

Quantum Field Theory II: Gauge Theories

Lecture in SS2024 at the KFU Graz

Axel Maas

Contents

1	Introduction	1
2	Gauge theories	3
2.1	Global symmetries	3
2.2	Classical Maxwell theory	5
2.3	Some basics of group theory	8
2.4	Orbit structure	13
2.4.1	Orbits	13
2.4.2	(Non-)linear representations	14
2.5	Yang-Mills theory	14
2.6	Where the gauge freedom comes from	18
2.6.1	The Abelian case	18
2.6.2	The non-Abelian case and the Gribov-Singer ambiguity	21
2.7	Matter, representations, and coupling universality	24
2.7.1	Matter fields in Abelian gauge theories	24
2.7.2	Matter fields and coupling universality in non-Abelian gauge theories	25
2.7.3	Explicit theories	29
2.8	Beyond Yang-Mills theory	30
2.8.1	Yang-Mills type	31
2.8.2	Closure only on-shell	31
2.8.3	Reducible gauge theories	32
2.8.4	Reducibility on-shell	32
2.8.5	Towards gravity	33
3	Classical solutions	34
3.1	Time evolution and initial conditions	34
3.2	Topological excitations	36
3.2.1	Instantons in quantum mechanics	37

3.2.2	Yang-Mills theory	40
3.2.2.1	Topology	40
3.2.2.2	Instantons	42
3.2.3	Other topological excitations	47
3.3	Classical tunneling	48
4	Perturbative quantization of Yang-Mills theory	49
4.1	The path integral	49
4.1.1	Heuristic introduction	50
4.1.2	Functional analysis	54
4.1.3	Sources	56
4.2	Quantization of Maxwell theory	57
4.3	Gauge-dependent correlation functions	63
4.4	Quantization of Yang-Mills theory	65
4.5	BRST and asymptotic states	68
4.5.1	BRST symmetry	68
4.5.2	Constructing the physical state space	71
4.5.2.1	Negative-norm states	71
4.5.2.2	BRST algebra	72
4.5.2.3	Physical states	73
4.5.2.4	Observable particles	74
4.6	General anti-field approach	76
4.6.1	BRST construction revisited	76
4.6.2	Observables	79
4.6.3	Application as gauge-fixing to the path integral	81
4.6.4	Beyond Yang-Mills theory	84
4.6.5	Quantum Master Equation	86
4.7	Ward-Takahashi identities and Slavnov-Taylor identities	89
4.7.1	Ward-Takahashi identities	90
4.7.2	Slavnov-Taylor identities	92
5	Perturbative treatment	95
5.1	Perturbation theory	95
5.1.1	Cross-sections and decays	95
5.1.2	General construction	97
5.2	Renormalization	108
5.2.1	Cutoff regularization	109

5.2.2	Renormalization prescription	110
5.2.3	Counter-term structure	111
5.2.4	Renormalization schemes	114
5.2.5	Renormalization group	116
5.2.6	BPHZ and MOM schemes	119
5.2.7	Gauge-symmetry-breaking regulators	120
5.2.8	Dimensional regularization	121
5.2.9	Lattice and infrared regularization	124
5.2.10	Composite operators	126
5.3	Examples of perturbative calculations	127
5.3.1	Gauge boson scattering	127
5.3.2	Ghost self-energy	129
5.3.3	Gauge boson self-energy and gauge invariance	134
5.4	Running couplings, Landau poles, and asymptotic freedom	135
5.5	Infrared singularities	137
6	Quantization beyond perturbation theory	139
6.1	The path integral beyond perturbation theory	139
6.2	Non-perturbative BRST symmetry	146
6.3	Gauge-invariance beyond perturbation theory	150
6.3.1	Abelian gauge theories and the Dirac phase	150
6.3.2	Composite states as physical degree of freedoms	153
6.3.3	Gauge invariance as redundancy	155
6.3.4	Asymptotic state space	155
6.4	Topology at the quantum level	156
6.4.0.1	θ vacua	156
6.4.0.2	Tunneling amplitude	157
6.4.1	Light fermions	158
7	Anomalies	159
7.1	Introduction	159
7.2	Global anomalies	159
7.2.1	Classical level	160
7.2.2	One-loop violation	162
7.3	Local anomalies	165
7.3.1	Anomalies as a quantum effect	166
7.3.2	Full expression for the anomaly	167

7.3.3	Anomaly cancellation	170
7.4	Relation to topology	171
7.5	Witten anomalies	173
7.6	Anomalies in the BRST framework	174
	Index	176

Chapter 1

Introduction

Gauge theories play a fundamental role in all of modern fundamental physics, but are at the same time enigmatic and challenging. The aim of this lecture is to provide an introduction to their quantum version, in the form of quantum gauge field theories. It turns out that a suitable language for these kind of theories is the path integral, which will be used throughout.

The paradigmatic classical gauge theory is electrodynamics. However, the range of gauge theories is much vaster than this. As the corresponding classical gauge theories are the starting point of a path-integral quantization, these will be investigated in some detail in chapter 2. It is also important to understand the classical solutions, which will be discussed in chapter 3.

Only in chapter 4 will Yang-Mills theory be quantized, as this requires much more techniques due to the non-Abelian nature. In fact, it will only be possible at this stage to quantize the theory at the level of perturbation theory. However, as both in practice and for the sake of concept formation perturbation theory is invaluable, this is anyhow a necessary step. This will be continued in chapter 5, where various perturbative processes will be discussed in detail. As at loop level also in gauge theories divergences arise, this will not be possible without an extensive discussion of renormalization and infrared problems.

After this, it is possible to delve into gauge theories beyond perturbation theory in chapter 6. This issue is indeed challenging, and not entirely solved until today. This will lead also to questions what are the physical degrees of freedom of gauge theories beyond perturbation theory, and will show that gauge symmetry leads to deeper ontological problems as well. Here QED will be again a role model, as it is possible to spell out both the problem and the solution in this case explicitly, although at the price of exponentially more complicated calculations. It will then be seen how this problem translates to non-Abelian gauge theories, and where the resolution becomes even more involved.

This lecture will not be able to cover actual calculations beyond perturbation theory, as the necessary methods are quite involved, and thus deserve lectures of their own. These are available for lattice gauge theory and for functional methods, although this is not an exhaustive list of possible methods. Some other methods will be encountered in more specialized lectures.

Finally, in contrast to non-gauge theories, which can always be quantized even if they are trivial, this is not the case for gauge theories. This is due to anomalies, which will therefore be treated in chapter 7, and will round out the discussion of gauge theories. This will show that the integral measure in the path integral can play a more important role than it usually appears to have.

However, gauge theories are a vast subject, and thus many issues can only be touched upon. Especially, concrete realizations of gauge theories as theories of, e. g., the strong force and weak force require their own dedicated lecture. There is also a vast literature on gauge theories. In the preparation of this course, the following books and review articles have been used:

- Peskin et al. “An introduction to quantum field theory”, Perseus
- Böhm et al. “Gauge theories”, Teubner
- O’Raifeartaigh, “Group structure of gauge theories”, Cambridge
- Weinberg, “The quantum theory of fields I+II”, Cambridge
- DeWitt, “The global approach to quantum field theory”, Oxford
- Haag, “Local quantum physics”, Springer
- Lavelle et al., “Constituent quarks from QCD”, Phys. Rept. 279
- Maas, “Gauges bosons at zero and finite temperatures”, Phys. Rept. 524
- Maas, “Brout-Englert-Higgs physics”, PPNP 106

Of course, these are my personal choices, and not necessarily the best ones to learn the topic from. Thus, I strongly recommend that any student should test various books and choose the one best suited for their own purposes.

Chapter 2

Gauge theories

As will be seen, gauge theories come in different types. However, in all cases, in the end the same geometrical structures appear, just their geometric complexity differs. Thus, it is always helpful to start out with the theory with the simplest geometrical structure. As will be seen, this is the quantum version of Maxwell theory, i. e. a theory just of electric fields and magnetic fields. In fact, this theory is trivial in the sense that it is not interacting. Thus, akin to the case of the free scalar and fermions, this will be the simplest example of a gauge theory.

It is worth mentioning, and is shown in the lecture quantum field theory I, that every theory describing elementary particles with a spin greater than 1/2 seems to necessarily involve gauge degrees of freedom, if a local description, i. e. one in which no multi-position integrals appear in the action, is required. Why this is so is as of yet unknown. But what this implies will be discussed in great detail in due course.

2.1 Global symmetries

Before discussing gauge symmetries it is useful to repeat a few statements about global symmetries. This will be taken up again in more details in section 2.4, but some preliminary remarks are necessary for the following.

In the simplest case at the classical level, a symmetry is any (infinitesimal) transformation of a set of fields $\phi^a(x)$

$$\phi^a \rightarrow \phi'^a = \phi^a + \delta A^{ab} \phi^b = (1 + \delta A)^{ab} \phi^b \quad (2.1)$$

such that the action is invariant. It is important here that it is sufficient that the action is invariant, not the Lagrangian itself. Thus, the Lagrangian can be allowed to change by a total derivative, which vanishes upon integration. Herein δ is an infinitesimal, constant

parameter, and the matrix A is usually required to be Lie-algebra-valued for a fixed choice of Lie algebra, but this depends on the action in question. Consider, as the arguably simplest case, the Lagrangian for N interacting, real scalar fields

$$\mathcal{L} = \frac{1}{2}\phi^a(\partial^2 + m^2)\phi^a - \lambda(\phi^a\phi^a)^2. \quad (2.2)$$

This so-called linear σ -model is invariant for transformations in which $1 + \delta A$ is an $O(N)$ matrix. Such a symmetry is called global, as it changes the fields at every space-time point in the same way. Since it does not affect space-time it acts in an internal space and is thus called an internal symmetry. In contradistinction, the gauge symmetry of classical electrodynamics is called a local symmetry, as it changes the vector potential at every space-time point in a different way.

From such a situation follows Noether's theorem. This theorem is one of the central reasons why symmetries are so important in physics, as it states that there is a conserved charge associated with any such global symmetry. Such conserved quantities are of central importance for both theoretical and practical reasons. Practical, because exploiting conservation laws is very helpful in solving problems. Theoretical, because conserved quantities define the properties of a system. In addition, the generalized Wigner-Eckart theorem allows to deduce complicated tensor structures based on such symmetries alone.

Essentially, Noether's theorem boils down to the fact that if the action is invariant under the transformation, and thus a variation of a generalized coordinate, the variation can be used to derive a conserved quantity. If A is an element of a Lie algebra a finite transformation is obtained by exponentiation, yielding

$$\phi'^a = (e^{\delta A})^{ab} \phi^b = (e^{\delta A} \phi)^a$$

Because A is hermitian, it defines an observable in a quantum theory. It is thus usually called the associated charge operator, and the observables are the charge. Indeed, electric charge will be defined in this way later.

But for the moment the aim is to see that this implies a conserved quantity. To construct it, it is necessary to analyze the transformation properties of the Lagrangian under the infinitesimal transformation, which is given by

$$\delta\mathcal{L} = \partial_\mu K^\mu,$$

i. e. zero up to a total derivative is allowed, as mentioned above. Of course, $K_\mu = 0$ is also allowed. It must also hold that

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi^a} \delta\phi^a + \frac{\partial\mathcal{L}}{\partial\partial^\mu\phi^a} \partial^\mu(\delta\phi^a) = \partial_\mu K^\mu.$$

This can be simplified using the equation of motion for ϕ^a ,

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

to yield

$$\left(\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \partial_\mu \phi^a} \right) \right) \delta \phi^a + \frac{\partial \mathcal{L}}{\partial \partial^\mu \phi^a} \partial^\mu (\delta \phi^a) = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \partial_\mu \phi^a} \delta \phi^a \right) = \partial_\mu J^\mu. \quad (2.3)$$

This defines a current. The current which has been thus defined is the symmetry current, or Noether current. Finally, this implies that

$$j^\mu = J^\mu - K^\mu$$

needs to define a conserved current, as

$$\delta S = \int d^d x \partial_\mu j^\mu = 0$$

for the transformation to act as a symmetry. Especially, this implies that

$$Q = \int d^d x j^0$$

defines a conserved charge.

Any quantity not invariant under the symmetry will be made up from tensor products of representations of the Lie algebra or Lie group, and thus has to obey the Wigner-Eckart theorem. This essentially says that any quantity transforming non-trivially can be rewritten in terms of the invariant tensors of the symmetry. E. g.

$$\phi^a \phi^b = \langle \delta^{ab} f \rangle,$$

where f is invariant under transformations. It is an exercise in group theory to find the full structure for more complex objects. The quantity f , which is invariant under symmetry transformations, is usually called a form factor or dressing function.

If the symmetry is not exact for whatever reason, this becomes more involved. This is beyond the scope of this lecture, and more can be learned in the lecture on (beyond) standard model physics. Here, except for the case of anomalies, all symmetries will be assumed to be unbroken.

2.2 Classical Maxwell theory

Switching now from theories with a global symmetry to gauge symmetries, the archetypal gauge theory is Maxwell theory. This is the theory of electromagnetic fields without

matter. It is encapsulated in the classical Lagrangian

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu\end{aligned}\tag{2.4}$$

with the gauge field A_μ and the field-strength tensor $F_{\mu\nu}$. From it the three-dimensional electric field and magnetic field can be derived as

$$E_i = F_{0i} = -F_{i0} = \partial_0 A_i - \partial_i A_0\tag{2.5}$$

$$B_i = \epsilon_{ijk}F_{jk} = \epsilon_{ijk}\partial_j A_k\tag{2.6}$$

with the Latin indices running only over the spatial components.

This theory is a gauge theory, because its Lagrangian is invariant under the so-called gauge transformation

$$A_\mu \rightarrow A_\mu^g = A_\mu + \partial_\mu g(x)\tag{2.7}$$

with $g(x)$ an arbitrary function. This is similar to (2.1), except for the appearance of a function rather than a constant. Because the transformation is thus different at every space-time point, this is called a local, or gauge, symmetry. Because the $g(x)$, being ordinary functions, are commuting at every space-time point, this is called an Abelian gauge theory.

It is noteworthy that $F_{\mu\nu}$, and thus \vec{E} and \vec{B} , are invariant under gauge transformations, i. e. they do not change when a gauge transformation is applied,

$$F_{\mu\nu}^g = \partial_\mu A_\nu^g - \partial_\nu A_\mu^g = F_{\mu\nu} + \partial_\mu \partial_\nu g(x) - \partial_\nu \partial_\mu g(x) = F_{\mu\nu}.$$

A quantity which behaves like this is called gauge-invariant. Since gauge transformations are a free choice, and thus gauge-dependent quantities depend on this choice, gauge invariance is a necessary condition for a quantity to be physical. Especially, since the electromagnetic fields are measurable, they need to be gauge-invariant.

The classical equations of motion for this theory are the Maxwell equations

$$\partial_\mu F^{\mu\nu} = 0\tag{2.8}$$

$$\frac{1}{2}\epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} = 0,\tag{2.9}$$

where the latter identity is sometimes also called (a) Jacobi identity. The quantity $*F_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$ is called also the Hodge dual or the dual field strength tensor. Both follow immediately from the antisymmetry of the field strength tensor.

The gauge freedom (2.7) can be used to implement conditions, so-called gauge conditions, on the field. A general gauge condition can be formulated as a functional $C[A_\mu, x]$ fulfilling

$$C[A_\mu, x] = 0$$

E. g., the Landau (or in the classical case also Lorentz and sometimes radiation) gauge is defined as

$$C_{\text{Landau}}[A_\mu, x] = \partial_\mu A^\mu = 0. \quad (2.10)$$

I. e. the gauge freedom (2.7) is used to transform any given gauge field such that the condition (2.10) applies.

It is not possible to implement any gauge condition. It is only possible, if for any arbitrary field configuration A_μ a function $g(x)$ exists such that $C[A_\mu^g, x_\mu] = 0$ can be satisfied. Depending on the form of $C[A_\mu, x_\mu]$ guaranteeing this can be an exercise in algebra, (partial) differential equations, or worse.

To see that the Landau gauge condition is well defined, consider

$$0 = \partial^\mu A_\mu^g = \partial_\mu A^\mu + \partial^2 g.$$

This is a wave equation for g with an inhomogeneity. Since the theory of partial differential equations implies that there exists a solution for any inhomogeneity, this implies that the Landau gauge is well defined.

Note, however, that it may happen that a gauge condition does not fully fix the gauge freedom. E. g., the Landau gauge condition does not do so, as any gauge transformation satisfying a free-wave equation $\partial^2 g = 0$, a so-called harmonic gauge transformation, does not change it. Imposing as boundary conditions that the (gauge) fields vanishes at (spatial) infinity removes also this ambiguity, and fixes the gauge completely.

In fact, it can be useful to leave a certain degree of gauge freedom. E. g., the Coulomb gauge condition $\partial_i A_i = 0$ is only defined up to purely time-dependent gauge transformations. There is thus a whole class of functions satisfying the Coulomb gauge condition. Another examples are axial gauges with some fixed vector n_μ such that $n_\mu A^\mu = 0$. Such gauges can already simplify calculations considerably in certain cases.

In classical physics, it is not necessary to discuss further the residual freedom. Any will do equally well, and often one can do without entirely. This will need to be discussed in more detail in the quantum theory, which will be done in section 2.6.2, where the situation is somewhat more involved.

It should, however, be noted that any term which would give a mass to the gauge bosons, e. g. $m A_\mu A^\mu$, cannot be added to the Lagrangian, as it will immediately break gauge invariances, as can be seen immediately by calculation. In fact, (2.4) is the unique

gauge-invariant Lagrangian perturbatively renormalizable by power-counting in four dimensions. Thus, any additional interactions compatible with gauge invariance would lead to a much more involved theory. Thus, Abelian gauge theories are non-interacting and exactly massless without matter

2.3 Some basics of group theory

Maxwell theory was based on the idea that at every space-time points the elements of a real vector could be changed by a gauge transformation given by a real, scalar function. To cover also the other forces observed so far in nature requires to generalize this concept. It turns out that it requires to allow for fields which are tensor-valued in some abstract internal space, just like with global symmetries, rather than just real functions. This requires some basic amount of group theory and the associated theory of algebras, which will be collected here. In fact, also the global case of section 2.1 can be put usually in the context of group theory.

To repeat, start with a group. A group \mathcal{G} is a set of elements g with an associative operation \cdot , which maps two elements g and h of the set into an element of the set, $g \cdot h \in \mathcal{G}$, and there exists a unique identity element e as well as a unique inverse element g^{-1} to every element g , such that $g \cdot g^{-1} = g^{-1} \cdot g = e$ and $e \cdot g = g \cdot e = g$. Note that the operation does not need, but can, be commutative. If it is not, $g \cdot h \neq h \cdot g$, the group is called non-Abelian. Otherwise the group is Abelian.

A mapping of the elements g of the set to a linear operator or tensor $D(g)$ in a vector space such that the group structure is conserved, $D(g)D(h) = D(g \cdot h)$, is called a representation. Such a representation needs not to be faithful, i. e. $D(g) = D(h)$ is allowed, provided the action of the operation is still represented correctly. However, note that not every group can be represented non-trivially, i. e. other than setting $D(g) = 1$ for all g , in every vector space. Sometimes the vector space needs to have a certain number of dimensions and/or needs to be complex.

Generalizations of Maxwell theory are then constructed by having fields which no longer map a space-time position to a real number, but to elements related to the group structure. It is important to note that the corresponding vector space is not position space, but an additional vector space.

For quantum field theory the most relevant subclass of groups are the Lie groups. They are characterized in the following way: Given a matrix representation of a Lie group, it is possible to write

$$D(g) = e^{i\tau^a \phi^a}, \quad (2.11)$$

with the τ^a called the generators of the group and the ϕ^a are real numbers. The generators can be chosen hermitian, i. e., $\tau_a = \tau_a^\dagger$. Note that, by definition¹, $D(g^{-1}) = D(g)^\dagger$. The τ^a form the basis of a vector space, with the real numbers as field. Thus, the exponent in (2.11) can be any element of this vector space. The algebras of Lie groups are called Lie algebras. Note that the representation (2.11) defines also a representation of the generators, and thus the algebra in the same vector space. In the following, groups will be denoted by capital letters, and algebras by small letters. Within this lecture only compact Lie groups will be considered for non-space-time symmetries, for which in addition $\text{tr}(\tau^a \tau^a)$ is a real and symmetric matrix with positive spectrum.

For finite-dimensional Lie algebras there exists a definite number N of generators τ^a , $a = 1, \dots, N$, giving the dimension of the algebra vector space. It is also defined to be the dimension of the group, $N = \dim \mathcal{G}$. It is important to note that the number N is independent of the actual dimension of the vector space of the representation of the group always the same. Thus, depending on the vector space in which the representation is constructed, also the τ^a will be matrices of different dimensions. Note that because of (2.11), a generic group element can be expanded for infinitesimal ϕ_a as $1 + i\tau^a \phi_a$. Thus the algebra describes infinitesimal group elements around the identity. This implies that the group is continuous. Continuous groups are a special case of groups, and compact Lie groups are again a special group of continuous groups, as further restrictions apply.

It should be noted that the relation (2.11) is actually multivalued. For the same algebra it is possible to groups differing by a center, i. e. a set of elements commuting with all other elements. These groups differ in terms of their connectedness. There is for any algebra always a unique covering group, i. e. a simply connected group of which all other groups belonging to the same algebra are subgroups.

Moreover, for the representation to respect the group structure it can be shown that the τ^a are required to respect the commutation relations

$$[\tau^a, \tau^b] = i f_{abc} \tau^c.$$

with the totally anti-symmetric structure constants f_{abc} . This algebra is fulfilled no matter the dimensionality of the actual matrices. Thus, the algebra can eventually be defined also in an abstract way without choosing actual matrices. Because of this, an actual matrix form of the generators is also called a representation. The structure constants fulfill the Jacobi identity

$$f_{abe} f_{cde} + f_{ace} f_{dbe} + f_{ade} f_{bce} = 0. \quad (2.12)$$

This identity will be used multiple times in this lecture.

¹Again, this is not necessary, but all that is needed for this lecture.

It should be noted that multiple groups can have the same algebra. Groups having the same algebra differ by their center, Z_G , which is the set of group elements $z \in Z_G$ which satisfy $[z, D(g)] = 0$ for all group elements g . E. g., it can be shown that $SU(2)$ and $SO(3)$ both have the same algebra, and differ only by the discrete group $Z_2 = \{1, -1\}$. The difference in the groups imply in which vector spaces a (faithful) representation is possible in which form, and whether the representations form a simply connected space or not. While interesting, these subtle points will only be returned to if need be, which is in most cases in (quantum) field theory not the case.

An example is the $su(2)$ algebra, with $N = 3$, having thus three generators. As a two-dimensional, complex representation the Pauli matrices,

$$\begin{aligned}\tau^1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ \tau^2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ \tau^3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.\end{aligned}\tag{2.13}$$

are usually chosen. Of course, any unitarily equivalent representation will do as well, as this only corresponds to a change of basis in the vector space defined by the generators. Here, it should be noted once more that several vector spaces play a role: The space-time vector space, the vector space in which the representation of a group exists, and the vector space which is defined by the algebra. In this example, the space-time vector space is the usual d -dimensional one, the representation vector space is two-dimensional, and the vector space defined by the generators is three-dimensional. For this particular case

$$D(g) = 1\omega_0 + \tau^a\omega_a$$

with $\omega_i^2 = 1$. However, it is not true in general that the generators can be used so simply to create the representation of the group elements.

The most relevant algebras for the standard model are $su(2)$ and $su(3)$. Though potentially any (Lie) algebra/group can be used to construct a valid quantum field theory, this lecture will mostly concentrate on the case of $su(N)$. Thus, for the sake of simplicity, in the following mostly the expressions for $su(N)$ will be given. Others can be obtained in a straightforward way. Maxwell theory can be considered as being derived from the algebra formed by the real unity, yielding as algebra space the real numbers, called $u(1)$, which by exponentiation yields the unit circle. The $u(1)$ group is the special case of all f^{abc} being zero, and the algebra being one-dimensional. This algebra is equivalent to $so(2)$, and it is the only non-compact group to be encountered in this lecture for an internal space.

The $\mathfrak{su}(2)$ algebra has the total-antisymmetric Levi-Civita tensor as structure constant, $f^{abc} = \epsilon^{abc}$ with $\epsilon^{abc} = 1$. The algebra $\mathfrak{su}(3)$ 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 $\mathfrak{su}(3)$ can be constructed, the so-called Gell-Mann matrices,

$$\begin{aligned} \tau^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \tau^2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \tau^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \tau^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} & \tau^5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} & \tau^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \tau^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} & \tau^8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \quad (2.14)$$

In general, there are $N^2 - 1$ base vectors for $\mathfrak{su}(N)$, but the dependency for the other algebras is different.

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 $\mathfrak{su}(2)$, this is only one generator, for $\mathfrak{su}(3)$ there are two. The size of the Cartan subalgebra is called the rank of the group.

The lowest-dimensional realizations of the commutation relations are the fundamental representations of the algebra or group, of which there are rank many, but not necessarily of the same dimensionality. 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 in the same vector space. For $\mathfrak{su}(2)$, there is only one. For $\mathfrak{su}(3)$, they are two, the second one created by using the $-\lambda^{a*}$ matrices.

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

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

which are three-dimensional for $\mathfrak{su}(2)$ and eight-dimensional for $\mathfrak{su}(3)$. There are also cases in which the fundamental and the adjoint representation coincide. There are an infinite number of further representations with increasing dimensionality, 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

$$R_{ij}^a R_{jk}^a = \delta_{ik} C_R, \quad (2.15)$$

being for $\mathfrak{su}(N)$ $(N^2 - 1)/(2N)$ 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 actual Haar integral measure. The Haar measure is invariant under a variable transformation using a different group-element, i. e., for $g \rightarrow gg'$ 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

$$\int_G dg = 1,$$

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

$$\begin{aligned} I(\theta) &= \frac{1}{V_G} \det M \\ g^{-1} \frac{\partial g}{\partial \theta^a} &= i \tau^b M^{ba} \\ V_G &= \int_G \det M \Pi_{a=1}^N d\theta^a \end{aligned} \quad (2.16)$$

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

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

$$g = e^{i \frac{\theta_a \tau_a}{2}} = \begin{pmatrix} \cos \frac{\eta_1}{2} + i \sin \frac{\eta_1}{2} \cos \eta_2 & i e^{-i \eta_3} \sin \frac{\eta_1}{2} \sin \eta_2 \\ i e^{i \eta_3} \sin \frac{\eta_1}{2} \sin \eta_2 & \cos \frac{\eta_1}{2} - i \sin \frac{\eta_1}{2} \cos \eta_2 \end{pmatrix}, \quad (2.17)$$

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 (2.16), 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 $16\pi^2$.

2.4 Orbit structure

This brief section will analyze some implications of groups and group structures when applying group elements to something, usually the vectors of the vector space in which a representation is defined. It should be noted, however, that the following could be also defined without an actual representation, as long as the action of a group element on something is defined. But within this lecture this will always be realized in a representation. This formalism will later be used to represent gauge theories as collections of orbits.

2.4.1 Orbits

The basic concept is that of an orbit. Take a single vector v of the vector space in which exists a representation. Then an orbit, sometimes also called group orbit, $\mathcal{O} = \{D(g)v\}$ is defined as the set of all vectors obtained under application of all group elements from this single vector. In a sense, this generalizes the concept of a ray. An example are the hyperspheres which are obtained under the application of the rotation groups on a single vector pointing to the surface of the hypersphere, e. g. in \mathbb{R}^2 the orbit of $(1, 0)^T$ would be $(\cos \alpha, \sin \alpha)^T$ under the two-dimensional rotation group.

This structure has a number of consequences. First, any vector space on which a representation acts can be decomposed into distinct orbits, though the number of orbits does not need to be countable. Second, group invariants, i. e. elements of the vector space satisfying $D(g)v = v$, are trivial, i. e. one element, orbits.

From the point of physics, it is useful that orbits can also be classified by their invariant tensors. An invariant tensor of rank n is a tensor T in the representation space of the group which satisfies $D_1(g) \dots D_n(g)T = T$, i. e. it remains invariant under application of group elements. E. g. in the simplest case of SO(n) and the vector representation, the invariant tensor is proportional to the unit matrix. An invariant tensor yields a characterization of an orbit, i. e. a quantity which is different for different orbits. In the example this is the length of the vector, which is invariant under rotation, and thus characterizes an orbit.

Since representations can have more than one invariant tensor, orbits may have more than one invariant to characterize them. Then, two orbits do only agree if they agree in all the quantities characterized by invariant tensors. E. g. for the tensor representation of $SU(3)$, build from matrices $V_{ab} = v_a v_b$ with v in the fundamental representation, there are two invariant tensors. Therefore, there are two quantities for any orbit which are invariant under the action of the group. One is again the length. The other is a more complicated quantity, which follows from the fact that every matrix of this representation satisfies $d_{ab}^c V_{ab} = 0$, and d is hence a traceless, symmetric tensor.

2.4.2 (Non-)linear representations

Though most of the previous has dealt with linear representations, i. e. representations where group elements g have been mapped to matrix representations $D(g)$ such that they act as

$$D(g)_{ab} x_b$$

on the elements x of the representation (Hilbert) space. This is not necessary.

The only requirement for a representation is actually that it maintains the group composition law. Thus, it is possible to formulate a non-linear representation

$$f_a(x, g),$$

provided

$$f(f(x, g), h) = f(x, g \cdot h)$$

where the group composition is maintained. Then, this map is a non-linear representation.

Though non-linear representation of groups (and algebras) are rare, they are sometimes encountered. This will be, e. g., the case in section 4.5.

2.5 Yang-Mills theory

It is now possible to generalize Maxwell theory to the basic building block of essentially all gauge theories, Yang-Mills theory. This is done by upgrading the gauge fields of Maxwell theory to a tensor product of a space-time vector field times a Lie algebra. This will be done here for an arbitrary gauge algebra, though the dimension of the algebra will be referred to as colors. Specializing the gauge algebra will then define a concrete gauge theory. E. g. for QCD it is $\mathfrak{su}(3)$. Effectively, the gauge field will assign for each point in space-time for every direction an element of a Lie algebra.

How this happens is actually 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 the generator will be replaced by the generators of the group, i. e., the gauge fields will be given as vector in the algebra at every space-time point

$$A_\mu(x) = A_\mu^a(x)\tau_a. \quad (2.18)$$

Note that at this point the generators are not in any specific representation. Rather, the gauge fields are currently vectors in the vector space of the algebra, and the A_μ^a are the components. The x -dependence merely states that a different vector of the algebra space is associated with every space-time point. I. e. gauge fields are elements of the vector space spanned by the generators of the Lie algebra.

Hence, there are $\dim G$ gauge fields in a Yang-Mills theory. Most of the following can be done essentially completely abstract, i. e. without choosing a particular representation for the generators. However, it is usually practical to chose a representation, usually the lowest-dimensional one. Then the τ^a are matrices of size the lowest-dimensional fundamental representation. E. g., for $su(N)$ these are $N \times N$ matrices, like the Pauli matrices (2.13) or the Gell-Mann matrices (2.14).

It then remains to construct a gauge-invariant action for Yang-Mills theory, which generalizes Maxwell theory. This is again an axiomatic process, which can be motivated by various geometric arguments, which will be discussed in section 2.6, but in the end remain a postulate. It is experiment which vindicates the following.

Since the gauge field is now Lie-algebra-valued, so will be any gauge transformation function be a vector in algebra space, $\tau^a\omega_a(x)$, from which a group-valued unitary gauge transformation

$$G(x) = \exp(ig\tau^a\omega_a(x)) \quad (2.19)$$

is obtained. Again, this does not yet require to specify a representation. The gauge transformation is now defined to act on the gauge fields as

$$A_\mu \rightarrow GA_\mu G^{-1} + G\partial_\mu G^{-1}. \quad (2.20)$$

In infinitesimal form this reads for the gauge fields A_μ^a

$$A_\mu^a \rightarrow A_\mu^a + D_\mu^{ab}\omega_b \quad (2.21)$$

$$D_\mu^{ab} = \delta^{ab}\partial_\mu - gf_c^{ab}A_\mu^c. \quad (2.22)$$

It is visible that this change is independent of the actual representation of G . It is only necessary that in an actual calculation both τ^a and G are in the same representation. In

fact, by involving both G and G^{-1} , the gauge transformation (2.20) is blind to the center of the gauge group, and will thus be identical for the covering group or any covered group.

