Hadron physics

Lecture in SS 2023 at the KFU Graz

Axel Maas

Contents

Chapter 1

Introduction

The strong nuclear force is of central importance both in our understanding of nature as well as in the understanding of field theory.

Its basic phenomenological importance is gained from the fact that nuclei are an important and indispensable part of our world. Understanding them is therefore of paramount relevance. Studying nuclei showed that they are bound states of the more elementary nucleons, the common name of protons and neutrons. The binding force is of considerable complexity, even if only viewed as the interaction creating these bound states. However, it was experimentally observed that there are many other particles, which interact through the same force. In a similar effort as with the periodic table of elements, it was understood that all of these particles, and several hundreds are known today, have an underlying structure. More are still discovered, with about seventy new ones in the last ten years. This underlying structure involves the quarks, which are bound together by gluons to form all particles interacting by this strong nuclear force: The hadrons.

This theory is called, for reasons to become clear later, quantumchromodynamics, or QCD, for short. While the theory can be formulated in the language of a quantum field theory, performing calculations are by far not trivial. The main reason is that, except at high energies, QCD is not accessible to perturbative methods, but requires a full non-perturbative treatment. Even with decades of development, this turned out to be a formidable task, still not solved. Hence, though some quantities can be calculated with reasonable accuracy, even the calculation of the average of proton and neutron mass required almost 40 years.

This already underlines the second important property of QCD. So far, it is the only accessible theory of this type. Given the problems encountered with it, it is clear that similar problems will emerge with any kind of (gauge) theories which require a non-perturbative treatment. It is therefore the test case to develop methods, as only here sufficient experi-

mental results are available to provide reliable checks of the calculations.

Given the vastness of complications in QCD, it is not surprising that the literature is similarly vast. Thus, there is no unique choice of textbook. For the sake of completeness, the following books have been used in the preparation of this lecture:

- Böhm et al., Gauge theories
- Cheng & Li, Gauge theory of elementary particle interactions
- Dissertori et al., QCD
- Peskin & Schröder, Quantum field theory
- Weinberg, Quantum field theory I and II
- Yndurian, The theory of quark and gluon interactions

Besides these, a number of various recent review articles, available from the arxiv, arxiv.org, are available, in particular 1404.3723. It is also recommendable to have a look at the most recent reviews in the Physics Reports journal on the topic. Also, as always, the PDG at pdg.lanl.gov is an essential repository.

Chapter 2

Basic phenomenology

As a first step, in the following the phenomenology of the strong interactions will be collected, establishing a repository of names and concepts. In the remainder of the lecture the underlying field theory will be developed, along with suitable techniques to deal with it. Ultimately, it will be sketched how the presented phenomenology emerges from the theory.

2.1 Nuclear physics

As noted, all atomic nuclei consists out of differing numbers of protons and neutrons, the former having a positive electric charge of the same size as the electron and the latter no electric charge at all. The simplest atomic nucleus is a single proton. However, single neutrons are not stable, but decay in a few minutes, and are therefore only appearing bound in nuclei.

Experimentally, this has been found by scattering experiments. When scattering with an electromagnetic probe, e. g. an electron, the characteristic quantity is the total momentum transfer $-q^2 = Q^2$ between the electron and the nuclei, usually transmitted by a photon. This defines furthermore the fraction of momentum

$$x = \frac{Q^2}{2pq},$$

where p is the momentum of the target. In the rest frame of the target the denominator becomes Mq_0, where the transferred energy q_0 is also often called the energy transfer ν. The variable x is called the Bjorken variable, or Bjorken-x, for short. These variables will also play an important role later when discussing hadronic collisions. Since in a two-body collision at fixed energy Q^2 is entirely characterized by a scattering angle, a description in terms of x is equally well suited.

In the case of elastic scattering, x becomes one. Hence, the crosssection as a function of x will be just a single peak at 1. Increasing the energy transfer into the inelastic regime, x becomes smaller than one. Then, additional peaks will appear, which correspond to excited states of the nuclei. Such excited states already show the presence of an internal structure, together with the finite extension of the nuclei. In addition, around $x \approx 1/A$, with A the number of nucleons in the nuclei, there is a broad peak. This broad peak is due to the elastic scattering with the nucleons in the nuclei. The broadness comes from the Fermi motion, i. e. from the fact that a nucleon confined to a nuclei has, due to the uncertainty relation, at least a momentum of roughly $1/R$, where R is the nuclei radius. With $R \sim 1 - 10$ fm, these are energies of order 100 MeV, and therefore not negligible on nuclear energy scales.

Because of electromagnetic repulsion it became quickly clear that the nuclear force must be very much stronger than QED to create quite compact nuclei, not much larger than the protons and neutrons, with more than one proton. It must also act on different charges, as the neutron is electrically neutral. One additional observation was made, when investigating nuclei and the nuclear force. It was not only because of its strength very different from QED and gravity, but it was also of a finite range. Though being much stronger than electrodynamics, it dropped almost to zero within a very short range of a few Fermi.

How can such a short-range force emerge? One suggestion comes from the relativistic description of bosons. Take the Klein-Gordon equation for a scalar boson of mass m,

$$(\partial^2 + m^2)\psi = 0, \tag{2.1}$$

with the wave-function ψ as the solution. Investigating a static, spherically symmetric situation yields

$$\psi(\vec{r}) = \frac{g}{4\pi|\vec{r}|} e^{-m|\vec{r}|}, \tag{2.2}$$

where g is an integration constant. Such a potential is called a Yukawa potential.

Thus, the wave-function decays exponential towards large distances, where the characteristic distance of propagation is given by the inverse mass of the particle. Hence, a short-range interaction can arise if the exchange particles are massive. This lead to the prediction of an exchange boson for the nuclear force with a mass of the order of roughly hundred MeV, given the range of the nuclear force of about a few fm.

The interpretation of the integration constant g becomes clear when the zero-mass limit is considered. In this case the wave-function becomes Coulomb-like, with g corresponding to the fine-structure constant. Thus, also in the massive case g can be interpreted as a coupling constant, which characterizes the strength of the strong nuclear force. It turns

out that characteristic values for it are about two orders of magnitude larger than for the electromagnetic force.

The Yukawa potential yields the right idea. Instead of having massless photons as force particles, massive particles must mediate the strong nuclear force. They were indeed found in the form of the mesons.

2.2 Mesons

While the protons and neutrons are fermions with spin 1/2, the force carrier of the nuclear force were identified to be actually bosons. Bosonic particles, which are affected by the strong nuclear force, are collectively called mesons, while such fermions are called baryons. Together, they are denoted as hadrons. Therefore, protons and neutrons are baryons.

The lightest of the mesons are the pions with quantum numbers $J^P = 0^-$, i. e. they are pseudoscalars. They come as a neutral one, π^0, and two oppositely charged ones, $\pi^\pm$. The range of the nuclear force is about 1 fm, which indicates that the force carrier, according to the Yukawa potential (2.2), should have a mass around 100-200 MeV. Indeed, the pions are found to have masses of 135.0 and 139.6 MeV for the uncharged and charged ones, respectively, and are thus much lighter than either protons or neutrons. These pions are not stable, but decay either dominantly electromagnetically into photons for the neutral one or like the neutron for the charged ones. Their life-time is of the order of 10^{-8} seconds and 10^{-17} seconds for the charged and uncharged ones, respectively. Therefore the charged ones live long enough to be directly detectable in experiment.

One of the surprises is that the neutral one decays into two photons, as usually photons are expected to couple only to electromagnetically charged objects. While this can be thought of as a neutral pion virtually splitting into two charged pions, and then annihilation under emission of photons, this is somewhat awkward. A more elegant resolution of this will be given in the quark model below in section 2.4.

With these pions it was possible to describe the overall properties of nucleons, especially long-range properties. At shorter range and for finer details it turned out that a description only with pions as force carriers was impossible. This was resolved by the introduction, and also observation, of further mesons. Especially the vector[1] meson ρ with a mass of 770 MeV, spin one, and a very short life-time of roughly 10^{-24} seconds and the vector meson ω with a mass of about 780 MeV, but with a 20 times longer life-time than the ρ, play an important role. This larger number of mesons is also at the core of apparent three-body forces observed in nuclear interactions, which are, e. g., necessary to describe deuterium

[1] Vector mesons are mesons with spin 1.

adequately. In fact, many more mesons have been discovered, and some more will appear later.

Describing how these various mesons create the strong nuclear force is in detail very complicated, but it in principle can be systematically performed, e. g. in the form of the so-called chiral perturbation theory. This will be detailed more in section 5.2.6. What is remarkable is that out of nowhere appeared several different mesons, all contributing to the nuclear force, and actually all of them also affected by the nuclear force. Such a diversity of force carriers is distinctively different from the case of QED, where only the photon appears.

2.3 Isospin and baryons

In the endeavor to find the carriers of the nuclear force, several other observations have been made. The first is that most nuclear reactions show an additional approximate symmetry, the isospin symmetry. The name is a historical relic, which stems from the fact that the underlying symmetry group is SU(2).

This symmetry is manifest in the almost degenerate masses of the proton and the neutron. Both particles can therefore be considered to be a doublet of this new symmetry. It is furthermore found that also the three pions fit into this scheme. This new symmetry is called isospin symmetry, and the nucleon form a doublet, while the pions a triplet. Similar, also the ρ and the ω can be fitted into this scheme. An interesting observation is that the value of the isospin is related to the electric charge as

$$Q = I_3 + \frac{B}{2} \tag{2.3}$$

if a new quantity, called baryon number, is introduced. It is one for the baryons and zero for all mesons. This is so far a phenomenological identification, but will become quite relevant in section 2.4. Of course, the anti-particle of the nucleons, the antiproton and antineutron, carry negative baryon number.

According to this rule, it is possible to attempt to construct a quadruplet, having four states with $I_3 = -3/2, -1/2, 1/2, 3/2$. To get integer charges, it must then have a baryon number, like the nucleons. These particles should therefore have electric charge -1, 0, $+1$, and $+2$, and corresponding anti-particles. These particles have been observed experimentally, and again the different states have almost the same mass. They are called Δ, have masses of about 1232 MeV, and are fermions, as are the nucleons. However, their spin is 3/2. In fact, they are not the only further baryons, and many more have been found experimentally.

2.4 The quark model

The number of baryons and mesons found by now numbers several hundreds. Already decades ago, when only a few dozens were known, it appeared unlikely that all of them should be elementary. This was very quickly confirmed by experimental results which showed that the proton had a finite size of about 1 fm, and Rutherford-like experiments found that there are scattering centers inside the proton, which appeared point-like. In fact, similar to the case of the nuclei, the experiments showed, to leading order, that the number of constituents in baryons is three, and in mesons it is two. However, beyond leading order, things were quite different. In the case of nuclei, as long as the energy transfer Q^2 is not large enough to disintegrate the nuclei, the number of constituents is essentially constant, and the elastic peak of the sub-structure scattering is essentially always at $1/A$, independent of Q^2. This turned out not to be the case for the nucleons. Depending on Q^2, the peak started to shift, and an increase is found towards small x, even though some peak structure remains at a third and a half, for baryons and mesons, respectively. This would indicate the presence of many more constituents, called partons, if the nucleon is probed at large energies. This kind of so-called scaling violation cannot be obtained from an interaction like (2.2), but requires something new.

In addition, in contrast to the nuclei these constituents were not almost isolated in essentially free space, but very tightly packed. Furthermore, while the neutron is electrically neutral, it was found to have a magnetic dipole moment, a feature beforehand believed to be only existing if there is an electrically charged substructure present.

This evidence together suggested that the elementary particle zoo could possibly be obtained from simpler constituents and put into a scheme like the periodic table of chemical elements, which originates from a few different particles.

Playing around with quantum numbers showed a number of regular features of the hadrons. This gave rise to the quark model. This model introduces as the particles making up hadrons the so-called quarks. In the beginning two quarks were needed to explain the regularities observed, the up quark and down quark, abbreviated by u and d, as well as their anti-particles. Since both the bosonic mesons and fermionic hadrons must be constructed from them, it requires them to be fermions themselves. Since all of the hadrons have an extension, none of them can be identified with a single quark, just like the periodic table does not contain a single proton, neutron or electron. However, in contrast to the latter no free quarks are observed in nature, a phenomenon to be discussed in section 5.5.

The simplest possibility to construct then a hadron would be from two quarks. This must be a boson, as two times a half-integer spin can only be coupled to an integer spin, and

therefore a meson. Since no free quarks are seen, the nucleons must contain at least three quarks to get a half-integer spin. These considerations turn out to be correct. However, they lead to the conclusion that quarks cannot have integer electric charges. This is most easily seen by looking at the nucleons.

Furthermore, scattering experiments yielded that the nucleons have no uniform substructure, but have a two-one structure, that is two quarks of one type and one quark of the other type. Since it is found that the down quark is heavier than the up quark, the heavier one, i. e. the neutron, should have two down quarks. This yields a composition of uud for the proton and udd for the neutron. The only solution for the observed electric charges of the proton and the neutron are then an assignment of $2/3$ of the (absolute value of the) electron charge for the up quark, and $-1/3$ of the (absolute value of the) electron charge for the down quark. This consistently yields the required positive and zero proton and neutron electric charges, respectively. This also explains the magnetic dipole moment of the neutron.

At the same time, the baryon number of quarks must be $1/3$ for up quark and down quark. This implies also that the isospin of the up quark is $+1/2$ and that of the down quark is $-1/2$. The isospin of hadrons can then be obtained by a tensor product of the isospin of the quarks, which belong to the fundamental representation of the SU(2) isospin group. Two quarks can therefore either form a singlet or a triplet, while three quarks can form either a doublet or a quadruplet. All of these particles have been seen in the previous sections already.

The pions are then constructed as a combination of a quark and an anti-quark, $u\bar{d}$ for the π^+, $\bar{u}d$ for the π^-, and a mixture of $\bar{u}u$ and $\bar{d}d$ for the π^0, i. e. the state of the π^0 is

$$\left|\pi^0\right\rangle = \cos\alpha \left|\bar{u}u\right\rangle + \sin\alpha \left|\bar{d}d\right\rangle,$$

where α is a mixing angle. This mixing angle can be determined experimentally or theoretically. Experimentally, this is possible by using the different decay patterns for different mixing angles. E. g., because of the different electric charges of up and down, the electromagnetic decays of the π^0 should depend on the mixing angle. Theoretically, by calculating the left-hand side and the right-hand side in different bases. However, in practice both possibilities are highly challenging. An assignment of two quarks instead of a quark and anti-quark is not possible, as this cannot give the required baryon number of zero.

Finally, to obtain reasonable agreement with experiment requires a quite different effective interaction of two quarks, a potential of the type

$$V(r) = \frac{a}{r} + br, \tag{2.4}$$

the so-called Cornell potential. I. e. in addition to the Coulomb term $1/r$ an additional term, which rises linearly in distance, something unseen before QCD. Its origin not fully resolved, though the potential itself is well established. This will be discussed at length in section 5.5.

Particles like the ρ meson are then also combinations of a quark and anti-quark, but where the quarks have relative orbital angular momentum, creating their total spin of one. The Δ, however, turns out to pose a serious challenge.

2.5 Color and gluons

At first glance, the Δ appears simple enough. The double-positive state Δ^{++} is just three up quarks, and with decreasing charge always one up quark is replaced by one down quark, until reaching the Δ^- with three down quarks. To obtain the observed 3/2 spin requires to align the spins of all three quarks. Of course, it could be possible that there would be a relative orbital angular momentum, but experimentally this is not found. In fact, there exists an excited version of the Δ with such an orbital angular momentum and total angular momentum of 5/2, which is also experimentally confirmed.

And this is, where the problem enters. Since the Δ is a fermion, its wave-function must be totally antisymmetric. Since the spins are aligned and all three quarks are of the same type in the ground-states of Δ^{++} and Δ^-, no wave-function can be constructed which is anti-symmetric. Thus the existence of the Δ appears to violate the Pauli principle at first sight. But this is not so. Originally introduced to resolve that problem, and later experimentally verified, another conserved quantum number is attached to quarks: Color. The wave-function can then be anti-symmetric in this new quantum number, saving the Pauli principle and the quark model at the same time.

Since this new quantum number of the quarks is not observed for the Δ, or any other hadron, the hadrons must all be neutral with respect to this new quantum number. For the mesons, consisting of a particle and an anti-particle, this is simple enough, as just both have to have the same charge. This is not the case for baryons. Assigning just positive or negative charges, like the electrical charge, it is not possible to construct neutral states out of three particles. Attempts to do so with fractional charges also do not succeed in the attempt to make the proton and neutron color-neutral simultaneously. It is therefore necessary to depart from the simple structure of the electromagnetic charge.

As a consequence, it is assumed that there are three different charges, suggestively called red, green (or sometimes, especially in older literature, yellow), and blue. It is furthermore assumed that not only a color and the corresponding anti-color is neutral,

but also a set of one of each of the colors is neutral. Then there are three quarks for each flavor: red, green, and blue up quarks, and red, green, and blue down quarks, totaling six quarks. A color-neutral baryon is containing a quark of each color, e. g. a proton contains a red and a blue up quark, and a green down quark. In fact, since the total color charge of a proton is zero, it is a mixture of any possible combination of color assignments to each three quarks, which are consistent with neutrality and the Pauli principle. Similar, the Δ^{++} consists of a red up quark, a green up quark, and a blue up quark.

One other important ingredient, now that there is a new charge, is what mediates the force between the charges. In electromagnetism it was the massless photon. It is therefore reasonable to assume that there is also a mediator of the force between color charges. These were indeed found, and named gluons. As the photon these are massless[2] bosons with spin one. However, they differ from photons in a very important property. While photons are only mediating the electromagnetic force, they are not themselves affected by it, since they carry no electric charge. But gluons carry color charge. In fact there are 8 different charges carried by gluons, and none of these eight are either the quark charges, nor is there any simple relation to the quark charges. Especially, it is impossible to add a single quark charge with any combination of the gluon charges to obtain a neutral object. To achieve this, at least two quarks have to be added to one or more gluons.

Nonetheless, the idea of gluons has been experimentally verified, and they have been identified as the carrier of the strong interaction, binding quarks into color-neutral hadrons. The exchange of mesons to bind nucleons into nuclei can be viewed as a higher-order effect of the gluon interaction. This is similar to a Van-der-Waals force, though the details are very different, as here not a color dipole moment enters, and the details are not yet fully resolved. Still, the interaction of nucleons in a nuclei can be traced back to the gluons.

Hence, the combination of quarks, gluons, and colors can explain the structure of all known hadrons, similar to the periodic table. Unfortunately, the strong force binding quarks by gluon exchange is not accessible using perturbation theory, at least when it comes to describing hadrons. Its treatment is therefore highly non-trivial. Because of the color, this underlying theory of hadrons is called chromodynamics, its quantum version quantum chromodynamics, or QCD in brief.

2.6 Strangeness

So far two quark flavors, up and down, have been introduced. This implies the presence of an approximate SU(2) flavor symmetry, since the masses of both quarks are similar. It

[2]The notion of massless is murky here, an will be returned to later.

was possible to describe all hadrons introduced so far using just these two quarks with the quark model.

However, already before QCD was formulated, hadrons were observed, which did not fit into this picture. The most well-known of them are the kaons $K^{\pm}$, K^0, and $\bar{K}^0$, four mesonic hadrons of masses 494 MeV for the charged ones and 498 MeV for the two uncharged ones. Most remarkably, these new mesons were more stable than those of similar masses made from the two quarks inside the quark model.

The reason is that there are more than the two quark flavors necessary to construct the proton and neutron. These additional quark flavors do not occur in naturally observed atomic nuclei[3] but can be produced in natural or artificial collisions, so-called hypernuclei.

The quark to obtain the kaons in the quark model has been called the strange quark s. The s quark has an electric charge of $-1/3$, just like the d quark. The charged kaons are therefore the combinations $u\bar{s}$, $\bar{u}s$, and the neutral ones $d\bar{s}$ and $\bar{s}d$, explaining their small mass difference, and their multiplicity. Two more such mesons are the η and η' mesons, which are made from $\bar{s}s$ combinations, and some admixtures from neutral combinations of u and d quarks. They are therefore even heavier, the η having a mass of 550 MeV. Somewhat peculiar, the mass of the η' is much higher, about 960 MeV. The reason is that the η' also receives mass from another source, the so-called axial anomaly. The latter will be discussed below in section 5.6.

This enlarges the isospin symmetry to the so-called flavor symmetry. Since it involves up, down, and strange quarks, it is an SU(3) group. Since the quarks have different masses, the group is not unbroken, but reduced to three counting symmetries, i. e. $U(1)^3$. Hence, the individual quark flavors are conserved, but bound states with differing quark content have differing masses. The isospin group is the almost unbroken SU(2) subgroup of this SU(3) flavor group.

This conservation of quark flavor by the strong interaction is also at the origin of the name strangeness. When the kaons were discovered, the quark model was yet to be established. The kaons, and also strange baryons, called hyperons, with a single or more strange quarks included, showed a different decay patterns than ordinary hadrons, due to the conservation of strangeness. Thus, they did not fit into the scheme, and were therefore considered strange.

The presence of the strange quark adds many more possible combinations to the quark model, which all have very similar masses. Thus, there is a very large number of hadrons with masses between 500 MeV, and roughly 3000 MeV, where the states become too unstable to be still considered as real particles. In fact, the number of states N turns out

[3]With the possible exception of the inner core of neutron stars, though this is not yet settled.

to rise exponentially with mass, $N \sim \exp(M/T_H)$. This is a so-called Hagedorn spectrum, where T_H is called the Hagedorn temperature. The reason for the name is that naively a system with such a spectrum has the property that at a finite temperature, the Hagedorn temperature T_H, an infinite number of particles is created, and thus the system cannot be heated beyond this point. For the strong interactions, the Hagedorn temperature is about 160 MeV. Of course, it is in practice possible to go beyond this temperature. What happens is that at this point the quark substructure can no longer be ignored, and this effect limits the number of states growth. This has originally lead to the idea that at this temperature a phase transition has to occur, which signals the change from a hadronic system to one were the quark substructure becomes more important. However, the most reliable results to date rather indicate that it is a crossover. This transition plays nonetheless a role in the development of the early universe, though a rather small one. This will be discussed in detail in chapter 7.

2.7 Hybrids, glueballs, and other exotics

When studying the quark model and the possible color charges, it becomes quickly clear that three quarks or one quark and one anti-quark are not the only possibilities how to create color-neutral objects. Tetraquarks, made from two quarks and two anti-quarks, as well as pentaquarks, made from four quarks and one anti-quark, are equally possible. There is indeed no a-priori reason why such bound states should not exist. However, the are not yet any unequivocally observed tetraquarks or pentaquarks, though some states exist for which the evidence has become very strong, especially for tetraquarks. That an experimental observation is so complicated can be motivated theoretically by two arguments.

The first effect is mixing. E. g. for a tetraquark, it is almost always possible to construct a meson with the same quantum numbers, i. e. the same spin, parity, charge-parity and electric charge. There is also the possibility to construct equally well a dimeson molecule. One of the most infamous examples is the σ meson[4]. It is a light neutral meson, with quantum numbers compatible, e. g., with the states $\bar{u}u$, $\bar{d}d$, $\bar{u}u\bar{d}d$, $\bar{u}u\bar{u}u$, $\bar{d}d\bar{d}d$, $\pi^+\pi^-$, and $\pi^0\pi^0$. Since it is a quantum state, it follows the quantum rules, which in particular imply that all states with the same quantum numbers mix. It is therefore a superposition of all such states. The question which of these states contribute most is highly non-trivial. It can, in principle, be experimentally measured or theoretically calculated. There is no really reliable way of estimating it. The results found so far indicate that the combination

[4]The official name is $f_0(500)$, though the historical name of σ meson is still commonly used.

of two pions is most dominant, it is therefore likely a molecule. For most other states the two-quark components appears to be the dominant one. Similarly, almost all baryons turn out to be completely dominated by the three-quark state. There are only few cases, where this may not be the case.

The second possibility is to investigate one of those possibilities where the quantum numbers of the tetraquarks cannot be created using a two-quark system. Such cases are rare, but they exist. In principle, it would therefore be sufficient to just observe such a state. Unfortunately, almost all of these states are highly unstable. They are therefore experimentally hard to observe, and it is thus challenging to establish their properties beyond doubt. Only very few candidates have been found so far, but some of them, to be discussed later in section 2.8, appear very promising.

This problem becomes more complicated due to the gluons. Though it is not possible to create a color-neutral state from a quark and a gluon, it is possible to combine a quark, an anti-quark and one or more gluons to obtain a colorless state. It is similarly possible to combine three quarks and a number of gluons to obtain a colorless state. Such states are called hybrids. However, the gluons can at most add angular momentum, but no other charges to the state. Therefore, there is always a state with the same quantum numbers, but just made from quarks. Since adding a particle or orbital angular momentum to a state usually increases its mass, these states are unstable against the decay to a state with just the minimum number of quarks. Though these hybrids are therefore formally admixtures to any state, it is essentially always a small one, and therefore hybrids are very hard to identify both experimentally and theoretically.

The last class of states which can come into the mix are bosonic glueballs, which combine only gluons to a colorless objects. The usual counting rules of the quark model do not apply to them, but as a rough estimate even the simplest state is made out of four gluons. Such states carry no electric charge, and most of them have the same J^{PC} quantum numbers as mesons, and therefore mix. However, there are some candidates, particularly the so-called f(x) mesons, with x around 1500 MeV, which appear to have a large admixture from glueballs. This is experimentally identified by the possible decays. Since gluons are, in contrast to quarks, electrically neutral, decays into electrically neutral decay products, except for photons, should be favored if there is a large glueball contribution in the state. This has been observed, especially when comparing the partial decay widths of decays to uncharged particles to the one to photons. However, even at best these states are only partially glueballs.

There are some glueball states which have quantum numbers which cannot be formed by only quarks, at least not in the simple quark model. Unfortunately, all theoretical

estimates place theses states at masses of 2.5-3 GeV, and therefore far above any hadronic decay threshold. They are therefore highly unstable, and decay quickly. Hence, there is not yet any experimental evidence for them, though new dedicated searches are ongoing or are prepared.

In the search for the above listed so-called exotic states strange quarks play an important role in the quest for tetraquarks and pentaquarks. Since strangeness is a conserved quantum number of the electromagnetic and strong interaction, it is possible to construct states which do not have the quantum numbers of an ordinary state, e. g. a meson with total strangeness of 2, which therefore must contain two strange quarks, and cannot be a simple quark-anti-quark state, or a baryon with strangeness -1, due to a single anti-strange quark, which therefore must be a pentaquark. Searches for such signatures are intensively pursued. Note that, e. g., tetraquarks could also consist out of two strange and two antistrange quarks. In this case, the excess strangeness is only visible in the decays. Such states are therefore called cryptoexotic.

2.8 Charm, bottom, and top

It appears at first rather surprising that there should be just one other quark, which has the same electric charge as the down quark. This appears unbalanced, and another quark with the electric charge of the up quark appears to be necessary. Indeed, this is correct, and there is also a heavier copy of the up quark, which is called charm[5] quark c . However, while the strange quark has a rather similar mass, despite its larger current mass, as the light quarks, the charm quark has not. Hadrons involving a charm quark are much heavier than hadrons containing only the lighter quarks. Charm is again a conserved quantum number in both the electromagnetic and strong interaction.

This conservation of charm has very interesting consequences. Of course there are hadrons were only one of the quarks is a charm quark, which are called open charm hadrons. The best known ones are the D mesons, with masses of about 1870 MeV mass and having the structure of a single charm quark and either an up or down quark. These are the lightest particles with a charm quark.

But there are also particles, especially mesons, which consists only of charm quarks. In the meson case, where the total charm is zero if they consist out of a charm and an anti-charm quark, these are called hidden charm. The latter states are particularly interesting, because they show a very interesting mass spectrum. In fact, the lightest $\bar{c}c$ states have

[5]The name originates from the fact that it solves several experimental mysteries observed in the weak interactions, and because at that time it appeared to complete the quark picture.

a mass which is quit a bit below threshold for the decay into two hadrons with a charm quark and an anti-charm quark each, the $\bar{D}D$ threshold. They can therefore not decay directly. Of course, the charm and anti-charm quarks can annihilate. But because of how quark and gluon color charges are arranged, such a process is substantially suppressed in QCD compared to the decay with a production of an additional quark-anti-quark pair: Because a gluon carries color, an annihilation in one gluon is impossible. In two gluons it is not possible because of spin-parity conservation, and into three gluons is already strongly suppressed. This is the so-called Okubo-Zweig-Iizuka (OZI) suppression, or sometimes just Zweig rule.

Hence, decays occur very slowly. Therefore, these hadrons are extremely stable compared to hadrons made from lighter quarks, where the pions offer a simpler decay channel. Thus, these meta-stable states carry the name of charmonia, which have masses of about 3 GeV, but decay widths of around a few 100 keV.

Because of this fact, the charmonia states turn out to present a very good realization of the possible states permitted by the potential (2.4). Similarly to the hydrogen atom, this potential creates states distinguished by a main quantum number and orbital quantum numbers. The most well-known state is the J/Ψ, at about 3097 MeV with a decay width of 93 keV, which is a state with one unit of angular momentum. However, the ground state of the system is the η_c, with a mass of 2984 MeV and a decay width of 320 keV. That the ground state decays quicker is mainly due to kinematic effects from the angular momentum. Simply put, the ground state is in an s-wave, and thus the wave-functions of the two charm quarks have a large overlap. Thus an annihilation into photons is much more likely than in the case with angular momentum, where the overlap of the wave functions is much smaller. Right now about 8 states are known, which are below the $\bar{D}D$ threshold, the heaviest the so-called $\psi(2S)$, with a mass of 3690 MeV and a decay width of 303 keV.

These charmonia have been very instrumental in understanding the potential (2.4), and thus the strong interactions. The very well-defined spectrum, which provides the opportunity of a true spectroscopy, including many angular momentum states[6], permits a much cleaner study than in case of the light hadrons, where the ubiquitous decays into pseudo-Goldstone bosons make resonances decay very quickly.

However, not all of the states in this spectrum are easily explained within the framework of the quark model and the potential (2.4). These are the so-called X, Y, and Z mesons, with masses above the $\bar{D}D$ threshold, and some also with open charm. These states do

[6]Angular momentum is, strictly speaking, not well-defined in a Lorentz-invariant manner. What is meant here is the angular momentum in the quark model in the rest frame of the particle.

not fit into the quark model spectroscopic scheme, and especially some may have quantum numbers, which are not in agreement with a simple quark-anti-quark system. This is still under experimental and theoretical investigation. At the same time, several states expected from the quark model are either missing, or substantially shifted compared to the quark model predictions. These absent or shifted states cannot explain the X, Y, and Z mesons, as they are all predicted to be very broad, while the latter are very narrow, of order a few MeV, while the others are expected at tens of MeV. This already shows that the simple quark model is not able to explain all hadrons.

With the charm quark, it may appear that everything is complete and symmetric. However, nature did not decide to stop at the charm quark, but added another quark: The bottom (or in older texts beauty) quark. It is a heavier copy of the down quark, and has therefore an electric charge of -1/3. Its mass is about three times that of the charm quark, with a current mass of 4.2 GeV. It therefore introduces another quark flavor, but like for the charm quark, it does not play any role at low energies.

Other than the mass and the electric charge, the bottom quark behaves essentially as the charm quark. Especially, there is a rich spectroscopy with open and hidden beauty[7], the latter also called bottonium in analogy to charmonium. Similar to the case for the charm quark, the lightest meson with open beauty is rather heavy, $B^{\pm}$ and B^0 being at 5.3 GeV. As a consequence, the bottonium spectrum has a large number of quasi-stable states, the lightest being the η_b with a mass of 9.4 GeV and a decay width of roughly 10 MeV, the Υ playing the role of the J/ψ with a mass of 9.5 GeV and a width of 54 keV, and then even 15 states to the heaviest $\chi_b(3P)$ with 10.5 GeV observed so far. There are also heavier states, including bottom versions of the X, Y, and Z mesons, which do not fit easily into a simple quark model explanation. Thus, an even richer spectroscopy is possible, though the production of these bottonia in so-called beauty farms requires more resources than for the charmonia.

Of course, for both bottom quarks and charm quarks there exist also baryons, with one or more of these quarks, also with both charm and bottom quarks. These are rather complicated to produce, but have been observed, though baryons with multiple charm or bottom quarks only very recently. These baryons are not as stable as the mesons, but are still sufficiently stable that their production and decay take place so far apart from each other, a few μm, that both processes can be experimentally resolved. They play therefore an important role to identify the production of charm and especially bottom quarks in high-energy collisions (so called c- and b-tagging).

Together, charmonia and bottonia are usually referred to as quarkonia. Studying theses

[7]For the flavor quantum number the term beauty still survives.

states is also interesting for other reasons than to understand QCD. Because the J/ψ and Υ are very long-lived, they are very well suited for precision measurements.

With the introduction of the bottom quark the situation appears again as unbalanced as with the introduction of the strange quark. This is indeed true, and the picture is extended by the top (or in old texts truth) quark t. This is the last quark, which so far has been found, though there is no a-priori reason to expect that there may not be further quarks, and searching for them is a very active field in experimental physics.

The top quark is a heavy version of the up quark, and thus has an electric charge of 2/3. The fact which is really uprising about it, is its mass of 173 GeV. It is therefore much heavier than any other quark or hadron. It is an interesting side remark, and a triumph of theory, that this mass has been established within 10 GeV before the direct observation of the top quark, by measuring other processes sensitive to the top quark very precisely, and then using theory to make a prediction.

The enormous mass of the top quark makes it the heaviest particle detected so far. Due to its large mass, it decays much quicker than the lighter quarks, with a decay width of 2 GeV, by weak effects. Hence, the formation even of short-lived hadrons with a top quark is almost impossible, and no hadron with a top quark has so far been directly observed. Whether a quasi-stable toponium is possible is not clear, but so far there is no experimental evidence for it. But, due to the mass, it is also not trivial to produce large numbers of top quarks, and thus the study of top quarks is rather complicated. Hence, the top quark is so far more of interest for its own properties, particularly its mass, rather than for its relation to QCD.

Chapter 3

Field theory

After having now the basic phenomenology for QCD available, it is now necessary to formulate it as a quantum field theory, i. e. a fully quantized theory. As it turns out, QCD has a number of particular features, and it is therefore not advisable to directly set out with a field-theoretical formulation of QCD. Rather, it is better to first gather the necessary ingredients.

3.1 Path integral

Though it is possible to perform canonical quantization for QCD, this is a rather cumbersome approach. A more elegant option is the path integral formalism, which is equivalent, at least at the theoretical physics level of rigor. Especially for gauge theories, which will turn out to be the relevant type of theories for QCD, it is the most convenient way of quantization.

3.1.1 Heuristic introduction

Since the path integral formulation is as axiomatic as is canonical quantization, it cannot be deduced. However, it is possible to motivate it.

A heuristic reasoning is the following. Take a quantum mechanical particle which moves in time T from a point a of origin to a point b of measurement. This is not yet making any statement about the path the particle followed. In fact, in quantum mechanics, due to the superposition principle, a-priori no path is preferred. Therefore, the transition amplitude U for this process must be expressible as

$$U(a, b, T) = \sum_{\text{All paths}} e^{i \cdot \text{Phase}}$$

which are weighted by a generic phase associated with the path. Since all paths are equal from the quantum mechanical point of view, this phase must be real. Thus it remains only to determine this phase. Based on the correspondence principle, in the classical limit the classical path must be most important. Thus, to reduce interference effects, the phase should be minimal for the classical path. A functional which implements this is the classical action S, determined as

$$S\left[C\right] = \int\limits_{C} dtL,$$

where the integral is over the given path C from a to b, and the action is therefore a functional of the path S and the classical Lagrange function L. Of course, it is always possible to add a constant to the action without altering the result. Rewriting the sum as a functional integral over all paths, this yields already the definition of the functional integral

$$U\left(a,b,T\right) = \sum_{C} e^{iS[C]} \equiv \int \mathcal{D}C e^{iS[C]}.$$

This defines the quantum mechanical path integral.

It then remains to give this functional integral a more constructive meaning, such that it becomes a mathematical description of how to determine this transition amplitude. The most useful approach so far for non-trivial interacting theories is the intermediate use of a lattice, i. e., a discretized space-time with a finite volume. However, even in this case there are still conceptual and practical problems, so that the following remains often an unproven procedure.

To do this, discretize the time interval T into N steps of size ε. Since any kind of path is permitted, it requires that all possible intermediate steps are admitted, even if the resulting path is non-differentiable or non-causal. In fact, it can be shown that the non-differentiable paths are the most important ones for a quantum theory. The reason for this is that there are many more non-differentiable paths than differentiable ones, and thus the differentiable ones are just an irrelevant subset. They are of measure zero.

Thus, at each time step, it is necessary to admit all positions in space. This yields

$$\int \mathcal{D}C = \int \mathcal{D}\left[\vec{r}\left(t\right)\right] = \int \frac{d^3\vec{r}_1}{M\left(\varepsilon\right)} ... \int \frac{d^3\vec{r}_N}{M\left(\varepsilon\right)},$$

as a discretization of the path integral, with some yet-to-be-determined integral measure $M(\varepsilon)$. It is furthermore assumed that particles move freely from time step n to $n+1$. To obtain the final expression, the limit $N \to \infty$ must be taken, implying that the path integral is an infinite number of ordinary integrals. Since the action is determined

classically, the phase can then be split into the phases for the individual time slices, and expanded to lowest order in ε. This gives a calculational prescription for the path integral.

Of course, when changing from a point-particle theory to a theory of a field ϕ, the corresponding action has to be used, which implies the replacement

$$\int dt L(x,t) \to \int d^d x \mathcal{L}(\phi(x,t)),$$

with the Lagrangian density $\mathcal{L}$. In this case, it is also no longer the paths of the particles over which it is integrated, but now it is necessary to integrate over all possible field configurations

A more detailed description of how to calculate this functional integral in quantum mechanics can be found elsewhere. Here, the main aim is the field theoretical case, in which the path integral reads

$$Z = \int \mathcal{D}\phi \exp\left(i \int d^d x \mathcal{L}(\phi)\right), \tag{3.1}$$

where Z is called the partition function. This is an integral over functions. Of course, in the same way it is possible to make a discretization for the field-theoretical case as for the quantum mechanical case, and this yields an operative definition of the path integral. To deal with it more elegantly requires some functional analysis, which will be discussed now in more detail.

3.1.2 Functional analysis

The following can be made mathematical more rigorous using the theory of distributions, in which functionals are defined by conventional integrals over appropriate test functions. However, this level of mathematical rigor is not necessary, and thus the following will be made as definitions. In general, under most circumstances in particle physics, this is sufficient. However, situations may arise, where it is necessary to go back to a more mathematical formulation.

The starting point before defining functional integration is the definition of a functional derivative. The basic ingredient for a functional derivative δ are the definitions

$$\begin{aligned}
\frac{\delta 1}{\delta\phi(x)} &= 0 \\
\frac{\delta\phi(y)}{\delta\phi(x)} &= \delta(x-y) \\
\frac{\delta}{\delta\phi(x)}(\alpha(y)\beta(z)) &= \frac{\delta\alpha(y)}{\delta\phi(x)}\beta(z) + \alpha(x)\frac{\delta\beta(z)}{\delta\phi(x)},
\end{aligned}$$

in analogy to conventional derivatives.

Consequently, a power series of a functional is defined as

$$F[\phi] = \sum_{n=0}^{\infty} \int dx_1 ... dx_n \frac{1}{n!} T(x_1, ..., x_n) \phi(x_1) ... \phi(x_n),$$

where the coefficients of an ordinary power series are now replaced by coefficient functions T. In particular, they can be obtained as

$$T(x_1, ..., x_n) = \frac{\delta^n}{\delta\phi(x_1) ... \delta\phi(x_n)} F[\phi]\Big|_{\phi=0}.$$

This defines the most important concepts for differentiation.

Concerning the functional integrals, they are as usually defined to be the inverse operation to functional derivatives. Therefore, integration proceeds as usual. In most practical cases, the relevant functional are either polynomials or can be expanded in a power series, and then functional integrals are straight-forward generalizations of the usual integrals. In particular

$$\begin{aligned} \int \mathcal{D}\phi &= \phi(x) \\ \int \mathcal{D}\phi\phi &= \frac{1}{2}\phi(x)^2, \end{aligned}$$

where the first expression implies that $\delta \int$ equals not to zero, but equals a δ-function.

Of particular importance are Gaussian integrals, i. e. the generalization of

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{\pi}} e^{-ax^2} = \frac{1}{\sqrt{a}}. \tag{3.2}$$

The result can be either obtained from the power series expansion or directly gleaned from the finite-dimensional generalization of Gaussian integrals, which is given by

$$\int_{-\infty}^{\infty} \frac{dx_1}{\sqrt{\pi}} ... \int_{-\infty}^{\infty} \frac{dx_n}{\sqrt{\pi}} e^{-x^T A x} = \frac{1}{\sqrt{\det A}},$$

with an arbitrary matrix A, though for a finite result the square-root of its determinant must be invertible, i. e., no zero eigenvalues may be present.

The functional generalization is then

$$\int \mathcal{D}\phi e^{-\int dxdy\phi(x)A(x,y)\phi(y)} = \frac{1}{\sqrt{\det A(x,y)}},$$

where A may now be operator-valued. Especially derivative operators will appear in this context later. The determinant of such an operator is given by the infinite product of its eigenvalues, and can alternatively be evaluated by the expression

$$\det A = \exp \operatorname{tr} \log(A),$$

just like for matrices, which is of great practical relevance. Alternatively, $\det A$ can be expressed in terms of the solutions of the eigenvalue equations

$$\int dy A(x,y)\phi(y) = \lambda\phi(x),$$

where the eigenvalues λ form a continuous manifold. The determinant is then given as the product of all eigenvalues.

An important property is the definition that a functional integral is translationally invariant. Thus, for an arbitrary functional F and an arbitrary function η and constant α

$$\int \mathcal{D}\phi F\left[\phi + \alpha\eta\right] \overset{\phi\to\phi-\alpha\eta}{=} \int \mathcal{D}\phi F\left[\phi\right] \tag{3.3}$$

holds by definition.

From these properties follows the validity of the substitution rule as

$$\int \mathcal{D}\phi F\left[\phi\right] = \int \mathcal{D}\psi \det\left(\frac{\delta\phi}{\delta\psi}\right) F\left[\phi\left[\psi\right]\right],$$

where the Jacobi determinant $\det(\delta\phi/\delta\eta)$ appears. In case of a linear transformation

$$\phi(x) = \int dy \eta(x,y)\psi(y),$$

the determinant is just $\det \eta(x,y)$ of the infinite-dimensional matrix $\eta(x,y)$ with the indices x and y.

With these definitions it is then possible to write down a closed expression for the connected correlation functions $\langle\phi(x_1)...\phi(x_n)\rangle$, which contain all knowledge on any given theory, for a theory with a single field ϕ with action S. They are given by

$$\langle T\phi(x_1)...\phi(x_n)\rangle = \frac{\int \mathcal{D}\phi\phi(x_1)...\phi(x_n)e^{iS[\phi]}}{\int \mathcal{D}\phi e^{iS[\phi]}}. \tag{3.4}$$

This is essentially the basic axiom of the path-integral formulation of a quantum field theory.

Such a writing permits also a more elegant way to express the correlation functions by[1]

$$\begin{aligned}\langle T\phi(x_1)...\phi(x_n)\rangle &= \frac{1}{Z\left[0\right]}\int \mathcal{D}\phi e^{iS[\phi]+\int d^dx\phi(x)j(x)}\ \phi(x_1)...\phi(x_n)\Big|_{j=0}\\ &= \frac{1}{Z\left[0\right]}\int \mathcal{D}\phi\frac{\delta^n}{\delta j(x_1)...\delta j(x_n)}e^{iS[\phi]+\int d^dx\phi(x)j(x)}\Big|_{j=0} = \frac{1}{Z\left[0\right]}\ \frac{\delta^n}{\delta j(x_1)...\delta j(x_n)}Z\left[j\right]\Big|_{j=0},\end{aligned}$$

where the quantities j are denoted as sources. From this generating functional $Z\left[j\right]$ it is possible to determine also generating functionals for connected and amputated vertex functions. Furthermore, this permits to reconstruct the original path-integral as

$$Z[j] = \sum_{n=0}^{\infty}\int d^dx_1...d^dx_n\langle T\phi(x_1)...\phi(x_n)\rangle j(x_1)...j(x_n),$$

which can be proven by comparing both expressions in an expansion term-by-term. This construction can be readily generalized to theories with more than one field.

3.2 Matter fields

The previous treatment permits the description of both scalar fields and gauge fields. However, it is insufficient when treating fermionic fields. The reason is that the classical action appears, which in its current form cannot take into account the Pauli principle, and thus that fermions have to anticommute. In the canonical quantization procedure, this is imposed by the canonical anti-commutation relation. In the path integral formulation, this achieved by replacing the classical fermionic fields with classical anti-commuting fields. This is achieved by replacing ordinary numbers with Grassmann numbers.

3.2.1 Grassmann variables

The starting point is to define anti-commuting numbers, α^a, by the property

$$\{\alpha^a, \alpha^b\} = 0$$

where the indices a and b serve to distinguish the numbers. In particular, all these number are nilpotent,

$$\left(\alpha^a\right)^2 = 0.$$

[1] There are subtleties involved, if the source breaks any symmetry explicitly. Then the limit of vanishing source may be different from the situation at zero source. This will not play a role in this lecture, but does sometimes in QCD.

Hence, the set $\mathcal{S}$ of independent, so-called, Grassmann numbers with $a = 1, ..., N$ base numbers are

$$\mathcal{S} = \{1, \alpha^a, \alpha^{a_1}\alpha^{a_2}, ..., \alpha^{a_1} \times ... \times \alpha^{a_N}\},$$

where all α_i are different. This set contains therefore only 2^N elements. Of course, each element of $\mathcal{S}$ can be multiplied by ordinary complex numbers c, and can be added. This is very much like the case of ordinary complex numbers. Such combinations z take the general form

$$z = c_0 + c_a\alpha^a + \frac{1}{2!}c_{ab}\alpha^a\alpha^b + ... + \frac{1}{N!}c_{a_1...a_N}\alpha^{a_1} \times ... \times \alpha^{a_N}. \tag{3.5}$$

Here, the factorials have been included for later simplicity, and the coefficient matrices can be taken to be antisymmetric in all indices, as the product of α^as are antisymmetric. For $N = 2$ the most general Grassmann number is therefore

$$z = c_0 + c_1\alpha^1 + c_2\alpha^2 + c_{12}\alpha^1\alpha^2,$$

where the antisymmetry has already been used. It is also common to split such numbers in their (Grassmann-)odd and (Grassmann-)even part. Since any product of an even number of Grassmann numbers commutes with other Grassmann numbers, this association is adequate. Note that there is no possibility to invert a Grassmann number, but products of an even number of Grassmann numbers are ordinary numbers and can therefore be inverted.

The conjugate of a product of complex Grassmann-numbers, with independent real and imaginary part, is defined as

$$(\alpha^a...\alpha^b)^* = (\alpha^b)^*...(\alpha^a)^* \tag{3.6}$$

Note that the Grassmann algebra is different from the so-called Clifford algebra

$$\{\beta^a, \beta^b\} = 2\eta^{ab}$$

which is obeyed, e. g., by the γ-matrices appearing in the Dirac-equation, and therefore also in the context of the description of fermionic fields.

To do analysis, it is necessary to define functions on Grassmann numbers. First, start with analytic functions. This is rather simple, due to the nilpotency of Grassmann numbers. Hence, for a function of one Grassmann variable

$$z = b + f$$

only, with b even and f odd, the most general function is

$$F(z) = F(b) + \frac{dF(b)}{db}f.$$

Any higher term in the Taylor series will vanish, since $f^2 = 0$. Since Grassmann numbers have no inverse, all Laurent series in f are equivalent to a Taylor series. For a function of two variables, it is

$$F(z_1, z_2) = f(b_1, b_2) + \frac{\partial F(b_1, b_2)}{\partial b_1} f_1 + \frac{\partial F(b_1, b_2)}{\partial b_2} f_2 + \frac{1}{2} \frac{\partial^2 F(b_1, b_2)}{\partial b_1 \partial b_2} f_1 f_2.$$

There are no other terms, as any other term would have at least a square of the Grassmann variables, which therefore vanishes. Note that the last term is not zero because $F(b_1, b_2) \neq F(b_2, b_1)$ in general, but even if this is the case, it is not a summation.

This can be extended to more general functions, which are no longer analytical in their arguments,

$$F(b, f) = F_0(b) + F_1(b) f \tag{3.7}$$

and correspondingly of more variables

$$F(b_1, b_2, f_1, f_2) = F_0(b_1, b_2) + F_i(b_1, b_2) f_i + F_{12}(b_1, b_2) f_1 f_2.$$

The next step is to differentiate such functions. Note that the function F_{12} has no definite symmetry under the exchange of the indices, though by using an antisymmetric generalization this term can be again written as $F_{ij} f_i f_j$ if F_{ij} is anti-symmetric.

Differentiating with respect to the ordinary variables occurs as with ordinary functions. For the differentiation with respect to Grassmann numbers, it is necessary to define a new differential operator by its action on Grassmann variables. As these can appear at most linear, it is sufficient to define

$$\begin{aligned} \frac{\partial}{\partial f_i} 1 &= 0 \\ \frac{\partial}{\partial f_i} f_j &= \delta_{ij} \end{aligned} \tag{3.8}$$

Since the result should be the same when $f_1 f_2$ is differentiated with respect to f_1 irrespective of whether f_1 and f_2 are exchanged before derivation or not, it is necessary to declare that the derivative anticommutes with Grassmann numbers:

$$\frac{\partial}{\partial f_1} f_2 f_1 = -f_2 \frac{\partial}{\partial f_1} f_1 = -f_2 = \frac{\partial}{\partial f_1}(-f_1 f_2) = \frac{\partial}{\partial f_1} f_2 f_1.$$

Alternatively, it is possible to introduce left and right derivatives. This will not be done here. As a consequence, the product (or Leibnitz) rule reads

$$\frac{\partial}{\partial f_i}(f_j f_k) = \left(\frac{\partial}{\partial f_i} f_j \right) f_k - f_j \frac{\partial}{\partial f_i} f_k.$$

Likewise, the integration needs to be constructed differently.

In fact, it is not possible to define integration (and also differentiation) as a limiting process, since it is not possible to divide by infinitesimal Grassmann numbers. Hence it is necessary to define integration. As a motivation for how to define integration the requirement of translational invariance is often used. This requires then

$$\begin{aligned} \int df &= 0 \\ \int df\, f &= 1 \end{aligned} \tag{3.9}$$

Translational invariance follows then immediately as

$$\int df_1 F(b, f_1 + f_2) = \int df_1 (h(b) + g(b)(f_1 + f_2)) = \int df_1 (h(b) + g(b) f_1) = \int df_1 F(b, f_1)$$

where the second definition of (3.9) has been used. Note that also the differential anti-commutes with Grassmann numbers. Hence, this integration definition applies for $f df$. If there is another reordering of Grassmann variables, it has to be brought into this order. In fact, performing the remainder of the integral using (3.9) yields $g(b)$. It is an interesting consequence that integration and differentiation thus are the same operations for Grassmann variables, as can be seen from the comparison of (3.8) and (3.9).

3.2.2 Fermionic matter

To describe fermionic matter requires then to replace all fields describing fermions, e. g. the electron fields ψ in the QED Lagrangian (3.15) below, by fields of Grassmann variables. I. e., a fermion field associates each space-time point with a spinor of Grassmann variables.

The most important relation necessary later on is again the Gaussian integral over Grassmann fields. To illustrate the use of Grassmann functions, this will be calculated in detail. The starting point is the integral

$$\int d\alpha^* d\alpha \exp(\alpha^* A \alpha),$$

with some ordinary number A. The Taylor expansion of this expression is

$$\int d\alpha^* d\alpha \exp(\alpha^* A \alpha) = \int d\alpha^* d\alpha \alpha^* A \alpha,$$

and any terms linear or constant in the Grassmann variables will vanish during the integration, and likewise, all higher-order terms will be zero, since $\alpha^2 = \alpha^{*2} = 0$. In the next step, it is necessary to be very careful in the ordering of the integrals, as also the

differentials anti-commute with the variables. To act with $d\alpha$ on the variable α requires to anti-commute it with α^* and $d\alpha^*$ first, giving a factor of $(-1)^2$,

$$\int d\alpha d\alpha^* \alpha^* A \alpha = -\int d\alpha^* d\alpha \alpha^* A \alpha = \int d\alpha^* \alpha^* A d\alpha \alpha = \int d\alpha^* \alpha^* A = A \tag{3.10}$$

which is remarkably different from the normal Gaussian integral (3.2), as it returns A instead of $A^{-1/2}$. It can be likewise shown, that the generalization to many variables yields $\det A$ instead of $(\det A)^{-1/2}$. Similarly, it can be shown that for the substitution rule the inverse Jacobian appears. All these results will be useful now when quantizing QED.

3.3 Quantization of QED

In principle, quantizing a theory is now performed by writing down the path integral (3.1) and use (3.4) to calculate the correlation function. That's it. Unfortunately, there is a twist to this for gauge theories, which comes in two levels of escalation.

Start with the naive quantization of the free Maxwell theory with the classical Lagrangian

$$\begin{aligned} \mathcal{L} &= -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu \end{aligned}$$

by writing down the generating functional

$$\begin{aligned} Z[j_\mu] &= \int \mathcal{D}A_\mu \exp\left(iS[A_\mu] + i \int d^d x j_\mu A^\mu \right) \\ S[A_\mu] &= \int d^d x \mathcal{L}, \end{aligned}$$

where the normalization has been absorbed into the measure for convenience. This integral is just a Gaussian one. Hence, it should be possible to integrate it. It takes the form

$$Z[j_\mu] = \int \mathcal{D}A_\mu \exp\left(i \int d^d x \left(\frac{1}{2} A^\mu (g_{\mu\nu} \partial^2 - \partial_\mu \partial_\nu) A^\nu + j^\mu A_\mu \right) \right).$$

However, it is not possible to perform this integral, since this would require the matrix

$$g_{\mu\nu} \partial^2 - \partial_\mu \partial_\nu \tag{3.11}$$

to be invertible, which is not the case. This can be seen directly by the fact that its momentum-space version $g_{\mu\nu} k^2 - k_\mu k_\nu$ is a projection operator which vanishes when contracted with k_μ.

An alternative way to see this is to note that any gauge transformation

$$A_\mu \to A^g_\mu = A_\mu + \partial_\mu g(x) \tag{3.12}$$

with g arbitrary leaves S invariant, and, as a shift, also does not influence the measure. Thus, there are flat directions of the integral, namely along such a gauge orbit, and thus the integral diverges. There are only few possibilities to escape. One is to perform the quantization on a discrete space-time grid in a finite volume, determine observables and only after this take the continuum and infinite-volume limit. This is in most cases only feasible numerically, but then a rather successful approach, and will be detailed below in section 5.1.5. Another one is to determine only quantities which are invariant under gauge transformations. However, it turns out that these are always including non-local expressions beyond perturbation theory, making them very hard to handle in practical calculations. The most convenient choice is very often performing gauge-fixing, i. e., cutting off the flat directions of the integral. This latter possibility will be used here, as it is very illustrative.

Select, as in classical electrodynamics, thus a gauge condition $C[A_\mu, x] = 0$ which selects uniquely exactly one gauge copy. I. e., for a set of gauge-fields related by gauge transformations (3.12) there is one and only one, but also at least one, which satisfies the condition C. An example of such a condition is, e. g., the Landau gauge

$$C = \partial^\mu A_\mu. \tag{3.13}$$

This is not as simple as it may seem, and only works for an Abelian gauge theory, as will be explored in detail in section 5.5.5.2.

To make the path integral well-defined, it is necessary to factor off the irrelevant number of field configurations equivalent under the gauge transformation (3.12), and just remain with one representative for each physically inequivalent field configuration. An alternative, given below by covariant gauges, is to average over all copies with a uniquely defined and integrable weight for each gauge copy.

To do this consider the functional generalization of the Dirac-δ function. The expression

$$\Delta^{-1}[A_\mu] = \int \mathcal{D}g \delta(C[A^g_\mu])$$

contains an integration over all gauge-transformations g for a fixed physical field configuration A_μ, but by the δ-function only the weight of the one configuration satisfying the gauge condition is selected. Hence, when performing the change of variables $g \to g + g'$ with some gauge transformation g' it remains unchanged by definition: The functional integral

is translationally invariant. As a consequence, Δ is actually gauge-invariant. Evaluating it at the gauge-transformed configuration $A_\mu^{g'}$ yields

$$\begin{aligned}\Delta[A_\mu^{g'}]^{-1} &= \int \mathcal{D}g\delta(C[A_\mu^{g+g'}]) = \int \mathcal{D}(g-g')\delta(C[A_\mu^{g}]) \\ &= \int \mathcal{D}g\delta(C[A_\mu^{g}]) = \Delta[A_\mu]^{-1}.\end{aligned}$$

Inverting Δ, the relation

$$1 = \Delta[A_\mu] \int \mathcal{D}g\delta(C[A_\mu^g]) \tag{3.14}$$

is found.

Inserting this into the functional integral yields

$$\begin{aligned}Z &= \int \mathcal{D}A_\mu \Delta[A_\mu] \int \mathcal{D}g\delta(C[A_\mu^g]) \exp(iS[A_\mu]) \\ &= \int \mathcal{D}g \int \mathcal{D}A_\mu^{g'} \Delta[A_\mu^{g'}]\delta(C[A_\mu^{g+g'}]) \exp(iS[A_\mu^{g'}]) \\ &= \int \mathcal{D}g \int \mathcal{D}A_\mu \Delta[A_\mu]\delta(C[A_\mu]) \exp(iS[A_\mu])\end{aligned}$$

In the second line, a gauge transformation of the integration variable A_μ is performed. In the last line the inner variables of integration have been changed from $A_\mu^{g'}$ to $A_\mu^{-g-g'}$ and it has been used that all expressions, except the δ-function, are invariant. Hence, the integral is not influencing anymore the remaining integral, and contributes only a factor, which can be removed by appropriate normalization of the functional integral. In addition, it would have been possible to also replace the action by any gauge-invariant functional, in particular expressions involving some observable f in the form $f[A_\mu]\exp(iS[A_\mu])$. Thus, gauge-fixing is not affecting the value of gauge-invariant observables. Due to the δ-function, on the other hand, now only gauge-inequivalent field configurations contribute, making the functional integral well-defined.

It remains to clarify the role of the functional Δ. It is always possible to resolve the condition $C[A_\mu^g] = 0$ to obtain g as a function of C. Hence, by exchanging C and g as variables of integration yields

$$\Delta[A_\mu]^{-1} = \int \mathcal{D}C \left(\det \frac{\delta C}{\delta g}\right)^{-1} \delta(C) = \left(\det \frac{\delta C[A_\mu, x]}{\delta g}\right)^{-1}_{C=0},$$

where it has been used that for satisfying C there is one and only one g for any gauge-inequivalent field configuration, a so-called gauge orbit. The appearing determinant is just the corresponding Jacobian. Thus, the function Δ is given by

$$\Delta[A_\mu] = \left(\det \frac{\delta C[A_\mu, x]}{\delta g(y)}\right)_{C=0} = \det M(x,y).$$

The Jacobian has the name Faddeev-Popov operator, abbreviated by M, and the determinant goes by the name of Faddeev-Popov determinant.

A more explicit expression is obtained by using the chain rule

$$\begin{aligned} M(x,y) &= \frac{\delta C[A_\mu,x]}{\delta g(y)} = \int d^dz \frac{\delta C[A_\mu,x]}{\delta A_\mu(z)}\frac{\delta A_\mu(z)}{\delta g(y)} \\ &= \int d^dz \frac{\delta C[A_\mu,x]}{\delta A_\mu(z)}\partial^y_\mu\delta(y-z) = -\partial^y_\mu\frac{\delta C[A_\mu,x]}{\delta A_\mu(y)}. \end{aligned}$$

To proceed further, a choice of C is necessary. Choosing, e. g., the Landau gauge $C = \partial^\mu A_\mu = 0$ yields

$$M(x,y) = -\partial^2\delta(x-y).$$

Due to the presence of the δ-function the functional $\det\Delta$ can then be replaced by $\det M$ in the path integral. Note that this result is independent of the field variables, and thus can also be absorbed in the normalization constant. Thus, at this point everything is complete. However, the resulting integral has always the implicit Landau gauge condition to be taken into account. To have rather an explicit condition, general covariant gauges are more useful.

These are obtained by selecting the condition $C = D[A_\mu,x] + \Lambda(x)$ for some arbitrary function Λ. In general, this will make Lorentz symmetry not manifest. This can be recovered by integrating the path integral over all possible values of Λ with some arbitrary integrable weight function. Since the path integral will not depend on Λ, as this is a gauge choice, the integration is only an arbitrary normalization. Using a Gaussian weight, the path integral then takes the form

$$\begin{aligned} Z &= \int \mathcal{D}\Lambda\mathcal{D}A_\mu \exp\left(-\frac{i}{2\xi}\int d^dx\Lambda^2\right)\det M\delta(C)\exp(iS) \\ &= \int \mathcal{D}A_\mu \det M \exp\left(iS - \frac{i}{2\xi}\int d^dx D^2\right), \end{aligned}$$

where the δ-function has been used in the second step. For the most common choice $D = \partial_\mu A^\mu$, the so-called covariant gauges, this yields the final expression

$$Z = \int \mathcal{D}A_\mu \exp\left(iS - \frac{i}{2\xi}\int d^dx(\partial_\mu A^\mu)^2\right).$$

This additional term has the consequence that the Gaussian integral is now well-defined, since the appearing matrix is changed to

$$g_{\mu\nu}\partial^2 - \partial_\mu\partial_\nu \to g_{\mu\nu}\partial^2 - \left(1-\frac{1}{\xi}\right)\partial_\mu\partial_\nu,$$

which can be inverted. The appearing parameter ξ, the so-called gauge parameter, is arbitrary, and can be chosen at will, defining the gauge. Furthermore, the ever-so popular Landau gauge corresponds to the limit $\xi \to 0$, as this is corresponding to the case where all of the weight of the weight-function is concentrated only on the gauge copy satisfying $\partial^\mu A_\mu = 0$. However, in principle this limit may only be taken at the end of the calculation, as it appears to recover the non-invertible original operator (3.11) in Maxwell theory. The choice $\xi = 1$ is known as Feynman gauge.

This process has at no place involved explicitly any matter fields. It therefore can be performed in the same way in the presence of matter fields. Since the local gauge freedom has been taken care of already, no further problems arise, and to quantize QED, it is only necessary to replace the action by the one of QED, and to also integrate about the (Grassmann-valued) fermion fields, yielding

$$Z = \int \mathcal{D}A_\mu \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp\left(-i \int d^d x \left(\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \bar{\psi}(i\gamma^\mu D_\mu - m)\psi + \frac{1}{2\xi}(\partial_\mu A^\mu)^2\right)\right), \quad (3.15)$$

from which now calculations can be performed. How this can be done in practice will be discussed after extending the quantization process to QCD.

Of course, this is only the result for a particular class of gauges, and many others exist. In particular, it is possible to chose conditions C which include also the matter fields explicitly or which are not even Lorentz invariant, like the Coulomb gauge $\sum_{i=1}^{3} \partial_i A_i = 0$.

3.4 Group theory

To treat QCD requires some group theory as group-theoretical structures will be required. These will be collected here before formulating QCD itself.

The basic element will be to represent the operators associated with color charge by generators of an algebra, a so-called Lie algebra G, which form the base vectors of a vector space. Hence, if there should be N independent charges, there must be N independent base vectors τ^a with $a = 1...N$ and $N = \dim G$, and the Lie algebra must therefore be N-dimensional. The defining property of such an algebra are the commutation relations

$$\left[\tau^a, \tau^b\right] = i f^{ab}_{\ \ c} \tau^c.$$

with the anti-symmetric structure constants f^{abc}, which fulfill the Jacobi identity

$$f^{abe} f^{cd}_{\ \ e} + f^{ace} f^{db}_{\ \ e} + f^{ade} f^{bc}_{\ \ e} = 0.$$

These base vectors can be chosen hermitian, i. e., $\tau_a = \tau_a^\dagger$. Note that the position at top or bottom (covariant and contravariant) of the indices is of no relevance for the Lie algebras encountered in the standard model, but can become important in more general settings.

