.
- [35] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.
- [36] P. C. Hohenberg. Existence of Long-Range Order in One and Two Dimensions. *Phys. Rev.*, 158:383–386, 1967.
- [37] J. Iliopoulos, C. Itzykson, and Andre Martin. Functional Methods and Perturbation Theory. *Rev. Mod. Phys.*, 47:165, 1975.
- [38] R. Jackiw. Functional evaluation of the effective potential. *Phys. Rev. D*, 9:1686, 1974.
- [39] Gunnar Kallen. On the definition of the Renormalization Constants in Quantum Electrodynamics. *Helv. Phys. Acta*, 25(4):417, 1952.
- [40] Zohar Komargodski and Adam Schwimmer. On Renormalization Group Flows in Four Dimensions. *JHEP*, 12:099, 2011.
- [41] Benjamin W. Lee, C. Quigg, and H. B. Thacker. Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass. *Phys. Rev.*, D16:1519, 1977.
- [42] H. Lehmann. On the Properties of propagation functions and renormalization constants of quantized fields. *Nuovo Cim.*, 11:342–357, 1954.
- [43] H. Lehmann, K. Symanzik, and W. Zimmermann. On the formulation of quantized field theories. *Nuovo Cim.*, 1:205–225, 1955.
- [44] H. Lehmann, K. Symanzik, and W. Zimmermann. On the formulation of quantized field theories. II. *Nuovo Cim.*, 6:319–333, 1957.
- [45] Andre Martin. Extension of the axiomatic analyticity domain of scattering amplitudes by unitarity. 1. *Nuovo Cim.*, A42:930–953, 1965.
- [46] N. D. Mermin and H. Wagner. Absence of ferromagnetism or antiferromagnetism in one-dimensional or two-dimensional isotropic Heisenberg models. *Phys. Rev. Lett.*, 17:1133–1136, 1966.
- [47] Yoichiro Nambu. Axial vector current conservation in weak interactions. *Phys. Rev. Lett.*, 4:380–382, 1960.
- [48] Giorgio Parisi. *Statistical Field Theory*. Frontiers in Physics. Addison-Wesley, 1988.
- [49] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.
- [50] H. David Politzer. Power Corrections at Short Distances. *Nucl. Phys.*, B172:349–382, 1980.

- [51] T. Regge. Introduction to complex orbital momenta. *Nuovo Cim.*, 14:951, 1959.
- [52] T. Regge. Bound states, shadow states and Mandelstam representation. *Nuovo Cim.*, 18:947–956, 1960.
- [53] Julian S. Schwinger. On gauge invariance and vacuum polarization. *Phys. Rev.*, 82:664–679, 1951.
- [54] K. Symanzik. Renormalizable models with simple symmetry breaking. 1. Symmetry breaking by a source term. *Commun. Math. Phys.*, 16:48–80, 1970.
- [55] K. Symanzik. Small distance behavior in field theory and power counting. *Commun. Math. Phys.*, 18:227–246, 1970.
- [56] Gerard 't Hooft. Dimensional regularization and the renormalization group. *Nucl. Phys.*, B61:455–468, 1973.
- [57] Gerard 't Hooft and M. J. G. Veltman. Regularization and Renormalization of Gauge Fields. *Nucl. Phys.*, B44:189–213, 1972.
- [58] I. V. Tyutin. Gauge Invariance in Field Theory and Statistical Physics in Operator Formalism. 1975.
- [59] Steven Weinberg. New approach to the renormalization group. *Phys. Rev.*, D8:3497–3509, 1973.
- [60] Steven Weinberg. *The Quantum Theory of Fields*, volume 1-2. Cambridge University Press, 1995.
- [61] Kenneth G. Wilson. Renormalization group and critical phenomena. 1. Renormalization group and the Kadanoff scaling picture. *Phys. Rev.*, B4:3174–3183, 1971.
- [62] Kenneth G. Wilson. Renormalization group and critical phenomena. 2. Phase space cell analysis of critical behavior. *Phys. Rev.*, B4:3184–3205, 1971.
- [63] Kenneth G. Wilson and Michael E. Fisher. Critical exponents in 3.99 dimensions. *Phys. Rev. Lett.*, 28:240–243, 1972.
- [64] W. Zimmermann. Convergence of Bogolyubov's method of renormalization in momentum space. *Commun. Math. Phys.*, 15:208–234, 1969. [Lect. Notes Phys.558,217(2000)].