Note that (2.20) is inhomogeneous, as is the Abelian version (2.7), in the sense that a field-independent term arises. The D_μ^{ab} (2.22) is called the covariant derivative in the adjoint representation of the gauge group. The adjoint stems from the appearance of the structure constants. Note, however, that this is a relation in terms of coordinates in the algebra space, and is not yet requiring a choice of representation.

There are two remarkable facts about this. On the one hand, there appears an arbitrary constant g in both (2.19) and (2.22). 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 . It arises from the non-Abelian structure of the gauge algebra, yielding non-zero structure constants f^{abc} . This non-linearity gives rise to all kind of technical complications.

This more involved structure 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

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu] = F_{\mu\nu}^a \tau_a = (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc} A_{b\mu} A_{c\nu}) \tau_a. \quad (2.23)$$

The field-strength tensor is again a vector in the gauge algebra, but a symmetric tensor in coordinate space. 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 straightforward 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

$$\text{tr}(F_{\mu\nu} F^{\mu\nu}) = F_{\mu\nu}^a F_a^{\mu\nu}$$

is. To see this, it is useful to first establish explicitly the behavior of $F_{\mu\nu}$. Under an

infinitesimal transformation, it follows that the additional terms transform as

$$\begin{aligned}
& \frac{1}{g} \partial_\mu D_\nu^{ab} \omega^b - \partial_\nu D_\mu^{ab} \omega^b + f^{abc} (A_\mu^b D_\nu^{cd} \omega^d + A_\nu^c D_\mu^{bd} \omega^d) \\
= & \frac{1}{g} (\partial_\mu \partial_\nu - \partial_\nu \partial_\mu) \omega^a + f^{abc} (\partial_\mu A_\nu^c \omega^b - \partial_\nu A_\mu^c \omega^b) + f^{abc} (A_\mu^b \partial_\nu \omega^c + A_\nu^c \partial_\mu \omega^b) \\
& + g f^{abc} (f^{cde} A_\mu^b A_\nu^e \omega^d + f^{bde} A_\nu^c A_\mu^e \omega^d) \\
= & f^{abc} F_{\mu\nu}^b \omega^c,
\end{aligned}$$

where it has been used that

$$\begin{aligned}
& (\partial_\mu \partial_\nu - \partial_\nu \partial_\mu) \omega^a = 0 \\
& (\partial_\mu A_\nu^c - \partial_\nu A_\mu^c) \omega^b + A_\nu^c \partial_\mu \omega^b - A_\mu^c \partial_\nu \omega^b + A_\mu^b \partial_\nu \omega^c + A_\nu^b \partial_\mu \omega^c = (\partial_\mu A_\nu^c - \partial_\nu A_\mu^c) \omega^b \\
& f^{abc} f^{cde} A_\mu^b A_\nu^e \omega^d + f^{abc} f^{bde} A_\mu^c A_\nu^e \omega^d = (f^{adb} f^{ceb} + f^{aeb} f^{bdc}) A_\mu^d A_\nu^e \omega^c \\
& = f^{abc} f^{deb} A_\mu^d A_\nu^e \omega^c.
\end{aligned} \tag{2.24}$$

Here the anti-symmetry of the structure constants and the Jacobi identity (2.12) has been used. Note that it was essential that the coupling constant appearing in the gauge transformation and in the field strength tensor was identical.

Thus it remains to form the product of the two field strength tensors. To leading order, only one is transformed, yielding

$$f^{abc} F_{\mu\nu}^b F^{a\mu\nu} \omega^c.$$

Since $F_{\mu\nu}^a$ is antisymmetric in the Lorentz indices, this vanishes, since the quadratic contraction is then symmetric in a and b . This implies that the field-strength tensor transforms homogeneously under an infinitesimal gauge transformation, in contrast to the fields themselves. As a consequence, it transforms under a finite transformation as

$$F_{\mu\nu} \rightarrow G F_{\mu\nu} G^{-1},$$

i. e. like a similarity transformation in the algebra space.

Thus, in analogy to Maxwell theory, the Lagrangian

$$\mathcal{L} = -\frac{1}{4} \text{tr}(F_{\mu\nu} F^{\mu\nu}) = -\frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu}, \tag{2.25}$$

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 the color-electric field and the color-magnetic field 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 necessarily composites of the gauge fields, like the action. 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. Much more will be said on this topic during this lecture.

There is an interesting consequence of (2.23). It is possible to rescale $A \rightarrow gA$, thereby rescaling $\bar{F} = gF$. As a consequence, the rescaling allows to rewrite (2.25) as

$$\mathcal{L} = -\frac{1}{4g^2} \text{tr}(\bar{F}_{\mu\nu} \bar{F}^{\mu\nu}), \quad (2.26)$$

where thus the coupling constant only acts as a mere prefactor. This remains true in the presence of matter, as will be seen in section 2.7. Thus, the amplitude of the gauge field and the coupling become exchangeable concepts. Moreover, any expansion in the coupling can be equally well be considered as an amplitude expansion.

2.6 Where the gauge freedom comes from

2.6.1 The Abelian case

As noted in section 2.5, it is possible to attach a geometrical interpretation to gauge fields. For this, it is best to understand a little bit more what gauge transformations and gauge fixing really are.

The definition (2.18) is actually the statement that at every point in space a tensor in a vector space is attached. I. e., for every point in space exists a tensor, in this case a rank one tensor and thus a vector, in an additional vector space. This is the vector space formed by the Lie algebra. Note that the gauge fields are the components of the vector, while the generators are the (base) vectors. At that point, no representation needs to be chosen for the generators themselves, and they can remain abstract.

Consider first the Abelian case. Then, the generator is the unit operator, and the Lie algebra space attached is a real vector space. At first sight, it appears that there is one such vector for each direction. This is, however, not true. A Lorentz vector is in the $(1/2, 1/2)$ representation of the Lorentz group. According to the Wigner-Eckart theorem

there can be thus only two non-group-theoretical degrees of freedom, multiplying the two invariant tensors, one for each of the two SU(2) subgroups. That is to be expected from classical physics, where it had already been seen that electromagnetic waves have only two physical degrees of freedom. So, it needs to be possible to decompose the vector potential accordingly.

The Abelian gauge field transforms according to (2.7), $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu g$. If the gauge field is contracted with a transverse projector, this yields for the gauge-transformed field in momentum space

$$\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) A'^\mu = \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) (A^\mu + p^\mu g(p)) = \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) A^\mu.$$

Thus, the transverse projected gauge-field is invariant under a gauge transformation. The longitudinal part behaves like

$$\frac{p_\mu p_\nu}{p^2} A'^\mu = \frac{p_\mu p_\nu}{p^2} A^\mu + \frac{p_\nu}{p^2} g(p)$$

and thus carries all the changes, but does also include (longitudinal) information from the gauge field. This reduces the degrees of freedom, leaving one. Since the gauge field is furthermore in the massless representation of the translation group, any momentum of a physical degree of freedom is necessarily lightlike. Thus, for momentum $P = (p, 0, 0, p)$, the gauge field can be decomposed as $A = (A_0, A_1, A_2, A_0) = A^\perp + A^0$. Because

$$F_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma} p^\rho A^\sigma$$

in momentum space the light-like part A^0 drops out of the field strength tensor, and thus the electric field and the magnetic field are independent of A^0 , and thus all observables.

This can also be seen in a different way. To get rid of the gauge part, define a physical photon field A_μ^P as

$$A_\mu^P = \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) A^\nu,$$

which is manifestly transverse, and thus has only two degrees of freedom, and is invariant under gauge transformations. Seems to be pretty simple, but there is a catch.

This catch appears when trying to invert the Fourier transform, and return to position space. This yields

$$\int d^4 p A_\mu^P(p) e^{ipx} = \int d^4 p \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) A^\nu(p) e^{ipx}.$$

Now, this is no longer a simple expression, because of the factor $1/p^2$. Rather, what is needed is an integral over all of space of $A_\mu(x)$,

$$\begin{aligned} A_\mu^P(x) &= g_{\mu\nu} \int d^4y \delta(x-y) A^\nu(y) - \frac{1}{4\pi} \int d^4y \frac{\partial_\mu \partial_\nu}{|x-y|} A^\nu(y) \\ &= g_{\mu\nu} A^\nu(x) - \frac{1}{4\pi} \int d^4y \frac{\partial_\mu \partial_\nu}{|x-y|} A^\nu(y). \end{aligned} \quad (2.27)$$

Because

$$\begin{aligned} &\left(g_{\mu\nu} \int d^4y \delta(x-y) - \frac{1}{4\pi} \int d^4y \frac{\partial_\mu \partial_\nu}{|x-y|} \right) \partial^\nu g(y) \\ &= \partial_\mu g(x) - \frac{1}{4\pi} \int d^4y \frac{\partial^2 \partial_\mu g(y)}{|x-y|} = \partial_\mu g - \partial_\mu g = 0, \end{aligned} \quad (2.28)$$

the so defined field is indeed gauge-invariant. However, it is not a local object anymore, as the second integral in (2.27) cannot be eliminated. Thus passing to a physical vector field implies even in QED giving up locality. Note that the physical field is entirely transverse, as any other contraction of it vanishes by construction. In particular, the light-like component A_0 drops out as well. Thus, it carries only two degrees of freedom (in four dimensions). In fact, as the non-local term is essentially a solution to the partial differential equations determining the vector potential at fixed electric and magnetic field, it is possible to regain an expression in terms of these fields.

In lower dimensions, the Lorentz group reduces first to $SO(1,2)$, and thus there is only a single non-trivial degree of freedom, due to the Wigner-Eckart theorem. Again, the longitudinal component and timelike component can be eliminated in the same way. Finally, in two dimensions, the Lorentz group is itself Abelian, $SO(1,1)$. The situation is quite odd, as spin can not be continuous, as the representations are labelled continuously. However, longitudinal components and time-like components do not contribute for the same reasons, and thus the gauge field has no physical degrees of freedom in two dimensions. Likewise, in higher dimensions, the degrees of freedom increase to be always dimensionality-2.

This leaves the question open, why only one gauge condition is necessary, rather than two. In fact, the previous example showed this already. Imposing, e. g., Landau gauge $\partial_\mu A^\mu = 0$ is making the gauge field already transverse without projection. But then Lorentz symmetry requires that the light-like part drops out. In a sense, the gauge condition fixes the relative distribution of freedom between both redundant degrees of freedom. The elimination of them is ultimately performed by the underlying Lorentz symmetry. This is not necessary, as is demonstrated by the fact that the vector potential is never necessary to introduce. Thus, it is possible to work without fixing this relation. But if

the vector potential should be used in calculation, this distribution is necessary to have a unique definition of the field. As this fixes the relative orientation of the vectors in algebra space, this can be thought of as fixing the coordinates in algebra space.

In principle, this can be done ultralocally. E. g., axial gauges work by introducing a vector n and demanding $nA = c$, where c is a constant, often set to zero. However, these conditions are not frame-independent, as introducing a fixed vector n makes Poincaré symmetry hidden. As a consequence, degrees of freedom need to be decomposed with respect to n , making calculations often quite demanding, and especially finding consistent approximations become challenging. Gauges like Landau gauge, on the other hand, allow to have manifest Lorentz Poincaré symmetries at all stages, at the expense of no longer being ultralocal, but merely local.

Thus, the number of degrees of freedom is determined due to the Lorentz representation. Each of the invariant tensors carries one vector in the algebra. For the same reason the situation is the same in the Yang-Mills case. However, because of its non-linear transformation (2.21), this becomes less obvious. Especially, (2.27) does not create a gauge-invariant field, and also is the field-strength tensor gauge-invariant. In fact, because of the non-linear part, it will explicitly depend on the other components as well.

But why bother at all? Why not work exclusively with invariant quantities, like (2.27)? After all, this is what classical mechanics teaches: Any remaining degree of freedom is nothing but a redundancy which can be eliminated by passing to suitable generalized coordinates. And classical electrodynamics shows that is indeed possible. The answer to this is one of practicality. Some things have even been calculated in this way, especially the (in)famous Aharonov-Bohm effect. But in general, this becomes practically essentially impossible. The reasons are the arising non-localities, such as those in (2.27). Thus, introducing redundancy by having gauge fields as auxiliary degrees of freedom, paying with the appearance of the internal vector space, is practically the better choice. While non-locality is not a real conceptual problem, this is an almost insurmountable practical one. This applies to general statements even more than to actual calculations. The obvious follow-up question is then why this kind of redundant variables, and not some other. The answer is that this has been the one that works. Others may exist, but none has been found making it easier.

2.6.2 The non-Abelian case and the Gribov-Singer ambiguity

When passing to Yang-Mills theory, the situation becomes only technically more involved. The simple translations of (2.7) become replaced by the non-linear transformation (2.20) in the vector space spanned by the Lie algebra space. While on a conceptual level, this

does not change anything, there arises a serious technical problem. This makes it also harder to see that indeed there are only two degrees of freedom per generator even in this case. The problem arises from the fact that a Lie algebra describes a manifold, and for non-Abelian (compact) Lie algebras this manifold does not allow for a trivial atlas. Thus, transfer functions are needed. However, because implementing a gauge choice is becoming space-time dependent, the transfer functions will also become space-time dependent. In this way, the global structure of the Lie algebra space affect also space-time quantities.

At the root of the problem is the so-called Gribov problem, or more generally the Gribov-Singer ambiguity. It is a genuine non-perturbative feature, and therefore does not arise in perturbative treatments of Yang-Mills theory.

The main problem starts with the realization that local gauge conditions do not have unique solutions beyond perturbation theory in the non-Abelian case. In general, the question of uniqueness needs to be formulated in the following way. Consider first the Abelian case. Given a gauge condition $C[A_\mu, x] = 0$ and arbitrary gauge field $\bar{A}_\mu(x)$, does there exist a unique solution $g(x)$ to

$$C[\bar{A}_\mu + \partial_\mu g, x] = 0?$$

Of course, the answer to this question will depend on the gauge condition. Provided that C takes the form of a differential equation for A_μ , this is turned into a differential equation for $g(x)$. The question can thus be answered by the theory of partial differential equations. E. g., in the case of the Landau gauge, this is an elliptic partial differential equation for which uniqueness theorems exist, proving that there is one and only one solution for fixed boundary conditions.

In the non-Abelian case, this takes the form

$$C[\bar{A}_\mu^a + \partial_\mu g, x]^a = 0?$$

with g group-valued. Thus, these are coupled equations. Because the manifold has a non-trivial atlas², the existence theorems can no longer be applied. If only the surrounding of an element is probed, which is the equivalent of a small amplitude expansion, this is no problem. This is the situation for perturbative expansions in the coupling constant, due to the equivalence of amplitude and coupling expansion. But if the whole manifold needs to be probed, this no longer holds true. Then transfer functions are needed, and thus the

²The 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 are necessary. 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.

global manifold structure plays a role, obstructing a purely local gauge condition to yield a unique answer.

Thus, by only implementing such a gauge, the vector potential is no longer single-valued. This can be most easily seen in gauges where perturbatively there is one and only solution, like in the Landau gauge. The additional solutions are called Gribov copies. 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. It is for this important to realize that any non-perturbative treatment has to start from the gauge transformation (2.20) rather than from the infinitesimal one (2.21). Using the latter implies the assumption that it is always possible to construct gauge transformations from consecutive infinitesimal ones. It is this assumption which actually breaks down in non-Abelian gauge theories.

To be concrete, consider $\text{su}(2)$ Yang-Mills theory. An example is given by the so-called instanton field configuration

$$A_\mu^a = 2\eta_{a\mu\nu} \frac{x^\nu}{g^2(x^2 + \rho^2)}$$

$$\eta_{a\mu\nu} = \begin{cases} \varepsilon_{a\mu\nu} & \text{for } \mu, \nu = 1, 2, 3 \\ \delta_{a\mu} & \text{for } \nu = 4 \\ -\delta_{a\nu} & \text{for } \mu = 4 \end{cases},$$

which will appear in chapter 3. It satisfies the Landau gauge condition, because it is essentially a four-dimensional rotation. Acting with the gauge transformation

$$G(x) = \frac{\tau^\mu r_\mu}{r}$$

$$\tau_\mu = \begin{pmatrix} \vec{\tau} \\ -i \end{pmatrix},$$

on it transforms it into

$$A_\mu = \frac{2\lambda^2}{g^2 r^2 (r^2 + \lambda^2)} \bar{\tau}_{\mu\nu} r^\nu$$

$$\bar{\tau}_{\mu\nu} = \frac{1}{4i} (\bar{\tau}_\mu \tau_\nu - \bar{\tau}_\nu \tau_\mu).$$

Though this field configuration has a different radial behavior, it remains essentially a rotation, and therefore also satisfies the Landau gauge condition. The field configuration depends inversely on the gauge coupling, and can thus not be expanded in it in a perturbative series, showing again that this is a genuinely on-perturbative effect.

This can also be seen from the fact that the gauge transformation (2.29) 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.

Of course, any space-time point is still associated to unique algebra vector along the gauge orbit. It is thus still possible to specify a unique gauge copy. But the gauge condition then usually takes the form of an integral equation. E. g., requiring to the gauge field to absolutely minimize

$$\int d^4x A_\mu^a A^{\mu a} \quad (2.29)$$

can be shown to yield a unique gauge copy on every gauge orbit.

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 some of these gauge conditions in numerical 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 analytical situation is even worse, and little is known about practical solutions to the problem. However, the situation is still of technical nature in the classical theory. The situation will be revisited in the quantized version in chapter (6), where the implications are more far reaching.

2.7 Matter, representations, and coupling universality

2.7.1 Matter fields in Abelian gauge theories

So far, the theory only contained the gauge fields. It remains to introduce matter particles. Again, this requires a postulate, which is then justified empirically. It is again useful to start with the Abelian case first.

Empirically, matter fields are required to transform under a gauge transformation as

$$\phi^g(x) = e^{ieg(x)}\phi(x), \quad (2.30)$$

i. e. by a pure phase factor. In this a new, arbitrary, constant, e appears. It will be seen that it plays the same role as the self-coupling in the Abelian case, but in the Abelian case only pertains to the interactions of matter with the gauge field. This also requires the matter fields to be complex. A term not involving derivatives, e. g. a mass term $\phi^\dagger\phi$, is trivially invariant under a gauge transformation.

This is different for expressions involving derivatives, like kinetic terms,

$$\partial_\mu\phi^g = e^{ieg} (\partial_\mu + ie\partial_\mu g) \phi.$$

Typical kinetic terms are found to be $(\partial_\mu\phi)^\dagger\partial^\mu\phi$ and $\bar{\psi}_i\gamma_{ij}^\mu\partial_\mu\psi_j$, where the Γ_μ are fixed matrices, the so-called Dirac matrices. These are terms for bosonic particles and fermionic particles, respectively, which will be discussed in more detail in the lecture quantum field theory I. At the moment, details beyond the derivative nature are not relevant.

What is relevant is that neither term can be alone gauge-invariant. This can be changed by introducing a covariant derivative

$$D_\mu^g\phi^g = (\partial_\mu - ieA_\mu^g)\phi^g = e^{ieg} (\partial_\mu + ie\partial g - ieA_\mu + ie\partial_\mu g) \phi = e^{ieg} D_\mu\phi, \quad (2.31)$$

which makes derivatives transform like like a matter field. Thus, by replacing all derivatives by covariant derivatives a gauge-invariant Lagrangian of Maxwell fields and matter is obtained. As this is the simplest, but not the only, possibility to do so, this is called minimal coupling. Indeed, it is again experiment which tells that this is correct.

It should be noted that (2.31) induces an interaction vertex between the matter and the gauge field, with a coupling strength e . At lowest order in the low-energy limit, the so-called Thomson limit, this is just the classical interaction of an electron with the electromagnetic field. This justifies to call this quantity electric charge. It should be noted that for every matter field there can be arbitrarily different electric charges, as long as the respective covariant derivatives and gauge transformations use the same. Thus, electric charge is not fixed by the theory, but is an experimental input.

2.7.2 Matter fields and coupling universality in non-Abelian gauge theories

After having the coupling of matter fields to Abelian gauge fields the next step is to generalize it to coupling to Yang-Mills fields. Again, it is for the moment irrelevant whether the matter is (pseudo)scalars, fermions, or have other spin assignments, though this of course influences details of the Lagrangian structure to form Lorentz scalars.

In addition, the following will use the concept of minimal coupling, which is a direct generalization of minimal coupling in the Abelian case, (2.31). I. e. it will work using a

covariant derivative only slightly generalized from the Abelian version. While there may be other possibilities to introduce gauge theories so far this is the only type consistent with any experimentally known theory. For this reason, this lecture is exclusively devoted to this class of gauge theories. They are also arguably the simplest version, and still more than complicated enough to be not fully understood until today.

For an Abelian group structure, every irreducible representation is one-dimensional. I. e. everything boils down to sets of independent functions, not vectors. Thus, everything ultimately boils down to allow for phase rotations of the matter fields under a gauge transformation, $\phi \rightarrow e^{ig\omega(x)}\phi$, where g is some arbitrary constant. This implies that ϕ needs to be complex. In fact, from the vantage point of global symmetries of section 2.1, this is having a global U(1) symmetry, and making it local. This is indeed also the usual way of creating a gauge theory: Take a matter theory with global symmetry, and making the group, or some subgroup of it, local³. Now, any generalization of phase transitions will still leave combinations of the type $\phi^\dagger\phi$ invariant under a gauge transformation, but not a kinetic term involving ∂_μ . Thus, ordinary kinetic terms involving $\partial_\mu\phi$ are no longer gauge-invariant. But this need to be.

The origin of this problem is that ∂_μ is not only involving the field ϕ at some position x . Because of the limiting procedure defining a derivative, it also involves information from some nearby region. But because of the gauge freedom, the coordinate systems in algebra space or group can even at two nearby points be different. And since the field components are the vector components in algebra space they can only be combined if calculated in the same coordinate system.

It is thus necessary to somehow compensate for the possible change of coordinate system. In the Abelian case this can be achieved by introducing the covariant derivative D_μ , which has the property that under a gauge transformation it behaves as $D_\mu\phi \rightarrow e^{ig\omega(x)}D_\mu\phi$, i. e. that applied to the field the whole object transforms as a field. Then, the combinations appearing in kinetic terms become again gauge-invariant. Because of this the covariant derivative is also called parallel transporter, because it makes the changes of the field comparable to changes of the field locally.

This requires that the covariant derivative itself transforms as

$$D_\mu \rightarrow e^{ig\omega(x)}D_\mu e^{-ig\omega(x)}.$$

If it would be just the ordinary derivative, this would yield

$$e^{ig\omega(x)}\partial_\mu e^{-ig\omega(x)} = \partial_\mu - ig\partial_\mu\omega(x).$$

³If a subgroup is made local, and the original group was not simple, some global group remains. However, this will not be relevant throughout most of this lecture, but plays a very important role in actual theories.

Thus, a mismatch $\partial_\mu\omega(x)$ arises. But this is precisely of the same form as (2.7). Thus, a covariant derivative defined as

$$D_\mu = \partial_\mu + igA_\mu \quad (2.32)$$

has exactly the desired properties. Because this implies that the gauge field connects the nearby points to compensate for the change of coordinate system, the gauge field is also called connection in a more mathematical context. And this is exactly what appears also in (2.31) in the Abelian case. It should be noted that the constant g has been arbitrary, but parametrizes the strength of the coupling between the gauge field and the matter field. However, as long as the constant is the same for the gauge transformation and the covariant derivative, it is arbitrary, and needs to be measured to be fixed. At the same time, it can be different for different matter fields.

This is the procedure of minimal coupling: Chose a matter field with global symmetry group, gauge the symmetry group or a subgroup, and fill in covariant derivatives with couplings taken from experiment. That's it, this is how the standard program for constructing a gauge theory from a non-gauge theory works. It should be noted, however, that even if the gauge fields are redundant degrees of freedom, the ungauged and gauged theories differ qualitatively. If passing to appropriate generalized coordinates it is not the ungauged theory, which arises, but a more complicated one. Moreover, the matter fields are now gauge-dependent. Thus, just as the gauge fields themselves, they depend on a choice of coordinate system, and are hence no longer physical degrees of freedom, as they were in the ungauged theory. How to remove the gauge field again will be discussed later, and what then are physical degrees of freedom.

The recipe to put matter into non-Abelian Yang-Mills theories is quite similar, but for two important details. The first is that a non-Abelian gauge theory has many different (irreducible) representations, and especially many higher-dimensional ones. It is thus necessary to chose the representation for the theory with the global symmetry. Any choice is possible, and experiment will dictate which one is necessary.

Choosing some linear representation R , this representation is n -dimensional. The generators in this representation are then N $n \times n$ matrices τ_a^R , and group elements can be written as $\exp(ig\tau_a^R\omega_a(x)/2)$. The factor 2 is here a convenient normalization. Any matter field is then an n -dimensional vector⁴ ϕ_i . A gauge transformation then acts as

$$\phi_i \rightarrow \left(\exp \left(ig \frac{\tau_a^R}{2} \omega_a(x) \right) \right)_{ij} \phi_j. \quad (2.33)$$

⁴Depending on algebra and representation this vector can be real or complex, which will not matter here.

Note that this is still a translation in the vector space spanned by the algebra, but in the vector space of the representation, i. e. in which the matter fields are defined, this is a rotation⁵. There can be multiple matter fields, each of which can be in a different representation. However, as will be seen below, the constant g in (2.33) is needed to be the same.

Just as before, combinations like $\phi_a^\dagger \phi_a$ are invariant under such transformations, but $\phi_a^\dagger \partial_\mu \phi_a$ are not, and require the introduction of a covariant derivative. This covariant derivative is defined as

$$D_\mu^{ij} = \delta^{ij} \partial_\mu - i \frac{g}{2} A_\mu^a (\tau_a^R)^{ij}.$$

The important difference to the covariant derivative in the QED case (2.32) is that here the same coupling constant g appears as in the Yang-Mills field strength tensor (2.23). This is the second important difference to the Abelian case, and is called coupling universality. The reason is that in the gauge transformation of the Yang-Mills field (2.21) the covariant derivative in the adjoint representation appears, which also involves the coupling constant. This term is absent in the Abelian case, as the structure constants vanish for an Abelian gauge theory.

To see that this is necessary, consider the fundamental case. Assume that there would be a difference, and the matter would couple with the coupling constant q instead of g . For good measure, also allow that the gauge transformation for the matter could be modified by some factor. Then

$$\begin{aligned} & \left(\delta^{ij} \partial_\mu - i \frac{q}{2} A_\mu^a \tau_{ij}^a \right) \left(\delta_{jk} + i \frac{p}{2} \tau_{jk}^b \alpha^b \right) \\ &= \delta^{ik} \partial_\mu - \frac{iq}{2} A_\mu^a \tau_{jk}^a + \frac{ip}{2} \tau_{ik}^b (\partial_\mu \alpha^b + \alpha^b \partial_\mu) + \frac{pq}{4} A_\mu^a \tau_{ij}^a \tau_{jk}^b \alpha^b - \frac{iq}{g^2} D_\mu^{ab} \tau_{jk}^a \alpha^b \\ &\stackrel{p=q}{=} \delta^{ik} \partial_\mu - \frac{iq}{2} A_\mu^a \tau_{jk}^a + \frac{q^2}{4g} A_\mu^a \tau_{ij}^a \tau_{jk}^b \alpha^b - \frac{iq}{2} f^{abc} A_\mu^c \tau_{jk}^a \alpha^b + \frac{iq}{2g} \tau_{ik}^b \alpha^b \partial_\mu \\ &\stackrel{q=g}{=} \delta^{ik} \partial_\mu - \frac{ig}{2} A_\mu^a \tau_{jk}^a + \frac{g}{2} A_\mu^a \tau_{ij}^b \tau_{jk}^a \alpha^b + i \tau_{ik}^b \alpha^b \partial_\mu \\ &= \left(\delta_{ij} + \frac{i}{2} \tau_{ij}^b \alpha^b \right) \left(\delta_{jk} \partial_\mu - \frac{ig}{2} A_\mu^a \tau_{jk}^a \right). \end{aligned}$$

Thus, the gauge transformation needs to be of the form (2.33) and the coupling constants need to be the same.

This has very far reaching consequences for any particle physics theory. It determines that the interaction strength of the gauge bosons and between the matter and the gauge bosons is uniquely related. This is a unambiguous prediction, which is experimentally testable. The other is that not every classical non-Abelian gauge theory can be consistently

⁵Since all Lie algebras correspond to rotations in spaces of division algebras.

quantized, as will be discussed in chapter 7. The latter has, e. g., implications to the equality of the electric charges of electrons and protons, as is discussed in the lectures on electroweak physics or physics beyond the standard model.

In much of the following, the gauge algebra/group will be $\mathfrak{su}(N)/\mathrm{SU}(N)$ with matter in the fundamental representation. This is the situation encountered in the standard model. It is also technically often the simplest case. A generalization to arbitrary representations and gauge algebras/group is usually straightforward, often just a replacement of numerical factors and constant matrices.

There is one important distinction to note. If a representation is not complex, as is the case for all representations of $\mathrm{SU}(N > 2)$, but (pseudo)real, matter particles and antiparticles are not independent. Rather, they are the same, possibly up to a phase. This will play an important role in section 6.3.2, where physical states, especially bound states, will be constructed.

It has also a second implication, if there are multiple fields in the same representation. If there are multiple fields in the same representation and with otherwise the same quantum numbers they can be transformed into each other by an additional symmetry transformation⁶. This is called a flavor symmetry⁷. The group associated with this symmetry will depend on whether the gauge group representation of the matter fields is complex, real, or pseudoreal. If the matter is fermionic, this flavor symmetry can also be combined with the corresponding chiral symmetry in different ways, depending on the type of representation. However, as this is otherwise not part of the gauge symmetry, and the possibilities quickly proliferate, this issue is mostly treated in lectures for particular implementations of gauge theories.

2.7.3 Explicit theories

With the information so far, it is possible to construct some general theories. The primary example is electrodynamics. For N_f electrically charged particles it takes the form

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \sum_f^{N_f} \bar{\psi}_f (i\gamma^\mu D_\mu^f - m_f) \psi_f.$$

Herein both the masses of the matter particles m_f and their coupling constants e_f can be chosen for each matter independently. In contrast, for N_f particles in the same represen-

⁶Provided no other terms in the Lagrangian, like mass-terms or non-gauge interaction terms, break this symmetry explicitly.

⁷Or, especially when it comes to the Higgs, custodial symmetry.

tation, in the non-Abelian case this takes the form

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} - \sum_f^{N_f} \bar{\psi}_f^i (i\gamma^\mu D_\mu^{ij} - m_f \delta_{ij}) \psi_f^j,$$

where the covariant derivative is now universal, and only the masses may differ. If the representation is fundamental, this is quantum chromodynamics. If there are multiple representations N_r are involved, this generalizes to

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} - \sum_r^{N_r} \sum_f^{N_f} \bar{\psi}_f^{ri} (i\gamma^\mu D_\mu^{rij} - m_f \delta_{ij}) \psi_f^{rj}.$$

It is noteworthy that the term for the gauge fields remains unchanged, as it exists purely in the algebra. It should be noted that in either case the representations must be those of the true groups, to keep the matter fields single-valued.

In all of the cases if the masses, and in the Abelian cases the coupling constants, degenerate, the theory has an additional global symmetry for rotating the $N_f^{(r)}$ matter fields into each other.

Bosonic fields can also be added, yielding

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} - \sum_r^{N_r} \sum_f^{N_f} \bar{\psi}_f^{ri} (i\gamma^\mu D_\mu^{rij} - m_f \delta_{ij}) \psi_f^{rj} - \sum_r^{M_r} \sum_f^{M_f} \left((D_\mu^{rij} \phi_j^r)^\dagger D_{rik}^\mu \phi_k + m_f^b \phi_{ri}^\dagger \phi_{ri} \right).$$

It is furthermore possible to add interactions terms between the matter fields, provided they are gauge-invariant. E. g. a term $(\phi_{ri}^\dagger \phi_{ri})^4$ is a typical self-interaction term for bosons. The general case is dependent on the fields and representations, and quickly becomes an exercise in group theory applications. This will not be pursued further here, but will be a substantial part in the lectures applied to specific theories, e. g. the standard model.