Such a Lie algebra can be represented, e. g., by a set of finite-dimensional matrices. An example is the su(2) algebra with its three generators, which can be chosen to be the Pauli matrices,

$$\begin{aligned} \tau^1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ \tau^2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ \tau^3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned}$$

Furthermore, to each algebra one or more groups can be associated by exponentiation, i. e.,

$$\lambda^a = e^{i\tau^a}, \tag{3.16}$$

provides group elements for the associated group, which are by definition unitary and thus $\lambda_a^{-1} = \lambda_a^\dagger$. For su(2), these are again the Pauli matrices, generating the group SU(2). However, the relation (3.16) is not necessarily a unique relation, and there can be more than one group representation. E. g., for su(2) there are two possible groups related to the algebra by the relation (3.16), the group SU(2) and the group $\mathrm{SU}(2)/\mathrm{Z}_2$, which is isomorphic to SO(3), where matrices which differ only be a negative unit matrix are identified with each other.

Because of the exponential relation, a generic group element $\exp(i\alpha_a\tau^a)$ with real numbers α_a can be expanded for infinitesimal α_a as $1 + i\alpha_a\tau^a$. Thus the algebra describes infinitesimal transformations in the group. This will play an important role when introducing gauge transformations for non-Abelian gauge theories.

There is only a denumerable infinite number of groups which an be constructed in this way. One are the N-dimensional special unitary groups with algebra su(N), and the simplest group representation SU(N) of unitary, unimodular matrices. The second set are the symplectic algebras sp($2N$) which are transformations leaving a metric of alternating signature invariant, and thus are even-dimensional. Finally, there are the special orthogonal algebras so(N), known from conventional rotations. Besides these, there are five exceptional algebras g_2, f_4, and e_6, e_7, and e_8. The u(1) algebra of Maxwell theory fits also into this scheme, the u(1) group is the special case of all f^{abc} being zero, and the algebra being one-dimensional. This is equivalent to so(2).

The most relevant algebras for QCD are su(2) and su(3). The su(2) algebra has the total-antisymmetric Levi-Civita tensor as structure constant, $f^{abc} = \epsilon^{abc}$ with $\epsilon^{abc} = 1$.

The algebra su(3) has as non-vanishing structure constants

$$\begin{aligned} f^{123} &= 1 \\ f^{458} = f^{678} &= \frac{\sqrt{3}}{2} \\ f^{147} = -f^{156} = f^{246} = f^{257} = f^{345} = -f^{367} &= \frac{1}{2}, \end{aligned}$$

and the corresponding ones with permuted indices. There is some arbitrary normalization possible, and the values here are therefore conventional.

From these, also the generators for the eight-dimensional algebra su(3) can be constructed, the so-called Gell-Mann matrices,

$$\begin{aligned} \tau^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \tau^2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \tau^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \tau^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} & \tau^5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} & \tau^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \tau^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} & \tau^8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \tag{3.17}$$

In general, there are $N^2 - 1$ base vectors for su(N), but the dependency for the other algebras is different. For the sake of simplicity, in the following only the expressions for su(N) will be given.

Generators, which are diagonal as matrices, and therefore commute with each other, are said to be in the Cartan subalgebra or subgroup of the algebra or group, respectively. For su(2), this is only one generator, for su(3) there are two.

These lowest-dimensional realization of the commutation relations are called the fundamental representations of the algebra or group. Since the commutation relations are invariant under unitary transformations, it is possible to select a particular convenient realization. Note, however, that there may be more than one unitarily inequivalent fundamental representations. For su(2), there is only one. For su(3), they are two, the second one created by using the $-\lambda^{a*}$ matrices.

It also possible to give representations of the algebras with higher-dimensional matrices. The next simple one is the so-called adjoint representations with the matrices

$$(A^a)_{ij} = -if^a_{ij},$$

which are three-dimensional for su(2) and eight-dimensional for su(3). There are also cases in which the fundamental and the adjoint representation coincide. This can be continued to an infinite number of further representations, which will not be needed here.

Further useful quantities are given by the Dynkin index T_R for an arbitrary representation R (R being e. g. τ or A)

$$\mathrm{tr} R^a R^b = \delta^{ab} T_R,$$

and the Casimirs C_R

$$R^a_{ij} R^a_{jk} = \delta_{ik} C_R,$$

being for su(N) $(N^2-1)/(2N)$ for the fundamental representation and N for the adjoint representation.

To perform a path-integral quantization of a theory involving Lie algebras, it will be necessary to integrate over a group. This can be done using the Haar measure, defined for group elements $g = \exp(\theta^a \tau_a)$ as

$$dg = I(\theta) \Pi_{a=1}^N d\theta^a,$$

where I is the integral measure. The Haar measure is invariant under a variable transformation using a different group-element, i. e., for $g \to gg'$ with an arbitrary different, but fixed, group element g' no Jacobian appears. This replaces the translational invariance of the measure of ordinary integrals. Furthermore, the Haar measure is defined such that

$$\int\limits_G dg = 1,$$

i. e., an integral over the complete group yields unity. The measure I can be shown to be

$$\begin{aligned} I(\theta) &= \frac{1}{V_G} \det M \\ g^{-1} \frac{\partial g}{\partial \theta^a} &= i \tau^b M^{ba} \\ V_G &= \int\limits_G \det M \Pi_{a=1}^N d\theta^a \end{aligned} \tag{3.18}$$

where the second line is an implicit definition, and V_G is the appropriate normalization function.

As an example, group elements of SU(2) can be written as

$$g = e^{i \frac{\theta_a \tau^a}{2}} = \begin{pmatrix} \cos\frac{\eta_1}{2} + i \sin\frac{\eta_1}{2} \cos\eta_2 & i e^{-i\eta_3} \sin\frac{\eta_1}{2} \sin\eta_2 \\ i e^{i\eta_3} \sin\frac{\eta_1}{2} \sin\eta_2 & \cos\frac{\eta_1}{2} - i \sin\frac{\eta_1}{2} \cos\eta_2 \end{pmatrix}, \tag{3.19}$$

where the three-dimensional vector θ^a with maximum length of 2π due to the periodicity of the exponential has been decomposed into polar coordinates with η_1 and η_3 ranging from 0 to 2π and η_2 to π. Using the implicit definition (3.18), this yields the Haar measure for SU(2)

$$dg = \frac{4}{V_G} \sin^2 \frac{\eta_1}{2} d\eta_1 \sin \eta_2 d\eta_2 d\eta_3,$$

and the group volume V_G has a value of $16\pi^2$.

This completes the required group theory necessary to describe QCD.

3.5 Yang-Mills theory

This will permit to formulate QCD in the next chapter. In a first step, this will require the introduction of a gauge theory being a tensor product of a vector field times a Lie algebra, a so-called Yang-Mills theory. This will be done here for arbitrary gauge groups, though the dimension of the algebra will be referred to as colors. After that, the algebra will be specialized to the ones needed in QCD, su(3), with the group representation $SU(3)/Z_3$. However, the division by Z_3 will not be relevant in perturbation theory.

The replacement is rather direct. In Maxwell theory, the gauge fields were a product of the (trivial) generator of u(1), being 1, and the gauge field A_μ. Thus, for a theory including a Lie algebra, a so-called Yang-Mills theory, just the generator will be replaced by the generators of the group, i. e., the gauge fields will be given as

$$A_\mu = A^a_\mu \tau_a$$

with the generators τ^a being in some representation of the gauge algebra. Hence, there are $\dim G$ gauge fields in a Yang-Mills theory, which are essentially algebra-valued. In the case of the standard model, the representation is the fundamental one, though most of the following is general.

It then remains to construct a gauge-invariant action for Yang-Mills theory. This is again an axiomatic process, which can be motivated by various geometric arguments, but in the end remains a postulate.

Since the gauge field is now Lie-algebra-valued, so will be any gauge transformation function, $\tau^a \omega_a(x)$, from which the group-valued unitary gauge transformation $G = \exp ig\tau^a\omega_a$ is obtained. The gauge transformation acts now on the gauge fields as

$$A_\mu \to G A_\mu G^{-1} + G\partial_\mu G^{-1}, \tag{3.20}$$

which in infinitesimal form reads for the gauge fields A^a_μ

$$\begin{aligned} A^a_\mu &\rightarrow A^a_\mu + D^{ab}_\mu \omega_b \\ D^{ab}_\mu &= \delta^{ab}\partial_\mu - g f^{ab}_{\ \ c} A^c_\mu, \end{aligned}$$

where D^{ab}_μ is the covariant derivative in the adjoint representation of the gauge group. There are two remarkable facts about this. One the on hand, there appears an arbitrary constant g in this relation. This constant will later turn out to take the place of the conventional electric charge as the coupling constant of Yang-Mills theory. The second is that the transformation is no longer linear, but there appears a product, even in the infinitesimal case, of the gauge field and the gauge transformation function ω^a. This non-linearity gives rise to all kind of technical complications.

This more lengthy expression requires also a change of the field-strength tensor, to obtain a gauge-invariant theory in the end. The field strength tensor of Yang-Mills theory is

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu] = F^a_{\mu\nu}\tau_a = (\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A_{b\mu} A_{c\nu})\tau_a.$$

There are two more remarkable facts about this field strength tensor. One is that it is no longer linear in the gauge fields, but that there appears an interaction term: Gauge fields in a Yang-Mills theory interact with each other, and the theory is even without matter non-trivial. Furthermore, the appearance of g confirms its interpretation as a coupling constant. The second is that a quick calculation shows that this expression is not gauge-invariant, in contrast to Maxwell theory. The reason is the non-commutativity of the algebra-valued gauge fields.

However, the combination

$$\mathrm{tr}(F_{\mu\nu}F^{\mu\nu}) = F^a_{\mu\nu}F^{\mu\nu}_a$$

is. Thus, in analogy to Maxwell theory, the Lagrangian

$$\mathcal{L} = -\frac{1}{4}\mathrm{tr}(F_{\mu\nu}F^{\mu\nu}) = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a,$$

defines a suitable gauge-invariant object, which defines Yang-Mills theory. Though it looks simple at first, it is a highly non-trivial theory, as it includes the interaction of three and four gauge bosons.

Another consequence of the gauge-variance of the field strength tensor is that color-electric and color-magnetic fields are gauge-variant as well, and they can thus not be measured: Yang-Mills theories do not manifest themselves as observable fields nor as observable color waves. From this follows also that color charge is gauge-variant and thus

not observable, in contrast to electric charge. The only type of gauge-invariant observables in Yang-Mills theory are bound-states, the aforementioned glueballs, and the interactions and behaviors of these glueballs. However, after fixing a gauge, it is of course possible to make statements also about the gauge bosons, and even use experiments to indirectly say something about the properties of gauge bosons in a particular gauge.

Chapter 4

Perturbative QCD

This now provides the basis to finally formulate QCD as a field theory.

4.1 QCD as a field theory

Based on section 3.5, it is straightforward to include also fermions (or scalars) in a Yang-Mills theory. To do this, it is necessary to select a representation for the matter field ψ. They then receive an additional index i, which counts the dimensionality of representation of the gauge group: this enumerates the color. The matter fields, in contrast to the gauge fields, are elements of the group, and not of the algebra. In particular, fundamental matter fields are not invariant under center transformations, and thus explicitly break the group down to its centerless subgroup, if it has non-trivial center.

In analogy to QED they then transform under gauge transformations as

$$\psi_i \to (G\psi)_i = G_{ij}\psi_j = \exp\left(-ig\tau^a\omega_a\right)_{ij}\psi_j,$$

which is therefore a group transformation. The generator τ^a is now in the chosen representation of the matter-fields.

To construct a gauge-invariant action, the covariant derivative for minimal coupling has to include this as well, and it reads now

$$(D_\mu)_{ij} = \delta_{ij}\partial_\mu + igA^a_\mu(\tau_a)_{ij}.$$

with again the generators τ^a in the representation of the matter fields.

QCD contains the six quark flavors, counted by a further index f of quarks, which are all in the fundamental representation of the gauge group, which turns out to be $\mathrm{SU}(3)/\mathrm{Z}_3$ by comparison to experiment. Since the center is absent, it is not broken by the matter

fields. This also explains the difference of gluon and quark colors: The respective indices belong to different representations of the gauge group.

Thus, the Lagrangian of QCD finally reads

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} + \sum_f \bar{\psi}^f_i(i\gamma^\mu D^{ij}_\mu - m_f)\psi^f_j, \tag{4.1}$$

where the generators τ^a appearing in the covariant derivative are the Gell-Mann matrices (3.17). Here the tree-level mass of the quarks m_f has been included explicitly, as the Brout-Englert-Higgs effect in calculations concentrating on QCD phenomena is usually approximated by such tree-level masses.

It should be noted that the coupling appearing in the covariant derivatives is the same for all flavors. The difference in color comes only from the different group indices, e. g., index $i = 1$ is red, and so on. This is in contradistinction to QED, where the different flavors (may) have a different charge. The reason for this is the non-linearity of the gauge transformation of the gauge field (3.20). For the non-linear part to cancel in the covariant derivative of the matter fields, the value of the gauge coupling g for the gauge fields and all matter fields has to be the same, a fact which is known as coupling universality. Otherwise, the theory would not be gauge-invariant. This is not the case for QED, since there is no non-linear part of the gauge field gauge transformation in QED.

4.2 Quantization of Yang-Mills theories

The quantization of QCD is essentially performed by quantizing Yang-Mills theory. This follows essentially the same steps as for QED. However, the non-linearity will introduce further complications, and will necessitate the introduction of further auxiliary fields, the so-called ghost fields, to obtain a description in terms of a local quantum field theory, at least in perturbation theory. For QCD, it is as convenient as for QED to not include the matter fields in the gauge condition, and only this class will be considered. Furthermore, only the class of covariant gauges will be discussed in detail, though other (non-covariant) gauges have also been studied intensively in the literature, e. g. Coulomb or axial gauges. This is beyond the scope of this lecture.

First of all, since the gauge transformations (3.20) leave the action invariant, there are again flat directions similar to QED, thus giving the same reason to implement a gauge-fixing procedure. However, in the following only the procedure for the perturbative case will be discussed. The extension to the non-perturbative case is far from obvious for non-Abelian gauge theories, due to the presence of the so-called Gribov-Singer ambiguity, and will be discussed later in section 5.5.5.2.

The first step is again to select a gauge condition, but this time one for every gauge field, $C^a[A^a_\mu, x] = 0$, e. g. again the Landau gauge[1] $C^a = \partial^\mu A^a_\mu$. The next steps are then the same as for QED, only keeping in mind to drag the additional indices alongside, and that the integration over gauge transformations is now performed using the Haar measure. This continues until reaching the expression

$$Z = \int \mathcal{D}G \int \mathcal{D}A^a_\mu \Delta[A^a_\mu]\delta(C^a[A^a_\mu]) \exp(iS[A^a_\mu])$$

in which for QED Δ could essentially be absorbed in the measure. For non-Abelian gauge theories, this is not possible. For a non-Abelian gauge field, the function Δ is given by[2]

$$\Delta[A^a_\mu] = \left(\det \frac{\delta C^a[A_\mu, x])}{\delta\theta^b(y)}\right)_{C^a=0} = \det M^{ab}(x,y). \tag{4.2}$$

with the non-Abelian Faddeev-Popov operator M^{ab}.

A more explicit expression is again obtained using the chain rule

$$\begin{aligned} M^{ab}(x,y) &= \frac{\delta C^a(x)}{\delta\theta^b(y)} = \int d^4z \frac{\delta C^a(x)}{\delta A^c_\mu(z)} \frac{\delta A^c_\mu(z)}{\delta\theta^b(y)} \\ &= \int d^4z \frac{\delta C^a(x)}{\delta A^c_\mu(z)} D^{cb}_\mu \delta(y-z) = \frac{\delta C^a(x)}{\delta A^c_\mu(y)} D^{cb}_\mu(y). \end{aligned} \tag{4.3}$$

To proceed further, a choice of C^a is necessary, which will again be covariant gauges, selected by the condition $C_a = D_a + \Lambda_a (= \partial_\mu A^\mu_a + \Lambda_a)$ for some arbitrary functions Λ^a. The path integral then takes the form

$$\begin{aligned} Z &= \int \mathcal{D}\Lambda^a \mathcal{D}A^a_\mu \exp\left(-\frac{i}{2\xi}\int d^4x \Lambda^a\Lambda_a\right) \det M \delta(C) \exp(iS) \\ &= \int \mathcal{D}A^a_\mu \det M \exp\left(iS - \frac{i}{2\xi}\int d^4x D^a D_a\right), \end{aligned} \tag{4.4}$$

for a Gaussian weight, and the δ-function has been used in the second step. The arbitrary parameter ξ is called the gauge parameter. As the determinant of an operator is a highly non-local object, the current expression is unsuited for most calculations.

This non-locality can be recast, using auxiliary fields, as an exponential. Using the rules of Grassmann numbers it follows immediately that

$$\det M \sim \int \mathcal{D}c^a \mathcal{D}\bar{c}^a \exp\left(-i\int d^4x d^4y \bar{c}_a(x) M^{ab}(x,y) c_b(y)\right), \tag{4.5}$$

[1] For simplicity, here only gauge conditions linear in the group indices are used. Of course, in general gauge conditions can also depend on gluon fields with a different index than their own.

[2] This determinant can be zero outside perturbation theory, see again section 5.5.5.2.

where the auxiliary Faddeev-Popov ghost and antighost fields c and $\bar{c}$ are Grassmann-valued scalar fields. Since these are just auxiliary fields, this is not at odds with the spin-statistics theorem. The fields are in general gauges not related, but may be so in particular gauges. This is, e. g., the case in Landau gauge where there exists an associated symmetry. If the condition D^a is local in the fields, the Faddeev-Popov operator will be proportional to $\delta(x-y)$, and this ghost term will become local.

It is furthermore often useful to introduce an additional auxiliary field, the Nakanishi-Lautrup field b^a. This is obtained by rewriting

$$\exp\left(-\frac{i}{2\xi}\int d^4x D^a D_a\right) \sim \int \mathcal{D}b^a \exp\left(i\int d^4x\left(\frac{\xi}{2}b^a b_a + b_a D^a\right)\right).$$

Upon using the equation of motion for the b field, the original version is recovered.

The final expression then reads

$$Z = \int \mathcal{D}A^a_\mu \mathcal{D}b^a \mathcal{D}c^a \mathcal{D}\bar{c}^a \exp\left(iS + \int d^4x\left(\frac{\xi}{2}b^a b_a + b_a D^a\right) - \int d^4x d^4y \bar{c}^a(x) M^{ab}(x,y) c^b(y)\right). \tag{4.6}$$

Choosing the gauge $D^a = \partial^\mu A^a_\mu = 0$, this takes the form

$$Z = \int \mathcal{D}A^a_\mu \mathcal{D}b^a \mathcal{D}c^a \mathcal{D}\bar{c}^a \exp\left(iS + i\int d^4x\left(\frac{\xi}{2}b^a b_a + b^a \partial_\mu A^\mu_a\right) - i\int d^4x \bar{c}^a \partial^\mu D^{ab}_\mu c^b\right).$$

Furthermore, the ever-so popular Landau gauge corresponds to the limit $\xi \to 0$, as this is corresponding to the case where all of the weight of the weight-function is concentrated only on the gauge copy satisfying $\partial^\mu A^a_\mu = 0$. However, in principle this limit may only be taken at the end of the calculation.

To return to QED, it is sufficient to notice that in this case

$$D^{ab}_\mu \phi^b \to \partial_\mu \phi,$$

and thus the ghost term takes the form

$$-i\int d^4x \bar{c}^a \partial^2 c^a.$$

Hence, the ghosts decouple, and will not take part in any dynamical calculations. However, their contribution can still be important, e. g., in thermodynamics. The decoupling of the ghosts is not a universal statement. Choosing a condition which is not linear in the gauge fields will also in an Abelian theory introduce interactions. Furthermore, from the sign of this term it is also visible that the kinetic term of the ghosts has the wrong sign compared to ordinary scalars, a sign of their unphysical spin-statistic relation.

This program can be performed in a much more formal and general way, the so-called anti-field method, and also using canonical quantization. Both are beyond the scope of this lecture.

To include fermions requires just to add their contribution to the Lagrangian, and to integrate over them, as they have not been involved in the gauge-fixing procedure.

4.3 BRST and asymptotic states

As stated, the chromoelectric and chromomagnetic fields themselves are no longer physical in a Yang-Mills theory. It thus requires some other method to identify physical degrees of freedom, and a more general construction of the physical state space is required.

A possibility to establish the physical state space is by use of the BRST (Becchi-Rouet-Stora-Tyutin) symmetry, which is a residual symmetry after gauge-fixing. Perturbatively, it permits to separate physical from unphysical fields. In the so-called Kugo-Ojima construction it is attempted to extend this construction beyond perturbation theory, though whether this is possible has not yet been settled. This will be briefly discussed in section 5.5.5.

4.3.1 BRST symmetry

The starting point for the discussion is the gauge-fixed Lagrangian with Nakanishi-Lautrup fields included

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F_a^{\mu\nu} + \frac{\xi}{2}b^a b_a + b^a D_a - \int d^4z \bar{u}_a(x)\frac{\delta D^a}{\delta A^c_\nu}D^{cb}_\nu u_b(z).$$

Herein the gauge condition is encoded in the condition $C^a = 0$. Furthermore, matter fields are ignored, as they will not alter the discussion qualitatively. These contributions will be reinstantiated later.

This Lagrangian furnishes actually a global symmetry, as it is invariant under the transformation δ_B defined as

$$\delta_B A^a_\mu = \lambda D^{ab}_\mu u_b = \lambda s A^a_\mu \tag{4.7}$$

$$\delta_B u^a = -\lambda\tfrac{g}{2} f^{abc}u_b u_c = \lambda s u^a \tag{4.8}$$

$$\delta_B \bar{u}^a = \lambda b^a = \lambda s \bar{u}^a \tag{4.9}$$

$$\delta_B b^a = 0 = \lambda s b^a. \tag{4.10}$$

Herein, λ is an infinitesimal Grassmann number, i. e., it anticommutes with the ghost fields.

As a consequence, this so-called BRST transformation s has to obey the generalized Leibnitz rule

$$s(FG) = (sF)G + (-1)^{\text{Grassmann parity of } F} FsG.$$

The Grassmann parity of an object is 1 if it is Grassmann odd, i. e. contains an odd number of Grassmann numbers, and 0 otherwise.

Showing the invariance is simple for the classical Lagrangian, as the transformation for the gauge boson is just an ordinary gauge transformation with gauge parameter λu^a, which is an ordinary real function.

That the remaining gauge-fixing part of the Lagrangian is invariant under a BRST transformation can be seen as follows. The quadratic term in b^a is trivially invariant. The second term from the gauge-fixing part transforms for a linear gauge condition D_a as

$$s(b^a D_a) = b^a \int d^4y \frac{\delta D^a}{\delta A^b_\mu} sA^b_\mu = b^a \int d^4y \frac{\delta D^a}{\delta A^b_\mu} D^{bc}_\mu u_c.$$

To determine the transformation of the ghost-part, there are four components on which the transformation acts. The first is when s acts on the anti-ghost. This yields

$$-s(\bar{u}_a(x)) \int d^4z \frac{\delta D^a}{\delta A^b_\nu} D^{bc}_\nu u_c(z) = -b_a \int d^4z \frac{\delta D^a}{\delta A^b_\nu} D^{bc}_\nu u_c(z).$$

It therefore precisely cancels the contribution from the second part of the gauge-fixing term.

The next is the action on the gauge-fixing condition,

$$\begin{aligned} &\int d^4y s \left(\frac{\delta D^a}{\delta A^b_\nu(y)} \right) D^{bc}_\nu u_c = \int d^4y d^4z \frac{\delta D^a}{\delta A^b_\nu(y) \delta A^d_\rho(z)} (sA^d_\rho(z)) D^{bc}_\nu u_c(y) \\ = &\int d^4y d^4z \frac{\delta D^a}{\delta A^b_\nu(y) \delta A^d_\rho(z)} D^{de}_\rho u_e(z) D^{bc}_\nu u_c(y) = 0. \end{aligned}$$

In linear gauges, like the covariant gauges, it immediately vanishes since the second derivative of the gauge condition is zero. In non-linear gauges, this becomes more complicated, and in general requires the exploitation of various symmetry properties, depending on the actual gauge condition.

The two remaining terms can be treated together as

$$\begin{aligned} s(D^{ab}_\mu u_b) &= \partial_\mu s u^a - g f^{ab}_{\ c}((sA^c_\mu) u_b + A^c_\mu s u_b) \\ &= -\frac{g}{2} \partial_\mu (f^{abc} u_b u_c) - g f^{ab}_{\ c} D^{cd}_\mu u_d u_b - g f^{ab}_{\ c} f_{bde} A^c_\mu u^d u^e \\ &= \frac{g}{2} f^{ab}_{\ c} (\partial_\mu (u_b u^c) - 2 u_b \partial_\mu u_c - 2g (f^{cd}_{\ e} A^e_\mu u_d u_b + g f_b^{\ de} A^c_\mu u_d u_e)). \end{aligned}$$

The first two terms cancel each other, after adequate relabeling of indices. The last two terms can be rearranged by index permutation such that the Jacobi identity can be used so that they vanish as well,

$$\begin{aligned} &= \frac{g}{2} f^{abc}(u_b \partial^\mu u_c + (\partial^\mu u_b) u_c - 2 u_b \partial^\mu u_c) \\ &\quad + g(f^{abc} f^{de}{}_c A^\mu_e u_d u_b + f^{abc} f^{de}{}_c A^\mu_e u_d u_b + f^{abc} f^{de}{}_c A^\mu_b u_d u_e) \\ &= g(f^{abc} f^{dec} A^\mu_e u_d u_b + f^{adc} f^{ebc} A^\mu_e u_d u_b + f^{aec} f^{dbc} A^\mu_e u_d u_b) \\ &= g(f^{abe} f^{cd}{}_e + f^{ace} f^{db}{}_e + f^{ade} f^{bc}{}_e) A^\mu_c u_d u_b = 0, \end{aligned}$$

for which a number of index rearrangements and relabellings are necessary, taking always the Grassmannian nature of the ghosts duly into account. Hence, indeed the gauge-fixed Lagrangian is BRST-invariant.

An amazing property of the BRST symmetry is that it is nil-potent, i. e., $s^2 = 0$. This follows immediately from a direct application. The previous calculation already showed that

$$0 = s(D^{ab}_\mu u^b) = s^2 A^a_\mu.$$

It is trivial for the anti-ghost and the auxiliary b^a field by construction. For the ghost it immediately follows by

$$s^2 u^a \sim s(f^{abc} u_b u_c) \sim f^{abc} f_b{}^{de} u_d u_e u_c - f^{abc} f_c{}^{de} u_b u_d u_e = f^{abc} f_b{}^{de}(u_d u_e u_c + u_c u_d u_e) = 0.$$

The last step is not trivial, but follows from the fact that the ghost product is Grassmannian in nature, and only non-zero if all three indices are different, and thus behaves as an anti-symmetric tensor ϵ_{cde}.

There is even more possible. It holds that the gauge-fixing part of the Lagrangian can be written as

$$\begin{aligned} \mathcal{L}_f &= s\left(\bar{u}^a \left(\frac{\xi}{2} b^a + D^a\right)\right) \\ &= \frac{\xi}{2} b^a b_a + b^a D_a + \bar{u}_a \int d^4 y \frac{\delta D^a}{\delta A^b_\mu(y)} D^{bc}_\mu u_b(y). \end{aligned}$$

Hence, the gauge-fixing part of the Lagrangian is BRST-invariant, since $s^2 = 0$. This can be generalized to other gauge conditions by adding arbitrary so-called BRST-exact terms $s(\bar{u}^a F_a)$ with F^a arbitrary to the Lagrangian. The factor of $\bar{u}$ is necessary to compensate the ghost of the BRST transformation, since any term in the Lagrangian must have a net number of zero ghosts. This extension leads to the so-called anti-field formalism for gauge-fixing. This will not be pursued further here.

The BRST transformation for matter fields also take the form of a gauge transformation with the parameter λu^a. Therefore, all matter Lagrangian contributions automatically satisfy invariance under a BRST transformation. For a fermionic or bosonic matter field ϕ in representation τ^a it takes the form

$$\begin{aligned}\delta_B\phi^i &= \lambda i g u^a \tau_a^{ij}\phi_j \\ s^2\phi^i &= i g \tau_a^{ij} s(u^a\phi_j) = i g \tau_a^{ij}\left(\frac{g}{2}f^{abc}u_b u_c\phi_j + i g u^a u_b \tau_{jk}^b \phi^k\right) \sim g^2\left\{\tau^a,\tau^b\right\}^{ij} u_a u_b \phi_j = 0,\end{aligned}$$

where in the second-to-last step the relation between structure constants and generators has been used backwards, permitting to combine both terms into the symmetric anti-commutator. The combination with the anti-symmetric ghost product yields then zero.

4.3.2 Constructing the physical state space

The following discussion shows how to explicitly construct the state space using BRST symmetry. It extends thereby the Gupta-Bleuler construction of QED, and it can be directly extended to include also matter fields.

The first concept in constructing the physical state space is the presence of states which do not have a positive norm. The simplest example is already given in Maxwell theory. Choose, e. g., Feynman gauge, i. e. $\xi = 1$. The corresponding propagator is then given by Gaussian integration as

$$\langle A_\mu^\dagger(x) A_\nu(y)\rangle = \delta^{ab} g_{\mu\nu}\int \frac{d^4p}{(2\pi)^4}\frac{e^{ip(x-y)}}{p^2+i\epsilon} = -\delta^{ab}g_{\mu\nu}\int\frac{d^3p}{2(2\pi)^3|\vec{p}|}e^{-ip_i(x-y)_i}.$$

The norm of a state

$$\Psi(x) = \int d^4x f(x) A_0(x)\left|0\right\rangle = \int \frac{d^4x d^4p}{(2\pi)^4} e^{ip_0x_0 - p_ix_i} f(p) A_0(x)\left|0\right\rangle,$$

with $f(x)$ an arbitrary weight function, created from the vacuum by the operator A_μ then reads

$$|\Psi|^2 = \int d^4x \int d^4y \langle A_0^\dagger(x) A_0(y)\rangle f^\dagger(x) f(x) = -\int\frac{d^3p}{2|\vec{p}|} f^\dagger(p) f(p) < 0.$$

Hence, there are negative (and zero) norm states present in the state space. These cannot contribute to the physical state space, or otherwise the probability interpretation of the theory will be lost. Or at least, it must be shown that the time evolution is only connecting physical, i. e. with positive definite norm, initial states to physical final states.[3]

[3]The precise characterization of what is a final state beyond perturbation theory is open. One possibility, discussed before, is a non-perturbative extension of the construction to follow. Another one

That they indeed do not contribute can be shown using the BRST symmetry. In fact, it will be shown that

$$\begin{aligned} Q_B \left|\psi\right\rangle_{\text{phys}} &= 0, \\ \left[Q_B, \psi\right]_{\pm} &= s\psi. \end{aligned} \tag{4.11}$$

will be sufficient to define the physical state space, where the second line defines the BRST charge Q_B. The $\pm$ indicate commutator or anticommutator, depending on whether ψ is bosonic (commutator) or fermionic (anticommutator). The BRST charge Q_B can also be defined from the Noether current. It is given by

$$Q_B = \int d^3x \left(b_a D_0^{ab} u_b - u_a \partial_0 b^a + \frac{1}{2} g f^{abc} u_b u_c \partial_0 \bar{u}_a \right).$$

It is fermionic. Since $s^2 = 0$ it directly follows that $Q_B^2 = 0$ as well.

The BRST charge has evidently a ghost number of 1, i. e., the total number of ghost fields minus the one of anti-ghosts is 1. This ghost number, similarly to fermion number, is actually a conserved quantum number of the theory. It is due to the invariance of the Lagrangian under the scale transformation

$$\begin{aligned} u^a &\rightarrow e^{\alpha} u^a \\ \bar{u}^a &\rightarrow e^{-\alpha} \bar{u}^a, \end{aligned}$$

with real parameter α. Note that such a scale transformation is possible since u^a and $\bar{u}^a$ are independent fields. Furthermore for a hermitian Lagrangian the relations

$$\begin{aligned} u^\dagger &= u \\ \bar{u}^\dagger &= -\bar{u} \end{aligned}$$

hold. As a consequence, also the BRST transformation and charge have ghost number 1 and are Hermitian. Together, they form the BRST algebra

$$\{Q_{\text{BRST}}, Q_{\text{BRST}}\} = 0 \tag{4.12}$$

$$[iQ_G, Q_{\text{BRST}}] = Q_{\text{BRST}} \tag{4.13}$$

$$[iQ_G, iQ_G] = 0 \tag{4.14}$$

as can be seen from an explicit calculation.

characterizes all physical states by the necessary condition to be invariant under renormalization - after all, physics should be independent of the scale at which it is measured. However, whether this condition is sufficient, in particular beyond perturbation theory, is also not clear. Bound states with non-zero ghost number, e. g., may also possess this property, though may not be a viable physical state.

Since the Lagrangian is invariant under BRST transformation, so is the Hamiltonian, and therefore also the time evolution and thus the S-matrix,

$$\begin{aligned}[Q_B, H] &= 0 \\ [Q_B, S] &= 0.\end{aligned}$$

Hence, if in fact the BRST symmetry is manifest[4], and the condition (4.11) defines the physical subspace that is already sufficient to show that physical states will only evolve into physical states. It remains to see what kind of states satisfy (4.11).

Because the BRST charge is nilpotent the state space can be separated in three subspaces:

- States which are not annihilated by Q_B, $V_2 = \{|\psi\rangle \, | Q_B \, |\psi\rangle \neq 0\}$.
- States which are obtained by Q_B from V_2, $V_0 = \{|\phi\rangle \, | \, |\phi\rangle = Q_B \, |\psi\rangle \, , |\psi\rangle \, \epsilon V_2\}$. As a consequence $Q_B V_0 = 0$.
- States which are annihilated by Q_B but do not belong to V_0, $V_1 = \{|\chi\rangle \, | Q_B \, |\chi\rangle = 0, |\chi\rangle \neq Q_B \, |\psi\rangle \, , \forall \, |\psi\rangle \, \epsilon V_2\}$.

The states in V_2 do not satisfy (4.11), and therefore would not be physical. The union of the two other states form the physical subspace.

$$V_p = V_0 \cup V_1.$$

It is this subspace which is invariant under time evolution. It is not trivial to show that all states in this space have positive semi-definite norm, but this is possible. This will be used here without proof. However, all states in V_0 have zero norm, and have no overlap with the states in V_1,

$$\begin{aligned}\langle\phi|\phi\rangle &= \langle\phi|Q_B|\psi\rangle = 0 \\ \langle\phi|\chi\rangle &= \langle\psi|Q_B|\chi\rangle = 0.\end{aligned}$$

Since matrix elements are formed in this way the states in V_0 do not contribute, and every state in V_p is thus represented by an equivalence class of states characterized by a distinct state from V_1 to which an arbitrary state from V_0 can be added, and thus a ray of states. Therefore, the physical Hilbert space H_p can be defined as the quotient space

$$H_p = V_p/V_0 = \frac{\mathrm{Ker} Q_B}{\mathrm{Im} Q_B},$$

[4]The consequences of a not manifest BRST are far from trivial, and the non-perturbative status of BRST symmetry is still under discussion, though there is quite some evidence that if it can be defined it is well defined. But how to define it is not finally settled.

the so-called cohomology of the operator Q_B. Therefore, all states in H_p have positive norm, provided that the states in V_1 have.

To define the theory in the vacuum, use can be made of asymptotic states, in perturbation theory usually known as in and out states. A corresponding physical asymptotic states ψ_p^a must therefore obey

$$s\psi_p^a = 0.$$

In the following, the classification of the fields will be done in this form for perturbation theory. In this case, this will finally amount to discarding essentially all composite fields. Beyond perturbation theory, this is no longer possible, as cluster decomposition in general no longer holds in gauge theories. How to proceed beyond perturbation theory is therefore not completely understood.

To obtain the asymptotic fields, start with the BRST variation of a given Green's functions. Asymptotic states are defined to be the pole-part of the asymptotic field. To obtain these, start with the formula

$$\langle T(s\psi_i)\psi_{i_1}...\psi_{i_n}\rangle = \langle T(s\psi_i)\psi_k\rangle \langle T\psi_k\psi_{i_1}...\psi_{i_n}\rangle .$$

In this case, the indices i sum all space-time and internal indices and T is the time-ordering. Essentially, a one has been introduced. Since in perturbation theory all interactions are assumed to cease for asymptotic states, the BRST transformation become linear in the fields

$$s\psi_i \to s\psi_i^a = C_{ik}\psi_k^a.$$

Furthermore, by comparison with the previous calculation, the coefficients can be defined as

$$C_{ik} = \langle T(s\psi_i)\psi_k\rangle = \frac{1}{Z[0]}\frac{i\delta^2}{\delta J_{s\psi_i}\delta J_{\psi_k}}Z[J]$$

at least asymptotically. Note that the source coupled to $s\psi_i$ is necessarily the one for a composite operator. Since in this case the Green's functions will be dominated by the on-shell (pole) part, only those coefficients will be relevant where $s\psi_i$ and ψ_k have the same mass.

As a consequence, this condition reads

$$J_i^p s\psi_i^{ap} = J_i^p \frac{1}{Z[0]}\frac{i\delta^2 Z[J]}{\delta J_{s\psi_i}\delta J_{\psi_k}}\psi_k^a = 0,$$

since the BRST-variation of physical fields vanish.

The interesting question is then the form of these asymptotic propagators appearing. In case of the gauge field

$$sA_\mu^{la}(x) = \int d^4y R_\mu^{lm}(x,y)u^{ma}(y), \tag{4.15}$$

where the index a stands for asymptotic. That only u appears is due to the fact that the ghost is the parameter of the BRST transformation. The propagator then has the form

$$R_\mu^{ab} = \left\langle T(sA_\mu^{aa})\bar{u}^b \right\rangle = -\left\langle TA_\mu^{aa} s\bar{u}^b \right\rangle .$$

The later identity is correct, since

$$s(AB) = (sA)B + (-1)^{g_B} AsB \tag{4.16}$$

and the fact that a physical vacuum expectation value for any pure BRST variation, $s(AB)$ vanishes, $\langle s(AB)\rangle = 0$. It then follows further

$$-\left\langle TA_\mu^a b^b \right\rangle = -\frac{1}{\xi}\left\langle TA_\mu^a C^b \right\rangle = \frac{1}{\xi}\left\langle TA_\mu^a A_\nu^c \right\rangle \phi_{bc}^\nu = \frac{1}{\xi} D_{\mu\nu}^{ac}\phi_{bc}^\nu \tag{4.17}$$

where it was assumed in the second-to-last step that the gauge-fixing condition C^a is linear in the field, $C^a = \phi_\nu^{bc} A^\nu$, and the appearance of partial derivatives has been compensated for by a change of sign. This is therefore a statement for all contributions not-orthogonal to ϕ_ν^{bc}.

Now, because of Lorentz and (global) gauge invariance, it must be possible to rewrite

$$R_\mu^{ab} = \delta^{ab}\partial_\mu R.$$

Therefore, asymptotically

$$\delta^{ab}\partial_\mu R = \frac{1}{\xi} D_{\mu\nu}^{ac}\phi_{cb}^\nu = -\langle TA_\mu^a b^b\rangle \tag{4.18}$$

must hold. The gluon propagator is asymptotically the free one. The right-hand side equals precisely the mixed propagator of the free A_μ and b^a field. This one is given by $\delta^{ab}\partial_\mu\delta(x-y)$, as can be read off directly from the Lagrangian. Therefore, $R = \delta(x-y)$ to obtain equality. Reinserting this into (4.15) yields

$$sA_\mu^{aa} = \partial_\mu u^a.$$

For the ghost the asymptotic BRST transformation vanishes, since its BRST transform is of ghost number 2. There is no single particle state with such a ghost number. The BRST transformed of the anti-ghost field is already linear, yielding

$$\begin{aligned} sA_\mu^{aa} &= \partial_\mu u^a \\ su^{aa} &= 0 \\ s\bar{u}^{aa} &= b^{aa} \\ sb^{aa} &= 0, \end{aligned}$$

for the full list of asymptotic BRST transformed fields. Unsurprisingly, these are exactly the BRST transformations of the free fields.

From this follows that the longitudinal component of A_μ, since ∂_μ gives a direction parallel to the momentum, is not annihilated by s, nor is the anti-ghost annihilated by the BRST transformation. They belong therefore to V_2. The ghost and the Nakanishi-Lautrup field are both generated as the results from BRST transformations, and therefore belong to V_0. Since they are generated from states in V_2 it is said they form a quartet with parent states being the longitudinal gluon and the anti-ghost and the daughter states being the ghost and the Nakanishi-Lautrup field. Therefore, these fields not belonging to the physical spectrum, are said to be removed from the spectrum by the quartet mechanism. Note that the equation of motion for the field b^a makes it equivalent to the divergence of the gluon field, which can be taken to be a constraint for the time-like gluon. Therefore, the absence of the Nakanishi-Lautrup field from the physical spectrum implies the absence of the time-like gluon. Finally, the transverse gluon fields are annihilated by the BRST transformation but do not appear as daughter states, they are therefore physical. In general gauges, the second unphysical degree of freedom will be the one constrained by the gauge-fixing condition to which b^a is tied, while the two remaining polarization directions, whichever they are, will belong to V_1.

Of course, the gauge bosons cannot be physical, since they are not gauge-invariant. Therefore, their removal from the spectrum must proceed by another mechanism, which is therefore necessarily beyond perturbation theory. A proposal for a similar construction also applying to the gauge bosons has been given by Kugo and Ojima, though its validity has not yet satisfactorily been established.

The introduction of quarks (or other matter) fields ψ follows along the same lines. It turns out that all of the components belong to V_1, i. e., $s\psi = 0$, without ψ appearing on any right-hand side, and therefore all fermionic degrees of freedom are perturbatively physically. This can be directly seen as their gauge, and consequently BRST, transformation is

$$\delta\psi^a = igu^a\tau^a_{ij}\psi_j,$$

and hence its free-field ($g = 0$) result is $s\psi^a_i = 0$. This is expected, since no asymptotic physical bound-state with ghost and fermion number one exists.

Similar as for the gauge boson, this cannot be completely correct, and has to change non-perturbatively.

4.4 Ward-Takahashi and Slavnov-Taylor identities

If a theory has a symmetry, irrespective whether it is global or local and whether it is explicit or hidden, this symmetry implies that certain changes can be made to the fields with well-defined consequences. From this results similar well-defined consequences for the correlation functions. In particular, this implies certain relations between combinations of correlation functions, so-called Ward-Takahashi or rigid identities for global symmetries, and Slavnov-Taylor identities for local symmetries.

These identities have two particular useful purposes. One is that it is possible from the knowledge of some correlation functions to infer knowledge about other correlation functions. The second use is that by checking the identities after a calculation, it is possible to determine whether errors occurred, being them either of numerical origin, by some glitch in the calculation, or by the approximations made. Unfortunately the fulfillment of the identities is only a necessary condition for the absence of errors, not a sufficient one. It is always possible that some errors cancel each other in the identities, so care has to be taken when interpreting a check using such identities.

4.4.1 Ward-Takahashi identities

Take a theory with only bosonic fields for simplicity, otherwise additional factors of minus one will appear due to the Grassmann nature of fermionic fields. Let the theory be symmetric under the infinitesimal change

$$\phi \to \phi' = \phi + \delta\phi = \phi + \epsilon f(\phi, x), \tag{4.19}$$

with ϵ infinitesimal. Then the generating functional $Z[J]$ should not change, i. e., δZ should be zero. This variation

$$\delta F(\phi) = \frac{\delta F}{\delta \phi}\delta\phi = \frac{\delta F}{\delta \phi}\epsilon f$$

acts on two components in the path integral. One is the action on the action itself, which yields

$$\frac{1}{\epsilon}\delta\left(e^{iS+i\int d^dxJ\phi}\right) = i\left(\frac{\delta S}{\delta\phi} + J\right)\epsilon f e^{iS+i\int d^dxJ\phi},$$

to first order in ϵ. The second is the measure. The shift (4.19) is a variable transformation, which generates a Jacobian determinant. This Jacobian determinant can also be expanded in ϵ, yielding

$$\det\frac{\delta\phi'}{\delta\phi} = \det\left(1 + \frac{\delta\epsilon f}{\delta\phi}\right) = 1 + \epsilon\frac{\delta f}{\delta\phi} + \mathcal{O}(\epsilon^2).$$

Together, this yields the variation

$$0 = \frac{1}{\epsilon}\delta Z = \int \mathcal{D}\phi \left(\frac{\delta f}{\delta \phi} + i\left(\frac{\delta S}{\delta \phi} + J\right) f\right) e^{iS + i\int d^d x J\phi}. \tag{4.20}$$

Differentiating this expression once with respect to the source and setting the sources afterwards to zero yields an expression connecting different correlation functions. E. g., performing a single derivative will yield

$$\left\langle T\phi(y)\frac{\delta f(\phi, x)}{\delta\phi(x)}\right\rangle + i\left\langle \phi(y)\frac{\delta S}{\delta\phi(x)} f\right\rangle + \langle Tf\rangle = 0.$$

In general, there will not only be one field involved, but many fields, numerated by a field index i. In this case, expression (4.20) takes the form

$$0 = \int \mathcal{D}\phi_i \left(\frac{\delta f_k}{\delta \phi_k} + i\left(\frac{\delta S}{\delta \phi_k} + J_k\right) f_k\right) e^{iS + i\int d^d x J\phi},$$

i. e., it becomes a sum over all fields. Deriving this expression in total n times for any sequence of field types i_l yields the set of all Ward-Takahashi identities

$$\begin{aligned}&\left\langle T\Pi_{l=1}^n \phi_{i_l}(x_l)\frac{\delta f_k}{\delta\phi_k(y)}\right\rangle + i\left\langle T\Pi_{l=1}^n \phi_{i_l}(x_l)\frac{\delta S}{\delta\phi_k(y)} f_k\right\rangle \\ &+ \sum_{m=1}^{n}\left\langle \Pi_{l=1}^{m-1}\phi_{i_l}(x_l) f_{i_m}\Pi_{r=m+1}^{n}\phi_{i_r}(x_r)\right\rangle = 0.\end{aligned} \tag{4.21}$$

To obtain practical cases requires to insert an action with a certain invariance.

Take as an example the action for so-called linear σ-model (or ϕ^4 model) with positive mass squared,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^\dagger\partial^\mu\phi + \frac{1}{2}m^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2.$$

The transformation function is then $f_i = \mp i\phi_i$, where $i = 1$ refers to ϕ and $i = 2$ refers to $\phi^\dagger$. The derivative of f actually vanishes in this case, since the Jacobian matrix under a linear shift of the fields is zero, by the definition of translational invariance of the path integral (3.3). This is not necessarily the case, and when treating anomalies a case will be encountered where the Jacobian is non-vanishing.

Furthermore, the action is invariant under the global symmetry transformation. This implies

$$\frac{\partial S[\phi_i + \epsilon f_i]}{\partial\epsilon} = 0 = \int d^d x \frac{\delta S}{\delta\phi_i}\frac{\partial(\phi_i + \epsilon f_i)}{\partial\epsilon} = \int d^d x \frac{\delta S}{\delta\phi_i} f_i,$$

and thus also the second term in (4.20) vanishes. Hence, only the third term remains, which can be conveniently written as

$$0 = \delta\left\langle T\Pi_{l=1}^n\phi_{i_l}\right\rangle, \tag{4.22}$$

which are called Ward identities in this context. E. g., at level $n = 2$, this identity implies

$$\langle T(\delta\phi(x))\phi(y)^\dagger\rangle + \langle T\phi(x)\delta\phi(y)^\dagger\rangle = \langle\phi(x)\phi(y)^\dagger\rangle - \langle\phi(x)\phi(y)^\dagger\rangle = 0,$$

which seems rather trivial. However, when rewriting the theory in terms of σ and χ fields, i. e. $\phi = \sigma + i\chi$, this implies

$$\langle T\delta\sigma\chi\rangle + \langle T\sigma\delta\chi\rangle = \langle\chi\chi\rangle - \langle\sigma\sigma\rangle = 0,$$

which implies that the propagators of both fields are identical, when, as here, no global symmetry breaking is included. At tree-level, this is immediately visible, but gives a constraint for the results beyond tree-level.

Of course, this is a rather simple result, and much more interesting ones are obtained at higher order and/or for more complicated theories. E. g., when the transformation is taken to be field-independent, but local, the quantum version of the equations of motion, the Dyson-Schwinger equations, are obtained, as will be exploited below.

4.4.2 Slavnov-Taylor identities

Of course, it is possible to perform the same for a local symmetry, a gauge symmetry. This yields the so-called Slavnov-Taylor identities (STIs). However, it is rather useful to take a different route to obtain them. In particular, the BRST symmetry will be very useful to obtain them much more directly than before.

Take a gauge-fixed theory, in which a BRST symmetry is well-defined and local, i. e., with a gauge-fixing condition at most linear in the fields. Since the vacuum state is physical and thus BRST-invariant, $s|0\rangle = 0$, it follows immediately that

$$0 = \langle s(T\Pi_l\phi_l)\rangle = \sum_k \sigma_k\langle T((\Pi_{l<k}\phi_l)(s\phi_k)(\Pi_{m>k}\phi_m))\rangle$$

where ϕ_l stands for any of the fields in the theory, σ_k is $+1$ if the expression $\Pi_{l<k}\phi_l$ is Grassmann-even, and -1 if it is Grassmann-odd.

A non-trivial example for the usefulness of such an identity is given when regarding the BRST transformation of the two-point correlator $\langle T\bar{u}^a(x)D^b\left[A^a_\mu, y\right]\rangle$, where D^b is the gauge-fixing condition. This yields

$$\begin{aligned}
0 &= \langle s\left(T\bar{u}^a(x)D^b\left[A^a_\mu, y\right]\right)\rangle = \langle T(s\bar{u}^a(x))D^b\left[A^a_\mu, y\right]\rangle - \langle T\bar{u}^a(x)(sD^b\left[A^a_\mu, y\right])\rangle \\
&= \langle Tb^a(x)D^b\left[A^a_\mu, y\right]\rangle - \langle T\bar{u}^a(x)(sD^b\left[A^a_\mu, y\right])\rangle \\
&= -\frac{1}{\xi}\langle TD^a\left[A^a_\mu, x\right]D^b\left[A^a_\mu, y\right]\rangle - \langle T\bar{u}^a(x)(sD^b\left[A^a_\mu, y\right])\rangle
\end{aligned}$$

where in the last line the equation of motion for the Nakanishi-Lautrup field has been used. The next step is to identify the action of the BRST transformation on the gauge-fixing condition. The BRST transformation annihilates any pure functions not depending on the fields by definition. Thus, it requires only to specify the action on the gauge condition D^a.

The result depends on the choice of this condition, and here one will be chosen which is linear in the gauge fields, $D^a = f^{ab}_\mu A^\mu_b$, though f may contain derivatives, though it will not contain integrals in the following. This yields

$$\begin{aligned} 0 &= -\frac{1}{\xi} f^\mu_{ac} f^\nu_{bd} \langle T A^c_\mu A^d_\nu \rangle - \langle T \bar{u}^a(x) (s f^{bc}_\mu A^\mu_c) \rangle \\ &= -\frac{1}{\xi} f^\mu_{ac} f^\nu_{bd} D^{cd}_{\mu\nu} - \langle T \bar{u}^a(x) f^{bc}_\mu D^\mu_{cd} u^d \rangle = -\frac{1}{\xi} f^\mu_{ac} f^\nu_{bd} D^{cd}_{\mu\nu} - \left\langle T \bar{u}^a(x) \frac{\delta S}{\delta \bar{u}^c} \right\rangle, \end{aligned} \quad (4.23)$$

where $D^{cd}_{\mu\nu}$ is the gauge boson's propagator.

To determine the second expression, the quantum equations of motion, the so-called Dyson-Schwinger equations, can be used. Since the path integral is by definition translational-invariant, it follows that

$$0 = \int \mathcal{D}\phi \frac{\delta}{i\delta\phi} e^{iS + i\int d^dx J\phi} = \int \mathcal{D}\phi \left(\frac{\delta S}{\delta \phi} + J \right) e^{iS + i\int d^dx J\phi} = \left\langle T \left(\frac{\delta S}{\delta \phi(x)} + J(x) \right) \right\rangle.$$

Differentiating this expression with respect to $J(y)$ yields

$$0 = \left\langle T \left(i\phi(y) \frac{\delta S}{\delta \phi(x)} + iJ(x)\phi(y) + \frac{\delta J(x)}{\delta J(y)} \right) \right\rangle \overset{J=0}{=} i \left\langle T \phi \frac{\delta S}{\delta \phi} \right\rangle + \delta(x-y), \quad (4.24)$$

where the limit of $J \to 0$ has been taken in the last step.

Thus, an expression like the second term in (4.23) is just a δ function. In the present case, taking the color indices and the Grassmannian nature of the ghost into account, this finally yields

$$f^\mu_{ac} f^\nu_{bd} D^{cd}_{\mu\nu} = i\xi \delta^{ab} \delta(x-y),$$

or for the linear covariant gauges $f^{ab}_\mu = \delta^{ab}\partial_\mu$ in momentum space

$$p^\mu p^\nu D^{ab}_{\mu\nu}(p) = -i\xi\delta^{ab}. \quad (4.25)$$

Thus, the gauge boson propagator's longitudinal part has only a trivial momentum-dependence. This result could also be derived using functional derivatives or directly from the gauge condition, and therefore holds irrespective of the calculational scheme, and in particular beyond perturbation theory.

In the same manner more complicated STIs can be derived. In general, they connect n-point, $n+1$-point, and $n+2$-point correlation functions. They are very useful in perturbation theory, as the $n+2$-point contributions turn out to be always of higher order in the coupling constant than the order at which a perturbative calculation is performed. Beyond perturbation theory, however, their usefulness diminishes quickly.

4.5 Perturbation theory

The most basic method to calculate physical quantities is perturbation theory. The prime example are cross-sections. However, as will become clear when deriving their perturbative evaluation, there are many problems which cannot be addressed using this method.

4.5.1 Cross-sections and decays

The primary quantities of interest at experiments are cross-sections and decay processes. The basic starting point is the quantum mechanical formula for the partial differential cross section $d\sigma$, which has for two incoming particles of mass M_i and momenta p_i and n outgoing particles of momenta q_i the form

$$d\sigma = \frac{1}{4\sqrt{(p_1 p_2)^2 - M_1 M_2}} (2\pi)^4 \delta\left(p_1 + p_2 - \sum_i q_i\right) \frac{d^3\vec{q}_1}{(2\pi)^3 E_{q_1}} \times ... \times \frac{d^3\vec{q}_n}{(2\pi)^3 E_{q_n}} |\mathcal{M}_{fi}|^2,$$

where $\mathcal{M}$ is the transition matrix element between the incoming state i and the outgoing state f. This formula can be generalized to also more than two incoming particles. However, in practice it is very hard in experiments to get any appreciably amount of three-particle collisions, so this plays little role in experimental physics. It is of much more importance in other environments, like the interior of a sun, where the enormous particle fluxes can compensate for the difficulties of colliding three or more particles. However, even in these cases four and more particle collisions are unlikely.

More interesting is the situation with a single particle in the initial state, which decays into an n-particle final state. The corresponding cross section is then called $d\Gamma$, and given by

$$d\Gamma = \frac{1}{2E_p} (2\pi)^4 \delta\left(p - \sum_i q_i\right) \frac{d^3\vec{q}_1}{(2\pi)^3 E_{q_1}} \times ... \times \frac{d^3\vec{q}_n}{(2\pi)^3 E_{q_n}} |\mathcal{M}_{fi}|^2.$$

To get the total values for $d\sigma$ and $d\Gamma$, they have to be integrated over the final momenta q_i for a particular channel, i. e., a particular final state. If identical particles occur, their interchange has to be taken into account, which adds a factor $1/m!$ where m is the number of such identical particles. These give the cross-section for a particular channel, i. e., set of particles in the final state. The final results are obtained after summing over all possible channels.

The central question of quantum field theory is therefore reduced to the calculation of the transition matrix elements $\mathcal{M}$. These are defined as

$$\langle f|\mathcal{S}|i\rangle = \langle f|i\rangle + i(2\pi)^4 \delta(p_i - p_f)\mathcal{M}_{fi}, \tag{4.26}$$

where p_f and p_i are the total initial and final state momenta, and $\mathcal{S}$ is the S-matrix, which is defined as the time-ordered product of the interaction Lagrangian as

$$\mathcal{S} = Te^{i\int d^dx\mathcal{L}_I}$$

where $\mathcal{L}_I$ contains only the parts of the Lagrangian which are more than quadratic in the fields, and the time-ordering operator T is defined as

$$T(\psi(x)\psi(y)) = \theta(x_0 - y_0)\psi(x)\psi(y) \pm \theta(y_0 - x_0)\psi(y)\psi(x),$$

where the minus sign applies if both fields are fermionic. The generalization to an arbitrary number of fields leads then to Wick's theorem.

Since the S-matrix is nothing more than just the time evolution operator, this expression is just given by correlation functions of the operators creating and annihilating the initial and final state, respectively. E. g., for a two muon to two electron process, the expression is the correlation function

$$\langle\mu\mu|S|ee\rangle = R_eR_\mu\langle T(\mu\mu e^\dagger e^\dagger)\rangle,$$

where the R_i are field normalization factors to be discussed latter. The resulting expression is a vacuum-to-vacuum transition amplitude, a so-called correlation function or Green's function. Calculation of these functions is therefore everything necessary to calculate the transition matrix element. This will now be done in perturbation theory using the path integral formalism.

4.5.2 General construction

Already by construction, time-ordered correlation functions can be calculated using the path integral as

$$< T\phi_1...\phi_n >= \left.\frac{\int \mathcal{D}\phi\phi_1...\phi_n e^{iS[\phi,J]}}{\int \mathcal{D}\phi e^{iS[\phi,J]}}\right|_{J=0}. \tag{4.27}$$

However, this is so far only a tautology, as this gives no constructive way of calculating actually the correlation functions. The method of choice used here will be perturbation theory. This essentially boils down to expanding the exponential in the fields, giving essentially an infinite series of quasi-Gaussian integrals. The result is that the transition matrix elements are determined by a sum over correlation functions in a theory with quadratic action. Such an expansion of the field is essentially an expansion around zero field values, and thus assumes that the field amplitudes are small. Hence, this is a perturbative approach.

To do this, split the Lagrangian into a quadratic part $\mathcal{L}_2$ and a remainder part $\mathcal{L}_I$, which includes all the interactions. This yields for the generating functional

$$\begin{aligned} Z[J] &= \int \mathcal{D}\phi e^{i\int d^dx\mathcal{L}_I} e^{i\int d^dx(\mathcal{L}_2+J\phi)} \\ &= e^{i\int d^dx\mathcal{L}_I\left[\frac{\delta}{i\delta J}\right]} \int \mathcal{D}\phi e^{i\int d^dx(\mathcal{L}_2+J\phi)}. \end{aligned}$$

This is only a rewriting of the expression, and is still exact. The argument of $\mathcal{L}_I$ is just indicating that all appearances of the field have been replaced by the derivative with respect to the source. To see the equivalence, take as an example a theory with cubic interaction term

$$\mathcal{L}_I = \frac{\lambda}{3!}\phi^3$$

and expand the exponential

$$\begin{aligned} & e^{i\int d^dx\mathcal{L}_I\left[\frac{\delta}{i\delta J}\right]} \int \mathcal{D}\phi e^{i\int d^dx(\mathcal{L}_2+J\phi)} \\ = & \int \mathcal{D}\phi \sum_n \frac{1}{n!}\left(\frac{\lambda}{3!}i\int d^dy \frac{\delta^3}{i\delta J(y)^3}\right)^n e^{i\int d^dx(\mathcal{L}_2+J\phi)} \\ = & \int \mathcal{D}\phi \left(1+i\int d^dy \frac{\lambda}{3!}\frac{\delta^2}{i\delta J(y)^2}\frac{\delta i\int d^dxJ\phi}{i\delta J(y)}+...\right) e^{i\int d^dx(\mathcal{L}_2+J\phi)} \\ = & \int \mathcal{D}\phi \left(1+i\int d^dy \frac{\lambda}{3!}\frac{\delta^2}{i\delta J(y)^2}\int d^dx\phi\delta(x-y)+...\right) e^{i\int d^dx(\mathcal{L}_2+J\phi)} \\ = & \int \mathcal{D}\phi \left(1+i\int d^dy \frac{\lambda}{3!}\phi\frac{\delta^2}{i\delta J(y)^2}+...\right) e^{i\int d^dx(\mathcal{L}_2+J\phi)} \\ = & \int \mathcal{D}\phi \left(1+i\int d^dy \frac{\lambda}{3!}\phi^3+...\right) e^{i\int d^dx(\mathcal{L}_2+J\phi)} \\ = & \int \mathcal{D}\phi \sum_n \frac{1}{n!}\left(\frac{\lambda}{3!}i\int d^dy\phi^3(y)\right)^n e^{i\int d^dx(\mathcal{L}_2+J\phi)} \\ = & \int \mathcal{D}\phi e^{i\int d^dx\mathcal{L}_I} e^{i\int d^dx(\mathcal{L}_2+J\phi)}. \end{aligned}$$

Such manipulations are very helpful in general.

To proceed it is necessary to perform the remaining shifted Gaussian integral. This can be readily generalized from the formula for ordinary numbers,

$$\int dx e^{-\frac{1}{4}xax+bx} = 2\sqrt{\frac{\pi}{a}}e^{\frac{b^2}{a}}.$$

This yields

$$\begin{aligned} \int D\phi e^{i\int d^dx(\phi(x)(\Omega-i\epsilon/2)\phi(x)+J(x)\phi(x))} &= \int D\phi e^{i\int d^dx(\phi(x)\Omega\phi(x))} e^{-\frac{i}{2}\int d^dxd^dyJ(x)\Delta(x-y)J(y)} \\ &= Z_2[0]\, e^{-\frac{i}{2}\int d^dxd^dyJ(x)\Delta(x-y)J(y)}. \end{aligned} \tag{4.28}$$

There are a number of points to take into account. Ω is just the quadratic part of the Lagrangian, e. g., for a free scalar field it is just $(-\partial^2 - m^2)/2$. The addition of the term $i\epsilon$ is actually needed to make the integral convergent, and has to be carried through all calculations. This can also be formally justified when using canonical quantization. Secondly, the so-called Feynman propagator Δ is defined such that

$$(2\Omega - i\epsilon)\Delta(x-y) = i\delta^d(x-y).$$

That it depends only on the difference $x - y$ comes from the assumption of translational invariance, which applies to the standard model. For a scalar particle of mass M and thus $\Omega = (-\partial^2 - M^2)/2$, this Feynman propagator takes, after Fourier transformation,

$$\begin{aligned}(-\partial^2 - M^2 + i\epsilon)\int d^d p e^{ip(x-y)}\Delta(p) &= i\int d^d p e^{ip(x-y)}\\ \int d^d p e^{ip(x-y)}(p^2 - M^2 + i\epsilon)\Delta(p) &= i\int d^d p e^{ip(x-y)}\end{aligned} \tag{4.29}$$

the form

$$\Delta(p) = \frac{i}{p^2 - M^2 + i\epsilon}, \tag{4.30}$$

which is more useful for a calculation than the rather involved position space expression, which can actually only be described in form of a tempered distribution. Thirdly, the factor $Z_2[0]$ in front of the integral containing the Feynman propagator is just the factor $1/a$ in the conventional integral, conveniently rewritten as an exponential. This factor will cancel partly the denominator in (4.27) when taking the limit $J \to 0$ at the end of the calculation.