2.8 Beyond Yang-Mills theory

While the bulk of this lecture focuses on Yang-Mills theory, it is useful to generalize the concept of gauge theories. These are often technically more involved, and more will be said in section 4.6. Here, primarily, other possible gauge theory structures will be exposed. In the following, matter will not be considered.

2.8.1 Yang-Mills type

To set the stage, consider again the type encountered so far, Yang-Mills type. The gauge transformation takes the form

$$A_\mu^a(x)' = A_\mu^a(x) + \delta_\epsilon A_\mu^a \quad (2.34)$$

$$\delta_\epsilon A_\mu^a(x) = D_\mu \epsilon^a(x) = \partial_\mu \epsilon^a + f_{bc}^a A_\mu^b \epsilon^c(x), \quad (2.35)$$

where $\epsilon^c(x)$ is a set of arbitrary functions, the gauge transformations. These are algebra-valued vectors. In these theories, the commutator of infinitesimal gauge transformations reads

$$[\delta_\epsilon, \delta_\eta] X = \delta_\xi X, \quad \xi^a = f_{bc}^a \epsilon^b \eta^c \quad (2.36)$$

with ϵ , η , ξ are arbitrary gauge transformations, and where X can be any field. It follows that the algebra of the gauge transformations closes, i. e. is valid, off-shell as the commutator of the gauge transformations is again a gauge transformation of the same type. Alternatively, this could be stated in the form that the algebra holds for any values of the fields.

2.8.2 Closure only on-shell

Off-shell closure holds for Yang-Mills gauge theories but is not a general feature of gauge systems. The gauge transformations might close only when the equations of motion hold. Notable examples where this is the case are extended supergravity theories. Rather than discussing the gauge structure of supergravities, which is rather intricate, “closure only on-shell” is illustrated in the case of a much simpler (but of no direct physical interest) system. Start with the Lagrangian

$$\mathcal{L} = \frac{1}{2}(\dot{q}^1 - \dot{q}^2 - \dot{q}^3)^2 = \frac{1}{2}\dot{y}^2, \quad y = q^1 - q^2 - q^3 \quad (2.37)$$

for a model with three coordinates q^1 , q^2 and q^3 . \mathcal{L} is invariant under two different sets of gauge transformations, which can be taken to be

$$\delta_\epsilon q^1 = \epsilon + \epsilon q^2 \dot{y}, \quad \delta_\epsilon q^2 = \epsilon, \quad \delta_\epsilon q^3 = \epsilon q^2 \dot{y} \quad (2.38)$$

and

$$\delta_\eta q^1 = 0, \quad \delta_\eta q^2 = \eta, \quad \delta_\eta q^3 = -\eta \quad (2.39)$$

The commutators of the gauge transformations on the fields are given by

$$[\delta_\epsilon, \delta_\eta] q^1 = \epsilon \eta \dot{y}, \quad [\delta_\epsilon, \delta_\eta] q^2 = 0, \quad [\delta_\epsilon, \delta_\eta] q^3 = \epsilon \eta \dot{y}. \quad (2.40)$$

From (2.37), the equation of motion for y is $\ddot{y} = 0$. Thus, the algebra of the gauge transformations (2.38) and (2.39) is closed (and, in fact, Abelian) only up to equations of motion, i.e., only on-shell. Otherwise, it cannot be expressed only in terms of the fields q^i , but requires their derivatives, and thus canonical conjugated momenta, as well.

2.8.3 Reducible gauge theories

The gauge transformations might also be “reducible”, i.e., dependent. Consider a theory of an Abelian antisymmetric tensor field $B_{\mu\nu} = -B_{\nu\mu}$. Following the construction of Yang-Mills theory, the field strength tensor is now rank three and is found to be

$$H_{\mu\nu\rho} = \partial_\mu B_{\nu\rho} + \partial_\nu B_{\rho\mu} + \partial_\rho B_{\mu\nu}.$$

A Lagrangian

$$\mathcal{L} = -\frac{1}{12} H_{\mu\nu\rho} H^{\mu\nu\rho} \quad (2.41)$$

can be constructed which is invariant under gauge transformations

$$\delta_\Lambda B_{\mu\nu} = \partial_\mu \Lambda_\nu - \partial_\nu \Lambda_\mu$$

where Λ is the gauge parameter. These transformations vanish for a class of parameters

$$\Lambda_\mu = \partial_\mu \epsilon \quad (2.42)$$

meaning that the gauge transformations Λ_μ are not all independent. Such gauge transformations are called reducible, and the corresponding gauge theory is said to be reducible. Such antisymmetric tensor fields define a natural generalization of electromagnetism, $A_\mu \rightarrow B_{\mu\nu}$ and occur in many models of unification. The main difference with electromagnetism being the irreducibility of the latter.

Reducibility could be stacked. If, e. g., ϵ in (2.42) in turn obeys some equation, this would be a second level of reducibility, and so on.

2.8.4 Reducibility on-shell

The last version is the possibility that the reducibility of the gauge transformations holds only on-shell. This therefore combines the previous two cases. Adding to (2.41) an auxiliary field A_μ leads to

$$\mathcal{L} = \frac{1}{12} A_\mu \varepsilon^{\mu\nu\rho\sigma} H_{\nu\rho\sigma} - \frac{1}{8} A_\mu A^\mu.$$

This Lagrangian reduces to (2.41) by inserting the equation of motion for A_μ ,

$$A_\mu = \frac{1}{3} \varepsilon^{\mu\nu\rho\sigma} H_{\nu\rho\sigma}.$$

The Lagrangian is invariant under gauge transformations of the form

$$\delta_\Lambda B_{\mu\nu} = \partial_\mu \Lambda_\nu - \partial_\nu \Lambda_\mu \quad (2.43)$$

$$\delta_\Lambda A_\mu = 0. \quad (2.44)$$

Thus, nothing yet happened for the Abelian case. Upgrading this to a non-Abelian case by considering, like in the Yang-Mills case, Lie-algebra-valued fields $A_\mu = A_\mu^a T_a$ and $B_{\mu\nu} = B_{\mu\nu}^a \tau_a$ requires again to introduce covariant derivatives instead of partial derivatives, $\partial_\mu \rightarrow D_\mu$. This is the so-called Freedman-Townsend model.

It can now be shown that in this model the gauge transformations vanish for parameters $\Lambda_\mu = D_\mu \epsilon$. But now, this vanishing occurs only if the equations of motion are satisfied. This is due to $[D_\mu, D_\nu] \propto F_{\mu\nu}$, and $F_{\mu\nu} = 0$ is the equation of motion for B . The theory is said in that case to be reducible on-shell.

2.8.5 Towards gravity

Finally, it should be remarked that if one would allow for position-dependent translations in \mathbb{M}^4 , very much like the position dependent transformations in algebra space, this would yield general relativity. Thus, the concept coming with gauge theories underlie the two central theories of modern physics, the standard model of particle physics and general relativity. All of the things said here then also apply to space-time and space-time coordinates, and it is necessary to pass to coordinate-invariant quantities. This is already the case in classical general relativity, as much as in classical electrodynamics. Also there one needs to work with gauge-invariant quantities, the generalization of electromagnetic fields, and no longer with coordinates, as they are no longer physical. How to do so at the quantum level has not yet been fully understand, and this is the quest for quantum gravity. This is the topic of a different lecture.

Chapter 3

Classical solutions

While gauge theories are needed to be quantized, classical solutions of them can still be of great relevance. This is usually true for two reasons. One is that quantum effects can be small, and be treated as perturbations around the classical solutions. Typical examples of this approach is already the WKB approximation in quantum mechanics. Another feature is that properties of classical solutions may still be useful even in the quantized theory. E. g. the wavelength of classical electromagnetic waves can be transferred to characterize the properties of radiation in the quantum theory.

3.1 Time evolution and initial conditions

Classical gauge theories are partial differential equations, usually (non-linear) wave equations. Due to their definition on Minkowski space-time, they are accessible to existence and uniqueness proofs. However, there is a twist for gauge theories, due to the gauge freedom.

Consider first the Abelian case. The Maxwell equations (2.8-2.9). The equation (2.8) constitutes four equations of motion, while (2.9) provides four more constraint equations. In terms of the gauge-invariant electric field and magnetic field they take the form of the usual Maxwell equations. In the end, only two, out of the six, degrees of freedom remain independent. Four are eliminated by constraint equations, and two of the dynamical equations turn out to be identically fulfilled if the other two are fulfilled.

The remaining second-order differential equations are relativistic wave-equations. To find a unique solutions requires initial condition. These can be chosen to give the independent field components and their derivatives on a space-like hypersurface. The Maxwell equation allows to evolve the field values then along time-like trajectories. The initial conditions fully specific the fields at all space-time points. While this may be technically

cumbersome in practice, this is the full solution. E. g. a spatial plane wave on a space-like hypersurface will give rise to free electromagnetic plane waves.

The situation changes drastically once the problem is expressed in terms of the vector potential. The constraint equation (2.9) is then trivially satisfied. The dynamical equation (2.8) takes the form

$$\partial^2 A_\nu - \partial_\nu \partial^\mu A_\mu = 0, \quad (3.1)$$

and thus take the form of a coupled set of equations. However, even when specifying the vector potential on a hypersurface at $t = 0$, any vector potential of the form

$$A_\mu(t, \vec{x}) + \partial_\mu g(t, \vec{x}) \text{ with } \partial_\mu g(0, \vec{x}) = 0 \quad (3.2)$$

is an equally good solution. Thus, the solution becomes many-valued. The reason is that the dynamical equations with the trivially resolved constraint equation can no longer uniquely determine the time evolution. Of course, at the physical level this does not matter, as (3.2) is just a gauge transformation. Put it differently, the time evolution of the vector potential is only unique up to gauge transformation. This can (only) be resolved by requiring that the gauge field satisfies a suitable gauge condition at all times. Thus, a gauge condition is an additional constraint equation for the time evolution of the vector potential, which needs to be satisfied by the initial conditions as well. Thus, a human choice, the gauge condition, is required to make the time evolution of the vector potential unique.

This, probably better than anything else, shows that the vector potential is an auxiliary degree of freedom introduced to simplify calculations. However, without that, many calculations would be impossible. If, e. g., the Landau gauge condition is specified, the last term in (3.1) drops out, and the equation is a free wave equation. Its time evolution is then again unique, and determined by the initial condition. In particular, it is deterministic. The whole problem can also be seen in a different way. If the time evolution is done in terms of the electric and magnetic field, it is, without gauge condition, possible to introduce an infinite number of different gauge fields. Only once a gauge condition is added, this becomes unique.

This provides, however, another important lesson: It is impossible to preserve gauge freedom and having a unique time evolution of the vector potential simultaneously. A unique, deterministic time evolution of the vector potential has a fixed gauge, and thus the gauge symmetry is no longer manifest.

3.2 Topological excitations

All of these statements about the Abelian case apply equally well to the non-Abelian case, except that even the non-Abelian electromagnetic fields are no longer gauge-invariant, and also do not evolve uniquely in time without fixing a gauge. Only quantities like the action are invariant, and show a unique evolution in time. It requires twice the number of generators of such quantities, for which the equations of motions need to be reconstructed. For them to be gauge-invariant requires (at least) products of fields, for which then time evolution equations need to be deduced from the Maxwell equations.

There is one issue, however, which makes this more problematic. Due to the Gribov-Singer ambiguity, gauge conditions are not local in space-time. A gauge condition like (2.29) is not local in time. Fixing such a gauge turns the time evolution equation into an integral equation, and makes the causal structure no longer appearing. Even when fixing the gauge only on a spatial hyper surface, this will generally require non-local conditions, connecting casually disconnected quantities. Of course, the vector potential is unphysical, and may have acausal correlations, as long as observables evolve entirely causal. This makes such an approach very involved indeed.

As will be seen in chapter 4, in the quantum theory entanglement will anyhow require to deal with such correlations, as in all quantum theories.

After these preliminaries, the question is what are classical solutions. For this, it will be useful to take a somewhat indifferent stand: Finding solutions for given initial conditions, but allow time evolution to change the gauge, without imposing it. As it turns out, for suitable chosen initial conditions there exists a time evolution such that (2.29) is maintained.

The results will indeed be relevant even in the quantum theory, especially in chapter 7. While matter particles, gauge bosons and bound states are states which consists out of a few valence particles, i. e. particles which define their quantum numbers, non-Abelian gauge theories can also support other states. Especially, it can support states which are genuine collective excitations of gauge bosons, with or without matter, 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. Such topological solutions are indeed the building blocks of classical solutions of Yang-Mills theory, very much like the plane waves are in Abelian gauge theories.

The most prominent of these excitations are so-called instantons. Though it is not possible to directly observe instantons, their effects can be identified. To understand them, it is best to first investigate similar excitations in quantum mechanics.

3.2.1 Instantons in quantum mechanics

The starting point is a one-dimensional double-well potential

$$V(x) = \lambda (x^2 - \eta^2)^2$$

This potential has two degenerate ground states centered around $\pm\eta$. These are two classically separated vacua for particles with kinetic energies less than $\lambda\eta^4$. The ground state can be obtained using the WKB approximation. Then the wave function is assumed as

$$\psi(x) = e^{i\Phi(x)}$$

where

$$\begin{aligned}\Phi(x) &= \pm \int^x dx' p(x') + O(\hbar) \\ p(x) &= \sqrt{2m} \sqrt{E - V(x)}\end{aligned}\tag{3.3}$$

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 this 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 antisymmetric 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

$$m \frac{d^2 x}{d\tau^2} = \frac{dV}{dx}$$

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

$$d\tau = \frac{1}{\sqrt{\frac{2\lambda}{m}(x^2 - \eta^2)^2}} dx = \frac{1}{\eta^2 \sqrt{\frac{2\lambda}{m}}} \frac{1}{1 - \frac{x^2}{\eta^2}} dx = \frac{1}{\eta \sqrt{\frac{2\lambda}{m}}} \frac{1}{1 - u^2} du.$$

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

$$x_{cl}(\tau) = \eta \tanh \left(\frac{\omega}{2} (\tau - \tau_0) \right) \quad (3.4)$$

$$\omega^2 = \frac{8\eta^2 \lambda}{m} \quad (3.5)$$

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 (3.4) 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 $\frac{\omega}{2} (\tau - \tau_0) \rightarrow \tau$ this yields

$$\begin{aligned} S_0 &= \int_{-\infty}^{\infty} d\tau \left(\frac{m\eta^2\omega}{4} \frac{1}{\cosh^4 \tau} + \frac{2\lambda\eta^4}{\omega} (\tanh^2 \tau - 1)^2 \right) \\ &= \frac{m\eta^2\omega}{3} + \frac{8\lambda\eta^4}{3\omega} = \frac{m^2\omega^3}{12\lambda} \end{aligned}$$

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

$$\langle 0|e^{-H\tau}|2\eta\rangle = e^{-S_0} \int D[\delta r] e^{-\frac{1}{2}\delta x \frac{\delta^2 S}{\delta x^2} \Big|_{x_{cl}} \delta x + O(\delta x^3)} \quad (3.6)$$

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.

As already mentioned, anti-instantons are instanton paths traversed in reversed time direction. A pair of instanton and anti-instanton relates over a topologically 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

$$S_{IA} = \frac{1}{2\lambda} \left(\frac{1}{3} - 2e^{|\tau_I - \tau_A|} + O(e^{2|\tau_I - \tau_A|}) \right)$$

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 (semi-)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.

3.2.2 Yang-Mills theory

3.2.2.1 Topology

It is convenient for the treatment of Yang-Mills theories to use the following definitions, normalizations

$$\begin{aligned}
 A_i &= A_i^a \frac{\lambda^a}{2} \\
 [\lambda^a, \lambda^b] &= 2i f^{abc} \lambda^c \\
 \text{tr}(\lambda^a \lambda^b) &= 2\delta^{ab} \\
 G_{\mu\nu}^a &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + f^{abc} A_\mu^b A_\nu^c
 \end{aligned} \tag{3.7}$$

and gauge condition

$$A_0 = 0 \implies E_i = \partial_0 A_i$$

where E_i are the chromoelectric fields, which are for this gauge, the so-called temporal gauge¹, the conjugate momenta of the A_i . The Hamiltonian is then given by

$$H = \frac{1}{2g^2} \int d^3\vec{x} (E_i^2 + B_i^2)$$

Thus, minimal energy has zero chromoelectromagnetic fields. In the Abelian this lowest-energy state would also automatically have constant vector potential. 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 take the form

$$A_i = iU(\vec{x}) \partial_i U(\vec{x})^\dagger$$

where $U(\vec{x})$ can be any possible gauge transformation. Consider those satisfying $U(\vec{x}) = 1$ for $x \rightarrow \infty$. These can be classified by the winding number

$$n_w = \frac{1}{24\pi^2} \int d^3\vec{x} \varepsilon^{ijk} \text{tr}((U^\dagger \partial_i U)(U^\dagger \partial_j U)(U^\dagger \partial_k U)) \tag{3.8}$$

which counts the number of times the group manifold of the gauge transformation is covered. This will be seen below. If (3.8) is expressed in terms of the gauge fields, it is called the Chern-Simons characteristic

$$n_{CS} = \frac{1}{16\pi^2} \int d^3\vec{x} \varepsilon^{ijk} \left(A_i^a \partial_j A_k^a + \frac{1}{3} f^{abc} A_i^a A_j^b A_k^c \right). \tag{3.9}$$

¹Note that this gauge breaks Lorentz symmetry and has thus a variety of subtleties. Also, the Gribov-Singer ambiguity needs to be addressed. However, neither affects the calculations at the level presented, and thus will be ignored.

This is obtained from (3.8) by partial integration and the fact that $\partial(UU^\dagger)$ vanishes by unitarity.

The quantity (3.9) can also be rewritten more generally as

$$\frac{1}{64\pi^2} \int d^4x \epsilon^{\mu\nu\rho\sigma} G_{\mu\nu}^a G_{\rho\sigma}^a = -\frac{i}{512\pi^4} \int d^4x \text{tr} \epsilon^{\mu\nu\rho\sigma} \partial_\mu \left(i A_\nu^a \partial_\rho A_\sigma^a + \frac{2}{3} f^{abc} A_\nu^a A_\rho^b A_\sigma^c \right),$$

where with $\epsilon_{\mu\nu\rho\sigma} G_a^{\rho\sigma}/2 = *G_{\mu\nu}^a$ the dual field strength tensor appears. This is called the topological charge. That (3.10) is the same as (3.9) can be seen as follows. Start from the topological charge Q ,

$$Q = \frac{1}{32\pi^2} \int d^4x G_{\mu\nu}^a * G_{\mu\nu}^a.$$

Expand explicitly as

$$\begin{aligned} Q &= \frac{1}{32\pi^2} \int d^4x \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} (\partial_\nu A_\mu^a - \partial_\mu A_\nu^a - f^{abc} A_\mu^b A_\nu^c) (\partial_\sigma A_\rho^a - \partial_\rho A_\sigma^a - f^{ade} A_\rho^d A_\sigma^e) \\ &= \frac{1}{32\pi^2} \int d^4x \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} (\partial_\nu A_\mu^a \partial_\sigma A_\rho^a - \partial_\nu A_\mu^a \partial_\rho A_\sigma^a + \partial_\nu A_\mu^a f^{ade} A_\rho^d A_\sigma^e - \partial_\mu A_\nu^a \partial_\sigma A_\rho^a \\ &\quad + \partial_\mu A_\nu^a \partial_\rho A_\sigma^a + \partial_\mu A_\nu^a f^{ade} A_\rho^d A_\sigma^e - f^{abc} A_\mu^b A_\nu^c \partial_\sigma A_\rho^a + f^{abc} A_\mu^b A_\nu^c \partial_\rho A_\sigma^a \\ &\quad + f^{abc} A_\mu^b A_\nu^c f^{ade} A_\rho^d A_\sigma^e) \end{aligned}$$

First have a look at the double derivatives

$$\frac{1}{2} \epsilon_{\mu\nu\rho\sigma} (\partial_\nu A_\mu^a \partial_\sigma A_\rho^a - \partial_\nu A_\mu^a \partial_\rho A_\sigma^a - \partial_\mu A_\nu^a \partial_\sigma A_\rho^a + \partial_\mu A_\nu^a \partial_\rho A_\sigma^a)$$

By partial integration this results in

$$\begin{aligned} &\frac{1}{2} \epsilon_{\mu\nu\rho\sigma} (\partial_\nu (A_\mu^a \partial_\sigma A_\rho^a) - A_\mu^a \partial_\nu \partial_\sigma A_\rho^a - \partial_\nu (A_\mu^a \partial_\rho A_\sigma^a) + A_\mu^a \partial_\nu \partial_\rho A_\sigma^a \\ &\quad - \partial_\mu (A_\nu^a \partial_\sigma A_\rho^a) + A_\nu^a \partial_\mu \partial_\sigma A_\rho^a + \partial_\mu (A_\nu^a \partial_\rho A_\sigma^a) - A_\nu^a \partial_\mu \partial_\rho A_\sigma^a) \end{aligned}$$

The total derivatives can be connected by

$$\epsilon_{\nu\mu\sigma\rho} \partial_\nu (A_\mu^a \partial_\sigma A_\rho^a) + \epsilon_{\nu\mu\rho\sigma} \partial_\nu (A_\mu^a \partial_\rho A_\sigma^a) + \epsilon_{\mu\nu\sigma\rho} \partial_\mu (A_\nu^a \partial_\sigma A_\rho^a) + \epsilon_{\mu\nu\rho\sigma} \partial_\mu (A_\nu^a \partial_\rho A_\sigma^a) = 4\epsilon_{\mu\nu\rho\sigma} \partial_\mu (A_\nu^a \partial_\rho A_\sigma^a)$$

The other terms result in

$$4\epsilon_{\mu\nu\rho\sigma} A_\mu^a \partial_\nu \partial_\rho A_\sigma^a = 0$$

since it is symmetric about the inner two indices. For the terms with one power of f^{abc} the result is

$$\begin{aligned} &-\epsilon_{\mu\nu\rho\sigma} (\partial_\nu A_\mu^a A_\rho^b A_\sigma^c - \partial_\mu A_\nu^a A_\rho^b A_\sigma^c + A_\mu^b A_\nu^c \partial_\sigma A_\rho^a - A_\mu^b A_\nu^c \partial_\rho A_\sigma^a) \\ &= -\epsilon_{\mu\nu\rho\sigma} (\partial_\nu (A_\mu^a A_\rho^b A_\sigma^c) - A_\mu^a A_\sigma^c \partial_\nu A_\rho^b - A_\mu^a A_\rho^b \partial_\nu A_\sigma^c - \partial_\mu (A_\nu^a A_\rho^b A_\sigma^c) + A_\nu^a A_\rho^b \partial_\mu A_\sigma^c \\ &\quad + A_\nu^a A_\sigma^c \partial_\mu A_\rho^b + \partial_\sigma (A_\rho^a A_\mu^b A_\nu^c) - A_\rho^a A_\mu^b \partial_\sigma A_\nu^c - A_\rho^a A_\nu^c \partial_\sigma A_\mu^b - \partial_\rho (A_\sigma^a A_\mu^b A_\nu^c) \\ &\quad + A_\sigma^a A_\mu^b \partial_\rho A_\nu^c + A_\sigma^a A_\nu^c \partial_\rho A_\mu^b) \end{aligned}$$

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

$$\frac{1}{16\pi^2} \int d^4x \partial_\mu \varepsilon_{\mu\nu\rho\sigma} \left(A_\nu^a \partial_\rho A_\sigma^a + \frac{1}{3} f^{abc} A_\nu^a A_\rho^b A_\sigma^c \right)$$

The last term vanished, because the square of the potential cannot 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.

Evidently, (3.10) 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 fields in 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. However, this can be seen intuitively. Because the gauge fields are given by (3.8) with U becoming the identity at infinity, at most the derivatives of the U can play a role. These depend only on trigonometric functions of angles, when sweeping the surface at infinity. To have a non-zero value, it is thus necessary that suitable multiples of periods are sampled, thus yielding a discretized result.

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. At the quantum level tunneling can connect topological different vacuum states. But there are also classical finite energy solutions.

3.2.2.2 Instantons

The classical solutions to the euclidean action will provide access. It is convenient to rewrite the action as

$$\begin{aligned} S &= \frac{1}{4g^2} \int d^4x G_{\mu\nu}^a G_{\mu\nu}^a \\ &= \frac{1}{4g^2} \int d^4x \left(\pm G_{\mu\nu}^a \Gamma_{\mu\nu}^a + \frac{1}{2} (G_{\mu\nu}^a \mp \Gamma_{\mu\nu}^a)^2 \right) \end{aligned}$$

where $\Gamma_{\mu\nu}^a = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}G_{\rho\sigma}^a$ is the dual field strength tensor. In this tensor the roles of the magnetic and electric field are interchanged. The first term is the topological charge (3.10). It is fixed if the gauge field vanishes up to a gauge transformation at infinity. To have a finite-energy configuration. But then the topological charge is fixed, and cannot change. If the field is (anti) self-dual, i.e. $G_{\mu\nu}^a = \pm\Gamma_{\mu\nu}^a$, the action is hence minimal.

In fact, this implies that the action is given by

$$S = \frac{8\pi^2 |Q|}{g^2}. \quad (3.10)$$

It is hence determined entirely in terms of the topological charge. 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 (3.8)

$$\begin{aligned} A_i^a &= 2(\lambda^a)^{-1}iU(\vec{x})\partial_iU(\vec{x})^\dagger \\ &= 2i(\lambda^a)^{-1}\frac{x_\mu}{x}\tau_\mu\tau_\nu^\dagger\frac{\partial}{\partial x_i}\frac{x_\nu}{x} \\ &= 2i(\lambda^a)^{-1}\tau_\mu\tau_\nu^\dagger\left(\frac{xx_\mu\delta_{\nu i} + x_\mu x_\nu x_i}{x^3}\right) \end{aligned}$$

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. This is a pure gauge solution with fixed topological charge.

To obtain a non-pure gauge solution with the same action, start with a parametrized ansatz

$$A_\mu^a = \eta_{a\mu\nu}\frac{x_\nu}{x^2}f(x^2). \quad (3.11)$$

The function f needs to be fixed by the self-duality requirement and also requires $f(x^2) \rightarrow 1$ as $|x| \rightarrow \infty$ to let the result be a pure gauge at infinity.

Start from the self-duality equation

$$G_{\mu\nu}^a = \Gamma_{\mu\nu}^a$$

It is worthwhile to note that this is a first-order partial differential equation, rather than the non-Abelian Maxwell equations, which are second-order differential equations. This requires the derivatives

$$\partial_\mu \left(\frac{f x_\nu}{x^2} \right) = \frac{2x_\nu x_\mu}{x^2} f' + \frac{f}{x^2} \delta_{\mu\nu} - \frac{2f x_\mu x_\nu}{x^4}. \quad (3.12)$$

The appearing contractions are given by

$$\begin{aligned} \varepsilon_{\mu\nu\rho\sigma} \eta_{a\rho\alpha} \partial_\sigma \frac{x_\alpha}{x^2} f &= \eta_{a\nu\sigma} \partial_\sigma \frac{x_\mu}{x^2} f + \eta_{a\mu\nu} \partial_\sigma \frac{x_\sigma}{x^2} f + \eta_{a\sigma\mu} \partial_\sigma \frac{x_\nu}{x^2} f \\ \varepsilon_{\mu\nu\rho\sigma} \eta_{a\sigma\alpha} \partial_\rho \frac{x_\alpha}{x^2} f &= -\eta_{a\nu\rho} \partial_\rho \frac{x_\mu}{x^2} f - \eta_{a\mu\nu} \partial_\rho \frac{x_\rho}{x^2} f - \eta_{a\rho\mu} \partial_\rho \frac{x_\nu}{x^2} f \end{aligned}$$

And the result is then

$$\begin{aligned} &\eta_{a\mu\alpha} \partial_\nu \frac{x_\alpha}{x^2} f - \eta_{a\nu\alpha} \partial_\mu \frac{x_\alpha}{x^2} f - \eta_{a\nu\sigma} \partial_\sigma \frac{x_\mu}{x^2} f - \eta_{a\mu\nu} \partial_\sigma \frac{x_\sigma}{x^2} f - \eta_{a\sigma\mu} \partial_\sigma \frac{x_\nu}{x^2} f \\ &= -2f^{abc} \frac{x_\alpha x_\beta}{x^4} f^2 \left(\frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} \eta_{b\rho\alpha} \eta_{c\sigma\beta} - \eta_{b\mu\alpha} \eta_{c\nu\beta} \right). \end{aligned} \quad (3.13)$$

All derivatives are symmetric in the indices by virtue of (3.12) and hence they can be changed to

$$\begin{aligned} &\eta_{a\mu\alpha} \partial_\nu \frac{x_\alpha}{x^2} f - \eta_{a\nu\alpha} \partial_\mu \frac{x_\alpha}{x^2} f - \eta_{a\nu\alpha} \partial_\mu \frac{x_\alpha}{x^2} f - \eta_{a\mu\nu} \partial_\sigma \frac{x_\sigma}{x^2} f + \eta_{a\mu\alpha} \partial_\nu \frac{x_\alpha}{x^2} f \\ &= -2f^2 \frac{x_\alpha x_\beta}{x^4} f^{abc} \left(\frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} \eta_{b\rho\alpha} \eta_{c\sigma\beta} - \eta_{b\mu\alpha} \eta_{c\nu\beta} \right). \end{aligned}$$

The right hand side can be changed using $f^{abc} = \epsilon_{abc}$ for SU(2) to

$$\begin{aligned} &\frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} (\delta_{\rho\sigma} \eta_{a\alpha\beta} - \delta_{\rho\beta} \eta_{a\alpha\sigma} + \delta_{\alpha\beta} \eta_{a\rho\sigma} - \delta_{\alpha\sigma} \eta_{a\rho\beta}) - \delta_{\mu\nu} \eta_{a\alpha\beta} + \delta_{\mu\beta} \eta_{a\alpha\nu} - \delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta} \\ &= \frac{1}{2} (-\delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta} - \delta_{\alpha\mu} \eta_{a\nu\beta} + \delta_{\beta\nu} \eta_{a\mu\alpha} - \delta_{\beta\mu} \eta_{a\nu\alpha}) - \delta_{\mu\nu} \eta_{a\alpha\beta} - \delta_{\mu\beta} \eta_{a\nu\alpha} - \delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta}. \end{aligned}$$

Multiplying with $x_\alpha x_\beta$ results in

$$\begin{aligned} &\frac{1}{2} (-\delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta} - \delta_{\alpha\mu} \eta_{a\nu\beta} + \delta_{\beta\nu} \eta_{a\mu\alpha} - \delta_{\beta\mu} \eta_{a\nu\alpha}) - \delta_{\mu\nu} \eta_{a\alpha\beta} - \delta_{\mu\beta} \eta_{a\nu\alpha} - \delta_{\alpha\beta} \eta_{a\mu\nu} + \delta_{\alpha\nu} \eta_{a\mu\beta} \\ &= -\frac{3}{2} \eta_{a\mu\nu} x^2 + 2\eta_{a\mu\alpha} x_\nu x_\alpha - 2\eta_{a\nu\alpha} x_\mu x_\alpha. \end{aligned}$$

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 (3.12)

$$\begin{aligned} & 2\eta_{a\mu\alpha}\partial_\nu\frac{x_\alpha}{x^2}f - 2\eta_{a\nu\alpha}\partial_\mu\frac{x_\alpha}{x^2}f - \eta_{a\mu\nu}\partial_\sigma\frac{x_\sigma}{x^2}f \\ &= \frac{4f'}{x^2}\left(\eta_{a\mu\alpha}x_\nu x_\alpha - \eta_{a\nu\alpha}x_\mu x_\alpha - \frac{1}{2}\eta_{a\mu\nu}x^2\right) - \frac{4f}{x^4}(\eta_{a\mu\alpha}x_\nu x_\alpha - \eta_{a\nu\alpha}x_\mu x_\alpha). \end{aligned}$$

Equating both to order $1/x^2$ yields a condition equation

$$\begin{aligned} 0 &= -2\eta_{a\mu\nu}f' + \frac{1}{x^2}(4f'\eta_{a\mu\alpha}x_\nu x_\alpha - 4f'\eta_{a\nu\alpha}x_\mu x_\alpha - 3f^2\eta_{a\mu\nu}) \\ &\quad + \frac{4f}{x^4}(\eta_{a\nu\alpha}x_\mu x_\alpha - \eta_{a\mu\alpha}x_\nu x_\alpha + f\eta_{a\mu\alpha}x_\nu x_\alpha - f\eta_{a\nu\alpha}x_\mu x_\alpha). \end{aligned}$$

Choosing, e. g., $a = 2$, $\nu = 3$, $\mu = 4$ this results in

$$\frac{1}{x^2}(-4f'x_3x_2 - 4f'x_1x_4) + \frac{4f}{x^4}(x_1x_4 + x_3x_2 - fx_3x_2 - fx_1x_4) = 0.$$

Equating the same factors of x_3x_2 results in

$$f(1 - f) - x^2f' = 0.$$

All other combinations yield the same condition.

This ordinary differential equation has the solution

$$f = \frac{x^2}{x^2 + \rho^2},$$

with ρ an integration constant. This is in combination with (3.11) the solution to the self-duality equation (3.12), yielding finally the one-instanton solution for Yang-Mills theory

$$A_\mu^a = 2\eta_{a\mu\nu}\frac{x_\nu}{(x^2 + \rho^2)} \quad (3.14)$$

The field strength tensor can then be calculated as

$$G_{\mu\nu}^a = \frac{-4\eta_{a\mu\nu}\rho^2}{(x^2 + \rho^2)^2}$$

The squared field strength is finally given by

$$(G_{\mu\nu}^a)^2 = \frac{192\rho^4}{(x^2 + \rho^2)^4}.$$

Direct integration shows that indeed $Q = 1$,

$$\begin{aligned} S &= \frac{1}{4g^2} \int d^4x (G_{\mu\nu}^a)^2 = \frac{192\rho^4}{4g^2} \int d\phi \int d\cos\theta \int dx \int dt \frac{x^2}{(x^2 + t^2 + \rho^2)^4} \\ &= \frac{60\pi^2\rho^4}{g^2} \int dx \frac{x^2}{(x^2 + \rho^2)^{\frac{7}{2}}} = \frac{60\pi^2\rho^4}{g^2} \frac{2}{15\rho^4} = \frac{8\pi^2}{g^2}. \end{aligned}$$

Thus, the instanton is the lowest-action configuration with non-trivial topological charge, and thus the classical vacuum solution. Since this is the minimal possible value for non-pure gauge configuration with pure-gauge configuration at infinity, it is the classical solution for this boundary condition.

The instanton solution is well localized, and is characterized by its size ρ . It is invariant under rotations in space. It can be shown that is not invariant under global color transformations, and its orientation has to be fixed. This was done implicitly in the ansatz (3.11), but any other would be equally good. Since the instanton can be placed at any position in space, its position is also a free parameter, and set implicitly to the origin in (3.11). 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.

In normal perturbation theory, the coupling constant is associated with the field and not the action. By changing $A \rightarrow gA$ the action (3.10) is made independent of g and the field strength tensor (3.7) is changed to

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c$$

The solution is then

$$\frac{x^2}{g(x^2 + \rho^2)},$$

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

$$\overline{\eta_{a\mu\nu}} = \begin{cases} \varepsilon_{a\mu\nu} & \text{for } \mu, \nu = 1, 2, 3 \\ -\delta_{a\mu} & \text{for } \nu = 4 \\ \delta_{a\nu} & \text{for } \mu = 4 \end{cases} \quad (3.15)$$

which is unchanged besides this replacement.

The instanton solution can be continued analytically to undo the Wick rotation. This yields the sphaleron solution,

$$A_\mu^a = 2\eta_{a\mu\nu} \frac{x_\nu}{x^2 - \rho^2} \quad (3.16)$$

Thus, the sphaleron is singular on a space-like trajectory $0 < x^2 = \rho^2$. In Minkowski space is is therefore an explosive event. In a dynamical situation, which is no longer characterized by a minimal action solution, it is associated with the production of gluons, or classically waves.

3.2.3 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. While they can be induced by suitable initial conditions classically, they really come into their own in the quantum case, where fluctuations can occur making them more relevant.

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 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.,

$$A_{\mu}^a = -\delta^{3a} \frac{1}{g} \vec{e}_{\varphi} \frac{1 + \cos \theta}{r \sin \theta},$$

which is given in spatial spherical coordinates. It is only static, in contrast to the instanton, which is an event 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

$$A_{\eta}^a = \delta^{3a} \frac{1}{g} \frac{\mu(\rho)}{\rho}$$

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 gauge theories.

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.

3.3 Classical tunneling

The instanton solutions can also be translated, and thus instantons can be located anywhere in space-time. Provided their distance is sufficiently large, a sum of the instanton configuration (3.11) will approximately be a solution to the self-dual equation with topological charge being a sum of the individual ones.

A particularly interesting situation arises when combining a far separated pair of an instanton and an anti-instanton. Their total topological charge is then approximately zero. While formally not an exact solution of the self-duality equation, this configuration has an interesting practical interpretation. Such a configuration will have a topological charge density characteristic for the $Q = 1$ case close to the instanton, for $Q = -1$ close to the antiinstanton, and essentially zero far away from both.

Thus, a trajectory starting at infinity and moving towards either instanton will effectively move from an empty vacuum to a topological vacuum, and vice versa. Thus, this can be considered to be a classical tunneling between states, which are usually separated by a non-deformable topological barrier, a kind of classical tunneling process. Similar arguments can be made with more complex configurations of topological states. As will be seen in section 6.4, this has far reaching consequences at the quantum level.

Chapter 4

Perturbative quantization of Yang-Mills theory

So far, Yang-Mills theory was treated entirely classically. It is now time to quantize it. In principle, this would appear to be a straightforward extension of the quantization of non-gauge theories. However, this is not the case, for two reasons. The first is the appearance of the self-interaction in (2.23). This will require additional steps, and will ultimately lead to the introduction of auxiliary fields. The other is that the non-Abelian nature will introduce a problem when constructing the integration range. This second set of problems turn out to be actually not appearing in perturbation theory. Therefore, to deal with one problem at a time, here only a perturbative quantization, i. e. only valid when doing perturbation theory, will be done. The extension beyond perturbation theory will be done in chapter 6. This makes sense, as essentially all the concepts introduced in the perturbative quantization will find their counter-part beyond perturbation theory, but at the expense of more technical difficulties. Thus, such a two-stage approach is particularly suitable for concept building.

4.1 The path integral

This section is brief recap of the path integral for non-gauge theories, together with the necessary tools to treat real-valued, scalar fields. If familiar from the elcture quantum field theory I, this section can be skipped.

4.1.1 Heuristic introduction

The path integral formulation is as axiomatic as is canonical quantization, it cannot be derived. However, it is possible to motivate it.

This is done most easily by defining it as a limiting procedure, using a so-called lattice regularization. As the name implies, it works by replacing space-time with a finite lattice of discrete points. This is just a tool to define a limiting process.

It is best to start with the quantum mechanics version. The heuristic reasoning is then as follows. 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

$$U(a, b, T) = \sum_{\text{All paths}} e^{i \cdot \text{Phase}} \quad (4.1)$$

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 function which implements this is the classical action S , determined as

$$S = \int_{C_a^b} dt L,$$

where the integral is over the classical path C_a^b from a to b . If the classical path is now replaced with some arbitrary path C_a^b , again connecting the points a and b , the action can be considered to be a functional of the path C_a^b and the classical Lagrange function L . Rewriting the sum as a functional integral over all paths, this yields already the definition of the functional integral

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

This defines the quantum mechanical path integral in a symbolic way.

It then remains to give this functional integral a mathematically concise meaning, such that it becomes a mathematical description of how to determine this transition amplitude. It is here where the lattice comes into play. However, for arbitrary interacting theories, there are still conceptual and practical problems, so that the following remains often an unproven procedure.

The starting point was the transition amplitude. In quantum mechanics, this amplitude is given by

$$U(a, b, T = t_N - t_0) = \langle a, t_N | e^{-iHT} | b, t_0 \rangle.$$

In the next step, insert at intermediate times a sum, or integral in cases of a continuous spectrum, over all states

$$U(a, b, T) = \sum_i \langle a, t_N | e^{-iH(t_N-t_1)} | i, t_1 \rangle \langle i, t_1 | e^{-iH(t_1-t_0)} | b, t_0 \rangle.$$

By this, the transition amplitude is expressed by a sum over all possible intermediate states, already quite in the spirit of (4.1). To fully embrace the idea, divide the time interval into N steps of size $\epsilon = T/N$, where N is large and will later be sent to infinity. That is actually already the latticized time. This yields

$$\begin{aligned} U(a, b, T) &= \sum_j \sum_{i_j} \langle a, t_N | e^{-iH\epsilon} | i_{N-1}, t_{N-1} \rangle \dots \langle i_1, t_1 | e^{-iH\epsilon} | b, t_0 \rangle \\ &= \int \prod_i dq_i \langle q_a, t_N | e^{-iH\epsilon} | q_{N-1}, t_{N-1} \rangle \dots \langle q_1, t_1 | e^{-iH\epsilon} | q_b, t_0 \rangle, \end{aligned} \quad (4.2)$$

where in the second line the result was rewritten in terms of a set of continuous eigenstates of the (generalized) position operator Q_i . These are therefore $N - 1$ integrals. Space is still a continuum.

If, as is the case for all systems of interest in the following, the Hamiltonian separates as

$$H = \frac{1}{2} P_i^2 + V(Q),$$

where the P_i and Q_i are the M canonically conjugated momenta, then for ϵ arbitrarily small the Baker-Campbell-Hausdorff formula

$$\exp F \exp G = \exp \left(F + G + \frac{1}{2} [F, G] + \frac{1}{12} ([[F, G], G] + [F, [F, G]]) + \dots \right).$$

yields

$$e^{-iH\epsilon} \approx e^{-\frac{i\epsilon}{2} P_i^2} e^{-i\epsilon V},$$

i. e. for infinitesimally small time steps the exponentials can be separated. Assuming the states to be eigenstates of the position operator and furthermore inserting a complete set of (also continuous) momentum eigenstates allows to rewrite the transition matrix elements as ordinary functions

$$\langle q_{i+1}, t_{i+1} | e^{-iH\epsilon} | q_i, t_i \rangle = e^{-\epsilon V(q_i)} \int \prod_j \frac{dp_j^i}{2\pi} \prod_k e^{-i\epsilon \left(\frac{p_k^2}{2} - ip_k \frac{q_{k+1} - q_k}{\epsilon} \right)}, \quad (4.3)$$

where products run over the number of independent coordinates M . The infinitesimal step (4.3) is also known as the transfer matrix, which transfers the system from one time to another. In fact, even if the Hamilton operator is not known, but only the transfer matrix, it is possible to construct the full transition amplitude, as this only requires to create a product over all transfer matrices.

Defining

$$\mathcal{D}p\mathcal{D}q = \prod_i^N \prod_j^M \frac{dp_j^i dq_j^i}{2\pi}, \quad (4.4)$$

and thus in total $2NM$ integration measures yields the first formulation of the path integral

$$U(a, b, T) = \int \mathcal{D}p\mathcal{D}q e^{-\epsilon p_k \frac{q_j^{i+1} - q_j^i}{\epsilon}} e^{-i\epsilon H(p_j^i, q_j^i)}$$

Taking the limit

$$\frac{q_j^{i+1} - q_j^i}{\epsilon} = d_t q_j^i$$

and performing the Gaussian integrals over the momenta yields

$$U(a, b, T) = \int \mathcal{D}q e^{i \sum^N \epsilon L(q_j^i, d_t q_j^i)} \stackrel{N \rightarrow \infty}{\cong} \int \mathcal{D}q e^{iS}$$

$$\mathcal{D}q = \prod_i^N \prod_j^M \frac{dq_j^i}{\sqrt{2\pi\epsilon}},$$

where L is the Lagrange function of the system, thus arriving at the original idea (4.1).

Considering the result in detail, it is important to note one important feature. The definition requires to chose any straight line between every point at every time. Thus, in general paths will contribute which are not differentiable. This is a very important insight: Quantum physics differs from classical physics not only by including all possible paths, but also by including not only differentiable paths. This is in stark contrast to Hamilton's principle of classical mechanics.

Passing now to a field theory, the transition is the same as in classical mechanics: The paths are replaced by the fields, the Lagrange function by the Lagrangian density, and the action is an integral over space-time. Of particular importance is then the partition function

$$Z = \int \mathcal{D}\phi e^{i \int d^d x \mathcal{L}(\phi, \partial_\mu \phi)}, \quad (4.5)$$

where the integral is over all possible field configurations, i. e. the set of all possible values of the fields at every space-time point, including any non-differentiable ones¹. Since any

¹In fact, it can be shown that those are the dominating ones. Making sense out of this expression in the continuum is highly non-trivial and requires to pass from Riemann integrals to different definitions of integrals, but this is not the subject of this lecture.

field configuration includes the time-dependence, the path-integral can be considered as an integral over all possible field configurations, and thus histories of the universe described by the Lagrangian \mathcal{L} , from the infinite past to the infinite future. Thus, the path integral makes the absence of locality in quantum physics quite manifest. The partition function (4.5) is essentially the transition function from the vacuum to the vacuum. It is important to note that in the whole setup the field variables are no longer operators, like in canonical quantization, but ordinary functions.

It is also important to note that the integration is over all possible field configurations. Thus, in contrast to the beginning and end of the paths in the quantum-mechanical case, there are no boundary conditions imposed. Of course, if desired, additional boundary conditions can be imposed, as can be any kind of underlying space-time manifold. But for any set of boundary conditions and/or any change of manifold a different theory will result. Neither of this will be done in this lecture.

While the vacuum-to-vacuum transition amplitude is a very useful quantity, what is really important are the expectation values of the correlation functions, i. e. expectation values of products, or other functions, of more than one field. These can be determined in a very similar way as before to be

$$\langle T(\phi(x_1)\dots\phi(x_n)) \rangle = \int \mathcal{D}\phi \phi(x_1)\dots\phi(x_n) e^{i \int d^d x \mathcal{L}(\phi, \partial_\mu \phi)}. \quad (4.6)$$

Here, T means time-ordering, i. e. that the time arguments x_i^0 are ordered in descendant order, i. e.

$$T(\phi(t_1)\phi(t_2)) = \theta(t_1 - t_2)\phi(t_1)\phi(t_2) \pm \theta(t_2 - t_1)\phi(t_2)\phi(t_1), \quad (4.7)$$

where the minus sign appears if the ϕ anticommute. Note that this time-ordering does not appear inside the path integral, but is created automatically. Inside the path integral, the ordering is arbitrary.

It is here implicitly assumed that $Z = 1$, i. e. that the measure of the path integral is normalized such that the expectation value of unity is one, $\langle 1 \rangle = 1$. Otherwise

$$\langle T\phi(x_1)\dots\phi(x_n) \rangle = \frac{\int \mathcal{D}\phi \phi(x_1)\dots\phi(x_n) e^{iS[\phi]}}{\int \mathcal{D}\phi e^{iS[\phi]}}, \quad (4.8)$$

holds.

There are two important remarks. This can be seen by noting that the fields are evaluated at fixed times, and therefore evaluate to functions of the positions in their respective transfer matrix elements (4.3). Thus, any expectation value is a path integral over the fields as ordinary functions weighted by the phase.

4.1.2 Functional analysis

So far, any explicit calculation with the path integral would require to go back to the limiting prescription (4.2). This would be quite awkward. Fortunately, this is usually not necessary, and it is often possible to work with the path integral in much the same way as with ordinary integrals.

How to do so falls into the mathematical purview of functional analysis and distribution theory to provide any level of rigor. For most applications in physics, little of this rigor is necessary. Thus, here results will be collected, which are useful to work in an operational way with the path integral, but they will not be proven. Rather, the corresponding mathematical literature and/or lectures can provide this, where desired and necessary. Nonetheless, it is advised that, with all simplicity the following seems to bring, one should always be wary that many issues run much deeper.

The starting point before defining functional integration is the definition of a functional derivative. Essentially, this is an extension of variational calculus in classical mechanics. The basic ingredient for a functional derivative δ with respect to a function (field) $\phi(x)$ 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. The last identity is known as the Leibnitz rule.

Consequently, a power series of a functional is defined as

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

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

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

This defines the most important concepts for differentiation. If not stated otherwise, it will be assumed in the following that any functional can be written as functional Taylor series. Just as with ordinary functions, this is not guaranteed.

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 $\delta \int$ does not equal to one, but to a δ -function.

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

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{\pi}} e^{-ax^2} = \frac{1}{\sqrt{a}}. \quad (4.9)$$

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

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

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

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

where A may now be operator-valued. Especially derivative operators may appear in this context. The determinant of such an operator can be evaluated by the expression

$$\det A = \exp \operatorname{tr} \ln(A), \quad (4.10)$$

just like for matrices, which is of great practical relevance. Herein, both the logarithm and the exponential of an operator are defined by the usual power series of these operations². Alternatively, $\det A$ can be expressed in terms of the solutions of the eigenvalue equation

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

²This implies that problems may arise, as this is not always justified. In fact, there exist operators for which even the trace is not well defined. Even though this rarely plays a role in the following, caution is in general advised.

where the eigenvalues λ can form a (complex) continuum, and finite, or even infinite, degeneracies are possible. 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

$$\int \mathcal{D}\phi F[\phi + \alpha\eta] \stackrel{\phi \rightarrow \phi - \alpha\eta}{=} \int \mathcal{D}\phi F[\phi] \quad (4.11)$$

holds by definition.

From these properties follows the validity of the substitution rule as

$$\int \mathcal{D}\phi F[\phi] = \int \mathcal{D}\psi \det\left(\frac{\delta\phi}{\delta\psi}\right) F[\phi[\psi]], \quad (4.12)$$

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

$$\phi(x) = \int dy \eta(x, y) \psi(y), \quad (4.13)$$

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

4.1.3 Sources

Combining all of the above allows for a different way of calculating correlations functions than by direct evaluation of (4.8). For this, introduce³ a so-called (external) source $j(x)$, and replace

$$iS[\phi] \rightarrow iS[\phi] + \int d^d x \phi(x) j(x).$$

Thus, the partition function will become a functional of $j(x)$. Then a more elegant way to express correlation functions is by

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

Note that it was assumed that functional derivation and integration commute, and that often the normalization $Z[0] = 1$ is chosen.

³There 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 applications.

Furthermore, this permits to reconstruct the original path-integral, or generating functional, as

$$Z[j] = \sum_{n=0}^{\infty} \int d^d x_1 \dots d^d x_n \langle T \phi(x_1) \dots \phi(x_n) \rangle j(x_1) \dots j(x_n), \quad (4.14)$$

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

4.2 Quantization of Maxwell theory

In principle, quantizing Maxwell theory is now performed by writing down the path integral and calculate correlation functions. 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 (2.4) 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. Explicitly, it takes the form

$$Z[j_\mu] = \int \mathcal{D}A_\mu \exp \left(i \int d^d x \left(\frac{1}{2} A^\mu (g_{\mu\nu} \partial^2 - \partial_\mu \partial_\nu) A^\nu + j^\mu A_\mu \right) \right). \quad (4.15)$$

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

$$g_{\mu\nu} \partial^2 - \partial_\mu \partial_\nu \quad (4.16)$$

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 understand the problem is to note that any gauge transformation (2.7) leaves S invariant, and, as a shift, also does not influence the measure. Thus, there are flat directions of the integral, and thus the integral diverges when integrating along this direction. Thus, at first sight, it seems to be impossible to find a well-defined path integral for Maxwell theory.

There are only few possibilities to escape. One is to perform the quantization on a discrete space-time grid in a finite volume, determine only gauge-invariant 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, so called Lattice Monte Carlo simulations of lattice gauge theory. This approach is a subject of a separate lecture.

Another one is to determine only quantities which are invariant under gauge transformations. Classically, this is perfectly feasible, as an explicit formulation only in terms of electric and magnetic fields, both manifestly gauge-invariant, shows. At the quantum level, however, this is not so easy. After all, this amounts to replacing the vector potential by the electromagnetic fields. The Jacobian then involves solving the partial differential equations (2.5-2.6), yielding non-trivial integrals, and ultimately leading to a non-local action. Still, this is possible, though very, very painful. It is, however, far too painful for the kind of generalization of gauge theories needed in particle physics, and even for QED in practical problems far too complicated. Thus, while reassuring that it would be conceptually possible, is not a practical solution.

In this context one may wonder why not directly integrating over the electromagnetic fields in the path integral (4.15) instead of the gauge fields. This is an inequivalent path integral to the one obtained by solving (2.5-2.6) as a variable transformation in (4.15). Interestingly, while mathematically perfectly feasible and doable, this does not yield a quantum theory compatible with experiments. Why this is so is a very good, and unsolved, question. It is a purely empirical outcome. This is even more surprising as the gauge-fields are classically an irrelevant technical tool, and actually introduce redundancies. They are thus quite the opposite to what is usually done in classical mechanics, the reduction to generalized coordinates. Still, if one wants to base the construction of the theory on experiment, one is therefore stuck with (4.15).

There is a third option, and the one to be used in this lecture. As the possibility to introduce gauge conditions such as (2.10) shows, the additional degrees of freedom over which is integrated do not carry physical information. The aim is therefore to separate off this unphysical degrees of freedom in a way which allows to remove it. In the end, this will lead to a(n infinite) prefactor of the path integral, which drops out when normalizing the partition function at zero source to unity. Even though unphysical information is removed, this is not equivalent to introducing a gauge-invariant formulation. Any gauge condition will define one distinct way of removing the superfluous degrees of freedom, but eventually one still works with the gauge fields A_μ . But this field will differ for every gauge condition, and therefore depends on the gauge condition. But this difference depends only on the gauge condition, and will therefore drop out in gauge-invariant quantities, making them physical.

In a very rough way, the three options can also be cast into this language: Lattice Monte

Carlo simulations sample a subset of coordinate systems, and averages over the choices. By using a finite number of choices the infinity of choices does not pose a problem. Removing the gauge fields is calculating coordinate-independent, which is always more painful. And gauge-fixing is just working in a choice of coordinate which is fixed once and for all.

After these initial remarks, it is now time to actually perform this gauge-fixing, and quantize Maxwell theory. To this end select, as in classical electrodynamics, a gauge condition $C[A_\mu, x] = 0$ which selects uniquely the gauge field. I. e., for a set of gauge-fields related by gauge transformations (2.7) there is one and only one, but also at least one, which satisfies the condition C when going through all possible g s. This is actually not a necessary condition to make progress, and it will be relaxed later on. But for the moment, it makes the calculation simplest.

An example of such a condition is, e. g., the Landau gauge (2.10), which in this language takes the form, suppressing from now on possible explicit dependencies on x_μ ,

$$0 = C[A_\mu] = \partial^\mu A_\mu. \quad (4.17)$$

To make the path integral well-defined, it is then necessary to get rid of all the gauge transformed fields which do not satisfy this condition. Then just one representative for each physically inequivalent field configuration is left⁴.

The question is now how to enforce this in practice. To do this consider the functional generalization of the Dirac- δ function. The expression

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

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 $g \rightarrow g + g'$ 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

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

⁴Note that this does not fix the harmonic part of the gauge transformations. However, they form a measure zero part of the whole set of gauge transformations, and therefore do not matter. Also, in any perturbative calculation the fields drop to zero at infinity, and thus in this case by definition they are excluded.

Inverting Δ , the relation

$$1 = \Delta[A_\mu] \int \mathcal{D}g \delta(C[A_\mu^g]) \quad (4.18)$$

is found.

Inserting this into the functional integral yields

$$\begin{aligned} Z &= \int \mathcal{D}A_\mu \Delta[A_\mu] \int \mathcal{D}g \delta(C[A_\mu^g]) \exp(iS[A_\mu]) \\ &= \int \mathcal{D}g \int \mathcal{D}A_\mu^{g'} \Delta[A_\mu^{g'}] \delta(C[A_\mu^{g+g'}]) \exp(iS[A_\mu^{g'}]) \\ &= \int \mathcal{D}g \int \mathcal{D}A_\mu \Delta[A_\mu] \delta(C[A_\mu]) \exp(iS[A_\mu]) \end{aligned} \quad (4.19)$$

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 over g 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 was demanded that it is always possible to resolve the condition $C[A_\mu^g] = 0$ to obtain g as a function of C . Then, by exchanging C and g as variables of integration, it is found that

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

The appearing determinant is just the corresponding Jacobian. Thus, the function Δ is given by

$$\Delta[A_\mu] = \left(\det \frac{\delta C[A_\mu, x]}{\delta g(y)} \right)_{C=0} = \det M(x, y). \quad (4.20)$$

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

$$\begin{aligned} M(x, y) &= \frac{\delta C[A_\mu, x]}{\delta g(y)} = \int d^d z \frac{\delta C[A_\mu, x]}{\delta A_\mu(z)} \frac{\delta A_\mu(z)}{\delta g(y)} \\ &= \int d^d z \frac{\delta C[A_\mu, x]}{\delta A_\mu(z)} \partial_\mu^y \delta(y - z) = -\partial_\mu^y \frac{\delta C[A_\mu, x]}{\delta A_\mu(y)}. \end{aligned} \quad (4.21)$$

To proceed further, a choice of C is necessary. If C is a local gauge condition, i. e. only involving the fields and their derivatives, the Faddeev-Popov operator will be local, i. e.

$$M \sim \delta(x - y).$$

If this is not the case, the resulting theory will be inherently non-local. Such a local choice is always possible in perturbation theory. Choosing, e. g., the Landau gauge $C = \partial^\mu A_\mu = 0$ yields

$$M(x, y) = -\partial^2 \delta(x - y). \quad (4.22)$$

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 the result (4.22) is independent of the field variables. Thus, this factor can be absorbed in the normalization constant. But then the original problem is solved and everything is complete. However, the resulting integral has always the implicit Landau gauge condition to be taken into account. Especially, this implies that the gauge field is always transverse.

This is implicit, and thus somewhat cumbersome. It can be made more explicit by taking a detour. To do so select as gauge condition

$$C = D[A_\mu, x] + \Lambda(x) \quad (4.23)$$

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

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

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 or R_ξ gauges, this yields the final expression

$$Z = \int \mathcal{D}A_\mu \exp\left(iS - \frac{i}{2\xi} \int d^d x (\partial_\mu A^\mu)^2\right). \quad (4.24)$$

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

$$g_{\mu\nu} \partial^2 - \partial_\mu \partial_\nu \rightarrow g_{\mu\nu} \partial^2 - \left(1 - \frac{1}{\xi}\right) \partial_\mu \partial_\nu, \quad (4.25)$$

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 $\xi \rightarrow 0$, 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 (4.16) in Maxwell theory. The choice $\xi = 1$ is known as Feynman gauge.

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 $\sum_{i=1}^3 \partial_i A_i = 0$ or axial gauges $n^\mu A_\mu = 0$ with some arbitrary vector n_μ . However, these kind of gauges have particular problems, as the gauge conditions themselves are not invariant under Lorentz transformation. E. g., the Coulomb gauge condition can be removed by a suitable Lorentz boost. They therefore introduce additional problems. To avoid these, in this lecture only Lorentz covariant gauges will be discussed. A source for more information on non-covariant gauges in perturbation theory is, e. g., Burnel, Springer (2008).

This process has at no place involved explicitly any matter fields. They will be added now.

Quantizing this theory is now straightforward, as at no point the gauge freedom is interfering with the additional term, though, of course, any gauge transformation applied to the gauge field needs to be applied also to the matter field. Other than that, quantization can be performed in the same way in the presence of matter fields, no matter whether they are fermionic or bosonic. 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

$$Z = \int \mathcal{D}A_\mu \mathcal{D}\psi_f \mathcal{D}\bar{\psi}_f \exp \left(-i \int d^d x \left(\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \sum_f \bar{\psi}_f (i\gamma^\mu D_\mu^f - m_f) \psi_f + \frac{1}{2\xi} (\partial_\mu A^\mu)^2 \right) \right), \quad (4.26)$$

from which now calculations can be performed. How this can be done in practice will be discussed after extending the quantization process to Yang-Mills theories, in chapter 5. The index f indicates that multiple fermion fields, so-called flavors, can be present. E. g. proton and electron to describe a hydrogen atom. Each of them has potentially different masses m_f and different electric charges e_f (thus the index at the covariant derivative).

4.3 Gauge-dependent correlation functions

There is now one subtlety. What is the now the value of the partition function, the answer to the question that started all of this? Because the matrix (4.25) is now invertible, this seems to be obvious. But it is not.

To understand what is happening, it is best to consider an explicit correlation functions, i. e. an expectation value. The photon propagator is the expectation value

$$\langle A_\mu(x)A_\nu(y) \rangle = \int \mathcal{D}A_\rho A_\mu A_\nu e^{iS}.$$

Irrespective of the issues with gauge symmetry, this expression will yield zero. The reason is, as noted before, that because space-time is isotropic, there is no preferred direction. Thus, a tensor cannot be non-zero. To avoid this problem, it is necessary to contract $A_\mu A_\nu$ with an invariant tensor of the Lorentz group. There are two symmetric rank two tensors suitable for the task, $g_{\mu\nu}$ and $p_\mu p_\nu$.

Denoting either of them as $P_{\mu\nu}$, the actual object to calculate is

$$\langle P^{\mu\nu} A_\mu(x)A_\nu(y) \rangle = \int \mathcal{D}A_\rho P^{\mu\nu} A_\mu A_\nu e^{iS}, \quad (4.27)$$

twice, once for each possible choice of $P_{\mu\nu}$. This object is well-defined from the space-time perspective, but not from the gauge symmetry perspective. This puts one on the path of section 4.2.

But before gauge-fixing, it is an interesting question to ask what happens if one tries to calculate the expression (4.27) by brute force. This can be done, and the result is actually $\sim \delta(x - y)$. The reason is that for any gauge field configuration with some value $A_\mu(x_0)$ at the fixed position x_0 , there exists a gauge transformation, which is only non-vanishing at x_0 , such that the value of the gauge transformed gauge field is $-A_\mu(x_0)$. In this way, any integration over the full gauge group yields zero. The only exception can happen if $x = y$, because then this is essentially the square of the gauge field, and thus positive. Hence, the propagator is only non-vanishing at coinciding space-time points. This argument can be extended to any gauge-dependent correlation function. Thus, without gauge-fixing all gauge-dependent quantities vanish up to expressions proportional to $\delta(x_i - x_j)$ for its arguments.

But this should be remedied by gauge-fixing. But it is not as simple. The decisive step is the expression (4.19). Here, the integral $\int \mathcal{D}g$ was absorbed in the normalization, because none of the remaining expressions depended on them, because all were gauge-invariant. This is no longer true, if the integral is taken over gauge-dependent quantities, like the photon propagator. Thus, the gauge transformation integral can then no longer be

separated as a factor, and be removed. Thus, the whole procedure of gauge-fixing seems to break down for gauge-dependent quantities.

There if, fortunately, a solution to this. The expression (4.19) can also be interpreted differently for any gauge-invariant observable $f(A_\mu)$. Denoting the set of all gauge field configurations, including all gauge-transformed field configurations, as Ω , the following expressions are identical

$$\begin{aligned}
& \frac{1}{\mathcal{N}} \int_{\Omega} \mathcal{D}A_\mu f(A_\mu) \exp(iS[A_\mu]) \\
&= \frac{1}{\mathcal{N}} \int_{\Omega/\Omega_c} \mathcal{D}g \int_{\Omega_c} \mathcal{D}A_\mu \Delta[A_\mu] \delta(C[A_\mu]) f(A_\mu) \exp(iS[A_\mu]) \\
&= \frac{1}{\mathcal{N}'} \int_{\Omega} \mathcal{D}A_\mu \Delta[A_\mu] \delta(C[A_\mu]) f(A_\mu) \exp(iS[A_\mu]) \\
&= \frac{1}{\mathcal{N}''} \int_{\Omega_c} \mathcal{D}A_\mu \Delta[A_\mu] f(A_\mu) \exp(iS[A_\mu]),
\end{aligned} \tag{4.28}$$

where Ω_c is the set of all gauge field configurations satisfying the gauge condition C_c . Thus, gauge-invariant expectation values are the same, whether they are calculated over the whole gauge field configuration space or whether over a restricted gauge-field configuration space, provided the Faddeev-Popov determinant is included to compensate for the geometric structure of the restricted space. Thus, on the level of gauge-invariant quantities, all expressions describe the same theory. The normalization constants differ, but can always be chosen such that $\langle 1 \rangle = 1$.