This is then sufficient to write down a perturbative calculation of an arbitrary correlation function. Take, for example, a model of a single scalar particle with interaction

$$\mathcal{L}_I = -\frac{\lambda}{4!}\phi^4.$$

The perturbative expression up to linear order in $\lambda/4!$ for a process involving two particles in the initial and final state, essentially elastic scattering, is then

$$\begin{aligned}&\langle T\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\rangle\\ =\;& \left.\frac{\int \mathcal{D}\phi\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)e^{iS[\phi,J]}}{\int \mathcal{D}\phi e^{iS[\phi,J]}}\right|_{J=0}\\ =\;& \frac{1}{Z[0]}\frac{\delta^4}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)}\left.\int \mathcal{D}\phi e^{iS[\phi,J]}\right|_{J=0}\\ =\;& \left.\frac{Z_2[0]}{Z[0]}\frac{\delta^4}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)}e^{i\int d^d x\mathcal{L}_I\left[\frac{\delta}{i\delta J}\right]}e^{-\frac{i}{2}\int d^d x d^d y J(x)\Delta(x-y)J(y)}\right|_{J=0}.\end{aligned}$$

The next step is to expand both exponentials, the first in a formal power series in $\mathcal{L}_I$, and the second one in the conventional exponential series,

$$\begin{aligned} = & \frac{Z_2[0]}{Z[0]} \frac{\delta^4}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)} \left(\sum_n \frac{1}{n!} \left(\frac{\lambda}{4!} i \int d^d y \frac{\delta^4}{i\delta J(y)^4} \right)^n \right) \times \\ & \times \left(\sum_m \frac{1}{m!} \left(-\frac{i}{2} \int d^d x d^d y J(x) \Delta(x-y) J(y) \right) \right) \Bigg|_{J=0} . \end{aligned}$$

Both are polynomial in the sources. The expansion of the exponential of the interaction Lagrangian yields terms with zero, four, eight,... derivatives with respect to the sources. The second term produces terms with zero, two, four,... powers of the sources. Since the sources are set to zero at the end, only terms without sources will remain. Thus, to order zero in the interaction Lagrangian only the term with four sources will survive the external derivative. To first order in the interaction Lagrangian only the term with eight powers of the sources will survive.

To this order in the expansion, the expression takes therefore the form

$$\begin{aligned} = & \frac{Z_2[0]}{Z[0]} \frac{\delta^4}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)} \left(\frac{1}{2!} \left(-\frac{i}{2} \int d^d x d^d y J(x) \Delta(x-y) J(y) \right)^2 \right. \\ & \left. -\frac{i\lambda}{4!} \int d^d z \frac{1}{4!} \frac{\delta^4}{\delta J(z)^4} \left(-\frac{i}{2} \int d^d x d^d y J(z) \Delta(x-y) J(y) \right)^4 + \mathcal{O}(\lambda^2) \right) . \end{aligned}$$

In principle, taking the derivatives is straight-forward. However, e. g., the first term is given by the expression

$$\frac{\delta^4}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)} \int d^d y_1 d^d y_2 d^d y_3 d^d y_4 J(y_1)\Delta(y_1-y_2)J(y_2)J(y_3)\Delta(y_3-y_4)J(y_4).$$

The first derivative, with respect to $J(x_4)$ could act equally well on all four sources under the integral. It will therefore provide four terms. Correspondingly, the second derivative can act on three different terms, making this 12 terms, and so on, giving in total 24 terms, with all possible combinations, or partitions, of the four arguments.

To illustrate the process, two steps for a particular combination will be investigated. The first derivative acts as

$$\begin{aligned} & \frac{\delta}{\delta J(x_4)} \int d^d x d^d y J(x) \Delta(x-y) J(y) \\ = & \int d^d x d^d y \delta^d(x-x_4) \Delta(x-y) J(y) + ... = \int dy \Delta(x_4-y) J(y) + ..., \end{aligned} \tag{4.31}$$

where the points indicate further contributions. For the action of the next derivative, there are two possibilities. Either it acts on the same factor of the product of the integrals, or

on a different one. Take first the possibility of the same factor. If it is a distinct factor, this just provides the same action. If it is the same factor, this immediately yields

$$\frac{\delta}{\delta J(x_3)}\int d^d y \Delta(x_4-y)J(y) = \int d^d y \Delta(x_4-y)\delta^d(y-x_3) = \Delta(x_4-x_3)$$

In total, this yields for the term proportional to $\lambda^0 = 1$

$$A = -\sum_{P_{ijkl}} \Delta(x_i-x_j)\Delta(x_l-x_k), \tag{4.32}$$

where P_{ijkl} indicates that the sum is over all 4! possible permutations of the index set $\{ijkl\}$.

The situation becomes somewhat more complicated for the terms proportional to λ, since now multiple derivatives with respect to the same source $J(x)$ appears. Again, a single such derivative acts like (4.31). A difference occurs when the second derivative occurs. This can either act again on another factor, but it could also act on the same factor. The first case just produces another factor of type (4.31). The second situation is different, and yields

$$\frac{\delta}{\delta J(z)}\int d^d y \Delta(z-y)J(y) = \int d^d y \Delta(z-y)\delta^d(y-z) = \Delta(z-z) \tag{4.33}$$

which appears to look like $\Delta(0)$. However, this not quite the case, as will be visible later. In particular, the expression $\Delta(0)$ cannot be easily interpreted. Furthermore, an integral over z still appears. It is therefore useful to keep first explicit terms of $\Delta(z-z)$ in the following.

After a slight change in notation, there will then be 8! possibilities for the order λ contribution. However, many of them turn out to be identical, yielding in total three further contributions

$$\lambda B = -i\lambda \int d^d x \Delta(x-x_1)\Delta(x-x_2)\Delta(x-x_3)\Delta(x-x_4) \tag{4.34}$$

$$\lambda C = -\frac{i\lambda}{2}\sum_{P_{ijkl}} \Delta(x_i-x_j)\int d^d x \Delta(x-x)\Delta(x-x_k)\Delta(x-x_l) \tag{4.35}$$

$$\lambda A D = -\frac{i\lambda}{8}\int d^d x \Delta(x-x)\Delta(x-x)\sum_{P_{ijkl}} \Delta(x_i-x_j)\Delta(x_k-x_l). \tag{4.36}$$

These four terms have simple interpretations, if each factor of Δ is considered to be a particle propagation along the connecting line of $x-y$. Then, the first term (4.32) corresponds to the interference pattern of identical particles when they are observed at two different initial and final positions: Since the particles are identical, any combination

is possible, including that one particle vanishes and the other one appears. This can be visualized by using a line to symbolize a factor of Δ, and draw all possible combinations between the four points.

Similar interpretations hold for the three remaining terms (4.34-4.36). The expression (4.34) contains for each factor of Δ a common point. This can be taken to be just a meeting of all four particles at a common vertex point x. Since there appears a pre-factor of λ, it can be said that the four particles couple with a strength λ, thus also the name coupling constant for λ. Such an interaction vertex could be denoted by a dot.

The third term (4.35) can be seen as one particle just propagating, while the second particle has an interesting behavior: It emits at an intermediate point a particle, an reabsorbs it then. Such a virtual particle contributes to a cloud of virtual emission and absorption processes, which becomes more common at higher orders. Pictorially, this corresponds to a loop in the propagation, which again harbors an interaction vertex.

The last contribution is different, as when drawing lines there appears an additional graph, which is disconnected from the initial and final positions, and has the form of the number eight. Such a disconnected diagram is also called vacuum contribution, as it is not connected to any external input, and is thus a property of the vacuum alone.

In general, the expression (4.32-4.36) are very cumbersome to deal with in position space. It is therefore more useful to perform a Fourier transformation, and perform the calculations in momentum space. In particular, this removes many of the cumbersome sums over partitions. How to switch to momentum space will be discussed in more detail after taking care of the remaining factor $Z_2[0]/Z[0]$.

Since the current calculation is a perturbative calculation, it is adequate to also expand $Z_2[0]/Z[0]$ in λ. This can be most directly done again using the formula (4.28). Thus, the factor Z_2 cancels immediately, and the remaining expansion terms are, up to combinatorial factors, very similar as before. Its inverse is thus given, to order λ, by

$$\frac{Z[0]}{Z_2[0]} = 1 + \frac{i\lambda}{2^3}\int d^dx \Delta(x-x)\Delta(x-x) + \mathcal{O}(\lambda^2) = 1 + \lambda D.$$

This term is easily identified as the prefactor appearing in (4.36). To order λ, this yields

$$\begin{aligned}
&\langle T\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\rangle = \frac{A+\lambda(B+C+AD)}{1+\lambda D} + \mathcal{O}(\lambda^2)\\
=\;& (A+\lambda(B+C+AD))(1-\lambda D) + \mathcal{O}(\lambda^2) = A+\lambda(B+C+AD) - \lambda AD + \mathcal{O}(\lambda^2)\\
=\;& A+\lambda(B+C) + \mathcal{O}(\lambda^2).
\end{aligned}$$

Thus, to order λ, the term with a disconnected contribution is canceled. It turns out that this is a generic result, and that all diagrams with disconnected contribution in a

perturbative expansion always cancel, and a general proof can be constructed in a very similar way to this evaluation in ϕ^4 theory up to leading order. However, this is beyond the scope of this lecture.

As stated, the explicit expression in position space turns out to be very awkward to use in actual calculation, and their evaluation in momentum space is preferable. This can be done using the expression for the Feynman propagator in momentum space, (4.30). The total sum then becomes

$$\begin{aligned}
&(2\pi)^d \int \frac{d^d p_1}{(2\pi)^{\frac{d}{2}}} \frac{d^d p_2}{(2\pi)^{\frac{d}{2}}} \frac{d^d p_3}{(2\pi)^{\frac{d}{2}}} \frac{d^d p_4}{(2\pi)^{\frac{d}{2}}} e^{-i(p_1 x_1 + p_2 x_2 + p_3 x_3 + p_4 x_4)} \times \\
&\times \left(\sum_{P_{ijkl}} (2\pi)^d \delta^d(p_k + p_l) \delta^d(p_i + p_j) \frac{i}{p_i^2 - m^2} \frac{i}{p_k^2 - m^2} \right. \\
&-i\lambda \delta^d(p_1 + p_2 + p_3 + p_4) \frac{i}{p_1^2 - m^2} \frac{i}{p_2^2 - m^2} \frac{i}{p_3^2 - m^2} \frac{i}{p^2 - m^2} \\
&\left. -\frac{(2\pi)^d \lambda}{2} \sum_{P_{Ijkl}} \delta^d(p_i + p_j) \delta^d(p_k + p_l) \frac{i}{p_i^2 - m^2} \frac{i}{p_k^2 - m^2} \frac{i}{p_l^2 - m^2} \int \frac{d^d q}{(2\pi)^{\frac{d}{2}}} \frac{i}{q^2 - m^2} \right) .
\end{aligned}$$

Note that the $i\epsilon$ contributions have not been written explicitly in the propagators, but left implicit. This is the standard conventions for such a representation of a perturbative expression. Of course, if the result is desired in momentum space rather than position space, which is normally the case, the Fourier transformation can be dropped.

The result already shows a number of regularities, which can be generalized to the so-called Feynman rules, which permit to directly translate from a graphical representation to the mathematical expression in perturbation theory. These can be derived rather generally, though this becomes rather cumbersome. Here, these will be stated simply without proof:

- Select the type and number of all external lines
- Determine the order (in all coupling constants, i. e., in all vertices) to which the process should be evaluated
- Draw all possible diagrams connecting in all possible ways the external lines with up to order vertices, and add them
- For each line, write a propagator of this particle type
- For each vertex, write the interaction vertex, i. e. essentially $\delta^n \mathcal{L}_I / \delta \phi^n$, for each
- Impose the conservation of all quantities, including momentum, conserved by a giving vertex at each vertex. This can be most directly done by following each input conserved quantity through the whole diagram until its final result

- Integrate over all undetermined momenta, i. e., each momentum running through a loop
- For each closed fermion loop, multiply the term by minus one, because of the Grassmann nature
- Lines, which are attached to the outside of a diagram, receive a further propagator of the corresponding type

Two things can further facilitate the result. On the one hand, any diagram will be zero, if any conservation law is not respected by the transition from initial to final state. However, in such cases also the perturbative vertices vanish identically. Secondly, there are many diagrams, which are identical up to reordering, as in the previous example. They can be collected, and normalized using so-called symmetry factors.

It can be immediately shown that the previous results can be obtained from these rules, as an explicit example of the more general Feynman rules.

Perturbative calculations in QCD with the Lagrangian (4.1) can be performed in essentially the same way. It is just necessary to replace it with the corresponding propagators and vertices derived in the same way from the Lagrangian. These are the propagators for the gluons, ghosts, and quarks

$$\begin{aligned} D^{ab}_{\mu\nu} &= \frac{-i\delta^{ab}}{k^2+i\epsilon}\left(g_{\mu\nu}-(1-\xi)\frac{k_\mu k_\nu}{k^2}\right) \\ D^{ij} &= \delta_{ij}\frac{i(p_\mu\gamma^\mu+m)}{p^2-m^2+i\epsilon} = \frac{i\delta_{ij}}{p_\mu\gamma^\mu-m+i\epsilon} \\ D^{ab} &= \frac{i\delta^{ab}}{p^2+i\epsilon}, \end{aligned}$$

respectively. There are in addition 3-point vertices of quarks and gluons, three gluons, and ghost and gluons

$$\begin{aligned} \Gamma^{gq\bar{q}}(p,q,k)^{aij}_{\mu} &= ig\gamma^\mu t^a_{ij} \\ \Gamma^{ggg}(p,q,k)^{abc}_{\mu\nu\rho} &= gf^{abc}(g^{\mu\nu}(k-p)^\rho + g^{\nu\rho}(p-q)^\mu + g^{\rho\mu}(q-k)^\nu) \\ \Gamma^{\bar{u}gu}(p,q,k)^{abc}_{\mu} &= gf^{abc}p^\mu, \end{aligned}$$

respectively, and the four-gluon vertex

$$\begin{aligned} \Gamma^{gggg}(p,q,k,l)^{abcd}_{\mu\nu\rho\sigma} &= -ig^2(f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma}-g^{\mu\sigma}g^{\nu\rho}) \\ &\quad +f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma}-g^{\mu\sigma}g^{\nu\rho})+f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma}-g^{\mu\rho}g^{\nu\sigma})), \end{aligned}$$

and all momenta defined incoming. Note that because ghosts obey Fermi statistics closed ghost loops also receive a minus sign.

Returning to the scalar theory, perturbative calculations can be further simplified by passing to connected, amputated diagrams.

The so-called connected diagrams are diagrams in which all lines are connected with each other. In the previous case, the result can be symbolically written as

$$\Delta\Delta + \Delta\Delta' + \Pi,$$

where Δ is a propagator, Δ' is a propagator with a loop attached, and Π is the graph where all four lines are connected. This is called the full correlation function. Of course, Δ and Δ' can also be determined from the two-point function $\langle T\phi\phi\rangle$, to the same order, and therefore contain no new information. The only new contribution for the four-point function at this order of perturbation theory is Π. It would therefore be useful, if it is possible to only calculate this contribution, instead of the whole one. Indeed, it can be shown that for a correlation function with n external legs

$$\begin{aligned} G(x_1,...,x_n) &= G_c(x_1,...,x_n) + \sum G_c(x_i,...,x_j)G_c(x_j,...,x_k) \\ &\quad + \sum G_c(x_i,...,x_j)G_c(x_k,...,x_l)G_c(x_m,...,x_n) + ... \end{aligned} \tag{4.37}$$

where the sums are over all possible ways to split the index set $\{x_i\}$ in two, three,... subsets. Furthermore, every connected correlation functions is a series in the coupling constant. Thus, in the present case,

$$\begin{aligned} G(x_1,x_2,x_3,x_4) &= G_c(x_1,x_2,x_3,x_4) + \sum_{P_{ijkl}} G_c(x_i,x_j)G_c(x_k,x_l) \\ G_c(x,y) &= \Delta(x-y) + \Delta'(x-y), \end{aligned}$$

where again the δ is the propagator, and Δ' is the propagator to order λ, which includes the attached loop, and G_c is the only diagram with all points connected. Contributions proportional to $\Delta^{'2}$ have to be dropped, as they are of higher order in the perturbative expansion. This relation can be inverted to obtain the connected functions from the other, but it is more interesting to calculate just the connected, and then calculate the complete one by the formula (4.37).

Finally, all external lines have the propagators attached to them, they are called non-amputated. Removing this yields the amputated correlation functions Γ, which can immediately yield again the non-amputated one. Thus, it is sufficient to calculate the amputated ones. In the same way, explicit momentum conserving factors can always be reinstantiated.

Thus, the calculation of the four-point function boils finally down to the calculation of the amputated, connected two-point function to order λ, and the amputated, connected four-point function. These are just given by

$$\begin{aligned}\Gamma_c(p,q) &= -i\lambda\int\frac{d^dr}{(2\pi)^4}\frac{i}{r^2-m^2}\\ \Gamma_c(p,q,k,l) &= -i\lambda,\end{aligned}$$

rather simple expressions indeed. These are also called proper or vertex correlation functions.

There is a further possibility to reduce the effort of perturbative calculations, though these do not reduce it further for the present example. It is rather simple to imagine situations, were it is possible to cut a single internal line to obtain two separate graphs. Such graphs are called one-particle reducible. It can be shown that it is sufficient two know all graphs, which cannot be separated in such a way, so-called one-particle irreducible graphs (1PI), to obtain all relevant results, and to reconstruct also the one-particle reducible ones. The generic connection can again be illustrated. Take two graphs which are 1PI, say graphs $A(p,q)$ and $B(k,l)$. They can be joined to a one-particle reducible graph by

$$A(p,q)\Delta(q)B(q,l),$$

i. e., by the insertion of a propagator. This can be repeated as necessary.

Thus, the final addition to the Feynman rules is

- Identify in all the diagrams the connected, amputated 1PI graphs. Calculate these, and the result can be obtained by just multiplying and adding the results together such as to obtain the original graphs

Note that the construction can be extended further, to so-called nPI graphs. However, their recombination is in general no longer possible by multiplications, but usually involves integration over intermediate momenta. This is beyond the scope of this lecture.

From this construction it follows that there are two distinct classes of perturbative calculations. One is the class of so-called tree-level calculations, in which no loops appear. Since graphs without loops are always one-particle reducible, they can always be cut so long as only to consist out of vertex and propagator expressions. On the other hand, this implies that a tree-level calculation can always be written as just a multiplication of propagators and vertices, without any integration. These contributions turn out to be furthermore the classical contribution, i. e., whatever remains when taking the limit of $\hbar \to 0$. Nonetheless, even tree-level calculations, in particular for many external particles, can become very cumbersome, and both a technical as well as a logistical problem.

The second type of diagrams are all graphs with loops. Since they vanish in the classical limit, this implies that these are the quantum, or also radiative, corrections to a process. The integrals make an evaluation much more complicated. Furthermore, the integrals are usually not finite, leading to the necessity of the renormalization process.

4.6 Renormalization

So far, all calculations have been at tree-level, i. e., no integrations have been necessary, as are required by the Feynman rules if loops appear. Such loop expressions are always of higher order in the coupling constants than the corresponding tree-level diagrams. However, experimental precision is sufficiently high to be sensitive to loop contributions, so-called radiative corrections. In fact, for QCD next-to-next-to-next-to-leading order (N^3LO) precision is already required, and for some quantities even N^4LO results are available[5].

One of the generic problems of such loop corrections is that the corresponding integrals are usually divergent. At first sight, this might seem to invalidate the theory. However, it turns out that it is possible to make the integrals convergent without introducing additional parameters into the theory, albeit at the price that the theory still looses its validity at some high cutoff-scale. Since this scale can be pushed to very high energies, this is of little practical importance, as it can anyway not be assumed that the standard model is a theory of everything, since it does not include gravity.

To make sense out of such a theory requires then two basic steps. One is a prescription how to regularize integrals, i. e., how to map their divergent value to a finite value. For this purpose of regularization the integrals are made convergent by the introduction of some parameter, and the original divergence is recovered when sending this parameter to a particular limit. As a result, all quantities calculated will depend on this parameter.

The second step, the so-called renormalization program, gives a prescription how to redefine the theory such as to loose the dependence on this extra parameter, the so-called renormalization scheme, without recovering the original divergence. The consequence of this program, and the particular renormalization scheme used, is that quantities like masses or coupling constants can no longer be interpreted as static quantities, but will depend on the scale at which they are measured. It is said that they become running. However, measurable quantities, like a cross-section, turn out not to depend on the measurement

[5]Note that in the standard model context calculations may be at different order in the various appearing couplings, as the quantitative contributions, due to the differing sizes of the relevant couplings, are vastly different. Thus, often calculations are done to the orders such that the quantitative contributions of all involved interactions are of roughly the same size, rather than of the same order.

scale, at least for an exact calculation. Unfortunately, most calculations are not exact in general, and in particular for the standard model. As a consequence, a dependence on the scale may be left.

4.6.1 Cutoff regularization

To illustrate the concept of regularization, it is useful to go to a simple model, the Yukawa model of a scalar ϕ and a fermion χ. This model appears throughout strong interaction physics, where the fermions are often interpreted as either quarks or nucleons, and the bosons as mesons. The simplest case with just one flavor each has a Lagrangian given by

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \bar{\chi}i(\gamma^\mu\partial_\mu - m)\chi - \frac{M^2}{2}\phi^2 - \frac{\lambda}{4!}\phi^4 - y\phi\bar{\chi}\chi.$$

Hence there are two masses, m and M, and two coupling constants y and λ. With two flavors of fermions, representing the nucleons, and three flavors of mesons, representing the pions, this model gives already a first reasonable approximation of nuclear physics.

Start with the self-energy of the scalar particle to order $\mathcal{O}(\lambda^1, y^0)$. In this case, there is only one diagram contributing, a so-called tadpole diagram. Its value is

$$\Pi^\lambda_\phi = -\frac{\lambda}{2}\int\frac{d^4p}{(2\pi)^4}\frac{1}{p^2 - M^2 + i\epsilon}, \tag{4.38}$$

where the factor $1/2$ is a symmetry factors. The integration over p_0 can be performed first by contour-integration and using the Cauchy theorem, since

$$\begin{aligned}\Pi^\lambda_\phi &= -\frac{\lambda}{2}\int\frac{d^3\vec{p}}{(2\pi)^4}\int dp_0\frac{1}{p_0^2 - \vec{p}^2 - M^2 + i\epsilon} \\ &= -\frac{\lambda}{2}\int\frac{d^3p}{(2\pi)^4}\int_{-\infty}^{\infty} dp_0\frac{1}{(p_0 + \sqrt{\vec{p}^2 + M^2})(p_0 - \sqrt{\vec{p}^2 + M^2}) + i\epsilon}\end{aligned}$$

This has a pole in the upper half-plane, and vanishes sufficiently fast on a half-circle at infinity. The residue at the simple poles $p_0 = \pm\sqrt{\vec{p}^2 + M^2}$ is $1/(p_0 \mp \sqrt{\vec{p}^2 + M^2})$, dropping the small contribution of $i\epsilon$, which only served to not have the pole on the axis. The Cauchy theorem then yields, using polar coordinates in the final expression,

$$\Pi^\lambda_\phi = \frac{i\lambda}{4\pi^2}\int_0^\infty \vec{p}^2 d|\vec{p}|\frac{1}{\sqrt{\vec{p}^2 + M^2}}.$$

This integral is divergent, as announced. It is also the only contribution at this order of perturbation theory, so there is no cancellation possible to remove this divergence. To

make sense of it, it is necessary to regularize it. The most straight-forward possibility is to replace the upper integral limit ∞ by a large, but finite number Λ, the so-called cutoff regularization.

The integral can then be calculated explicitly to yield

$$\Pi_\phi^\lambda = \frac{i\lambda}{4\pi^2}\left(\Lambda^2\sqrt{1+\frac{M^2}{\Lambda^2}} - M^2\ln\left(\frac{\Lambda+\Lambda\sqrt{1+\frac{M^2}{\Lambda^2}}}{M}\right)\right). \tag{4.39}$$

As can be seen, the integral diverges with the cutoff Λ quadratically, and has in addition a sub-leading divergence logarithmically in Λ. Still, as long as the limit is not performed, the result is finite, independent of the momentum, but explicitly dependent on Λ.

4.6.2 Renormalization prescription

To remove this dependence, it is worthwhile to investigate the total structure of the two-point function $\Gamma_{\phi\phi}$, which is just the propagator $D_{\phi\phi}$. Amputation of the unamputated equation

$$\Gamma_{\phi\phi} = D_{\phi\phi}(p^2 - M^2 + \Pi_\phi)D_{\phi\phi}$$

yields the expression for the amputated and connected two-point function by division, giving

$$\frac{1}{D_{\phi\phi}} = p^2 - M^2 + \Pi_\phi.$$

However, in a perturbative setting the self-energy is assumed to be small. Thus, it is possible to expand the self-energy, and replace it as

$$\frac{1}{D_{\phi\phi}} = p^2 - M^2 + \Pi_\phi^\lambda. \tag{4.40}$$

To leading order the propagator is then given by

$$D_{\phi\phi} = \frac{p^2 - M^2 + \Pi_\phi^\lambda}{(p^2 - M^2 + i\epsilon)^2}.$$

Instead of using this approximate expression it is possible to use the inversion of the expression (4.40). This results in the so-called resummed propagator, as it contains contributions which are of higher-order in the coupling constant.

Diagrammatically, it corresponds to an infinite series of diagrams with an ever-increasing number of tadpole attachments. This already illustrates that this is only a partial resummation of the perturbative series, since at order λ^2 there are also other types of diagrams

contributing. Thus, this looses some of the systematics of the perturbative expression, and it is necessary to be wary with it.

Nonetheless, for the current purpose, it is more transparent to work with the expression (4.40). As is seen from the result (4.39), the contribution Π_ϕ^λ is momentum-independent and dependent on the cutoff Λ. If it would be finite, it could be interpreted as a change of the mass M, since then the expression would have the form

$$p^2 - M^2 - \delta M^2 \to p^2 - M_R^2$$

with the renormalized mass

$$M_R = \sqrt{M^2 + \delta M^2}.$$

The actual mass of a ϕ particle, which would be measured in an experiment, would then be M_R, instead of the bare mass M. In fact, since the experimental measurement is the only knowledge available on the theory, it is mandatory that the bare parameters of the theory, like the bare mass M, are adjusted such that the resulting renormalized mass M_R agrees with experiment[6].

Now, since the actual bare parameters cannot be measured, there is nothing which prevents us to set it to

$$M^2 = M_R^2 - \delta M^2,$$

with the experimental input M_R. This automatically fulfills the requirement to reproduce the experiment. In particular, since M is not an observable quantity, there is no reason for it to be finite, and independent of the cutoff Λ. Thus, it is possible to absorb the infinity of the divergent integral in unobservable bare parameters of the theory. This can be arranged already at the level of the Lagrangian by replacing

$$\frac{M^2}{2}\phi^2 \to \frac{{M_R}^2}{2}\phi^2 - \frac{\delta M^2}{2}\phi^2.$$

The second term is a so-called counter-term, and it depends on the actual order of the calculation. E. g., at tree-level, it would be zero. This replacement is called a renormalization scheme.

4.6.3 Counter-term structure

It is actually not the the only contribution which appears. If the calculation is extended to also include corrections up to $\mathcal{O}(\lambda, y^2)$, there is a second diagram contributing to the

[6]This implies that the bare parameters have to be adapted at each order of perturbation theory calculated.

self-energy, which is due to a loop of the fermions. The expression then takes the form

$$\Pi_\phi^{\lambda,y^2} = \Pi_\phi^\lambda + \Pi_\phi^{y^2},$$

with the fermionic contribution given by

$$\Pi_\phi^{y^2} = -\frac{y^2}{2}\int \frac{d^4p}{(2\pi)^4}\frac{\mathrm{tr}((\gamma_\mu p^\mu + M)(\gamma_\nu(p^\nu - q^\nu) + M))}{(p^2 - M^2 + i\epsilon)((p-q)^2 - M^2 + i\epsilon)}.$$

Using the trace identities tr1 = 4, trγ_μ=0, and tr$\gamma_\mu\gamma_\nu = 4g_{\mu\nu}$ this simplifies to

$$-\frac{y^2}{2}\int \frac{d^4p}{(2\pi)^4}\frac{p(p-q) + M^2}{(p^2 - M^2 + i\epsilon)((p-q)^2 - M^2 + i\epsilon)}.$$

Since the numerator scales with p^2, the integral is quadratically divergent. Suppressing the $i\epsilon$, the expression can be rewritten by introducing a zero and then shifting the integration argument, as

$$\begin{aligned}
& -\frac{y^2}{2}\int \frac{d^4p}{(2\pi)^4}\frac{(p^2 - m^2) + ((p-q)^2) - m^2) - q^2 + 4m^2}{(p^2 - m^2)((p-q)^2 - m^2)} \\
= & -\frac{y^2}{2}\int \frac{d^4p}{(2\pi)^4}\left(\frac{1}{(p-q)^2 - m^2} + \frac{1}{p^2 - m^2} + \frac{4m^2 - q^2}{(p^2 - m^2)((p-q)^2 - m^2)}\right) \\
= & -\frac{y^2}{2}\int \frac{d^4p}{(2\pi)^4}\left(\frac{2}{p^2 - m^2} + \frac{4m^2 - q^2}{(p^2 - m^2)((p-q)^2 - m^2)}\right),
\end{aligned}$$

Such integrals can be performed using a number of analytical tricks. However, for the present purpose this will not be necessary. It is sufficient to observe that the resulting integral, just by counting powers of integration momenta, will have the form

$$\Pi_\phi^{y^2} = c_1\Lambda^2 + (c_2 m^2 + c_3 q^2)\ln\frac{\Lambda}{m} + f(m^2, q^2),$$

where f is some finite function when Λ is send to infinity, and depends on both λ and y, as do the constants c_i.

The first two terms have again the same structure as the tadpole contribution (4.38). However, the third term is different, as it does depend explicitly on the momentum. Therefore, it cannot be absorbed into a mass renormalization. However, it can be absorbed in a renormalization of the kinetic term. If in the Lagrangian the modification

$$\partial_\mu\phi\partial^\mu\phi \to \partial_\mu\phi\partial^\mu\phi + \delta Z_\phi\partial_\mu\phi\partial^\mu\phi = Z_\phi\partial_\mu\phi\partial^\mu\phi,$$

is performed, the kinetic term of the field ϕ has been renormalized by a factor Z_ϕ. Choosing

$$\delta Z_\phi = -c_3\ln\frac{\Lambda}{m},$$

this will remove the divergence. By this the field amplitude is arranged to agree with the experimental one by the introduction of the wave-function renormalization $Z_\phi^{\frac{1}{2}}$.

Performing further calculations, it turns out that similar changes have to be performed for the remaining bare parameters m, λ, and y, yielding a renormalized fermion mass m_R, and renormalized couplings λ_R and y_R. Thus, including these counter-terms yields the renormalized Lagrangian

$$\begin{aligned}\mathcal{L}_R &= \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \bar{\chi}i(\gamma^\mu\partial_\mu - m_R)\chi - \frac{{M_R}^2}{2}\phi^2 - \frac{\lambda_R}{4!}\phi^4 - y_R\phi\bar{\chi}\chi \\ &+\frac{\delta Z_\phi}{2}\partial_\mu\phi\partial^\mu\phi + \bar{\chi}i(\delta Z_\chi\gamma^\mu\partial_\mu - \delta m)\chi - \frac{\delta{M}^2}{2}\phi^2 - \frac{\delta\lambda}{4!}\phi^4 - \delta y\phi\bar{\chi}\chi.\end{aligned}$$

It should be noted that always certain products of fields appear together with a parameter of the theory. Thus, often explicit factors of various Zs are introduced such that not kinetic terms are renormalized, but rather the field itself, in the sense of an amplitude renormalization. In this case, explicit factors of $Z_i^{1/2}$ are multiplied for each field in the counter-term Lagrangian, and the counter-terms δM, δm, $\delta\lambda$, and δy are redefined by appropriate factors of $Z_i^{-1/2}$. This is, however, conventional, but the more common case for QCD.

Also, it is usual that δx is rather defined as

$$\delta x = Z_x x = (1 + \delta Z_x)x,$$

i. e. as a multiplicative factor to the original quantity. However, Z_x may then depend again on x, even in the form of $1/x$. E. g., renormalized QED reads then

$$\begin{aligned}\mathcal{L}_{QED} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2\xi}(\partial^\mu A_\mu)^2 + \bar{\psi}(i\gamma_\mu + m)\psi + eA_\mu\bar{\psi}\gamma^\mu\psi \\ &-\frac{\delta Z_A}{4}F_{\mu\nu}F^{\mu\nu} - \frac{\frac{Z_A}{Z_\xi} - 1}{2\xi}(\partial_\mu A^\mu)^2 + \delta Z_\psi\bar{\psi}i\gamma^\mu\partial_\mu\psi - (Z_\psi Z_m - 1)m\bar{\psi}\psi \\ &+(Z_e Z_A^{-\frac{1}{2}} Z_\psi - 1)eA_\mu\bar{\psi}\gamma^\mu\psi.\end{aligned}$$

In this case all parameters, m, e, and ξ, as well as A_μ and ψ have been multiplicatively renormalized. It should be noted that also the ghost fields would have to be renormalized, if they would not decouple in QED. The extension of this to QCD is hence a straightforward generalization.

The remaining question is then whether this is sufficient, or whether further terms, e. g. a sixth power of the fields, would be necessary, or whether non-multiplicative terms would appear. It can be shown that in perturbation theory in four dimensions for any gauge theory of the type of QCD with parameters with at least zero energy dimension, i. e.

dimensionless couplings, masses, or couplings with dimensions of energy to some positive power, it is always possible to perform the renormalization with a finite number of terms. Thus, the process is finite, and for QED actually complete at this stage. However, this is only proven in perturbation theory, and though it is commonly assumed to hold also beyond perturbation theory, a proof is lacking.

The general proof, also for dimensions different than four, and more complex theories is possible, but beyond the current scope. However, for all known quantum gauge theories in four dimensions with non-trivial dynamics and observable bound states renormalization is necessary.

4.6.4 Renormalization schemes and dimensional transmutation

So far, the counter terms have been identified by direct comparison. However, assume that the propagator has finally the form

$$D = \frac{c^2 - d^2 + 2p^2}{p^4 + (d^2 - c^2)p^2 - c^2 d^2}.$$

Such a propagator has no longer the form of a conventional free particle. It is thus not clear how to determine, e. g., δm, such that it represents the mass of a particle. Thus, it is necessary to give a more precise definition of what physical mass means. Since such a mass would be expected as a pole, one possibility would be to choose it as the smallest momentum at which the propagator has a simple pole. In this case, this would imply

$$m_R = d,$$

and thus the counter-terms can be arranged such that this equality holds. This is called a pole scheme.

It becomes much more ambiguous for the coupling constants, as they are not associated with some pole. For the electromagnetic charge, it still seems reasonable to choose its macroscopic value, i. e., the one known from classical physics, which is the so-called Thomson limit. A similar definition cannot be made for, e. g. the strong coupling. Another possibility is therefore, e. g., to choose

$$\Gamma^{A\bar{\psi}\psi}(p, q, p + q) \stackrel{!}{=} e,$$

for two arbitrarily chosen momenta. This already shows that a certain ambiguity is introduced, because a scale μ is introduced, which is proportional to p, at a fixed ratio of p and q. It is even more ambiguous when it comes to identify conditions for the wave-function renormalization.

In fact, it turns out that the conditions chosen are arbitrary, i. e., any genuinely measurable quantity is not depending on this choice[7]. Thus, any choice will do. Any such set of choices is called a renormalization scheme, and it is possible to express quantities using one renormalization scheme by results in a different renormalization scheme. For QED, e. g., it is possible to define the following set of renormalization conditions

$$\left(p^2 g_{\mu\nu} - p_\mu p_\nu\right)_{p^2=\mu^2} D^{\mu\nu}_{AA}(\mu^2) = i \tag{4.41}$$

$$\mu^2 g_{\mu\nu} D^{\mu\nu}_{AA}(\mu^2) = i\xi \tag{4.42}$$

$$\mathrm{tr} D_{\psi\psi}(\mu^2) = i m(\mu^2) \tag{4.43}$$

$$(\mathrm{tr}\gamma_\mu p^\mu D_{\psi\psi})(\mu^2) = i16\mu^2 \tag{4.44}$$

$$\Gamma^{A\bar{\psi}\psi}(\mu^2, \mu^2, \mu^2) = ie(\mu^2) \tag{4.45}$$

Note that there is no condition that involves a mass of the photon. It can be shown that such a mass would violate gauge invariance. The condition (4.42) follows actually directly from the QED version of the STI (4.25). There are two remarkable, and generic, features in this description.

One is that in the definition of the renormalization constants appears a scale μ, the so-called renormalization scale. Its value is arbitrary, but it cannot be removed. Of course, it would be possible to choose for each of the five conditions (4.41-4.45) a different scale, but these would then differ only by constant prefactors multiplying the single scale. Since this scale is arbitrary, nothing which is observable can depend on it. This observation is the basis for the so-called renormalization-group approach, which uses this knowledge and by forming derivatives on renormalization-scale-invariant quantities determines (functional) differential equations, which are useful for determining properties of correlation functions. This will be exploited later in section 5.1.7.

There is a further consequence of this scale. If a theory like Yang-Mills theory is taken, there appears no dimensionful parameter at the classical level, and the theory is classically scale-invariant. However, when the renormalization conditions are imposed, this is no longer the case, since they involve this scale. Since this scale is a manifestation of the ultraviolet divergences, and thus incompleteness of the theory, it is thus created in the quantization process. It is thus said that the classical scale invariance is broken by quantum effects, a process also referred to as dimensional transmutation. In a sense, it is a global anomaly, as the quantization process itself is breaking the classical scale symmetry[8]. This perspective will be further discussed in section 5.6.

[7] Actually, any quantity which is renormalized cannot be measured directly. The only direct measurements possible measure either cross sections, masses, or decay rates in one form or the other, and permit then an indirect determination of the parameters.

[8] As a side remark, it should be noted that the exact masslessness of the photon can be shown to

The second feature is that the mass of the electron and the electric charge now depend on the renormalization, and thus energy scale, by virtue of the renormalization conditions (4.43) and (4.45). Thus, the parameters of the theory become energy-dependent, and out of a set of theories with fixed parameters e and m a single theory with energy-dependent parameters emerges. These energy-dependent quantities are therefore called running. Some more properties of this feature will be discussed in section 4.7. Since, as stated, quantities depending on the renormalization scale are no longer observable, neither the masses nor the charges of the elementary particles are, in fact, observable. They are only given implicitly in a fixed renormalization scheme via renormalization conditions such as (4.41-4.45). Of course, this still permits to plot the energy-dependence of such a quantity. However, the plot is only meaningful after fixing the renormalization scheme.

When changing to QCD, there are more renormalization conditions, due to the different flavors and the ghosts. As a consequence, standardized renormalization schemes have been developed, which are commonly used, and are therefore usually not made explicit. These schemes have been tailored for particular purposes, and must be looked up, if a calculation is to be compared to preexisting results. However, to compare to the commonly used schemes, it is necessary to introduce the concept of dimensional regularization.

4.6.5 Dimensional regularization

The cut-off regularization discussed in section 4.6.1 is by no means the only possibility. There exist quite a plethora of different regularization schemes, which are all consistent. However, almost all of these prescriptions hide symmetries, in particular gauge symmetries. This modifies the STIs and introduces additional counter-terms, making them rather cumbersome in many practical applications. The cut-off regularization is one example of such a regularization prescription which hides gauge symmetry.

However, for the case of perturbation theory, it is possible to find a regularization prescription, which leaves gauge symmetry explicit. This simplifies many calculations tremendously. The price to be paid is that the analytic structure of the appearing correlation functions has to be known, and that the presence of anomalies and chiral symmetries requires very special attention. The prior of these two requirements makes this prescription almost useless beyond perturbation theory. Nonetheless, in the perturbative treatment of QCD, it is almost always employed. Especially the standard renormalization schemes usually explicitly reference it.

The name of this prescription is dimensional regularization. Its name stems from the

be a consequence of this broken scale symmetry in massless QED. In this case the photon becomes the Goldstone boson of the breaking of the global scale symmetry.

fact that an integral is analytically continued away from the number of dimensions in which it should be evaluated to a dimensionality in which it is finite, then evaluated, and finally the result is analytically continued back to the original number of dimensions. In this process, the change of dimensions is entirely formal, and therefore not restricted to an integer number of dimensions. The original divergences then appear as poles of the type $1/\delta$ with δ being the distance to the desired dimensionality. These poles correspond to the explicit appearances of the cutoffs, e. g. in equation (4.38), when a cutoff regularization is performed.

The rules for dimensional regularization can be given mathematically quite precisely. The first part of the prescription is to set any integral to zero, which does not depend explicitly on a scale,

$$\int d^d k (k^2)^\alpha = 0.$$

For integrals involving a scale, take the following example, which is continued to D being different from the target number of dimensions d

$$A = \frac{1}{i\pi^2}\int d^d k \frac{1}{(k^2 - m^2 + i\epsilon)^r} \to A^r = \frac{M^{d-D}}{i\pi^2}\int \frac{d^D k}{(2\pi)^{D-d}} \frac{1}{(k^2 - m^2 + i\epsilon)^r}.$$

The original, unregularized integral is obtained in the limit $D \to d$. Since this is only a regularization, the total value of A^r should not change its energy dimensions, and therefore a dimensional regularization scale M is introduced. This integral is convergent for $D < 2r$. Performing a Wick rotation, i. e., replacing formally $k_0 \to ik_0$, yields

$$A^r = \frac{(2\pi M)^{d-D}}{\pi^2}\int d^D k \frac{(-1)^r}{(k^2 + m^2 - i\epsilon)^r} = \frac{(2\pi M)^{d-D}}{\pi^2}\int k^{D-1} d|k| d\Omega_D \frac{(-1)^r}{(k^2 + m^2)^r},$$

which is for a finite integral always permitted. Using the rotational invariance, the angular integral can be performed yielding the volume of a D-dimensional unit-sphere,

$$\begin{aligned} \int d\Omega_D &= \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} \\ \Gamma(z) &= \int_0^\infty t^{z-1} e^{-t} dt. \end{aligned} \tag{4.46}$$

Of course, a sphere is only a geometric object in the conventional sense for D being integer. The expression (4.46) is therefore taken to define the volume of a sphere in non-integer dimensions.

The remaining integral is then elementary, and can be solved using Cauchy's theorem, to yield

$$A^r = (4\pi M^2)^{\frac{d-D}{2}} \frac{\Gamma\left(r - \frac{D}{2}\right)}{\Gamma(r)} (-1)^r (m^2)^{\frac{D}{2} - r}.$$

So far, this result is valid in $D < 2r$ dimensions. To obtain the originally desired dimensionality, replace $D = d - 2\delta$,

$$A^r = (4\pi M^2)^\delta \frac{\Gamma\left(r - \frac{d}{2} + \delta\right)}{\Gamma(r)} (-1)^r (m^2)^{\frac{d}{2}-r-\delta}.$$

A^r is now expanded for small δ, as the desired limit is $\delta \to 0$. For the case of $r = 1$, i. e., for a massive tadpole like (4.38), the expansion in δ yields

$$A^r = m^2 \left(\frac{1}{\delta} - \gamma + \ln 4\pi - \ln \frac{m^2}{M^2} + 1\right) + \mathcal{O}(\delta),$$

where γ is the Euler constant ≈ 0.577. This expression has a simple pole in δ, replacing the divergence on the explicit cut-off.

From now on, the procedure is essentially identically to the cut-off regularization: The divergent terms are absorbed in counter-terms, and then renormalization is performed. If just the term $1/\delta$ is absorbed the corresponding renormalization scheme is called minimal subtraction (MS), but more commonly the (almost always appearing) combination

$$\frac{1}{\delta} - \gamma + \ln 4\pi$$

is absorbed by the counter-terms. This is the so-called modified minimal subtraction scheme, denoted by $\overline{\text{MS}}$, the standard scheme of most perturbative calculations.

Similarly, it is possible to calculate any kind of other diagram. For example, a massless loop integral in the $\overline{\text{MS}}$-scheme takes the form

$$\int \frac{d^d q}{(2\pi)^d} q^{2\alpha} (q-p)^{2\beta} = \frac{1}{(4\pi)^{\frac{d}{2}}} \frac{\Gamma(-\alpha-\beta-\frac{d}{2})\Gamma(\frac{d}{2}+\alpha)\Gamma(\frac{d}{2}+\beta)}{\Gamma(d+\alpha+\beta)\Gamma(-\alpha)\Gamma(-\beta)} (p^2)^{2\left(\frac{d}{2}+\alpha+\beta\right)},$$

and so on. It is usually possible to reduce given loop integrals by appropriate transformations into one of several master integrals, for which the dimensional regularization results are known, and can be found either in books or some tables in review articles.

4.6.6 Composite operators

When dealing with QCD, one is often interested in observable quantities, and especially in hadrons. However, hadrons are bound states, and thus composite objects. They are hence described by composite operators, which are usually evaluated at the same space(-time) point. E. g. the simplest operator to describe a scalar meson is $\bar{\psi}(x)\psi(x)$. However, such a product is usually ill-defined at the quantum level. As a consequence, such operators

are usually not renormalized just by renormalizing the field operators individually, but require further (multiplicative) renormalization, i. e.

$$(\bar{\psi}(x)\psi(x))_R = Z_\psi Z_c \bar{\psi}(x)\psi(x),$$

where Z_ψ are the wave-function renormalizations of the (anti-)quark field ψ, and Z_c is the additionally required renormalization constant.

But since in a renormalizable theory only a finite number of counter-terms are required, this divergence cannot be an independent one in QCD. It must thus be expressible in terms of the other renormalization constants. In fact, in the present case it turns out to be just given by the inverse of the quark mass renormalization, and hence $m\bar{\psi}(x)\psi(x)$ is actually finite and renormalization-group and renormalization-scheme invariant, and thus observable.

This still leaves the question of how to calculate Z_c. Of course, one possibility, often the simplest one, is to not search explicitly for the relation to the other renormalization constants. Instead, the composite operator is directly calculated and renormalized independently. Since the divergence structure is unique, this will introduce at most an ambiguity in the prefactor, which can be fixed by comparison to an observable quantity. Actually, this may even be the quickest path to determine the relation to the other renormalization constants by reverse engineering the functional dependency on the other renormalization constants based on the divergence structure.

It is, of course, also possible to systematically derive the relations. The starting point is to add a source term for the composite operator to the action, e. g.

$$\mathcal{L}_s = \int d^4x \chi(x)\bar{\psi}(x)\psi(x).$$

Correlation functions of the composite operator are then obtained by derivatives with respect to the new source $\chi(x)$. In this case, the term has the same structure as a mass-term with x-dependent mass $\chi(x)$. Thus, this term contributes to the mass of the quarks, and thus its renormalization is fixed together with the mass renormalization of the quarks. This is how the relation to the mass-renormalization of the quarks comes about. In case of composite operators with more than two fields, however, the situation quickly deteriorates. Since formally the term is non-renormalizable, at least for quarks, this implies that formally all possible counter-terms with the appropriate mass-dimension have to be included. These are fixed in the same way as in any non-renormalizable theory. The only exception is as the sources of the composite operators are send to zero at the end of the calculations, these counter-terms are actually not free, but fixed by the remainder. Nonetheless, calculating them is as complicated as in any non-renormalizable theory, and therefore in practice

this way is rarely done. Especially, since there is no general construction principle for interesting theories which avoids these troubles.

4.7 Running couplings, Landau poles, and asymptotic freedom

An equation like (4.45) defines an energy dependence of the coupling constant, a so-called running coupling constant. Generically, resummed perturbation theory to second order yields an expression like

$$\alpha(q^2) = \frac{g(q^2)^2}{4\pi} = \frac{\alpha(\mu^2)}{1 + \frac{\alpha(\mu^2)}{4\pi}\beta_0 \ln \frac{q^2}{\mu^2}} \equiv \frac{4\pi}{\beta_0 \ln \frac{q^2}{\Lambda^2}} \tag{4.47}$$

for the gauge coupling g.

The equation (4.47) implies that once the coupling is fixed to experiment at μ, and an expression like (4.45) is evaluated at a different momentum q, the right-hand-side is given in terms of $\alpha(\mu^2)$ by this expression (4.47). Besides the explicit value of the renormalization scale and the experimental input at this scale there appears a pure number β_0. This is the so-called first coefficient of the β-function, which is defined by the ordinary differential equation fulfilled by g as

$$\frac{dg}{d\ln\mu} = \beta(g) = -\beta_0 \frac{g^3}{16\pi^2} + \mathcal{O}(g^5),$$

and it can be determined, e. g., by evaluating perturbatively to this order the right-hand-side of (4.45). The values of β_0 depends on the gauge group, as well as the type and representation of the matter fields which couple to the interaction in question. Actually, β_0 could in principle depend on the renormalization scheme, but does not do so in QCD. This actually also applies to the next expansion coefficient of the β function, β_1, but is no longer true for higher orders. It is also only true ins schemes which are mass-independent, i. e. where all renormalization conditions do not involve explicitly any mass.

Before evaluating β_0, the right-hand-side of (4.47) should be noted. There, the various constants have been combined into a single scale Λ, making the dependency of the theory on a single input parameter manifest. This is the so-called scale of the theory, which also sets a typical scale for processes in the theory. E. g., it is about a 50 MeV-1 GeV for QCD, though its precise value depends on the renormalization scheme and the order of the perturbative calculation, and then called Λ_{QCD}. It also makes manifest the dimensional transmutation, as it makes explicit that a dimensionless constant, the gauge coupling, is actually given in terms of a dimensionful quantity, Λ.

Returning to β_0, it can be evaluated to yield in general

$$\beta_0 = \frac{11}{3}C_A - \frac{2}{3}N_f \tag{4.48}$$

where C_A is the adjoint Casimir of the gauge group, and N_f counts the number of quark flavors. Plugging these in for QCD yields 7, if all masses are neglected, i. e., at very high energies.

First of all, since these constants are non-vanishing, the running couplings have divergences at momenta $q^2 = \Lambda^2$. These are artifacts of perturbation theory, and called Landau poles. They indicate that at the latest at momenta $q^2 \approx \Lambda^2$ perturbation theory will fail. Beyond perturbation theory these Landau poles vanish for all theories which can be defined reasonably beyond perturbation theory. For QCD, the effects of this pole become relevant at approximately the scale of hadronic physics, about 1-2 GeV.

Of course, the perturbative expansion makes only sense in the energy domain in which the coupling is small and positive. This provides another surprise. This is the domain above Λ. Hence, the theory becomes weaker interacting at large energies, until the interactions cease altogether at infinite energy. Such a behavior is known as asymptotic freedom, since the theory is non-interacting for asymptotically large energies.

Similar equations like (4.48) actually hold also for all other renormalization-dependent parameters. E. g., the masses of the particles all decrease with the measured momenta. Thus, the masses of particles become less and less relevant the higher the energy.

Chapter 5

QCD beyond perturbation theory

So far, the treatment of QCD was purely perturbative. As noted, this treatment breaks down once the relevant energies become of the order of 1 GeV or less. Also, perturbation theory expands around non-interacting quarks and gluons. Especially, all asymptotic states are only collections of quarks and gluons, and therefore perturbation theory has no stable bound states, though theses exist definitely in nature. Hence, non-perturbative methods are ultimately necessary. However, non-perturbative physics is, though much richer, also much more complicated, especially on a technical level. As a consequence, it is often necessary to make rather strong assumptions, and hence systematic errors are usually at best lower limits to the actual errors. In fact, the term systematic error is rather the statement that some analytically not controllable error exists, which can only be heuristically estimated, based on some general principle. Analytic control over systematic errors would require in almost all cases an analytical understanding of the theory, which is unattainable without an exact solution of the theory, which in turn would obliberate the requirement to assess systematic errors. Hence, systematic errors remain lower limits to errors due to our (lack of) analytic control over a theory. Most often this problem turns up in the form that a functional dependency on some error parameter is known, but the prefactor can only be estimated.

5.1 Methods

There are a multitude of different methods, each with individual strengths and weaknesses, to address QCD beyond perturbation theory. Some of the more popular ones will now be discussed in turn.

5.1.1 Operator-product expansion

The operator-product expansion (OPE) is not really a full non-perturbative method, it is rather perturbation theory augmented with non-perturbative information, which is determined from other methods or experiments.

The basic idea of the OPE is rather straight-forward. Take an expectation value

$$\langle O_1(x)O_2(0)\rangle,$$

with two operators O_1 and O_2. Perturbatively, it is a reasonable assumption that the most dominant part to this expression comes from the contribution $x \approx 0$. Furthermore, the composite operator $O_1O_2(0)$ has certain quantum numbers, and can therefore be rewritten in a suitable basis with the same quantum numbers. Combining this implies the replacement

$$O_1(x)O_2(0) \to \sum_n C_{12}^n(x)O_n(0), \tag{5.1}$$

where the O_n are now a suitable basis of operators with the same quantum numbers, and all dependence on x is in the coefficient functions C. These can be Fourier-transformed, to provide functions of the momentum q. These can be calculated in perturbation theory. To leading order they will be determined by balancing the mass dimensions on both side.

Assume for example that the combination O_1O_2 is gauge-invariant, scalar, and has dimension 6. This dimension would be present, e. g., if both O_1 and O_2 would be scalar meson currents. Then, the leading contributions are[1]

$$c^1(q) \times 1 + \frac{c^2(q)}{(q^2)^2} m_f \bar{q}_f q_f + \frac{c^3(q)}{(q^2)^2} F^a_{\mu\nu} F_a^{\mu\nu} + \mathcal{O}\left(\frac{1}{(q^2)^3}\right).$$

The higher orders have increasingly larger powers of $1/q^2$, and therefore become less and less relevant at high energies, where the OPE works best. In principle, it is possible that quantum effects increase the degree of divergences. In theories, such as QCD, which are asymptotically free, these corrections are at most logarithmic, and therefore do not change the counting of divergences for divergent contributions. Of course, in cases with a finite leading term, this can have substantial importance.

Another thing to worry about is that the momentum q in Minkowski space-time are not necessarily large, if their components are. E. g., they could be on the light-cone. Thus, the argumentation is most relevant for space-like momenta. In position-space, this is more evident. Large momenta correspond to small distances x^2. However, x^2 can also become small for light-like separation. Then, this expansion is rather a light-cone expansion.

[1] The explicit appearance of the quark mass is necessary to obtain the correct renormalization properties.

The coefficient functions c^i, known as Wilson coefficients, are all finite, to leading-order, at large q^2. They can be calculated in perturbation theory, by evaluating the expansion (5.1). These coefficient functions are not universal, as they depend on the operators in questions. The non-perturbative improvement now appears when taking expectation values of (5.1):

$$\langle O_1(x)O_2(0)\rangle \to \sum_n C_{12}^n(x)\langle O_n(0)\rangle. \tag{5.2}$$

The expectation values $\langle O_n(0)\rangle$ are now universal, the same for all pairs of operators on the left-hand side. In the previous examples, they would be 1, the so-called chiral or quark condensate $\langle\bar{q}q\rangle$, and the gluon condensate $\langle F^a_{\mu\nu}F_a^{\mu\nu}\rangle$. However, these quantities are zero to all orders in perturbation theory, at least for vanishing quark masses. Thus, they have to be obtained either by non-perturbative methods, or from experiment. Fitting these with, e. g., experimental data or obtaining them from one of the other methods below hence provides an improvement over the purely perturbative result. Since their values are universal, this will improve, at least for large momenta, the results in many calculations simultaneously. This incidentally also shows that the non-perturbative contributions are suppressed by powers in q^2 at high energies, a consequence of asymptotic freedom. This justifies again the use of perturbation theory at sufficiently large energies, especially as the perturbative corrections in the functions c_i turn out to be at most logarithmic in q.

As an example, the calculation of the Wilson coefficient to leading order yields for the quark propagator the additional contribution

$$S_{ij} \approx -(2\pi)^d \frac{\delta_{ij}\langle\bar{q}q\rangle}{4N_c}\left(1 - \frac{m}{d}\gamma^\mu\partial_\mu\right)\delta(p),$$

and thus a correction to the quark propagator at zero momentum. Though this will be substantially modified with non-perturbative methods in 5.2, this already shows an important qualitative feature: The quark condensate modifies the quark propagator at zero momentum, and thus contributes to an effective mass of the quarks. Going to higher order yields

$$S_{ij} = -i\delta_{ij}\frac{g^2 C_F\langle\bar{q}q\rangle}{12p^4}\left(d - \xi - \frac{2(d-2)}{d}(1-\xi)\frac{m_q\gamma_\mu p^\mu}{p^2}\right).$$

Though this order explicitly depends on the gauge by the appearance of the gauge parameter ξ, this now shows a (gauge-dependent) effective mass for the quarks over all momenta, proportional to the value $\langle\bar{q}q\rangle$. Since this quantity can only be non-zero if chiral symmetry is spontaneously broken, this implies that breaking spontaneously chiral symmetry will influence the quark mass. Whether it is broken, however, is left to non-perturbative methods and, of course, experiment.

In some cases, it is also possible to invert the argumentation. If there is a particle with a large (tree-level) mass, its momentum-space propagator at tree-level can be expanded as

$$\langle\phi^\dagger\phi\rangle = \frac{1}{p^2 - m^2 + i\epsilon} \approx \frac{1}{m^2}\left(1 + \frac{p^2}{m^2} + \mathcal{O}\left(\frac{p^4}{m^4}\right)\right).$$

Hence, to leading order any process in which the particle appears will be suppressed by powers of it mass. Generically, if the process has again a characteristic energy scale Q^2, the effect will be of order Q^2/m^2, which is small if $Q^2 \ll m^2$. This condition is, e. g., very well fulfilled in QCD for the top quark at typical hadronic energies. Hence, the top quark can almost throughout QCD be neglected with an error of order $1/175^2$, tiny in comparison to other error sources.

This more heuristic argument can be formalized, and is known as the Applequist-Carrazone theorem. It is at the heart of heavy-quark effective theory (HQET). This is an approximation where heavy quarks, e. g. the bottom and even the charm quark, are treated as so heavy that they can essentially be treated like non-relativistic particles, or even inert particles. This makes especially bound states of two heavy quarks accessible to an approximation with quantum mechanics, where the gluon interactions between quasi-static quarks is essentially given by the pure Yang-Mills interaction. Even in cases of heavy-light systems such an approximation is often rather good.

There are various ways for deriving heavy-quark effective theory. One possibility is to integrate out the heavy quark fields, like in lattice calculations to be discussed in section 5.1.5. Expanding afterwards the determinant of the Dirac operator in $1/m$ yields then an operator which can be reintegrated using new effective fields. In general, the typical structure of the effective theory is of type

$$\mathcal{L} = \bar{h}\left(iD_0 + \frac{1}{2M}(\gamma_i D_i)^2 + i\frac{1}{4M^2}\mathcal{O}(\gamma_i D_i \gamma_j D_j D_0)\right)h$$

where h is the heavy quark field, the summation is over the spatial part, and the precise form of the higher order operator $\mathcal{O}$ depends on the type of expansion, like in speed, mass or others, performed, while the first part is almost universal. The second part is also always vanishing on-shell, and therefore describes different consequences of quantum corrections in the different expansion schemes at finite orders. It is evident then that the expansion becomes non-relativistic, as now the time-derivative remains first order, while the spatial derivatives become second order. Of course, this theory is non-renormalizable, and as such will require an increasing amount of experimental input in each order of perturbation theory.

It should be noted that there are two limitations to the Appelquist-Carrazone theorem. One is that certain assumption on long-range effects are made, which strictly speaking are

not fulfilled by QCD. They have to do with topological aspects to be introduced later in section 5.4. Still, though this is formally a problem, in practical applications the induced error is irrelevant.

A more severe practical problem appears when it comes to renormalization. A mass-independent renormalization scheme, e. g. $\overline{\text{MS}}$, is defined in a way which includes automatically all particles in a theory, no matter the mass. Thus, renormalizing a theory in such a mass-independent scheme and then performing a heavy-mass expansion, or just neglecting heavy particles, necessarily fails. The most often implemented solution is to perform the heavy-mass approximation before renormalization. This, however, limits the energy range for which a calculation is possible, and will lead to non-analyticities at the thresholds of the heavy particles. The alternative is to use an explicitly mass-dependent renormalization scheme, where then a consistent approximation can be done. Though this is conceptually the cleanest approach, it is technically substantially more involved, and therefore often disfavored.

5.1.2 Sum rules

Sum rules are, in principle, rather simple statements, which are nonetheless exactly true, and they derive usually from rather simple statements. A simple example of a sum rule is, e. g.,

$$\int d^4x \left(\frac{2}{3}2u_p(x) - \frac{1}{3}d_p(x)\right) = 1,$$

where u_p and d_p are the density of up quarks and down quarks inside a proton, and the factor 2 takes into account that there are twice as many up quarks than down quarks. Hence, this sum rule is just the statement that the total charge of the proton, in units of the electron charge, is 1.

The real usefulness of such statements comes into play by combining several of them. E. g., another sum rule is

$$\int d^4x \left(\frac{2}{3}u_n(x) - \frac{1}{3}2d_n(x)\right) = 0,$$

stating that the neutron is electrically neutral. Subtracting both yields

$$\int d^4x \left(\frac{2}{3}(2u_p - u_n) - \frac{1}{3}(d_p - 2d_n)\right) = 1.$$

This is a non-trivial statement about the difference of the quark distributions inside the proton and neutron. Assuming that the difference, only due to isospin-breaking effects, is

small, and hence the distributions in the proton and neutron are similar yields

$$\int d^4x\,(2u+d) = 3.$$

This shows that there are three quarks inside the nucleon, or that the baryon number of quarks is 1/3. Neglecting the small difference between up and down quarks finally yields

$$\int d^4x u = 1,$$

which is nothing but the statement that the distribution is normalized.

Though this example was quite simple, it generalizes to more complicated statements. This is particular true if, e. g. by electromagnetic measurements, the distributions are measured and then the results are used in different sum rules. A famous example is the following. Assume that u and d are known. Then, if the naive quark picture would be correct, it follows that

$$\int dp_0 \int d^4x e^{ip_0 t}(2u(x)+d(x)) = E$$

should give the total energy of a nucleon. However, it was found that the sum-rule was violated by more than 50%, which is direct evidence that the gluons and sea quarks carry a substantial amount of the total energy of the nucleon. Similar results have also been found, e. g., for the nucleon spin, the so-called spin crises. Though, of course, it is not a real crises. There is no a-priori reason why the gluons and sea quarks should not carry a substantial amount of the spin. However, it was yet not possible to fully satisfactorily resolve how much they actually carry separately, both experimentally and theoretically.

5.1.3 Scattering theory and Roy equations

5.1.4 Dispersion relations

Dispersion relations are relations which are established on basis of the analyticity of two-point functions, i. e. functions which depend only on a single external momentum variable, or two space-time positions, e. g. $\langle O_1(x)O_2(y)\rangle = \Pi(q)$. The operators O_i can be composite, but the important restriction is that they are physical, especially gauge-invariant. An example could be two hadron currents, and then this would be the interaction of two hadrons with an exchange of a momentum q.