However, for gauge-dependent observables the expressions are not equivalent, for the arguments given above. In the first two lines, they vanish, but not in the third line and fourth line. Thus, from a purely mathematical point of view, these theories are distinct. From the point of physics, there is just an infinite number of equivalent quantum theories, the one without gauge fixing and the infinitely many choices of Ω_c , which all give the same gauge-invariant observables.

Alternatively, this can also be taken to imply that any choice of theory with the action

$$S' = S - i \ln \Delta_c \tag{4.29}$$

gives the same observable quantities, provided they are integrated over the corresponding configuration set Ω_c , either directly implemented as integration range or by a δ -function. In either way, this leads ultimately to the expression (4.24). Note that gauges like the linear-covariant gauges then can be considered to be averages over theories with different Ω_c .

This infinite degeneracy of quantum theories is a consequence of working with redundant variables. If it would be technically possible to go to generalized variables, just one theory would remain. Hence, the degeneracy should rather be considered to be a choice of suitable variables for technical purposes rather than having any physical meaning. At any rate, for the purpose of this lecture, expressions like (4.24) will be used as the definition of the quantum theory, while, e. g., lattice calculations rather start at (4.28). But, as was just shown, both yield eventually the same results for observables.

4.4 Quantization of Yang-Mills theory

For the same reasons as for Maxwell theory in section 4.2 also in Yang-Mills theory flat directions appear. As there, different possibilities are available to cope with the problem, of which gauge-fixing will be the method of choice here. As in the case of Maxwell theory this will be restricted to gauge conditions, which will not involve the matter fields in a manifest way. Thus, the following is the same for any matter content of the theory⁵, and the matter fields need not be included explicitly. 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 gauge or axial gauges. As they introduce a host of additional technical complexities, without yet any obvious and simple gains, this topic is beyond the scope of this lecture.

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

$$Z = \int \mathcal{D}G \int \mathcal{D}A_\mu^a \Delta[A_\mu^a] \delta(C^a[A_\mu^a]) \exp(iS[A_\mu^a]) \quad (4.30)$$

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 functional Δ is given by

$$\Delta[A_\mu^a] = \left(\det \frac{\delta C^a[A_\mu, x]}{\delta \theta^b(y)} \right)_{C^a=0} = \det M^{ab}(x, y). \quad (4.31)$$

with the non-Abelian Faddeev-Popov operator M^{ab} . What still can be absorbed into the measure is the infinite constant $\int \mathcal{D}G$. This, at least, renders the integral finite.

⁵Examples of gauge conditions which involve the matter fields explicitly are discussed in the lecture on electroweak physics.

To proceed, a more explicit expression for the Faddeev-Popov operator is again obtained using the chain rule

$$\begin{aligned} M^{ab}(x, y) &= \frac{\delta C^a(x)}{\delta \theta^b(y)} = \int d^4 z \frac{\delta C^a(x)}{\delta A_\mu^c(z)} \frac{\delta A_\mu^c(z)}{\delta \theta^b(y)} \\ &= \int d^4 z \frac{\delta C^a(x)}{\delta A_\mu^c(z)} D_\mu^{cb} \delta(y - z) = \frac{\delta C^a(x)}{\delta A_\mu^c(y)} D_\mu^{cb}(y). \end{aligned} \quad (4.32)$$

The important difference to the Abelian case (4.21) is the appearance of the covariant derivative instead of the ordinary one. As a consequence, the gauge field itself appears in the Faddeev-Popov operator for any gauge condition⁶. This is what makes it impossible to get rid of the Faddeev-Popov determinant in the same way as in Maxwell theory.

To proceed further, a choice of C^a is necessary. This 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

$$\begin{aligned} Z &= \int \mathcal{D}\Lambda^a \mathcal{D}A_\mu^a \exp\left(-\frac{i}{2\xi} \int d^4 x \Lambda^a \Lambda_a\right) \det M \delta(C) \exp(iS) \\ &= \int \mathcal{D}A_\mu^a \det M \exp\left(iS - \frac{i}{2\xi} \int d^4 x D^a D_a\right), \end{aligned} \quad (4.33)$$

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

This determinant can be recast, by introducing auxiliary fields, as an exponential. Using the rules of anticommuting Grassmann numbers it follows immediately that

$$\det M \sim \int \mathcal{D}c^a \mathcal{D}\bar{c}^a \exp\left(-i \int d^4 x d^4 y \bar{c}_a(x) M^{ab}(x, y) c_b(y)\right), \quad (4.34)$$

where the auxiliary (Faddeev-Popov) ghost and antighost fields c and \bar{c} are Grassmann-valued, i. e. anti-commuting, scalar fields. Note that $c^\dagger = c$ and $(i\bar{c})^\dagger = i\bar{c}$ holds, and there is no relation between ghost and antighosts, despite their (historic) names⁷. Since the ghosts are just auxiliary fields, spin and Grassmann parity assignments are not at odds with the spin-statistics theorem. As a consequence the sign of the kinetic term of the ghosts is wrong compared to ordinary scalars, a sign of this unphysical spin-statistic

⁶The covariant derivative has even perturbatively zero modes, and can thus not be inverted to get rid of this problem.

⁷Though there is a relation in Landau gauge, but not in general covariant gauges.

relation. If the condition D^a is local in the fields, the Faddeev-Popov operator will be proportional to $\delta(x - y)$, and the ghost action will become local.

It is furthermore often useful to introduce an additional auxiliary field, the real, scalar 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 for the gauge-fixed path integral then reads

$$Z = \int \mathcal{D}A_\mu^a \mathcal{D}b^a \mathcal{D}c^a \mathcal{D}\bar{c}^a \exp\left(iS + \int d^4x \left(\frac{\xi}{2}b^a b_a + b_a D^a\right) - \int d^4x d^4y \bar{c}^a(x) M^{ab}(x, y) c^b(y)\right). \quad (4.35)$$

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

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

Furthermore, the ever-so popular Landau gauge corresponds to the limit $\xi \rightarrow 0$, as this is corresponding to the case where all of the weight of the weight-function is concentrated only on the gauge copy satisfying $\partial^\mu A_\mu^a = 0$. 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

$$D_\mu^{ab} \phi^b \rightarrow \partial_\mu \phi,$$

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 gauge condition which is not linear in the gauge fields will also in an Abelian theory introduce interactions.

This program can be performed in a much more formal and general way, the so-called anti-field method. This will be introduced in section 4.6. It is also possible, but much more tedious and less developed, using canonical quantization. This will therefore not be done in this lecture.

To include matter fields 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, just like in section 4.2. Also, the same applies for gauge-dependent correlation functions as for the Abelian theory in section 4.3.

4.5 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. The situation beyond perturbation theory is addressed in chapter 6.

4.5.1 BRST symmetry

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

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

Herein the gauge condition is encoded in the condition $C^a = D^a[A_\mu, x] - \Lambda^a(x) = 0$. Furthermore, matter fields are ignored, as they will not alter the discussion qualitatively. Their contributions will be reinstated later.

Gauge-fixing actually did not fix the gauge completely. One freedom is that if $D^a \equiv D^a[A_\mu^a]$ only depends on gauge fields of the same color a explicitly there is a global gauge transformation still possible, i. e. a space-time independent rotation $G \neq G(x)$. This just states that there is no global color direction preferred by the gauge choice.

But there is also another symmetry. As will be seen, it is related to the fact that covariant gauges, except for Landau gauge, average over gauge copies. The symmetry will be related to gauge transformations between two different copies over which is averaged, accompanied by a readjustment of the Jacobian, and thus the ghost fields. It is hence a global symmetry, as the averaging was not done in a space-time-dependent fashion. This will be the BRST symmetry. The existence of this symmetry is far from obvious.

It is here best to work backwards, i. e. starting from the known form of the BRST symmetry. The associated infinitesimal BRST transformation δ_B is defined as

$$\delta_B A_\mu^a = \lambda D_\mu^{ab} u_b = \lambda s A_\mu^a \quad (4.36)$$

$$\delta_B u^a = -\lambda \frac{g}{2} f^{abc} u_b u_c = \lambda s u^a \quad (4.37)$$

$$\delta_B \bar{u}^a = \lambda b^a = \lambda s \bar{u}^a \quad (4.38)$$

$$\delta_B b^a = 0 = \lambda s b^a. \quad (4.39)$$

Herein, λ is an infinitesimal Grassmann number, i. e., it anticommutes with the ghost fields. This is a very different form than usual symmetries, which involve only non-Grassmann parameters, and has thus some quite unique properties.

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

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

with 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 trivial for the classical Lagrangian, as the transformation for the gauge boson is just an ordinary infinitesimal gauge transformation with the odd gauge parameter λu^a , which is an ordinary real function. This is the manifestation of what has been said above that this is a transformation between the different gauge copies.

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 as

$$s(b^a D_a) = b^a \int d^4y \frac{\delta D^a}{\delta A_\mu^b} sA_\mu^b = b^a \int d^4y \frac{\delta D^a}{\delta A_\mu^b} D_\mu^{bc} u_c.$$

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

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

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,

$$\begin{aligned} & \int d^4y s \left(\frac{\delta D^a}{\delta A_\nu^b(y)} \right) D_\nu^{bc} u_c = \int d^4y d^4z \frac{\delta D^a}{\delta A_\nu^b(y) \delta A_\rho^d(z)} (sA_\rho^d(z)) D_\nu^{bc} u_c(y) \\ & = \int d^4y d^4z \frac{\delta D^a}{\delta A_\nu^b(y) \delta A_\rho^d(z)} D_\rho^{de} u_e(z) D_\nu^{bc} u_c(y) = 0. \end{aligned}$$

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

$$\begin{aligned}
s(D_\mu^{ab} u_b) &= \partial_\mu s u^a - g f_c^{ab} ((s A_\mu^c) u_b + A_\mu^c s u_b) \\
&= -\frac{g}{2} \partial_\mu (f^{abc} u_b u_c) - g f_c^{ab} D_\mu^{cd} u_d u_b - g f_c^{ab} f_{bde} A_\mu^c u^d u^e \\
&= \frac{g}{2} f_c^{ab} (\partial_\mu (u_b u^c) - 2 u_b \partial_\mu u_c - 2g (f_e^{cd} A_\mu^e u_d u_b + g f_b^{de} A_\mu^c u_d u_e)).
\end{aligned}$$

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,

$$\begin{aligned}
&= \frac{g}{2} f^{abc} (u_b \partial^\mu u_c + (\partial^\mu u_b) u_c - 2 u_b \partial^\mu u_c) \\
&\quad + g (f^{abc} f_c^{de} A_e^\mu u_d u_b + f^{abc} f_c^{de} A_e^\mu u_d u_b + f^{abc} f_c^{de} A_b^\mu u_d u_e) \\
&= g (f^{abc} f^{dec} A_e^\mu u_d u_b + f^{adc} f^{ebc} A_e^\mu u_d u_b + f^{aec} f^{dbc} A_e^\mu u_d u_b) \\
&= g (f^{abe} f_c^{cd} + f^{ace} f_e^{db} + f^{ade} f_e^{bc}) A_c^\mu u_d u_b = 0,
\end{aligned}$$

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 nilpotent, i. e., $s^2 = 0$. This follows immediately from a direct application. The previous calculation already showed that

$$0 = s(D_\mu^{ab} u^b) = s^2 A_\mu^a.$$

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

$$\begin{aligned}
\mathcal{L}_f &= s \left(\bar{u}^a \left(\frac{\xi}{2} b^a + D^a \right) \right) \\
&= \frac{\xi}{2} b^a b_a + b^a D_a + \bar{u}_a \int d^4 y \frac{\delta D^a}{\delta A_\mu^b(y)} D_\mu^{bc} u_b(y).
\end{aligned} \tag{4.40}$$

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, which will be discussed in more detail in section 4.6. This is also very much in line with section 4.3: The additional action added for the different quantum versions of the theory in (4.29) is a BRST-exact term.

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

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

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.

It should be noted that if the Nakanishi-Lautrup fields would have not been introduced, the BRST transformation would turn out to be nilpotent of order three, $s^2 \neq 0$ but $s^3 = 0$. This yields ultimately to a non-linear representation of the underlying symmetry group.

4.5.2 Constructing the physical state space

The following discussion shows how to explicitly construct the state space using BRST symmetry. It can be directly extended to include also matter fields.

4.5.2.1 Negative-norm states

The first concept in constructing the physical state space is the realization of 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

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

The expectation value of the probability density of the photon field

$$\Psi(x) = \int d^4 x f(x) A_0(x) = \int \frac{d^4 x d^4 p}{(2\pi)^4} e^{ip_0 x_0 - p_i x_i} f(p) A_0(x),$$

with $f(x)$ an arbitrary weight function, then reads

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

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.

4.5.2.2 BRST algebra

That they indeed do not contribute can be shown using the BRST symmetry. To this end, the generator Q_B of BRST transformations

$$[Q_B, \psi]_{\pm} = s\psi,$$

the BRST charge, Q_B will be useful tool to define the physical states. 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 noted above. This symmetry is a new symmetry, in contrast to the residual global color symmetry and BRST symmetry itself. It is not related to a gauge transformation,

as it does not change the gauge fields, it acts as an identity transformation on them. Its origin is that an additional auxiliary degree of freedom has been introduced with the ghost fields, their relative length. Hence, the relative length of the ghost fields is a redundant degree of freedom, and could be integrated out without affecting the physical content of the theory. In reality, this reduces again practicality, and it is not done.

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

$$\{Q_{\text{BRST}}, Q_{\text{BRST}}\} = 0 \quad (4.41)$$

$$[iQ_G, Q_{\text{BRST}}] = Q_{\text{BRST}} \quad (4.42)$$

$$[iQ_G, iQ_G] = 0 \quad (4.43)$$

as can be seen from an explicit calculation.

4.5.2.3 Physical states

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

$$[Q_B, H] = 0$$

$$[Q_B, S] = 0.$$

Hence, if in fact the BRST symmetry is manifest⁸, then BRST charge is conserved in scattering process. Especially that implies that initial states with fixed BRST charge will only evolve into final states of the same BRST charge. Since s is a gauge transformation, this implies any quantity which changes non-trivially under s is not invariant under gauge transformations, and can thus not be physical. Thus, any set of fields Ω describing a physical initial state must satisfy $s\Omega = 0$. This implies that the S matrix will only scatter gauge-invariant states into gauge-invariant states. That is as it needs to be by construction of gauge invariant quantities. Note that it still makes mathematical sense to calculate the S matrix between non-invariant states, even if this does not describe physical, i. e. gauge-invariant, states.

Because the BRST charge is nilpotent any object on which the BRST transformation can act can be separated in three subsets:

- Objects which are not annihilated by s , $V_2 = \{\psi | s\psi \neq 0\}$.

⁸The 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.

- Objects which are obtained by s from V_2 , $V_0 = \{\phi | \phi = s\psi, \psi \in V_2\}$. As a consequence $sV_0 = 0$.
- States which are annihilated by s but do not belong to V_0 , $V_1 = \{\chi | s\chi = 0, \chi \neq s\psi \forall \psi \in V_2\}$.

The states in V_2 change non-trivially under s , and therefore would not be physical as they are not gauge-invariant. The union of the two other states form the physical subspace,

$$V_p = V_0 \cup V_1.$$

Elements from V_p will evolve, or scatter, into states in V_p . However, given any correlation function build from fields Ω in V_p and thus satisfying $s\Omega_i = 0$ can always be replaced by $\Omega_i + \phi$ without altering a correlation function,

$$\Omega_1\Omega_2 \rightarrow (\Omega_1 + \phi)\Omega_2 = \Omega_1\Omega_2 + s\psi\Omega_2 = \Omega_1\Omega_2 + s(\psi\Omega_2) + \psi s\Omega_2 = \Omega_1\Omega_2.$$

The last term vanishes by construction. The second-to-last term is a gauge transformation of a quantity. However, since $(\Omega_1 + \phi)\Omega_2$ is invariant under a gauge transformation by construction, this term needs to be either gauge-invariant, in conflict with the starting point, or zero.

Therefore, the set of physical operators H_p can be defined as the quotient set

$$H_p = V_p/V_0 = \frac{\text{Ker } s}{\text{Im } s}, \quad (4.44)$$

the so-called cohomology of the operator s . Therefore, H_p describes gauge-invariant quantities. Using a suitable regulator, e. g. lattice, it can be shown than all expectation values of gauge-invariant fields have positive norm, solving the problem raised in section 4.5.2.1.

It is useful to introduce the notation that an element of the kernel of a nilpotent operator is called closed or a cocycle. Thus, a closed element vanishes when acted upon with the operator. An element in the image of a nilpotent operator is called exact or coboundary.

4.5.2.4 Observable particles

It remains to sort the particles to whetehr they are physical or not. In principle, this would be needed to be done for every possible quantum number channel using the LSZ formalism. However, in perturbation theory, there is a simpler procedure.

Perturbatively, by construction, asymptotically the theory is the free theory. Especially, as sufficiently large space-like separations, all interaction ceases⁹. Thus, any observable

⁹Haag's theorem forbids this beyond perturbation theory, see chapter 6.

particle can only exist in the quantum number channels of the elementary ones. Moreover, the LSZ construction allows to consider any field with the same quantum numbers for that purpose. Hence, in perturbation theory it is sufficient to consider the free elementary fields.

Applying the BRST transformation on asymptotic fields therefore will identify whether a state is physical or not. And since asymptotically all interactions ceases, it is therefore sufficient to consider the BRST transformations at vanishing gauge coupling.

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

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

for the full list of asymptotically BRST transformed fields. Unsurprisingly, these are exactly the BRST transformations of the free fields, and especially would be the same as in Maxwell theory.

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 so-called (Kugo-Ojima) 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 .

This result is in complete agreement with the general construction for the Maxwell theory in section 2.6.1. The difference between Yang-Mills theory and Maxwell theory is that the statement in Maxwell theory also holds in the non-asymptotic case. Of course,

all of this is only true perturbatively. Beyond perturbation theory the situation is more involved, see section 6.3.4.

The introduction of 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 matter degrees of freedom are perturbatively physical. This can be directly seen as their gauge, and consequently BRST, transformation is

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

and hence its free-field ($g = 0$) result is $s\psi_i^a = 0$. Similar as for the gauge bosons this changes non-perturbatively.

4.6 General anti-field approach

The previous construction can be generalized in a way which is both illuminating the underlying structures more as well as is useful in many practical calculations, especially in more complicated theories and/or other gauges. This is the so-called anti-field formalism.

4.6.1 BRST construction revisited

In section 4.5.1 it was observed that the complete gauge-fixing part could be rewritten as a BRST transformation, (4.40). It is possible to generalize this observation, and to show that every gauge-fixing prescription can be rewritten in terms of a BRST-exact term. This could be alternatively phrased as the trivial statement that it is possible to specify a gauge condition by the part of the gauge symmetry it does not fix.

To do so, consider a gauge theory of fields φ^i described by a classical action $S_0(\varphi^i)$. The equations of motion constrain the fields classically to certain configurations, denoted by Σ . Now, let the action be invariant under (infinitesimal) gauge transformations

$$\delta_\varepsilon\varphi^i = R_\alpha^i\varepsilon^\alpha. \quad (4.45)$$

where R can contain the fields and derivatives or the fields themselves, as has been the case so far. However, it is allowed that the gauge transformation functions ε could also be Grassmann-type, which occurs, e. g., in supergravity theories. But here this will only require to keep track of exchanges of fields, introducing an occasional power of -1 .

Assume the theory to be (on-shell) reducible, with no reducibility on the reducibility functions. In such a case there are relations among the gauge parameters but no relations

among the relations, i. e. there exists statements like (2.42), but there are no further relations involving the ϵ . In general, such relations take the form¹⁰

$$Z_{\Delta}^{\alpha} R_{\alpha}^i = C_{\Delta}^{ij} \frac{\delta S_0}{\delta \varphi^j},$$

with some matrices Z and C , and the on-shell relations appear due to the Euler-Lagrange equation on the right-hand side.

	Pure ghost number	Anti-field number	Total ghost number
φ^i	0	0	0
c^{α}	1	0	1
c^{Δ}	2	0	2
φ_i^*	0	1	-1
c_{α}^*	0	2	-2
c_{Δ}^*	0	3	-3

Table 4.1: Quantum numbers for the different field types. The total ghost number is defined as pure ghost number minus anti-field number.

Now, rather than to try to incorporate this into the path integral, and localize the results using ghost fields, this will be done in reverse order. I. e., get the ghost fields first, and show how they can be used to implement the gauge condition. For each commuting (anticommuting) gauge parameter ϵ^{α} fermionic (bosonic) ghost variable c^{α} are introduced. Also introduce ghosts of ghosts c^{Δ} , one for each (independent) reducibility identity of the theory. The set of original fields, ghosts and ghosts of ghosts are collectively denoted as Φ^A . In addition, the configuration space is doubled by adding conjugate fields with respect to each of the Φ s by introducing so-called anti-fields Φ_A^* . They are postulated to have opposite (Grassmann) parity. Gradings, i. e. charges under the various counting symmetries, are assigned to the various fields as displayed in table 4.1. Note that neither antighost fields nor Nakanishi-Lautrup fields appear at this stage. They will be added later.

For the following it is useful to define the antibracket for functionals of fields $F(\Phi^A, \Phi_A^*)$,

¹⁰Note that here, and in most of the expressions in this section, δ is a total derivative. Thus $\frac{\delta S}{\delta \phi} = \frac{\delta' S}{\delta' \phi} - \partial \frac{\delta' S}{\delta' \partial \phi} + \dots$ where the δ' are partial derivatives. Thus, the sum terminates for field theories usually after the second term, and yields the Euler-Lagrange equations, and thus the equations of motion. The minus sign appears from the partial integration when removing the variational part of the field, just like in classical mechanics, and leads to alternating signs in each term.

$G(\Phi^A, \Phi_A^*)$ as

$$(F, G) = \frac{\delta^R F}{\delta \Phi^A} \frac{\delta^L G}{\delta \Phi_A^*} - \frac{\delta^R F}{\delta \Phi_A^*} \frac{\delta^L G}{\delta \Phi^A}.$$

It is hence an extension of the usual Poisson bracket, but now involving fields and the corresponding anti-fields instead of conjugated variables. Here L (respectively R) refers to the standard left (respectively right) derivative. These are related as

$$\frac{\delta^R F}{\delta(\text{field})} \equiv (-1)^{\varepsilon_{\text{field}}(\varepsilon_F+1)} \frac{\delta^L F}{\delta(\text{field})},$$

where ε denotes the Grassmann-parity.

The antibracket has the following properties:

- $(F, G) = -(-1)^{(\varepsilon_F+1)(\varepsilon_G+1)}(G, F)$, where $\varepsilon_F = 0$ (1) for F bosonic (fermionic); the anti-bracket is symmetric if both F and G are bosonic, and antisymmetric otherwise.
- Jacobi identity, $(-1)^{(\varepsilon_F+1)(\varepsilon_H+1)}(F, (G, H)) + \text{cyclic permutations} = 0$
- $(FG, H) = F(G, H) + (F, H)G (-1)^{\varepsilon_G(\varepsilon_H+1)}$; $(F, GH) = (F, G)H + G(F, H) (-1)^{\varepsilon_G(\varepsilon_F+1)}$
- The ghost number of (F, G) is the sum of the ghost numbers of F and G

which can all be shown by straightforward explicit calculations.

Define the BRST transformation s of any functional $F(\Phi^A, \Phi_A^*)$ in terms of the antibracket,

$$sF = (S, F).$$

The generating function $S = S(\Phi^A, \Phi_A^*)$ of the BRST transformation is sometimes called the generalized action. The reason for this will become evident later.

The decisive demand on the BRST transformation will now be that it is nilpotent of order two, which takes the form

$$(S, S) = 0. \quad (4.46)$$

This equation is called the (classical) master equation. It will be the key ingredient to find the actual form of S , and thus the action of s .

The generalized action, and thus both the BRST transformation and the solution to the classical master equation, can be obtained constructively. It can be shown to be unique up to canonical transformations. This construction yields a, not necessarily finite sum, and is given by

$$S = S_0 + S_1 + S_2 + \dots \quad (4.47)$$

$$S_0 \equiv \text{classical action}, \quad S_1 = \varphi_i^* R_\alpha^i c^\alpha, \quad S_2 = c_\alpha^* Z_\Delta^\alpha c^\Delta + \dots \quad (4.48)$$

The proof of this statement (including the reducible case) is involved, and will be skipped, but yields all higher terms as well. It basically proceeds by inserting the expansion (4.47) into (4.46), and then solve the system iteratively. It requires a starting point. Evidently, the equations of motions imply that $(S_0, S_0) = 0$. The expression for (S_0, S_1) is then solved by utilizing (4.45), and so on. While in this way it is straightforward to see that (4.47) with (4.48) solves (4.46), the challenging part is showing the uniqueness of the solution. In addition, at the perturbative level S is under quite mild assumptions local in the fields and their derivatives.

The solution S of the master equation is key to the BRST-antifield formalism. It can be written down explicitly for the Yang-Mills theory and the Abelian antisymmetric model introduced earlier. In the Yang-Mills case it is given by

$$S = -\frac{1}{4} \int d^n x F_{\mu\nu}^a F_a^{\mu\nu} + \int d^n x A^{*\mu}_a D_\mu c^a + \frac{1}{2} \int d^n x c_a^* f^a_{bc} c^b c^c \quad (4.49)$$

and for the antisymmetric Abelian tensor by

$$S = -\frac{1}{12} \int d^n x H_{\mu\nu\rho} H^{\mu\nu\rho} + \int d^n x B^{*\mu\nu} (\partial_\mu c_\nu - \partial_\nu c_\mu) + \int d^n x c^{*\mu} \partial_\mu c$$

In those cases, the solution S of the master equation is linear in the antifields. For gauge systems with an ‘‘open algebra’’, i. e. for which the gauge transformations closes only on-shell, or for on-shell reducible gauge theories, the solution of the master equation is more complicated. It contains terms that are non-linear in the antifields. These terms are essential for getting the correct gauge fixed action below.

4.6.2 Observables

Before doing so, already at the current stage it is possible to generalize the statement (4.44),

$$H_p(s) \simeq \text{Observables}, \quad (4.50)$$

which depends on the concrete s . Again, a full proof of (4.50) is far too lengthy. Instead, only the key ingredients that underlie it will be given.

To start out, note that

$$s\Phi^A = (S, \Phi^A) = -\frac{\delta^R S}{\delta\Phi_A^*} = -\sum_k \frac{\delta^R S_k}{\delta\Phi_A^*} \quad (4.51)$$

$$s\Phi_A^* = (S, \Phi_A^*) = \frac{\delta^R S}{\delta\Phi^A} = \sum_k \frac{\delta^R S_k}{\delta\Phi^A}. \quad (4.52)$$

holds. Note that the ghost number of S is 0, the ghost number of the BRST transformation is 1 as well as the one of the antibracket, so that the gradings of both sides of the equation $sF = (S, F)$ match.

As a next step, the BRST transformations of all the variables is expanded according to the antifield number. This leads to

$$s = \Delta + \gamma + \sum_{i>0} s_i. \quad (4.53)$$

The first term Δ has antifield number -1 and is called the Koszul-Tate differential, the second term γ has antifield number 0 and is called the longitudinal differential, and the next terms s_i have antifield number i . Although the expansion stops at $\Delta + \gamma$ for Yang-Mills theory, see (4.49), higher order terms are present for gauge theories with an open algebra, or on-shell reducible theories. Note that

$$0 = s^2 = \Delta^2 + \{\Delta, \gamma\} + (\gamma^2 + \{\Delta, s_1\}) + \dots, \quad (4.54)$$

and in this equation each term has to vanish separately, as each term is of different antifield number.

Explicitly, the Koszul-Tate differential satisfies $\Delta\varphi^i = 0$, as there is no field operator of anti-field number -1 and $\Delta\varphi_i^* = \delta S_0/\delta\varphi^i = 0$ by virtue of the equations of motion. For the longitudinal differential it follows in the same way that¹¹

$$\gamma\varphi^i = R_\alpha^i c^\alpha. \quad (4.55)$$

Now, let A be a BRST-closed function(al), $sA = 0$. As seen already in section 4.5.2.3 the equivalence class

$$A \sim A + sB. \quad (4.56)$$

plays an important role. Since dealing with observables, the only relevant operators are of ghost number 0, thus A needs to have total ghost number zero and B total ghost number -1 . The latter can only be satisfied, if B contains at least one anti-field. Expanding A and B in antifield number yields

$$\begin{aligned} A &= \sum_{k \geq 0} A_k = A_0 + A_1 + A_2 + \dots \\ B &= \sum_{k \geq 1} B_k = B_1 + B_2 + B_3 + \dots \end{aligned}$$

¹¹In the following R will be assumed to be a local operator. However, in general it can contain also integrations, i. e.

$$\gamma\varphi^i(x) = \int d^d y R_\alpha^i(y, x) c^\alpha(y),$$

which would require somewhat more lengthy expressions throughout, though would change little.

Acting with s on A using the expansion (4.53) then gives

$$(\Delta + \gamma + \dots)(A_0 + A_1 + \dots) = (\gamma A_0 + \Delta A_1) + \dots,$$

where the term in parentheses collects all antifield number zero contributions. The condition $sA = 0$ implies that this term must vanish on its own and thus $\gamma A_0 = -\Delta A_1$. Furthermore,

$$A + sB = A + (\Delta + \gamma + \dots)B = (A_0 + \Delta B_1) + \dots$$

where the last term in parentheses is again the antifield zero contribution. Using (4.52),

$$\Delta B_1 = \frac{\delta B_1}{\delta \Phi^*} \frac{\delta S_0}{\delta \Phi^i},$$

implies that the second term of the antifield zero contribution in $A + sB$ vanishes when the equations of motions are fulfilled, i.e. on-shell. A similar property holds for ΔA_1 . There is therefore a clear connection of Δ to the dynamics and the equations of motions.

From (4.55) and the fact that γ is a derivative, this yields¹²

$$\gamma A_0 = \frac{\delta A_0}{\delta \varphi^i} R_\alpha^i c^\alpha.$$

As this is a gauge transformation (with gauge parameters replaced by the ghosts), γA_0 vanishes if A_0 is gauge-invariant. Thus, the longitudinal differential is associated with gauge transformations. A necessary condition for A to be BRST-closed is thus that its first term A_0 be gauge-invariant on-shell. And furthermore, two such A_0 s are equivalent when they coincide on-shell.

It turns out that the condition on A_0 is also sufficient for A to be BRST-closed, in the sense that given an A_0 that is gauge-invariant on-shell, it can be completed by terms $A_1, A_2 \dots$ of higher antifield number such that $sA = 0$.

To summarize: the term that determines the cohomological class of a BRST cocycle is the first term A_0 . This term must be an observable, in that it must be gauge invariant on-shell. It can therefore be concluded that $H_p(s)$ captures indeed the concept of observables. The differential Δ reduces therefore the on-shell functions to the set of all gauge-invariant functions.

4.6.3 Application as gauge-fixing to the path integral

Consider first the Yang-Mills case. To perform actual path-integral calculations, it is necessary to gauge-fix the theory, as discussed in section 4.4. In comparison to section 4.4

¹² A_0 can only depend on ϕ_i , as it has pure ghost number zero and antifield number zero.

so far the antighost field and the Nakanishi-Lautrup fields are missing, while having all the additional anti-fields. While both approaches can be used in this case independently, it is very convenient to combine them. Thus, add as additional fields the antighost \bar{c}_a and Nakanishi-Lautrup fields b_a , together with their antifields. They transform still as $s\bar{c}_a \sim b_a$ and $sb_a = 0$. Take \bar{c}_a and b_a have ghost number -1 and 0, respectively. The corresponding antifields \bar{c}^{a*} and b^{a*} have thus ghost number 0 and -1, respectively. It can then again be shown that both fields are BRST exact. Hence the cohomology is not altered by the introduction of these new variables. In particular, the set of observables is not affected. The solution of the master equation with the new variables included reads, for Yang-Mills theory,

$$S = -\frac{1}{4} \int d^n x F_{\mu\nu}^a F_a^{\mu\nu} + \int d^n x A_a^{\mu*} D_\mu c^a + \frac{1}{2} \int d^n x c_a^* f_{bc}^a c^b c^c - i \int d^n x \bar{c}^{*a} b_a. \quad (4.57)$$

The last term is called the non-minimal part.