For such gauge-invariant operators it can be shown that the only possible analytic structure is that they have one or more poles at some $q^2 \geq 0$ on the real axis, and a branch cut on the real axis starting at some $q^2 \geq 0$ which corresponds to the lightest

two-particle state in the same channel. There may also be possibly more poles on further Riemann sheets which do not play a role in the following. Consider then the integral

$$I_n = -4\pi\alpha \int \frac{dq^2}{2\pi i} \frac{1}{(q^2+Q_0^2)^{n+1}} \Pi(q^2) \tag{5.3}$$

where $Q_0^2 < 0$ and the integration paths encircles this pole. The theorem on residues then yields

$$I_n = \frac{1}{n!} \frac{d^n}{d(q^2)^n} \Pi(q^2)\big|_{q^2=-Q_0^2} .$$

Now expand in (5.3) Π using the OPE at Q_0^2, requiring $-Q_0^2$ to be sufficiently large. Then none of the terms in the expansion can grow faster than logarithmically. Hence, it is permissible to push the integration contour to infinity, except where the branch point is. A further theorem of complex analysis then ensures that the integral is entirely determined by the discontinuity across the cut, which is given by its imaginary part. This yields

$$\begin{aligned} I_n &= -4\pi\alpha \int \frac{dq^2}{2\pi i} \frac{1}{(q^2+Q_0^2)^n} \text{discontinuity}(\Pi(q^2)) \\ &= -8\pi\alpha \int \frac{dq^2}{2\pi} \frac{1}{(q^2+Q_0^2)^{n+1}} \Im\Pi(q^2). \end{aligned}$$

By virtue of the optical theorem, however, this is just an integral over the corresponding interaction cross-section σ

$$I_n = \frac{1}{\pi} \int_0^\infty ds \frac{s}{(s+Q_0^2)^{n+1}} \sigma(s).$$

Hence, this relates the coefficient of the OPE, non-perturbative information, directly to cross-sections, making them directly accessible for other purposes. Such a relation is called a dispersion relation. One of the arguably most famous ones is obtained from the cross-section of $e^+e^- \to$hadrons, which to leading order yields

$$\int_0^\infty ds \frac{s}{(s+Q_0^2)^{n+1}} \sigma_{e^+e^-\to\text{hadrons}}(s) = \frac{4\pi\alpha^2}{n(Q_0^2)^n} \sum_f \left(Q_f^2 + \mathcal{O}(\alpha^3, \alpha_s, (Q_0^2)^{-2})\right) \tag{5.4}$$

which counts the active quark species. Hence dispersion relations are valuable tools to obtain non-perturbative information from cross-sections.

5.1.5 Lattice

So far, the methods employed were, in a sense, at least semi-perturbative or rather indirect. Of much more interest are methods which permit to directly compute, at least in principle,

in an exact way arbitrary n-point functions. Several such methods exist, and some will be discussed in the following. All of them provide a means to obtain an exact result.

However, in practice, for any reasonably interesting theories, all of these methods face practical limitations, which take various forms. This requires in all cases to make approximations which induce systematic errors. It is in the nature of these errors that they cannot be determined accurately, as this would require to have a solution to the theory in question. However, all methods provide possibilities to give lower bounds on these errors in a systematic fashion, which, in principle but not in practice, can be arbitrarily refined, but will remain lower bounds.

One further noteworthy point is that though all of these methods can be formulated exactly in Minkowski space-time, it is in most practical applications, just as in perturbation theory, useful to go to Euclidean space-time in intermediate steps, or even obtain the final result in Euclidean space-time. Though this may seem to be a strong limitation, it is not. It can be shown that all knowledge about a quantum field theory like QCD in Minkowski space-time, can be obtained from the same theory in Euclidean space-time, by means of so-called reconstruction theorems. Though an exact reconstruction requires complete knowledge of the theory, it is still possible to obtain some subsets of results, e. g. the masses of particles, even under approximations, exactly, and the other one to an extent anyhow limited by the applied approximations. Hence, this is in many, but not in all cases, not a too strong limitation.

The first of these methods is the so-called lattice approach. As it will become apparent, all of these methods will require to violate in one way or the other one or more symmetries of the theory forcible due to the introduced approximations. For the lattice, this is rotational and translation symmetries.

The idea behind lattice gauge theory is to replace continuous space-time by a grid of points, usually a hypercubic[2] grid of points x. This reduces rotation symmetry from the Lorentz group to the hypercubic group, and the translation symmetry to the subgroup of lattice translations. In practical calculations, which are mostly done numerically, also the size of the lattice is made finite. This requires boundary conditions, which in the infinite-volume limit are, of course, irrelevant. They are hence chosen usually periodic or anti-periodic, for simplicity.

The points are then connected by directed vectors in μ direction between two lattice points, and which has length a, the lattice constant. The original continuum theory should be recovered in the limit $a \to 0$, which is a non-trivial problem.

[2]There have also been attempts using different geometries, or even random locations, but none of theses turned out to be superior to the simplest case of hypercubic lattices.

5.1.5.1 Yang-Mills theory

To implement gauge fields logically on the lattice, the naive idea would be to just discretize the gluon field $A_\mu = g\tau^a A^a_\mu$. However, this turns out to be an impractical choice in calculations. Rather, so-called links are used,

$$U_\mu(x) = e^{iaA_\mu(x)} \approx 1 + iaA_\mu(x) \tag{5.5}$$

which therefore permit to recover the original gauge field in the continuum limit. The name stems from the fact that the links have a direction, and therefore link the points x and $x + ae_\mu$, where e_μ is the unit vector in direction μ. Geometrically, it can be shown that this is just the parallel transport which is required to construct a covariant transport over a finite distance, in this case a. In the infinitesimal limit this parallel transport is given by the covariant derivative, which replaces the ordinary derivative giving the parallel transport without local gauge invariance.

Exponentiating the algebra element A_μ yields a group element, and hence $U^{-1} = U^\dagger$. This implies that the now also group-valued gauge transformations $g = \exp(i\omega)$ take the form

$$U_\mu(x) \to g(x)U(x)_\mu g^\dagger(x),$$

which can be shown by using the expansion (5.5), and noting that $(A_\mu(x) - A_\mu(x + ae_\mu))/a$ is just the discrete derivative.

Since for the su(3) algebra there are two possible groups, there is a choice involved. In QCD, this choice is arbitrary, but in the standard model the choice is unique by the requirement that all fermion fields must have single-valued wave-functions, and is $\mathrm{SU}(3)/\mathrm{Z}_3$. However, absence of the global factor $\mathrm{Z}_3 = \{1, e^{\frac{2\pi i}{3}}, e^{\frac{4\pi i}{3}}\}$ makes the theory very hard to treat numerically. Since the theory recovered in the continuum limit is again the algebra theory, however, it should not matter which group is used. Therefore, in practice rather the group SU(3) is used.

The action is constructed using the so-called plaquette

$$U_{\mu\nu}(x) = U_\mu(x)U_\nu(x + ae_\mu)U^\dagger_\mu(x + ae_\nu)U^\dagger_\nu(x)$$

yielding

$$S = -a^4\beta \sum_x \sum_{\nu<\mu} \left(1 - \frac{1}{3}\Re\mathrm{tr}U_{\mu\nu}(x)\right) \tag{5.6}$$

with $\beta = 2N_c/g^2 = 6/g^2$ related to the coupling constant. This lattice action is called the Wilson action. That (5.6) yields indeed the continuum action in the limit $a \to 0$ can be shown by using the expansion (5.5), and noting that the corrections to it are of order $\mathcal{O}(a^2)$, and therefore irrelevant in the continuum limit.

Taking the continuum limit also implies that any additions to the action (5.6), which vanish in this limit $a \to 0$ can be freely included in the action. This freedom has been widely used to design so-called improved actions, which have various desired properties, e. g. reducing the dependence on the lattice spacing. However, in general adding some terms to the lattice action will not always act in the same way for all observables, and hence an improvement in one quantity can very well be a degradation for another one, and hence this requires careful checks.

One of the particular advantages of this reformulation is that the group elements, in contrast to the algebra elements, are bounded, in the sense that $\mathrm{tr}U_\mu$ is finite for all group elements. This implies that the lattice path integral

$$Z = \int \mathcal{D}U(x) e^S$$

is actually finite, even when integrating over the whole group. Hence, gauge-fixing is not necessary to obtain finite results, as long as the regulator a is not removed. Furthermore, since a finite a also implies that there is a maximum momentum on the lattice of order $1/a$ there is also an intrinsic momentum cutoff, and hence the theory is (gauge-invariantly) regulated. Thus, the theory is completely finite. This makes the lattice formulation particularly accessible to numerical calculations, which has been widely exploited. Especially, any observable

$$\langle O \rangle = Z = \int \mathcal{D}U(x) O e^{-S}$$

can now be calculated by brute force numerically, by just explicitly evaluating the finite number of integrals explicitly[3].

This is particular useful to obtain masses. Since in a finite volume all energy levels are discrete, it is always possible to expand in energy eigenstates[4],

$$\langle O(t)O(0) \rangle = \sum a_n e^{-E_n t},$$

where the reality of the exponents is due to the Euclidean space-time. Hence, just by waiting long enough, only the ground-state will contribute appreciably, and can thus be read-off. Determining the higher contributions is possible, though even with exact results no more energy levels can be determined than a quarter of the points in time direction. Also here, several sophisticated methods beyond the scope of this lecture are available to extract more information.

[3]In practice, the integral is still high-dimensional, at least $V \times d \times (N_c^2 - 1)$, and sophisticated numerical algorithms are required.

[4]For periodic boundary conditions in a finite volume the exponentials have to be replaced by cosh functions.

5.1.5.2 Quarks

The formulation of fermions, and thus especially quarks, on the lattice is at the same time straightforward and cumbersome. Cumbersome it is because most applications of lattice calculations aim at a numerical evaluation. However, there is no numerical algorithm for Grassmann numbers. Hence, it is necessary to somehow get rid of them. The basic stratagem to do so can already be applied in the continuum theory. Since the fermions only appear as a bilinear in QCD, it is just a Gaussian integral, and hence the integral can be performed using (3.10). Then, only the determinant of the Dirac operator $D_\mu\gamma_\mu$ appears

$$\int \mathcal{D}A\bar{\psi}\psi e^{-S_{\rm YM}-\int\bar{\psi}D_\mu\gamma_\mu\psi} = \int \mathcal{D}A e^{-S_{\rm YM}}\det(D_\mu\gamma_\mu). \tag{5.7}$$

On a finite lattice also the eigenspectrum of the Dirac operator is discrete, and can be calculated using suitable numerical techniques.

However, there is a serious problem encountered when adding fermions. In the free case, the simplest discretization of the Dirac operator in Fourier space is

$$D_\mu\gamma_\mu = i\sum_\mu \gamma^\mu \sin(p_\mu a) + m. \tag{5.8}$$

This can be treated exactly, and this yields the exact fermion propagator

$$D^{\bar{\psi}\psi}(p) = \frac{am - i\sum_\mu \gamma_\mu \sin p_\mu a}{a^2m^2 + \sum_\mu \sin^2 p_\mu a}.$$

This propagator has the problem that not only for $p_\mu a \approx 0$ the sin can be approximated by $p_\mu a$, but also for the largest momenta on the lattice, $p_\mu a \approx \pi$. Hence, there are indeed 16 rather than one pole of this propagator, and hence there are far too many degrees of freedom in this naive discretization. This is the so-called fermion doubling problem.

It would at first sight appear possible that this is just due to a too simplistic discretization (5.8) of the Dirac operator. Unfortunately, this is not the case. The reason is the breaking of the rotation group by the lattice. Fermions do have an additional symmetry related to space-time symmetry, chiral symmetry. This symmetry is necessarily broken by the lattice in any naive discretization. As a consequence, it can be shown, this is the so-called Nielsen-Niomiya theorem, that it is not possible to evade the doubling problem without destroying chiral symmetry.

Of course, in the continuum limit chiral symmetry is restored. Hence, the solution to this problem is to give up a Dirac operator with explicit chiral symmetry at finite lattice spacing. This so-called Ginsparg-Wilson construction leads to a Dirac operator

without doubling, but also without chiral symmetry. However, it furnishes a replacement symmetry, which rather than fulfilling the ordinary chiral algebra

$$\{\gamma_5, D\} = 0$$

instead fulfills

$$\{\gamma_5, D\} = aD\gamma_5 D, \tag{5.9}$$

where the transition to the continuum symmetry is made explicit by the appearance of the lattice spacing a, and D is the Dirac operator. Explicit solutions to (5.9) have been constructed. The only drawback remaining is then that this introduces additional discretization errors, as the chiral symmetry is distorted at finite a, and all Dirac operators constructed in this way turn out to be numerically more expensive than the naive version by one to two orders of magnitude. As a consequence, several intermediate solutions have been constructed, which have less conceptual problems, are cheaper, but have larger lattice artifacts.

An interesting result obtained from such calculations is that ignoring the dynamics of fermions completely, i. e. setting the determinant of the Dirac operator to one in (5.7), the so-called quenched approximation, is a rather good approximation: Many results agree with the ones from full QCD at the few-percent level. This implies that QCD is essentially dominated by the gluon dynamics. Of course, this cannot be true in general. The leading order of the β function (4.48) changes qualitatively above a certain number of flavors, roughly 16, and asymptotic freedom is lost. In this case, also the quenched approximation can no longer be expected to give good results. In fact, it is expected that already much earlier, at about 10-12 flavors, qualitative changes set in, which invalidate the quenched approximation. However, given that only six quark flavors are known so far, this is not a real practical limitation. Also, since the quarks are massive, and especially for larger generations very massive, this implies that only the few lightest quarks will play an important role for most of hadronic physics.

5.1.6 Dyson-Schwinger equations

The Dyson-Schwinger equations (DSEs) are essentially the quantum version of the equations of motion. As such, they are a coupled set of equations for all the correlation functions of a given theory. Their solution therefore determines a theory completely[5]. Though it is

[5]In fact, the DSEs may have multiple solutions, as well as the FRGs to be introduced in the next section. Similar to ordinary differential equations boundary conditions may then be necessary to identify the appropriate solution. These may be either physical ones, e. g. at first order phase transition where two phases coexist, or unphysical ones, e. g. due to gauge conditions.

possible to formulate DSEs also using a lattice regularization, the focus will here be on the continuum version. As a consequence, DSEs face the same problem concerning gauge symmetry as does perturbation theory, in fact even more so as will be discussed later in section 5.5.5.2. Hence, it is necessary to fix a gauge. For technical convenience, in the following the Landau gauge will be chosen.

The most straightforward way to derive the DSEs for a generic field ϕ^a is by using the fact that the integral of a total derivative vanishes

$$\begin{aligned} 0 &= \int \mathcal{D}A^a \frac{\delta}{\delta A^a(y)} e^{-S+\int d^dx A^a(x) j^a(x)} \\ S &= \int d^dx \mathcal{L}. \end{aligned}$$

Here, S is the action, j^a is the source[6] of A^a and the integral is over full field space. Performing the derivative and pulling the resultant factor out of the integral by replacing A^a with $\delta/\delta j^a$, the equation

$$\left(-\frac{\delta S}{\delta A^a(x)}\bigg|_{A^a(x)=\frac{\delta}{\delta j^a(x)}} + j^a(x) \right) Z[j^a]\bigg|_{j^a=0} = 0 \tag{5.10}$$

is obtained. Performing further derivatives will create a sequence of equations.

To establish the meaning of (5.10), it is necessary to recast it further. The first step is to introduce the free energy W

$$W = \ln Z,$$

which changes (5.10) to

$$-\frac{\delta S}{\delta A_a}\left[\frac{\delta W}{\delta j} + \frac{\delta}{\delta j}\right] + j_a = 0.$$

Performing a Legendre transformation to obtain the effective action Γ

$$\begin{aligned} \Gamma[A] &= -W + \int d^dx \frac{\delta W}{\delta j_a} j_a \\ W[j^a] &= -\Gamma[\phi^a] + \int d^dx j^a(x) A^a(x) \end{aligned}$$

implies

$$\begin{aligned} A^a &= \frac{\delta W}{\delta j^a} \\ j^a &= \frac{\delta \Gamma}{\delta A^a}. \end{aligned}$$

[6]There are some subtleties when it comes to gauge-dependent sources. They will not play a role here, but one should be very careful in employing them in general, especially beyond perturbation theory.

Hence, the sources and the fields are related dependent and independent variables. Thus, the Legendre transform Γ of W is the generating functional of vertex functions. Especially, since the derivative of the effective action is just the connected and amputated one-point function, this shows explicitly that (5.10) is an equation for the one-point function. It also shows that further derivatives will yield equations for higher n-point functions. In particular

$$\begin{aligned}\Gamma^{A_{a_1}\ldots A_{a_n}}(x_1,\ldots,x_n) &= \frac{\delta^n\Gamma}{\delta A_{a_1}(x_1)\ldots\delta A_{a_n}(x_n)} \qquad (5.11)\\ \Gamma^{A_aA_b}(x,y) &= D^{A_aA_b-1}(x-y) = \frac{\delta^2\Gamma}{\delta A_a(x)\delta A_b(y)} = \left(\frac{\delta^2 W}{\delta j_a(x)\delta j_b(y)}\right)^{-1}.\end{aligned}$$

The order of the field indices A_i is relevant not only because of assignment of the arguments, but also if anti-commuting fields appear.

This yields finally

$$\frac{\delta\Gamma}{\delta A_a} - \frac{\delta S}{\delta A_a}\left[A_a + \frac{\delta^2 W}{\delta j_a \delta j_b}\frac{\delta}{\delta j_b}\right] = 0$$

as an equation to determine the DSEs for the vertex functions.

In the case of Grassmann fields u like ghosts and fermions, two independent sources are necessary. This modifies the above to

$$Z = \int \mathcal{D}u^a\bar{u}^a e^{-S[u^a,\bar{u}^a]+\int d^dx(\bar{\eta}^a(x)u^a(x)+\bar{u}^a(x)\eta^a(x))} \qquad (5.12)$$

$$u^a(x) = \frac{\delta W}{\delta\bar{\eta}^a(x)} \qquad \bar{u}^a(x) = -\frac{\delta W}{\delta\eta^a(x)} \qquad (5.13)$$

$$W(\eta^a,\bar{\eta}^a) = -\Gamma(u^a,\bar{u}^a) + \int d^dx(\bar{\eta}^a(x)u^a(x) + \bar{u}^a(x)\eta^a(x)) \qquad (5.14)$$

$$\eta^a(x) = \frac{\delta\Gamma}{\delta\bar{u}^a(x)} \qquad \bar{\eta}^a(x) = -\frac{\delta\Gamma}{\delta u^a(x)} \qquad (5.15)$$

where all derivatives with respect to Grassmann variables act in the direction of ordinary derivatives.

It is worthwhile to derive an explicit example, which will be the DSE for the ghost in Landau gauge. It is one of the simplest equations, but shows already all relevant mechanisms. The general procedure to obtain the corresponding Dyson-Schwinger equations is to calculate equation (5.10) and then derive once more with respect to the field or with respect to the anti-field in case of anti-commuting fields. The additional source term then yields the propagator while the right-hand-side of the equations are found by the derivative of the action.

Since in the course of the derivation, the source in equation (5.10) becomes the inverse full propagator, it makes sense to already rewrite (5.10) as[7]

$$j^a(x)Z = \frac{\delta S}{\delta \phi^a(x)}\Big|_{\phi^a(x)=\frac{\delta}{\delta j^a(x)}} Z \tag{5.16}$$

at the sources set equal to 0. The ghost sector of Yang-Mills theory is given by

$$S_{gh} = \int d^d z \bar{c}^c(z)\partial_\rho(\delta^{cd}\partial_\mu + g f^{cde} A^e_\rho(z))c^d(z) \tag{5.17}$$

Entering this into (5.10) and differentiating with respect to $\bar{c}^a(x)$ yields

$$\left(-\partial^{2x} c^a(x) - g f^{ade}\partial^x_\mu A^e_\mu(x) c^d(x) + \frac{\delta\Gamma}{\delta \bar{c}^a(x)}\right) e^W = 0$$

where the x-index on a ∂ indicates the variable with respect to which to derive. Replacing the fields by their respective derivatives and divide, after performing the derivation, by $\exp(W)$ yields

$$-\partial^{2x} c^a(x) - g f^{ade}\partial^x_\mu \left(\frac{\delta W}{\delta j^e_\mu(x)}\frac{\delta W}{\delta \bar{\eta}^d(x)} + \frac{\delta^2 W}{\delta j^e_\mu(x)\delta\bar{\eta}^d(x)}\right) + \frac{\delta\Gamma}{\delta\bar{c}^a(x)} = 0$$

As a general feature of such derivations, terms containing products of only single derivatives of W appear at this stage. Here, the aim is only an equation for the propagator. Hence, when deriving such terms again with respect to the fields, always at least one single derivative remains, which can be replaced by a classical field. When setting the classical sources to zero at the end, also the classical fields are set to 0 and therefore these terms always vanish. Hence they can be neglected already at this stage of the calculation, and will not appear furthermore. This is not true when deriving equations for higher n-point functions. E. g., when determining the equation for a four-point function, each single derivative may be acted upon, yielding a non-vanishing term.

So the remaining expression is

$$-\partial^{2x} c^a(x) - g f^{ade}\partial^x_\mu \frac{\delta^2 W}{\delta j^e_\mu(x)\delta\bar{\eta}^d(x)} + \frac{\delta\Gamma}{\delta\bar{c}^a(x)} = 0$$

To obtain the equation for the ghost propagator, this equation is derived once more with respect to $c^b(y)$ which leads to

$$-\partial^{2x}\delta^{ab}\delta(x-y) - g f^{ade}\partial^x_\mu \frac{\delta^3 W}{\delta c^b(y)\delta j^e_\mu(x)\delta\bar{\eta}^d(x)} + \frac{\delta^2\Gamma}{\delta c^b(y)\delta\bar{c}^a(x)} = 0$$

[7]It should be noted that the equations may differ in form depending on the order of derivatives, though of course this only corresponds to rearrangements due to identities relating different vertex functions.

The last term defines the inverse ghost propagator as

$$\frac{\delta^2\Gamma}{\delta c^b(y)\delta\bar{c}^a(x)} = D_G^{ab-1}(x-y)$$

The propagator is then given by

$$\frac{\delta^2 W}{\delta\eta^b(y)\delta\bar{\eta}^a(x)} = D_G^{ab}(x-y),$$

which can be proven as

$$\int d^dz \frac{\delta^2 W}{\delta\eta^c(z)\delta\bar{\eta}^a(x)}\frac{\delta^2\Gamma}{\delta c^b(y)\delta\bar{c}^c(z)} = \int d^dz \frac{\delta c^a(x)}{\delta\eta^c(z)}\frac{\delta\eta^c(z)}{\delta c^b(y)} = \frac{\delta c^a(x)}{\delta c^b(y)} = \delta^{ab}\delta(x-y).$$

The third term yields the interaction part. Using

$$\frac{\delta^2 W}{\delta j_\mu^e(x)\delta\bar{\eta}^d(x)} = -\int d^dz d^dw \frac{\delta^2 W}{\delta j_\nu^f(z)\delta j_\mu^e(x)}\frac{\delta^2\Gamma}{\delta\bar{c}^g(w)\delta A_\nu^f(z)}\frac{\delta^2 W}{\delta\eta^g(w)\delta\bar{\eta}^d(x)}$$

where the minus arises due to the anti-commuting derivatives. Using further the fact, that

$$\frac{\delta^2\Gamma}{\delta\bar{c}^g(w)\delta A_\nu^f(z)}|_{j=\eta=\bar{\eta}=0} = 0,$$

which generalizes to all mixed two-point functions, it is possible to write down the result in position space

$$\begin{aligned} D_G^{ab-1}(x-y) &= \partial^{2x}\delta(x-y) \\ &+ gf^{ade}\partial_\mu^x \int d^dz d^dw D_{\mu\nu}^{ef}(x-z) D_G^{dg}(x-w)\Gamma_\nu^{c\bar{c}A;bgf}(y,w,z). \end{aligned}$$

Herein the gluon propagator $D_{\mu\nu}$ is defined as

$$\begin{aligned} D_{\mu\nu}^{ab}(x-y)^{-1} &= \frac{\delta^2\Gamma}{\delta A_\nu^b(y)\delta A_\mu^a(y)} \\ D_{\mu\nu}^{ab}(x-y) &= \frac{\delta^2 W}{\delta j_\nu^b(y) j_\mu^a(x)}. \end{aligned} \tag{5.18}$$

and the full ghost-gluon vertex

$$\frac{\delta^3\Gamma}{\delta c^a(x)\delta\bar{c}^b(y)\delta A_\mu^c(z)} = \Gamma_\mu^{c\bar{c}A;abc}(x,y,z).$$

Replacing all expressions with their Fourier-transformed[8] and afterwards dropping $\int d^dp/(2\pi)^d \exp(-ip(x-y))$, produces the result in momentum space as

$$D_G^{ab-1}(p) = -\delta^{ab}p^2 - igf^{ade}\int \frac{d^dq}{(2\pi)^d} p_\mu D_{\mu\nu}^{ef}(p-q) D_G^{dg}(q)\Gamma_\nu^{c\bar{c}A;bgf}(-p,q,p-q), \tag{5.19}$$

[8] All momenta are always defined incoming and momentum conservation at the vertices is taken into account. Hence in principle one of the arguments of the vertices could be dropped, but since this depends on conventions, all are kept.

where momentum conservation at the vertex has been used.

From (5.19) it is also possible to directly read of the tree-level vertex

$$\Gamma_{\mu}^{tl;c\bar{c}A;abc}(p,q,k) = igf^{abc}q_{\mu}\delta(p+q+k) \tag{5.20}$$

Rewriting (5.19) using (5.20) gives

$$\begin{aligned} D_G^{ab-1}(p) &= -\delta^{ab}p^2 \\ &+ \int \tfrac{d^dq}{(2\pi)^d}\Gamma_{\mu}^{tl;c\bar{c}A;dae}(-q,p,q-p)D_{\mu\nu}^{ef}(p-q)D_G^{dg}(q)\Gamma_{\nu}^{c\bar{c}A;bgf}(-p,q,p-q). \end{aligned} \tag{5.21}$$

which is the final form.

In a similar way, all DSEs can be derived. As is visible, this is algorithmic, and can therefore be automatized. Three important statements have to be made.

The first is that the equations couple different correlation functions. In the present case, the equation for the ghost propagator is coupled to the gluon propagator and the ghost-gluon vertex, and thus an n-point vertex of higher order. Generically for a theory which in four dimensions is renormalizable by superficial power-counting the equations for an n-point function involves the $n+1$ and $n+2$ point functions. Furthermore, in general in such theories also expressions with two integrals in momentum space appear.

The DSEs therefore are a coupled system of non-linear integral equations. Solving them exactly is therefore in all but the most trivial cases impossible. Expanding all correlation functions in the coupling is another way to generate ordinary perturbation theory, but will thus not provide any further insights. An alternative is to approximate the higher n-point functions in such a way that a finite number of unknowns and a finite number of equations remain, which is called a truncation. Such a system can then be solved using appropriate techniques, though this is still in general by no means trivial. Furthermore, it is not guaranteed that a given such truncation will produce equations which have a solution at all, or it can have multiple solutions of which some or all can be spurious solutions which disappear outside the truncation. Therefore careful checks, most desirably by comparison to experiments, should be made.

Second, the DSEs involve all information, in contrast to the STIs. Thus, an exact solution of the DSEs will automatically also satisfy all STIs[9]. This is, of course, not true for some truncation, including the perturbative expansion. They will in some way violate the STIs. In perturbation theory, this happens at the next order of the coupling constant, and is therefore self-consistent. This is in general not true for any other truncation. The

[9]Though in general there are quantities only appearing in the STIs, but not the DSEs, which are then pure gauge artifacts. They are determined uniquely by the solutions of the DSEs, though actually calculating them may still be technically non-trivial.

converse is actually not true: The STIs only make a statement about relations due to the symmetries of the theory, and therefore only encode redundant information. Thus, knowledge of the solutions of the STIs is not sufficient to solve the DSEs.

Third, though the equation derived is for a gauge-dependent quantity, equations can also be derived for gauge-invariant quantities. Equations describing mesons and baryons are special kinds of the DSEs, and are known as Bethe-Salpeter (BSEs) and Faddeev equations, respectively.

5.1.7 Renormalization group equations

As discussed in section 4.6, in the renormalization process a scale μ is introduced. However, the choice of this scale was arbitrary, at least as long as $\mu > 0$. As a consequence, any physical observables should not depend on this scale. Especially, it must be permissible to shift this scale $\mu \to \mu + \delta\mu$ without affecting observables. Of course, since this scale was used to define the renormalized parameters, this will induce a shift in all the parameters p of the theory $p \to p + \delta p$, in QCD the gauge coupling g and the masses of the quarks m. This will also induce a shift in the field amplitudes δZ, as these are also renormalized.

Consider a single field. The latter implies that a connected correlation function

$$\Gamma^n = \langle \phi_1 ... \phi_n \rangle$$

will be shifted under an infinitesimal transformation in μ to first order as

$$\Gamma^n \to (1 + n\delta Z)\Gamma^n \tag{5.22}$$

A dependence on different fields will just induce a sum of the different shifts. However, simultaneously, the correlation function is a function of the parameters of the theory, and of the renormalization scale. E. g., the renormalized propagators depend only on the renormalized parameters g and m, and the renormalization scale μ. Also any wave-function renormalization is entirely given in terms of these quantities. Thus, (5.22) must also be given by

$$d\Gamma^n = n\delta Z\Gamma^n = \frac{\partial\Gamma^n}{\partial\mu}\delta\mu + \frac{\partial\Gamma^n}{\partial g}\delta g + \frac{\partial\Gamma^n}{\partial m}\delta m.$$

To remove the explicit dependence on $\delta\mu$, usually the redefinitions

$$\begin{aligned} \beta &= \mu\frac{\delta g}{\delta\mu} \\ \beta_m &= \frac{\mu}{m}\frac{\delta m}{\delta\mu} \\ \gamma &= \mu\frac{\delta Z}{\delta\mu} \end{aligned} \tag{5.23}$$

are performed. Since the particular correlation function no longer appears, these are universal functions of the coupling constant, masses, and the renormalization scale only. However, they are not unique. Since the dependence on the renormalization scale implies that they depend on the way how it is fixed, and therefore on the renormalization scheme.

This form is known as the/a Callan-Symanzik equation

$$\left(\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}+\beta_m m\frac{\partial}{\partial m}-n\gamma\right)\Gamma^n=0, \tag{5.24}$$

which is an exact equation the correlation function has to fulfill. Especially, if the functions β, β_m, and γ would be known, then it would be possible to determine its running with the renormalization scale μ. Setting e. g. all momenta equal to μ, like it is done in certain renormalization schemes, this would provide limited knowledge on the full momentum dependence of Γ^n.

These equations can alternatively be derived in an integrated form. Since for any multiplicatively renormalized vertex function[10]

$$\Gamma_0=\Pi_i Z_i^{n_i}\Gamma,$$

where i enumerates the fields, and the n_i counts the times a field appears in the related expectation value. The vertex function Γ_0 only depends on the unrenormalized quantities, while Γ is the renormalized one. Since the left-hand side does not depend on μ, neither does the right-hand side. Taking a total derivative on the right hand side with respect to μ also yields (5.24).

The situation becomes particularly simple in a mass-independent scheme, i. e. where all renormalization conditions do not involve any mass explicitly, like the $\overline{\text{MS}}$ scheme[11]. Then in the equation (5.24) the functions do not depend on any ratios μ/m, and the equation can be integrated using the method of characteristics, yielding

$$\Gamma(\mu,\alpha,m,Q)=\exp\left(-\sum_i n_i\int\limits_\alpha^{\bar\alpha}dx\frac{\gamma_i(x)}{\beta(x)}\right)\Gamma(\bar\mu,\bar\alpha,\bar m,Q),$$

where the barred quantities are the initial conditions. The most important point is that the dependence on Q is not involved. The importance of this becomes even more evident

[10]Note that under certain conditions there may be mixing between various vertex functions under renormalization. Then this is a matrix equation. This complication will not be considered here. Also a dependence on the gauge-parameter is skipped over here, but can be considered if necessary.

[11]Such schemes have their own problems due to threshold effects, but this of no importance for the present case.

when rescaling all dimensionful quantities by a common factor, especially by $\bar{\mu}$ such that the initial condition becomes $\bar{\mu}$ independent. This yields

$$\begin{aligned}\Gamma(\mu,\alpha,m,Q) &= \bar{\mu}^{d_\Gamma}\exp\left(-\sum_i n_i\int_\alpha^{\bar{\alpha}}dx\frac{\gamma_i(x)}{\beta(x)}\right)\Gamma\left(1,\bar{\alpha},\frac{\bar{m}}{\bar{\mu}},\frac{Q}{\bar{\mu}}\right)\\ &= \exp\left(\int_\alpha^{\bar{\alpha}}dx\left(\frac{d_\Gamma}{\beta(x)}-\sum_i n_i\frac{\gamma_i(x)}{\beta(x)}\right)\right)\Gamma\left(1,\bar{\alpha},\frac{\bar{m}}{\bar{\mu}},\frac{Q}{\bar{\mu}}\right)\end{aligned}$$

where d_Γ is the naive/canonical/engineering dimension of Γ, i. e. the one obtained by just counting the dimensions of the fields in the expectation value. In the second line, the relation (5.23) has been used to include the canonical dimension into the exponent. Since without renormalization all γ_i vanish, they modify the canonical dimension. Thus, this combination is also called anomalous dimension.

A very interesting case arises if the function either has only one momentum scale Q or all momenta are equal, i. e. the so-called symmetric configuration. Then by setting $\bar{\mu}=Q$, the equation describes the dependency on Q alone. Since due to the γ_i, this is not a scaling just with Q^{d_Γ}, since the limits of the integral now implicitly depend on Q, this shows how the classical scaling at $m=0$ is broken due to quantum corrections.

Of course, the dependence of the relevant functions can, e. g., be obtained first in fixed-order perturbation theory, and then be used to integrate the equation. This yields so-called renormalization-group improved results, where the name 'group' is here of historical origin only. Still, the whole process can be remapped to a half-group.

E. g. for QCD it follows to NLO

$$\begin{aligned}Q^2\frac{d\alpha}{dQ^2} &= \beta(\alpha(Q^2))\approx-\alpha^2\sum_i\beta_i\alpha^i\\ \beta_0 &= \frac{33-2n_f}{12\pi}\\ \beta_1 &= \frac{153-19n_f}{24\pi^2}\\ \frac{Q^2}{\bar{m}(Q^2)}\frac{d\bar{m}(Q^2)}{dQ^2} &= \gamma_m(\alpha(Q^2))\approx-\alpha\sum_i\gamma_i\alpha^i\\ \gamma_0 &= \frac{1}{\pi}\\ \gamma_1 &= \frac{303-10n_f}{72\pi^2}\end{aligned}\tag{5.25}$$

where β_0, β_1, and γ_0 are the same in all mass-independent renormalization schemes, while the value of γ_1 is already specific to $\overline{\text{MS}}$. The approximation indicates that this will not

capture any non-perturbative contributions. As noted before in section 4.7, this yields the running of the coupling.

So far, this provides no way to determine the relevant functions, except other means. Non-perturbatively, these can be determined using lattice methods, or DSEs. However, the idea behind this approach can be extended to yield the so-called functional renormalization group (FRG), which will yield a hierarchy of equations similar to the DSEs. In fact, they are related by functional integration.

The idea is the following. Instead of performing regularization at the level of the correlation functions, e. g. by putting cutoffs somewhere, already the theory itself is regulated. This is performed by replacing the partition function

$$Z = \int \mathcal{D}\phi e^{iS} \to Z_k = \int \mathcal{D}\phi^{iS+i\Delta S_k} \tag{5.26}$$

where ΔS_k will regulate the action, a so-called regulator. For this purpose, a regulator scale k is introduced to control the regularization. Especially, it will be necessary that in the limit $k \to 0$ the regulator term vanishes, and the original action is recovered. Performing such a replacement can be done in several ways, and here only the most commonly used one will be discussed.

It makes use of the fact that all loops include propagators, and thus it suffices to modify the propagators to achieve its purpose. As an added possibility, it will be used not only to regulate the ultraviolet part of the theory, but also the infrared part, such as to avoid any infrared singularities introduce by tampering with the ultraviolet degrees of freedom. A suitable choice, already in Fourier space for simplicity, is then

$$\Delta S_k = \int \frac{d^4p}{(2\pi)^4} \phi(p)^\dagger R_k(p) \phi(p),$$

with suitable extensions for gauge fields or fermions in terms of tensor structures in color, flavor, and Lorentz space, usually the one of the ordinary action. Note that gauge theories play here a complicated role, and it is best to introduce this regularization after gauge-fixing, when the gluon propagator is well-defined.

The regulator $R_k(p)$ has to fulfill certain conditions to play its role,

$$\lim_{\frac{k^2}{p^2}\to 0} R_k(p) = 0 \tag{5.27}$$

$$\lim_{\frac{p^2}{k^2}\to 0} R_k(p) > 0 \tag{5.28}$$

$$\lim_{p^2\to\infty} R_k(p) \to \infty \tag{5.29}$$

These conditions hold true in Euclidean space-time. Formulating them in Minkowski space-time is complicated, and an implicit Wick rotation is made in the following wherever

necessary to give expressions a well-defined meaning. The first condition ensures that in the limit $k \to 0$ the original theory is recovered. The second condition provides an effective and finite mass-term, regulating the infrared. The last condition requires this mass to become infinitely large at large momenta, damping out the fields.

It is now possible to introduce again the free energy and the quantum effective action in much the same way as before, but all of them will depend on this regulator, and especially the regulator scale k. Especially the quantum effective action Γ_k can now be used to derive an explicit set of equations. Deriving it with respect to $t = \ln k$ yields

$$\begin{aligned}
\partial_t \Gamma_k &= -\partial_t W_k = -\partial_t \ln Z_k = -\frac{1}{Z}\partial_t Z_k = -\frac{1}{Z_k}\int \mathcal{D}\phi \partial_t e^{iS+i\Delta S_k} \\
&= -\frac{1}{Z_k}\int \frac{d^4p}{(2\pi)^4}\partial_t R_k(p) \int \mathcal{D}\phi\phi^\dagger(p) R_k(p)\phi(p) e^{iS+i\Delta S_k} \\
&= -\frac{1}{2}\int \frac{d^4p}{(2\pi)^4} D_k^2(p)\partial_t R_k(p) = \frac{1}{2}\int \frac{d^4p}{(2\pi)^4}\frac{1}{\Gamma_k^2(p)+R_k(p)}\partial_t R_k(p) \qquad (5.30)
\end{aligned}$$

where in the last line (5.11) has been used, and the fact that the appearance of the regulator will induce a shift in the Legendre transformation by the regulator, as in this case the functional dependency is fixed, and therefore this part of the propagator is known exactly. The generalization to more particles is almost trivial. Since all fields have to be regulated, (5.26) just receives a term for each field, and hence (5.30) is then a sum of terms.

The expression (5.30) is an equation for the quantum effective action as a function of the full 2-point function, and thus a coupled system. To solve it, it is possible to proceed as in the DSE case. By performing functional derivatives of (5.30), a tower of equations for the n-point functions is created, which in the present case are then integro-differential equations, for all the n-point functions. These equations are structurally different from the DSEs, as e. g. only one momentum integration appears, and there is always an explicit regulator. The latter is, on a technical level, quite convenient, as it renders all integrals automatically finite, and the equations are hence already the one for the renormalized quantities. Like in the DSE case, a full solution to these equations provides all information on a theory. Hence, the solutions at $k = 0$ will then automatically also solve the DSEs, and vice-versa.

However, as in the DSE case, these equations can still not be solved exactly. Furthermore, all n-point functions depend on an additional parameter, k. Hence, it is in general not possible to solve these equations exactly, and approximations, again called truncations, are required. These can be done, e. g. in the same way as in the case of DSEs, i. e. by neglecting or making assumptions on the behavior of certain n-point functions, until a

solvable system is obtained. In addition, the additional dependence on k permits a further type of approximation. In this case, the n-point functions are only regarded as functions of k, and the equations (5.30) then become ordinary partial differential equations in k, which are more accessible to solutions. Still, in general no solution of the whole system is possible.

In such an approximation, the n-point functions in the limit $k \to 0$ will be just constants. This seems to be a drastic ansatz, but here a link to the idea behind the Callan-Symanzik equation (5.24) shows that much more information is included. The scale k acts like the renormalization scale. Thus, the development of n-point functions with k can be regarded as the development with the renormalization scale, and hence provides limited information on the n-point functions as well.

It should be noted that in the limit $k \to 0$ the original theory is recovered, no matter which regulator is chosen, provided it fulfills the conditions (5.27-5.29). But this is only true without approximations. Approximations will in generally hamper this. Moreover, when approximations are performed, the results will in general depend on the regulator. This is an additional type of systematic error compared to DSEs. Though several approaches have been developed over time to minimize this effect, as a systematic error this is in principle a uncontrollable one, just like all approximations in a non-perturbative setting. Hence, special care has to be exercised when choosing the regulator, especially when it is unavoidable that it breaks one or more symmetries of a theory.

Still, this requires to provide a starting condition for the differential equations. In an asymptotically free theory, like QCD, this is rather simple, since by choosing k large enough, these will be just the perturbative (tree-level) action. Things are more involved in other theories, and some of the biggest successes of the FRG approach arose there. However, this is beyond the scope of this lecture.

5.1.8 Low-energy effective theories

The non-perturbative methods introduced so far were all quite powerful in that they all provide an exact solution. However, they are all limited by the ability to perform the corresponding calculations exactly, which at the current time is impossible for QCD.

An alternative approach are effective theories. In this approach, it is not attempted to solve the full theory. Rather, a simpler theory is designed, which should describe certain phenomena of the original theory in a certain energy regime, but should be treatable using simpler methods, e. g. perturbation theory. In QCD, due to asymptotic freedom, such theories are primarily low-energies ones.

Since such theories should be limited to a very small energy range, and usually only

to a few orders in perturbation theory, requirements like renormalizability are of less importance. Hence, in each order of perturbation theory additional counter-terms will appear. These must be fixed, e. g., by experimental input. However, then all calculations in such an effective theory are uniquely determined.

Such effective theories can be systematically developed by including all operators in the (quantum effective) action compatible with the global symmetries of an original theory. Hence, this low-energy theory is usually not a gauge theory. Further reduction can be achieved, but at the cost of no longer being a systematic expansion. Examples are the Nambu-Jona-Lasinio model to be used in the next section, the linear σ-model used previously, or the quark-meson model. A full systematic theory is chiral perturbation theory introduced in the next section.

The reason for postponing the explicit construction of such models now is that they are mostly motivated by chiral symmetry to be discussed next.

5.2 Chiral symmetry breaking

5.2.1 Flavor symmetry and chiral symmetry and the hadron multiplet structures

5.2.2 Symmetry at the quark level

One of the most important features of QCD is the presence, or rather the absence of a symmetry: Chiral symmetry. The QCD Lagrangian (4.1) exhibits one interesting additional symmetry if the masses of all the fermions are set to zero, besides the then manifest flavor symmetry. This additional symmetry emerges by the combination of a flavor or fermion number transformation and an axial transformation. Axial transformations are a special property of fermions, and there is no analogue for bosons of arbitrary spin. It is mediated by multiplying every fermion field by the matrix $\exp(i\alpha\gamma_5)$, where α is a real parameter, and $\gamma_5 = -i\gamma_0\gamma_1\gamma_2\gamma_3$ is a combination of the γ-matrices. This can be shown using the fact that γ_5 anti-commutes with all γ_μ. The anti-fermion field is transformed by the corresponding hermitian conjugated phase factor.

This phase symmetry adds an additional U(1) symmetry to the theory, which is called the axial symmetry $\mathrm{U}_A(1)$. In addition, like the generalization of the fermion number symmetry U(1) to the flavor symmetry $\mathrm{SU}(N_f)\times\mathrm{U}(1)=\mathrm{U}(N_f)$ for N_f flavors, it is possible to enlarge the axial symmetry to an axial flavor symmetry, called chiral symmetry. This name stems from the fact that it turns out that it connects fermions with spin projections along and opposite to their momentum direction, i. e. of different helicities. Since

these projections yield classically a left-handed and right-handed screw[12], the name chiral, Greek for handedness, is assigned. The total symmetry of the theory is therefore $\mathrm{SU}(N_f)\times\mathrm{SU}_A(N_f)\times\mathrm{U}(1)\times\mathrm{U}_A(1)$ for N_f flavors of quarks.

Of these symmetries, the axial symmetry is actually broken by an anomaly during quantization. This will be discussed in detail in section 5.6. Non-zero quarks masses break the chiral symmetry explicitly to a diagonal subgroup of the full symmetry group, and the non-degenerate quark masses then finally break the flavor symmetry just to a diagonal flavor number symmetry, $\mathrm{U}(1)^{N_f}$. Hence, little is left from the classical symmetries of massless QCD. But because the masses of the up and down quarks are small it turns out that the consequences of the explicit breaking are such that the symmetry is still a good guide-line. In fact, on top of the explicit breaking the symmetry receives contributions from a spontaneous breakdown, and this fact actually shapes the low-energy dynamics of QCD.

One of the major consequences is the relation of the masses of the hadrons to the masses of the quarks. The mass of the proton is known very precisely to be 938.3 MeV, and the neutron to be 939.6 MeV. This implies that the mass difference between up and down quarks must be tiny. The Δ is somewhat heavier, about 1230 MeV. This can be understood as an excited state, and it is therefore heavier. Most ground state mesons have a mass of about 600 MeV or more. All this suggest a mass of about 300 MeV for the up quark and down quark, with very little difference[13]. But two mysteries appear. One is that the pions are very light, just about 140 MeV. The second is that any attempt to directly measure the quark masses yield consistently a mass of about 2.3(7) MeV for the up quark, and 4.8(5) MeV for the down quark. Though the difference is consistent, the absolute values are much smaller than the suggested 300 MeV from the nucleon properties.

The resolution of this puzzle is found in the spontaneous breaking of chiral symmetry. To understand it, it is convenient to neglect the small masses of the up and down quarks, and also all heavy flavors. Then, chiral symmetry is exact on the level of the Lagrangian. However, it has to be broken in nature, as many indirect evidence shows. E. g. chiral symmetry implies that bound states of opposite parity, but otherwise identical content, should have the same mass. But the mass splitting between such bound states for hadrons is large. E. g., the parity partner of the nucleon is called the $N(1535)$ and has, as its name suggests a mass of 1535 MeV, about 50% heavier than the nucleons. This is much larger

[12]Note that the question of what is left-handed or right-handed depends upon whether you look at a screw from the top or the bottom. Since different conventions on how to look at a screw are in use, care should be taken.

[13]Though this difference is crucial for making the proton lighter and thus (more) stable (than the neutron), and therefore chemistry possible.

than expected due to the explicit breaking of the chiral symmetry because of the small current quark masses. Thus, chiral symmetry must be much stronger broken than just from the current quark masses. The strong interactions must spontaneously break it. The details of this breaking are governed by Goldstone's theorem.

5.2.3 The Goldstone theorem

5.2.3.1 At tree-level

To lay out the Goldstone theorem, it is helpful to investigate a very simplified model, before returning to QCD: Once again, the linear-σ model. In this course, it will also become clear why the linear-σ model is such a good low-energy approximation to QCD, and it will justify why it will later be used as a starting point to develop chiral perturbation theory as a more comprehensive description of low-energy QCD in section 5.2.6.

Take four scalar particles, which can interact with themselves. Such particles can be described by four real scalar fields ϕ_i, arranged in a four-dimensional vector in an internal space. The linear σ-model then has a classical Lagrangian of

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - U(\phi^\dagger \phi), \tag{5.31}$$

where U describes the potential for the field, which can only come from interactions between the two fields, and $\dagger$ is here just taking the transpose. This theory has an O(4) symmetry, as rotations of the four real components do not affect the scalar products. The simplest possibility to obtain a symmetry for this theory is that the potential depends only on the product $\phi^\dagger \phi$.

To obtain the simplest example for a symmetry, take the potential

$$U(\phi^\dagger \phi) = \frac{\mu^2 v^2}{4} - \frac{\mu^2}{2}\phi^\dagger \phi + \frac{\mu^2}{4v^2}(\phi^\dagger \phi)^2. \tag{5.32}$$

The pre-factors, i. e. coupling constants, as well as the irrelevant constant term have been chosen judiciously such that the result will be looking simple. This potential, as well as the kinetic term, is invariant under the O(4) rotation $\phi \to \exp(i\alpha^a \tau^a)\phi$, where τ^a are the generators of the O(4) group. It is a bit odd theory, as the quadratic term, usually associated with a mass, is negative, and thus the mass is purely imaginary. This would be called a tachyon. This is, however, not a problem. The sign of the quartic term is positive. Therefore, the energy is bounded from below classically, and the theory remains stable. It is therefore just an odd term in the potential energy.

To proceed, an interesting question is what the classical lowest energy state is. Since the kinetic term is positive, any spatial or temporal variation would increase the total energy.

Hence, the state of lowest energy is necessarily a field ϕ_0 constant throughout space and time, which minimizes the potential (5.32). This constant is found to be $\phi_0 = vn$, where n is an arbitrary unit vector in the four-dimensional internal space, i. e. for all directions of n it is a minimum. Thus, the solution manifold is highly degenerate. This is a consequence of the symmetry: Any change in n can be offset by a symmetry transformation, without changing the physics.

One can proceed by specifying n further. However, any choice is physically indistinct, and therefore arbitrary. But for further calculations keeping n manifest is awkward, and therefore in the following the explicit choice $n = (0, 0, 0, 1) = e_1$ is made. Since the choice is arbitrary, the symmetry is not violated. However, the symmetry is no longer manifest either. One therefore speaks of hiding the symmetry, or a hidden symmetry. However, in an abuse of language, though a common one, this is called a broken symmetry.

To make the situation more transparent, the next step is to shift the field ϕ by its value at the minimum, which in the quantum theory is $\langle\phi\rangle = ve_1$ and thus its vacuum expectation value (vev), i. e. replace

$$\phi(x) \to ve_1 + \eta(x)e_1 + \xi_i(x)e_i, \tag{5.33}$$

with $i = 2...4$, and the e_i being unit vectors. In this way, fluctuations η and ξ of the fields around the classical vacuum v can be studied. Inserting this into the Lagrangian (5.31) yields

$$\mathcal{L} = \partial_\mu\eta\partial^\mu\eta + \partial_\mu\xi_i\partial^\mu\xi_i - \mu^2\eta^2 + \frac{\mu^2}{v}\eta^3 + \frac{\mu^2}{v}\eta\xi_i\xi_+\frac{\mu^2}{4v^2}\eta^4 + \frac{\mu^2}{4v^2}(\xi_i\xi_i)^2. \tag{5.34}$$

This Lagrangian shows now a number of very interesting features, which are very generic.

The first is that the two fields η and ξ behave differently. While there is a mass term, now with the correct sign, for η, giving it a mass of $\sqrt{2}\mu$, there is no mass for the ξ. Pictorially, one can think of η as excitations which describe fluctuations out of the minimum, while the ξ, which are orthogonal to the direction of the chosen vacuum, move between the different minima of the potential. Since the vacua all have the same energy, this does not cost any energy, and therefore these modes are massless. This is a generic feature of such situations, and is known as Goldstone's theorem. The massless modes are therefore called Goldstone bosons. In a nutshell, it is the statement that there are as many massless particles as there are directions in which the minima are equivalent. The precise formulation is that there are as many massless particles as there are generators of the symmetry group minus the number of generators of symmetry transformations after all possible remaining choices have been made. Here, this is a breaking from O(4) to O(3), the remaining invariance group. These are $4(4-1)/2 - 3(3-1)/2 = 6 - 3 = 3$, and thus there are 3 massless modes. This will be proven for the quantum theory below.

The second is that there are now many different interactions between the fields η and ξ. However their couplings, i. e. their pre-factors, are not all different, but completely determined by the original parameters. The reason is that the symmetry is just hidden. To ensure that any symmetry transformation is still valid requires that the various interactions cannot have arbitrary pre-factors, because otherwise it would no longer be invariant under the symmetry transformation

$$(v + \eta(x))e_1 + \xi_i(x)e_i \to e^{i\alpha^a\tau^a}((v + \eta(x))e_1 + \xi_i(x)e_i), \tag{5.35}$$

where it should be noted that the vacuum solution v is also transformed accordingly. For that to work out, it is necessary to keep track when changing from (5.31) to (5.34), which occurrence of v stem from the original coupling constants in (5.31), and which from the shift (5.33), since only the latter are affected.

If at any point a term is added to the Lagrangian, which violates the symmetry, the symmetry becomes explicitly broken. The most obvious way is to add a mass term for the ξ field to the Lagrangian (5.34). Then, the Lagrangian is no longer invariant under the symmetry transformation (5.35). Of course, this can be translated back into the original Lagrangian (5.31), where it takes the form of an additional quadratic term in the potential of type $-(r\phi)^2$, where r is an arbitrary vector in the internal space. The effect is essentially that the potential is tilted, and the vacuum state has now a unique solution vr, which has no longer an invariance. This gives also a physical explanation for the mass: Since there are no degenerate vacuum solutions anymore, any movement increases the energy.

This illustrates how pions gain their small mass. The breaking of chiral symmetry would lead to a number of Goldstone bosons. In case of two quark flavors[14], there will be three Goldstone bosons, which are the three pions. These would be massless, if the quarks would be massless. However, because of the small current mass of the quarks, the symmetry is not exact, but rather explicitly broken. This gives the mass to the pions. That the masses of the pions are still large compared to current masses of the quarks is a dynamical effect. Approximately, the pion masses scale linearly with the current masses of the quarks, the so-called Gell-Mann-Oaks-Renner (GMO) relation

$$m_\pi^2 = \frac{m_q \langle\bar{q}q\rangle}{f_\pi^2}, \tag{5.36}$$

where $\langle\bar{q}q\rangle$ is the quark condensate and f_π is the so-called pion decay constant, and has in these conventions the value of roughly 92 MeV. Note that this pre-factor is large.

Before continuing on, a few words about subtleties and semantics must be said. When going to the quantum theory, quantum effects do what they always do, and they will mix

[14]It can be shown that the number of colors does not matter.

all the degenerate vacuum states, and no vacuum will be preferred. Therefore, the vacuum state will exhibit a perfect symmetry, in contrast to the classical case. But the quantum system also carries the seed of the classical physics within, as it is metastable. If any arbitrarily small external perturbation, e. g. an infinitesimal mass for the ξ_i from some other physics process, arises, it will immediately have a unique vacuum state, in which the symmetry is no longer realized. Hence, though strictly speaking the system without external influence is perfectly symmetric, the presence of this metastability has lead to the expression that the symmetry is nonetheless spontaneously broken.

In fact, even though the symmetry is exact, a full non-perturbative calculation shows that the system has both an ordinary massive and a massless excitation, and hence the most pertinent feature of the Goldstone theorem are realized even with the symmetry present. Of course, in a perturbative calculation this will not show. Since perturbation theory only permits very small deviations from the vacuum state, the particles will still appear to be tachyons, as the relevant Lagrangian is (5.31). To cure this problem, one can introduce a weak external perturbation to the theory, which prefers a single vacuum, perform perturbation theory around this vacuum, i. e. using the Lagrangian (5.34) instead, and remove the external perturbation at the end[15]. Then, also in perturbation theory the system exhibits a massive and a massless particle. Especially because of this trick, which is extremely useful in the standard model, the more appropriate notion of hidden symmetry is nowadays very rarely used, and almost always the situation will be denoted by spontaneous symmetry breaking.

A more formal statement can be obtained as follows. Take as the symmetry group a semi-simple Lie-group G. Then the symmetry transformation of the associated real fields transforming under a real representation of the symmetry group are given by

$$\delta\phi_i = iT_a^{ij}\phi_j\theta^a \tag{5.37}$$

with arbitrary infinitesimal parameters θ^a counting from 1 to dim$G$. The Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi_i\partial^\mu\phi^i - V(\phi)$$

must be invariant under a group transformation. Since the kinetic term is trivially so, this implies for the potential

$$0 = \delta V = \frac{\partial V}{\partial\phi_i}\delta\phi^i = i\frac{\partial V}{\partial\phi_i}T_{ij}^a\phi^j\theta_a.$$

[15]There are some subtleties associated with non-analyticities in this case, but they are irrelevant in perturbation theory.

Since the parameters are arbitrary, this can only be satisfied if

$$\frac{\partial V}{\partial \phi_i} T^a_{ij} \phi^j = 0$$

holds. Differentiating this equation with respect to ϕ_k yields

$$\frac{\partial^2 V}{\partial \phi_k \partial \phi_i} T^a_{ij} \phi^j + \frac{\partial V}{\partial \phi_i} T^a_{ik} = 0.$$

The symmetry is hidden by expanding around the minimum of the potential, and therefore the first derivatives have all to vanish. The symmetric matrix of second derivatives is positive at a minimum, i. e., has only positive or zero eigenvalues

$$\frac{\partial^2 V}{\partial \phi_k \partial \phi_i} = (M^2)^{ki}.$$

Expanding now, as before, the field around the classical minimum at $\psi_i = \phi_i - f_i$, the quadratic order of the Lagrangian reads

$$\mathcal{L} = \frac{1}{2} \partial_\mu \psi_i \partial^\mu \psi^i - \frac{1}{2} (M^2)_{ki} \psi_k \psi_i + ...$$

Since the matrix M, the mass matrix, is semi-definite positive, all particles have at tree-level only positive or zero mass.

The conditional equation for a classical minimum reads

$$(M^2)^{ki} T^a_{ij} f^j = 0.$$

If the classical minimum is invariant under a subgroup H of G, this subgroup is called the stability group of G. As a consequence for generators t_a out of H

$$t^a_{ij} f^j = 0$$

holds. Therefore, the value of the mass matrix is irrelevant for these directions, and there can be dim$H$ massive modes. However, for the coset space G/H with generators τ^a, the corresponding equations

$$\tau^a_{ij} f^j \neq 0.$$

are not fulfilled, and therefore the corresponding entries of the mass-matrix have to vanish. Since these represent dim(G/H) equations, there must be dim(G/H) massless modes, the Goldstone modes.

5.2.3.2 Quantized Goldstone theorem

To determine the consequence of hiding symmetry at the quantum level, it is useful to investigate the normalized partition function

$$T[J_i] = \frac{Z[J_i]}{Z[0]} = \frac{1}{Z[0]} \int \mathcal{D}\phi_i \exp\left(i \int d^4x (\mathcal{L} + J_i\phi_i)\right),$$

with the same Lagrangian as before. Since the Lagrangian and the measure are invariant under a symmetry transformation[16], the variation of the partition function must vanish

$$0 = \delta Z[J_i] = \int \mathcal{D}\phi_i e^{iS + i\int d^4x J_i\phi_i} \int d^4x \left(\frac{\partial \delta\phi_i}{\partial \phi_j} + \delta\left(iS + i\int d^4x J_i\phi_i\right)\right).$$

The first term is the deviation of the Jacobian from unity. As the measure is invariant, it vanishes. The second is the variation of the action, which also vanishes. Only the third term can contribute. Since all variations are arbitrary, it thus follows

$$\int d^4x J_i T^a_{ik} \frac{\delta T[J_i]}{i\delta J_k} = 0,$$

where it has been used that $Z[0]$ is a constant, and the order of functional and ordinary integration has been exchanged, and

$$\frac{\delta T[J_i]}{i\delta J_i} = \frac{1}{Z[0]} \int \mathcal{D}\phi_i\phi_i \exp\left(i \int d^4x (\mathcal{L} + J_i\phi_i)\right).$$

Furthermore, it has been used that all variations are independent, thus delivering dim$G$ independent equations.

Since

$$\delta T \equiv \delta\left(e^{T_c}\right) = e^{T_c}\delta T_c,$$

and the factor $\exp(T_c)$ is not depending on x, since it is a functional, this can be rewritten in terms of the generating functional T_c for connected Green's functions as

$$\int d^4x J_i T^a_{ik} \frac{\delta T_c[J_i]}{i\delta J_k} = 0.$$

This can furthermore transformed into an equation for the vertex (i. e., connected and amputated Green's functions) generating functional Γ, which is related to the connected

[16] If the measure would not be invariant, this would lead to an anomaly. This happens, e. g., in the case of the axial symmetry.

one by a Legendre transformation[17]

$$\begin{aligned} i\Gamma[\phi] &= -i\int d^4x J_i\phi_i + T_c[J] \\ \langle\phi_i\rangle &= \frac{\delta T_c[J]}{i\delta J_i} = \langle 0|\phi_i|0\rangle[J_i] \\ J_i &= -\frac{\delta\Gamma[\phi]}{i\delta\phi_i}, \end{aligned} \tag{5.38}$$

by exchanging the derivative and the source. This yields finally

$$\int d^4x \frac{\delta\Gamma}{\delta\phi_i} T^a_{ik}\langle\phi_k\rangle = 0. \tag{5.39}$$

For the fields developing a vacuum-expectation it then holds

$$\begin{aligned} f_i &= \langle 0|\phi_i|0\rangle = \frac{\delta T_c}{i\delta J_i}[0] \\ 0 &= J_i = -\frac{\delta\Gamma}{i\delta\phi_i}[f_i]. \end{aligned} \tag{5.40}$$

The inverse propagator of the fields ϕ_i is given by

$$\frac{i\delta^2\Gamma}{\delta\phi_i(x)\delta\phi_j(y)}[f_i] = -(D^{-1})_{ik}(x-y). \tag{5.41}$$

An expression for this object can be obtained by differentiating (5.39) with respect to the field once more yielding

$$\int d^4x \left(\frac{\delta^2\Gamma}{\delta\phi_i(x)\delta\phi_j(y)} T^a_{ik}\langle\phi_k\rangle + \frac{\delta\Gamma}{\delta\phi_i} T^a_{ii}\delta(x-y)\right).$$

The last term vanishes since the generators are Hermitian and traceless and, even if not, because $\delta\Gamma/\delta\phi_i = 0$, while the first one is just the Fourier-transform of the inverse propagator at zero momentum, yielding

$$(G^{-1})_{ij}(p=0)T^a_{ik}f_k = 0.$$

Thus, there must vanish as many inverse propagators as there are non-zero f_i. At tree-level the inverse propagator is given by

$$(G^{-1})_{ij} = \delta_{ij}(p^2 + m^2),$$

[17] Using the same notation for the field and its one-point Green's function.

and thus this implies that the pole mass must vanish, the propagator becomes that of a massless particle, just as classically. However, in the full quantum theory, the mass becomes momentum-dependent, and the full propagator takes the form

$$G_{ij}^{-1} = Z_{ij}(p^2)(p^2 + M_{ij}(p^2)).$$

Thus, only the combination $Z(0)/M(0)$ must vanish. The propagator still has a pole at this point. However, there can also be further poles at non-zero momentum. Therefore, in a Goldstone channel can be also additional particles, which can even be stable. In the case of QCD, where chiral symmetry dictates that the channel has to have quantum numbers 0^-, a pseudoscalar, there are no additional stable states, but several unstable ones, the excited pions.

It should be noted here that this derivation only applies to a global symmetry. When the symmetry becomes local, the Jacobian determinant cannot be ignored anymore, and modifications will be necessary. From an axiomatic point of view, the reason for failure is that implicitly a positive definite Hilbert space has been assumed, which is not the case for a gauge theory.

5.2.4 The NJL model version

While the linear-σ model is an example of how the Goldstone theorem works, it does not provide any insight into chiral symmetry breaking; fermions are required. It is, however, not necessary to use full QCD to understand the underlying mechanism of chiral symmetry breaking. A simpler example of the concept is given by the so-called Nambu-Jona-Lasinio (NJL) model. Its Lagrangian for N_f flavors of quarks is given by

$$\mathcal{L} = \sum_f \bar{\psi}_f(i\gamma^\mu\partial_\mu - m_f)\psi_f + G\left((\bar{\psi}_f\psi_f)^2 - (\bar{\psi}i\gamma_5\psi)^2 - (\bar{\psi}\tau_f^i\psi)^2 + (\bar{\psi}i\gamma_5\tau_f^i\psi)^2\right).$$

Every quark field ψ_f has also $N_c = 3$ components, and hence the theory has a SU(N_c) global color symmetry, rather than the gauge symmetry of QCD. The first two terms are iso(flavor)scalar interactions, while the last terms are iso-vector interactions. Since the theory involves a coupling constant of energy dimension two, it is perturbatively (and likely also non-perturbatively) not renormalizable in four dimensions. However, since it should only serve as a low-energy effective theory, there is no problem in including an explicit regularization description into the theory, e. g. an explicit cutoff Λ.

The full quark propagator

$$S(p) = Z_f(p)(\gamma^\mu\partial_\mu - M_f(p) + i\epsilon)^{-1}$$

has two tensor structures, described by a wave-function renormalization $Z(p)$ and a mass function $M(p)$. At tree-level, $Z = 1$ and $M_f = m_f$. The DSE for the quark propagator can be derived in the same way as for QCD. In the simplest approximation, all terms but the tadpole are neglected, and the self-interaction are taken at tree-level. The quark propagator in this Hartree approximation has then still two tensor structures, but only the scalar part is modified while Z remains 1. After taking a trace in Dirac, color, and flavor space this yields the Hartree approximation after Wick rotation

$$M_f = m_f + 8N_f N_c G \int_0^\Lambda \frac{d^4p}{(2\pi)^4} \frac{M_f}{p^2 + M_f^2}. \tag{5.42}$$

Thus, in this approximation the mass is at most shifted by a constant, but remains momentum-independent. This integral equation cannot be solved analytically, but is accessible to numerical methods.

The results are rather interesting. Setting $m_f = 0$ is the so-called chiral limit of massless quarks. In this case, the equation has the trivial solution $M_f = 0$. However, it turns out that if G rises above a certain critical limit, or alternatively the product $N_c N_f$ does, a second solution appears, with $M_f > 0$. Hence, a sufficiently strong interaction yields a non-vanishing effective mass for the quarks. Assuming now that the four-quark interaction is an effective description of the gluon interaction between quarks at low energies, this implies that a sufficiently strong gluon interaction will create an effective mass for the quarks, giving them their constituent mass. If the quarks have a finite current mass m_f, there is only a solution with non-vanishing mass-shift, increasing the constituent quark mass above the current quark mass.

That this indeed breaks chiral symmetry can be seen by two arguments. First, a non-vanishing effective mass is already indicative of the loss of chiral symmetry - massive particles are no longer helicity, and thus chirality, eigenstates. The second is the explicit calculation of the quark, or chiral, condensate $\langle\bar{\psi}(0)\psi(0)\rangle$. If chiral symmetry is intact, this quantity is zero. However, the operator content shows that this is the integrated quark propagator

$$\langle\bar{\psi}_f(0)\psi_f(0)\rangle = -i \int_0^\Lambda \frac{d^4p}{(2\pi)^4} \mathrm{tr} S_f(p) = -\frac{M_f - m_f}{G}, \tag{5.43}$$

where the second equality stems from resolving the DSE for the quark propagator in the Hartree approximation for the also appearing trace. Hence, in this approximation there is a chiral condensate if and only if $M_f \neq m_f$, and thus this is sufficient for a breaking of chiral symmetry. This also implies that any non-zero current mass explicitly breaks chiral symmetry, in agreement with the explicit investigation at the Lagrangian level.

Of course, at least in the chiral limit this leaves open the question whether the chirally broken solution is indeed the realized one. To decide this, it is, like for the Goldstone theorem, necessary to investigate what is the solution with the lowest potential energy. To do so, the simplest approximation is the so-called mean-field approximation. For this, in the potential

$$V = -G\left((\bar{\psi}\psi)^2 - (\bar{\psi}i\gamma_5\psi)^2 - (\bar{\psi}\tau_f^i\psi)^2 + (\bar{\psi}i\gamma_5\tau_f^i\psi)^2\right)$$

all non-linear terms are replaced like

$$(\bar{\psi}_f\psi_f)^2 \to \langle\bar{\psi}_f\psi_f\rangle^2,$$

as with the vacuum expectation value at the classical level of the Goldstone theorem. Since only the first term acquires a non-vanishing vacuum expectation value, this is the only relevant one. Since it has a negative sign, the solution is more preferred where chiral symmetry is broken.

As a side-remark, the NJL model was inspired by the BCS theory of superconductors. Since in superconductors the chiral condensate becomes the meaning of the gap between two bands, the corresponding equation is called gap equation, and likewise is (5.42) called gap equation, as are the more complicated ones in other theories.

5.2.5 Full QCD

Of course, this is a rather crude approximation, even of the very simplified NJL model. However, these results hold in the NJL model even in much more sophisticated truncations, without changing the qualitative picture. A similar calculation in QCD with a fully resolved gluon interaction is currently at the forefront of research[18]. In QCD, it is found that the results follow a similar pattern as in the NJL model, though here it appears that many subtle and balancing effects in the various tensor structures of propagators and vertices play a role. In a more direct way, these quantities can also be evaluated using lattice gauge theory, and show the same results. Here, however, the differences to the NJL model are completely obscured.

[18]There are several earlier calculations with effective gluon interactions, i. e. some kind of modeled momentum-dependent interaction between the quarks, which all show a qualitatively similar picture to the NJL case. Such calculations are often dubbed rainbow truncation, as the missing back-reaction of the quarks to the gluons yields that the quark DSE can be represented as a full line of all possible gluon exchanges over the quark line.

The most interesting result is, however, the modification to the running mass. The perturbative result (5.25) yields

$$\bar{m}(Q^2) = \bar{m}_0 \alpha(Q^2)^{\frac{\gamma_0}{\beta_0}} \left(1 + \frac{\beta_1}{\beta_0}\alpha(Q^2)\right)^{\frac{\beta_0\gamma_1 - \beta_1\gamma_0}{\beta_0\beta_1}} .$$

Hence, if the mass is zero at tree-level $\bar{m}_0 = 0$, i. e. in the chiral limit, it will remain so, in contradistinction to the non-perturbative result. At the same time, outside the chiral limit, it will diminish logarithmically in the ultraviolet, but will have a Landau pole at a finite value. The non-perturbative result, however, generically, and rather independent of truncation, cures this Landau pole, and leads to an infrared finite value of $\bar{m}(0) = M$, which plays roughly the same role as the mass obtained previously in the NJL model. Due to asymptotic freedom, the behavior at large Q^2 will still be the same as in perturbation theory, and especially the pre-factor is the one of the current mass, not of the constituent mass M.

Note that neither perturbatively nor non-perturbatively $\bar{m}$ is a well-behaved mass for the quark, and neither is M. These quantities depend on the renormalization scheme, and are therefore not physical. If a pole mass for the quark exists, a problem to be addressed in section 5.5.4, it will be scheme-independent, and solely defined by the poles of the quark propagator. This does not happen even in perturbation theory.

Considerably different from the NJL model, however, is the dependence on N_f. Increasing N_f in full QCD actually weakens, and at some point removes, chiral symmetry breaking, while it strengthens it in the NJL model. The reason is a back-coupling of an increase of the quarks flavors to the strength of the gluon-quark interaction, which in this form is not captured by the NJL model. This directly shows also one of the many limitations of the NJL model, despite its many phenomenological successes. Still, it gives at low N_f a reasonable qualitative picture of chiral symmetry breaking.

However, in full QCD additional complications arise. One is that the chiral condensate (5.43) is no longer a good observable. Since QCD is renormalizable, this quantity is, as noted in section 4.6.6, not renormalization-group independent. Hence, this condensate cannot be observed, and its value is meaningless. However, the combination $m_f\langle\bar{\psi}_f(0)\psi_f(0)\rangle$ is renormalization-group invariant, and is physical meaningful. This quantity is, in fact, as good an indicator for chiral symmetry breaking, e. g. in the context of the Gell-Mann-Oakes-Renner relation (5.36).

As the pion is the Goldstone boson, its properties are determined directly by the breaking of chiral symmetry. Especially, in the chiral limit it becomes massless. Since there are no other massless particles in QCD in which it could decay, and for the charged ones even not in the standard model, it becomes also stable. Thus, effects mediated by the

pions will become effectively long-range, and the low-energy behavior of QCD will become quite different. As long as the chiral symmetry is explicitly broken, however, the pion is massive, and pion exchange in QCD remains short-ranged.

5.2.6 Chiral perturbation theory

Noticing that QCD has a mass gap due to the pion mass, it appears reasonable to use the Appelquist-Carrazone theorem to construct an effective low-energy theory of hadronic interaction, and thus possibly even nuclear interaction. The simplest example of this is the (non-)linear σ-model. However, this theory cannot cover the whole complexity of QCD. Thus, it needs to be expanded. Especially, as a low-energy effective theory, it is no longer required to be renormalizable. In principle, it is possible to thus write down an effective theory of the pions which include all, i. e. infinitely many, terms in the Lagrangian, which conserve the symmetries of QCD. These would be the approximate flavor and chiral symmetry, the former broken explicitly by the difference of up and down masses and electric charges, the latter by the dynamical breaking of chiral symmetry.

With an infinite number of terms, this would be not useful. However, the Appelquist-Carrazone theorem then gives an ordering principle. The expansion should be in powers of Q^2/m^2, where Q is the energy scale, and m is the pion mass. At the Lagrangian level, this would be an ordering in momentum space. In addition, there can be interactions of any order without momenta. These are in this naive counting always suppressed by some effective coupling constant of ratios of the pion mass to some larger mass scale. If these effective coupling constants are assumed to be of typical hadronic size, i. e., some tens of MeV, then they are of roughly the same importance as the derivative terms of the same canonical dimensions. Hence, the expansion is in the canonical dimension.

Of course, as soon as the canonical dimension exceeds four, i. e. anything beyond the σ-model, this theory is no longer (perturbatively) renormalizable. Hence, the definition requires the inclusion of new counter-terms at every order, with independent pre-factors to be determined by comparison to experiment. However, after this finite number of inputs, the theory becomes predictive to the accuracy of the order of the expansion.

The resulting effective theory is known as chiral perturbation theory. It has been successfully exploited to determine various interactions between hadrons at low energies. One particular triumph is that it permits to calculate the interaction between nucleons, and thus provide input to nuclear physics from a level close to full QCD. There exist extensions which also include strange particles as well as baryons. However, the larger the energy scale, the worse the approximations become.

In fact, it is not entirely clear to which extent chiral perturbation theory can be con-

sidered to be a really good approximation to QCD. Formally, it is clear that the expansion cannot be correct from the outset, as any perturbative expansion. However, this qualitative statement does not imply that it is quantitatively bad. One of the unique predictions which confirms whether chiral perturbation theory actually works is that there are logarithms in the pion mass. These control the approach to the chiral limit, and the logarithms then yield the long-range behavior for a vanishing pion mass. These so-called chiral logs are a unique feature of the chiral dynamics underlying chiral perturbation theory. Therefore, one test of chiral perturbation theory is to perform calculations in full QCD for different pion masses, e. g. using lattice calculations, and then identify these chiral logs. If this behavior is not observed, chiral perturbation theory would not be a good approximation. At the current time, it appears that for the physical pion mass the situation is at the border line of the reliability of chiral perturbation theory, but this is not yet finally settled.

This question seems to be not so terribly important, as experiment can always decide the reliability of an approximation. But chiral perturbation theory has the unique feature that it can analytically access any regime in its range of applicability. This is, e. g., important when going to nuclear astrophysics, where there are many processes which are decisive for the stars, but are experimentally for all practical purposes impossible to realize, e. g. cross-sections with three or four particles in the initial state. Such quantities can so far also not be calculated using lattice calculations or functional methods. Hence, if chiral perturbation theory is sufficiently reliable, it is the method of choice for these problems. So far, derived observations indicate that it works out, but then these are so emergent results influenced by many factors that an independent confirmation would be invaluable.

5.3 Hadrons as field-theoretical bound states

5.3.1 Generalities

One of the most remarkable observations in QCD is that only bound states of quarks and gluons appear, for a reason called confinement to be discussed in more detail in section 5.5. The consequence is that only bound states are observable, and therefore hadron physics has to describe such bound states. Bound states are thus necessarily composite operators. The simplest such operator would be $\bar{q}q$, describing a scalar particle, the so-called σ or $\kappa(600)$ meson. However, it is not as simple.

One of the problems immediately arising is that in a quantum mechanical system mixing is possible. Thus, once a quantum number channel is specified by all conserved quantum numbers, any state $|n\rangle$ is in general represented by an operator O_n consisting of

an infinite series

$$O_n = \sum c_n^i O_n^i. \tag{5.44}$$

n is here the collection of quantum numbers uniquely specifying the states. In QCD, leaving aside the rest of the standard model, these are at most J^{PC}, flavor, total momentum and possible relative momenta between gauge-invariant constituents and relative spins as far as they can be measured. Including the remainder of the standard model, e. g. to model weak decays, modifies this list. However, irrespective what the list is, it specifies uniquely the channel. The sum then contains all operator O_n^i compatible with this list. This is an infinite number, and especially can contain operators made up of an infinite number of gluon, quark, and, possibly, ghost fields, as long as the total operator is gauge invariant. The coefficient c_n^i are in fact functions of, e. g., momenta.

Of course, since any set of operators forms a basis, it possible to rewrite the series. The question is only, in which way it should be rewritten. An especially useful way is to rewrite it in such a way as that operators of fixed quantum numbers and fixed energies are obtained, essentially obtaining for the infinite set of levels an infinite number of operators,

$$O_n(E) = \sum c_n^i O_n^i(E).$$

On the right-hand side two general types of operators will contribute. Some are genuine bound-state operators, i. e. operators which cannot be split into more gauge-invariant operators. An example is the aforementioned σ-operator. The other operators are states, for which this is possible. In case of the σ-meson, this would be, e. g., the two-pion operator $(\bar{q}\gamma_5 q)(p)(\bar{q}\gamma_5 q)(k)$, where the relative momenta p and k yield the total momentum $p + k = P$, such that $P_0 = E$, but are otherwise arbitrary. Therefore, in general the relative momenta should be integrated over. Note that a bound-state operator needs not be an operator with the minimum number of fields possible to create a set of quantum numbers. E. g. tetraquark operators for the σ-channel can have the form $\bar{q}\lambda^a q\bar{q}\lambda^a q$, and are such cases. Operators which can be split into a number of gauge-invariant operators can be classified into two categories. One is scattering states. These are states where the total energy is larger than the sum of the masses of the two individual operators in their respective quantum number channels. These are unbound objects. The other is so-called hadronic molecules, i. e. a self-bound system of two hadrons, similar to a conventional atom, where the binding yields a mass defect.

The energy levels appearing can be sorted into several possibilities. Every channel has a ground state, i. e. one lowest energy level E. This state can have zero energy, and may be degenerate. In QCD, such a state only exists in the chiral limit, where the pions becomes massless. Otherwise, all states are massive, and there is thus a mass gap in the spectrum.

The lowest energy can be a stable state, as is the case for pions, but need not be, as is the case in the σ-channel, where in QCD it turns out that two pions are the lightest state, and which are unbound. In general, this is a dynamical question, and the energy levels have to be calculated.

There are then an infinite number of further energy levels. There may be additional discrete levels below the first scattering state in the given channel, the so-called elastic threshold. These are states which are stable, excited states. Above the elastic threshold, there is a continuum of states. These consists of all possible relative momenta of the particles in the scattering states. In addition, there may be further surplus states which are resonances, i. e. states which have a finite life-time before decaying. Their distinction from scattering states is subtle, and will be discussed further below. If the energy is high enough such that there is more than one possible scattering state in the channel, the inelastic threshold is reached, and the number of possible states then quickly proliferates.

5.3.2 Analytical structure

No matter whether a particle is elementary or composite, as long as it is an observable particle, it is described by a correlation function, which is a function of a single four-momentum. Hence, an elementary particle q is described by the (1PI) correlation function

$$D(p) = \langle q^\dagger(-p)q(p)\rangle$$

while a composite particle $q^\dagger q$ is described by

$$D(p) = \langle (qq^\dagger)(-p)(q^\dagger q)(p)\rangle.$$

In any case, the correlation function D has a special analytical structure. If the correlation function is obtained from an operator set which includes all information, then D can have the following analytical structures

- There can be zero or more poles on the real axis at $-p^2 = m^2$. These signal stable particle excitations
- There can be zero or more poles on the second, but not first, Riemann sheet at $-p^2 = (M + i\Gamma)^2$, describing unstable resonances of mass M and decay width Γ
- There is a cut starting at $-p^2 = (M' + M'' + ...)^2$, where the M'... are the mass of the particles into which a decay at the elastic threshold is possible

E. g., for the proton there is a real pole at $m = 938$ MeV, a first resonance at $M = 1440$ MeV with $\Gamma \approx 300$ MeV and the elastic threshold for purely hadronic decays is at 1080 MeV for a decay in a p wave into a proton and a pion.

Besides the pole, however, the structure of the correlation function is rather complicated. If the width satisfies $\Gamma \ll M$, then the propagator is well approximated by a Breit-Wigner shape

$$D(p) \approx \frac{1}{p^2 - m^2 + i\sqrt{p^2}\Gamma}, \tag{5.45}$$

similar to the situation in quantum-mechanical scattering. Note, however, that the width is not really a constant, but in general a function of p^2. Widths quoted are usually the value of this function at the resonance's position, which in this case is well approximated by $\Gamma(M^2)$.

In cases of rather narrow resonances, most of the difference is due to phase space, and it is better to trade the width for the coupling of a particle to its decay channel by

$$\Gamma(p^2) = \frac{g^2}{6\pi}\sqrt{p^2}$$

In general, however, it is better to express the width by the phase shift δ as

$$\Gamma(p^2) = \frac{m^2 - p^2}{\sqrt{p^2}\cot\delta(p^2)}$$

The phase shift passes through $\pi/2$ at the resonance position.

However, these approximations are in general true only for narrow resonances. If the width $\Gamma(m^2)$ becomes large, the phase shift no longer necessarily passes through $\pi/2$, and the propagator, though still having a pole, may be very different from (5.45), including strong asymmetries. In a strongly interacting field theory like QCD, this is easily possible. E. g. for the σ-meson, to the best of our knowledge, the phase shift $|\delta|$ never exceeds $\pi/3$, and is actually negative. Such broad resonances require therefore substantial care. In fact, it is more a philosophical question whether such states should still be regarded as particles.

In any perturbative calculation, like in chiral perturbation theory, such an analytical structure can be immediately identified. Beyond analytical methods, this becomes more challenging.

The situation is more subtle for gauge-dependent degrees of freedom, like quarks and gluons, and this will therefore be reiterated in section 5.5.4

5.3.3 Bound states on the lattice

Lattice calculations lend themselves rather straightforwardly to the calculation of bound states, as operators can be directly evaluated. However, in any practical calculation only a finite number of operators can be selected, and therefore it can never be guaranteed that the overlaps c_n^i in (5.44) with the desired state is sufficiently large, or existent at all, to detect its contribution in a numerical simulation. Though several sophisticated methods exist to isolate levels, this remains a challenging task with substantial systematic errors. Furthermore, multi-particle states, e. g. nuclei with many quarks, are very expensive, as the number of operators required for a full gauge-invariant expression for a state of fermions increases factorial with the number of quark fields, and the statistical noise increases exponentially with the mass of a state compared to the lightest state.

However, especially ground-states, once a suitable operator basis is found, are relatively straightforward to extract. Since a lattice has a finite volume, the energy levels are no longer a continuum, but are discrete. Furthermore, scattering states with non-zero momentum have a kinetic energy of at least the inverse lattice spacing, making them especially on small lattices heavy. At the same time, it can be shown that stables states with non-zero mass suffer at most exponentially suppressed volume-corrections.

Finally, because lattice calculations are performed in Euclidean space-time, the following holds true. For any operator O in the Heisenberg picture

$$C(t) = \langle O^\dagger(t) O(0)\rangle = \sum_n \langle O(t)|n\rangle\langle n|O\rangle = \sum_n \langle e^{Ht} O(0) e^{-Ht}|n\rangle\langle n|O\rangle = \sum_n |\langle 0|O(0)|n\rangle|^2 e^{-Et}$$

where n is a complete set of energy eigenstates. Thus, at sufficiently long times only the state of lowest energy represented by the operator survives. Of course, on a finite and periodic lattice there is no infinite time, and this changes the result to

$$C(t) = \langle O^\dagger(t) O(0)\rangle = \sum_n |\langle 0|O(0)|n\rangle|^2 \cosh(Et)$$

but on sufficiently large lattices, this can still be fit reasonably well. Very useful are the effective masses

$$m(t) = -\ln \frac{C(t)}{C(t+a)},$$

which will become almost constant if only a single state dominates. However, in actual numerical calculations also subleading levels can often be fitted. These results did not make any statement about the three-momentum of the state. Usually, this is done at zero momentum. If not, then the energy is increased by the kinetic energy. The result also shows that the signal dies off exponentially with time, indicating serious statistical

challenges if the mass is large in units of the lattice spacing, as the situation deteriorates exponentially. Vice versa, it may happen that the lowest level is superimposed by some other level, which decays also very slowly. It can then happen that the lowest mass cannot be isolated for a finite lattice, or it may even be that the presence of a further state cannot be detected at all. This is another systematic error, which is especially hard to detect if no experimental input for a given quantum-number level are available, and thus a true prediction is attempted.

Though this result is obtained in Euclidean space-time, it can be shown that for any gauge-invariant correlation function the so-obtained energy is indeed the total energy in Minkowski space-time, and especially at zero kinetic energy its mass.

The situation is somewhat more challenging for massless particles, as e. g. for the pions in the chiral limit. The finite volume introduces a mass even for such particles. The reason is that only a constant mode on a periodic lattice will have zero energy, and the discreteness of the spectrum then forces any other state to a non-zero value. Since a pseudoscalar zero mode is forbidden, the vacuum is scalar, the pion is always massive in a finite volume. Its masslessness in the chiral limit is then only recovered in the infinite-volume limit. It can be shown that the effective mass of a massless particle is polynomial dependent on the volume. Hence, to identify a massless particle would require to identify, and reliably fit, the polynomial dependence of the mass.

The situation becomes more complex beyond stable states. First of all, as for massless states, finite-volume corrections to masses of excited states can be shown to only decay polynomially with the volume, and are therefore harder to control. The second is that they will mix with scattering states. In principle, the latter problem is straight-forward. Since the scattering states are just a combinatorial effect, it suffices to remove all known scattering states, and the remainder are then the resonances and excited states. In practice, it is often difficult to identify all levels for statistical reasons, and often overlap problems in the operator basis make states practically invisible in a calculation.

An alternative for states below the inelastic threshold is the explicit calculation of the phase shifts, which can be obtained by the so-called Lüscher method from the volume-dependence of the energy levels between the elastic and the inelastic thresholds. This procedure will not be detailed here, but is rather robust. It essentially uses the fact that the energy levels in a finite volume, ignoring discretization artifacts, can be related not only to single particle states, but also to multi-particle states with the same quantum numbers. In a finite volume, these multi-particle states are never really isolated, and therefore the energy levels are modified by the interaction energies. Thus, the energy levels are a statement about hadronic interactions. Since the size of the volume is then a

statement about the maximum distance of the two particles, this is all what is needed to obtain the phase shift in the infinite volume. In fact, this method works best on rather small volumes, where the energy levels can be clearly distinguished.

Since the phase shift also provides the information on the decay-width, the whole analytical structure can be obtained in theses cases, though of course with a finite number of points available from any numerical lattice calculations strictly speaking there is no mathematical guarantee to be not missing something. In cases like the σ-meson even this can still provide conceptual problems. The extension of the Lüscher method beyond the inelastic threshold is expensive, and for multiparticle channels still a topic of current research.

5.3.4 Bound states in functional methods

Bound states in functional calculations are entirely different than in lattice calculations. In principle, their treatment appears at first straight-forward. Since the operators describing bound states are just correlation functions, it appears sufficient to solve the corresponding equation in a given truncation, and the analytical structure will yield the states in this channel. Though this is correct in principle, and has been done is some rare model theories, it is usually not possible in a theory as complex as QCD. One of the problems is the requirement to solve the theory in the whole complex plane, which is rather challenging. Another is that mesonic bound states are already four-point functions, baryons six-point functions, and glueballs eight-point functions. Thus, the complexity even in simple truncations is enormous.

A trick to make this problem more tractable is to evaluate the equations on-shell, and only keep those terms becoming singular. This the so-called Bethe-Salpeter equation (or Faddeev equation for baryons). It takes the form

$$\Gamma_{\alpha\beta}(p,P) = \int dq K_{\alpha\gamma\delta\beta}(p,q,P)\left(S(q+\sigma P)\Gamma(q,P)S(q+(1-\sigma)P\right)_{\gamma\delta}.$$

The equation is again a self-consistency equation for the Bethe-Salpeter amplitude Γ, which depends on the total momentum p with $P^2 = -M^2$ and M the mass of the meson, and the relative momenta of the quarks p. The quark propagator S must be supplied from elsewhere, e. g. its own DSE or from an ansatz. The kernel K depends on the quantum numbers in question, and involves all the quark-gluon interaction vertices, as well as gluon propagators. Again, all of them must be provided from their respective DSEs or from ansätze. The momentum partition parameter σ distributes the relative momenta between the quarks. Since the quarks themselves are not independent degrees of freedom, the correct solution will be independent of σ.

The simplest ansatz is the so-called rainbow-ladder ansatz, where

$$K \sim \alpha(k^2)\Gamma^{\mathrm{tl}}\Gamma^{\mathrm{tl}},$$

where α is, despite its name, an effective interaction strength, and Γ^{tl} are the tree-level quark-gluon vertices. Selecting then, e. g., $\Gamma = i\gamma_5$ yields the equation for a pion, which in this approximation indeed becomes massless in the chiral limit. Solving this integral equation is still a numerically non-trivial task. The values M for which the equation has a solution indicates the hierarchy of states in the channel.

The general problem is that any approximation can produce spurious solutions. Furthermore, these results do not provide insights into these stability of states. These are questions of current research.

5.3.5 The hadron spectrum

5.3.5.1 Light hadrons

5.3.5.2 Charmonium and bottonium

5.3.5.3 Exotics

5.4 Topological excitations

While quarks, gluons and hadrons are states which consists out of a few valence particles, i. e. particles which define their quantum numbers, QCD can also support other states. Especially, it can support states which are genuine collective excitations of gluons, which can no longer be identified individually. Such states are usually highly unstable. However, some such excitations are stabilized by geometric effects, and then called topological excitations, for reasons to be explored in the following.

The most prominent of these excitations are so-called instantons. Though it is not possible to directly observe instantons, their effects can be identified, e. g., in lattice calculations through various techniques. To understand them, it is best to first investigate similar excitations in quantum mechanics.

5.4.1 Instantons in quantum mechanics

5.4.1.1 Leading order

The starting point is a one-dimensional double-well potential

$$V(x) = \lambda\left(x^2 - \eta^2\right)^2$$

Classically, This potential has two degenerate ground states with energy $\pm\eta$. These are two classically separated vacua for particles with kinetic energies less than $\lambda\eta^4$. The quantization can be done using the WKB approximation. Then the wave function is assumed as

$$\psi(x) = e^{i\Phi(x)}$$

where

$$\begin{aligned} \Phi(x) &= \pm \int^x dx' p(x') + O(\hbar) \\ p(x) &= \sqrt{2m}\sqrt{E - V(x)} \end{aligned} \tag{5.46}$$

where m is the mass of the particle and E is its energy. The WKB approximation results from an expansion in $\hbar$, here only taken to lowest order. If quantization is done around both minima, two degenerate solutions are found. However, this is not quantum mechanically true, and tunneling will mix both states. The new ground state will be the symmetric combination of both single states, and the first excited state will be the anti-symmetric combination. During tunneling, the WKB solutions do not oscillate, but decay exponentially.

An analogous expansion of the quantum-mechanical path integral around the classical path can be done. The imaginary values of p of the WKB approximation can equivalently be replaced by imaginary time, a so-called Wick rotation. The Green's function describing the tunneling is then given by

$$\langle -\eta | e^{-iHt} | \eta \rangle = \int \mathcal{D}r e^{-S},$$

where r is the path and S the action. The path associated with the solution of the classical equations of motion in imaginary time gives then the maximum contribution to the path integral.

The path integral can be expanded around this path as a correction factor. After the Wick rotation is performed, the classical equation of motion for the tunneling path is given by

$$m\frac{d^2x}{d\tau^2} = \frac{dV}{dx}$$

or equivalently by the conservation of energy for the lowest classical energy state, where $E = 0$,

$$\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2 - V(x) = 0$$

yielding after separation of variables

$$d\tau = \frac{1}{\sqrt{\frac{2\lambda}{m}\left(x^2-\eta^2\right)^2}}dx = \frac{1}{\eta^2\sqrt{\frac{2\lambda}{m}}}\frac{1}{1-\frac{x^2}{\eta^2}}dx = \frac{1}{\eta\sqrt{\frac{2\lambda}{m}}}\frac{1}{1-u^2}du.$$

This can be integrated and yields

$$\tau - \tau_0 = \frac{1}{\eta\sqrt{\frac{2\lambda}{m}}}\tanh^{-1}\frac{x}{\eta}$$

where both integration constants have been absorbed in τ_0. This yields for the classical path

$$x_{cl}(\tau) = \eta\tanh\left(\frac{\omega}{2}(\tau-\tau_0)\right) \tag{5.47}$$

$$\omega^2 = \frac{8\eta^2\lambda}{m} \tag{5.48}$$

The solution goes from $x_{cl}(-\infty) = -\eta$ to $x_{cl}(\infty) = \eta$, and hence describes the tunneling of the lowest classical energy state from one classical allowed position to the other. τ_0 is a free parameter and describes the localization of the tunneling in imaginary time. The solution is well localized in time. The path (5.47) is referred to as an instanton. Its reverse $-x_{cl}(\tau)$ is referred to as an anti-instanton.

The action of this path is

$$S_0 = \int_{-\infty}^{\infty} d\tau\left(\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2 + V(x)\right)$$

Rescaling $\frac{\omega}{2}(\tau-\tau_0)\to\tau$ this yields

$$\begin{aligned} S_0 &= \int_{-\infty}^{\infty} d\tau\left(\frac{m\eta^2\omega}{4}\frac{1}{\cosh^4\tau} + \frac{2\lambda\eta^4}{\omega}\left(\tanh^2\tau-1\right)^2\right) \\ &= \frac{m\eta^2\omega}{3} + \frac{8\lambda\eta^4}{3\omega} = \frac{m\eta^2\omega}{3} + \frac{m\omega\eta^2}{3} = \frac{2m\omega\eta^2}{3} \\ &= \frac{2m\omega}{3}\frac{m\omega^2}{8\lambda} = \frac{m^2\omega^3}{12\lambda} \end{aligned}$$

It is convenient to scale such, that $m=\omega=1$ and to change the potential to $V(x) = \lambda x^2(x+2\eta)^2$. The action is then $\frac{1}{12\lambda}$ and the condition $\lambda \ll 1$ controls the validity of the semi-classical expansion.

To calculate the tunneling amplitude, the exponent of the path integral is expanded around the classical path

$$\left\langle 0|e^{-H\tau}|2\eta\right\rangle = e^{-S_0}\int D[\delta r]\, e^{-\frac{1}{2}\delta x\frac{\delta^2 S}{\delta x^2}\Big|_{x_{cl}}\delta x + O\left(\delta x^3\right)} \tag{5.49}$$

Note that the linear term vanishes, since the classical path already minimizes the action. It is implicitly assumed, that τ is smaller than the typical lifetime of an instanton, otherwise modes with more than one instanton would have to be taken into account. It is also assumed that τ is nevertheless large enough to allow the semi-classical WKB approximation. The tunneling is hence proportional to $\exp(-S_0)$. The proportionality constant requires the calculation of higher orders.

5.4.1.2 Higher-order corrections

To take fluctuations into account, it is necessary to calculate the second factor up to order $O(\delta x^2)$ in (5.49). The functional derivative of the action gives the operator

$$O = -\frac{d^2}{d\tau^2} + \left.\frac{d^2V}{dx^2}\right|_{x=x_{cl}}$$

The path integral is Gaussian in δx. Expanding the path integral in an integral about a base s with $s \to \infty$, and the operator O in an arbitrary basis, the prefactor f for $\exp(-S_0)$, using a functional integration theorem for small fluctuations, becomes

$$\prod_s \left(\int dx_s\right) e^{-\frac{1}{2}\sum\limits_{i,j} x_i O_{ij} x_j} = (2\pi)^{\frac{n}{2}} \frac{1}{\sqrt{\det O}}$$

The determinant can be calculated by finding the eigenvalues of O. This task is rather typical for calculations involving topological excitations, and warrants therefore a rather detailed investigation.

Since determinants are products of eigenvalues, this amounts to solving the eigenvalue equation

$$\begin{aligned}
Ox_n(\tau) &= \varepsilon_n x_n(\tau) \\
\left(-\frac{d^2}{d\tau^2} + \lambda\frac{d}{dx}\left(4x\left(x^2-\eta^2\right)\right)\Big|_{x=x_{cl}}\right) x_n(\tau) &= \varepsilon_n x_n(\tau) \\
\left(-\frac{d^2}{d\tau^2} + \lambda\left(12x^2-4\eta^2\right)\big|_{x=x_{cl}}\right) x_n(\tau) &= \varepsilon_n x_n(\tau) \\
\left(-\frac{d^2}{d\tau^2} + 12\lambda\eta^2\tanh^2\left(\frac{\omega}{2}(\tau-\tau_0)\right) - 4\lambda\eta^2\right) x_n(\tau) &= \varepsilon_n x_n(\tau) \\
\left(-\frac{d^2}{d\tau^2} + \frac{3}{2}\omega^2\tanh^2\left(\frac{\omega}{2}(\tau-\tau_0)\right) - \frac{1}{2}\omega^2\right) x_n(\tau) &= \varepsilon_n x_n(\tau)\,.
\end{aligned}$$

Using $1-\frac{1}{\cosh^2 x}=\tanh^2 x$ and resetting $\tau-\tau_0\to\tau$ and $x_n(\tau+\tau_0)\to x_n(\tau)$ yields

$$\begin{aligned}\left(-\frac{d^2}{d\tau^2}+\omega^2-\frac{3\omega^2}{2\cosh^2\left(\frac{\omega\tau}{2}\right)}\right)x_n(\tau) &= \varepsilon_n x_n(\tau)\\ \left(-\frac{d^2}{d\tau^2}+\omega^2\left(1-\frac{3}{2\cosh^2\left(\frac{\omega\tau}{2}\right)}\right)\right)x_n(\tau) &= \varepsilon_n x_n(\tau)\end{aligned}$$

The solutions to this Schrödinger-type equation are known. They can be found by the following procedure

$$\begin{aligned}0 &= \left(-\frac{d^2}{d\tau^2}-2\frac{\frac{3\omega^2}{4}}{\cosh^2\left(\frac{\omega\tau}{2}\right)}-2\frac{1}{2}\left(\varepsilon_n-\omega^2\right)\right)x_n\\ &= \left(\frac{d^2}{d\tau^2}+\frac{3\omega^2}{2\cosh^2\left(\frac{\omega\tau}{2}\right)}+E_n\right)x_n\\ &= \frac{d}{dy}\left(\left(1-y^2\right)\frac{dx_n}{dy}\right)+\left(s(s+1)-\frac{E_n}{1-y^2}\right)y\end{aligned}$$

where $e_n=\frac{1}{2}(\varepsilon_n-\omega^2)$, $E_n=\frac{2}{\omega}\sqrt{-2e_n}$, and $y=\tanh\omega\tau/2$. Making then the ansatz

$$x_n=\left(1-y^2\right)^{\frac{E_n}{2}}w(y)$$

yields

$$0=u(1-u)\frac{d^2w}{dy^2}+(E_n+1)(1-2u)\frac{dw}{dy}-(E_n-s)(E_n+s+1)$$

with $u=(1-y)/2$.

This is the differential equation for a hypergeometric function and the solutions are

$$x_n=(1-y)^{\frac{E_n}{2}}F\left(E_n-s,E_n+s+1,E_n+1,\frac{1-y}{2}\right)$$

For a finite wave function it is required that $s=E_n+n$ with $n<2$.

Hence only two bound states arise, 0 and 1. The original energy is then

$$\begin{aligned}\varepsilon_n &= \omega^2\left(-\frac{1}{4}(s-n)^2+1\right)\\ \varepsilon_0 &= 0\\ \varepsilon_1 &= \frac{3}{4}\omega^2\end{aligned}$$

There is in addition a scattering continuum. The normalized eigenfunction for ε_0 is

$$x_0(\tau)=\sqrt{\frac{3\omega}{8}}\frac{1}{\cosh^2\frac{\omega\tau}{2}}$$

which can be verified by insertion.

The presence of a zero-mode, i. e. $\varepsilon_0 = 0$, results in a vanishing determinant and hence in an infinite tunneling amplitude. This can only be interpreted in such a way that the fluctuations in direction of the zero mode are not small, and the expansion in terms of a Gaussian is invalid. The integration in the direction of the zero mode path must hence be done exactly. To do this, the paths are parametrized by

$$x(\tau) = \sum_n c_n x_n(\tau)$$

Noting further that

$$\begin{aligned}\frac{d}{d\tau_0}x_{cl}(\tau-\tau_0) &= -\frac{\omega\eta}{2}\frac{1}{\cosh^2\left(\frac{\omega}{2}(\tau-\tau_0)\right)} = -\frac{\omega\eta}{2}\sqrt{\frac{8\omega}{3}}x_0(\tau)\\ &= -\sqrt{\frac{8\omega^3\eta^2}{12}}x_0(\tau) = -\omega\sqrt{\frac{\omega^3}{12\lambda}}x_0(\tau) = -\omega\sqrt{S_0}x_0(\tau)\end{aligned}$$

where again $\tau - \tau_0$ was taken into τ. By replacing the integration over the expansion part of c_0 by

$$dx = \frac{dx}{d\tau_0}d\tau_0$$

and using that the expansion takes only into account fluctuations around the classical path

$$\begin{aligned}dx &= \frac{dx}{d\tau_0}d\tau_0 \approx \frac{dx_{cl}}{d\tau_0}d\tau_0 = -\omega\sqrt{S_0}x_0(\tau)\,d\tau_0\\ \frac{dx}{x_0} &= -\omega\sqrt{S_0}d\tau_0\end{aligned}$$

simplifies the problem further. Finally, using

$$dx = x_0 dc_0 \Longrightarrow dc_0 = -\omega\sqrt{S_0}d\tau_0$$

the functional integral can now be rewritten from integrals about different space points to an integral over all possible combinations of c_i and hence

$$\prod_s \int dx_s = \prod_n \int dc_n$$

each coefficient contributes its eigenvalue to the operator. This method where the continuous variable is replaced by a product over eigenfunction expansion coefficient will also be a recurring procedure later on.

Hence the result for the propagator to this order is

$$-\left(\prod_{n>0}\frac{2\pi}{\varepsilon_n}\right)^{\frac{1}{2}}\omega\sqrt{\frac{S_0}{2\pi}}\int d\tau_0$$

The first factor is the determinant with the zero mode excluded. The minus sign is of no importance, since the tunneling probability is given by the square of the amplitude. The result increases linearly with time, the size of the instanton, leading to a finite transition probability per unit time.

To calculate the remaining determinant, it is convenient to restrict the calculation to a finite time interval $[-\tau_m/2, \tau_m/2]$ and impose boundary conditions at the edge of the interval as $x_n(\pm\tau_m/2) = 0$. Since there is a continuum of eigenvalues, the product diverges to 0. This divergence is controlled by the largest eigenvalues, which are independent of the specific shape of the potential and do not correspond to tunneling events. To factor these continuum states independent of the potential out, the determinant can be normalized by the harmonic oscillator, which exhibits the same shape at large eigenvalues. This is again a routine manipulation often used, and yields

$$\left(\frac{\det\left(-\frac{d^2}{d\tau^2}+V''(x_{cl})\right)}{\det\left(-\frac{d^2}{d\tau^2}+\omega^2\right)}\right)^{-\frac{1}{2}} = -\omega\sqrt{\frac{S_0}{2\pi}}\left(\int d\tau_0\right)\left(\frac{\det'\left(-\frac{d^2}{d\tau^2}+V''(x_{cl})\right)}{\frac{1}{\omega^2}\det'\left(-\frac{d^2}{d\tau^2}+\omega^2\right)}\right)^{-\frac{1}{2}}$$

where $\det'$ denotes the determinant without the lowest mode. The lowest eigenvalue of the harmonic oscillator equation is given simply by $\omega^2/2$. The next eigenvalue of the upper determinant is $3\omega^2/4$. For the harmonic oscillator the next value is ω^2 up to corrections of order $1/\tau_m$, which are not important for the limit of $\tau_m \to \infty$. For these eigenvalues of the oscillator the boundary conditions are important. Both exhibit furthermore a continuous spectrum. The solutions to the $V''(x_{cl})$ potential are known. The potential itself is localized, so the solutions are plane waves for $\tau \to \pm\infty$. One solution is hence $x_p(\tau) \sim \exp(ip\tau)$ for the limit of $\tau \to \infty$. For $\tau \to -\infty$ the wave becomes

$$\begin{aligned} x_p(\tau) &\sim e^{ip\tau+i\delta_p} \\ e^{i\delta_p} &= \frac{1+i\frac{p}{\omega}}{1-i\frac{p}{\omega}}\frac{1+2i\frac{p}{\omega}}{1-2i\frac{p}{\omega}} \end{aligned} \tag{5.50}$$

where the phase shift is provided by the potential. The phase shift is known from scattering theory. There is no reflection for the plane wave. The second independent solution is obtained when reversing $\tau \to -\tau$. The spectrum of eigenfunctions is then obtained by imposing the boundary conditions

$$p_n\tau_m - \delta_{pn} = \pi n$$

The solutions of this equation will be denoted by p'. In case of the harmonic oscillator there is no phase shift, since the second derivative is constant and the boundary conditions impose

$$p_n\tau_m = \pi n \tag{5.51}$$

The ratio of the determinants is then given by

$$\prod_{n>1}\frac{\omega^2+p_n'^2}{\omega^2+p_n^2}=e^{\sum\limits_n \ln\frac{\omega^2+p_n'^2}{\omega^2+p_n^2}}$$

Changing from summation over n to integration over p, which delivers by (5.51) a factor τ_m/π, and setting $\delta_p/\tau_m = p_n' - p_n$, which is small, it is possible to obtain by expanding

$$e^{\frac{\tau_m}{\pi}\int_0^\infty dp \ln\frac{\omega^2+p^2+2p\frac{\delta_p}{\tau_m}}{\omega^2+p^2}}=e^{\frac{\tau_m}{\pi}\int_0^\infty dp\frac{2p\delta_p}{\tau_m(\omega^2+p^2)}}$$

δ_p depends also on p. Integrating by parts delivers

$$\begin{aligned}\int\limits_0^\infty dp\frac{2p\delta_p}{(\omega^2+p^2)} &= \delta_p\ln\left(\omega^2+p^2\right)\big|_0^\infty-\int\limits_0^\infty dp\ln\left(\omega^2+p^2\right)\frac{d\delta_p}{dp}\\ &= \delta_p\ln\left(\omega^2+p^2\right)\big|_0^\infty-\int\limits_0^\infty dp\ln\left(\omega^2+p^2\right)\frac{6\omega^3+12\omega p^2}{\omega^4+5\omega^2p^2+4p^4}\end{aligned}$$

This leads to

$$\frac{1}{\pi}\int\limits_0^\infty dp\frac{2p\delta_p}{\tau_m\left(\omega^2+p^2\right)}=-\frac{1}{6}\left(27\tanh^{-1}\frac{1}{2}-6+2{}_2F_1\left(\frac{1}{2},2,\frac{5}{2},\frac{1}{4}\right)+6\ln 3+24\ln\omega\right)=C\left(\omega\right)$$

Where ${}_2F_1$ is again a hypergeometric function. Additionally the first term vanishes by virtue of (5.50). This value will be referred to as $C\left(\omega\right)$. The value of $\exp\left(C\left(\omega\right)\right)$ is for $\omega=1$ just $1/9$. The final tunneling amplitude is then

$$\begin{aligned}\left\langle-\eta|e^{-H\tau_m}|+\eta\right\rangle &= {}_O\left\langle-\eta|e^{-H_O\tau_m}|\eta\right\rangle_O\left(\sqrt{\frac{S_0}{2\pi}}\sqrt{\frac{4}{3}}\sqrt{\frac{1}{e^{C(\omega)}}}\right)e^{-S_0}\left(\omega\tau_m\right)\\ &= \left(\left\langle0,-\eta|x\right\rangle\left\langle x|0,\eta\right\rangle e^{-\frac{\omega\tau_m}{2}}\right)\left(\sqrt{\frac{S_0}{2\pi}}\sqrt{\frac{4}{3}}\sqrt{\frac{1}{e^{C(\omega)}}}\right)e^{-S_0}\left(\omega\tau_m\right)\\ &= \left(\phi_{-\eta}\left(-\eta\right)\phi_\eta\left(\eta\right)e^{-\frac{\omega\tau_m}{2}}\right)\left(\sqrt{\frac{S_0}{2\pi}}\sqrt{\frac{4}{3}}\sqrt{\frac{1}{e^{C(\omega)}}}\right)e^{-S_0}\left(\omega\tau_m\right)\\ &= -\left(\sqrt{\frac{\omega}{\pi}}e^{-\frac{\omega\tau_m}{2}}\right)\left(\sqrt{\frac{2S_0}{3\pi}}\sqrt{\frac{1}{e^{C(\omega)}}}\right)e^{-S_0}\left(\omega\tau_m\right)\end{aligned}\tag{5.52}$$

where ϕ_a is the ground state wave function localized at a. The last factor comes from the zero mode integration, the factor in front of it from the instanton solution. The first factor comes from the lowest state harmonic oscillator and the second is the ratio of the

two determinants. The result is proportional to the time interval τ_m and hence there is a finite tunneling probability per unit time. The harmonic oscillator part comes from the normalization.

From (5.52) the ground state splitting ΔE can be read off, by using the following expansion. The two ground states can be approximated by $\psi_0 = \left(\phi_{-\eta}(x) + \phi_\eta(x)\right)/\sqrt{2}$ and $\psi_1 = \left(\phi_{-\eta}(x) - \phi_\eta(x)\right)/\sqrt{2}$. Then the tunneling amplitude is approximately

$$\begin{aligned}\left\langle -\eta|e^{-H\tau_m}|\eta\right\rangle &\approx \psi_0^*(-\eta)\,\psi_0(\eta)\,e^{-E_0\tau_m} + \psi_1^*(-\eta)\,\psi_1(\eta)\,e^{-(E_0+\Delta E)\tau} \\ &= \frac{1}{2}(\phi_{-\eta}^*(-\eta)\,\phi_{-\eta}(\eta) + \phi_{-\eta}^*(-\eta)\,\phi_\eta(\eta) + \phi_\eta^*(-\eta)\,\phi_{-\eta}(\eta) + \phi_\eta^*(-\eta)\,\phi_\eta(\eta)) \\ &\quad + \frac{1}{2}(\phi_{-\eta}^*(-\eta)\,\phi_{-\eta}(\eta) - \phi_{-\eta}^*(-\eta)\,\phi_\eta(\eta) - \phi_\eta^*(-\eta)\,\phi_{-\eta}(\eta) + \phi_\eta^*(-\eta)\,\phi_\eta(\eta))e^{-\Delta E\tau_m}\end{aligned}$$

If $\Delta E\tau_m$ is small, the result is then

$$\left\langle -\eta|e^{-H\tau_m}|\eta\right\rangle \approx \frac{1}{2}\phi_{-\eta}^*(-\eta)\,\phi_\eta(\eta)\,(\Delta E\tau_m) \tag{5.53}$$

here uneven terms in η and $-\eta$ in argument and index can be neglected, since they are small compared to the even terms. The first part is the harmonic oscillator overlap, and the level splitting is hence approximately

$$\Delta E \approx \omega\left(\sqrt{\frac{2S_0}{3\pi}}\sqrt{\frac{1}{e^{C(\omega)}}}\right)e^{-S_0}.$$

If the assumptions break down, multiple instantons and their interactions become important.

5.4.1.3 Corrections at two loop order

The procedure described can be continued to higher orders. The mathematics involved get then more and more tedious, so only some remarkable results from two loop order will be cited here. The loop correction is essentially achieved by calculating the correlation function in presence of one instanton. The modifications lead to

$$\exp(-S_0) \to \exp\left(-S_0 - \frac{71}{72}\frac{1}{S_0}\right)$$

Since the correction is negative and of order one implies, that for $S_0 \approx 1$, the one loop approximation breaks down. Since $S_0 \sim \frac{\omega^3}{\lambda}$, this is already the case for moderately high barriers. The minus sign implies that the one loop order overestimates the tunneling probability.

5.4.1.4 Instanton-Anti-Instanton interactions

As already mentioned, anti-instantons are instanton paths traversed in reversed time direction. A pair of instanton and anti-instanton relates over a topological trivial path (the same path traversed once in each direction), the same vacua. This interaction leads to the tunneling amplitude

$$\langle \eta | e^{-H\tau_m} | \eta \rangle = \frac{\tau_m}{2\pi\lambda} \int d\tau e^{S_{IA}(\tau)}$$

The prefactor comes from the instanton density (there must be two in the time interval τ_m). The instanton-anti-instanton action S_{IA} can be calculated from a path as

$$S_{IA} = \frac{1}{2\lambda} \left(\frac{1}{3} - 2e^{|\tau_I - \tau_A|} + O\left(e^{2|\tau_I - \tau_A|}\right) \right)$$

following the classical tunneling path in both directions. The parameter τ_0 may be different for instanton τ_I and anti-instanton τ_A. If $\tau_I - \tau_A \gg 1$, S_{IA} tends to $2S_0$ while for $\tau_I - \tau_A \approx 0$ it tends to 0. In the latter case however the instanton solution is not a good approximation, since the relevant time scales do not allow anymore a classical treatment. There are methods using steepest ascent to solve this problem numerical. By analytical continuation in the coupling constant $\sqrt{2\lambda}$,and after redefining the perturbative expansion, an analytical calculation is possible for $(E_0 + E_1)/2$. This can be compared to the single instanton solution above, where ΔE can be calculated.

5.4.1.5 Fermions

If light fermions are introduced and coupled by V'' to the potential, there are two interesting effects visible: First, tunneling is only possible, if the fermion number changes during the tunneling. Second, the fermions introduce long range attractive forces between instantons and anti-instantons leading to formation of instanton-anti-instanton states.

5.4.2 Yang-Mills theory

5.4.2.1 Topology

It is convenient for the treatment of Yang-Mills theories to use the following definitions, normalizations and gauges:

$$\begin{aligned} A_i &= A_i^a \frac{\lambda^a}{2} \\ \left[\lambda^a, \lambda^b\right] &= 2if^{abc}\lambda^c \\ \mathrm{tr}\left(\lambda^a\lambda^b\right) &= 2\delta^{ab} \\ G_{\mu\nu}^a &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + f^{abc}A_\mu^b A_\nu^c \\ A_0 &= 0 \Longrightarrow E_i = \partial_0 A_i \end{aligned}$$

where E_i are the chromoelectric fields, which are for this gauge the conjugate momenta of the A_i. The Hamiltonian is then given by

$$H = \frac{1}{2g^2}\int d^3\vec{x}\left(E_i^2 + B_i^2\right)$$

The classical vacua have zero field strengths, but the potential A does not need to be constant for non-Abelian theories. The gauge fields are in this case limited to pure gauge configurations, which can be enumerated as

$$A_i = iU\left(\vec{x}\right)\partial_i U\left(\vec{x}\right)^\dagger$$

where $U\left(\vec{x}\right)$ can be any possible gauge transformation. It is possible to restrict these to those satisfying $U\left(\vec{x}\right) = 1$ for $x \to \infty$. These can be classified by the winding number

$$n_w = \frac{1}{24\pi^2}\int d^3\vec{x}\varepsilon^{ijk}\mathrm{tr}\left(\left(U^\dagger\partial_i U\right)\left(U^\dagger\partial_j U\right)\left(U^\dagger\partial_k U\right)\right) \tag{5.54}$$

which counts the number of times the group manifold of the gauge transformation is covered. If (5.54) is expressed in terms of the gauge fields, it is called the Chern-Simmons characteristic

$$n_{CS} = \frac{1}{16\pi^2}\int d^3\vec{x}\varepsilon^{ijk}\left(A_i^a\partial_j A_k^a + \frac{1}{3}f^{abc}A_i^a A_j^b A_k^c\right) \tag{5.55}$$

The quantity (5.55) can also be rewritten more generally as

$$\frac{1}{64\pi^2}\int d^4x\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}^a F_{a\rho\sigma} = -\frac{i}{512\pi^4}\int d^4x\mathrm{tr}\epsilon^{\mu\nu\rho\sigma}\partial_\mu\left(iA_\nu^a\partial_\rho A_\sigma^a + \frac{2}{3}f^{abc}A_\nu^a A_\rho^b A_\sigma^c\right)$$

Evidently, this is a total derivative, and hence can be cast into a surface integral at infinity. It is therefore independent of the internal structure of the space-time it is integrated over,

but depends only on the contribution from the boundary. Furthermore, the expression has the same color structure as the usual Lagrangian, and the Lorentz indices do not play a role in gauge transformations of the field-strength tensor. Hence, this quantity is gauge-invariant. Thus, it is an observable quantity. It is the so-called topological charge, or Chern class, of the gauge field configuration. Furthermore, the quantity is evidently invariant under any continuous distortions of the gauge fields inside the volume. It is less obvious that this is true for any continuous deformations of the gauge fields on the boundary, and that all of these possible deformations fall into distinct classes, the so-called Chern classes, such that the integral is an integer k, characterizing this class. This fact is stated here without proof.

Hence n_w is an integer, and enumerates an infinite set of classical vacua. Since they are topological different, there is no path from one vacuum to the other where the energy remains zero all the way. But tunneling can connect topological different vacuum states.

5.4.2.2 Tunneling

The first question is, are there really any tunneling events? The answer is yes. Again classical solutions to the euclidean action will provide access. It is convenient to rewrite the action as

$$\begin{aligned} S &= \frac{1}{4g^2}\int d^4x G^a_{\mu\nu}G^a_{\mu\nu} \\ &= \frac{1}{4g^2}\int d^4x \left(\pm G^a_{\mu\nu}\Gamma^a_{\mu\nu} + \frac{1}{2}\left(G^a_{\mu\nu} \mp \Gamma^a_{\mu\nu}\right)^2\right) \end{aligned}$$

where $\Gamma^a_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}G^a_{\rho\sigma}$ is the dual field strength tensor. In this tensor the roles of the magnetic and electric field are interchanged. The first term is invariant under topological changes, see below, the second is always positive. If the field is (anti) self-dual, i.e. $G^a_{\mu\nu} = \pm\Gamma^a_{\mu\nu}$, the action is hence minimal. The first term is the topological charge

$$Q = \frac{1}{32\pi^2}\int d^4x G^a_{\mu\nu}\Gamma^a_{\mu\nu}$$

and the invariance can be seen from the equivalence to n_{CS}.

$$\begin{aligned} Q &= \frac{1}{32\pi^2}\int d^4x \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}\left(\partial_\nu A^a_\mu - \partial_\mu A^a_\nu - f^{abc}A^b_\mu A^c_\nu\right)\left(\partial_\sigma A^a_\rho - \partial_\rho A^a_\sigma - f^{ade}A^d_\rho A^e_\sigma\right) \\ &= \frac{1}{32\pi^2}\int d^4x \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}(\partial_\nu A^a_\mu\partial_\sigma A^a_\rho - \partial_\nu A^a_\mu\partial_\rho A^a_\sigma + \partial_\nu A^a_\mu f^{ade}A^d_\rho A^e_\sigma - \partial_\mu A^a_\nu\partial_\sigma A^a_\rho \\ &\quad +\partial_\mu A^a_\nu\partial_\rho A^a_\sigma + \partial_\mu A^a_\nu f^{ade}A^d_\rho A^e_\sigma - f^{abc}A^b_\mu A^c_\nu\partial_\sigma A^a_\rho + f^{abc}A^b_\mu A^c_\nu\partial_\rho A^a_\sigma \\ &\quad +f^{abc}A^b_\mu A^c_\nu f^{ade}A^d_\rho A^e_\sigma) \end{aligned}$$

First have a look at the double derivatives

$$\frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}\left(\partial_\nu A^a_\mu \partial_\sigma A^a_\rho - \partial_\nu A^a_\mu \partial_\rho A^a_\sigma - \partial_\mu A^a_\nu \partial_\sigma A^a_\rho + \partial_\mu A^a_\nu \partial_\rho A^a_\sigma\right)$$

By partial integration this results in

$$\begin{aligned}&\frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}(\partial_\nu\left(A^a_\mu \partial_\sigma A^a_\rho\right) - A^a_\mu \partial_\nu \partial_\sigma A^a_\rho - \partial_\nu\left(A^a_\mu \partial_\rho A^a_\sigma\right) + A^a_\mu \partial_\nu \partial_\rho A^a_\sigma\\ &-\partial_\mu\left(A^a_\nu \partial_\sigma A^a_\rho\right) + A^a_\nu \partial_\mu \partial_\sigma A^a_\rho + \partial_\mu\left(A^a_\nu \partial_\rho A^a_\sigma\right) - A^a_\nu \partial_\mu \partial_\rho A^a_\sigma)\end{aligned}$$

The total derivatives can be connected by

$$\varepsilon_{\nu\mu\sigma\rho}\partial_\nu\left(A^a_\mu \partial_\sigma A^a_\rho\right)+\varepsilon_{\nu\mu\rho\sigma}\partial_\nu\left(A^a_\mu \partial_\rho A^a_\sigma\right)+\varepsilon_{\mu\nu\sigma\rho}\partial_\mu\left(A^a_\nu \partial_\sigma A^a_\rho\right)+\varepsilon_{\mu\nu\rho\sigma}\partial_\mu\left(A^a_\nu \partial_\rho A^a_\sigma\right) = 4\varepsilon_{\mu\nu\rho\sigma}\partial_\mu\left(A^a_\nu \partial_\rho A^a_\sigma\right)$$

The other terms result in

$$4\varepsilon_{\mu\nu\rho\sigma}A^a_\mu \partial_\nu \partial_\rho A^a_\sigma = 0$$

since it is symmetric about the inner two indices. For the terms with one power of f^{abc} the result is

$$\begin{aligned}&-\varepsilon_{\mu\nu\rho\sigma}\left(\partial_\nu A^a_\mu A^b_\rho A^c_\sigma - \partial_\mu A^a_\nu A^b_\rho A^c_\sigma + A^b_\mu A^c_\nu \partial_\sigma A^a_\rho - A^b_\mu A^c_\nu \partial_\rho A^a_\sigma\right)\\ = \;&-\varepsilon_{\mu\nu\rho\sigma}(\partial_\nu\left(A^a_\mu A^b_\rho A^c_\sigma\right) - A^a_\mu A^c_\sigma \partial_\nu A^b_\rho - A^a_\mu A^b_\rho \partial_\nu A^c_\sigma - \partial_\mu\left(A^a_\nu A^b_\rho A^c_\sigma\right) + A^a_\nu A^b_\rho \partial_\mu A^c_\sigma\\ &+A^a_\nu A^c_\sigma \partial_\mu A^b_\rho + \partial_\sigma\left(A^a_\rho A^b_\mu A^c_\nu\right) - A^a_\rho A^b_\mu \partial_\sigma A^c_\nu - A^a_\rho A^c_\nu \partial_\sigma A^b_\mu - \partial_\rho\left(A^a_\sigma A^b_\mu A^c_\nu\right)\\ &+A^a_\sigma A^b_\mu \partial_\rho A^c_\nu + A^a_\sigma A^c_\nu \partial_\rho A^c_\nu)\end{aligned}$$

The not-total-derivative terms are again symmetric in two indices and hence vanish. The total derivatives are symmetric about three indices and hence do not vanish in all cases, but only in two out of three, and hence the result is 4/3. The total result is then

$$\frac{1}{16\pi^2}\int d^4x \partial_\mu \varepsilon_{\mu\nu\rho\sigma}\left(A^a_\nu \partial_\rho A^a_\sigma + \frac{1}{3}f^{abc}A^a_\nu A^b_\rho A^c_\sigma\right)$$

The last term vanished, because the square of the potential can not provide a contribution over the whole space because otherwise the fields would not vanish. The remaining last term is a result of the commutator of the generators and the factors of the generators, the gauge fields.