From this generalized action the gauge-fixed action S_ψ will be constructed. It will be shown later that the generating functional

$$\mathcal{Z} = \int \mathcal{D}\Phi^A \exp\left(\frac{i}{\hbar} S_\psi[\Phi^A]\right), \quad (4.58)$$

does then not depend on the choice of the functional ψ . Here, ψ is called the (yet unspecified) gauge-fixing fermion, and has Grassmann-parity 1 (hence its name) and ghost number -1. In (4.58), the notation

$$\begin{aligned} \Phi^A &= (A_\mu^A, c^A, \bar{c}^A, b_A) \\ \Phi_A^* &= (A_\mu^{A*}, c^{A*}, \bar{c}^{A*}, b^{A*}), \end{aligned}$$

has been used.

The gauge-fixed action $S_\psi[\Phi^A]$ is now defined as

$$S_\psi[\Phi^A] = S \left[\Phi^A, \Phi_A^* = \frac{\delta\psi}{\delta\Phi^A} \right],$$

i. e. it is the generalized action with the anti-fields fixed by the derivatives of the gauge-fixing fermions with respect to the fields. Before turning to the proof that this is correct, it is helpful to illustrate (4.58) by showing how the gauge-fixing fermion ψ can be used to reproduce the familiar expressions for the path integral of Yang-Mills theory, (4.35).

A possible choice, which leads to non-degenerate propagators for all fields and ghosts, is given by

$$\psi = i \int d^n x \bar{c}^a \left(D^a + \frac{\alpha}{2} b^a \right), \quad (4.59)$$

where D^a is the gauge condition, e.g. $D^a = \partial^\mu A_\mu^a$ for covariant gauges and α is the gauge parameter. This yields

$$\begin{aligned}\bar{c}^{a*} &= \frac{\delta\psi}{\delta\bar{c}_a} = i\left(D^a + \frac{\alpha}{2}b^a\right) \\ A_{a\mu}^* &= \frac{\delta\psi}{\delta A_\mu^a} = -i\partial_\mu\bar{c}_a \\ b^{a*} &= \frac{\delta\psi}{\delta b_a} = i\frac{\alpha}{2}\bar{c}^a \\ c_a^* &= \frac{\delta\psi}{\delta c^a} = 0.\end{aligned}$$

This leads to the familiar gauge-fixed Yang-Mills action

$$S_\psi[A_\mu^a, c^a, \bar{c}_a, b_a] = \int d^n x \left(-\frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} - i\partial^\mu\bar{c}_a D_\mu c^a + \left(D^a + \frac{\alpha}{2}b^a\right)b_a \right),$$

as was obtained by the Faddeev-Popov procedure in section 4.4. The conventional Landau gauge is recovered by taking the limit $\alpha \rightarrow 0$. As the resulting path integral can be written as

$$\int \mathcal{D}[A_\mu^a \bar{c}^a c^a] \delta(\partial^\mu A_\mu^a) e^{\left(\frac{i}{\hbar} S_\psi[A_\mu^a, \bar{c}^a, c^a]\right)},$$

the transversality of the gauge boson is directly implemented. Landau gauge is thus called a strict gauge. An example of a non-strict gauge is Feynman gauge, $\alpha = 1$, in which case the equations of motion do not imply $D^a = 0$ but yield instead $b^a \sim D^a$.

The choice of the gauge-fixing fermion is arbitrary. E. g. to ψ a term $\sim \bar{c}_a \bar{c}_b c^c$ can be added, which yields quartic ghost couplings. Quartic ghost renormalizations may even be needed without such explicit terms, e.g. when using $D^a = \partial^\mu A_\mu^a + d_{bc}^a A_\mu^b A^{\mu c}$, where d_{bc}^a is a symmetric tensor in color space.

By appropriately choosing the gauge fixing fermion, the path integral can be reduced to an expression that involves only the (perturbatively) physical (transverse) degrees of freedom and which is manifestly unitary in the (perturbative) physical subspace. Independence on the choice of ψ (still to be proved) guarantees then that the expression (4.58) is correct.

As a final point, note that in order for (4.58) to be indeed independent on the choice of ψ , it is necessary that the measure is BRST invariant. This can be investigated using the operator Ω ,

$$\Omega = (-1)^{\epsilon_A} \frac{\delta^L}{\delta\Phi^A} \frac{\delta^L}{\delta\Phi_A^*}. \quad (4.60)$$

The BRST transformation can be written as

$$\Phi^A \rightarrow \Phi^{A'} = \Phi^A + (\mu S, \Phi^A) = \Phi^A - (\Phi^A, S)\mu = \Phi^A - \frac{\delta^L S}{\delta\Phi_A^*} \mu,$$

where μ is a constant, anti-commuting parameter. The Jacobian¹³ of this transformation is given by

$$J_{AB} = \frac{\delta^L \Phi^{A'}}{\delta \Phi^B} = \delta^{AB} - \frac{\delta^L}{\delta \Phi^B} \frac{\delta^L S}{\delta \Phi_A^*} \mu.$$

Therefore the measure transforms as

$$\mathcal{D}\Phi^A \rightarrow \det J \mathcal{D}\Phi^{A'}.$$

For an infinitesimal transformation, this can be approximated by the super-trace, i. e. a trace with negative signs for fermionic quantities,

$$\det J \approx 1 + \text{str} \left(-\frac{\delta^L \delta^L S}{\delta \Phi_B \delta \Phi_{A^*}} \mu \right) = 1 - (-1)^{\epsilon_A} \frac{\delta^L \delta^L S}{\delta \Phi_A \delta \Phi_{A^*}} \mu = 1 + \Omega S.$$

It follows that the measure is BRST-invariant if, and only if, $\Omega S = 0$. The property $\Omega S = 0$ can be shown by explicit calculation for pure Yang-Mills theory. The more general case will be treated later.

4.6.4 Beyond Yang-Mills theory

The results of section 4.6.3 generalizes straightforwardly to gauge theories other than Yang-Mills theory. The solution of the master equation S is of the form

$$S = S_0 + \phi_i^* R_\alpha^i c^\alpha + \dots,$$

where S_0 is the classical action. The second term is uniquely determined by the gauge transformation (4.45), and all further terms depend on the specific theory. While the expansion of the solution of the master equation stops at antifield number one in the Yang-Mills case, higher order terms appear in the case of open gauge systems. It is again often convenient to extend to the non-minimal sector by introducing \bar{c}_α and b_α and the corresponding antifields in a similar manner to what has been done in the case of Yang-Mills theory.

Assuming again $\Omega S = 0$, the quantized theory follows from a path integral

$$\mathcal{Z} = \int \mathcal{D}\Phi^A \exp \left(\frac{i}{\hbar} S_\psi[\Phi^A] \right).$$

S_ψ is the solution S of the master equation

$$(S, S) = 0, \tag{4.61}$$

¹³As the Jacobian involves commuting and anti-commuting fields, the Jacobian ‘‘determinant’’ is actually a so-called super-determinant sdet , where additional minus signs arise from anticommuting fermionic quantities.

in which the antifields have been eliminated by use of the gauge-fixing fermion ψ as before,

$$S_\psi[\Phi^A] = S_\psi \left[\Phi^A, \Phi_A^* = \frac{\delta\psi}{\delta\Phi^A} \right]. \quad (4.62)$$

The gauge-fixing fermion has again odd Grassmann-parity and ghost number -1. It is in perturbation theory given by a local expression

$$\psi = \int d^n x \chi(\Phi^A, \partial_\mu \Phi^A, \dots, \partial_{\mu_1} \dots \partial_{\mu_k} \Phi^A).$$

For theories with an open algebra, the terms quadratic in the antifields will lead to quartic (or higher) ghost-antighost vertices in the gauge-fixed action. While these terms are a gauge-dependent option in the Yang-Mills case, they have an unavoidable character (in relativistic gauges) for open gauge algebras. These terms, which follow directly from the general construction of the gauge-fixed action, cannot be obtained through the exponentiation of a determinant, since this procedure always produces an expression which is quadratic in the ghosts.

A useful concept is that of gauge-fixed BRST transformation, which is what s becomes after gauge-fixing. It is denoted by s_ψ and defined as

$$s_\psi \Phi^A = (s\Phi^A)|_{\Phi_A^* = \delta\psi/\delta\Phi^A}.$$

Note that $s_\psi^2 = 0$ is in general only valid on the gauge-fixed shell, i.e. for field configurations satisfying $\delta S_\psi / \delta \Phi^A = 0$.

If S is linear in the antifields, i.e. if the gauge algebra closes off-shell, $s_\psi \Phi^A = s\Phi^A$. This follows directly from

$$s\Phi^A = (S, \Phi^A) = -\frac{\delta^R S}{\delta \Phi_A^*}, \quad (4.63)$$

which is independent of the antifields, if S depends only linearly on the antifields. In that case $s_\psi^2 = 0$ even off-shell provided all variables are kept. This is the case in Yang-Mills theory¹⁴.

If S is linear in the antifields, one may in fact write it as

$$S = S_0 - (s\Phi^A)\Phi_A^*,$$

by virtue of (4.63). The gauge-fixed version is then

$$S_\psi = S_0 - (s_\psi \Phi^A) \frac{\delta\psi}{\delta\Phi^A} = S_0 - s_\psi \psi. \quad (4.64)$$

¹⁴In Yang-Mills theory, often the auxiliary fields b_a are eliminated by means of their own equations of motion. Then off-shell nilpotency on the antighosts is lost, for which $s\bar{c}_a \sim b_a$, even though $s_\psi^2 = 0$ is true off-shell beforehand.

Therefore $s_\psi S_\psi = 0$, as the first term is BRST invariant, and the second term is annihilated by s_ψ by virtue of $s_\psi^2 = 0$, which holds off-shell when S is linear in the antifields. The property $s_\psi S_\psi = 0$ is actually quite general and holds even when S is not linear in the antifields. It can be proved directly as follows,

$$s_\psi S_\psi = (s_\psi \Phi^A) \frac{\delta^L S_\psi}{\delta \Phi^A} = -\frac{\delta^R S}{\delta \Phi_A^*} \frac{\delta^L S_\psi}{\delta \Phi^A}.$$

The left-derivative in this expression is a total derivative, as S_ψ depends on Φ_A directly and through the gauge-fixing fermion. Using the chain rule, this yields by virtue of (4.62)

$$-\frac{\delta^R S}{\delta \Phi_A^*} \left(\frac{\delta S}{\delta \Phi^A} + \frac{\delta^2 \psi}{\delta \Phi^A \delta \Phi^B} \frac{\delta S}{\delta \Phi_B^*} \right) = 0.$$

The second term vanishes because the product of the functional derivatives of S have a symmetry in (A, B) opposite to that of the second functional derivative of ψ . The first term vanishes by the master equation, thus proving the claim.

4.6.5 Quantum Master Equation

In order to prove perturbative gauge independence of the expressions given above, it is necessary to discuss two important features of the path integral.

First, assume that a theory of fields χ^α is given, governed by the action $S[\chi^\alpha]$, with no gauge invariance. This could be the gauge-fixed action. Expectation values are, after proper normalization, calculated as

$$\langle F \rangle = \int \mathcal{D}\chi F \exp\left(\frac{i}{\hbar} S[\chi]\right), \quad (4.65)$$

where \hbar has been reinstated temporarily. The so-called Dyson-Schwinger equations can be directly derived from the vanishing of the path integral of a total derivative¹⁵

$$\int \mathcal{D}\chi \frac{\delta}{\delta \chi^\alpha} \left(F e^{\frac{i}{\hbar} S} \right) = 0.$$

This leads to

$$\left\langle \frac{\delta F}{\delta \chi^\alpha} + \frac{i}{\hbar} F \frac{\delta S}{\delta \chi^\alpha} \right\rangle = 0,$$

which is equivalent to

$$\left\langle F \frac{\delta S}{\delta \chi^\alpha} \right\rangle = i\hbar \left\langle \frac{\delta F}{\delta \chi^\alpha} \right\rangle. \quad (4.66)$$

¹⁵Which is itself a consequence of translation invariance of the measure, and which is part of the definition of the path integral.

This expression contains on the left-hand side the expectation values of the classical equations of motions. In the classical limit $\hbar \rightarrow 0$, the right-hand side vanishes and the classical equations of motion hold.

Second, if χ^α is changed under a transformation, $\chi^\alpha \rightarrow \chi^\alpha + \epsilon^\alpha$, where ϵ^α depends on the fields, the expectation value (4.65) is in general not invariant. Furthermore, invariance of the classical action is not sufficient to guarantee that the path integral is invariant. Also the measure needs to be invariant.

Especially, this raises the question whether the measure is BRST invariant, which can be tested with the methods outlined in section 4.6.3. If this test failed, it is necessary to modify (4.65) to reestablish a meaningful quantum theory. This may be possible, in some cases, by taking a different measure¹⁶. This requirement is crucial, since BRST symmetry guarantees gauge-independence of the results.

Assume that there are non-trivial measure terms. They can be exponentiated¹⁷ into the action. Since there is an overall $(1/\hbar)$ in front of S , the measure terms appear as quantum corrections to S . So, this replaces effectively the classical action by a “quantum action”

$$W = S + \hbar M_1 + \hbar^2 M_2 + \dots, \quad (4.67)$$

where the functionals M_i stem from non-trivial measure factors. The theorem discussed below states that quantum averages are gauge-independent if the master equation (4.61) is replaced by the “quantum master equation”

$$\frac{1}{2}(W, W) = i\hbar\Omega W, \quad (4.68)$$

where Ω is defined in (4.60). Note that if $\Omega S = 0$, the Jacobian is unity for the BRST transformation and W might then be taken equal to S . The quantum master equation then reduces to the classical master equation (4.61), which is solved by S . However, while there is always a solution to the classical master equation, the solution to the quantum master equation might get obstructed, creating a gauge anomaly to be discussed in chapter 7. For the moment it is assumed that this is not the case.

This leads to the general rules for computing expectation values of observables (including of 1). These are the quantum averages, weighted by $\exp(iW/\hbar)$, of the BRST observables corrected by the addition of appropriate \hbar (and possibly also higher) order

¹⁶Usually, the measure is taken to be the naive measure. However, it is part of the definition of a quantum theory, and may actually be chosen differently. This leads potentially to different results for observables, and ultimately experiment would decide which one describes a given system.

¹⁷This requires the absence of zero modes in a suitable sense. It is exactly here where (additional) non-perturbative problems may arise.

terms. Namely, consider a classical observable A_0 . Construct its (in fact, one of its) BRST-invariant extension $A = A_0 + \text{ghost terms}$, so that $(S, A) = 0$. The BRST cocycle A has to be augmented as

$$A \rightarrow \alpha = A_0 + \hbar B_1 + \hbar^2 B_2 + \dots$$

where the terms of order \hbar and higher must be such that $\sigma\alpha = 0$, where σ is defined as

$$\sigma\alpha = (W, \alpha) - i\hbar\Omega\alpha,$$

with W the solution of the quantum master equation (4.68). Note that the B -terms supersede the ghost terms needed classically to fulfill $(S, A) = 0$. The operator σ is then the quantum generalization of the BRST variation s . The ψ -independent expectation value $\langle A_0 \rangle$ of the observable $A_0[\phi^i]$ is computed from α as

$$\langle A_0 \rangle = \int \mathcal{D}\Phi^A \alpha \left(\Phi^A, \Phi_A^* = \frac{\delta\psi}{\delta\Phi^A} \right) \exp \left(\frac{i}{\hbar} W \left[\Phi^A, \Phi_A^* = \frac{\delta\psi}{\delta\Phi^A} \right] \right). \quad (4.69)$$

The claim is that this expectation value does not depend on the choice of the gauge-fixing fermion, i. e.

$$\begin{aligned} \psi' &= \psi + \delta\psi \equiv \psi + \mu \\ \langle A_0 \rangle_{\psi'} &= \langle A_0 \rangle_{\psi}, \end{aligned} \quad (4.70)$$

where μ is an arbitrary modification of ψ .

To prove the claim, denote the argument of the path integral (4.69) by V for convenience in the following. It can be shown straightforwardly that

$$\Omega V = 0 \Leftrightarrow \sigma\alpha = 0.$$

Now, perform the variation of the gauge fixing functional (4.70). The variation of the quantum average (4.69) is equal to

$$\int \mathcal{D}\Phi \frac{\delta^L \mu}{\delta\Phi^A} \frac{\delta^L V}{\delta\Phi_A^*}. \quad (4.71)$$

To evaluate this expression, note that

$$\frac{\delta^L}{\delta\Phi^A} \left(\mu \frac{\delta^L V}{\delta\Phi_A^*} \right) = \frac{\delta^L \mu}{\delta\Phi^A} \frac{\delta^L V}{\delta\Phi_A^*} + (-1)^{\epsilon_A} \frac{\delta^L \delta^L V}{\delta\Phi^A \delta\Phi_A^*}.$$

The derivatives are total ones. Denoting partial derivatives by a prime ', the last term can be rewritten as

$$\frac{\delta^L \delta^L V}{\delta\Phi^A \delta\Phi_A^*} = \frac{\delta'^L \delta^L V}{\delta'\Phi^A \delta\Phi_A^*} + \frac{\delta^2 \psi}{\delta\Phi^A \delta\Phi^B} \frac{\delta^L \delta^L V}{\delta\Phi_B^* \delta\Phi_A^*} = \Omega V = 0.$$

As in the case of equation (4.65), the second term vanishes by parity arguments. Thus the integral (4.71) can be rewritten as a total derivative in field space, which vanishes in view of translation invariance of the standard measure. Therefore the path integral does not get modified if one changes the gauge-fixing fermion, as claimed.

Given A_0 , its BRST extension is determined up to a BRST exact term sB , see (4.56). This ambiguity can be extended to higher orders in \hbar as

$$\alpha \rightarrow \alpha + \sigma\beta.$$

It can be shown straightforwardly that

$$(\sigma\beta) \exp\left(\frac{i}{\hbar}W\right) \sim \Omega\left(\beta \exp\left(\frac{i}{\hbar}W\right)\right).$$

As the right-hand side is a total derivative by its definition, (4.60), the path-integral over the left-hand side vanishes

$$\int \mathcal{D}\Phi \sigma\beta \exp\left(\frac{i}{\hbar}W\right) = 0.$$

Therefore adding any element in the image of σ does not alter the quantum averages. The path integral associates a unique answer to a given cohomological class of σ , i.e., does not depend on the choice of representative.

Note that the ambiguity in α , given A_0 , is more than just adding a σ -trivial term to α . At each order in \hbar a new non-trivial observable can be added since this does not modify the classical limit. This addition is relevant, in the sense that it changes the expectation value by terms of order \hbar or higher. A similar ambiguity exists for the quantum measure (i.e., the M_1, M_2 etc. terms in W). These terms do not spoil BRST invariance and must be determined by other criteria, e. g. experiment.

4.7 Ward-Takahashi identities 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 identities 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.

4.7.1 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 change

$$\phi \rightarrow \phi' = \phi + \delta\phi = \phi + \epsilon f(\phi, x), \quad (4.72)$$

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 on the action itself, which yields

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

to first order in ϵ . The second is the measure. The shift (4.72) 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

$$0 = \frac{1}{\epsilon} \delta Z = \int \mathcal{D}\phi \left(\frac{\delta f}{\delta\phi} + i \left(\frac{\delta S}{\delta\phi} + J \right) f \right) e^{iS+i \int d^d x J\phi}. \quad (4.73)$$

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

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

In general, there will not only be one field involved, but many fields, numerated by a field index i . In this case, expression (4.73) takes the form

$$0 = \int \mathcal{D}\phi_i \left(\frac{\delta f_k}{\delta \phi_k} + i \left(\frac{\delta S}{\delta \phi_k} + J_k \right) f_k \right) e^{iS + i \int d^d x J \phi}, \quad (4.74)$$

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, sometimes also called rigid identities given that global symmetries are sometimes called rigid symmetries,

$$\begin{aligned} & \left\langle T \Pi_{l=1}^n \phi_{i_l}(x_l) \frac{\delta f_k}{\delta \phi_k(y)} \right\rangle + i \left\langle T \Pi_{l=1}^n \phi_{i_l}(x_l) \frac{\delta S}{\delta \phi_k(y)} f_k \right\rangle \\ & + \sum_{m=1}^n \left\langle \Pi_{l=1}^{m-1} \phi_{i_l}(x_l) f_{i_m} \Pi_{r=m+1}^n \phi_{i_r}(x_r) \right\rangle = 0. \end{aligned} \quad (4.75)$$

This also eliminates the term $\langle f \rangle$. To obtain practical cases requires to insert an action with a certain invariance.

Take as an example the action for the 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 ϕ^\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. This is not necessarily the case, and they cause so-called anomalies to be discussed in chapter 7.

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 (4.73) vanishes. Hence, only the third term remains. As the invariance of both the action and the measure is the usually the case for symmetries. However, because only (4.74) needs to vanish, they can be individually still non-zero, if their sum is zero. Such a case is called anomaly cancellation, see chapter 7, if a contribution of the measure is cancelled by the other. If actually (4.74) is non-zero, the symmetry is explicitly broken. The Ward Takahashi identities will in this case not be fulfilled, but will yield some non-zero results. In this case they are called broken or deformed. They may still be useful, technically.

In the present case of the linear σ -model, the only remaining third term can be conveniently written as

$$0 = \delta \langle \Pi_{l=1}^n \phi_{i_l} \rangle, \quad (4.76)$$

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 a scalar σ field and a pseudoscalar χ field, 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 (4.66), are once more obtained, as will be exploited below.

4.7.2 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 expectation value of any BRST variation vanishes, see section 4.5.2.3, it follows immediately that

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

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 $\langle \bar{u}^a(x) D^b [A_\mu^a, y] \rangle$, where D^b is the

gauge-fixing condition. This yields

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

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, $D^a = f_\mu^{ab}A_\mu^b$, though f may contain derivatives, though it will not contain integrals in the following. This yields

$$\begin{aligned}
0 &= -\frac{1}{\xi} f_{ac}^\mu f_{bd}^\nu \langle A_\mu^c A_\nu^d \rangle - \langle \bar{u}^a(x)(s f_\mu^{bc} A_\mu^c) \rangle \\
&= -\frac{1}{\xi} f_{ac}^\mu f_{bd}^\nu D_{\mu\nu}^{cd} - \langle \bar{u}^a(x) f_\mu^{bc} D_{cd}^\mu u^d \rangle = -\frac{1}{\xi} f_{ac}^\mu f_{bd}^\nu D_{\mu\nu}^{cd} - \left\langle \bar{u}^a(x) \frac{\delta S}{\delta \bar{u}^c} \right\rangle, \quad (4.77)
\end{aligned}$$

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

To determine the second expression, the Dyson-Schwinger equations (4.66) can be used. It is useful to derive them here somewhat more explicitly than in section 4.6.5. Since the path integral is by definition translational-invariant, it follows that

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

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

$$0 = \left\langle \left(i\phi(y) \frac{\delta S}{\delta\phi(x)} + iJ(x)\phi(y) + \frac{\delta J(x)}{\delta J(y)} \right) \right\rangle \stackrel{J \rightarrow 0}{=} i \left\langle \phi \frac{\delta S}{\delta\phi} \right\rangle + \delta(x-y), \quad (4.78)$$

where the limit of $J \rightarrow 0$ has been taken in the last step¹⁸.

Thus, an expression like the second term in (4.77) 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

$$f_{ac}^\mu f_{bd}^\nu D_{\mu\nu}^{cd} = i\xi \delta^{ab} \delta(x-y),$$

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

$$p^\mu p^\nu D_{\mu\nu}^{ab}(p) = -i\xi \delta^{ab}. \quad (4.79)$$

¹⁸Beyond perturbation theory there are subtleties associated with this limit.

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.

While the BRST transformation has been used, the same statements are also found when starting from a localized version of (4.72). This is not surprising, as BRST transformations are basically gauge transformations extended to the ghost sector.

Chapter 5

Perturbative treatment

5.1 Perturbation theory

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

5.1.1 Cross-sections and decays

The primary quantities of interest at experiments are cross-sections and decay processes. The basic starting point is the same object as in non-gauge theories, 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

$$d\sigma = \frac{1}{4\sqrt{(p_1 p_2)^2 - M_1 M_2}} (2\pi)^4 \delta\left(p_1 + p_2 - \sum_i q_i\right) \frac{d^3 \vec{q}_1}{(2\pi)^3 E_{q_1}} \times \dots \times \frac{d^3 \vec{q}_n}{(2\pi)^3 E_{q_n}} |\mathcal{M}_{fi}|^2, \quad (5.1)$$

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 appreciable 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

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

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

$$\langle f|\mathcal{S}|i\rangle = \langle f|i\rangle + i(2\pi)^4 \delta(p_i - p_f) \mathcal{M}_{fi}, \quad (5.2)$$

where p_f and p_i are the total initial and final state momenta, and \mathcal{S} is the S-matrix, which is perturbatively¹ defined as the time-ordered product of the interaction Lagrangian as

$$\mathcal{S} = T e^{i \int d^d x \mathcal{L}_I}$$

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 state and the final state, respectively. E. g., for a two muon to two electron process, the expression is the correlation function

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

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.

¹Non-perturbatively one is stuck as formulating the process using correlation functions, and then evaluating the full expression for a given process. This leads also to interesting questions about how to define the initial state and the final state, some of which will be discussed in chapter 6.

5.1.2 General construction

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

$$\langle T\phi_1\dots\phi_n \rangle = \frac{\int \mathcal{D}\phi \phi_1\dots\phi_n e^{iS[\phi,J]}}{\int \mathcal{D}\phi e^{iS[\phi,J]}} \Big|_{J=0}. \quad (5.3)$$

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

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

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

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

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,

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

This yields

$$\begin{aligned}
\int D\phi e^{i \int d^d x (\phi(x)(\Omega - i\epsilon/2)\phi(x) + J(x)\phi(x))} &= \int D\phi e^{i \int d^d x (\phi(x)\Omega\phi(x))} e^{-\frac{i}{2} \int d^d x d^d y J(x)\Delta(x-y)J(y)} \\
&= Z_2[0] e^{-\frac{i}{2} \int d^d x d^d y J(x)\Delta(x-y)J(y)}.
\end{aligned} \tag{5.4}$$

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,

$$\begin{aligned} (-\partial^2 - M^2 + i\epsilon) \int d^d p e^{ip(x-y)} \Delta(p) &= i \int d^d p e^{ip(x-y)} \\ \int d^d p e^{ip(x-y)} (p^2 - M^2 + i\epsilon) \Delta(p) &= i \int d^d p e^{ip(x-y)} \end{aligned} \quad (5.5)$$

the form

$$\Delta(p) = \frac{i}{p^2 - M^2 + i\epsilon}, \quad (5.6)$$

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 (5.3) when taking the limit $J \rightarrow 0$ 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 identical particles in the initial state and the final state, essentially elastic scattering, is then

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

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 series²,

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

²It is this operation - expanding in a power series and swapping summation and (functional integration)

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

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

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.

- which is not valid beyond perturbation theory. The mechanism can be seen by an example,

$$\begin{aligned} \infty &> \frac{e^{\frac{1}{\lambda}}}{2\sqrt{\lambda}} K_{\frac{1}{4}} \left(\frac{1}{8\lambda} \right) = \int dx e^{-x^2 - \lambda x^4} = \int dx \sum_i \frac{(-1)^i \lambda^i x^{4i}}{i!} e^{-x^2} \neq \sum_i \frac{(-1)^i \lambda^i}{i!} \int dx x^{4i} \exp^{-x^2} \\ &= \sum_i \frac{(-1)^i \lambda^i (1 + (-1)^{4i})}{i! 2} \Gamma \left(2i + \frac{1}{2} \right) = \sum_i \frac{\sqrt{\pi} (-1)^i \lambda^i (4i - 1)!!}{i! 2^{\frac{2i+1}{2}}} \\ &= \sqrt{\pi} \left(1 - \frac{3\lambda}{2\sqrt{2}} + \frac{105\lambda^2}{8\sqrt{2}} - \frac{3465\lambda^3}{16\sqrt{2}} + \frac{675675\lambda^4}{128\sqrt{2}} + \dots + \mathcal{O} \left(\frac{(\lambda i)^i}{\sqrt{i}} \right) \right) = \infty, \end{aligned}$$

where K_n is the modified Bessel function of the second kind. The reason is the lack of absolute convergence of the series, which does not permit the exchange of integration and summation. Note that for $\lambda < 1$ the first few lines appear to converge, but this changes starting from terms of order $1/\lambda$. Conversely, this shows that the limit $\lambda \rightarrow 0$ is not smooth. The generalization of this is known as Haag's theorem, that for any non-interacting field theory the free theory and the interacting one are not unitarily equivalent, as will be encountered again in chapter 6. In fact, note that the correct result has an essential singularity in λ , making the absence of a smooth limit explicit.

The first derivative acts as

$$\begin{aligned} & \frac{\delta}{\delta J(x_4)} \int d^d x d^d y J(x) \Delta(x-y) J(y) \\ &= \int d^d x d^d y \delta^d(x-x_4) \Delta(x-y) J(y) + \dots = \int dy \Delta(x_4-y) J(y) + \dots, \end{aligned} \quad (5.7)$$

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 $\lambda^0 = 1$

$$A = - \sum_{P_{ijkl}} \Delta(x_i-x_j) \Delta(x_l-x_k), \quad (5.8)$$

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 (5.7). 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 (5.7). The second situation is different, and yields

$$\frac{\delta}{\delta J(z)} \int d^d y \Delta(z-y) J(y) = \int d^d y \Delta(z-y) \delta^d(y-z) = \Delta(z-z) \quad (5.9)$$

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

$$\lambda B = -i\lambda \int d^d x \Delta(x-x_1) \Delta(x-x_2) \Delta(x-x_3) \Delta(x-x_4) \quad (5.10)$$

$$\lambda C = -\frac{i\lambda}{2} \sum_{P_{ijkl}} \Delta(x_i-x_j) \int d^d x \Delta(x-x) \Delta(x-x_k) \Delta(x-x_l) \quad (5.11)$$

$$\lambda AD = -\frac{i\lambda}{8} \int d^d x \Delta(x-x) \Delta(x-x) \sum_{P_{ijkl}} \Delta(x_i-x_j) \Delta(x_k-x_l). \quad (5.12)$$

These four terms have simple interpretations, if each factor of Δ is considered to be a particle propagating along the connecting line of $x - y$. Then, the first term (5.8) 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 (5.10-5.12). The expression (5.10) 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 (5.11) can be seen as one particle just propagating, while the second particle has an interesting behavior: It emits at an intermediate point a particle, and 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 (5.8-5.12) 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 (5.4). 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

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

This term is easily identified as the prefactor appearing in (5.12). 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, (5.6). The total sum then becomes

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

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. In general, the necessary propagators and vertices can be derived from the classical (possibly gauge-fixed) action as

$$i\Gamma^{\phi_1^{a_1} \dots \phi_n^{a_n}}(x_1, \dots, x_n) = \left. \frac{i\delta^n S}{\delta \phi_1^{a_1}(x_1) \dots \delta \phi_n^{a_n}} \right|_{\phi_i=0},$$

which afterwards can be transformed to momentum space by a Fourier transformation. The conservation of momenta and other quantum numbers will then appear as suitable δ -functions, which are usually factored of and manually imposed afterwards. For the propagators, it is still necessary to invert the results.

Perturbative calculations in Yang-Mills theory with scalar and fermion fields in some representation 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 gauge bosons, ghosts, fermions, and scalars

$$D_{\mu\nu}^{ab} = \frac{-i\delta^{ab}}{k^2 + i\epsilon} \left(g_{\mu\nu} - (1 - \xi) \frac{k_\mu k_\nu}{k^2} \right) \quad (5.13)$$

$$D_G^{ab} = \frac{i\delta^{ab}}{p^2 + i\epsilon} \quad (5.14)$$

$$D_F^{ij} = \delta_{ij} \frac{i(p_\mu \gamma^\mu + m)}{p^2 - m^2 + i\epsilon} = \frac{i\delta_{ij}}{p_\mu \gamma^\mu - m + i\epsilon} \quad (5.15)$$

$$D_S^{ij} = \frac{i\delta^{ij}}{p^2 - m^2 + i\epsilon}, \quad (5.16)$$

respectively. There are in addition 3-point vertices of fermions and scalars and gauge bosons, three gauge bosons, and ghost and gauge bosons

$$\Gamma^{gq\bar{q}}(p, q, k)_\mu^{aij} = ig\gamma^\mu t_{ij}^a \quad (5.17)$$

$$\Gamma^{gs\bar{s}}(p, q, k)_\mu^{aij} = igt_{ij}^a (q + k)_\mu \quad (5.18)$$

$$\Gamma^{ggg}(p, q, k)_{\mu\nu\rho}^{abc} = gf^{abc}(g^{\mu\nu}(k - p)^\rho + g^{\nu\rho}(p - q)^\mu + g^{\rho\mu}(q - k)^\nu) \quad (5.19)$$

$$\Gamma^{\bar{u}gu}(p, q, k)_\mu^{abc} = gf^{abc}p^\mu, \quad (5.20)$$

respectively, and the four-gauge-boson vertex and the two-gauge boson-two-scalar vertex³

$$\Gamma^{gggg}(p, q, k, l)_{\mu\nu\rho\sigma}^{abcd} = -ig^2(f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho}) \quad (5.21)$$

$$+ f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma})) \quad (5.22)$$

$$\Gamma^{s\bar{s}gg}(p, q, k, l)_{\mu\nu}^{ijab} = ig^2 g_{\mu\nu} \delta_{ab} \delta_{ij} \quad (5.23)$$

and all momenta defined incoming. Note that because ghosts obey Fermi statistics closed ghost loops also receive a minus sign. If the Nakanishi-Lautrup field remains, further rules are appearing. Especially the propagators become matrix-like, as there are now two-point vertices changing the gauge field to an auxiliary field.

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

$$\Delta\Delta + \Delta\Delta' + \Pi, \quad (5.24)$$

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,

³If the scalar is in the adjoint or any other representation a more complex gauge structure emerges.

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

$$G(x_1, \dots, x_n) = G_c(x_1, \dots, x_n) + \sum G_c(x_i, \dots, x_j)G_c(x_j, \dots, x_k) + \sum G_c(x_i, \dots, x_j)G_c(x_k, \dots, x_l)G_c(x_m, \dots, x_n) + \dots \quad (5.25)$$

where the sums are over all possible ways to split the index set $\{x_i\}$ in two, three, ... subsets. Furthermore, every connected correlation function G_c , i. e. anyone which cannot be written as a product of two or more other connected correlation functions, is a series in the coupling constant. Thus, in the present case,

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

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 Δ'^2 have to be dropped, as they are of higher order in the perturbative expansion. This relation can be inverted to obtain the connected correlation functions from the other, but it is more interesting to calculate just the connected, and then calculate the complete one by the formula (5.25).

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 reinstated.

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

$$\Gamma_c(p, q) = -i\lambda \int \frac{d^d r}{(2\pi)^4} \frac{i}{r^2 - m^2} \\ \Gamma_c(p, q, k, l) = -i\lambda, \quad (5.26)$$

rather simple expressions indeed. Only missing is the explicit form of the contribution from Δ' . However, it requires renormalization, and will be discussed in section 5.2. These are also called proper or vertex correlation functions. In fact, for the four-point vertex function to leading order in λ , the entire result is given by (5.26).

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 to 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 n PI 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 $\hbar \rightarrow 0$. 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 to be discussed now.

5.2 Renormalization

So far, all calculations have been (essentially) 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 in QED even $N^{10}LO$ results are available⁴.

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 loses 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 conventional gauge theories could be theories of everything, since they do 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, gives a prescription how to redefine the theory such as to lose 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 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.

⁴Note 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.

5.2.1 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. This is also exactly the diagram appearing in the evaluation of Δ' in section 5.1 in (5.24). Hence, the result for this quantity will be obtained now.

Its value is

$$\Pi_\phi^\lambda = -\frac{\lambda}{2} \int \frac{d^4 p}{(2\pi)^4} \frac{1}{p^2 - M^2 + i\epsilon}, \quad (5.27)$$

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

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

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,

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

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

$$\Pi_\phi^\lambda = \frac{i\lambda}{4\pi^2} \left(\Lambda^2 \sqrt{1 + \frac{M^2}{\Lambda^2}} - M^2 \ln \left(\frac{\Lambda + \Lambda \sqrt{1 + \frac{M^2}{\Lambda^2}}}{M} \right) \right). \quad (5.28)$$

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 .

5.2.2 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

$$\frac{1}{D_{\phi\phi}} = p^2 - M^2 + \Pi_\phi^\lambda. \quad (5.29)$$

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 (5.29),

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

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 loses 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 (5.29). As is seen from the result (5.28), 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

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

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 experiment⁵.

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

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

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.

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

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

⁵This implies that the bare parameters have to be adapted at each order of perturbation theory calculated.

with the fermionic contribution given by

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

Using the trace identities $\text{tr}1 = 4$, $\text{tr}\gamma_\mu = 0$, and $\text{tr}\gamma_\mu\gamma_\nu = 4g_{\mu\nu}$ 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. This will be taken up in section 5.3. It is sufficient to observe that the resulting integral, just by counting powers of integration momenta, will have the form

$$\Pi_\phi^{y^2} = c_1\Lambda^2 + (c_2m^2 + c_3q^2) \ln \frac{\Lambda}{m} + f(m^2, q^2),$$

where f is some finite function when Λ is sent 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 (5.27). 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

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

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 Z s 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.

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 the non-Abelian case 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 Yang-Mills type 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. This is called

renormalizability by power-counting. 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 most quantum gauge theories in four dimensions with non-trivial dynamics and observable bound states renormalization is necessary.

5.2.4 Renormalization schemes

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

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

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 choose⁶

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

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.

⁶The Thomson limit is this expression at $p = q = 0$.

In fact, it turns out that the conditions chosen are arbitrary, i. e., any genuinely measurable quantity is not depending on this choice⁷. 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

$$(p^2 g_{\mu\nu} - p_\mu p_\nu)_{p^2=\mu^2} D_{AA}^{\mu\nu}(\mu^2) = i \quad (5.30)$$

$$\mu^2 g_{\mu\nu} D_{AA}^{\mu\nu}(\mu^2) = i\xi \quad (5.31)$$

$$\text{tr} D_{\psi\psi}(\mu^2) = im(\mu^2) \quad (5.32)$$

$$(\text{tr} \gamma_\mu p^\mu D_{\psi\psi})(\mu^2) = i16\mu^2 \quad (5.33)$$

$$\Gamma^{A\bar{\psi}\psi}(\mu^2, \mu^2, \mu^2) = ie(\mu^2) \quad (5.34)$$

Note that there is no condition that involves a mass of the photon. The corresponding counter-term would violate gauge-invariance, as a term $\sim A_\mu^2$ is not gauge-invariant. The condition (5.31) follows actually directly from the QED version of the STI (4.79). 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 (5.30-5.34) 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.

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 symmetry⁸.

⁷Actually, 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.

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

This perspective will be further discussed in chapter 7.

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 (5.32) and (5.34). 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 5.4. Note that masses manifest themselves as poles in correlation functions, and only in the pole scheme this explicitly enters, while (5.32) is actually only indirectly related to this pole. There is also a problem with masses for gauge-dependent elementary fields, which will be discussed in more detail in chapter 6. In the end, everything which depends on the choice of scheme and scale are not physical. However, measurements can nonetheless be used to determine them within a fixed scheme as a function of the scale, using a prescription like (5.30-5.34). And thus one can, e. g., plot the energy-dependence of such a quantity. However, the plot is only meaningful after fixing the renormalization scheme.

When changing to Yang-Mills theory, there are more renormalization conditions, due to the ghosts, and if there is more than one matter type. However, because of coupling unification further vertices do not all renormalize independently. Only those appearing at the Lagrangian level require usually independent conditions. A vertex like a two-fermion-two-ghost vertex, which arises at loop level, will not need an independent counter term, though it will need to be renormalized. This distinguishes the so-called primitively divergent quantities from those whose renormalization could be determined in terms of others. In practice, however, this may actually be very involved.

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 another type of regularization, dimensional regularization. Before doing so in section 5.2.8, it is useful to linger a little more on general properties of renormalization.

5.2.5 Renormalization group

As discussed in section 5.2.4, in the renormalization process a renormalization 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 $\mu \rightarrow \mu + \delta\mu$ 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 $p \rightarrow p + \delta p$, in Yang-Mills theory the gauge coupling g and the masses of the matter fields 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 vertex function

$$\Gamma^n = \langle \phi_1 \dots \phi_n \rangle$$

will be shifted under an infinitesimal transformation in μ to first order as

$$\Gamma^n \rightarrow (1 + n\delta Z)\Gamma^n \quad (5.35)$$

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 wavefunction renormalization is entirely given in terms of these quantities. Thus, (5.35) must also be given by⁹

$$n\delta Z\Gamma^n = d\Gamma^n = \frac{\partial\Gamma^n}{\partial\mu}\delta\mu + \frac{\partial\Gamma^n}{\partial g}\delta g + \frac{\partial\Gamma^n}{\partial m}\delta m.$$

To remove the explicit dependence on $\delta\mu$, usually the redefinitions

$$\begin{aligned} \beta &= \mu \frac{\delta g}{\delta\mu} \\ \beta_m &= \frac{\mu}{m} \frac{\delta m}{\delta\mu} \\ \gamma &= \mu \frac{\delta Z}{\delta\mu} \end{aligned} \quad (5.36)$$

are performed, called β functions or γ anomalous dimensions. The particular correlation function no longer appears. Hence, 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

$$\left(\mu \frac{\partial}{\partial\mu} + \beta \frac{\partial}{\partial g} + \beta_m m \frac{\partial}{\partial m} - n\gamma \right) \Gamma^n = 0, \quad (5.37)$$

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

⁹If there are more parameters, like the gauge parameter, further terms appear. This is right now an unnecessary complication.

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 function¹⁰

$$\Gamma_0 = \Pi_i Z_i^{n_i} \Gamma,$$

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 (5.37).

The situation becomes particularly simple in a mass-independent scheme, i. e. where all renormalization conditions do not involve any mass explicitly¹¹. Then in the equation (5.37) the functions do not depend on any ratios μ/m , and the equation can be integrated using the method of characteristics, yielding

$$\Gamma(\mu, \alpha, m, Q) = \exp \left(- \sum_i n_i \int_{\alpha}^{\bar{\alpha}} dx \frac{\gamma_i(x)}{\beta(x)} \right) \Gamma(\bar{\mu}, \bar{\alpha}, \bar{m}, Q),$$

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 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

¹⁰Note 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.

¹¹Such schemes have their own problems due to threshold effects, but this of no importance for the present case.

relation (5.36) has been used to include the canonical dimension into the exponent. Since without renormalization all γ_i vanish, they modify the canonical dimension. Thus the name 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_r} , 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 Yang-Mills theory coupled to N_f fermions in the fundamental representation it follows to NLO

$$Q^2 \frac{d\alpha}{dQ^2} = \beta(\alpha(Q^2)) \approx -\alpha^2 \sum_i \beta_i \alpha^i$$

$$\beta_0 = \frac{11}{3} C_A - \frac{4}{3} N_f T_R \quad (5.38)$$

$$\frac{Q^2}{\bar{m}(Q^2)} \frac{d\bar{m}(Q^2)}{dQ^2} = \gamma_m(\alpha(Q^2)) \approx -\alpha \sum_i \gamma_i \alpha^i \quad (5.39)$$

$$\gamma_0 = C_R \xi$$

where β_0 and γ_0 (and also β_1) are the same in all mass-independent renormalization schemes, while the further expansion coefficients are not. The approximation indicates that this will not capture any non-perturbative contributions. As discussed later in section 5.4, this yields, e. g., the running of the coupling.

5.2.6 BPHZ and MOM schemes

The general renormalization procedure becomes quite intricate beyond one loop. The reason is that with multiple nested loop integrals it is possible that sub integrals diverge. But if the theory is perturbatively renormalizable at all, i. e. by power counting, the Bogoliubov-Parasiuk-Hepp-Zimmermann (BPHZ) theorem guarantees that it is sufficient to include the counter terms for all possible vertices up to the relevant order in power counting. E. g., in four dimensions these are up to mass dimension four. Since counter-terms appear only at loop order, they are of higher order in the coupling constant. Thus, say, at two-loop orders the counter terms generate one-loop diagrams with the counter

term value at the vertex. That has already been seen explicitly in section 5.2.1. The structure was

$$\Pi(p^2) + \delta m^2,$$

where the self-energy was a loop diagram, but the counter term δm^2 was just a constant. Thus, the BPHZ theorem generally allows to renormalize by creating all diagrams to a given loop order and add to them all diagrams with one loop less, but all possible insertions of counter terms. This can be systematized to leading to the so-called forest-and-groove construction.

This result immediately suggests a particular regularization strategy. Because tree-level vertices yield power-like momentum behavior, in general the divergence structure of a diagram will be of the form

$$\Pi(p^2) = c_0 + c_1 p^2 + c_2 p^4 + \pi(p^2)$$

where the coefficients c_i are divergent, and π is finite. Such a situation can be regularized with a momentum subtraction,

$$\Pi(p^2) - \Pi(0) - p^2 \partial_{p^2} \Pi(p^2)|_{p^2=0} - \frac{1}{2} p^4 \partial_{p^2}^2 \Pi(p^2)|_{p^2=0} = \pi(p^2),$$

and thus automatically finite. This is a momentum subtraction scheme (MOM), more precisely the BPHZ subtraction scheme. It generates automatically a finite result, but requires that Π is (reasonably) regular at $p^2 = 0$. Otherwise, an infrared regulator is needed. Utilizing further properties also allows to choose $p^2 = \mu^2$ as subtraction point, yielding a general momentum subtraction scheme.

Of course, if the quantity in question depends on n independent momenta, this needs to be taken care of. However, as again the maximum divergence is limited by the BPHZ theorem, it is sufficient to consider a fixed momentum arrangement, e. g. the symmetric point $p_1^2 = p_2^2 = \dots = p_n^2 = q^2$, and perform the derivatives with respect to q^2 .

5.2.7 Gauge-symmetry-breaking regulators

The cut-off regularization discussed in section 5.2.1 is not gauge invariant in the sense that the necessary counter term is found to depend on the gauge. However, almost all regularization prescriptions are in the same sense in conflict with symmetries, in particular gauge symmetries. This is not a problem in principle. It will only yield that there are more counter-terms, which break the symmetries. Thus the symmetries are not manifest anymore at the level of the renormalized Lagrangian, but only at the level of physical observables, like cross-sections. This also modifies the STIs and WTIs, as they reflect this breaking.

How this works out can be best seen in an explicit example. E. g., when using a pure cutoff-regularization with cutoff Λ , the gauge boson tadpole term will yield a term proportional¹² $g_{\mu\nu}\Lambda^{d-2}$. Likewise, the other one-loop parts have similar structures.

As an example, doing the calculations explicitly for the three-dimensional case, the full self-energy to one-loop order in Landau gauge reads

$$\begin{aligned}\Pi_{\mu\nu}^{ab} &= -\delta^{ab}\frac{g^2C_A}{64p}\left(g_{\mu\nu} + \frac{p_\mu p_\nu}{p^2}\right) - \frac{g^2C_A\Lambda}{6\pi^2p^2}\delta^{ab}g_{\mu\nu} \\ &\quad -\delta^{ab}\frac{g^2C_A}{32p}\left(5g_{\mu\nu} - 6\frac{p_\mu p_\nu}{p^2}\right) - \frac{4g^2C_A\Lambda}{6\pi^2p^2}\delta^{ab}g_{\mu\nu} + \frac{2g^2C_A\Lambda}{3\pi^2p^2}\delta^{ab}g_{\mu\nu} \\ &= \delta^{ab}\frac{11g^2C_A}{64p}\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) - \frac{g^2C_A\Lambda}{6\pi^2p^2}\delta^{ab}g_{\mu\nu},\end{aligned}\tag{5.40}$$

where in the first step the first two term comes from the ghost loop, the third and fourth one from the gluon loop, and the fifth from the tadpole. This violates the not modified Slavnov-Taylor identity (4.79), as

$$p_\mu p_\nu D_{\mu\nu}^{ab} = -\frac{g^2C_A\Lambda}{6\pi^2}\delta^{ab}$$

which should yield zero on the right-hand side in Landau gauge. Thus, cutoff regularization breaks gauge invariance. This can be remedied by adding a counter term

$$\mathcal{L}_c = \frac{g^2C_A\Lambda}{6\pi^2}A_a^\mu A_\mu^a$$

to the Lagrangian, which also violates gauge invariance, as it is effectively an (infinite) mass term for the gauge boson. This additional counter-term cancels the effect, and restores the Slavnov-Taylor identity, and thus gauge invariance at the quantum level. Thus, gauge invariance is allowed to be broken by the regulator, provided the renormalization ensures that gauge invariance is maintained in the renormalized theory. As this example shows, this is usually technically involved.

5.2.8 Dimensional regularization

One of the very few regulators to keep all symmetries intact is dimensional regularization. It leaves all symmetries manifest. The price to be paid is the assumption that correlation functions are analytic functions of the dimensionality. That this is violated can be seen from the fact that two-dimensional Yang-Mills theory is actually a trivial theory, which can be solved exactly. There are no physical degrees of freedom, as there are no transverse

¹²In two dimensions this will be logarithmic.

polarizations. Therefore, even at the level of perturbation theory, it is only a theory of gauge-fixing, and the physical state space is empty. Because of the exact solvability, this result can be demonstrated non-perturbatively: There is only one physical state: An empty vacuum. As the theory is interacting for higher dimensions, this implies a non-analyticity at two dimensions, and thus the assumption is violated. Thus, this regulator will not work non-perturbatively without also inducing problems.

But for perturbation theory it is without problems. In fact, it is perturbatively so well working that it has become the mainstay of perturbative calculations, even though it requires 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 name dimensional regularization stems from the 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 (5.27), 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,

$$\int d^d k (k^2)^\alpha = 0. \quad (5.41)$$

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

$$A = \frac{1}{i\pi^2} \int d^d k \frac{1}{(k^2 - m^2 + i\epsilon)^r} \rightarrow A^r = \frac{M^{d-D}}{i\pi^2} \int \frac{d^D k}{(2\pi)^{D-d}} \frac{1}{(k^2 - m^2 + i\epsilon)^r}. \quad (5.42)$$

The original, unregularized integral is obtained in the limit $D \rightarrow d$. 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 is nothing but the usual dimensional transmutation. This integral is convergent for $D < 2r$. Performing a Wick rotation, i. e., replacing formally $k_0 \rightarrow ik_0$, yields

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

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,

$$\int d\Omega_D = \frac{2\pi^{\frac{D}{2}}}{\Gamma(\frac{D}{2})} \quad (5.43)$$

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt.$$

Of course, a sphere is only a geometric object in the conventional sense for D being integer. The expression (5.43) 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

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

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

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

A^r is now expanded for small δ , as the desired limit is $\delta \rightarrow 0$. For the case of $r = 1$, i. e., for a massive tadpole like (5.27), the expansion in δ yields

$$A^r = m^2 \left(\frac{1}{\delta} - \gamma + \ln 4\pi - \ln \frac{m^2}{M^2} + 1 \right) + \mathcal{O}(\delta), \quad (5.44)$$

where γ is the Euler constant $\gamma \approx 0.577$. 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: Renormalization is performed, with the divergent terms being absorbed in counter-terms. 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

$$\int \frac{d^d q}{(2\pi)^d} q^{2\alpha} (q-p)^{2\beta} = \frac{1}{(4\pi)^{\frac{d}{2}}} \frac{\Gamma(-\alpha - \beta - \frac{d}{2}) \Gamma(\frac{d}{2} + \alpha) \Gamma(\frac{d}{2} + \beta)}{\Gamma(d + \alpha + \beta) \Gamma(-\alpha) \Gamma(-\beta)} p^{2(\frac{d}{2} + \alpha + \beta)}, \quad (5.45)$$

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.

An important insight is that an expansion of Γ will always yield terms like $1/\delta$, and never terms like $1/\delta^2$. Thus, the apparent divergences are of different orders than, e. g., in cutoff regularization, showing explicitly the difference.

In comparison, the result (5.40) in dimensional regularization reads

$$\begin{aligned}\Pi_{\mu\nu}^{ab} &= -\delta^{ab}\frac{g^2C_A}{64p}\left(\delta_{\mu\nu} + \frac{p_\mu p_\nu}{p^2}\right) - \delta^{ab}\frac{g^2C_A}{32p}\left(5\delta_{\mu\nu} - 6\frac{p_\mu p_\nu}{p^2}\right) + 0 \\ &= \left(\delta_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right)\delta^{ab}\left(1 - \frac{11g^2C_A}{64p}\right),\end{aligned}$$

and thus has the same, finite¹³, transverse part than the one obtained in cutoff regularization, but is directly compatible with the STI (4.79).

5.2.9 Lattice and infrared regularization

An alternative to violating gauge invariance with a cutoff regulator is violating Poincaré symmetry using a lattice as a regulator. There are two distinct advantages compared to the cutoff regulation. The consequences of Poincaré symmetry are much more easily to estimate, like angular momentum, than of gauge invariance violation. The reason is that Poincaré invariance is a global symmetry, rather than a local one. This also implies the second advantage. Once gauge invariance is broken, the results even for gauge-invariant quantities will depend on the choice of gauge until gauge invariance is restored by renormalization. This has been exemplified in section 5.2.7. Because most calculations are approximate, this restoration is usually not exactly possible, leading to residual violations, and thus dependence on the choice of gauge.

Of course, there is a disadvantage. A lattice is obtained by making space-time discrete. This implies that integrals become sums and various directions are different, e. g., diagonally between two lattice points compared to along the edge of a lattice. This leads to much more involved tensor structures, and severely more complicated calculations. Thus, this approach usually lends itself easier for numerical calculations, so-called lattice simulations, than for analytical calculations. Even though there are exceptions to this rule. As even the numerical simulations are very involved, a description of them will be relegated to the corresponding lecture.

¹³In three dimensions Yang-Mills theory is superrenormalizable, and thus the correlation function itself is, after renormalization, finite.

Once this is done, the results will be depending on the distance of the lattice points, the lattice spacing. In the renormalization this dependence has then to be removed, just as always a dependence on the regulator has to be removed. However, this has now also to be done numerically, leading to technical problems, which are also beyond the scope of the present lecture.

There is, however, one element of further importance. For any numerical simulations the number of points has to become finite. Thus, in addition the volume has to be made finite.

However, a finite volume has also an interesting further advantage. Because a finite volume restricts the maximum distance this will regulate also the infrared, not only the ultraviolet, of integrals. Thus, in theories in which there are spurious infrared divergences, e. g. in loop integrals with massless particles, they will be regulated. Because wave-lengths are limited, every (scalar¹⁴) particle is massive in a finite volume. Of course, this implies a dependence on the volume of the results. Taking the limit of the volume to infinity afterwards should then be possible without additional renormalization. After all, flat Minkowski space with infinite volume is the arena where particle physics is defined. Thus, these regularization is here only a technical tool. In fact, imposing a finite volume can also be combined with different regulators than a lattice regularization.

This becomes especially important when using approximate methods. Especially many infrared divergences in gauge theories emerge only in perturbative calculations, and are an artifact of them. Then, such regulators are necessary, and have to be dealt with carefully.

There is also another advantage. In dimensions less than four infrared divergencies become more severe. The reason is that propagators retain their same structure as in higher dimensions, as this is determined in terms of number of derivatives. Especially, the propagators of massless particles diverge as $1/p^2$ in all dimensions. On the other hand, in loop integrals the measure is proportional to p^{d-1} , and thus provides less and less suppression in lower dimensions. While this reduces ultraviolet divergencies, making many theories effectively ultraviolet finite in lower dimensions, this amplifies infrared divergencies due to the massless particles. This entails that perturbation theory in Yang-Mills theory needs further infrared regularization beyond one loop in three dimensions. This yields additional logarithmic structures on top of the power-like structure obtained in three-dimensional perturbation theory. In two dimensions the divergencies are so severe that perturbation theory breaks down. However, because Yang-Mills theory becomes trivial in two dimensions, this problem can usually be circumvented by exploiting this feature. This

¹⁴This is not true for massless particles with spine one or more, as here the number of degrees of freedom differs between the massive case and the massless case.

will not be detailed further here.

5.2.10 Composite operators

When dealing with gauge theories beyond perturbation theory, it will be seen that composite states, i. e. products of fields evaluated at the same point, become very important. E. g. the simplest composite operator to describe positronium is $\bar{\psi}(x)\psi(x)$, where ψ is the electron field. They describe thus, e. g., bound states. They also play a role in various perturbative settings.

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 fields 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, by virtue of the BPHZ theorem of section 5.2.6, this divergence cannot be an independent one. It must thus be expressible in terms of the other renormalization constants. In fact, in the present case it turns out to be related to the mass renormalization, as this is the same composite operator as the mass term in the Lagrangian. Thus, it must be given by the inverse of the fermion 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 not to 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 fermions, and thus its renormalization is fixed together with the mass renormalization of the fermions. This is how the relation to the mass-renormalization of the fermions 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 fermions, 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 that as the sources of the composite operators are sent 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.

5.3 Examples of perturbative calculations

In the following some processes will be discussed, to enhance familiarity with the details of the procedures and add a number of technical tricks to the tool set. Only Yang-Mills theory will be considered, though without explicitly specifying the gauge group. As the most relevant case is four dimensions, and issues like renormalization are specific for every dimension, the following will be done only in four dimensions, and using dimensional regularization and in the $\overline{\text{MS}}$ scheme.

Also, there are additional issues in both lower and higher dimensions, which would lead astray at this point. E. g., perturbation theory works in three dimensions only up to one-loop order, and in two dimensions not all, because of infrared singularities, and in five dimensions Yang-Mills theory is perturbatively non-renormalizable.

5.3.1 Gauge boson scattering

As a first example, consider elastic gauge boson scattering at tree-level. As seen in section 4.5.2.4, a physical initial state can only contain transverse gauge bosons. Thus, only transverse polarization vectors are admissible for the initial state. This also implies that the same is true for the final state. Also, it is necessary to sum over all colors to obtain a result, which respects the residual global gauge symmetry in linear covariant gauges, and likewise BRST symmetry and gauge invariance. This implies that gauge bosons can be exchanged in the final state, as the result is summed over all possible combinations of color charge.

This yields four diagrams, three from the gauge-boson three point vertex and one from the four-point vertex, which can be taken from section 5.1.2. The three diagrams with three-point vertex contain an s -channel exchange and a t -channel exchange. The latter appears twice, as the exchange of the final state particles creates a different diagram. Having as initial state the momenta p_1 and p_2 and colors a_1 and a_2 and as final state values q_1 , q_2 , b_1 and b_2 the relevant truncated matrix elements is

$$\mathcal{M}_{a_1 a_2 b_1 b_2}(p_i, q_i, \lambda_i, \omega_i) = \epsilon_\mu^*(p_1, \lambda_1) \epsilon_\nu^*(p_2, \lambda_2) \mathcal{N}_{a_1 a_2 b_1 b_2}^{\mu\nu\rho\sigma}(p_i, q_i) \epsilon_\rho(q_1, \omega_1) \epsilon_\sigma(q_2, \omega_2).$$

The polarization vectors $\epsilon(p, \lambda)$ depend on the helicities λ , and are transverse to its momenta $p_\mu \epsilon^\mu(p, \lambda) = 0$, and are space-like, complex unit vectors. Thus, there are two independent ones, as required, and they can be taken to be orthogonal. For a momentum along the z -axis, a possible form is

$$\epsilon_\mu(p, \Lambda) = \begin{pmatrix} 0 \\ 1 \\ \pm i \\ 0 \end{pmatrix}$$

where the different signs correspond to the two different values of the helicity. This satisfies $\epsilon_\mu^*(\lambda) \epsilon^\mu(\lambda') = -\delta_{\lambda\lambda'}$. If the polarizations are not measured, it is useful in the following that

$$\sum_\lambda \epsilon_\mu^*(p, \lambda) \epsilon_\nu(p, \lambda) = -g_{\mu\nu}, \quad (5.46)$$

the so-called polarization sum.

The unprojected matrix element is the truncated one, given by

$$\begin{aligned} \mathcal{N}_{a_1 a_2 b_1 b_2}^{\mu\nu\rho\sigma}(p_i, q_i) &= \Gamma^{ggg}(p_1, p_2, -p_1 - p_2)_{\mu\nu\alpha}^{a_1 a_2, c} D_{\alpha\beta}^{cd}(p_1 + p_2) \Gamma^{ggg}(q_1, q_2, -p_1 - p_2)_{\rho\sigma\beta}^{b_1 b_2 d} \\ &+ \Gamma^{ggg}(p_1, q_1, -p_1 - q_1)_{\mu\rho\alpha}^{a_1 b_1 c} D_{\alpha\beta}^{cd}(p_1 + q_1) \Gamma^{ggg}(p_2, q_2, -p_1 - q_1)_{\nu\sigma\beta}^{a_2 b_2 d} \\ &+ \Gamma^{ggg}(p_1, q_2, -p_1 - q_2)_{\mu\rho\alpha}^{a_1 b_2 c} D_{\alpha\beta}^{cd}(p_1 + q_1) \Gamma^{ggg}(p_2, q_1, -p_1 - q_2)_{\nu\sigma\beta}^{a_2 b_1 d} \\ &+ \Gamma^{gggg}(p_1, p_2, -q_1 - q_2)_{\mu\nu\rho\sigma}^{a_1 a_2 b_1 b_2}. \end{aligned} \quad (5.47)$$

Due to the lengthy expressions for the propagator (5.13) and vertices (5.19) and (5.22), this is a tedious, but straight-forward, evaluation. Note that the antisymmetry of the structure constant implies no identical combination of gauge bosons will appear.

For (5.1) the absolute value of the matrix element squared is needed, and thus

$$\begin{aligned} |\mathcal{M}|^2 &= \mathcal{N}_{a_1 a_2 b_1 b_2}^{\mu\nu\rho\sigma}(p_i, q_i) \mathcal{N}_{a_1 a_2 b_1 b_2}^{\alpha\beta\gamma\delta*}(p_i, q_i) \times \\ &\times \sum_{\lambda_i, \omega_i} (\epsilon_\mu^*(p_1, \lambda_1) \epsilon_\nu^*(p_2, \lambda_2) \epsilon_\alpha^*(p_1, \lambda_1) \epsilon_\beta^*(p_2, \lambda_2) \times \\ &\times \epsilon_\rho(q_1, \omega_1) \epsilon_\sigma(q_2, \omega_2) \epsilon_\beta(q_1, \omega_1) \epsilon_\gamma(q_2, \omega_2)). \end{aligned}$$

Using (5.46), this can be reduced to contractions of the Lorentz indices. Note that the initial state and final state gauge indices are also summed over, to obtain a gauge-invariant result. Eventually, this leads to the final result

$$\frac{d\sigma}{dt} = \frac{9g^4}{32\pi s^2} \left(3 - \frac{tu}{s^2} - \frac{su}{t^2} - \frac{st}{u^2} \right) = \frac{9g^4}{32\pi s^2} \frac{(s^2 + st + t^2)^3}{s^2 t^2 (s+t)^2},$$

in terms of the Mandelstam variables, where the solid angle has been exchanged for the the Mandelstam variable t , with ϕ being integrated over. It shows nicely poles at any of the Mandelstam variables, which arise due to the masslessness of the gauge bosons. They correspond to poles in the exchange of a gauge boson in any of the possible three channels of the one-particle exchange diagrams.

5.3.2 Ghost self-energy

The simplest possible 1-loop process in Yang-Mills theory is the one-loop correction of the ghost propagator, as it involves only a single diagram: a so-called rainbow diagram, where the ghost emits during propagation a gluon, which it later absorbs again. Using the Feynman rules of section 5.1.2, the relevant expression for this diagram is

$$\Pi_G^{ab}(p) = \frac{g^2 f^{adc} f^{bfe} \delta^{ce} \delta^{df}}{(4\pi)^d} \int d^4 q (p+q)_\mu \frac{1}{q^2 + i\epsilon} \left(g_{\mu\nu} - (1-\xi) \frac{q_\mu q_\nu}{q^2} \right) \frac{1}{(p+q)^2 + i\epsilon} p_\nu$$

where it has been used that ghost, antighost, and gluon are all distinct, and thus no symmetry factor appears. To simplify matters, $\xi = 0$, i. e. Landau gauge, will be used in the following. Because the ghost is a gauge-dependent auxiliary degree of freedom the result will be different for a different choice.