Mathematical, since A_μ is a pure gauge, $A_\mu A_\nu$ can be written as $U\partial_\mu U^\dagger U \partial_\nu U^\dagger$ and this can be changed by integration by parts to $\partial_\mu U U^\dagger \partial_\nu U U^\dagger$, which is by unitarity 0. The rest can be combined in a total derivative

$$\begin{aligned}K_\mu &= \frac{1}{16\pi^2}\varepsilon_{\mu\nu\rho\sigma}\left(A^a_\nu \partial_\rho A^a_\sigma + \frac{1}{3}f^{abc}A^a_\nu A^b_\rho A^c_\sigma\right)\\ Q &= \int d^4x \partial_\mu K_\mu = \int d^3\sigma_\mu K_\mu\end{aligned}$$

where in the last step the volume integration was replaced by a surface integration and the Gaussian theorem was used. Since the integration is equivalent to (5.55), the topological invariance has been shown. With this knowledge, the action can now be calculated for an arbitrary (anti-)self-dual field as

$$S = \frac{8\pi^2 |Q|}{g^2}$$

By

$$Q = \int_{-\infty}^{\infty} dt \frac{d}{dt} \int d^3\vec{x} K_0 = n_{CS}(t=\infty) - n_{CS}(t=-\infty)$$

it is visible, that configurations with $Q \neq 0$ connect different vacua. The simplest non-trivial solution can be obtained for SU(2) for $Q = 1$. An appropriate ansatz for U is

$$\begin{aligned} U(x) &= i\frac{x_\mu}{x}\tau_\mu \\ \tau_\mu &= \begin{pmatrix} \vec{\tau} \\ -i \end{pmatrix} \\ \tau_a \tau_b &= \delta_{ab} + i\varepsilon_{abc}\tau_c \end{aligned}$$

This leads to A_μ^a by virtue of equation (5.54)

$$\begin{aligned} A_i^a &= 2(\lambda^a)^{-1} iU(\vec{x}) \partial_i U(\vec{x})^\dagger \\ &= -iii2(\lambda^a)^{-1} \frac{x_\mu}{x}\tau_\mu\tau_\nu^\dagger \frac{\partial}{\partial x_i}\frac{x_\nu}{x} \\ &= 2i(\lambda^a)^{-1}\tau_\mu\tau_\nu^\dagger \left(\frac{x x_\mu \delta_{\nu i} + x_\mu x_\nu x_i}{x^3}\right) \end{aligned}$$

For SU(2), this can be simplified to

$$\begin{aligned} A_\mu^a &= 2\eta_{a\mu\nu}\frac{x_\nu}{x^2} \\ \eta_{a\mu\nu} &= \begin{cases} \varepsilon_{a\mu\nu} \text{ for } \mu,\nu = 1,2,3 \\ \delta_{a\mu} \text{ for } \nu = 4 \\ -\delta_{a\nu} \text{ for } \mu = 4 \end{cases} \end{aligned}$$

where $\eta_{a\mu\nu}$ is called the t'Hooft symbol. A parametrized solution

$$A_\mu^a = \eta_{a\mu\nu}\frac{x_\nu}{x^2} f\left(x^2\right)$$

can be used to fulfill the self-duality requirement for $Q \neq 1$. $f(x^2)$ is required to be 1 as $x \to \infty$ to let the result be a pure gauge. Inserting this leads to

$$\begin{aligned} G_{\mu\nu}^a &= \Gamma_{\mu\nu}^a \\ \partial_\nu A_\mu^a - \partial_\mu A_\nu^a - f^{abc}A_\mu^b A_\nu^c &= \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}\left(\partial_\sigma A_\rho^a - \partial_\rho A_\sigma^a - f^{abc}A_\rho^b A_\sigma^c\right) \end{aligned}$$

The derivatives can be calculated as

$$\partial_\mu \left(\frac{f x_\nu}{x^2} \right) = \frac{2 x_\nu x_\mu}{x^2} f' + \frac{f}{x^2} \delta_{\mu\nu} - \frac{2 f x_\mu x_\nu}{x^4} \tag{5.56}$$

The first two terms on the right hand side can be changed as

$$\begin{aligned}
\varepsilon_{\mu\nu\rho\sigma} \eta_{a\rho\alpha} \partial_\sigma \frac{x_\alpha}{x^2} f &= \varepsilon_{\mu\nu\sigma\rho} \eta_{a\alpha\rho} \partial_\sigma \frac{x_\alpha}{x^2} f \\
&= \left(\delta_{\alpha\mu} \eta_{a\nu\sigma} + \delta_{\alpha\sigma} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\sigma\mu} \right) \partial_\sigma \frac{x_\alpha}{x^2} f \\
&= \eta_{a\nu\sigma} \partial_\sigma \frac{x_\mu}{x^2} f + \eta_{a\mu\nu} \partial_\sigma \frac{x_\sigma}{x^2} f + \eta_{a\sigma\mu} \partial_\sigma \frac{x_\nu}{x^2} f \\
\varepsilon_{\mu\nu\rho\sigma} \eta_{a\sigma\alpha} \partial_\rho \frac{x_\alpha}{x^2} f &= -\varepsilon_{\mu\nu\rho\sigma} \eta_{a\alpha\sigma} \partial_\rho \frac{x_\alpha}{x^2} f \\
&= -\left(\delta_{\alpha\mu} \eta_{a\nu\rho} + \delta_{\alpha\rho} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\rho\mu} \right) \partial_\rho \frac{x_\alpha}{x^2} f \\
&= -\eta_{a\nu\rho} \partial_\rho \frac{x_\mu}{x^2} f - \eta_{a\mu\nu} \partial_\rho \frac{x_\rho}{x^2} f - \eta_{a\rho\mu} \partial_\rho \frac{x_\nu}{x^2} f \\
\varepsilon_{\mu\nu\rho\sigma} \eta_{a\rho\alpha} \partial_\sigma \frac{x_\alpha}{x^2} f - \varepsilon_{\mu\nu\rho\sigma} \eta_{a\sigma\alpha} \partial_\rho \frac{x_\alpha}{x^2} f &= 2\eta_{a\nu\sigma} \partial_\sigma \frac{x_\mu}{x^2} f + 2\eta_{a\mu\nu} \partial_\sigma \frac{x_\sigma}{x^2} f + 2\eta_{a\sigma\mu} \partial_\sigma \frac{x_\nu}{x^2} f
\end{aligned}$$

And the result is then

$$\begin{aligned}
& \eta_{a\mu\alpha} \partial_\nu \frac{x_\alpha}{x^2} f - \eta_{a\nu\alpha} \partial_\mu \frac{x_\alpha}{x^2} f - \eta_{a\nu\sigma} \partial_\sigma \frac{x_\mu}{x^2} f - \eta_{a\mu\nu} \partial_\sigma \frac{x_\sigma}{x^2} f - \eta_{a\sigma\mu} \partial_\sigma \frac{x_\nu}{x^2} f \\
= \; & -2 \left(\frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} f^{abc} \eta_{b\rho\alpha} \eta_{c\sigma\beta} \frac{x_\alpha x_\beta}{x^4} f^2 - f^{abc} \eta_{b\mu\alpha} \eta_{c\nu\beta} \frac{x_\alpha x_\beta}{x^4} f^2 \right) \\
& \eta_{a\mu\alpha} \partial_\nu \frac{x_\alpha}{x^2} f - \eta_{a\nu\alpha} \partial_\mu \frac{x_\alpha}{x^2} f - \eta_{a\nu\sigma} \partial_\sigma \frac{x_\mu}{x^2} f - \eta_{a\mu\nu} \partial_\sigma \frac{x_\sigma}{x^2} f - \eta_{a\sigma\mu} \partial_\sigma \frac{x_\nu}{x^2} f \\
= \; & -2 f^{abc} \frac{x_\alpha x_\beta}{x^4} f^2 \left(\frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} \eta_{b\rho\alpha} \eta_{c\sigma\beta} - \eta_{b\mu\alpha} \eta_{c\nu\beta} \right)
\end{aligned} \tag{5.57}$$

All derivatives are symmetric in the indices by virtue of (5.56) and hence they can be changed to

$$\begin{aligned}
& \eta_{a\mu\alpha} \partial_\nu \frac{x_\alpha}{x^2} f - \eta_{a\nu\alpha} \partial_\mu \frac{x_\alpha}{x^2} f - \eta_{a\nu\alpha} \partial_\mu \frac{x_\alpha}{x^2} f - \eta_{a\mu\nu} \partial_\sigma \frac{x_\sigma}{x^2} f + \eta_{a\mu\alpha} \partial_\nu \frac{x_\alpha}{x^2} f \\
= \; & -2 f^2 \frac{x_\alpha x_\beta}{x^4} f^{abc} \left(\frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} \eta_{b\rho\alpha} \eta_{c\sigma\beta} - \eta_{b\mu\alpha} \eta_{c\nu\beta} \right)
\end{aligned}$$

The right hand side can be changed using SU(2) f^{abc} to

$$\begin{aligned}
& \frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} \left(\delta_{\rho\sigma} \eta_{a\alpha\beta} - \delta_{\rho\beta} \eta_{a\alpha\sigma} + \delta_{\alpha\beta} \eta_{a\rho\sigma} - \delta_{\alpha\sigma} \eta_{a\rho\beta} \right) - \delta_{\mu\nu} \eta_{a\alpha\beta} + \delta_{\mu\beta} \eta_{a\alpha\nu} - \delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta} \\
= \; & \frac{1}{2} \left(-\delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta} - \delta_{\alpha\mu} \eta_{a\nu\beta} + \delta_{\beta\nu} \eta_{a\mu\alpha} - \delta_{\beta\mu} \eta_{a\nu\alpha} \right) - \delta_{\mu\nu} \eta_{a\alpha\beta} - \delta_{\mu\beta} \eta_{a\nu\alpha} - \delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta}
\end{aligned}$$

Multiplying with $x_\alpha x_\beta$ results in

$$\begin{aligned}
& \frac{1}{2} \left(-\delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta} - \delta_{\alpha\mu} \eta_{a\nu\beta} + \delta_{\beta\nu} \eta_{a\mu\alpha} - \delta_{\beta\mu} \eta_{a\nu\alpha} \right) - \delta_{\mu\nu} \eta_{a\alpha\beta} - \delta_{\mu\beta} \eta_{a\nu\alpha} - \delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta} \\
= \; & -\frac{3}{2} \eta_{a\mu\nu} x^2 + 2\eta_{a\mu\alpha} x_\nu x_\alpha - 2\eta_{a\nu\alpha} x_\mu x_\alpha
\end{aligned}$$

Hence the result is now

$$2\eta_{a\mu\alpha}\partial_\nu\frac{x_\alpha}{x^2}f - 2\eta_{a\nu\alpha}\partial_\mu\frac{x_\alpha}{x^2}f - \eta_{a\mu\nu}\partial_\sigma\frac{x_\sigma}{x^2}f = -2\frac{f^2}{x^4}\left(-\frac{3}{2}x^2\eta_{a\mu\nu} + 2\eta_{a\mu\alpha}x_\nu x_\alpha - 2\eta_{a\nu\alpha}x_\mu x_\alpha\right)$$

The left hand side can be expanded using (5.56)

$$\begin{aligned} & 2\eta_{a\mu\alpha}\partial_\nu\frac{x_\alpha}{x^2}f - 2\eta_{a\nu\alpha}\partial_\mu\frac{x_\alpha}{x^2}f - \eta_{a\mu\nu}\partial_\sigma\frac{x_\sigma}{x^2}f \\ = & \frac{4f'}{x^2}\left(\eta_{a\mu\alpha}x_\nu x_\alpha - \eta_{a\nu\alpha}x_\mu x_\alpha - \frac{1}{2}\eta_{a\mu\nu}x^2\right) - \frac{4f}{x^4}\left(\eta_{a\mu\alpha}x_\nu x_\alpha - \eta_{a\nu\alpha}x_\mu x_\alpha\right) \end{aligned}$$

Equating both to order $1/x^2$ yields a condition equation

$$\begin{aligned} 0 = & -2\eta_{a\mu\nu}f' + \frac{1}{x^2}\left(4f'\eta_{a\mu\alpha}x_\nu x_\alpha - 4f'\eta_{a\nu\alpha}x_\mu x_\alpha - 3f^2\eta_{a\mu\nu}\right) \\ & + \frac{4f}{x^4}\left(\eta_{a\nu\alpha}x_\mu x_\alpha - \eta_{a\mu\alpha}x_\nu x_\alpha + f\eta_{a\mu\alpha}x_\nu x_\alpha - f\eta_{a\nu\alpha}x_\mu x_\alpha\right) \end{aligned}$$

Choosing, e. g., $a = 2$, $\nu = 3$, $\mu = 4$ this results in

$$\frac{1}{x^2}\left(-4f'x_3x_2 - 4f'x_1x_4\right) + \frac{4f}{x^4}\left(x_1x_4 + x_3x_2 - fx_3x_2 - fx_1x_4\right) = 0$$

Equating the same factors of x_3x_2 results in

$$f\left(1-f\right) - x^2f' = 0$$

A solution is provided by the ansatz $f = x^2/\left(x^2+\rho^2\right)$,

$$\begin{aligned} & \frac{x^2}{x^2+\rho^2}\frac{\rho^2}{x^2+\rho^2} - x^2\left(\frac{1}{x^2+\rho^2} - \frac{x^2}{\left(x^2+\rho^2\right)^2}\right) \\ = & \frac{x^2\rho^2}{\left(x^2+\rho^2\right)^2} - \frac{x^2}{\left(x^2+\rho^2\right)} - \frac{x^4}{\left(x^2+\rho^2\right)^2} = 0 \end{aligned}$$

The field strength is then

$$A^a_\mu = 2\eta_{a\mu\nu}\frac{x_\nu}{\left(x^2+\rho^2\right)} \tag{5.58}$$

The field strength tensor can then be calculated as

$$\begin{aligned} G^a_{\mu\nu} = & \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu \\ = & 2\eta_{a\nu\alpha}\partial_\mu\frac{x_\alpha}{\left(x^2+\rho^2\right)} - 2\eta_{\alpha\mu\alpha}\partial_\nu\frac{x_\alpha}{\left(x^2+\rho^2\right)} + 4f^{abc}\eta_{b\mu\alpha}\eta_{c\nu\beta}\frac{x_\alpha x_\beta}{\left(x^2+\rho^2\right)^2} \\ = & \frac{-4\eta_{a\mu\nu}\rho^2}{\left(x^2+\rho^2\right)^2} \end{aligned}$$

The squared field strength is finally given by

$$\left(G^a_{\mu\nu}\right)^2 = \frac{192\rho^4}{\left(x^2+\rho^2\right)^4}.$$

Integrating over it shows that the total action is finite.

The resulting instanton solution is well localized, and is characterized by its size. It is invariant under rotation in space. It can be shown that is not invariant under color transformations, and its orientation has to be fixed (it can be changed by unitary transformations of $R^{ba}A^a_\mu$). Since the instanton can be placed at any position in space, its position is also a free parameter. Size, position, and color orientation are together known as the collective coordinates of the instanton, or moduli. The space of all possible values of these parameters is known as moduli space.

Direct integration shows that indeed $Q = 1$,

$$\begin{aligned} S &= \frac{1}{4g^2}\int d^4x \left(G^a_{\mu\nu}\right)^2 = \frac{192\rho^4}{4g^2}\int d\phi \int d\cos\theta \int dx \int dt \frac{x^2}{\left(x^2+t^2+\rho^2\right)^4} \\ &= \frac{60\pi^2\rho^4}{g^2}\int dx \frac{x^2}{\left(x^2+\rho^2\right)^{\frac{7}{2}}} = \frac{60\pi^2\rho^4}{g^2}\frac{2}{15\rho^4} = \frac{8\pi^2}{g^2} \end{aligned}$$

Thus, the instanton is the lowest-action configuration with non-trivial topological charge, and thus the classical vacuum solution.

In normal perturbation theory, the coupling constant is associated with the field and not the action. By changing $A \to gA$ the action (5.56) is made independent of g and the field strength tensor (5.54) is changed to

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$$

A new solution is then

$$\frac{x^2}{g\left(x^2+\rho^2\right)}$$

resulting in a change of the fields and the fields strength tensor by $1/g$, indicating that these fields are much stronger than perturbative fields. The instanton, and topological excitations in general, are genuine non-perturbative contributions.

It is possible to construct an anti-self-dual $Q = -1$ solution by replacing $\eta_{a\mu\nu}$ by $\overline{\eta_{a\mu\nu}}$ with

$$\overline{\eta_{a\mu\nu}} = \begin{cases} \varepsilon_{a\mu\nu} \text{ for } \mu,\nu = 1,2,3 \\ -\delta_{a\mu} \text{ for } \nu = 4 \\ \delta_{a\nu} \text{ for } \mu = 4 \end{cases} \tag{5.59}$$

which is unchanged besides this replacement.

The tunneling amplitude is again proportional to S_0, which is for the self-dual solution for $Q = 1$

$$P_{\text{tunnel}} = Ke^{-\frac{8\pi^2}{g^2}}\tau \tag{5.60}$$

The coefficient K can be determined analog to the one loop order calculations in the quantum mechanical case.

5.4.2.3 θ vacua

The previous section, showed that there are different vacua, and these vacua can be connected by tunneling events. The true vacuum is hence a superposition of the different vacuum states. If instanton solutions do not interact with each other, the tunneling amplitude to go from vacuum i to vacuum j can be determined as

$$\left\langle j|e^{-H\tau}|i\right\rangle = \sum_{N_+=0}^{\infty}\sum_{N_-=0}^{\infty}\frac{\delta_{N_+-N_--j+i}}{N_+!N_-!}\left(K\tau e^{-S_0}\right)^{N_++N_-}$$

where N_+ is the number of instantons and N_- the number of anti-instantons. Using

$$\delta_{ab} = \frac{1}{2\pi}\int_0^{2\pi} d\theta e^{i\theta(a-b)}$$

this becomes

$$\left\langle j|e^{-H\tau}|i\right\rangle = \frac{1}{2\pi}\int_0^{2\pi} d\theta \sum_{N_+=0}^{\infty}\sum_{N_-=0}^{\infty}\frac{e^{i\theta(N_+-N_--j+i)}}{N_+!N_-!}\left(K\tau e^{-S_0}\right)^{N_++N_-}$$

This can be rearranged as

$$\begin{aligned}\left\langle j|e^{-H\tau}|i\right\rangle &= \frac{1}{2\pi}\int_0^{2\pi} d\theta e^{i\theta(i-j)}\sum_{N=0}^{\infty}\frac{2^N\cos^N\theta}{N!}\left(K\tau e^{-S_0}\right)^N \\ &= \frac{1}{2\pi}\int_0^{2\pi} d\theta e^{i\theta(i-j)}e^{2K\tau e^{-S_0}\cos\theta}\end{aligned}$$

From this can be read of, that the vacuum state wave functions are $e^{i\theta n}\left|n\right\rangle$, and the energy is $E(\theta) = -2Ke^{-S_0}\cos\theta$. The full vacuum state, the θ vacuum, is

$$\left|\theta\right\rangle = \sum_n e^{i\theta n}\left|n\right\rangle \tag{5.61}$$

The lowest energy is negative, as expected, since tunneling lowers the ground state energy. There are nevertheless no excitation associated with this energy, since the value of θ is fixed and cannot be changed. A fixed term could be added to the Lagrangian as an additional parameter, the so-called topological term

$$L = \frac{i\theta}{32\pi^2} G^a_{\mu\nu} \Gamma^a_{\mu\nu}.$$

The significance is not completely clear, since the interaction violates CP and T, but it is a surface term, and may be associated with screening. The value of θ in QCD seems to be experimentally smaller than 10^{-10}, and the complex is known as the strong CP-problem. Note that such a term does not affect perturbative calculations, because surface terms do not alter the perturbative expansion.

5.4.2.4 Tunneling amplitude

The calculation for the tunneling amplitude is in principle the same as for the quantum-mechanical problem, but much more tedious. Zero modes also exist, and lead to infrared divergences additionally to the ultraviolet divergences. Hence only results will be provided here. There are $4N_c$ zero modes, one for each polarization. They can be replaced by integration about collective coordinates, and give a factor $\sqrt{S_0}$. Integration over color orientation yields a factor, but of size and position yields a non-converging integral. Hence only a differential tunneling amplitude can be determined,

$$\frac{d^5 n_I}{d\rho d^4 z} \sim \left(\frac{8\pi^2}{g^2}\right)^{2N_C} e^{-\frac{8\pi^2}{g^2}} \frac{1}{\rho^5},$$

where the exponent of ρ is determined from the dimension. The regularization of this expression can be done using the Pauli-Villar scheme. Some more calculations yield

$$\frac{d^5 n_I}{d\rho d^4 z} \sim \frac{1}{\rho^5} \exp\left(-\frac{8\pi^2}{g^2} + \frac{11}{3} N_c \log\left(M\rho\right)\right)$$

where M is the regulator mass and the prefactor cancels. The complete result is

$$\frac{dn_I}{d\rho d^4 z} = \frac{0.466 e^{-1.679 N_c}}{(N_c-1)!\,(N_c-2)!} \left(\frac{8\pi^2}{g^2}\right)^{2N_c} \frac{1}{\rho^5} e^{-\frac{8\pi^2}{g^2} + \frac{11}{3} N_c \log(M\rho)} \tag{5.62}$$

As a consequence, small instantons are strongly suppressed. For large instantons the approximation breaks down. One problem associated with such a distribution is that even with less approximations the distribution is not integrable. This implies that there is nothing what suppresses large instantons, and the result becomes unstable at large distances. Though results using lattice calculations show this not to happen, the mechanism is unknown.

5.4.3 Light quarks

The presence of different vacua can be connected to the axial charge.

In perturbation theory, renormalization is not possible while conserving all charges, if external charge and axial charge fluxes are involved. The divergence in arbitrary order for two gauge fields coupling to the flavor singlet axial current is

$$\partial_\mu j^5_\mu = \frac{N_f}{16\pi^2} G^a_{\mu\nu} \Gamma^a_{\mu\nu} \tag{5.63}$$

where the gauge fields are arbitrary background fields. This is essentially an anomaly, which will be discussed in more detail in section 5.6. The surface term on the right hand side is relevant, since the vacua of QCD are not equivalent. The effect is, that by instanton events the axial charge is not conserved and connected to the topological charge Q.

Essential the zero mode moves every state one level up, and hence the chirality in each state changes. The consequence is, that the ninth Goldstone boson η' in QCD acquires even in the chiral limit a mass. Another consequence is that in the presence of light quarks, single instantons cannot exist (although there is some exception possible), but correlated instantons are necessary, even in the limit of a dilute gas. Secondly, quarks can travel from instanton to instanton. They change their chirality and the same flavor may not travel together.

The fact that the chirality changes at an instanton also implies that in an instanton background chirality is not conserved. Thus, chiral symmetry breaking can be viewed as a consequence of the presence of instantons.

5.4.4 Other topological excitations

Instantons are special, because they are the classical minimal energy configuration. However, this by no means imply that they are the only topological stable solutions. Such other solutions could, in fact, be even minimizing the quantum action[19].

Two particularly relevant such configurations are monopoles and vortices. While instantons are point-like events in space-time, these objects are of higher dimensionality.

A (chromo)monopole is an object which has a world-line, and therefore forms a one-dimensional object in space-time. A possible field configuration is, e. g.,

$$A^a_\mu = -\delta^{3a} \frac{1}{g} \vec{e}_\varphi \frac{1+\cos\theta}{r\sin\theta},$$

[19]Strictly speaking, any topological solution alone can never be relevant, as only infinite-action configurations have a relevant measure in the path integral. However, it is always possible to add quantum noise to a topological configuration to increase its action to make it relevant.

which is given in spatial spherical coordinates. It is static, in contrast to the instanton, which is located in space-time. As the field strength diverges along a line, it is also called a defect.

Vortices are closed two-dimensional surfaces, with a typical field configuration like

$$A^a_\eta = \delta^{3a}\frac{1}{g}\frac{\mu(\rho)}{\rho}$$

in cylindrical coordinates, where the profile function μ varies from zero at $\rho = 0$, making the field configuration regular, to an odd integer $2n + 1$, where n is the number of flux quanta in the vortex, at $\rho = \infty$, but will go again to zero for a zero-flux vortex. The (fuzzy) surface of the vortex is essentially defined by the maximum of the field configuration in ρ direction.

There are also other topological field configurations, including volume-like ones, which will not be detailed further. These field-configurations are not unrelated. E. g. several investigations, mostly using lattice calculations, find that instantons cluster along the worldlines of monopoles, which in turn lie on the surfaces of vortices. It is therefore likely that no single type alone can be identified as the low-energy effective degree of freedom of QCD.

Like in the case of instantons, almost all other topological configurations contribute in the process of chiral symmetry breaking. In fact, the properties of chiral symmetry breaking can be rather well reproduced by topological excitations alone. It appears hence likely that they play the dominant role in terms of effective degrees of freedom in this process.

5.5 Confinement

One of the most remarkable features of QCD is probably the non-observation of free quarks and gluons. This phenomenon is called confinement. However, shaping a precise meaning of what confinement is in a theoretical language is a highly non-trivial question. Hence, probably simpler is first an experimental statement: Confinement is the fact that no massless, strongly interacting vector particles and no fermions with fractional electric charges with hadronic masses have been observed. These statements include two important requirements. One is that such particles should be strongly interacting. Otherwise, e. g. the photon would be a candidate. They secondly require the masses to be of hadronic size. Since a multitude of beyond-the-standard-model scenarios have also fractionally electrically charged particles, this distinction is necessary.

There are two problems with these statements. One is that they obviously do not refer in any way to color. The reason is that experimentally there is no color, as there is no detector for it. In fact, there is no experimental concept of color. This is theoretically obvious: Color is a gauge degree off freedom. There is a theorem, Elitzur's theorem, which states that essentially all quantities not-invariant under local gauge transformations necessarily vanish without gauge-fixing. Since gauge-fixing is a coordinate choice, this implies that any observable quantity must be gauge-invariant. Thus, color cannot be used in an experimental context.

The second is, that this condition can not be maintained for all siblings of QCD, so-called QCD-like theories. Here, a sibling are theories with the same structure as QCD, but different gauge group, different number of fermions, or different representations for the fermions. E. g., in theories with gauge group G_2 or with fermions in the adjoint representation, there are hybrids, i. e. hadrons which consist out of a finite number of gluons and a single quark, which are gauge-invariant. Such particles would carry fractional electrical charge. On the other hand, QCD with about 13-15 flavors of massless quarks would likely have a light, if not even massless, ρ meson. These examples show how hard it is to give even experimentally a well-defined meaning to the word confinement.

The situation in theory is both more simple and more complex.

Simple as the requirement of gauge-invariance seems to imply that gluons and quarks cannot be observed, period. However, it is by far non-trivial to show this to be correct in a more rigorous sense. In fact, for an Abelian gauge theory it is possible to construct the equivalent of a free quark by a dressing of a gauge-dependent source with an infinite number of photons, though with still a finite total energy. Axiomatic field theory seems to indicate that this, called bleaching, is impossible for non-Abelian gauge theories (there is no almost local colorless state with otherwise quark quantum numbers), but a proof is lacking[20].

Complex, as, except for just calculations using lattice, we have only a vague idea of how this effect works out dynamically, i. e. how it manifests in correlation functions.

In the following, more will be said about the theoretical perspective. These remarks should however make one aware that it is always necessary to first find a common definition of the word 'confinement' ere one discusses it with somebody else.

[20]The situation at finite temperature is worse, and thus it is unclear whether deconfinement, to be discussed in chapter 7, could actually occur.

5.5.1 Wilson criterion and the strong-coupling expansion

One of the most naive ways in which to investigate confinement is to investigate the following situation, motivated by the idea of test charges in classical electrodynamics. Reduce first the problem to the quenched case, i. e. Yang-Mills theory. Place then two test-charges into the system, one with fundamental and one with anti-fundamental charge. Since test-charges can be taken to be static, this situation is completely characterized by the spatial distance between the two test charges. Connect these with a gauge field such that the total setup is gauge-invariant. Finally, measure the total energy of this arrangement as a function of the distance of the test charges.

The interesting result is that this energy has the following form,

$$V(r) = \sigma r + c + d\frac{\alpha}{r} + \mathcal{O}(\alpha^2), \tag{5.64}$$

where c and d are some constants, α is the strong coupling constant, and σ is called the string tension for reasons to become clear soon. The Coulomb-like term as well as most of the higher order corrections are what is expected in perturbation theory. In fact, since in perturbation theory the asymptotic, non-interacting states are quarks and gluons, perturbation theory knows nothing of confinement. It is a purely non-perturbative phenomena.

The other two terms are significant. They imply that the energy is linearly rising with distance. In fact, $\sigma \sim (400\text{ MeV})^2$ is so large that moving the two charges even the size of a proton away from each other is already very expensive, and any macroscopical scale is absurdly so. There is a restraining force associated with such a potential, which attempts to keep the charges together. This is not a necessary consequence of the requirement of gauge invariance, but a genuine feature of Yang-Mills theory. The requirement of gauge invariance also occurs in theories where this behavior is not observed. It is thus a genuine phenomena of the current theory.

This behavior is actually the one expected if a (relativistic) string of tension σ would be put between the charges, and then elongated. In fact, even the sub-leading constant term c turns out to be in quite good agreement with such a result. This is the reason to call this the string tension[21].

That Yang-Mills theory can create such a behavior can be shown in the strong-coupling limit $\alpha \to \infty$, using a lattice-regularized version of Yang-Mills theory. To do so, it first requires to be a bit more explicit about the corresponding operator. It is given by the

[21]In fact, before the inception of QCD, this lead to the birth of string theory as a theory of hadronic interactions. String theory was then abolished due to conceptual problems and a wrong high-energy behavior, only to be reborn as a possibility for quantum gravity almost two decades later.

so-called Wilson line,

$$U(C) = P \exp\left(ig \int\limits_C ds^\mu A^a_\mu \frac{\lambda^a}{2} \right), \tag{5.65}$$

where C is a path, which starts at the first charge, moves to the second over the distance R, follows this charge for some time T, then returns to the first charge, and final closes on itself by going back in time. It is therefore a rectangle of size RT. That the path-ordered exponential (5.65) is actual the correct expression can be seen by exponentiating the covariant derivative, which connects two fundamental quark sources over an infinitesimal distance to give a gauge-invariant expression. However, it should be noted that nowhere enters anymore that the quarks sources are fermions, and the same expression is obtained for scalars or other objects. The reason is simply that for static test charges the Lorentz structure does not matter. Its (Euclidean) expectation value

$$W = \langle U \rangle = \frac{1}{Z} \int \mathcal{D}A_\mu \mathrm{tr} U e^{-S}$$

is gauge-invariant, as by expansion it can be shown that U transforms under a gauge-transformation g as gUg^{-1}. On the lattice, if taking a rectangle of size the lattice spacing, it coincides with the plaquette.

From this the Wilson potential is defined as

$$V(R) = -\lim_{T\to\infty} \frac{1}{T} \log W.$$

It coincides with the potential (5.64). Hence, the Wilson line must behave asymptotically as $\exp(-RT\sigma)$. Thus, an asymptotic non-vanishing string tension σ implies that the exponent behaves like the area enclosed by the curve C. This is called the area law, and taking a non-vanishing σ to be equivalent to confinement, this area law is a criterion, the so-called Wilson criterion, for confinement. In contrast, if the exponent only scales with the length of the curve C, a so-called perimeter law, the string tension is zero, and the potential (5.64) is qualitatively the same as the one of QED, and therefore there is no confinement according to the Wilson criterion.

Before showing the former, it is best to first convince oneself of the latter, i. e. that in QED this produces the familiar $1/r$ potential. Quenched QED is just Maxwell theory, and hence there are no interactions. Of course, it is necessary to gauge-fix, which can be most conveniently done in a covariant gauge, in the following the Feynman gauge. The

gauge-fixed Lagrangian together with the Wilson line becomes

$$\begin{aligned} W &= \frac{1}{Z}\int \mathcal{D}\exp\left(ie\int\limits_C ds_\mu A_\mu - \int d^4x\left(F_{\mu\nu}F_{\mu\nu} + \frac{1}{2}(\partial_\mu A_\mu)^2\right)\right) \\ &= \frac{1}{Z}\int \mathcal{D}\exp\left(ie\int\limits_C ds_\mu A_\mu - \frac{1}{2}d^4x\left(\partial_\mu A_\nu\right)^2\right). \end{aligned}$$

This integral, after shifting the gauge-field, is a standard Gaussian one, and can therefore be directly evaluated. This yields

$$\begin{aligned} W &= \exp\left(-\frac{e^2}{2}\int\limits_C ds_\mu \int dt_\nu S_{\mu\nu}(s-t)\right) \\ S_{\mu\nu} &= -\frac{1}{4\pi}\frac{\delta_{\mu\nu}}{x^2} = \frac{\delta_{\mu\nu}}{(2\pi)^4}\int d^4p\frac{e^{ipx}}{p^2} \end{aligned}$$

where S is also known as the Schwinger function, and nothing but the propagator in space-time rather than in momentum space.

Evaluating the line-integral along a rectangle is not entirely straightforward. The integrand has singularities at coinciding points of the paths. They stem from the singularities of the propagator, and have to be regularized. The simplest resolution is to deform the integration contours by some ϵ such that this does not occur. Then all the integrals are elementary and the Wilson line becomes

$$W = \exp\left(\frac{e^2}{2\pi^2}\left(\frac{T}{R}\tan^{-1}\frac{T}{R} + \frac{R}{T}\tan^{-1}\frac{R}{T} - \frac{1}{2}\ln\left(\left(1+\frac{T^2}{R^2}\right)\left(1+\frac{R^2}{T^2}\right)\right) - \frac{1}{\epsilon}(T+R) + 2 - \ln\frac{\epsilon^2}{RT}\right)\right.$$

which leads to the potential

$$V(R) = -\frac{e^2}{4\pi R} + \frac{e^2}{2\pi^2\epsilon}.$$

The first part is the familiar electromagnetic potential, justifying the identification as the potential. The second, constant, term depends on the regularization, and is essentially just a constant shift of the potential. After renormalization, this shift can be set to zero, which is then indeed the classical expression.

Such a calculation is spoiled for Yang-Mills theory by the self-interactions of the gluons. To make still progress, it is possible to perform a strong-coupling expansion, i. e. consider the case of $g \to \infty$. The strong-coupling expansion is essentially an expansion in suitable graphs. Start with the Wilson plaquette action (5.6). The strong-coupling expansion is equivalent to $\beta \to 0$. Hence, it is a series expansion of the Boltzman weight,

$$e^{-\beta S} = \sum \frac{1}{n!}(\beta S)^n = \sum_k \chi^k \int \mathcal{D}(U) e^{-\beta S}\chi^{k*}$$

where the characters χ are given by $\mathrm{tr}D^k$ where D is a representation of dimensionality k of the gauge group. Especially, the χ are numbers. Calculating them is an exercise in group theory.

Performing this calculation for the case of the Wilson loop leads to the results that essentially a covering of the area enclosed by C by plaquettes is the dominant contribution. The contribution is then proportional to the number of plaquettes, and thus to the area. The pre-factor can then be calculated by calculating the corresponding characters. This yields at leading order

$$W = \exp\left(\frac{\ln \frac{1}{3g^2}}{a^2} RT\right)$$

where a is the lattice spacing, and therefore a string tension of $1/a^2 \times \ln 3g^2$, where the lattice spacing just sets the dimensionality. Hence, the string tension is essentially given by the (large) coupling, as was to be expected given that no other parameter exist in the theory.

The drawback of this argumentation is that the strong-coupling expansion is possibly not connected analytically to the continuum limit of $g \to 0$. Hence, the proof of having an area-law in the strong-coupling limit has not necessarily any implications for the continuum theory, as reassuring as the result itself is.

Performing numerical calculations using lattice gauge theory did, however, show that for all practical purposes the string tension survives, provided there is no further non-analyticity involved in taking the continuum or infinite-volume limit.

It should be noted that a string also forms in baryons. The form was long a debate, but finally the results strongly favor a three-edged star, rather than a triangle or more complicated shapes. Thus, there is a common center (at least for three quarks of the same mass), from which strings to the three quarks emerge symmetrically.

5.5.2 Regge trajectories and quarkonia spectra

Though quarks are not infinitely heavy, the potential (5.64) is not without merit, as it has a number of implications observable even with finite quark masses.

One of the results can already be inferred from a classical calculation. A meson can in its simplest form be viewed as a rigid string of tension σ with the quark and anti-quark at the opposite ends of the string. The classical rotation energy of this assembly is given by

$$E = \frac{1}{2} I \omega^2.$$

where I is the moment of inertia, given by $ml^2/12$, where m is the mass of the assembly, and l the length. The angular velocity is given by $\vec{\omega} ml^2 = \vec{L}$, where $\vec{L}$ is the angular

momentum. This yields for the total energy

$$E = \frac{1}{24}\frac{1}{ml^2}\vec{L}^2.$$

Doing a semi-classical quantization gives $\vec{L}^2 = L(L+1)$ with L now the angular momentum quantum number. Identifying then ml^2 with $1/\sqrt{\sigma}$ and E with the rest mass of the particles, this implies that the mass of a meson is roughly proportional to its angular momentum, and thus total spin, squared, with the string tension as a constant of proportionality. Indeed, such a behavior is observed for the mesons, at least for angular excitations, and this behavior is called a Regge trajectory. In fact, also baryons follow this pattern. Even the constant of proportionality has a value of about 650 MeV, which is not to far off from the string tension proper. This is a great success to the string-picture of the Wilson potential.

Another spectacular success of the string picture are quarkonia, and to some extent also other mesons. Assuming just the simple potential (5.64) and treating the heavy quarks in quarkonia non-relativistically leads to a Schrödinger-type wave equation. More refined versions of this are obtained from heavy-quark effective theory, as described in section 5.1.1. Solving this Schrödinger equation, essentially with the same methods as in quantum mechanics, yields energy levels as well as spin assignments close to the true spectra. The agreement can be further improved by including additional phenomenological contributions in the potential, e. g. spin-orbit forces or flavor-dependent correction terms as well as relativistic corrections. It is also possible to extend this approach to baryons, in the form of a quantum-mechanical three-body problem. However, states like the X, Y, Z mesons are not captured.

Both results, together also with the unexpected success of quenched lattice calculations, show that the main contribution of the quark interactions is essentially a gluon-induced effective potential of the type (5.64). Of course, this is not all. E. g., quantitative precision, especially for excited states, requires a full treatment. Also, some states are known which do not fit into the quantum number classification according to this quantum-mechanical picture, and others are missing. But overall the agreement is good.

This can be understood when recalling the consequences of chiral symmetry breaking in section 5.2: The quarks receive an additional contribution to their mass of order several hundred MeV. As the relevant binding energies/mass defects, i. e. the difference between these effective masses of the constituents and the bound states masses, are usually (much) smaller than these effective quark masses, the quarks are usually almost inert at the relevant distance scales. Only the long-range gluons then play a role.

5.5.3 The problem with dynamical particles

The main reason for the string picture is two-fold. One is that the gluon exchange creates such a string. This is a dynamical result, and can only be obtained by calculation. The other is a group-theoretical effect, and thus can be read off from the Lagrangian: It is not possible to dress a fundamental charge with gluons, because there is no singlet in the combination of any number of gluons and a single quark (or anti-quark). Both are thus necessary conditions for the string picture, and only both together are a sufficient condition. The latter is by no means trivial. For other gauge groups (e. g. G_2) or other representations of the quarks (e. g. adjoint) it is possible to dress a quark with a finite number of gluons (3 in case of G_2 and 1 in case of adjoint quarks). Hence, in QCD a gauge-invariant dressing requires other (anti-)quarks. This is energetically impossible if the quarks are infinitely heavy.

An interesting effect happens when dressing becomes energetically possible, i. e. in case of QCD when there are dynamical quarks. Then, at some distance, enough energy is deposited in the string that it is possible to form out of the original meson two mesons by converting this energy into particles. Thus, a highly radially excited meson will decay into two mesons, rather than into two quarks. This is very different from the QED case, where ionization is possible when investing enough energy. That this effect is 100% effective can be taken as another definition of confinement in full QCD. The process of the decay of the string is also known as string breaking, and has been numerically confirmed by lattice calculations.

However, the appearance of string-breaking makes the Wilson condition for confinement void. QCD with dynamical quarks, as well as some other theories without dynamical quarks, are not confining. This is at odds with the definition of confinement according to the observation of quark-like and gluon-like particles as discussed at the beginning of this section. The Wilson criterion is therefore often considered to be a theoretical construct, especially as it makes only a statement about particles which are not part of the theories (there are no quarks in Yang-Mills theory proper). Even if abandoning the Wilson criterion as a statement about confinement, the question of the appearance of a string tension is highly non-trivial, and not entirely solved so far.

But the string is not entirely gone: For a substantial distance, about ten Fermi, there is still a linear rising pseudo-potential between dynamical quarks. Pseudo-potential, as in a theory with particle creation and annihilation there is no well-defined concept of a potential anymore, and the name is just kept from the static case. This is sometimes called an intermediate string tension, and its presence explains once more why the quenched description is so suitable for hadronic properties, as well as the success of Regge theory

for full QCD.

5.5.4 Positivity violation

Even if one is willing to accept that QCD is not confining according to the Wilson criterion, one other problem is not addressed by it: The confinement of gluons. Since gluons can screen a gluon, any attempt to move the gluons inside a glueball[22] far away from each other will just create more glueballs, since gluons can screen each other. String breaking always occurs for gluons. Nonetheless, also in case of gluons there is an intermediate string tension, though it requires adjoint charges rather than fundamental ones to calculate it. Therefore, a useful condition for gluon confinement is so far not available.

However, a necessary criterion for gluons to be confined is that there are at least no colored gluonic states in the asymptotic spectrum. That this is not sufficient is immediately clear, as no unstable particle will appear in the asymptotic spectrum either, even if it is as physical as, e. g., the ρ meson.

To understand this more closely requires an elaboration of the discussion in section 5.3.2. There, the analytic structure in momentum space was discussed. But a more direct access, as is also exploited in the lattice approach in section 5.3.3, is in terms of the position space. This is granted by the Schwinger function. This function is essentially the propagator in position space at zero three-momentum,

$$\Delta(t) = \frac{1}{\pi}\int\limits_0^\infty dp_0 \cos(tp_0) D(p_0^2) \tag{5.66}$$

where $D(p^2)$ is the propagator.

In case of a stable particle with a simple pole mass M, having an Euclidean propagator

$$D(p) = \frac{1}{p^2 + m^2}, \tag{5.67}$$

and thus a pole at $p = \pm im$, the Schwinger function is given by

$$\Delta(t) = \frac{1}{2m} e^{-mt}.$$

This recovers the lattice result from section 5.3.3. Therefore, a simple exponential decay is expected in Euclidean time for a massive, stable particle.

[22] Which involves at least four gluons, as the four gluons fields inside the field-strength tensor squared is the minimum number of gluon fields to construct a gauge-invariant operator.

Beyond tree-level, the propagator (5.67) is modified by the appearance of a cut, starting at[23] $p = \pm 2im$. At lowest order in perturbation theory, a possible analytic form for such a propagator in four dimensions is given by

$$\begin{aligned} D(p) &= \frac{1}{p^2 + m^2 + \Pi(p^2, m^2)} \\ \Pi(p^2, m^2) &= -g^2 \left(\frac{\pi}{2\sqrt{3}} + \sqrt{1 + \frac{4m^2}{p^2}} \operatorname{atanh} \left(\sqrt{\frac{p^2}{4m^2 + p^2}} \right) \right), \end{aligned} \tag{5.68}$$

where g has dimension of mass. This form is motivated by leading-order perturbation theory, and occurs, e. g., for a scalar theory with a three-point coupling.

If the particle is not stable, the poles are moved off the first Riemann sheet onto the second Riemann sheet at $m + i\Gamma/2$. The cut then starts at $\pm 2iM$, where M is the mass of the particles in which the original particle can decay, assuming for the moment only this two particles in the theory. In this case, the propagator (5.68) is modified to

$$\begin{aligned} D(p)^{-1} &= p^2 + \left(m + i\frac{\Gamma}{2} \right)^2 - g^2 \left(\Pi\left(p^2, m^2, \Lambda^2\right) - \Pi\left(\left(m + i\frac{\Gamma}{2} \right)^2, m^2, \Lambda^2 \right) \right) \\ &\quad - h^2 \left(\Pi\left(p^2, M^2, \Lambda^2\right) - \Pi\left(\left(m + i\frac{\Gamma}{2} \right)^2, M^2, \Lambda^2 \right) \right) \\ \Pi(p^2, m^2, \Lambda^2) &= -\sqrt{1 + \frac{4m^2}{p^2}} \operatorname{atanh} \sqrt{\frac{p^2}{4m^2 + p^2}} + \frac{(2\Lambda^2 + 4m^2 + p^2) \operatorname{atanh} \sqrt{\frac{p^2}{4\Lambda^2 + 4m^2 + p^2}}}{\sqrt{p^2}\sqrt{4\left(\Lambda^2 + m^2\right) + p^2}} \\ &\quad + \frac{1}{2} \ln \left(1 + \frac{\Lambda^2}{m^2} \right), \end{aligned} \tag{5.69}$$

where there are now two couplings of dimension mass, g and h, describing[24] a self-interaction g and a decay channel with strength h, and Λ is the cutoff, of which this renormalized propagator is independent. This is a more explicit example of the generic Breit-Wigner propagator (5.45).

However, since neither the gluon (nor the ghost or the quark) behaves necessarily like a physical particle, after all it is not gauge-invariant and thus not physical, its propagator may not be of either form (5.68) or (5.69). There have been various proposals, which form it may have instead. One, the so-called Gribov-Stingl type, has complex poles on the first Riemann sheet, but may have vanishing residues at the poles. This behavior can

[23] Or any other decay channel of the particle, which is below this one.

[24] Note that not all parameters are independent. The reality of the Euclidean propagator fixes one of the parameters m, Γ, M, g, and h as a function of the others, since the decay width is not an independent quantity.

be described by a meromorphic function

$$D(p) = \frac{e^2 + fp^2}{p^4 + 2m^2\cos(2\phi)\,p^2 + m^4}. \tag{5.70}$$

Its Schwinger function is given by

$$\Delta(t) = \frac{e^2}{2m^3\sin(2\phi)} e^{-tm\cos(\phi)} \left(\sin(\phi + tm\sin(\phi)) + \frac{fm^2}{e^2}\sin(\phi - tm\sin(\phi))\right). \tag{5.71}$$

The Schwinger function therefore exhibits oscillation with a period determined by the angle ϕ and the mass parameter m. Its positivity violations are signaling therefore the instability of the particle. The propagator (5.70) illustrates at the same time the difference between the concept of screening mass and pole mass. The screening mass is defined as the inverse square-root of the propagator at zero momentum, and given by

$$D(0)^{-\frac{1}{2}} = \frac{m^2}{e}.$$

Since e depends on the wave-function renormalization, this immediately shows that a screening mass is never renormalization-group-invariant, and can therefore not be a physical observable quantity. The pole mass, i. e., the location of the poles of (5.70) is given by

$$ime^{\pm i\phi},$$

and is thus a set of complex conjugate poles, but in general with non-zero residuum. This mass is renormalization-group invariant, if there is no additive mass-shift to m, and can thus be, at least in principle, a physical mass, though in the present case it may be gauge-dependent. Note that in the limit of real masses, e has to go to m and f to one, or otherwise a double-pole would emerge. However, such complex-conjugated poles do not fulfill the analyticity properties of physical particles, as discussed before, even if the mass makes sense.

There is another concept which appears when using the Euclidean correlation function. Assume that the propagator can be more generally written as

$$D(p) = \frac{Z}{p^2 + M(p^2)^2}, \tag{5.72}$$

with some wave-function renormalization constant Z. At first sight, the condition $p^2 = -M(p^2)^2$ could look like an indicator for the pole mass of the particle described by the propagator. However, this is only possible if the function $M(p^2)$ only depends on the real part of p^2 on the real p^2 axis. This is certainly the case for (5.67), but is not necessarily the case for (5.72). Thus, the point $-M(p^2)^2$ only gives a would-be pole mass, and the

correct pole mass is only obtained when taking into account the full dependence of $M(p^2)$ on complex momenta.

However, all of these concepts still have functions which do not offer a cut structure. Though the final behavior of the gluon propagator has not been settled, it appears possible that such a possibility should be kept. The results available from functional studies suggest a form which has a cut on the imaginary momentum axis starting at zero momentum. It is also a possibility that the gluon propagator vanishes at zero for some gauges. In this case, functional forms which provide such a structure are given by[25]

$$\begin{aligned} Z_1(p) &= \frac{A_z p^{2\kappa_{AA}}}{1+f+A_z p^{2\kappa_{AA}}} f_{\rm UV}(p^2) \\ Z_2(p) &= \frac{A_z p^{2\kappa_{AA}}}{(1+f+A_z p)^{2\kappa_{AA}}} f_{\rm UV}(p^2), \end{aligned} \tag{5.73}$$

where $f_{\rm UV}$ encodes the perturbative tail. Both have no simple complex or real poles on the first Riemann sheet, but have additional poles on further Riemann sheets. If the gluon propagator is finite at zero momentum a form like

$$Z_3(p) = \frac{p^2(m^2+p^2)^{\kappa_{AA}}}{\sigma(1+fp^2+gp^{2+2\kappa_{AA}})} f_{\rm UV}(p^2), \tag{5.74}$$

is suggested. This form behaves as (5.73) in an intermediate range, depending on the parameters. It has also the same analytic structure as (5.73). I. e., the only singularities in the complex plane is a cut along the real axis. In fact, the best established results for the gluon propagator in Landau gauge[26] are best described by (5.74).

Most importantly, all of the forms permit Wick rotation, essential to transfer the results back to Minkowski space. Despite the quite different forms, the propagators in momentum space are in all cases quite similar, which is one of the reasons why a final decision on the correct form is still out. The position-space form is quite different, but the differences only manifest at long times and thus give an exponentially suppressed signal, making it also complicated to decide. At short times, all agree, since this region is dominated by asymptotic freedom, and thus the behavior must be like in perturbation theory, i. e. a logarithmic decay in momentum space.

In any case, all propagators (5.70), (5.73), and (5.74), imply that positivity is manifestly violated for the gluon[27]. This implies the absence of the gluon from the asymptotically physical Hilbert space. But as noted, positivity violation is not equivalent to confinement.

[25]Note that similar functional forms are already encountered in QED

[26]Note that the definition of Landau gauge needs to be refined, see section 5.5.5.2.

[27]Note that a gluon propagator which vanishes at zero momentum is implying (maximal) positivity violation.

Whether confinement is implying positivity violation is not known. However, a confined particle is necessarily absent from the asymptotic physical Hilbert space. Thus, the positivity violation implies already this necessary condition, though this necessary condition may also be realized in other ways yet unknown.

Note that this does not prevent the existence of a stable asymptotic gluon state in the unphysical part of the Hilbert space of the theory; it is just excluded from the physical part of the Hilbert space[28]. Furthermore, the screening mass for an infrared non-vanishing gluon propagator is not necessarily implying the existence of a pole mass of the gluon[29], as illustrated by the form (5.74). The only implication of the positivity violation is thus just that the gluon will not appear as an asymptotic state in the physical part of the Hilbert space.

It should be noted that positivity violations of the Schwinger functions is necessarily implying positivity violations of the spectral function ρ, defined implicitly by

$$D(p) = \int_0^\infty dM^2 \frac{\rho(M^2)}{p^2 + M^2}, \tag{5.75}$$

where any possible one-particle pole is included in the spectral function. The implication is obtained by inserting into (5.66) the representation (5.75)

$$\Delta(t) = \frac{1}{\pi} \int_0^\infty dp_0 \cos(tp_0) \int_0^\infty dM^2 \frac{\rho(M^2)}{p_0^2 + M^2} = 2 \int dM e^{-Mt} \rho(M^2),$$

where it is assumed that both integrations can be exchanged.

Any positivity violations of the spectral function implies the absence of a Källen Lehmann representation. The spectral function of a unstable but otherwise physical particle, like described by (5.69), remains positive. Thus, so must be its Schwinger function. This is not necessarily the case for unphysical particles. This illustrates how sensitive the Schwinger function is to details of the propagator structure.

From a practical point of view, the Schwinger function yields a more direct access to the analytic properties than the spectral function. Also, the reconstruction of the spectral function is non-trivial, and in case of the propagator being only available on a finite number of (lattice) momenta necessarily not unique, leading to significant systematic uncertainties.

[28]This observation can already be indirectly inferred from the violation of the Oehme-Zimmermann superconvergence relation in perturbation theory.

[29]If one exists, it would be gauge-parameter-independent, as described by the Nielsen identities, though in general not gauge-independent.

A positivity violation of the spectral function immediately marks a state as unphysical. Such a violation follows also immediately if the propagator either vanishes at zero momentum or its derivatives w. r. t. p^2 are not of constant sign, since (5.75) implies

$$\frac{\partial^n D(p)}{\partial (p^2)^n} = (-1)^n \int\limits_0^\infty dM^2 \frac{\rho(M^2)}{(p^2+M^2)^n} \tag{5.76}$$

This provides a sufficient, but not necessary, condition to identify a particle as unphysical. The positivity violation in the Schwinger function removes a particle immediately from the physical asymptotic state space, which is a weaker statement.

This exhausts most of the possibilities for the gluon propagator. There is one remarkable additional option. I. e. a gluon propagator which behaves as

$$D(p) = \frac{Z}{p^4},$$

i. e. is a massless double-pole. Such a gluon propagator would immediately yield an interaction potential, which is linearly rising. However, since the exchanged particle is gauge-dependent, this is not the Wilson string tension, but a different form. Such a behavior is also known as infrared slavery. There are some gauges, most prominently the non-covariant Coulomb gauge[30] $\vec{\nabla}\vec{A}^a = 0$, in which such a behavior is observed. At least in Coulomb gauge it can be proven that the so obtained (Coulomb-)string tension is an upper limit to the Wilson string tension, i. e. only a non-vanishing Coulomb string tension permits the Wilson confinement criterion. It is not an equivalence, and the Coulomb string tension appears to be also non-zero if the Wilson string tension is.

It was long assumed that the gluon propagator is in any gauge of this form. However, this is certainly not the case, and e. g. Landau gauge is a counter-example. In Landau gauge, it appears to be infrared finite, and thus of any of the above mentioned form. The situation for the quark is more subtle, and it is not yet even in a single gauge unambiguously established whether it has a simple pole, multiple poles, or a cut-structure. However, it appears that this decision has little impact on observable quantities.

5.5.5 Kugo-Ojima and non-perturbative BRST

There are a number of more formal attempts to solve the confinement problem. One such attempt is the Kugo-Ojima construction. As was discussed in section 4.3, in perturbation

[30] Since the gauge condition is not Lorentz-invariant, there are frames in Minkowski space-time where the gauge condition is meaningless. Thus, Coulomb gauge alone is not well-defined, and will therefore not pursued here further. There is an extensive literature discussing these problems.

theory it is possible to classify states as being physical or unphysical by the BRST symmetry. The Kugo-Ojima construction is an attempt to extend this construction to all colored states, i. e. to show that being colored implies the absence from the physical Hilbert space.

5.5.5.1 The construction

This construction is rather extensive, and therefore only some of its salient features will be presented here. The total construction is essentially following those in section 4.3.2, just that it is not necessary to take the limit of $g \to 0$ to obtain the asymptotic states. This requires a number of preconditions to be fulfilled.

One is that there is no massless observable state. This may seem obvious, as no hadron is massless, but this is a rather non-trivial requirement: In the chiral limit the pions become massless. It is not clear what is the fate of the Kugo-Ojima construction then is, but if it works for a finite pion mass, there should be an extension to include the chiral limit as well. Note that photons are not relevant, as they do not couple (directly) to the strong interaction, and can therefore be factored out.

The second is that the construction has only be done in covariant gauges. Though this is a restriction, it would nonetheless be a big leap forward to have an explicit construction at least in one gauge. If a gauge is fixed, the color charge becomes a well-defined quantity, and global color symmetry is a valid concept. The second construction then translates into the requirement of an unbroken global color symmetry. This can be shown to be equivalent to a ghost propagator, which is always present in covariant gauges, which diverges stronger than a massless particle pole at zero momentum, i. e. $1/p^{2\alpha}$ with $\alpha > 1$. That is a straightforward criterion to check. This has been done in non-perturbative calculations. As it stands, this seems not to be the case, as a direct evaluation of the ghost, at least in Landau gauge, exhibits only a pole structure of a massless particle.

That this is not surprising can be seen from the last ingredient, the requirement of an unbroken BRST symmetry. It is a vital ingredient in the construction that the BRST symmetry is unbroken and has the same algebra as in the perturbative case, (4.12-4.14). This requires of course a further quartet to copy the BRST mechanism verbatim also for transverse gluons and quarks. Such states are not available as elementary states. Especially, the elementary states with the required non-vanishing ghost number have been all used up for the perturbative sector. Thus, the only possibility are composite states, involving ghosts, i. e. ghost-gluon or ghost-quark bound-states. That for the realization/breaking of a symmetry composite states are necessary is not a problem, as the case of the pion for chiral symmetry demonstrates.

But this elegant construction has a serious problem, which is caused by the Gribov-

Singer ambiguity.

5.5.5.2 The Gribov problem

The Gribov problem is genuine non-perturbative, and starts already with the quantization procedure. The reason is that some of the conditions in section 4.2 turn out to be incorrect.

The main problem starts with the realization that local conditions, like the covariant gauges, do not have unique solutions anymore. This can be most easily seen in gauges where perturbatively there is one and only solution, like in the Landau gauge. Thus, the statement of the existence of Gribov copies takes there the form that the gauge condition $\partial_\mu A^\mu_a = 0$ has more than one solution beyond perturbation theory.

An example is given by the instanton field configuration (5.58). It satisfies the Landau gauge condition, just because it is essentially a four-dimensional rotation. Acting with the gauge transformation

$$G(x) = \frac{\tau_\mu r_\mu}{r}, \tag{5.77}$$

on it transforms it into

$$\begin{aligned} A_\mu &= \frac{2\lambda^2}{r^2(r^2+\lambda^2)}\bar{\tau}_{\mu\nu} r_\nu \\ \bar{\tau}_{\mu\nu} &= \frac{1}{4i}(\bar{\tau}_\mu \tau_\nu - \bar{\tau}_\nu \tau_\mu). \end{aligned}$$

Though this field configuration has a different radial behavior, it remains essentially a rotation, and therefore also satisfies the Landau gauge condition. Since in the conventions used the field configuration does not depend on the gauge coupling, this is genuinely a non-perturbative effect. This can also be seen from the fact that the gauge transformation (5.77) is not continuously deformable to a unit matrix. It is hence not obtainable from a series of infinitesimal gauge transformation, it is a so-called large gauge transformation.

This is actually necessary. Since the perturbative construction proves that there is no infinitesimally adjacent gauge transformation to the Landau gauge, and it is therefore unique, any additional gauge copies have to be separated by a large gauge transformation. There are many more explicit examples known in the literature.

The origin of the problem is that a non-Abelian gauge group has a non-trivial structure. Especially, in a generic non-Abelian group it is impossible to cover the whole group with a single coordinate system[31]. The simplest example is the gauge group SU(2). It is equivalent to the surface of a 3-sphere. Such a surface cannot be described with a single coordinate system, as it would be ill-defined at least at one pole, and at least two coordinate systems

[31]This is also the reason why an Abelian gauge theory is not affected. U(1) is isomorphic to a circle, which can be covered by a single coordinate system.

are necessary. A unique gauge condition would require to identify at each point in space-time a single point on this sphere, to identify a unique representative of the gauge orbit. Perturbatively, this is no problem, as the assumption of a small coupling translates into the requirement of a small field amplitude, and therefore at all points in space and time the field values are close to the same point on the sphere. Non-perturbatively, large field amplitudes are possible, and therefore the field can be anywhere on the sphere. Thus, different coordinate systems are required for different points in space and time. This is in principle possible, but such a global information cannot be provided using a local gauge condition, i. e. a gauge condition which involves only the fields and derivatives, but requires global input, like integrals over the field. After all, the coordinate systems at different, possibly far, separated points in space and time are required. A formal proof of this for the class of covariant gauges has been established by Singer, but the argument already shows that the problem will surface for any local gauge condition.

Though this seems to be a formidable problem, it is rather a practical than a conceptual problem. Using gauge conditions which involve integrals over the gauge fields, it is always possible to construct gauge conditions, which are not ambiguous. It is straightforward to implement such gauge conditions in lattice calculations. However, they become very quickly very expensive in terms of computing time, it may even be exponentially expensive, when the volume is increased. The reason is that using an integrated gauge condition implies that an integral equation has to be satisfied, which becomes numerically more expensive the more lattice points there are.

The situation in the continuum is even worse, and little is known about practical solutions to the problem. To understand it better, it is best to concentrate on the best-studied case, the Landau gauge. To remedy the problem, further (non-local) constraints are required. To implement the constraints, a sequence of further conditions can be applied. In all cases, the first step taken is always to reduce the perturbative gauge freedom by implementing a local gauge condition, here the Landau gauge (3.13). This reduces the space of all gauge orbits to a hypersurface.

The remaining set of Gribov copies will be called the residual gauge orbit. Since the condition is perturbatively unique, only finite gauge transformations connect two different elements of the residual gauge orbit. This is trivially so, since any infinitesimal gauge transformation will move along the gauge orbit automatically out of the gauge-fixing hypersurface implemented by the perturbative gauge-fixing by construction. It thus remains to understand the structure of this hypersurface.

A possible first restriction of the residual gauge orbit is to constrain it to the so-called first Gribov region. This first Gribov region is defined by the requirement that the

Faddeev-Popov operator (4.3) is strictly positive semi-definite, i. e., all of its eigenvalues are zero or positive. This region can be shown to be bounded and convex, and the Faddeev-Popov operator has zero eigenvalues only on the boundary of this region, the so-called Gribov horizon. It can be shown that all gauge orbits pass at least once through the first Gribov region. That is very important, and must be fulfilled by any gauge condition, since otherwise physical information is lost by implementing it[32]. The boundedness is a remarkable fact, as it implies that when calculating physical observables no arbitrarily large field fluctuations have to be taken into account. It contains the origin of field-space, and thus perturbation theory, as well. This follows from the fact that in the vacuum case (4.3) reduces to the positive semi-definite Laplacian. Thus by restricting to the first Gribov region, ordinary perturbation theory is always included.

Besides this first Gribov region, the remainder of the residual gauge orbit is a set of further Gribov regions. These are separated by further concentric Gribov horizons, each having more and more negative eigenvalues. The number of negative eigenvalues increases by one by passing the boundaries of these regions, but stays constant inside. It is expected that every residual gauge orbit passes through every Gribov region, though there is not yet an explicit proof of this.

This restriction can be implemented using a θ-function in the perturbative gauge-fixed path integral (4.6)

$$\begin{aligned} \langle \mathcal{Q} \rangle &= \lim_{\xi \to 0} \int \mathcal{D}A_\mu \mathcal{D}c \mathcal{D}\bar{c} \mathcal{Q} \theta \left(-\partial_\mu D^{ab}_\mu \right) (A_\mu, c, \bar{c}) e^{-\int d^4x \mathcal{L}_g} \\ \theta \left(-\partial_\mu D^{ab}_\mu \right) &= \prod_i \theta(\lambda_i), \end{aligned} \tag{5.78}$$

where λ_i is the ith eigenvalue of the Faddeev-Popov operator (4.3). Thus, only if all eigenvalues are positive or zero the θ-function contributes, requiring that the definition $\theta(0) = 1$ has to be made for the step function. Note that the restriction is actually necessary, as otherwise the formal inversion of the Faddeev-Popov operator in (3.14) becomes problematic. The horizon also needs special care during the inversion. It is essentially a zero-over-zero problem, needing a well-define regularization, something which is also not yet fully formally under control. Finally, note that outside the first Gribov region the Faddeev-Popov determinant gets a sign, depending on the number of negative eigenvalues.

Unfortunately, a unique, method-independent prescription how to effectively implement this restriction to the first Gribov region explicitly has not yet been constructed. There are, however, a number of possibilities, which have been explored.

E. g. one proposal for how to implement this restriction using additional ghost fields, and thus in a similar way as in perturbation theory, has been made by approximating

[32]It may be possible to loose a measure zero of gauge orbits. This is not studied in any depth so far.

the θ-function by a δ-function with the argument that in a high-dimensional space only the boundary contains an appreciable part of the volume. This generates the so-called Zwanziger Lagrangian. However, due to subtleties related to the definition of the step-function it is not yet proven that this is a valid procedure, though it has many interesting properties. Furthermore, no Gribov copy, or any gauge copy in general, is preferred compared to another. It would thus be completely legitimate to always chose the innermost Gribov copy for each gauge orbit. If (almost) all gauge orbits have a representative away from the Gribov horizon, this would yield distinctively different results for gauge-dependent quantities, e. g. the expectation value of the lowest Faddeev-Popov eigenvalue. Thus, such a replacement is already implementing a certain selection of Gribov copies, and thus corresponds to an extended gauge-fixing procedure. This is completely correct, provided (almost) all gauge orbits have Gribov copies off the Gribov horizon. Though not proven, this appears very likely.

After restricting to the first Gribov region, the remainders of the residual gauge orbits still possess a large number of Gribov copies. This set will also be denoted as the residual gauge orbit in the following, to avoid the term residual of the residual gauge orbit. In fact, in an infinite volume this number is likely infinite, and in a finite volume V it appears to be a rapidly rising function of V. Actually, counting Gribov copies is in practice a non-trivial problem, since two Gribov copies are different if and only if they differ at least at one space-time point after factorizing all possible global gauge transformations and all space-time transformations. This implies that for the decision whether two representatives of a gauge orbit are identical or Gribov copies, it is required to compare their field values at every space-time point[33]. It is also in general non-trivial how to find all Gribov copies, so that they can be counted[34].

Once more, it should be noted that one Gribov copy has no intrinsic difference compared to another Gribov copy, since they are physically equivalent. Thus any choice of a Gribov copy to represent the residual gauge orbit is equally acceptable. This is nicely illustrated by using stochastic quantization. Stochastic quantization is an alternative, but equivalent, formulation, of the path integral as a stochastic process, where additional dynamics occur in an additional, fictitious time. The equilibrated results for correlation functions are then the ordinary correlation functions. In this approach it is found that

[33]It appears that Gribov copies differ from each other over some large domain, so in practice already a coarse search can yield that two candidates are different. However, to ensure that they are the same requires a check of the whole space-time point by point.

[34]It should be noted that, though two dimensions has trivial dynamics, gauge fixing has the same subtleties as in higher dimensions. Two-dimensional Yang-Mills theory is therefore an ideal laboratory to study these issues without the obscuring dynamics.

there is no stochastic force acting along a gauge orbit, and thus in the stochastic equilibration process no point on a gauge orbit is preferred. This, of course, is just in disguise the problems encountered when defining the path integral, which require to introduce a gauge condition in the first place.

As stated, the residual gauge orbits inside the first Gribov horizon possess further Gribov copies. It is therefore necessary to specify a gauge further. There are two strategies mainly in use currently for that purpose. The first method is essentially a stochastic approach. In this case, instead of specifying conditions for selecting a Gribov copy, a random Gribov copy is chosen for each residual gauge orbit. This prescription, termed the minimal Landau gauge, therefore averages over Gribov-copy-dependent properties when calculating correlation functions. Assuming the choice to be ergodic, unbiased, and well-behaved, this implies that this prescription is equivalent to averaging over the residual gauge orbit. However, a constructive prescription how to make this choice in a path integral formulation is only developing. Precise definitions of this gauge therefore exist only as operational definitions in terms of algorithms in lattice gauge theory. The second approach attempts to characterize Gribov copies and make a choice based on these characteristics. One possibility will be discussed as an example.

The central element of all operational definitions of the Landau gauge is the fact that any Gribov copy in the first Gribov region maximizes the functional

$$\begin{aligned} F[A] &= 1 - \frac{1}{V}\int d^dx A^a_\mu A^a_\mu \\ \langle F[A]\rangle &= 1 - \frac{N_g}{2^d\pi^{d/2}\Gamma\left(1+\frac{d}{2}\right)V}\int dp p^{d-1} D^{aa}_{\mu\mu}(p) \\ D^{ab}_{\mu\nu} &= \langle A^a_\mu A^b_\nu\rangle, \end{aligned} \tag{5.79}$$

on each configuration, where $D^{ab}_{\mu\nu}$ is again the gluon propagator. This implies that this gauge minimizes the integrated weight of the gluon propagator. That this is indeed satisfying the Landau gauge conditions follows from the fact that the first derivative of (5.79) is the Landau gauge condition, and the Hessian is the Faddeev-Popov operator. If any given algorithm finds one of all the maxima with equal probability, it would be a faithful representation of the distribution along the residual gauge orbit, and also be ergodic. The fact that there are multiple Gribov copies inside the first Gribov region translates into the statement that the functional (5.79) has multiple maxima.

An alternative way to choose a representative on the residual gauge orbit is the absolute Landau gauge, which makes a very definite choice rather than a random choice. This gauge choice is derived from the following observation. The functional (5.79) has, up to topological identifications, a unique absolute maximum. The resulting set of absolute

maxima, called the fundamental modular domain or region is by definition embedded in the first Gribov region, and includes the origin. It is less trivial to show that it is also convex and bounded, and thus connected. It can furthermore be shown that part of the boundary of the fundamental modular domain coincides with the Gribov horizon in the thermodynamic limit only. All possibly remaining degenerate absolute minima are on the boundary. This boundary has actually a quite rough structure, including wedge singularities, and topological configurations, like e. g. instantons, are located there. By construction, in this region the gluon propagator has its least integrated weight.

Based on this observation, the absolute Landau gauge is defined as selecting the Gribov copy which belongs to the fundamental modular domain. This condition can be realized by either checking the absolute minimization of (5.79) explicitly or by the introduction of a suitable weight factor in the path integral. In case the residual gauge orbit has more than one Gribov copy on the boundary of the fundamental modular domain, again a random choice is made. It should be noted that if the thermodynamic arguments made before were correct, the absolute Landau gauge and the minimal Landau gauge will coincide in the thermodynamic limit.

Both these descriptions belong to a larger class of descriptions. It is obtained by rewriting the path integral in the expression (5.78) as

$$\langle\mathcal{O}\rangle = \lim_{\xi\to 0}\int \mathcal{D}A_\mu\mathcal{D}c\mathcal{D}\bar{c}\mathcal{O}(A_\mu,c,\bar{c})\Theta(-\partial_\mu D^{ab}_\mu)e^{-\int d^4x\mathcal{L}_g}w(A_\mu,c,\bar{c}),$$

where w is an appropriately chosen weight functions, which includes a normalization such that any observable remains unchanged. The minimal Landau gauge corresponds to the choice $w=1$, i. e. averaging over the first Gribov region with a flat weight. The absolute Landau gauge takes the form

$$w_2 = \exp\left(\mathcal{N}_2 - \frac{\lambda_2}{V}\int d^dx A^a_\mu A^a_\mu\right). \tag{5.80}$$

where the $\mathcal{N}_i$ are appropriately chosen normalizations and the limit of $\lambda_2\to\infty$ has to be taken. The opposite limit $\lambda_2\to 0$ recovers the minimal Landau gauge. These weights can also depend on other fields. Another possibility discussed in the literature is

$$w_1 = \exp\left(\mathcal{N}_1 + \frac{\lambda_1}{V}\int d^dx d^dy \partial^x_\mu\bar{c}^a(x)\partial^y_\mu c^a(y)\right) \tag{5.81}$$

involving the ghost fields. It is important to note that all these gauges include non-local information, both from the Θ function, as well as from the integrals. In a sense, minimal Landau gauge is special, as there all non-localities stem from the Θ function alone, another justification of its name.

In principle, it would be possible to average, in a well-defined way, over the Gribov regions to obtain also non-perturbatively a well-defined gauge, similar to, e. g., covariant gauges in perturbation theory. However, such Hirschfeld gauges induce significant cancellations. The reason is that the determinant of the Faddeev-Popov operator can now have either signs and is part of the weight factor when the ghosts are integrated out. It therefore poses a sign problem. Hence, no practical implementation has been constructed so far, but only conceptually developed. Thus, the alternative to select by some prescription a single representative or a smaller subset for each gauge orbit, which therefore satisfies further constraints as described above is more popular. This gauge has, however, a conceptually importance to be discussed next.

5.5.5.3 Non-perturbative BRST symmetry

To understand this importance, it is necessary to interject a few more words on the concept of broken symmetries.

A symmetry of a classical system, i. e. of the Lagrangian, which remains unbroken in the quantization process can still be broken dynamically, i. e. spontaneously. To describe this, in the following the language of correlation functions will again be used. In terms of correlation functions, a symmetry is unbroken, if under a symmetry transformation of the Lagrangian for fields ϕ_i

$$\phi_i \to \phi_i + \delta\phi_i \tag{5.82}$$

all correlation functions remain invariant as well. This implies that the correlation functions of operators not invariant under (5.82), e. g. $\langle\phi_i\rangle$, have to vanish identical. If these conditions are not fulfilled, the symmetry is broken.

In case of a gauge theory, it is necessary to differentiate between gauge-invariant and non-gauge-invariant correlation functions. If only gauge-dependent correlation functions show a behavior indicating the breakdown of a symmetry, this cannot have any measurable consequences. If the affected symmetry is a global part of the gauge symmetry[35], it can be argued that this is a mere artifact of the description, and the symmetry is, in fact, intact. This is the combination of observations which will be used here.

Before continuing, it is worthwhile to discuss the question of observing a broken symmetry. Naively, when just performing the path integral, all correlation functions non-invariant under some global symmetry will always vanish[36]. The trivial reason is just that no direc-

[35]Any local part cannot be broken anyhow. This can be proven exactly, and is known as Elitzur's theorem. In fact, explicit breaking would yield a gauge anomaly, and thus observable quantities would depend on the gauge, which is not desirable for a theory describing physics.

[36]The fact that that does not seem to be the case in numerical lattice simulations is just an artifact of

tion of the global symmetry is preferred without an explicit breaking. This can only be remedied by introducing an explicit breaking when determining the correlation functions, and performing the limit of zero explicit breaking afterwards. This is done, e. g., by the introduction of an external field.

Put it otherwise, spontaneous breaking is observed by preventing the system to reach all possible targets of a symmetry transformations by an external field. As a consequence, the correlation functions are no longer averages over all field configurations linked by a symmetry transformation, and therefore no longer invariant under such transformations.

In fact, in an experimental observation an absolute direction is never observed. E. g., take a magnet's magnetization. The orientation is measured relative to a reference orientation, e. g., compared to a different magnet. This reference orientation plays the role of the explicit breaking. Thus, an equivalent way of observing the breaking of a symmetry is to use operators invariant under a global symmetry transformation, but measuring the relative orientation of two operators not being separately invariant. One such operator for a magnet would be the averaged magnetization correlator,

$$\begin{aligned} C_{MM} &= \langle MM \rangle \\ M &= \sum_i s_i \end{aligned} \tag{5.83}$$

where the s_i are the local spins, and thus M is the configuration-wise polarization. Since the latter is only non-zero in the broken phase, the correlator can only then be non-zero.

The same concept together with the Hirschfeld gauge can be used to recover[37] a well-defined non-perturbative BRST symmetry, obeying the same algebra (4.12-4.14) as the perturbative version. The BRST transformations (4.7-4.10) are, in fact, just an ordinary gauge transformation for the gluon fields. This becomes evident when the ghost fields are integrated out. Leaving quarks aside, the integrand of the path integral only involves the gluon fields. The BRST symmetry is not changing physical observables. Thus, the only non-trivial action can be a gauge transformation. The explicit form after integrating out the ghosts is, even perturbatively, very complicated, and non-local. What happens is essentially just a gauge transformation between the different gauge copies satisfying the covariant gauge condition. This implies that the Gaussian weight function will also change such as to alter the weight of the gauge copy appropriately. Furthermore, in the Landau

the employed local algorithms, which fail to perform the full average over field configuration space in this case. The problem is that when starting in a certain sector of the theory, the update cycles often stay within this starting sector.

[37]The following is not mathematical rigorous, but there are possibilities, using lattice gauge theory, to make it much more so.

gauge case the perturbative BRST transformation can only be the identity transformation, as there is one and only one configuration perturbatively which satisfies the Landau gauge condition. This is especially so, as the BRST transformation does not change the gauge parameter.

Beyond perturbation theory, the situation changes by the appearance of Gribov copies. Especially, since there is now more than one gauge copy satisfying the Landau gauge condition, the BRST transformation is no longer an identity transformation in Landau gauge.

Now, concentrate again on Landau gauge. By construction, any BRST transformation will not change the Landau gauge condition, because of the anti-ghost equation of motion. This implies that BRST transformations mediate between different Landau-gauge Gribov copies. If an extended gauge condition, like minimal Landau gauge, selects only a subset of Gribov copies, it is possible that a BRST transformation leads out of this set, and thus BRST symmetry appears broken: The correlation functions are no longer invariant. However, this breaking is again introduced by an additional gauge condition, which specializes the Landau gauge further to the minimal Landau gauge.

To restore invariance, it is necessary to choose a gauge which is respecting BRST. This requires to include all[38] possible targets of a BRST transformation, i. e. all Gribov copies. But this is just a Hirschfeld-Landau type gauge. By averaging over all Gribov copies, the correlation functions are also averages over all possible BRST-transformed versions of them. Thus, they will be invariant under BRST transformations. Thus BRST symmetry can be regained just as any other global (gauge) symmetry. Whether there exist relative alignment operators and what their interpretation would be is an interesting question, in particular whether they show the existence of broken and unbroken phases, and if yes, under which circumstances. This is not yet known.

The bottom line of the previous discussion is that the realization of the Kugo-Ojima construction may be tied to the choice of a particular type of Landau gauge beyond perturbation theory. In this gauge, the Hirschfeld(-Landau) gauge, a non-perturbative BRST symmetry might be present which has the same algebra as the perturbative one. Thus, the Kugo-Ojima construction, essentially algebraic in nature, can go through unaltered. The original Kugo-Ojima construction was formulated also for arbitrary covariant gauges. However, similar insights as presented here for Landau gauge are not available yet for covariant gauges. It is thus unclear, whether this construction holds beyond Landau gauge, though there is no hint that this should not be the case.

[38]It would of course be interesting to know if there exist any subsets of Gribov copies such that BRST would remain intact within these subsets. But this is unknown.

This leaves the question what occurs in the other non-perturbative completions of the Landau gauge. That is not yet well understood. However, the following argument may give a guide-line. If the condition restricts to a set of Gribov copies, like the minimal Landau gauge, there is still a residual gauge symmetry which links this set. To the best of our current knowledge, this symmetry will likely not be a local transformation, i. e. its transformation rules will involve integrals. This makes any construction similar to the Kugo-Ojima one at least complicated, though likely not impossible. The only exception may be those cases where really a single gauge copy is identified, like in the case of the fundamental modular region. How to proceed in this case is yet unclear.

5.5.6 Confinement and topology

The Kugo-Ojima construction focuses on the state space. Even if it can be proven to be correct, it will not explain the string tension of Yang-Mills theory nor the intermediate distance string tension of QCD.

The opposite extreme is a perspective based on the topological excitations. They make no statement about the state space, and especially not about gluons. However, they can make a statement about the Wilson criterion.

The simplest example of how topological excitations can explain the origin of a string tension are vortices. There is unfortunately no simple method to show the following, and rather the insight is gained by either lattice simulations, or the numerical simulations of effective models, i. e. models only containing center vortices. The upshot is that as long as the number of vortices piercing the area enclosed by a Wilson line is proportional to the area, the free energy, and thus the Wilson potential, shows an area law. This will happen, if the vortices percolate in space-time, which has been observed in both kinds of numerical simulations. Conversely, if the vortices are prevented from a percolation, the string tension vanishes.

Similar results can be obtained with the world-lines of monopoles. If sufficiently many world-lines cross the area enclosed by a Wilson line a string tension arises. Instantons are somewhat different. If enough instanton events occur on the sheet spanned by the Wilson line then also a string tension would arise. However, in contrast to vortices and monopoles, the occurrence rate of instantons in full Yang-Mills theory, or QCD, is far too small to achieve this effect.

It should be kept in mind that all of these topological configurations are interrelated, and there is interaction between them. Furthermore, since the extraction of these entities from the full theory is not without ambiguities, and largely gauge-dependent, all of these results are more of an approximate nature yet.

5.6 Anomalies

Anomalies are quite interesting features of quantum theory, and play an important role in the low-energy physics of QCD, though the name itself is somewhat of a misnomer.

An anomaly is that some symmetry, which is present on the classical level, is not present when considering the quantum theory. The symmetry is said to be broken by quantum effects. Generically, this occurs if the action of a theory is invariant under a symmetry, but the measure of the path integral is not. While the breaking of a global symmetry by an anomaly is no conceptual problem, the breaking of a gauge symmetry would make a theory ill-defined, and must therefore be avoided. It is therefore necessary to show that QCD does not develop an anomaly of the latter type if it already develops one of the former.

5.6.1 Global anomalies

The most important global anomaly is the breaking of dilatation symmetry. This symmetry corresponds to rescaling all dimensionful quantities, e. g. $x \to \lambda x$. Maxwell theory, massless QED, Yang-Mills theory, and massless QCD are all invariant under such a rescaling at the classical level. This is no longer the case at the quantum level. By dimensional transmutation, surfacing in the renormalization process, an explicit scale is introduced into the theory, and thereby the quantum theory is no longer scale-invariant. Such global anomalies have very direct consequences. E. g., this dilatation anomaly leads to the fact that the gluon remains massless, at the perturbative level.

5.6.2 Axial anomaly

Another example is the so-called axial anomaly, which occurs due to the breaking of the global axial symmetry of fermions. A consequence of it is the anomalously large η' mass. While the dilatation anomaly is quite obvious, the chiral anomaly is much more subtle, and therefore deserves some more discussion.

5.6.2.1 Classical level

To prepare for this, it is worthwhile to consider the situation as it would be without anomalies, i. e. at the classical level. For this purpose, start with a gauge theory with fermions ψ being in some representation R of the gauge Lie group G with generators T and gauge fields in the adjoint representation. The fermionic part of the Lagrangian is

then given by

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu(\partial_\mu - igT^a A^a_\mu) - m)\psi = \bar{\psi}(i\gamma_\mu D^\mu - m)\psi$$

from which the Dirac equation

$$(i\gamma_\mu D^\mu - m)\psi = 0$$

follows as the equation of motion, and likewise for the anti-fermion.

The current carrying the charge is then

$$j^a_\mu = \bar{\psi}\gamma_\mu T^a \psi.$$

Due to the chiral symmetry, there is also a corresponding axial current

$$j^{5a}_\mu = \bar{\psi}\gamma_5\gamma_\mu T^a \psi.$$

In addition, there are also the singlet currents

$$\begin{aligned} j_\mu &= \bar{\psi}\gamma_\mu\psi \\ j^5_\mu &= \bar{\psi}\gamma_5\gamma_\mu\psi, \end{aligned}$$

which corresponds to the fermion current and the axial current.

Naively, the divergences of these equations can be calculated using the Dirac equation.

$$\begin{aligned} \partial^\mu j^a_\mu &= -i\bar{\psi}(g\tau^b\gamma_\mu A^\mu_b - m)\tau^a\psi - i\bar{\psi}\tau^a(-g\tau^b\gamma_\mu A^\mu_b + m)\psi \\ &= ig\bar{\psi}\left[\tau^a, \tau^b\right]\gamma_\mu A^\mu_b\psi = -gf^{abc}A^\mu_b\bar{\psi}\gamma_\mu\tau_c\psi = -gf^{ab}_{\ \ c}A^\mu_b j^c_\mu. \end{aligned}$$

This implies that the color current is not observed, as long as the current is gauged. For a non-gauge current, like a flavor current, g vanishes, and the current is conserved.

This is not surprising, as a non-Abelian gauge theory has no gauge-invariant charge. However, the current is a gauge-vector, and therefore covariantly conserved

$$D^{ab}_\mu j^\mu_b = 0. \tag{5.84}$$

In the same way, it is possible to calculate the situation of the axial color current. Because of the commutation relation between γ matrices, the result is

$$D^{ab}_\mu j^\mu_b = 2im\bar{\psi}\gamma_5\tau^a\psi = 2mip^a, \tag{5.85}$$

Here, p is the pseudo-scalar density, and not a momentum component. Thus, even in a non-gauge theory this current is only conserved for fermions without a mass term in the Lagrangian.

The calculations for the singlet current is simpler, and yields

$$\begin{aligned}\partial_\mu j^\mu &= 0\\ \partial^\mu j^5_\mu &= 2im\bar\psi\gamma_5\psi = 2imp^0.\end{aligned}$$

Hence, the number of fermion is, a expected, a conserved current. The axial current is only conserved for massless fermions. This is the result that chiral symmetry gets explicitly broken, already classically, by a mass-term.

5.6.2.2 One-loop violation

So far, this was the conservation at the classical level, which already requires the fermions to be massless. At the quantum level, this result is expressed by Ward-identities. In particular, take Ward identities for correlation functions of the form

$$T^{ijk}_{\mu\nu\rho} = \langle T j^i_\mu j^j_\nu j^k_\rho\rangle,$$

where i, j, and k can take the values V, A, and P, which require to replace the j by j^a, j^{5a}, and p^a, respectively, and the Lorenz index is dropped in the last case. Calculating the corresponding Ward identities for a local chiral transformation

$$\begin{aligned}\psi' &= e^{i\beta(x)\gamma_5}\psi(x)\\ \bar\psi' &= \bar\psi e^{i\beta(x)\gamma_5}\end{aligned}$$

yields the expressions

$$\partial^\mu_x T^{VVA}_{\mu\nu\rho}(x,y,z) = \partial^\nu_y T^{VVA}_{\mu\nu\rho}(x,y,z) = 0 \tag{5.86}$$

$$\partial^\rho_z T^{VVA}_{\mu\nu\rho}(x,y,z) = 2mT^{VVP}_{\mu\nu}(x,y,z), \tag{5.87}$$

directly implementing the relations (5.84) and (5.85). This is what should happen, if there would be no anomalies.

To check this, it is possible to calculate the leading-order perturbative correction. Since only fermion fields appear in the vacuum expectation value, this is a vacuum triangle graph, and the coupling is to external currents. In fact, it does not matter at this point whether the external currents are gauged or non-gauged, since to this order this only alters the presence or absence of color matrices at the external vertices. The only relevant part of the external vertices is their Dirac structure.

Evaluating all the Wick contractions yields two Feynman diagrams, which translate to

$$\begin{aligned}&T^{VVA}_{\mu\nu\rho}(p_1,p_2,p_3=-p_1-p_2) = \\ &-i^3\int\frac{d^4k}{(2\pi)^4}\Big(\mathrm{tr}\gamma_\mu(\gamma_\alpha k^\alpha - m)^{-1}\gamma_\nu(\gamma^\beta k_\beta - \gamma_\beta p_2^\beta - m)^{-1}\gamma_\rho\gamma_5(\gamma_\gamma k^\gamma + \gamma_\gamma p_1^\gamma - m)^{-1}\\ &+\mathrm{tr}\gamma_\nu(\gamma_\alpha k^\alpha - m)^{-1}\gamma_\mu(\gamma^\beta k_\beta - \gamma_\beta p_1^\beta - m)^{-1}\gamma_\rho\gamma_5(\gamma_\gamma k^\gamma + \gamma_\gamma p_2^\gamma - m)^{-1}\Big).\end{aligned} \tag{5.88}$$

This expression is linearly divergent. One of the most important points in anomalies, and in quantum field theories in general, is that the result is independent of the regulator employed. This will be discussed later how to show this. Here, it permits to use a Pauli-Villar regulator with a mass M, which is technically more simple than other possibilities. Using dimensional regularization makes the result subtle, as it depends on the way the matrix γ_5 is analytically continued. This problem will therefore be avoided here.

To test the vector Ward identity, the expression can be multiplied with p_1^μ. To simplify the so obtained expression it is useful to employ

$$\gamma_\mu p_1^\mu = -(\gamma_\mu k^\mu - \gamma_\mu p_1^\mu - m) + (\gamma_\mu k^\mu - m),$$

yielding

$$\begin{aligned}
&p_1^\mu T^{VVA}_{\mu\nu\rho}(p_1, p_2, p_3 = -p_1 - p_2) = \\
&-i^3 \int \frac{d^4k}{(2\pi)^4} \Big(\mathrm{tr} - (\gamma_\alpha k^\alpha - m)^{-1} \gamma_\nu (\gamma^\beta k_\beta - \gamma_\beta p_2^\beta - m)^{-1} \gamma_\rho \gamma_5 \\
&\mathrm{tr}(\gamma_\gamma k^\gamma + \gamma_\gamma p_1^\gamma - m)^{-1} \gamma_\nu (\gamma^\beta k_\beta - \gamma_\beta p_2^\beta - m)^{-1} \gamma_\rho \gamma_5 \\
&+ \mathrm{tr}(\gamma_\gamma k^\gamma + \gamma_\gamma p_2^\gamma - m)^{-1} \gamma_\nu (\gamma_\alpha k^\alpha - m)^{-1} \gamma_\rho \gamma_5 \\
&+ \mathrm{tr} - (\gamma_\gamma k^\gamma + \gamma_\gamma p_2^\gamma - m)^{-1} \gamma_\nu (\gamma^\beta k_\beta - \gamma_\beta p_1^\beta - m)^{-1} \gamma_\rho \gamma_5 + (m \to M) \Big) .
\end{aligned} \tag{5.89}$$

This rather length expression is now a finite integral. It is therefore permissible to reshuffle the momenta like $k \to k + p_2$ in the first term and $k \to k + p_2 - p_1$ in the second term. Then, the first and third and second and fourth term cancel each other, and likewise this happens for the regulator. Thus, the vector Ward identity is fulfilled. The result for the second identity in (5.86) works in the same way.

The situation changes drastically for the axial Ward identity (5.87). The expression (5.88) is still divergent, so before doing anything, it will again be regulated using a Pauli-Villar regulator, to make it well-defined. To evaluate (5.87) requires multiplication with $p_3 = -p_1 - p_2$, which can rewritten as

$$\begin{aligned}
\gamma_\mu p_3^\mu \gamma_5 &= (\gamma_\mu k^\mu - \gamma_\mu p_2^\mu - m)\gamma_5 + \gamma_5(\gamma_\mu k^\mu + \gamma_\mu p_1^\mu - m) + 2m\gamma_5 \\
&= (\gamma_\mu k^\mu - \gamma_\mu p_1^\mu - m)\gamma_5 + \gamma_5(\gamma_\mu k^\mu + \gamma_\mu p_2^\mu - m) + 2m\gamma_5.
\end{aligned}$$

This yields

$$
\begin{aligned}
&p_3^\rho T^{VVA}_{\mu\nu\rho}(p_1,p_2,p_3=-p_1-p_2) = 2i\int\frac{d^4k}{(2\pi)^4}\\
&\Big(m\mathrm{tr}\left(\gamma_\mu(k_\alpha\gamma^\alpha-m)^{-1}\gamma_\nu(\gamma_\beta k^\beta-\gamma_\beta p_2^\beta-m)^{-1}\gamma_5(\gamma_\gamma k^\gamma+p_1^\gamma\gamma_\gamma-m)^{-1}\right)\\
&m\mathrm{tr}\left(\gamma_\mu(k_\alpha\gamma^\alpha-m)^{-1}\gamma_\nu(\gamma_\beta k^\beta-\gamma_\beta p_1^\beta-m)^{-1}\gamma_5(\gamma_\gamma k^\gamma+p_2^\gamma\gamma_\gamma-m)^{-1}\right)\\
&M\mathrm{tr}\left(\gamma_\mu(k_\alpha\gamma^\alpha-M)^{-1}\gamma_\nu(\gamma_\beta k^\beta-\gamma_\beta p_2^\beta-M)^{-1}\gamma_5(\gamma_\gamma k^\gamma+p_1^\gamma\gamma_\gamma-M)^{-1}\right)\\
&M\mathrm{tr}\left(\gamma_\mu(k_\alpha\gamma^\alpha-M)^{-1}\gamma_\nu(\gamma_\beta k^\beta-\gamma_\beta p_1^\beta-M)^{-1}\gamma_5(\gamma_\gamma k^\gamma+p_2^\gamma\gamma_\gamma-M)^{-1}\right)\Big)
\end{aligned}
$$

There are two remarkable facts to be observed. The first is that this expression is finite. The projection with p_3 drops out the divergent terms. This can be seen using the Dirac matrix identity

$$\mathrm{tr}\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma\gamma_5 = -4i\epsilon_{\mu\nu\rho\sigma}. \tag{5.90}$$

Because of the anti-symmetry of the ϵ-symbol, any term containing two or more factors of k vanishes. Hence, the numerator is reduced by two powers of k, making the integral finite. This did not work in (5.89) as there one index less was uncontracted. However, the regulator still had to be present in the first place to make this projection well-defined. The second is that this expression, except for the regulator, is identical to T^{VVP} up to a factor of m, which is obtained by replacing $\gamma_\rho\gamma_5$ in (5.88).

The term involving the regulator can then be calculated, as when removing the regulator in the end, the external momenta and masses can always be neglected, and the integral becomes a simple tadpole integral. The final result is thus

$$
\begin{aligned}
ip_3^\rho T^{VVA}_{\mu\nu\rho}(p_1,p_2) &= 2miT^{VVP}_{\mu\nu}(p_1,p_2) + \lim_{M\to\infty} 8iM^2\epsilon_{\mu\nu\rho\sigma}p_\rho^1 p_\sigma^2 \times \frac{i}{16\pi^2}\frac{-1}{2M^2}\\
&= 2miT^{VVP}_{\mu\nu}(p_1,p_2) + \frac{1}{2\pi^2}\epsilon^{\mu\nu\rho\sigma}p_\rho^1 p_\sigma^2
\end{aligned} \tag{5.91}
$$

Thus, the Ward identity (5.87) is violated. The anomaly is both finite and independent of the masses of the involved particles. It is also independent of the structure of the external interaction, except for its Lorentz structure. The only thing changes is the appearance of corresponding pre-factor a^{abc} of the coupling matrices T^a in charge space, which turn out to be

$$a^{abc} = \frac{1}{2}\mathrm{tr}\left(\left\{T^a,T^b\right\}T^c\right), \tag{5.92}$$

a result which will become significant later. This is not the only anomaly, and a similar result holds for the case of three axial currents.

Without proof, it should be noted here that there is still a certain regulator dependency. It is possible by symmetries to add a finite term of form $C\epsilon_{\mu\nu\rho\sigma}(p_1-p_2)^\sigma$ to the counter-term

in (5.89). Though C can be tuned to absorb the anomaly, this term will also contribute to the vector identities, and induce there an anomaly for $C \neq 0$. Thus, it is only possible to shift the anomaly around, without removing it.

The most well-known consequence of this anomaly is the decay of a neutral pion into two photons. This is precisely of the type investigate here, where the photons play the role of the vector currents. The axial current is related to the pion field by a QCD relation

$$\partial^\mu j^a_\mu = \frac{f_\pi}{\sqrt{2}} M_\pi^2 \pi^a, \tag{5.93}$$

where a is an isospin index, counting the three pions, $a = 0, \pm$, where only $a = 0$ is relevant because of charge conservation. Since there are no massless hadrons, there can be no pole in the corresponding amplitude T^{VVA}, and thus the product with p_ρ has to vanish. As a consequence, the amplitude T^{VVP}, describing the transition, would vanish as well, because of the Ward identity, and therefore the pion would usually not decay into two photons, if at rest. However, due to the anomaly, this is not necessary, as the anomaly can balance the Ward identity. Hence, the pion at rest can decay into two photons, due to the anomaly, a process indeed observed in experiment.

It should be noted that (5.93) is actually not the original expression of the anomaly (5.91), as this is a statement involving expectation values. Its formulation using fields is known as the hypothesis of partially conserved axial current (PCAC), as it does not immediately follow. However, it is found to hold acceptably. This is highly non-trivial. After all, the left-hand side is from the electroweak interactions, while the right-hand side is a strong-interaction object.

By taking expectation values in different states of (5.93), it is possible to establish various conjectures for low-energy physics. E. g. taking it between a proton and a neutron state, the left-hand-side mediates essentially a β-decay. The right-handed-side then corresponds to the transformation of a neutron to a proton under pion exchange. Assuming that the latter is dominated by the tree-level process, and decomposing everything in suitable Lorentz tensors and form factors, yields the Goldberger-Treiman relation

$$f_\pi g_{\pi NN} = m_N g_A(0),$$

where $g_{\pi NN}$ is the coupling constant between nucleons and the pion, m_N is the (averaged) nucleon mass, and $g_A(0)$ is the axial form-factor of the nucleon at zero momentum transfer. This relation is experimentally found to hold at the 10% level, a typical value for such derivations.

5.6.3 Anomalies and WTIs

To understand the origin of the anomalies, it is important to remember how Ward identities are obtained in general, repeating some aspects of section 4.4. Any well-defined symmetry transformation should leave the partition function unchanged, i. e.

$$0 = \delta Z = \delta \int \mathcal{D}\phi e^{iS + i\int d^4x j\phi}, \tag{5.94}$$

where ϕ is for simplicity a non-Grassmann field, which changes under the transformation as $\phi \to \phi + \epsilon f(\phi, x)$, with f some arbitrary function and ϵ infinitesimal. Performing the variation yields

$$0 = \int \mathcal{D}\phi e^{iS + i\int d^4x j\phi} \int d^4x \left(i\left(\frac{\delta S}{\delta \phi} + j\right) f + \frac{\delta f}{\delta \phi}\right), \tag{5.95}$$

where the first two terms come from the exponent. At the classical level, the source term vanishes, and the derivative of the action just gives the equations of motion, yielding the classical Ward identities. The third term is new in the quantum theory, and gives the contribution of the Jacobian,

$$\det \frac{\delta(\phi + \epsilon f)}{\delta \phi} = \det\left(1 + \epsilon \frac{\delta f}{\delta \phi}\right) \approx 1 + \epsilon \frac{\delta f}{\delta \phi} + \mathcal{O}(\epsilon^2).$$

This is a genuine quantum contribution. It will be the source of the anomaly. Here it also becomes evident that the term anomaly is actually a misnomer. There is nothing anomalous about it. It is just a quantum effect.

To obtain Ward identities from (5.95), it is sufficient to derive with respect to the source some number of times, and then set the sources to zero at the end, yielding

$$0 = \left\langle T\Pi_l \phi_l \frac{\delta f}{\delta \phi}\right\rangle + i\left\langle T\Pi_l \phi_l \frac{\delta S}{\delta \phi} f\right\rangle + \sum_k \left\langle T\Pi_{l<k}\phi_l f \Pi_{m>k}\phi_m\right\rangle. \tag{5.96}$$

In this way an anomaly surfaces in Ward identities in the full quantum theory. This also shows that an anomaly is not a perturbative effect, since this is an exact result. However, it is still possible that the Jacobian is actually one, and a deviation from one in the one-loop calculation is just an artifact of perturbation theory.

5.6.4 Full expression for the anomaly

To check this, rotate first to Euclidean time, by replacing $t \to it$ and correspondingly in all covariant quantities the time component by i-times the time component and in all

contravariant quantities the time components by $-i$-times the time components. Then expand the fermion fields in orthonormal eigenfunctions ψ_n of the Dirac operator,[39]

$$\begin{aligned}\psi(x) &= \sum_n a_n \psi_n(x) \\ \bar{\psi}(x) &= \sum_n \psi_n^\dagger(x) \bar{b}_n,\end{aligned}$$

which satisfy

$$i\gamma_\mu D^\mu \psi_n = \lambda_n \psi_n \tag{5.97}$$

$$-i\gamma_\mu D^\mu \psi_n^\dagger = \lambda_n \psi_n^\dagger. \tag{5.98}$$

This permits to rewrite the path integral as an infinite product of integrations over the coefficients,

$$\mathcal{D}\psi\mathcal{D}\bar{\psi} = \Pi_m da_m d\bar{b}_m, \tag{5.99}$$

keeping in mind that these differentials are Grassmannian.

Now, a (formally) local chiral transformation $\beta(x)$

$$\psi \to e^{i\beta(x)\gamma_5}\psi,$$

then corresponds to a linear transformation of the coefficients

$$a_m \to C_{mn} a_n = a'_n,$$

which yields the Jacobian

$$\Pi_m da'_m d\bar{b}'_m = \frac{1}{(\det C)^2} \Pi_m da_m d\bar{b}_m,$$

or, formally,

$$\mathcal{D}\psi'\mathcal{D}\bar{\psi}' = \frac{1}{(\det C)^2}\mathcal{D}\psi\mathcal{D}\bar{\psi}.$$

This determinant can be rewritten as

$$\frac{1}{(\det C)^2} = e^{-2\mathrm{tr}\ln C} = e^{-2\mathrm{tr}\delta C}, \tag{5.100}$$

[39]Note that in general ψ and $\psi^\dagger$ are independent variables in Euclidean space-time, and not related by † as in Minkowski space-time, as the degrees of freedom are differently distributed. This is the reason for different coefficients a_n and b_n rather than a relation by †. The † on ψ should therefore be regarded rather as an index than an operation, as is common in the literature, to have expressions which formally look the same in both Euclidean and Minkowski space-time.

where in the last equality it was assumed that β is infinitesimal, and thus $C = 1 + \delta C$ is close to one. In this case, δC can be evaluated starting from

$$a'_m \psi_m = (1 + i\beta\gamma_5) a_n \psi_n$$

which can be reduced using the orthonormality of the eigenstates of the Dirac equation to

$$a'_m = \int d^4x \psi^\dagger_m (1 + i\beta\gamma_5)\psi_n a_n = (1 + \delta c_{mn}) a_n. \tag{5.101}$$

Inserting this result into (5.100) yields for the Jacobian of the infinitesimal transformation

$$J = \exp\left(-2i \int d^4x \beta \psi^\dagger_m \gamma_5 \psi_m\right), \tag{5.102}$$

where the trace has been evaluated.

Unfortunately, the expression, as it stands, is ill-defined. It is necessary to regularize it. A useful possibility to make the expression well-defined is by replacing the trace over the eigenstates as

$$\psi^\dagger_m \gamma_5 \psi_m \to \lim_{\tau\to 0} \psi^\dagger_m \gamma_5 e^{-\lambda^2_m \tau} \psi_m, \tag{5.103}$$

where the limit has to be performed at the end of the calculation only. Expanding the Gaussian and using the relations (5.97-5.98), this expression can be rewritten as

$$\lim_{\tau\to 0} \psi^\dagger_m \gamma_5 e^{-\lambda^2_m \tau} \psi_m = \lim_{\tau\to 0} \mathrm{tr}\left(\gamma_5 e^{-\tau(\gamma_\mu D^\mu)^\dagger \gamma_\nu D^\nu}\right). \tag{5.104}$$

The exponential can be rewritten as

$$(\gamma_\mu D^\mu)^+ \gamma_\nu D^\nu = -D_\mu D^\mu + \frac{i}{4}\left[\gamma^\mu, \gamma^\nu\right] F^a_{\mu\nu} \tau_a. \tag{5.105}$$

The limit is still ill-defined. It is necessary to regularize the expression in a suitable way. This is achieved by the heat-kernel regularization.

For a differential operator, here given by $\Delta = (\gamma_\mu D^\mu)^\dagger \gamma_\nu D^\nu$, it is possible to define a heat-kernel as

$$(\partial_\tau + \Delta_x) G(x, y, \tau) = 0 \tag{5.106}$$

$$G(x, y, 0) = \delta(x - y). \tag{5.107}$$

Which is solved by the formal expression

$$G(x, y, \tau) = e^{-\Delta_x \tau} = \sum_m e^{-\tau\lambda_m} \psi^\dagger_m(y)\psi_m(x).$$

This is already the expression (5.104). Without proof, it can now be shown that this heat kernel can be expanded for small τ as

$$G(x,y,\tau) \to_{\tau\to 0} \frac{1}{(4\pi\tau)^2} \exp^{-\frac{(x-y)^2}{4\tau}} \sum_{j=0}^{\infty} a_j(x,y)\tau^j.$$

Inserting this expansion into (5.102) yields

$$\ln J = -2i \lim_{\tau\to 0} \frac{1}{(4\pi\tau)^2} \int d^4x\beta \sum_j \tau^j \mathrm{tr}\gamma_5 a_j.$$

For $\tau \to 0$, the first term does not contribute, as a_0 has to be equal to one because of the condition (5.107). Terms with $j > 2$ will be irrelevant, because of the powers of τ. This leaves only $j = 1$ and $j = 2$. For these terms follows from the requirement that the expansion (5.106) satisfies a descent equation

$$-\Delta a_{j-1} = j a_j.$$

Since $a_0 = 1$, a_1 can be obtained algebraically from (5.105). Since all resulting terms have at most two γ matrices, the trace will vanish. Similarly, for a_2 only those terms can contribute to the trace where at least four γ matrices appear, which implies only the term quadratic in $F_{\mu\nu}$ will contribute. Which is precisely what is necessary to cancel the pre-factor.

Thus, the remainder is just

$$J = \exp\left(-\frac{i}{32\pi^2}\int d^4x\beta\epsilon^{\mu\nu\rho\sigma}F^a_{\mu\nu}F_{a\rho\sigma}\right). \tag{5.108}$$

Hence, the Jacobian is non-trivial, and will contribute in the Ward identities (5.96). However, this is still a rather complicated expression, which does not yet look like the one-loop result.

That this coincides with the one-loop anomaly can be obtained by an explicit calculation. Since this was for the global case, take β to be constant. The integral can then be rewritten as

$$\int d^4x \mathrm{tr}\epsilon^{\mu\nu\rho\sigma}\partial_\mu\left(iA^a_\nu\partial_\rho A^a_\sigma + \frac{2}{3}f^{abc}A^a_\nu A^b_\rho A^c_\sigma\right). \tag{5.109}$$

Since the perturbative case was the Abelian case, the second term can be dropped. The first term is then for the global case just two external fields, e. g. playing the roles of the photon field in the pion decay, and two momenta in Fourier space, which, after relabeling, yield the desired one-loop expression. Hence, indeed the full and the one-loop anomaly coincide. In gauge theories there are also anomalies in box and pentagon graphs with an odd number of axial insertions, which are again one-loop exact.

To obtain the final result including all color factors requires an explicit calculation, inserting the Jacobian (5.108) into the Ward identity (5.96). This will yield (5.91) with (5.92) inserted

To see how this can lead to a problem, consider for a moment the case that left-handed and right-handed fermions are coupled differently. Due to the different sign of γ_5 in the corresponding projector, this will reemerge as a different sign of the anomaly, yielding

$$k^\rho T^{V^a V^b A^c}_{\mu\nu\rho}(p,q,k) = 2m T^{V^a V^b P^c}_{\mu\nu}(p,q,k) + \frac{\mathrm{tr}\left\{\tau^a_L,\tau^b_L\right\}\tau^c_L - \mathrm{tr}\left\{\tau^a_R,\tau^b_R\right\}\tau^c_R}{2}\frac{1}{3\pi^2}\epsilon_{\mu\nu\rho\sigma}p^\rho q^\sigma,$$

were L and R indicate the representation of the left-handed and right-handed fermions. As a consequence, the classical gauge symmetry is broken by the anomaly, and results will depend on the choice of gauge. This can be directly understood form this expression. The tensors carry color. Thus they will vanish, if there is no vector-vector-axial/pseudo-scalar coupling in QCD. This is the case, as all couplings are vector-vector-vector and vector-fermion-fermion. It is not possible to construct others from the elementary vertices. The last term vanishes, because in QCD left-handed and right-handed particles couple in the same way, though this is not true for the full standard model.

There is an interesting twist for the quantity making up the Jacobian

$$\frac{1}{64\pi^2}\int d^4x \epsilon^{\mu\nu\rho\sigma}F^a_{\mu\nu}F_{a\rho\sigma} = -\frac{i}{512\pi^4}\int d^4x \mathrm{tr}\epsilon^{\mu\nu\rho\sigma}\partial_\mu\left(iA^a_\nu\partial_\rho A^a_\sigma + \frac{2}{3}f^{abc}A^a_\nu A^b_\rho A^c_\sigma\right)$$

Evidently, this is again the topological charge (5.56)

Since this quantity was obtained from the chiral transformation properties of the fermions, it suggest itself that it is connected to properties of the Dirac operator, and this is indeed the case. This topological charge is equal to the difference of the number of the left-handed n^- and right-handed n^+ zero modes of the (necessarily in the present context massless) Dirac operator D_μ, $\gamma_\mu D^\mu\psi = 0$, called the index of the Dirac operator. This is the celebrated (Atiyah-Singer) index theorem.

To see this, note first that because γ_5 anti-commutes with the other γ_μ it follows that that for any eigenmode of the Dirac operator ψ_m to eigenvalue λ_m that

$$i\gamma_\mu D^\mu\gamma_5\psi_m = -i\gamma_5\gamma_\mu D^\mu\psi_m = -\lambda_m\gamma_5\psi_m.$$

Hence, every non-zero eigenmode is doubly degenerate, and therefore the index is the same if all eigenmodes are included.

Start with an expression for this difference,

$$n^+ - n^- = \int d^4x \sum_{m,\lambda_m=0}\psi^\dagger_m\gamma_5\psi_m.$$

The inserted γ_5 will guarantee the correct counting. It is possible to use a very similar trick as before when regularizing the sums when doing the path integral calculation in section 5.6.4. The additional eigenvalues can be added as

$$\int d^4x \sum_m \psi_m^\dagger \gamma_5 \psi_m e^{-\lambda_m^2 \tau},$$

as the γ_5 symmetry ensures that all added terms vanish. But this is precisely expression (5.103), and thus this will lead to the same result as in section 5.6.4. Thus, the final answer is

$$n^+ - n^- = k = \frac{1}{64\pi^2} \int d^4x \epsilon^{\mu\nu\rho\sigma} F^a_{\mu\nu} F_{a\rho\sigma}$$

Hence, the anomaly has a certain connection to the topology of the gauge-fields.

This is in as far remarkable as the topology of gauge fields is an intrinsic property of Yang-Mills theory, and thus existing without any fermions, and hence in anomaly-free theories. At the same time, anomalies also exist without gauge fields, e. g. in the form of global anomalies. They are tied to the path-integral measure for theories with fermions. It is the unique property of the covariant derivative in the form of the Dirac operator for fermions which ties both effects together in the presented way. Other realizations than minimal coupling will not have this property, or at least in a different way. This connection is therefore deeply ingrained in the Dirac operator of gauge theories.

The two presented forms of anomalies are not the only ones. However, these are the only ones relevant to QCD.

5.6.5 The Witten-Veniziano relation

5.7 Large-N_c

While strictly speaking not QCD, there is a certain class of deformations of QCD, which showed to be very useful. These are theories in which N_c or N_c and N_f are send to a very large number, in the extreme case infinity, while at the same time maintaining asymptotic freedom, the so-called large-N_c limit[40]. One important convention is that the large values of N_c are odd. This ensures that baryons, which are made up of N_c quarks, remain fermions.

In perturbation theory, this leads to drastic simplifications. Because the contractions of structure constants, generators, and performing of traces gives every diagram a particular

[40]There are several possibilities, some also playing with representation aspects. The different versions highlight different properties, but the basic philosophy remains the same.

prefactor of a power of N_c (and N_f), it is possible to order diagrams accordingly. At sufficiently large N_c, just those with the highest power in N_c will contribute, the others being sub-leading corrections. The surviving diagrams have two special features. One is that they are planar, i. e. their Feynman diagrams can be painted without crossing propagator lines. More importantly, in most versions of the large-N_c limit, diagrams containing quark loops become suppressed, just because there are only N_c quarks, while there are $N_c^2 - 1 \approx N_c^2$ gluons. Thus, the large-N_c limit becomes quenched. However, at the same time the coupling diminishes, since the β-function depends also on N_c. Hence, naively the theory would become non-interacting. To avoid this, the coupling is increased, in the simplest form such that $\lambda = g^2 N_c$, the so-called 't Hooft coupling, stays constant. The perturbative expansion then becomes an expansion in λ.

Of course, perturbation theory is not everything, and non-perturbative contributions remain even in this limit[41]. They are also necessary, such that mesons and baryons remain bound. But the baryons will consist out of a very large number of quarks, and thus will be essentially infinitely heavy, and therefore static. The mesons, which are still made from a quark and an anti-quark, remain light.

Using the methods described in the next chapter 6 it can be shown that hadronic interactions become also suppressed in the large-N_c limit. As a consequence, the hadrons become stable, and scattering processes trivial.

Though all of this sounds very different from QCD, several statements can be obtained, provided that $N_c = 3$ is not that different from the large-N_c limit. Many lattice simulations have shown that this statement is surprisingly good, and thus the large-N_c limit gives another view on why the quenched approximation is as good as it is. Hence, it is possible to calculate many quantities in an $1/N_c$ expansion, and the results are indeed often, though not always, in surprisingly good agreement with the full result.

[41] Actually, it can be shown that a free theory and an interacting theory can never be unitarily equivalent, and therefore the limit of the coupling constant to zero cannot be analytic, like it is assumed in perturbation theory. This is known as Haag's theorem. But one should keep in mind that this does not imply that non-perturbative contributions must be qualitatively relevant, nor quantitative important.

Chapter 6

Hadronic interactions

The discussion so far centered on static properties of QCD, e. g. the bound state spectrum. Though these are already substantially important, it is the dynamics of reactions involving hadrons which put our understanding of the theory to the test. This is even more important, as that hadrons play a twofold important role in modern day experiments. On the one hand, hadrons, especially protons, are used as projectiles in scattering experiments, like the LHC. On the other hand, since quarks are confined, hadrons signal the presence of quarks and gluons in final states. A dynamical first-principle calculation of hadronic interactions in a practical manner is something which so far exceeds our possibilities. It is only now that we start to understand how to deal with low-energy scattering of hadrons and their decays. Full control over a high-energy collision of hadrons with multiple hadrons in the final state is yet far beyond a (quasi-)exact treatment, being it either with lattice, functional, or similar methods. The currently most useful, and most successful, approach, is to treat the interactions of hadrons phenomenologically while the interaction of quarks and gluons at sufficiently high energy is treated perturbatively.

As in many cases in the following scattering processes are described, it will often be useful to use instead of the individual particle momenta the so-called Mandelstam variables. These are defined for the process $ab \to cd$ as

$$\begin{aligned} s &= (p_a + p_b)^2 = (p_c + p_d)^2 \\ t &= (p_a - p_c)^2 = (p_b - p_d)^2 \\ u &= (p_a - p_d)^2 = (p_b - p_c)^2 \end{aligned}$$

Herein s is the center of mass energy. At Born-level, the variables t and u correspond to the momentum transferred by a single exchange particle and final-state-exchange reactions, respectively. The corresponding processes are therefore referred to as t-channel and u-channel, respectively. A process in the s-channel corresponds to a full annihilation of the

initial state into a particle, which then subsequently turns into the final state.

6.1 A sample process: $e^+e^- \rightarrow$hadrons

To start off, consider the annihilation of an e^+e^- pair, neglecting their masses, into hadrons. For a center-of-mass energy s the total cross-section is

$$\sigma(s) = \frac{1}{s} f\left(\frac{m_i}{s}\right),$$

with m_i the masses of the produced hadrons. At sufficiently large energies, this will be dominated by a single photon exchange, and therefore the cross-section will behave as $1/s$, the propagator of an effectively massless exchange particle. This is also called an inclusive cross-section, as it integrates over all final states.

A less inclusive cross-section will be the process $e^+e^- \rightarrow hX$, with h some specific hadron with momentum p. Such a reaction can be viewed as roughly a two step process, it factorizes. First, in a hard, i. e. perturbative, interaction a parton a, quark or gluon, is formed. In a second step, due to confinement, this parton turns into hadrons. Under this assumption, the corresponding partial cross-section would be given by a convolution

$$d\sigma^{e^+e^-\rightarrow hX}(p,s) = \sum_a \int_0^1 \frac{dz}{z} d\sigma^{e^+e^-\rightarrow aX}\left(\frac{p}{z}, s\right) D_a^h(z). \tag{6.1}$$

Since it is not possible to identify the original parton, the sum has to be over all partons. The cross-section $d\sigma^{e^+e^-\rightarrow aX}$ is the hard cross section, to be calculated perturbatively, to obtain a parton a with momentum p/z.

This hard process can be calculated straightforwardly at Born level using standard perturbation theory. It is a pure s-channel process. If the energy is small enough that contributions from the weak interaction do not play a role, the differential cross-section is

$$\frac{d\sigma}{dt} = \frac{\alpha^2 e_q^2 \pi N_c}{s} \frac{t^2 + u^2 - m_q^2(2s - m_q^2)}{s^2}$$

and the total cross-section is

$$\sigma = \frac{2\pi\alpha^2}{s} e_q^2 N_c \left(1 - \frac{\beta_q^2}{3}\right) \beta_q$$

which has to be summed over the different quark species q, if the flavor is not detected, and β_q is the speed of the quark. There are a few remarks. One is that gluons do not contribute at this order, as they do not carry electric (or weak) charges. The second is

that the only difference to the same process with leptons in the final state is just the factor N_c. Thus, this partonic cross-section is enhanced compared to the purely QED one. This can be used, by taking a ratio of the cross-sections of both processes, the so called R ratio, to determine experimentally the number of colors. However, one thing neglected here is that also any other particle having the same quantum number as the photon could be exchanged, provided s is large enough, and would require to exchange s in the denominator by $((\sqrt{s}-m)^2+\frac{\Gamma^2}{4})^2$, thus leading to a resonance structure. There are multiple electrically and flavor neutral hadrons having the required 1^- quantum numbers, the vector quarkonia. Examples are the ρ, the ϕ and the J/ψ. In the vicinity of their masses, the simple approximation breaks down.

The second term in (6.1), the fragmentation functions D_a^h, are non-perturbative in nature and describe how likely it is that a parton a with fraction z of the hadron momentum p hadronizes in said hadron h.

There are number of remarks to be made.

First, it is assumed that both subprocesses take part so far separated that they do not influence each other, hence the term factorization. Thus, the fragmentation function D becomes universal, and is process-independent. This idealization works in surprisingly many circumstances. It becomes invalidated, if there occurs strong interactions between final state particles, i. e. X and h, or final and initial state, which is due to the weakness of QED for this process not a problem, but may become one if the initial state is hadronic. Still, though such factorization violations have been observed experimentally, they are in many cases irrelevant, most notably if most of the integral weight is close to $z \approx 1$, and thus the parton is very hard. Such factorization violations will therefore be mostly neglected throughout.

The second is that the process is still specific for the final state X. In many cases, this remainder is, or cannot, be observed, and then also summation over X has to be performed. This is then called semi-inclusive. If, in the other extreme, all other particles are also measured, it is called an exclusive cross-section.

This leaves open how to determine D. As noted, it cannot yet be calculated with nowadays technology. Thus, it is usually obtained as a fit to experimental data and/or from a simplified model of QCD, fitted to experimental data. But since it is assumed to be processes-independent, this can be done in one experiment, and then be used to make predictions for all other experiments.

As it stands, (6.1) is correct at tree-level. Radiative corrections lead to two modifications. One is that renormalization is necessary. In principle, cross-sections as observables are renormalization-group invariant, and especially do not depend on the renormaliza-

tion scale μ_R. However, perturbative calculations do not preserve this property beyond one-loop order, and thus introduce an artificial dependence on μ_R. Of course, since the cross-section now depends on this scale, so does the fragmentation function, as it can only be determined taking this dependency into account. Second, it is not possible to account for arbitrary virtualities of the intermediate particle, since the fragmentation function is not known. Thus, this virtuality has to be cut off at some scale μ_F, the factorization scale, introducing a second scale. Hence, the full version of expression (6.1) reads

$$d\sigma^{e^+e^-\to hX}(p,s) = \sum_a \int_0^1 \frac{dz}{z} d\sigma^{e^+e^-\to aX}\left(\frac{p}{z}, s, \mu_R, \mu_F\right) D_a^h(z, \mu_R, \mu_F). \tag{6.2}$$

In practical calculations, this leaves to chose μ_R and μ_F. A common choice is to select a value where the dependence on the choice becomes minimal, which in most cases is $\mu_R = \mu_F = \sqrt{s}$. The residual dependence of the cross-section on variations, usually within a factor 2-3, of both scales is often quoted as a systematic error of the calculation.

Finally, the assumption that the initial particles have a sharp energy is often not warranted, as they can emit photons, which cannot be identified reliably. This is so-called initial-state radiation, ISR. This can be accounted for by averaging over all possible initial states

$$d\sigma_{ISR}^{e^+e^-\to hX}(p,s) = \int_0^1 dx_1 \int_0^1 dx_2 f_{e/e}(x_1,s) f_{e/e}(x_2,s) d\sigma^{e^+e^-\to hX}(p, x_1x_2s),$$

where the electron masses are neglected. The functions $f_{e/e}(x,\mu^2)$ describe the probability to find inside the virtual clouds of particles of an electron an electron with a fraction x of the total energy when it is probed at an energy μ^2, which is here the center-of-mass energy. This electron distribution function can be calculated, to good approximation, in perturbative QED, in a way which will be explained later. It becomes

$$\begin{aligned} f_{e/e}(x,\mu^2) &= \beta(\mu^2)(1-x)^{\beta(\mu^2)-1} \\ \beta(\mu^2) &= \frac{2\alpha}{\pi}\left(\ln\frac{\mu^2}{m_e^2} - 1\right), \end{aligned}$$

where m_e is the electron mass and α the fine-structure constant.

This complication is just the tip of the literal iceberg, especially when s lies close to, and especially just above, a resonance. Since the main emphasis here is on hadron-hadron reactions, this will not be further elaborated, but would reemerge when considering charged hadron collisions, like pp collisions.

6.2 Lepton-hadron scattering and DIS

To understand the structure of a hadron, leptonic probes are especially useful, as they interact weakly with the partons. Thus, to leading order there will be interactions with at most one parton by exchange of either a virtual photon or weak gauge boson. Hence, this probes the properties of the parton inside the hadron, a process known as deep-inelastic scattering (DIS). The information gained will play an important role when it comes to hadron-hadron interactions, since in this case partons from both hadrons interact, and their properties are modified by the enclosing hadron. The consequences of this can be inferred from lepton-hadron interactions.

The kinematics of this process are most conveniently described by the following set of variables. The center of mass energy s is determined as $s = (l + p)^2$ from the incoming electron and hadron 4-momenta l and p, respectively. In the collision process, the exchange particle carries the (space-like) momentum transfer $q = l - p$ with $Q^2 = -q^2 > 0$. Since the wave-length is given by $1/Q$, this characterizes the structure size which can be resolved by the probe.

Usually, the final lepton momenta l' can be measured. Neglecting the lepton's mass, this yields

$$Q^2 = -(l - l')^2 = 4E_l E_{l'} \sin^2 \frac{\theta}{2},$$

where θ is the scattering angle. As will be seen, very useful is also the quantity $\nu = (l - l')p/M_h$. If the final state can be, at least summarily, measured, more information are available. Especially helpful is the invariant mass of the hadronic system

$$W = (p + q)^2 = M_h^2 + 2M_h\nu - Q^2.$$

Note that for elastic scattering $W^2 = M_h^2$. Further useful quantities are

$$\begin{aligned} x &= \frac{Q^2}{2M_h\nu} \\ y &= \frac{q \cdot p}{l \cdot p} \\ \eta &= \ln \cot \theta \end{aligned}$$

where especially x plays an important role later on, not unlike the case of x in the previous section. The pseudo-rapidity η is derived from the rapidity

$$\frac{1}{2} \ln \frac{E + p_z}{E - p_z}$$

in the limit of vanishing mass of the particle in question, and describes the relative angle to the collision axis, i. e. of the projectile-target axis assuming a head-on collision.

Under the assumption that the complete interaction can be described as just an exchange of a single particle, the matrix element can be written as

$$M = \langle l'|J_\mu|l\rangle g_{lV} \frac{-\eta^{\mu\nu}}{q^2 - M_V^2} g_{hV} \langle X|J_\nu|h\rangle,$$

where J is a current operator like $\bar{e}\gamma_\mu e$ for an electron current, V characterizes the exchanged particle, and g the couplings to this particle, and X is the hadronic final state. If several particles could be exchanged, e. g. photon and Z boson, the matrix elements must be averaged. This is a straight-forward extension, which will be neglected here, and spin-orientation etc. will be implicitly averaged or summed.

This structure suggests to write the cross section as

$$\begin{aligned} d\sigma &= \frac{4\pi}{4l\cdot p} \frac{(g_{lV} g_{hV})^2}{(Q^2 + M_V^2)^2} L_{\mu\nu} H^{\mu\nu} \frac{d^3\vec{l'}}{2E_{l'}(2\pi)^3} \\ L_{\mu\nu} &= \frac{1}{2}\langle l|J_\mu^\dagger|l'\rangle\langle l'|J_\nu|l\rangle \\ H_{\mu\nu} &= \frac{1}{8\pi}\sum_h \langle X|J_\mu^\dagger|X\rangle\langle X|J_\nu|h\rangle (2\pi)^4 \delta(p_X - k - p), \end{aligned} \tag{6.3}$$

where k is the remainder momentum.

Assuming that this is the only interaction of the lepton makes the evaluation of the leptonic matrix element straight-forward,

$$L_{\mu\nu} = 2\left(l_\mu l'_\nu + l'_\mu l_\nu - \frac{Q^2}{2}\eta_{\mu\nu} + iC_{lV}\epsilon^{\rho\sigma}_{\mu\nu} l_\sigma l'_\tau\right) + 2D_{lV} m_l^2 \eta_{\mu\nu},$$

where the last two terms arise due to the parity-violation of the weak interaction, if the exchanged particle is either a W or a Z. Due to the complicated internal structure of the hadron, its tensor cannot be calculated as simply. However, its tensor structure is up to dimensionless, real coefficient functions, so-called structure functions, uniquely determined by Lorentz symmetry. It can thus be decomposed as

$$H^{\mu\nu} = -F^1\eta_{\mu\nu} + \frac{1}{p\cdot q}\left(F_2 p^\mu p^\nu + iF_3\epsilon^{\mu\nu}_{\rho\tau} p^\rho q^\tau + (F_4 + iF_5)p^\mu q^\nu + (F_4 - iF_5)q^\mu p^\nu + F_6 q^\mu q^\nu\right)$$

Note that there are also other definitions of the structure functions in existence, and in case of polarized particles or additional measured momenta additional structure functions appear.

This expression can be further simplified. If weak interaction effects on the internal structure of the hadron are neglected, $F_5 = 0$. Also, electromagnetic current conservation

demands $H_{\mu\nu}$ to be transverse with respect to q, up to corrections of order $(m_l m_q/Q^2)^2$, which are tiny. This yields

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi}{x}\frac{\alpha_{lV}\alpha_{hV}}{(Q^2+M_V^2)^2}\left(xy^2F_1 + \left(1-y-\frac{(xyM_h)^2}{Q^2}\right)F_2 - C_{lV}x\left(y-\frac{y^2}{2}\right)F_3\right)$$

with the coupling constants α defined as $\alpha = g^2/(4\pi)$. The structure functions then depend only on Q^2 and x as kinematic variables, as well as Λ_{QCD} and the quark masses.

At very low energies, the lepton probes the hadron as a whole, and is thus primarily sensitive to the charge density. It is found that the charge density is essentially exponentially decaying at long distances, and thus, up to kinematic factors, F_1 and F_2 have a dipole structure, while F_3 vanishes, as parity violation does not play a role. At energy scales sufficiently large to excite hadrons as a whole, resonance structures appear in the F_i reflecting this.

At very large Q^2, the behavior changes, as the internal structure is probed. Especially, the structure functions do not vanish at large Q^2, but rather tend in leading order to a function depending only on x, the so-called Bjorken scaling. This is the behavior expected for point-like, massless scattering centers. Since this independence implies $Q^2 - 2M_h\nu = 2xp{\cdot}q$, this looks like the probe would be elastically scattered at the scattering center, which therefore has no internal excitations and appears unbound, a true elementary particle. These results are only approximately true, and this behavior is slightly violated, so-called scaling violations. This will be ignored for the moment.

A dependence only on x and point-like, essentially free scattering centers is just what the parton model is. Thus, the scattering is described by parton density functions (PDF) $f(x)$ where f can be either q, $\bar{q}$, with q any quark flavor, or g for the gluons. Then $f(x)dx$ can be interpreted as the probability to find a parton of type q with momentum fraction x inside the hadron. This probability will be dependent on the hadron under scrutiny, especially for the different flavors.

A leading-order perturbative calculation this immediately yields the Callan-Gross relation, $2xF_1 = F_2$, one of the historically first indications for the point-like structure of the partons. The remainder structure functions then depend on the charges involved. For a photonic probe, e. g.

$$F_2 = x\sum_{D,U}\left(\frac{1}{9}(D+\bar{D}) + \frac{4}{9}(U+\bar{U})\right), \tag{6.4}$$

where D collects the down-type quarks d, s, and b, while U collects the up-type quarks u, c, and t. In this case $F_3 = 0$. Note that due to the Appelquist-Carrazone theorem, the PDFs are suppressed like Q^2/M_q^2 with the quark masses, and therefore heavy quarks

rarely play a role, though some experimental results even exist for the charm content of the nucleon.

Similarly, this leads back to the sum-rules of section 5.1.2. For the proton e. g.

$$\begin{aligned}\int dx(u(x) - \bar{u}(x)) = &\quad 2\\ \int dx(d(x) - \bar{d}(x)) = &\quad 1\\ \int dx(s(x) - \bar{s}(x)) = &\quad 0.\end{aligned}$$

and many others. It is often helpful to define formally valence quarks and sea quarks, $q_s(x) = q(x) - q_s(x)$ and $\bar{q}(x) = \bar{u}_s(x)$ to remove the trivial contribution of conserved quantum numbers from the quantum contributions. E. g. for the proton

$$\begin{aligned}\int dx u_v(x) = &\quad 2\\ \int dx u_s(x) = &\quad 0\\ \int dx(u_s(x) - \bar{u}_s(x)) = &\quad 0.\end{aligned}$$

However, relations like the last have to be taken with care. They are true for the C-invariant and flavor-conserving strong interactions, but are violated by the weak interactions and the CKM matrix. Though small, these effects have been experimentally detected. E. g.

$$\int dx(s(x) - \bar{s}(x)),$$

is not zero for the proton, despite appearance.

Of course, the total structure functions will have to be again summed over all possible partons with all possible momentum fractions,

$$F_i^{Vh}(x, Q^2) = \sum_f \int_0^1 \frac{dz}{z} f_h\left(\frac{x}{z}, \mu_F, \mu_R\right) F_i^{Vf}(Q^2, z, \mu_F, \mu_R),$$

where F^{Vf} is the structure constant for hitting the parton itself with the probe V. This treats the parton essentially as free, and only keeps the radiative corrections to the interaction with V. It is not sensitive to the environment, and therefore independent of the hadron h. The PDFs f_h encode all the information on the relation of the parton to the hadron, but are not specific to the reaction between the probe and the parton. Again, this factorization holds only under specific assumptions, though these are quite often fulfilled. An often convenient choice for the scales is $Q^2 = \mu_R^2 = \mu_F^2$, which e. g. yields $F_i^{Vf}(Q^2, z, Q^2, Q^2) \sim \delta(1 - z)$.

Formally, this expansion is only the leading term of a more precise operator product expansion description, as introduced in section 5.1.1, in the limit of $Q^2 \to \infty$. In this

limit, the PDFs also formally reduce to $\langle h|N_f(x)|h\rangle$, where N_f is the number operator. The sub-leading terms of the OPE can be ordered by their so-called twist, i. e. the difference between the (canonical) mass dimension and the spin of the operator. Sub-leading twist contributions have been measured and calculated. Note that, as with all expansions, also this OPE is only valid under certain conditions, and higher-twist corrections can become more important than the leading-twist contribution.

The deviations of the f from a dependence entirely on x can be determined as long as the scales $Q^2 = \mu^2 = \mu_F^2 = \mu_R^2$ is sufficiently large as to make the development be dominated by perturbation theory. Then, the renormalization group introduced in section 5.1.7 can be used. This yields the so-called DGLAP evolution equations

$$\mu^2\partial_{\mu^2}q = \int_x^1 \frac{\alpha_S(\mu^2)}{2\pi}\left(P_{qq}(z)q\left(\frac{x}{z},\mu^2\right) + P_{qg}(z)g\left(\frac{x}{z},\mu^2\right)\right) \tag{6.5}$$

$$\mu^2\partial_{\mu^2}\bar{q} = \int_x^1 \frac{\alpha_S(\mu^2)}{2\pi}\left(P_{qq}(z)\bar{q}\left(\frac{x}{z},\mu^2\right) + P_{qg}(z)g\left(\frac{x}{z},\mu^2\right)\right) \tag{6.6}$$

$$\mu^2\partial_{\mu^2}g = \int_x^1 \frac{\alpha_S(\mu^2)}{2\pi}\left(P_{gg}(z)g\left(\frac{x}{z},\mu^2\right) + \sum_{f=q,\bar{q}} P_{gq}(z)f\left(\frac{x}{z},\mu^2\right)\right) \tag{6.7}$$

The structure of these equations are rather straight-forward. They require the knowledge of the PDFs in some input range $(x, 1)$ for a given x. Since small x correspond to small momentum fractions, these are the unknown non-perturbative contributions. Then each term describes the probability of a parton described by the input PDF to split into a parton of the desired type at the probed x for every x. This occurs, of course, with a coupling strength α_S. If assumed that this process is governed by perturbation theory, the splitting functions P can be determined in perturbation theory. To leading order, they are given by

$$P_{qq} = C_F\left(\frac{1+z^2}{1-z}\right) \tag{6.8}$$

$$P_{qg} = T_F(z^2 + (1-z)^2) \tag{6.9}$$

$$P_{gg} = 2C_A\left(\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z)\right) + \frac{11C_A - 4n_fT_F}{6}\delta(1-z) \tag{6.10}$$

$$P_{gq} = C_F\frac{1+(1-z)^2}{z}, \tag{6.11}$$

where the + subscript is defined as

$$F(z)_+ = F(z) - \delta(1-z)\int_0^1 dyF(y).$$

Of course, higher-order corrections can be determined. With the DGLAP evolution equations it is no longer necessary to determine the PDFs for arbitrary momentum fractions x, but only in the non-perturbative domain. Up to sub-leading non-perturbative corrections, they can then be determined for all other x. Of course, in practice the input PDFs are from experiments, and therefore of limited precision. Thus, the PDFs determined through evolution will inherit this uncertainty, and it usually tends to blow up for smaller and smaller x. This is especially true for sea quark or gluon contributions, which are somewhat harder to extract than for the valence quarks. Similar equations can also be derived for the fragmentation functions. In practice, PDFs are parametrized in some way, and then the parameters are fitted such as to reproduce experiments best. There are many strategies for doing so, and this is far beyond the scope of this lecture.

Note that such evolution equations must maintain the Llewellyn-Smith sum-rule

$$\int_0^1 dx x \sum_f f(x) = 1, \tag{6.12}$$

which is just the statement that the partons make up the complete hadron[1]. Since the quark PDFs are found to saturate only the sum rule for the proton to about 50%, the presence of a large amount of gluons becomes explicit. This was one of the first indirect indications for the existence of gluons.

Of course, since the parton has been removed from the hadron by the interaction with V, it will afterwards fragment in the same way as for the lepton-lepton collision described before. The factorization ensures this multi-step process as three independent steps. The first is the selection of the parton by means of a PDF. The second is the interaction with the probe. And the third is finally the fragmentation. If the energy of the struck parton is large enough, the fragmentation will not be into a single hadron, but usually into many hadrons, which will share the kinetic energy of the original parton, as well as its quantum numbers[2]. If the kinetic energy is substantial compared to the total mass of the produced hadrons, the hadrons will be highly collimated in the original movement direction of the partons. Such a collimated spray of hadrons is called a jet.

Of course, the hard interaction between the probe and the parton does not need to be confined into the kicking of a single parton outside of its original hadron. In the process further partons with high energies can be produced, e. g. by gluon radiation, or

[1] In principle, all other particles have to be included besides quarks and gluons. In practice, their contribution is essentially always negligible.

[2] Soft partons from the struck hadron are needed to maintain color neutrality of the fragmentation products. This is usually assumed to be a negligible contribution to the final result.

the fusion of the probe with a gluon into a quark-anti-quark pair via a loop process. In this case, multiple jets may arise, which carry the relative kinematic information of the original partons. Note that the hard interaction may also produce particles other than partons, especially leptons or weak gauge bosons. They will not fragment immediately. If sufficiently stable, they will escape the collision. Otherwise, they will decay. If the decay is once more into partons, this may also induce secondary jets.

The remainder of the target hadron is of course still there, and will either also fragment in some way or be essentially unchanged and propagating almost on its original course, the latter happening quite often at small x. For this to happen, the struck parton must have been emitted essentially as a collinear, colorless virtual particle, which is struck instead. This hypothetical particle is called a pomeron, an object of which the precise nature is not fully understood, but must contain several colored constituents. This is called a diffractive process. Since essentially the pomeron's substructure is probed in this case, the PDFs from diffractive events are not the same as for non-diffractive events, and therefore not universally applicable for interactions involving the original hadron.

6.3 Hadron-hadron scattering

In hadron-hadron scattering at low energies, the interactions are mainly due the hadrons as a whole. They are therefore dominated by non-perturbative contributions, and have to be treated as such. This is the realm of chiral perturbation theory or of phase shift analysis in lattice calculations. Since these are thus either very phenomenological or treatable with the techniques already discussed, they will not be within the focus here. More interesting are high-energy reactions, which are not accessible with either methods, since both are either conceptually or computationally restricted to sufficiently small energies.

Of course, hadron-hadron interactions manifest all the behaviors seen in the previous two cases, and hence in the following only those aspects will be highlighted where differences occur. Particular interesting is the case of large center-of-mass energy $s \to \infty$ in a two-to-two process. The total cross-section is found to behave like

$$\sigma = \frac{1}{M_i^2} \left(\frac{s}{M_i^2} \right)^{\alpha_i - 1},$$

where i identifies some quantum number channel. The mass-scale M is just setting the scale. The $\alpha(0)$ describes the intercept of the Regge trajectory in the corresponding quantum number channel. Especially, if the exchange particle would have spin J, then without interaction $\alpha_i(0) = J$. However, QCD modifies this to a different number, of

order 1, which is characteristic for every Regge trajectory. If the exchanged particle is the ominous Pomeron, it is almost exactly one.

This is still the case of rather soft interactions, where the exchanged energy is small, but should give a flavor of non-perturbative contributions. At smaller energies s, e. g. resonance structures due to bound states appear, and a description beyond phenomenological approaches is so far nigh on impossible.

Thus, so far most useful are hard interactions with large momentum transfer, such that like in DIS the parton interaction can be factorized. A cross-section for the production of weakly interacting final state particles X, plus possible remnants of the projectiles, takes the form

$$d\sigma(hH \to X) = \int_0^1 dx_1 dx_2 \sum_{ab} f_{a/h}(x_1, \mu_F^2, \mu_R^2) f_{b/h_2}(x_2, \mu_F^2, \mu_R^2) d\sigma^{ab\to X}(Q^2, \mu_F^2, \mu_R^2),$$

where the momentum transfer Q^2 is assumed to be only between the two partons. If the initial state is a hadron-anti-hadron pair, and the final state a lepton-anti-lepton pair, this is called a Drell-Yan process.

The process factorizes in taking a parton from each hadron, and let them interact in a hard way to produce the final state. It is necessary to take all possible partons, with all possible energy fractions, into account. Of course, if the final state X includes also partons, then these need again to be fragmented.

Of course, there are again the remnants (or unchanged projectiles in case of diffractive interactions) of the original hadrons. However, in the usual head-on collisions, they essentially keep the original direction of movement, as their interaction is limited to small momentum transfers. The hard sub-processes has in general isotropically distributed large transverse momentum, in the center-of-mass system, with respect to the collision axis, and therefore can be isolated from the remnants. This transverse momentum, that is the magnitude of the transverse part of the momentum of the detected particles, is often a much better quantity to characterize produced particles than their full momentum. The total initial transverse momentum of the event is zero, and little is unobserved with the remnants. For the longitudinal part of the momentum, and thus the complete momentum, an unknown part is carried away with the fragments of the projectile, especially if undetected particles, like neutrinos, arise. In a similar vain, also transverse energy is defined as the energy from the rest mass and the transverse momentum.

6.4 Factorization schemes

PDFs appeared so far as somewhat physical objects, as they seem to describe the hadron's structure. However, this is only true at leading order.

Consider the process of photon scattering at sufficiently large Q^2, such that all masses can be neglected. To leading order[3], the hard process will be $q\gamma^* \to q'$, i. e. a quark will be hit by the necessarily virtual photon, and thus receive a kick. Gluons do not contribute at this order, as they do not couple electromagnetically. It is useful to split for the following the hadronic tensor $H_{\mu\nu}$, irrespective of whether on parton or hadron level, in two parts

$$\begin{aligned} H_\Sigma &= -\eta^{\mu\nu} H_{\mu\nu} \\ H_L &= p^\mu p^\nu H_{\mu\nu}. \end{aligned}$$

This permits to express the structure function as

$$\begin{aligned} \frac{F_2}{x} &= \sum_f \int_x^1 \frac{dz}{z} f\left(\frac{x}{z}\right) \left(\frac{1}{1-\epsilon} H_\Sigma^{Vf}(z) + \frac{3-2\epsilon}{1-\epsilon} \frac{4z^2}{Q^2} H_L^{Vf}(z) \right) \\ F_1 - \frac{F_2}{2x} &= -\sum_f \int_x^1 \frac{dz}{z} f\left(\frac{x}{z}\right) \frac{4z^2}{Q^2} H_L^{Vf}(z). \end{aligned}$$

The ϵ herein is the same as has been introduced in dimensional regularization in section 4.6.5, i. e. calculations are performed in $4-2\epsilon$ dimensions. Without loop-level calculations, ϵ can always be set to zero, but its explicit factorization is useful in actual calculations.

The calculation of the leading order hadronic tensors is straight-forward, and yields

$$\begin{aligned} H_\Sigma &= e_f^2 (1-\epsilon) x \delta(y-x) \\ H_L &= 0. \end{aligned}$$

Hence, the result is

$$2xF_1 = F_2 = x \sum_{q,\bar{q}} e_f^2 f(x).$$

This reproduces the Callan-Gross relation. It also shows that the structure function, which is observable, is directly related to the PDF. This result is the origin of relations like (6.4), and the idea of interpreting PDFs as physical quantities.

[3]There are many subtleties, similar as those discussed in the ISR, with collinear particles in the final state. This problem can be dealt with using perturbative techniques not specific to QCD. They will therefore not be detailed here explicitly, except where necessary.

This picture is eradicated at next-to-leading order. If the same flavor should be observed in the final state, only the additional process $\gamma^* q \to q' g$ contributes at the next order. There are four possibilities, two including gluon radiation and two including photon-gluon fusion. This is still a tree-level process, but it involves a certain problem, a collinear singularity. The result for H_Σ can be written in terms of the outgoing quark momentum

$$\begin{aligned} \hat{s} &= Q^2 \frac{1-z}{z} \\ q' &= \frac{\sqrt{\hat{s}}}{2}(1, -\sin\theta, 0, -\cos\theta), \end{aligned}$$

where thus θ is the angle between the incoming quark and the gluon momentum, as

$$H_\Sigma = 4e_q^2 \alpha_S C_F \frac{1}{16\pi} \int_{-1}^{+1} d\cos\theta \left(\frac{2(1-z)}{1-\cos\theta} + \frac{1-\cos\theta}{2(1-z)} + \frac{2z(1+\cos\theta)}{(1-z)(1-\cos\theta)} \right)$$

which is singular for $\cos\theta \to 1$. This corresponds to a gluon being essentially collinearly emitted, i. e. with transverse momentum $k_T^2 = \hat{s}/4\sin^2\theta$ being almost zero. To regulate the expression, a lower cutoff κ^2, can be imposed on the integral[4]. Performing the integral then yields

$$H_\Sigma = e_q^2 \frac{\alpha}{2\pi} \left(P_{qq}(z) \ln \frac{Q^2}{\kappa^2} + R_{ff}(z) \right),$$

where $P_{qq}(z)$ is given in (6.8), and $R_{ff}(z)$ is finite for $\kappa^2/Q^2 \to 0$ and vanishing terms have been suppressed.

For H_L to this order no such divergences arise, the calculation can be performed in a straightforward way, and it will thus only modify R_{ff}. This yields for F_2

$$\frac{F_2(x, Q^2, \kappa)}{x} = \sum_f \left(f(x) + \int_x^1 \frac{dz}{z} f\left(\frac{x}{z}\right) \frac{\alpha}{2\pi} \left(P_{qq}(z) \ln \frac{Q^2}{\kappa^2} + R_{ff}(z) \right) \right).$$

This shows that F_2 is beyond leading order no longer a measure of the PDFs. But this makes F_2 also dependent on the arbitrary cutoff κ.

The origin of this is that the soft gluon is strictly speaking not a perturbative object. It therefore should rather be part of the PDF. However, this leaves open the question of the scale from which point onwards on a gluon should be considered part of the PDFs. This scale μ_F is arbitrary, and is just the same factorization scale as introduced already

[4]A more formal approach would used dimensional regularization instead, but yields ultimately the same results.

in (6.2). It is here introduced by defining a new parton distribution function

$$f^F(x,\mu^2,R_f^F) = f(x,\kappa) + \int_x^1 \frac{dz}{z} q\left(\frac{x}{z},\kappa\right) \frac{\alpha}{2\pi}\left(P_{qq}(z)\ln\frac{Q^2}{\kappa^2} + R_f^F(z)\right).$$

Of course, since the splitting is arbitrary, also the original PDF must now be changed to include this arbitrariness. However, if the whole process is well-defined, as it is experimentally observed, the final result can not depend on κ, and therefore the modified PDF f^F cannot. Thus, the underlying assumption is that the dependencies on κ on the right-hand side cancel. Since in the end only f^F will be fitted to experiment, this is, in a sense, a self-fulfilling prophecy. This redefinition is also arbitrary in the sense, as nothing prevents from including finite parts of R into the PDF, and thus add a function R^F, which will become negligible at large Q^2. While the dependency on κ is fixed, this arbitrariness in R implies that there are different schemes to factorize soft gluons, and thus, similar as for a renormalization scheme, a factorization scheme is created. However, since soft gluons are assumed to be emitted in the same way for all processes, the procedure is afterwards fixed for all calculations. Still, the final result

$$\frac{F_2(x,Q^2,\mu_F^2)}{x} = \sum_f \left(f^F(x,\mu_F^2,R_f^F) + \int_x^1 \frac{dz}{z} f\left(\frac{x}{z},\mu_F^2,R_f^F\right) \frac{\alpha}{2\pi}\left(P_{qq}(z)\ln\frac{Q^2}{\mu_F^2} + R_{ff}(z) - R_f^F(z)\right)\right),$$

depends on this scheme. Due to this arbitrariness, PDFs are not physical, as they depend on the scheme. Of course, the scheme-dependence is only an artifact of our inability to provide a full non-perturbative calculation. However, as for renormalization schemes, higher order results do not manifest a full independence on μ_F, and thus remains always a scheme dependence. The residual dependency on this is a systematic error, for which the same considerations apply as for any systematic error.

Schemes are then defined by defining R_f^F. Commonly used schemes are minimal subtraction schemes without finite terms, $R_f^{\overline{\text{MS}}} = 0$ and the DIS scheme with $R_f^{\text{DIS}} = R_f$, such that only the leading part remains in the perturbative calculation. As noted before, a common choice is $\mu_F = Q^2$ in single-scale problems, as this eliminates the logarithm. In the DIS scheme, this yields that also to NLO $F_2 = x\sum_{q,\bar{q}} e_f^2 f^{\text{DIS}}(x)$, though an explicit calculations shows that this is no longer equal to F_1 at this order. However, it should be noted that the function R_f contain process-dependent contributions, as they are calculated based on the process at hand. Thus, while being universal for all processes of the same type, this remains no longer true, if, e. g. hadron-hadron interactions are considered, as in section 6.3. This does not happen for the $\overline{\text{MS}}$ scheme, as there no trace of the underlying process remains, at the expense of more complicated formulas for the structure functions.

In practice, the (at a rigorous level unjustified) estimate of the systematic error is obtained by varying μ_F^2 in a certain window, usually by factor of 2-3, and take the variation under this change as the error. This follows the procedure used for the residual dependency on the renormalization scale.

It should be noted that such logarithms do in general not only arise due to collinear emissions, but can also occur in the form of $\ln(1/x)$ when the momenta of the emitted particles become small, similar to the Landau pole of QCD. Especially, contributions can have both singularities simultaneously. These are therefore also non-perturbative in origin, and have to be treated likewise. Various schemes have been developed for these cases, similar to the DGLAP evolution equations, e. g. the BFKL evolution equations, which describe the behavior in x similar to the one in Q^2 of (6.5-6.7). In the end, all of them have to some extent to cut off the growth towards the non-perturbative domain, either with methods similar to those discussed above, or by introducing models for the PDFs which counteract the growth explicitly. This remains an active area of research.

It is, however, clear, that the ultimate resolution will be of non-perturbative nature. E. g., small x means a small momentum fraction, and therefore many partons. Since any parton has a cloud of virtual particle, it has an effective size. At some point, so many partons are present that the picture of almost non-interacting particles no longer holds, as they start to overlap, and interact strongly. Various ideas have been proposed to deal with such a situation, ranging from effective shadowing to limit the number of effectively present partons up to an effective condensate (color glass condensate) which describes the remaining partons as a background effect to give the single soft parton a well-defined behavior. Though these pictures show some success, none of them can satisfactorily explain all observed phenomena quantitatively equally well.

6.5 The nucleon spin

So far, the discussion has concentrated on a situation where the particles have not been polarized, and therefore the results were independent of the initial and final spin alignment. It is experimentally possible to use polarized particles, and also measure to some extent the final spin. This adds another reference direction, in addition to those signaled by the momenta, complicating things. Especially, it introduces on top of the structure functions polarized structure functions. Similar calculations can then be performed as before.

One of the major results of such experiments was that a sum-rule similar to the Llewellyn-Smith one (6.12) can be constructed for the nucleon spin. And again, it was found that the sum-rule is only saturated to about 50% by the valence quark contribution,

indicating once more the existence of additional structure. However, in contrast to (6.12), the contribution of the gluons, though quite hard to measure, was found to be small, such that even when including gluons the sum-rule is not fulfilled. In contrast to the momentum sum-rule, the spins of the partons do not need to saturate the possible contributions, as relative angular momentum of the partons can also contribute.

The problem is that, in contrast to ordinary quantum mechanics, it is in a non-Abelian quantum gauge field theory not possible to give individual physical meaning to orbital angular momentum and spin, as both parts become gauge-dependent. Therefore, the contributions cannot be measured separately, at least not in any gauge-invariant way. This has made the resolution of this problem, once inadequately termed spin crisis, somewhat involved, and not finally settled.

Though no similar results for other hadrons exist, it can be rather safely assumed that similar problems would arise as well.

6.6 Hadronization

One other input was the fragmentation functions which ultimately describe how partons struck out of a hadron are observed. This information is, e. g., relevant for how the energy deposited in a jet relates to the original energy of the parton, especially when it comes to particles escaping outside the jet cone, or how properties like flavor are transported. The latter is especially interesting to distinguish between gluon and (anti-)quark jets. However, because high-energetic partons fragment into many hadrons, it is practically impossible to both measure and parametrize the fragmentation functions, while this is still possible as long as the parton only hadronizes in very few particles.

Thus, the same procedures as for PDFs do not work. Since a first-principles calculation from QCD is at the time also prohibitively complicated, the current solution is to refer to simplified models, which can be treated at least stochastically. A wide range of such models exist, where the approximations and model parameters are adjusted such as to describe different experiments equally well. The so tuned models are then used in their fixed form for calculations. Though not a perfect solution, several different models yield compatible results, and describe the experiments reasonably well. Hence, often the deviations between different models is used as estimate of the systematic error.

Of course, as with PDFs, the fitting always leaves a residual risk that effects are included in the fit rather than being distinguished as a new phenomenon. To counteract this possibility, the fits can be performed on a subset of a set of tuning experiments, and tested against the remainder, to check for any systematic deviations.

6.7 Monte-Carlo generators

As is visible, the experimental description of collisions is usually performed in a three step process, as far as factorization is viable. The first step is to obtain the initial particles for the interesting process. These are described by the various PDFs of the initial projectiles. The second step is to calculate the elementary process, usually performed perturbatively, to a certain order in perturbation theory. For the LHC at this time most such hard processes are calculated to orders between tree-level and NNLO. The third step is then to transform the final state of the elementary reaction into the particles finally observed in experiments. This involves usually first a decay cascade of the particles obtained in the elementary process into electrons, positrons, muons, photons, and other particles stable and weakly interacting on the scale of the experiments or into quarks and gluons. In the latter case also the creation of jets and/or fragmentation must be calculated, to obtain the final state.

While the initial state is rather simple, already the possibly involved intermediate particles before the hard collisions can easily number dozens. Moreover, even at moderate orders of perturbation theory, very many contributions appear. The final state is even more complex, and can involve many final particles, and several jets - up to eight jets and about a dozen quasi-stable particles are currently feasible for experiments.

To deal with this complexity requires computers. For many theories, the outer part of the event, the creation and fragmentation of the particles entering into the elementary hard process, is the same. Hence, this part can be automatized. Such programs, in which only the elementary cross-section is used as an input, are called Monte-Carlo generators. They perform a stochastic simulation of all possible background for a given signature, to provide stochastic estimates of the expected event signature for a given hard subprocess, just like real experiments are stochastic samples selected by certain properties of their final states. Current examples are publicly available codes like Sherpa, Pythia, Herwig, or Whizard, and can be downloaded, e. g., at www.hepforge.org.

Even the elementary cross-sections can be calculated in many cases automatically, especially at tree-level. In these cases only the Lagrangian has to be provided, to permit an automated calculation from the initial to the final state.

6.8 Nuclear physics

The ultimate version of hadronic interactions is, of course, nuclear physics. While it should in principle be possible to calculate all nuclear processes directly from QCD, this is

in practice almost impossible, just as solid state physics is something not directly accessible from QED. The reason is that with the number of valence partons involved, the problem size usually grows factorial. This even prevented so far a successful calculations of many n-body hadron problems, such as determining the mass of the deuteron, let alone helium, even with the most powerful lattice calculations. The same is true for functional methods, where only now baryons become accessible.

Thus, nuclear physics has so far been mainly the purview of models. There are of course quantum-mechanical models, like the shell model, to describe the properties of nuclei. The approximation with quantum mechanics is to some extent justified, as the average binding energy and kinetic energy of nucleons is of the order of MeV, while the masses of the nucleons is of order GeV. The necessary potentials are rather involved, as simple two-body interactions only describe some bulk properties reasonably well, but already fail in describing several less prominent features. Thus 3-body, and partly 4-body, interactions are required.

To provide this potential, two possibilities are in widespread use. One is to model this potential on the most generic principles, and then fit the parameters using experimental data. The second is a derived approach, where the potential is attempted to be obtained from QCD. One possibility is again lattice calculations. Though in general having the same problem of factorial growth, the demand of having three, or possibly four, particles is more reasonable to be achievable. However, at the current time, the available precision is not yet sufficient to rival phenomenological estimates. The second option is by going through the chiral perturbation theory of section 5.2.6. This allows to derive nucleon interaction from the more basic general hadron interactions. However, this is to some extent limited, as the mass of nucleons, which need to be included into chiral perturbation theory leading to baryon chiral perturbation theory, are strictly speaking too large compared to the expansion parameter of the pion mass. Still, the obtained results are currently competitive with purely phenomenological fits.

Finally, if only a qualitative picture is required, somewhat simple models for the calculations of the most basic properties can be used. This may be a feasible compromise when more bulk properties are required for many-body problems. After all, each nucleon adds already in the quantum-mechanical case to the size of the Hilbert space. Since the problems are no longer analytical feasible already in quantum mechanics, and, e. g., variational methods are used, the size of this Hilbert space is a serious constraint in practice for many-nucleon problems, say substantially above iron.

To construct a sample model, some basic properties of nucleon interactions are required. Assuming the relevant energies sufficiently small to neglect excited nucleon states, i. e.

below the roughly 250 MeV needed to create a Δ, only the nucleons ψ are needed. For many questions, isospin breaking and electromagnetism are also only second-order effect, and thus just two flavors of nucleons with the same mass are required. The primary interaction will be due to the exchange of pions π, the lightest hadrons. They need to be included. But to obtain both attractive and repulsive channels requires at least one more hadron, usually a vector meson, the ρ. Since the cutoff will anyhow be small compared to the ρ mass, this particle can be modeled as an elementary vector boson with a mass, despite this leads to a superficial violation of gauge invariance. To model chiral symmetry-breaking effects along the lines of the linear-σ model in section 5.2.3 also the σ has to be included. This will help to have the correct π dynamics. The final Lagrangian of the nucleon-meson model then looks like

$$\begin{aligned}\mathcal{L} &= \bar{\psi}\left(i\partial_\mu\gamma^\mu - g_{\pi N}v - \frac{g'_{\rho N}}{2}\rho_\mu\gamma^\mu\right)\psi + \frac{1}{2}\left(\partial_\mu\sigma\partial^\mu\sigma - \lambda(3\sigma_0^2 - v^2)\sigma\right)\\ &+\frac{1}{2}\left(\partial_\mu\pi\partial^\mu\pi - \lambda(\sigma_0^2 - v^2)\pi^2\right) - \lambda\sigma_0(\sigma_0^2 - v^2)\sigma - \lambda\sigma_0(\sigma^2 + \pi^2)\\ &-\frac{\lambda}{4}\left((\sigma^2 + \pi^2)^2 + \sigma_0^2(\sigma_0^2 - 2v^2)\right) - \frac{1}{4}(\partial_\mu\rho^\nu - \partial_\nu\rho^\mu)^2 - \frac{m_\rho^2}{2}\rho^\mu\rho_\mu\\ &+\bar{\psi}\left(g_{\sigma N}\sigma + g_{\pi N}\pi\right), \end{aligned} \tag{6.13}$$

where the various appearing constants have to be fixed by experiment, and $\pi = \pi^a\tau^a$. A mass difference between proton and neutron could be introduced by an additional explicit mass term for the nucleons, which is not proportional to the unit matrix in isospin space.

Already the model (6.13) is quite elaborate, especially when compared to the simplicity of QCD itself (4.1). However, it is still simpler compared to the many-body problem of QCD. This should illustrate the problem of emergent phenomena in hadron physics.

Note that by reinterpreting the nucleon fields as quarks fields, supplemented by a global color symmetry, this model is also known as the quark-meson model.

Chapter 7

The QCD phase diagram

The ultimate question of nuclear physics are the properties of neutron stars. In fact, a neutron star can be considered as a gigantic, stable nucleus, though this is an oversimplification. Since the density inside a neutron star increases from the outside to the inside, the physics will change. While the outer part is indeed essentially a nuclear system, the situation in the interior is not clear. Because QCD is so strongly interacting, it is very hard to treat it in such an environment, and even lattice simulations fail.

The reason is actually not the strongness of QCD. It is rather a problem which has to do with the distinguishability of particles and anti-particles, and how this can be treated if there are more particles than anti-particles. For some theories, like QCD, but there are also such systems in nuclear physics and solid-state physics, this entails algorithmic problems, the so-called sign-problem, which prevents so far the development of efficient simulation algorithms, as the computation time scales exponentially with the system size. It is a technical problem, and not a physics problem. The origin of this problem comes from the determinant in (5.7). For QCD, this determinant develops a complex phase dependent on the chemical potential at finite density. In a numerical averaging procedure, such a phase implies oscillations, and the numerical averaging procedure becomes thus extremely unreliable or expensive. Other methods, especially functional methods, do not suffer from this problem. However, they are also facing (yet) problems, e. g. in terms of accurately describing baryons.

It is therefore still an open field what actually occurs near the core of a neutron star, and whether the state of matter there is still just many nuclei, or whether other hadrons, including strange ones, play a significant role.

The situation in neutron stars is only a part of a wider topic, the QCD phase diagram, i. e. what is the state of matter at densities of similar or larger size than in nuclei and temperatures of size of hadron masses.

Experimentally, these are very hard to address questions. For neutron stars, possibly gravitational wave astronomy, together with the X-ray spectrum of neutron stars as a function of time, as well as their sizes and masses, will be the only available experimental input for a long time to come.

At much smaller densities, the situation becomes essentially the one of nuclear physics. here, experiments with nuclei, especially collisions, can be used. In this way, it was possible to find that there is (very likely) a phase transition between a gas of nucleons and (meta-)stable nuclei at roughly the density of nuclei, which permits to form nuclei. Since the nuclei show properties that are best described as them being droplets of a liquid made up of nucleons, this is also known as the nuclear liquid-gas transition. This phase separation persist for a few MeV in temperature, up to about 15 MeV, where it ends in a critical end-point. Thus, both phases are not qualitatively distinguished, just like in the case of gaseous and liquid water.

From the point of view of particle physics, the involved energy scales of nuclei are very small, and the distinction between the liquid and gaseous phases is essentially irrelevant. Thus both phases are not regarded as different. This common phase is denoted as the hadronic or vacuum phase, the latter as for all practical purposes the phase consists of far separated hadrons with vacuum in between.

There are now several interesting directions to move on. Usually, they are signified by temperature and the baryon-chemical potential , i. e. the chemical potentialdistinguishing baryons and anti-baryons. Thus, zero (baryo-)chemical potential[1] denotes a situation in which there is the same number of baryons and anti-baryons, which includes the possibility of none of either type.

When neutron stars are interesting, the axis with zero temperature and finite baryo-chemical potential is most interesting. Neutron stars do have some temperature, but it is of the order of one MeV, and even during their formation in a supernova explosion or during a merger it does not exceed 10-20 MeV. On hadronic scales, this is essentially negligible. Thus, to good approximation it is a movement only along the baryo-chemical potential direction. Starting from the vacuum, i. e. zero baryo-chemical potential, it can be shown exactly that nothing will happen before reaching a chemical potential of roughly a third of the nucleon mass. This is due to some analyticity properties of the free energy

[1] As in most cases only the baryo-chemical potential is present, the word baryo is dropped. The only other chemical potentials of relevance in most cases are an isospin-chemical potential, which gives the difference between up and down quarks, and is therefore relevant in neutron stars, though it is even then only small because of the close similarity of up and down quark, strange-chemical potential, which determines how many more strange than anti-strange quarks appear, and electro-chemical potential, which indicates the presence of electrons.

as a function of the masses of the lightest baryon, and the fact that it is made out of three quarks. This feature is called the silver-blaze feature. However, any small temperature will change this, and it is then only approximately true for chemical potentials much larger than the temperature, but smaller than this silver-blaze point.

After this point soon the nuclear liquid-gas transition is encountered. After this experimentally established point, as noted above, no fully reliable results are available. There are some reasons to believe that at least one further phase transition will be encountered, though even this is not sure. It is furthermore unclear whether this will happen at densities sill relevant for a neutron star, or significantly above it. However, model calculations, i. e. calculations using simplified versions of QCD, as well as comparisons to other theories which are similar to QCD, indicate that there could even exist many different phases, some amorphous, and some crystalline, in which besides the nucleons also other hadrons, like pions, kaons, and hyperons, may play a role.

The situation is much better regarding zero chemical potential. This situation is relevant in the early universe. At that time all matter in the universe is already present, and there is thus a sizable amount of baryons, but the temperature is high enough to thermally produce baryon-anti-baryon pairs. This reduces the chemical potential to very close to zero.

This situation is good accessible experimentally by high-energy heavy-ion collisions, e. g. at the LHC with up to 2.4 TeV kinetic energy per nucleon for lead nuclei. In such an experimental setup, temperatures as high as 600-700 MeV with almost zero chemical potential, despite the 416 nucleons in the original nuclei, can be achieved. Furthermore, this situation poses no serious problems to numerical simulations. Hence, the knowledge of this axis is rather good.

It turns out that the physics depends significantly on the masses of the up and down quarks. Though this is also suspected for the remainder of the phase diagram, it is evident in this case. If both quarks are very heavy, there is a first-order phase transition at a temperature of about 250-300 MeV. As the quark masses become lighter, this temperature decreases, and the transition becomes a rapid cross-over at a temperature[2] of about 150-160 MeV. This is the situation for up and down quark masses observed in nature, the physical masses. The mass of the strange quark influences the precise values of the temperatures, but does not provide any qualitative influence, and the heavier quarks have even less relevance. If the quark masses are decreased further, the temperature still drops a little bit. More importantly, at some point the cross-over turns again into a phase transition,

[2]There is no unique definition of a cross-over temperature. This is just the temperature were most changes occur.

this time of second order, becoming first order for even lighter quark masses, and remains until zero quark masses.

The critical temperature can be inferred from a rather simple argument, due to Hagedorn. When plotting the number of hadronic states, including resonances, as a function of mass, the resulting curve is an exponential. Of course, at some point the states become very unstable, or further flavor thresholds open, such that this curve is not experimentally confirmed at arbitrarily high masses. Assuming that it continues nonetheless, there is an exponential density of states. Neglecting interactions, and just inserting this density of states into a thermodynamic calculation, the resulting free energy will diverge at a characteristic temperature, the Hagedorn temperature, being indeed roughly 150 MeV based on the known hadronic states. Of course, such a divergence is unphysical, and indicative of either a phase transition or that the density of states is modified by some new effect. For QCD, the latter is just that the substructure of the hadrons become relevant.

What happens can be understood already in a simple picture. Temperature is classically nothing more than the kinetic energy of particles. In a quantum theory, temperature is just energy, which can also be converted to new particles. This will be exponentially suppressed with the mass of the created particles. Hence, the lightest particles will be most copiously produced. In QCD, these are the pions. These particles will have large kinetic energies, and will rapidly and repeatedly collide. At very high temperatures, because of the asymptotic freedom of QCD, these scatterings will mainly be dominated by hard partonic scatterings, and thus be almost perturbative. Thus, QCD becomes essentially as it behaves at high-energies. Especially, this implies that the effects of chiral symmetry breaking become reduced, and the quarks lose their effective mass, though not their current mass, at the phase transition or cross-over. In fact, in the limit of zero quark mass, the second order transition becomes a symmetry transition where chiral symmetry becomes restored. At the same time, since most collisions are hard and partonic, excited states become very unstable and in most cases it does not matter anymore that quarks are confined into hadrons. They act effectively as if they would no longer be confined. Thus, one speaks also of a deconfined phase, and calls the transition a deconfinement transition. However, since it is a cross-over, it is clear that qualitatively nothing has changed, but quantitatively it is a completely different situation.

As a consequence of this dominance of the partonic degrees of freedom and asymptotic freedom actually the high-temperature thermodynamic behavior of the theory is essentially that of a free gas of quarks and gluons, a so-called Stefan-Boltzmann gas. The reason is mainly that the hard processes contribute to the free energy like the temperature to the fourth power, while all other effects contribute at most like the cube of the temperature,

or even less. Hence, they become for thermodynamic bulk properties, i. e. extensive properties, irrelevant. Still, there are certain observables which are sensitive to non-trivial effects. Furthermore, the transition is very slow, and even at a few times the transition temperature even the bulk quantities are not yet fully dominated by the partonic processes. The system then behaves not like a gas, but rather like an almost ideal fluid.

This is also the situation encountered in the early universe. While it cools done, it will go through this cross-over. Before that, it is essentially dominated by the quarks and gluons, and only afterwards it starts to be dominated by the hadrons. However, because the transition is a cross-over, it seems that the transition had little quantitative influence on the evolution of the universe. Still, the point where it became possible to form stable nucleons is an important point, as this fixed the relative abundances of elements in the early universe. This process is called nucleosynthesis. The relative amount of nuclei created at this time, essentially only hydrogen, helium, lithium, and their isotopes, have been both observed and calculated. Both theory and experiment agree for most isotopes rather well.

The situation in the remainder of the phase diagram is not yet clear. It is possible to map out parts of it with heavy-ion collisions at lower energies. Because then less energy is available, less particles are produced, and therefore the baryon chemical potential is larger. Still, the accessible region is that of rather high temperatures, above those characteristic for neutron stars, and likely below the relevant chemical potentials. Also, numerical simulations start again to fail the larger the chemical potential becomes. Hence, the situation becomes less and less clear. What seems to be certain at the current time is that for quite some distance into the chemical potential direction little changes, and the cross-over remains at a temperature only slowly decreasing with increasing chemical potential. There are some speculations about a critical end-point, from which a phase boundary starts, which eventually meets with the chemical potential axis, but this is not yet settled. Other than that, the field is still wide open.

7.1 Deconfinement

At several points it has been mentioned that hadrons should start to overlap. This can occur either through thermal particle production, essentially pions, or by increasing the net baryon density. At some point, so many particles will be present that the inter-particle distances will become substantially smaller than the size of a typical hadrons. At the same time, the typical energies, i. e. either the typical thermal energies or the typical energies

above the Fermi energy of the baryons, will become large compared to[3] $\Lambda_{\rm QCD}$. Thus, most interactions will become hard, and the effective interaction starts to become weak due to asymptotic freedom: Hard scattering processes start to dominate.

This is also manifest in thermodynamic bulk quantities. E. g., the free energy behaves at very large temperatures like

$$-\Omega = P = N_c \frac{7\pi^2 T^4}{180} + \mathcal{O}\left(\frac{T^4}{\ln T}, T^3\right),$$

and thus like a gas of free gluons and quarks. The sub-leading corrections stem from two contributions. The logarithmic first part originates from perturbative contributions, while the second part is from non-perturbative ones. The reason for both of them being sub-leading is that the hard interactions characterized by the scale T dominate due to asymptotic freedom. Hence, at sufficiently high temperatures the bulk thermodynamics of QCD appear as the one of a free gas of quarks and gluons, manifesting very much the idea that due to the overlap of the hadrons the quarks and gluons become liberated, and are therefore free, or deconfined.

Of course, this is only true because the hard processes dominate due to asymptotic freedom. Since the transition is just a cross-over, there can be no qualitative change, though arbitrarily large quantitative ones. Especially, by scattering with thermally excited particles, even usually stable hadrons are no longer so. Only some kind of collective excitations will survive, like in any thermal medium.

However, since the (valence) quarks and gluons are no longer necessarily located within a given hadron, like in the vacuum, and parton-exchange reactions become common-place, this distinction is almost semantical. Still, gauge-invariance requires that the medium remains strictly and locally color-neutral, but the effective excitations can have substantially different forms then those at zero temperature[4]. Another reason is that the Wilson string tension defined as one kind of confinement vanishes even over intermediate distance, and actually also vanishes in pure Yang-Mills theory. However, the necessary ingredient for interpreting the Wilson potential as a (quasi-)potential, that it describes interactions over a temporal extent, is lo longer valid in a thermodynamic setting, as there is strictly no time dependence in equilibrium.

Thus, though often heard, the notion of deconfinement at finite temperature remains somewhat murky.

[3] Actually, for reasons becoming clearer in a formal field-theoretical language, the relevant energy scales are πT for fermions and $2\pi T$ for bosons.

[4] It is long speculated whether in a thermal medium it is, e. g. possible to obtain a gauge-invariant quasi-particle with fractional baryon charge. This was not successful so far.

Similarly, at asymptotically large densities, the leading-order of the free energy is again just the Stefan-Boltzmann-type behavior

$$-\Omega = P = N_c \left(\frac{7\pi^2 T^4}{180} + \sum_f \left(\frac{\mu_f^2 T^2}{6} + \frac{\mu_f^4}{12\pi^2} \right) \right),$$

and similar arguments can be made on the notion of deconfinement. However, it is unknown whether there is a phase transition separating a high-density phase from the remainder of the phase diagram, and therefore, at least for now, a qualitative change appears at least not impossible.

7.2 Chiral symmetry restoration

Much less controversial than deconfinement is the fate of chiral symmetry: It becomes restored. At light quark masses, this actually occurs as a phase transition. This does not alleviate the problems with an analytical connection between the low-temperature and high-temperature phase, as the connection is still possible in the full quantum and thermal phase diagram by making a detour over the mass.

How chiral symmetry restoration acts can be seen best in the linear-σ model, but for simplicity actually just a simpler model with only two particles and a U(1) global symmetry suffices, with two fields, the massive σ and the massless π, with the vacuum expectation value of the original field f. This will also outline other aspects. A useful starting point is given by formulating the Lagrangian as

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma - \frac{1}{2}(6\lambda f^2 - \mu^2)\sigma^2 + \frac{1}{2}\partial_\mu\pi\partial^\mu\pi - \frac{1}{2}(2\lambda f^2 - \mu^2)\pi^2 \\ &\quad -\sqrt{2}\lambda f\sigma(\sigma^2+\pi^2) - \frac{1}{4}\lambda(\sigma^2+\pi^2)^2 - \mu^2 f^2 + \lambda f^4. \end{aligned} \tag{7.1}$$

In this case the explicit zero-energy contribution is kept for reasons that will become apparent shortly, but will be essentially the same as when treating non-relativistic Bose-Einstein condensation. Only terms linear in the fields have been dropped, as they will not contribute in the following. The situation is similar as before, but now the condensate f has not been specified by the minimization of the classical potential, but is kept as a free quantity, which will take its value dynamically.

To investigate the thermodynamic behavior it is useful to analyze the thermodynamic potential Ω in analogy to the non-relativistic case as

$$\Omega(T, f) = -P(T, f) = -T \ln \frac{Z}{V},$$

where P is the pressure, T the temperature, and V the volume. Z is the generating functional. For the following purposes, it is sufficient to use the so-called mean-field approximation. In this case, the interaction terms are neglected. Without going into details, the thermodynamic potential can be evaluated directly, since the functional integral becomes Gaussian. It reads

$$\begin{aligned}\Omega(T,f) &= -\mu^2 f^2 + \lambda f^4 \\ &\quad + \int \frac{d^3p}{(2\pi)^3}\left(\frac{\omega_1^2+\omega_2^2}{2} + T\left(\ln\left(1-e^{-\frac{\omega_1}{T}}\right)+\ln\left(1-e^{-\frac{\omega_2}{T}}\right)\right)\right) \\ \omega_1 &= \sqrt{6\lambda f^2 - \mu^2 + p^2} = \sqrt{m_\sigma^2 + p^2} \\ \omega_2 &= \sqrt{2\lambda f^2 - \mu^2 + p^2} = \sqrt{m_\pi^2 + p^2}.\end{aligned} \tag{7.2}$$

The frequencies ω consists of the momenta and the masses of the particles after hiding the symmetry, which are dependent on the value of the condensate f. There are three contributions. The first outside the integral is the classical contribution. The second are the first two terms inside the integral. They are the contributions from quantum fluctuations. The third term represents thermal fluctuations.

To recover the tree-level results, the second term must be neglected and the zero-temperature limit taken. This yields

$$\Omega(0,f) = -\mu^2 f^2 + \lambda f^4.$$

Tis potential has a minimum at non-zero f, $f^2 = \mu^2/(2\lambda)$. Inserting this into the Lagrangian makes the masses explicit.

Something new happens at finite temperature. At small temperatures it is possible to excite the σ or the pions, which then form a thermal bath of non-interacting bosons, and the total pressure is just the sum of their respective pressures. However, the value of f will become temperature-dependent: At each temperature it will take the value which minimizes the thermodynamic potential.

When going to higher temperatures, it is useful to make a high-temperature expansion for the thermodynamic potential. High temperature requires here T to be larger than the scale of the zero-temperature case, which is given by the condensate, which is of order $\mu/\sqrt{\lambda}$. In this case, it is possible to obtain an expansion for Ω. The leading terms up to order $\mathcal{O}(1)$ are given by

$$\Omega(T,f) = \lambda f^4 + \left(\frac{1}{3}\lambda T^2 - \mu^2\right) f^2 - \frac{\pi^2}{45}T^4 - \frac{\mu^2 T^2}{12}. \tag{7.3}$$

This results exhibits one interesting feature. The term of order f^2 has a temperature-

dependent coefficient, which changes sign at[5] $T_c^2 = 3\mu^2/\lambda$. As a consequence, the shape of the thermodynamic potential as a function of f changes. Below T_c, it has a minimum away from zero, as at zero temperature. With increasing temperature, this minimum moves to smaller and smaller temperatures, and arrives at zero at T_c. Hence, at T_c, the value of f changes from a non-zero to a zero value, and the symmetry becomes manifest once more. Above T_c, the minimum stays at zero, and for all higher temperatures the symmetry is manifest.

Replacing f with its temperature dependent value in (7.3) yields the expressions

$$\begin{aligned}\Omega_{T<T_c} &= \tfrac{\mu^2T^2}{12} - \left(\tfrac{\pi^2}{45} + \tfrac{\lambda}{36}\right)T^4 \overset{T=T_c}{=} -\frac{\pi^2\mu^2}{5\lambda^2}\\ \Omega_{T>T_c} &= \tfrac{\mu^4}{4\lambda} - \tfrac{\pi^2}{45}T^4 - \tfrac{\mu^2T^2}{12} \overset{T=T_c}{=} -\frac{\pi^2\mu^2}{5\lambda^2},\end{aligned}$$

which coincide at T_c. Also their first derivatives with respect to the temperature equal at T_c

$$\begin{aligned}\frac{d\Omega_{T<T_c}}{dT} &= -(8\pi^2T^2 + 10\lambda T^2 - 15\mu^2)\tfrac{T}{90} \overset{T=T_c}{=} -\frac{8\pi^2+5\lambda}{\sqrt{300}}\sqrt{\frac{\mu^2}{\lambda}}\\ \frac{d\Omega_{T>T_c}}{dT} &= -(8\pi^2T^2 + 15\mu^2)\tfrac{T}{90} \overset{T=T_c}{=} -\frac{8\pi^2+5\lambda}{\sqrt{300}}\sqrt{\frac{\mu^2}{\lambda}},\end{aligned}$$

but their second derivatives do not

$$\begin{aligned}\frac{d^2\Omega_{T<T_c}}{dT^2} &= \tfrac{\mu^2}{6} - (4\pi^2+5\lambda)\tfrac{T^2}{15} \overset{T=T_c}{=} -\frac{(25\lambda+24\pi^2)\,\mu^2}{30\lambda}\\ \frac{d^2\Omega_{T>T_c}}{dT^2} &= -\tfrac{8\pi^2T^2+5\mu^2}{30} \overset{T=T_c}{=} -\frac{(5\lambda+24\pi^2)\,\mu^2}{30\lambda}.\end{aligned}$$

Thus, a phase transition of second order occurs at T_c. Note that at very large temperatures only the term $\pi^2T^4/45$ is relevant, which is precisely the one of a free non-interacting gas of two boson species, a Stefan-Boltzmann-like behavior.

As stressed previously repeatedly, it is possible that quantum effects could modify the pattern considerably or even melt the condensate. It is therefore instructive to investigate the leading quantum corrections to the previous discussion.

This is also necessary for another reason. If the symmetry becomes manifest once more at large temperatures, the mass of scalar excitations becomes formally tachyonic,

[5]Note that strictly speaking using the high-temperature expansion at this temperature is doubtful. For the purpose here it will be kept since it makes the mechanisms more evident then the rather technical calculations necessary beyond the high-temperature expansion. The qualitative outcome, however, is not altered, at least within the first few orders of perturbation theory.

indicating a flaw of the theory. That can be seen directly by reading off the condensate-dependent masses of the excitations being as usual

$$\begin{aligned} m_\sigma^2 &= 6\lambda f^2 - \mu^2 = -\mu^2\theta(T-T_c) + (2\mu^2 - \lambda T^2)\theta(T_c - T) \\ m_\pi^2 &= 2\lambda f^2 - \mu^2 = -\mu^2\theta(T-T_c) - \frac{\lambda T^2}{3}\theta(T_c - T). \end{aligned}$$

Furthermore, also the Goldstone theorem is violated, as the mass of the Goldstone boson π is no longer zero below the transition temperature[6]. Both problems are fixed by quantum corrections, demonstrating the importance of quantum fluctuations even in the high-temperature phase.

In the expression for the free energy (7.2) the zero-point energy, and thus the quantum fluctuations have been neglected. Using a cutoff-regularization with cutoff Λ their contribution can be determined as

$$\int \frac{d^3p}{(2\pi)^3}\frac{\omega}{2} = \frac{1}{64\pi^2}\left(2m^2\Lambda^2 - m^4\ln\frac{\Lambda^2}{m^2} - \frac{m^4}{2}\right) + \mathcal{O}\left(1, \frac{1}{\Lambda}\right).$$

where the constant terms $\mathcal{O}(1)$ do not depend on the mass. This contribution is quadratically divergent and has to be regulated. This can be done by introducing into the Lagrangian (7.1) the necessary counter-terms

$$\delta\mu^2(\sigma^2+\pi^2) - \delta\lambda(\sigma^2+\pi^2)^2.$$

Repeating the calculation for the free energy yields at zero temperature the expression

$$\begin{aligned} \Omega(0,f) &= -(\mu^2+\delta\mu^2)f^2 + (\lambda+\delta\lambda)f^4 \\ &\quad + \frac{1}{64\pi^2}\left(2(m_\sigma^2+m_\pi^2)\Lambda^2 - m_\sigma^4\ln\frac{\Lambda^2}{m_\sigma^2} - m_\pi^4\frac{\Lambda^2}{m_\pi^2} - \frac{m_\sigma^4}{2} - \frac{m_\pi^2}{2}\right). \end{aligned}$$

To determine the renormalization constants two conditions will be implemented. One is that the free energy is finite when the cutoff is send to infinity. The second is that the Goldstone boson mass is zero, equivalent to requiring that $f = \mu^2/(2\lambda)$, and required by the Goldstone theorem. Both conditions can be satisfied by the choice

$$\begin{aligned} \delta\mu^2 &= \frac{\lambda\Lambda^2}{4\pi^2} + \frac{\lambda\mu^2}{4\pi^2}\ln\frac{\Lambda^2}{2\mu^2} + \mu^2\frac{\delta\xi}{\lambda} \\ \delta\lambda &= \frac{5\lambda^2}{8\pi^2}\ln\frac{\Lambda^2}{2\mu^2} + \delta\xi. \end{aligned}$$

[6] In a full quantum treatment, the role of the Goldstone boson could be played at finite temperature by some composite excitation instead. However, at the mean-field level no such excitations are available, and thus the Goldstone theorem is violated.

Herein the contribution $\delta\xi$ is not determined by these conditions, and can be set at will by other renormalization conditions. This indicates that both conditions are not independent. This fixes the thermodynamic potential at zero temperature. It can be shown that no new counter terms are necessary at non-zero temperature. Therefore, the high-temperature expansion can be performed as previously.

Performing once more a high-temperature expansion is possible. However, in this case also higher-order terms have to be kept, since the vacuum energy has now contributions of order $\mathcal{O}(m^4\ln(m^2/\mu^2))$. At higher order in the high-temperature expansion terms of order $\mathcal{O}(m^4\ln(m^2/T^2))$ appear, which combine to relevant terms. The result is

$$\begin{aligned}\Omega(T,f) &= -\frac{\pi^2}{45}T^4 - \frac{\mu^2T^2}{12} - \frac{(m_\sigma^3+m_\pi^3)T}{12} + \frac{\mu^4}{32\pi^2}\ln\frac{8\pi^2T^2e^{-2\gamma+\frac{3}{2}}}{\mu^2}\\ &\quad -\mu^2f^2\left(1+\frac{\delta\xi}{\lambda}+\frac{\lambda}{4\pi^2}\ln\frac{8\pi^2T^2e^{-2\gamma+1}}{\mu^2}-\frac{\lambda T^2}{3\mu^2}\right)\\ &\quad +\lambda f^4\left(1+\frac{\delta\xi}{\lambda}+\frac{5\lambda}{8\pi^2}\ln\frac{8\pi^2T^2e^{-2\gamma+1}}{\mu^2}\right).\end{aligned}$$

The critical temperature can be determined again as the point where f vanishes, yielding

$$T_c^2 = \frac{3\mu^2}{\lambda}\left(1+\frac{\delta\xi}{\lambda}+\frac{\lambda}{4\pi^2}\ln\frac{24\pi^2e^{-2\gamma+1}}{\lambda}\right).$$

To attach a final value it would be necessary to determine the value for $\delta\xi$ by some other renormalization condition. To order λ, which is the current order, the final result for T_c will then not depend on this renormalization prescription. One obvious possibility would be to give T_c its (hypothetically) experimentally measured value, as T_c may not depend on the renormalization process: As a physical observable, it is renormalization-group invariant.

To obtain the corrections for the masses, it is necessary to calculate the corresponding self-energies. Without going into the details, the result to the present order in λ is given at high temperatures and after renormalization by

$$\Pi_\sigma = \Pi_\pi = \frac{\lambda T^2}{3},$$

and thus momentum independent. It is therefore a correction to the mass. The complete mass to this order is therefore

$$\begin{aligned}m_\sigma^2 &= 2\mu^2\left(1-\frac{\lambda T^2}{3\mu^2}\right)\theta(T_c-T)+\frac{1}{3}\lambda\left(T^2-\frac{3\mu^2}{\lambda}\right)\theta(T-T_c)\\ m_\pi^2 &= \frac{\lambda}{3}\left(T^2-\frac{3\mu^2}{\lambda}\right)\theta(T-T_c).\end{aligned}$$

These results yield a number of interesting observation. First, since T_c is larger[7] than $3\mu^2/\lambda$, the mass of the σ is always positive, stabilizing the system. Secondly, in this case the mass of the Goldstone boson is always zero below the phase transition temperature, in agreement with the Goldstone theorem. Above the phase transition, the masses of both particle degenerate, and the symmetry is manifest once more also in the spectrum. These properties are generic for symmetries hiding by a condensate which thaws with increasing temperature. Also that the mean-field approximation is in general insufficient is a lesson which should be kept duly in mind. Of course, at the present time much more sophisticated methods are available to treat this problem, though they are in general very complicated.

It should be noted that in gauge theories, like QCD, an important problem arises when doing finite-temperature perturbative calculations. Due to the infrared effects of the massless gauge bosons higher orders become amplified. As a consequence, all order in perturbation theory from order $g^6 \ln g$ onwards contribute equally, and the perturbative series becomes meaningless. This is the so-called Linde problem. Thus, non-perturbative methods are necessary beyond this order. Also, there are non-extensive quantities, which are never dominated by hard scatterings, like a spatial version of the Wilson potential, and even at infinite temperature non-perturbative methods are required, as will be discussed next.

7.3 Very high temperatures

The natural question arising is then whether there are at all any effective difference between a gas of free quarks and gluons and actual QCD. The answer to this can be gleaned from a more field-theoretical investigation. In fact, Infinite temperature has two effects. First of all, due to an effective mass proportional to the temperature all fermions are effectively infinitely massive. Secondly, due to the absence of any temporal dynamics, the theory becomes essentially three-dimensional, where the superfluous gauge field degree of freedom acts like an additional scalar field. However, a gauged scalar in three dimensions is not a free theory, but rather manifests most of the pertinent features, especially confinement, of strongly interacting Yang-Mills theories. Thus, strong interactions remain, they just are completely subleading for thermodynamic bulk quantities.

In analogy to electrodynamics, based on the assignment of Lorentz indices in the field-strength tensor, such interactions are also often denoted as (chromo)magnetic interactions,

[7] It is not obvious that $\delta\xi$ cannot be negative and large, thus making the improved estimate for T_c smaller than before. However, it turns out not to be the case at this order for any renormalization prescription.

in contrast to the (chromo)electric interactions. After all, the effective three-dimensional theory, with parameters which can be determined from the four-dimensional theory, 'lives' in the three original spatial dimensions. That this implies an Euclidean signature is not a contradiction, since the four-dimensional theory in equilibrium has anyhow no time-dependence.

7.4 Small densities and the critical endpoint

When moving away from the zero temperature axis, the situation becomes more complicated. Generically, most model calculations find that when following the cross-over line it eventually turns into a real phase transition again at some density, which is somewhere around, but usually smaller than, nuclear density. In the chiral limit, the line just remains second order, and then turns into a first order transition. This point is known as a critical (end-)point. Its precise location is unknown, and strongly depends on the model in question. All calculations based on full QCD, either of lattice type or using functional methods, are so far restricted to the condition $T/\mu_q \lesssim 1$, μ_q being the quark chemical potential. where the effects of the sign problem and/or baryons are not too strong. They all agree that no critical point is found inside this cone. This is also confirmed in model studies accessible to genuine non-perturbative methods[8].

There is, however, no principle necessity for the existence of an endpoint, and models exist without one. It is hence an open question, whether one exists.

7.5 Nuclear matter

At low temperature, after crossing the gas-liquid transition, a region of nuclear matter is entered. While in pure QCD the distinction into proton and neutrons is, up to the mass difference of up and down quarks, meaningless, in real matter neutrons and protons can transform into each other by the weak interaction. The so-created nuclear matter is stable, with a free energy of about -16 MeV/nucleon, and the preferred state is hence infinitely extended nuclear matter. Below the transition, only lumps of not to many nucleons, the heaviest stable nuclei, are present. It is worthwhile to note that without electromagnetic repulsion and weak decays, in principle also much heavier nuclei could be stable.

Of course, other hadrons can now be created by hadronic interactions as well, and may be more stable. At sufficiently large densities, it is a possibility that pions and kaons

[8]E. g. strong coupling or gauge theories without sign problem, like 2-color QCD or G_2 QCD, i. e. QCD with the gauge group SU(3) replaced by the group SU(2) or the exceptional group G_2, respectively.

can be produced in appreciable number and condense. Whether this happens or not, and if it happens whether this induces a phase transition, is unknown. One possibility widely discussed was that possibly even strange matter, where net strangeness is created by flavor-violating weak interactions, could become more stable than non-strange matter. Though by now this appears unlikely, it cannot yet be fully excluded.

7.6 High densities

At very high densities, the hadrons again substantially overlap, this time only due to sheer number. Similar to the case of high temperatures, hard parton exchange reactions become very important. Hence, once more a localization of partons within a given hadron is no longer possible. In addition, rescatterings will make individual hadrons no longer distinguishable, and only collective excitations can again be stable. In this situation, it appears possible that the physics is dominated by the quarks. Gluon are playing here a possibly smaller part, as in contrast to finite temperature they do not couple to the chemical potential, and thus their is no direct increase of their number.

Since quarks are fermions, they can in principle act as fermions in ordinary matter, and especially form Fermi surfaces. However, it strongly depends on how the strong interactions are effectively modified by the dense medium, since any long-rang interaction immediately destabilizes a Fermi surface. But given the necessity of local gauge invariance and the number of quarks, the effective interaction between quarks may very well be short range. In fact, it may well be that in a sufficiently dense medium the quarks will become essentially localized by the Pauli principle, effectively quenching the theory, especially if crystal-like structures form. The question whether chiral symmetry becomes restored or not at finite density then becomes a question of details.

Another possibility is that the quarks form colored versions of Cooper pairs, inducing a version of a superconductor, a color superconductor. Indeed, most model calculations, as well as perturbative calculations, support such a phenomena. A physical interpretation of such exotic phases of matter at high baryon densities is, at best, difficult. The problem is that the nature and structure of the colored fluid can be changed by gauge transformations. It can therefore serve at most as a picture, and the true physics has to be determined using gauge-invariant quantities. This is also an important caveat when comparing to model studies, in which color becomes a global symmetry, e. g. the NJL model. In such models, color is observable, and therefore states like color superconductors are physical, which is not the case for QCD.

Asymptotically high densities behave in a very similar fashion as at high temperatures.

Again, the bulk quantities are determined by hard scatterings, and non-perturbative effects contribute only sub-leading to them. Hence, they are accessible to perturbative calculations. As in the case of high temperature, this situation is only reached at asymptotically high densities, and thus in regions where other interactions, like the weak interactions, may start to play a role. Nonetheless, the pure-QCD limit of very high density is essentially that of a weakly interacting degenerate Fermi gas of quarks.

In should be noted that in model studies, depending on the details of the model, the number of possible different phases quickly proliferates, especially if besides the usually assumed amorphous phases also crystalline phases are taken into account. The true situation for QCD is at the time of writing unknown. Results from non-perturbatively accessible gauge theories differ strongly, depending on the theory in question. Hence, in contrast to the high-temperature case, the phase structure at finite density seems to be highly non-generic.

7.7 Neutron stars

The physical important application of high densities and low temperatures are neutron stars. Though the typical densities inside a neutron star can be estimated from their size and mass, it is not clear, which type of phases could be encountered. At the surface and the outer crust the densities are certainly such that nuclear matter prevails, possible in different types of spatial structures, e. g. crystals or long-range ordered (so-called pasta phases). However, closer to the core, the situation becomes less obvious. The various model studies present a multitude of possible phases at the relevant densities, from purely nuclear ones over those including other hadrons, markedly strange ones, until pure quark phases like color superconductors.

So far, the best constraints come from astronomical observations, especially masses, spin-down rates, cooling rates, sizes and magnetic fields. They all, especially the presence of two-solar-mass neutron stars, strongly indicate a very stiff equation of state. Such a large stiffness is mainly found in nuclear phases, or at best nuclear phases with small contributions from other hadrons. Only in very few models have phases dominated by quarks and gluons a large stiffness, and are therefore compatible with such heavy neutron stars. This strongly indicates that the interior of neutron stars, provided QCD itself is similar to the majority of models, is likely hadronic.

However, only of a small number of neutron stars are all properties known with very good accuracy. This leaves open the possibility that several distinct types of neutron stars exist, which, depending on parameters like size-to-mass ratio or perhaps initial conditions,

could have differing internal structures. Particularly, it may be that besides the heavy neutron stars with a nuclear structure lighter neutron stars could exist with a predominantly strange hadron core (strange stars), or stars with a core and large part of the crust dominated by quark and gluons (quark stars).

Within the next few years, possibly as early as 2017, a new source of information will become available on neutron stars. Depending on the relative frequency of neutron star binaries then first neutron star mergers can be detected with the then available gravitational wave detectors. Already about 40 mergers will be sufficient to determine further properties, like the bulk modulus, of the interior matter of neutron stars, with ever increasing precision the more detected. This should further constrain the type of matter encountered.

7.8 Heavy-ion collisions

Besides astronomical observations, an experimental possibility to investigate the QCD phase diagram are collisions with heavy ions. The higher and more central, i. e. head-on, the collisions are, the higher the temperature of the so-created medium. At the same time, the large amount of energy will create a large amount of both baryons and anti-baryons, washing out the initial chemical potential to almost zero. On the other hand, low energies and off-central, or peripheral, collisions, will have a larger chemical potential, and lower temperature.

Such collisions are non-equilibrium situations. Naively they should not be describable using equilibrium physics, i. e. thermodynamics. Fortunately, the systems seem to equilibrate very quickly, in about less than a tenth of the collision time, and many features of low-energy particles are well described both with thermodynamics and also hydrodynamics. Many features of high-energetic particles are accessible to perturbation theory in a medium. Still, a full description from QCD is so far not possible, and most models, as well as semi-classical transport theory, can only describe part of the experimental observations. In many cases, the interpretation of various experimental observables are still under debate.

Still, the results strongly support a rapid transition into a very different medium at high temperatures and low chemical potentials. At larger chemical potentials, the temperatures still remain rather high, and it is not clear whether it could be possible to reach at least as low temperatures as encountered in supernovas, lest in neutron stars. Also, there is not yet any statistically reliable and unambiguous signal which would decide the presence of a critical point.

The problems arise mainly because all the observations are rather indirect. Instead of probing the medium directly, the medium first undergoes chemical freezeout, i. e. the hadrons cease to interact inelastically, and afterwards kinetic freezeout, i. e. also elastic interactions end, before the results of the initial collision, also called fireball, reach the detectors. Hence, many steps have to be traced back. An alternative are probes which do interact with the medium weakly, i. e. electromagnetic probes like photons or W and Z bosons. They can be detected, but they carry only limited information on the medium besides changes in the production mechanism. Hence, a full description of a heavy-ion collision remains one of the serious challenges in many-particle physics, which would go far beyond the current lecture.

A last interesting remark is the onset of thermodynamic properties of collisions. Originally, it was anticipated that very high collision energies and large heavy ions, like lead, are necessary to have enough particles involved for equilibration. More recent results show, however, that already proton-heavy ion, and partly even very high energetic proton-proton collisions with production of many soft particles, show thermodynamic behavior, as soon as there are about 20 or more particles in the final state. A possible explanation is that the strong increase of partons at low x effectively makes the protons already a medium, and when probed at that small x the result is rather a probe of a QCD medium, rather than of individual partons. The details of this are also open questions.

Index