Because the gauge boson propagator is a transverse projector, the part proportional to q_μ drops out, and the result is

$$\Pi_G^{ab}(p) = \frac{g^2 f^{adc} f^{bdc}}{(4\pi)^d} \int d^4 q \frac{1}{q^2 + i\epsilon} \left(p^2 - \frac{(pq)^2}{q^2} \right) \frac{1}{(p+q)^2 + i\epsilon}$$

The prefactor can be reduced using (2.15). This leads to two separate contributions

$$\Pi_G^{ab}(p) = \frac{g^2 C_A \delta^{ab}}{(4\pi)^d} \int d^4 q \left(\frac{p^2}{q^2(p+q)^2 + i\epsilon} + \frac{(pq)^2}{q^4(p+q)^2 + i\epsilon} \right) \quad (5.48)$$

The first term is already of the type (5.45), and could be directly calculated. The second has not yet the form.

The second term is a so-called tensor integral, call it $B^{\mu\nu}$ for the moment, as it is of the form $p_\mu p_\nu B^{\mu\nu}(p)$, if the external momentum is pulled out of the integral. Using the

fact that the integral will be eventually regulated, and thus is finite, Lorentz symmetry requires

$$B_{\mu\nu}(p) = g_{\mu\nu}B_{00}(p^2) + p_\mu p_\nu B_{11}(p^2).$$

Using transverse and longitudinal projectors, this yields

$$B_{00} = \frac{1}{d-1} \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2} \right) B_{\mu\nu} \quad (5.49)$$

$$B_{11} = \frac{1}{d-1} \frac{1}{p^2} \left(d \frac{p_\mu p_\nu}{p^2} - g_{\mu\nu} \right) B_{\mu\nu}, \quad (5.50)$$

where d is the dimension.

To get more generally applicable formulas for latter use, introduce for a moment masses m_1 and m_2 for the gluon and ghost. Rewrite then

$$qp = \frac{1}{2} \left(((q+p)^2 - m_2^2) - (k^2 - m_1^2) - p^2 + m_2^2 - m_1^2 \right).$$

With this, the integral kernel becomes that of a one-loop scalar integral, i. e. one with momentum dependencies only in the denominator. This structure can be tackled directly with Feynman parameters, and afterwards with dimensional regularization.

To do so, proceed as follow. Consider first a simpler integral

$$B_0 = \frac{(2\pi\mu)^{4-d}}{i\pi^2} \int d^d q \frac{1}{(q^2 - m_1^2)((q+p)^2 - m_2^2)},$$

which will anyway be needed to express the final solution in a compact way, and also is the first term in (5.48). Using the Feynman parametrization

$$\frac{1}{D_1 D_2} = \int_0^1 dx \frac{1}{((1-x)D_1 + xD_2)^2}$$

it follows that

$$B_0 = \frac{(2\pi\mu)^{4-d}}{i\pi^2} \int d^d q dx \frac{1}{(q^2 + 2xqp + x(p^2 - m_2^2) - (1-x)m_1^2)^2}$$

This integral is dimensionally regulated, and therefore finite. It is thus allowed to perform the shift $q \rightarrow q - xp$, yielding

$$B_0 = \frac{(2\pi\mu)^{4-d}}{i\pi^2} \int d^d q dx \frac{1}{(q^2 - x^2 p^2 + x(p^2 - m_2^2) - (1-x)m_1^2)^2}.$$

The momentum integral is now the tadpole integral of type (5.42) for $r = 2$, although with an x -dependent mass over which still has to be integrated,

$$B_0 = (4\pi\mu^2)^\delta \Gamma(\delta) \int_0^1 dx \frac{1}{(x^2 p^2 - x(p^2 - m_2^2 + m_1^2) + m_1^2)^\delta}.$$

Because δ is infinitesimal, the fraction can be expanded by

$$\frac{1}{a^x} = 1 - x \ln a + \frac{1}{2} (x \ln a)^2 + \mathcal{O}(x)$$

yielding to leading non-trivial order

$$B_0 = \frac{1}{\delta} - \gamma + \ln(4\pi) - \int dx \ln \frac{x^2 p^2 - x(p^2 - m_2^2 + m_1^2) + m_1^2 - i\epsilon}{\mu^2} + \mathcal{O}(\delta),$$

where the $i\epsilon$ has been made explicit, as this is important to select the correct Riemann sheet for the integration of the logarithm. The final integral is non-trivial but elementary, and can be found in either tables or solved by further, lengthy, substitutions. The final result is

$$\begin{aligned} B_0 &= \frac{1}{\delta} - \gamma + \ln(4\pi) + 2 - \ln \frac{m_1 m_2}{\mu^2} + \frac{m_1^2 - m_2^2}{p^2} \ln \frac{m_2}{m_1} - \frac{m_1 m_2}{p^2} \left(\frac{1}{r} - r \right) \ln r \\ r &= \frac{1}{2m_1 m_2} \left(m_1^2 + m_2^2 - p^2 - i\epsilon \pm \sqrt{(p^2 - m_1^2 - m_2^2 + i\epsilon)^2 - 4m_1^2 m_2^2} \right). \end{aligned}$$

This structure is the same as in non-gauge theories: The result has a cut, starting at $p^2 = (m_1 + m_2)^2$ with a non-vanishing imaginary part due to the decay of the particle as soon as sufficient energy is available. Because the gauge bosons and ghosts are massless, this will yield that there is a cut starting at zero. This feature gives rise to infrared singularities in more complicated processes, which need a more elaborate framework to be dealt with in general. Here, however, it will be sufficient to just keep the masses long enough to regularize all expressions. The final result will be independent of them.

Returning to the original problem, B_{00} from (5.49) can now be expressed as

$$B_{00} = \frac{1}{2(d-1)} (A_0(m_2^2) + 2m_1^2 B_0 + (p^2 + m_1^2 - m_2^2) B_1),$$

where A_0 is the tadpole integral (5.44)

$$A_0(m^2) = m^2 \left(\frac{1}{\delta} - \gamma + \ln(4\pi) - \ln \frac{m^2}{\mu^2} + 1 \right) + \mathcal{O}(\delta)$$

and B_1 is a similarly obtained integral as B_0 ,

$$\begin{aligned} B_1 &= \frac{(2\pi\mu)^{4-d}}{i\pi^2} \frac{1}{2p^2} \int d^d q \frac{2pq}{(q^2 - m_1^2)((p+q)^2 - m_2^2)} \\ &= \frac{1}{2p^2} (A_0(m_1^2) + A_0(m_2^2) - (p^2 + m_1^2 - m_2^2) B_0). \end{aligned}$$

Likewise, B_{11} from (5.50) is given by

$$B_{11} = \frac{1}{2(d-1)p^2} ((d-2)A_0(m_2) - 2m_1^2 B_0 - d(p^2 + m_1^2 - m_2^2) B_1).$$

and thus completes the needed integrals.

Before continuing, it will be necessary to obtain the zero-mass limits of these integrals. Because A_0 is a tadpole integral, it vanishes at zero mass, according to the dimensional regularization rule (5.41). For the B_i it is needed that, after setting $m_1 = m_2 = m$

$$r = \frac{i\epsilon - p^2 \pm \sqrt{(p^2 + i\epsilon)^2}}{2m^2} + 1 \mp \frac{i\epsilon(1 + p^2)}{\sqrt{(p^2 + i\epsilon)^2}} \mp \frac{m^2}{\sqrt{(p^2 + i\epsilon)^2}} + \mathcal{O}(m^3).$$

This yields eventually

$$B_0 = \frac{1}{\delta} - \gamma + \ln(4\pi) + 2 - \ln\left(\frac{p^2}{\mu^2}\right),$$

and thus up to a constant term, which can be absorbed in the renormalization, a logarithmic behavior with momentum. As B_1 and the B_{ii} are determined in turn by B_0 , this completely fixes the result at zero mass. Note that, as advertised above, the result has a cut starting at zero momentum. Physically, this corresponds to the splitting of a massless particle into two massless particles.

The renormalized result for the ghost self-energy can now be constructed along the same lines. However, because of the second term

$$\Pi_G^{ab}(p) = \frac{g^2 C_A \delta^{ab}}{(4\pi)^d} \int d^4q \left(\frac{p^2}{q^2(p+q)^2 + i\epsilon} + \frac{(pq)^2}{q^4(p+q)^2 + i\epsilon} \right)$$

this cannot be done using the expressions B_i and B_{ii} , as it effectively involves three propagators, a so-called triangle integral. It is possible to solve them for finite masses in the same way as the B_i integrals. However, this becomes quite tedious.

Fortunately, this is not necessary, because the ghost self-energy can be recast into the form of (5.45). For this note that pq can be rewritten as

$$pq = -\frac{1}{2} ((p-q)^2 - p^2 - q^2).$$

Then, with a change of $q \rightarrow -q$, the ghost self-energy becomes a sum of terms of the type (5.45),

$$\Pi_G^{ab}(p) = -\frac{g^2 C_A \delta^{ab}}{(4\pi)^d} \int d^4q \left(\frac{p^2}{q^2(p-q)^2 + i\epsilon} + \frac{((p-q)^2 - p^2 - q^2)^2}{4q^4(p-q)^2 + i\epsilon} \right).$$

Applying (5.45) yields

$$\begin{aligned} \Pi_G^{ab}(p) = & -\frac{g^2 C_A \delta^{ab} p^2}{(4\pi)^{\frac{d}{2}}} \left(\frac{p^2}{\mu^2}\right)^{\left(\frac{d}{2}-2\right)} \left(\frac{\Gamma\left(2-\frac{d}{2}\right)\Gamma\left(\frac{d}{2}-1\right)^2}{\Gamma(d-2)\Gamma(1)^2}\right. \\ & + \frac{1}{4} \left(\frac{\Gamma\left(\frac{d}{2}-1\right)\Gamma\left(\frac{d}{2}-1\right)\Gamma\left(\frac{d}{2}\right)}{\Gamma(0)\Gamma(1)\Gamma(d-1)} - 2\frac{\Gamma\left(2-\frac{d}{2}\right)\Gamma\left(\frac{d}{2}-2\right)\Gamma\left(\frac{d}{2}\right)}{\Gamma(0)\Gamma(2)\Gamma(d-2)}\right. \\ & \frac{\Gamma\left(3-\frac{d}{2}\right)\Gamma\left(\frac{d}{2}-2\right)\Gamma\left(\frac{d}{2}-1\right)}{\Gamma(1)\Gamma(2)\Gamma(d-3)} + \frac{\Gamma\left(1-\frac{d}{2}\right)\Gamma\left(\frac{d}{2}-2\right)\Gamma\left(\frac{d}{2}+1\right)}{\Gamma(-1)\Gamma(2)\Gamma(d-1)} \\ & \left. \left. - 2\frac{\Gamma\left(1-\frac{d}{2}\right)\Gamma\left(\frac{d}{2}-1\right)\Gamma\left(\frac{d}{2}\right)}{\Gamma(0)\Gamma(1)\Gamma(d-1)} + 2\frac{\Gamma\left(2-\frac{d}{2}\right)\Gamma\left(\frac{d}{2}-1\right)^2}{\Gamma(1)\Gamma(d-2)}\right)\right). \end{aligned}$$

Combining everything, and performing the limit of $d \rightarrow 4 - \epsilon$ and neglecting $\mathcal{O}(\epsilon)$, yields

$$\begin{aligned} \Pi_G^{ab}(p) &= -\frac{g^2 C_A \delta^{ab}}{(4\pi)^2} p^2 \left(1 + \epsilon \ln(4\pi) + \epsilon \ln \frac{p^2}{\mu^2}\right) \left(\frac{1}{\epsilon} + 2 - \gamma + \frac{1}{4} \left(\frac{3}{\epsilon} + \frac{13}{2} - \gamma\right)\right) \\ &= -\frac{g^2 C_A \delta^{ab}}{(4\pi)^2} p^2 \left(\frac{17}{\epsilon} + \frac{14}{8} \ln \frac{p^2}{\mu^2} + \frac{29 + \ln(4\pi)(14) - 10\gamma}{8}\right). \end{aligned} \quad (5.51)$$

It was here necessary to keep also terms of order ϵ in the process, as otherwise finite terms would have been missed, which arose from the combination with pole terms. The result (5.51) has precisely the anticipated structure,

$$\frac{a}{\epsilon} + b + c \ln \frac{p^2}{\mu^2}.$$

Setting the counterterm to

$$\delta \tilde{Z}_3 = -\frac{a}{\epsilon} - b$$

yields the final result for the renormalized selfenergy

$$\Pi_G^{ab}(p) = -\frac{g^2 C_A \delta^{ab}}{(4\pi)^2} p^2 \frac{7}{4} \ln \frac{p^2}{\mu^2},$$

which is, unsurprisingly, gauge-dependent.

Performing the resummation, or using alternatively the Dyson-Schwinger equation, yields the final ghost propagator to order g^2 using the tree-level propagator $G_0 = -1/p^2$ to obtain

$$G^{ab}(p) = \delta^{ab} G_0 \sum_{n=0}^{\infty} (\Pi G_0)^n = \frac{\delta^{ab}}{\frac{1}{G_0} - \Pi} = -\frac{\delta^{ab}}{p^2 \left(1 + \frac{g^2 C_A}{(4\pi)^2} \frac{7}{4} \ln \frac{p^2}{\mu^2}\right)}. \quad (5.52)$$

The result is a propagator, which is not an entirely massless free particle, as it does not behave as $1/p^2$, but has logarithmic corrections at large energies, $\sim 1/(p^2 \ln p)$. This is

a typical behavior for propagators in renormalizable four-dimensional theories, and also shows for massive particles once $p \gg m$. The propagator keeps its massless pole at $p = 0$, this still exhibiting the properties of a massless particle. Furthermore, there is another logarithmic pole at space-like $p^2 > 0$. It is only logarithmic, and originates from the fact that perturbation theory is strictly speaking not valid anymore at sufficiently small energies, due to a so-called Landau pole. This will be discussed in more detail in section 5.4. There is also the announced the cut at time-like momenta $p^2 < 0$.

5.3.3 Gauge boson self-energy and gauge invariance

The same process, but with substantially more technical complications, can be performed for the gauge boson. Without going into details, it is, however, instructive to study the general structure of the self-energy.

Because the gauge boson propagator is a tensor, the behavior will be at order g^2

$$\Pi_{\mu\nu}^{g^2} = \Pi_{\mu\nu}^c + \Pi_{\mu\nu}^A + \Pi_{\mu\nu}^T,$$

where these are the self-energy stemming from a ghost-loop, and a gauge boson tadpole. The latter will vanish in dimensional regularization. As in the previous case, it is useful to capitalize on Lorentz symmetry. Given that there is only one independent Lorentz vector, the momentum p , it is necessarily true that

$$\Pi_{\mu\nu}^{g^2}) = P_{\mu\nu}^T \Pi_T + P_{\mu\nu}^L \Pi_L = \left(g_{\mu\nu} + \frac{p_\mu p_\nu}{p^2} \right) \Pi_T + \frac{p_\mu p_\nu}{p^2} \Pi_L.$$

This introduces the transverse projector $P_{\mu\nu}^T$ and the longitudinal projector $P_{\mu\nu}^L$.

While the split appears arbitrary at first, its choice is based on the Slavnov-Taylor identity (4.79), which implies

$$\Pi_L = \xi + \delta\xi = \xi_R,$$

where ξ_R is the renormalized gauge parameter. Since

$$\Pi_L = P_L^{\mu\nu} \Pi_{\mu\nu}^c + P_L^{\mu\nu} \Pi_{\mu\nu}^A = \xi_R$$

this implies that either both of the terms need to be individually constant, or their non-constant terms need to cancel each other. Performing the actual calculations would show that the latter is the case. Thus, to ensure gauge invariance, as tested by the Slavnov Taylor identity, requires to consider both the ghosts and the gauge bosons.

This implies that only Π_T will contain non-trivial information. In fact, no Slavnov-Taylor identity constraints its behavior¹⁵. This is not surprising, as the Slavnov-Taylor identity express only constraints due to gauge invariance. Thus, the transverse part can be considered to contain the dynamical information. However, in general, the result will still depend on the gauge parameter.

Actually performing the calculation eventually yields a structure like (5.52) for Π_T , just with a different prefactor before the logarithm.

5.4 Running couplings, Landau poles, and asymptotic freedom

An equation like (5.34) defines an energy dependence of the coupling constant, a so-called running coupling constant. Generically, resummed perturbation theory to second order yields for (5.34) an expression like

$$\alpha(q^2) = \frac{g(q^2)^2}{4\pi} = \frac{\alpha(\mu^2)}{1 + \frac{\alpha(\mu^2)}{4\pi}\beta_0 \ln \frac{q^2}{\mu^2}} \equiv \frac{4\pi}{\beta_0 \ln \frac{q^2}{\Lambda^2}} \quad (5.53)$$

for the gauge coupling g , where β_0 depends on the object in question. However, Slavnov-Taylor identities will establish relations, and often yield identical values. This is as expected, as gauge invariance implied coupling universality.

The equation (5.53) implies that once the coupling is fixed to experiment at μ , and an expression like (5.34) is evaluated at a different momentum q , the right-hand-side is given in terms of $\alpha(\mu^2)$ by this expression (5.53). 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 (5.34). 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,

¹⁵To be precise, the transverse part appears in the Slavnov-Taylor identities for vertices, but there is non obtained from doing a BRST transformation of two fields. Thus, this usually rather considered to be Slavnov-Taylor identities for the vertices, which take the transverse part as input. However, as the equations are really coupled, this is not a clear-cut distinction.

β_0 could in principle depend on the renormalization scheme, but does not do so in theories with a single type of matter in the fundamental representation. 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 in 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 (5.53) 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, e. g. for a QCD-like theory,

$$\beta_0 = \frac{11}{3}C_A - \frac{2}{3}N_f \quad (5.54)$$

where C_A is the adjoint Casimir of the gauge group, and N_f counts the number of fermion flavors in the fundamental representation. Other types of particles or other representations would yield a different value. For QCD this yields $\beta_0 = 7$, if all masses are neglected, i. e., at very high energies.

There are a few generic features of (5.53). First of all, since Λ is a finite number, 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. In the region where the coupling remains positive, perturbation theory could be applicable. In QCD, this is at energies higher than Λ^2 for less than 17 flavors and the other way around for more flavors, in QED it is at energies lower than Λ^2 .

This provides another surprise. If it is the domain above Λ 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. Conversely, if the interaction strength increases at high energies, like in QED, this implies that perturbation theory will eventually break down. There are three known possibilities then.

The first is that this indicates that the theory at the quantum level is non-interacting, i. e. that quantum interferences averages out all interactions. In such a case the theory

is called trivial. This seems to be the case for QED. However, to show that this happens requires usually non-perturbative methods to cope with the strong increase in the interactions.

The second possibility is that just perturbation theory breaks down, and non-perturbative methods are necessary. This seems to be the case, e. g., for quantum gravity. The theory remains then well-defined.

The third is that just higher orders become relevant. E. g. the next order in the β -function in the same theory is

$$\beta_1 = \frac{34}{3}C_A^2 - \frac{20}{6}C_A N_f - 2C_F N_f.$$

It now depends on the actual value of the various quantities appearing, but this can change the sign of the β -function once more. In addition, there is also the possibility that the sum is vanishing. Then the running coupling achieves a finite value at the extremes of momenta ranges. This is called asymptotic safety. If this happens at the perturbative level, this is called a Banks-Zak fixed point. This behavior could also occur non-perturbatively, which is called a Wilson-Fisher fix point. The prior scenario is e. g. realized in some scenarios beyond the standard model in which the Higgs is composite, while the latter one seems to be realized in quantum gravity.

There is also the possibility that the β -function vanishes. In this case, the coupling does not run at all. Such a theory is called conformal. However, this also implies that in such a theory different scales all look the same. Thus, there is no distinction between high energies and low energies, and as a consequence no observable dynamics takes place. Certain supersymmetric theories have this property.

Similar equations like (5.54) 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.

5.5 Infrared singularities

There is one very generic feature of perturbative treatments of theories with massless particles. This is the appearance of infrared singularities, due to the propagators behaving as $1/q^2$. Note that formally this problem arises on the light-cone, but often the problem is treated after Wick rotation, where it indeed is purely infrared.

While in loop integrals this issue can often be dealt with by introducing a small mass as regulator, there is a situation where this becomes a more substantial problem. This is when considering initial states or final states. Evidently, an arbitrary number of massless

particles can be created at zero energy cost. At finite experimental energy resolution this is even possible for the emission of a number of finite, but low energy, particles. This is known as the infraparticle problem. A full non-perturbative solution requires to define effective particles, which include such a set of admixtures. However, a complete solutions has not yet been achieved.

There are two dimensions, in which the problem needs to be considered. One is kinematics. These kind of divergencies can arise from both the emission of a massless low-energy particle, as well as from a particle in which the emission angle vanishes, a so-called collinear singularity. The other is that emissions can be either virtual, i. e. in loops associated with external legs, or real, when the particle is emitted. In the latter case, this is especially relevant in combination with collinear emissions, as then they can experimentally not be distinguished.

Because of the rigid structures of perturbation theory, it is straightforward to treat these emissions systematically. They usually lead to large logarithms, i. e. logarithms, which behave as $\ln^n s/\Lambda^2$, so-called Sudakov logarithms. Here s is the typical energy scale squared, Λ a typical scale of the theory, and n some exponent, usually larger than 1. Thus, these corrections become very large at high energies, and can even surpass quantitatively terms suppressed by powers in the couplings. That would threaten the consistency of perturbation theory.

This issue seems to be particularly aggravated in gauge theories, where gauge symmetries necessarily add massless particles. However, here cancellations become important, which originate from gauge symmetry. What happens is that initial states or final states are gauge singlets, and a summation over all possible gauge states need to be performed. This includes also the gauge charges of every emitted particle, and has also impact on the exchange of gauge bosons in the final states, as they alter the gauge charges, at least in the non-Abelian case, of final state particles, and thus the summation.

The consequences are formulated in the Bloch-Nordsieck theorem (for initial states only) and in the Kinoshita-Lee-Naunberg theorem for simultaneously initial states and final states. They state that infrared singularities will cancel, when full gauge multiplets are summed over in initial states and initial states and final states, respectively. As a consequence, no large Sudakov logarithms arise. Thus, it is necessary that all initial states particles and final state particles are gauge singlets. However, this implies that for such cancellation to be possible those cannot be elementary particles. Hence, more complicated objects are required than the asymptotic states of perturbation theory, which leads to non-perturbative aspects.

Chapter 6

Quantization beyond perturbation theory

6.1 The path integral beyond perturbation theory

Perturbation theory can be cast in a multitude of different ways. There are, however, two very important features, which turn out to be crucial in seeing its limitation. One is that it is inherently a small field expansion. This is of less importance in the present context, but plays a role, e. g., in electroweak physics.

The other is that it assumes that correlation functions, and thus by the reconstruction theorem the partition sum, is (functionally) expandable in a Taylor series in the coupling constants and sources. As a consequence, the only asymptotically appearing states are the elementary ones, as discussed in section 4.5.2. That this cannot be the full story can be seen by a simple consideration: Consider QED with protons and electrons. The hydrogen atom is a stable state, but does not belong to the perturbative asymptotic state space, and should therefore not exist. This is, obviously, a contradiction.

The reason is that it can be shown that the state space of the free theory, to which perturbation theory is analytically connected by a smooth deformation, is not unitarily equivalent to the interacting theory. This is known as Haag's theorem, see also footnote 2 in chapter 5. Thus, the free theory cannot be analytically connected to the interacting one, and thus correlation functions are necessarily non-analytic functions of the parameters. Likewise, the quantum effective action does not need to be an analytic function of the sources.

Thus, to fully capture a theory requires more than just standard perturbation theory. This does not imply that these effects are necessarily quantitatively important for any question. In fact, many questions within the standard model have been treated entirely

perturbatively, and agree with experiment to enormous precision. But there are other cases, where the same approach fails miserably.

Thus, it is generically necessary to treat also gauge theories beyond perturbation theory. This induces a host of additional problems already in the quantization process, which will be discussed here. It should be noted that this is not a fully solved problem yet, and thus the following may change or need amendments in the future.

To understand it better, it is best to concentrate on the best-studied case, the Landau gauge. To understand how to solve the problem requires to better understand where it affects the perturbative quantization procedure in section 4.4. The problem is exactly located in the expression (4.30). It is obtained by the inversion of the operator Δ in (4.18). This object appears as the Faddeev-Popov determinant (4.31). Inverting it is only possible if, and only if, the Faddeev-Popov operator has no zero modes, i. e. in Landau gauge

$$\partial_\mu D_{ab}^\mu \omega^b = \lambda \omega^a,$$

where the functions ω^a appear in the gauge transformation (2.19) has no non-trivial solutions for $\lambda = 0$. But, e. g., (2.29) is exactly such a zero mode. This can be immediately seen, as for any gauge-transformed field A_μ^ω the Landau gauge condition evaluates infinitesimally as

$$\partial^\mu A_\mu^{a\omega} = \partial^\mu A_\mu^a + \partial^\mu D_\mu^{ab} \omega^b \quad (6.1)$$

and thus infinitesimally any Gribov copy is obtained from another Landau-gauge configuration exactly by a gauge transformation being a zero mode of the Faddeev-Popov operator. So far, it is unclear how many Gribov copies of any given field configuration exist, but any attempt at counting suggest that there may be an infinite number of copies.

Unfortunately, the problem becomes worse. It turns out that there are further Gribov copies which even have negative eigenvalues. These again arise only for large gauge transformations, and thus do not appear in perturbation theory. However, they imply cancellations in an expression like (4.30). Exactly these cancellations will ultimately modify which states are physical in non-Abelian gauge theories, and thus can be observed.

At this point, it is important to understand the scope of the problem: It is a practical problem. The basic idea of gauge fixing works just fine: Choosing a coordinate system. It is just that specifying one is more complicated than expected. Especially, the idea that a local condition, like (4.17), is insufficient, and there are many, probably infinitely many, coordinate systems satisfying this choice. Singer's result amplifies this problem by stating that a unique choice is not possible using a local condition in field space.

Thus, the problem is 'only' how to select a coordinate system, i. e. gauge, such that it is useful. Of course, just like covariant gauges (4.23) average eventually over all pos-

sible choices Λ , a first idea could be to just average over all Gribov copies. Such gauges are called Fujikawa-Hirschfeld gauges. While conceptually a possible solution the arising cancellations due to the sign of the Faddeev-Popov determinant has yet prevented any practical use of such a solution.

The alternative is to further tighten the perturbative gauge conditions by imposing further conditions. Of course, these conditions must still be such that at least one representative of every orbit can satisfy them¹. Showing that this is possible is occasionally already a non-trivial problem. So far, no final and proven solution exists. Therefore here an example, Landau gauge, will be discussed, which is among the best studied so far, to give a flavor of how to proceed.

To get a visual idea of what is going on, it is useful to imagine the field configuration space as a three-dimensional ordinary space. Then every gauge-orbit is a curve in this space, which do not intersect and fill the space densely. Every point along a curve is a full field configuration, and gauge transformations move along the curves. Imposing now a perturbative unique gauge condition, in the present example the Landau gauge condition (4.17), means to introduce a hypersurface into this space, which cuts every curve orthogonal². Because of the orthogonality any infinitesimal gauge transformation moves along the curves outside the hypersurface, and thus the corresponding gauge copies no longer satisfy the gauge condition.

Gribov copies arise now because the curves cut this perturbative gauge-fixing hypersurface multiple times, but always at finite distances on the curve and always orthogonal, and thus without violation of the perturbative gauge condition. The discrete, though possible denumerable infinite, set of intersection points makes up the residual gauge orbit. The aim must therefore be to add conditions, or a sequence of conditions, which remove these additional intersection points. For this it is useful to understand the structure of the hypersurface, which depends, of course, on the perturbative gauge condition.

As the appearance of Gribov copies is linked to the zero modes by (6.1), it is not surprising to find that these are also instrumental in structuring the hypersurface. The first Gribov region is defined by the requirement that the Faddeev-Popov operator (4.32) is strictly positive semi-definite, i. e., all of its eigenvalues are zero or positive. This region can be shown to be bounded in terms of the field amplitudes and convex, and the Faddeev-Popov operator has zero eigenvalues only on the boundary of this region, the so-called first

¹Various gauge conditions involving non-local information have been constructed over time. However, they are usually practical only in the numerical context of lattice simulations, and will therefore not be covered here, as the goal here are continuum methods. There are also some examples for the case of two dimensions, in which Yang-Mills theory is trivial and exactly solvable.

²In the context of general relativity, this is also known as a foliation.

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. 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 (4.32) 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 horizons, 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.

The first step is then to restrict the path integral to the first Gribov region. This restriction can be implemented using a functional θ -function in the perturbative gauge-fixed path integral (4.35)

$$\begin{aligned} \langle \mathcal{Q} \rangle &= \lim_{\xi \rightarrow 0} \int \mathcal{D}A_\mu \mathcal{D}c \mathcal{D}\bar{c} \theta(-\partial_\mu D_\mu^{ab}) \mathcal{Q}(A_\mu, c, \bar{c}) e^{-\int d^4x \mathcal{L}_g} \\ \theta(-\partial_\mu D_\mu^{ab}) &= \prod_i \theta(\lambda_i), \end{aligned} \quad (6.2)$$

where λ_i is the i th eigenvalue of the Faddeev-Popov operator (4.32). 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 (4.18) 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.

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 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 interest-

ing properties. Furthermore, no Gribov copy, or any gauge copy in general, is preferred compared to another. It would thus be completely legitimate to always choose 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 on the Gribov horizon. Though not proven, this appears very likely.

However, even after restricting to the first Gribov region, the remainders of the residual gauge orbits still possess a large number of Gribov copies. In particular, also on the horizon such additional copies arise, which make the replacement using a δ -function an even more non-trivial attempt.

This set of Gribov copies in the first Gribov region 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 point³. It is also in general non-trivial how to find all Gribov copies, so that they can be counted⁴.

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 there is no stochastic force acting along a gauge orbit, and thus in the stochastic equili-

³It 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.

⁴It 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.

bration 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

$$\begin{aligned}
 F[A] &= 1 - \frac{1}{V} \int d^d x A_\mu^a A_\mu^a \\
 \langle F[A] \rangle &= 1 - \frac{N_g}{2^d \pi^{d/2} \Gamma(1 + \frac{d}{2}) V} \int d p p^{d-1} D_\mu^{\mu a a}(p) \\
 D_{\mu\nu}^{ab} &= \langle A_\mu^a A_\nu^b \rangle,
 \end{aligned} \tag{6.3}$$

on each configuration, where $D_{\mu\nu}^{ab}$ is again the (full) gauge boson 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 (6.3) 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 (6.3) has multiple maxima.

An alternative way to choose a representative on the residual gauge orbit is then 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 (6.3) 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 the topological configurations to be discussed in chapter are located there. By construction, in this region the gauge boson 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 (6.3) 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 if the non-overlapping parts of the Gribov horizon are of measure zero.

Both these descriptions belong to a larger class of descriptions. It is obtained by rewriting the path integral in the expression (6.2) as

$$\langle \mathcal{O} \rangle = \lim_{\xi \rightarrow 0} \int \mathcal{D}A_\mu \mathcal{D}c \mathcal{D}\bar{c} \mathcal{O}(A_\mu, c, \bar{c}) \Theta(-\partial_\mu D_\mu^{ab}) e^{-\int d^4x \mathcal{L}_g} w(A_\mu, c, \bar{c}), \quad (6.4)$$

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

$$w_2 = \exp \left(\mathcal{N}_2 - \frac{\lambda_2}{V} \int d^d x A_\mu^a A_a^\mu \right). \quad (6.5)$$

where the \mathcal{N}_i are appropriately chosen normalizations and the limit of $\lambda_2 \rightarrow \infty$ has to be taken. The opposite limit $\lambda_2 \rightarrow 0$ recovers the minimal Landau gauge. These weights can also depend on other fields. Another possibility discussed in the literature is

$$w_1 = \exp \left(\mathcal{N}_1 + \frac{\lambda_1}{V} \int d^d x d^d y \partial_\mu^x \bar{c}^a(x) \partial_y^\mu c^a(y) \right) \quad (6.6)$$

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.

Even if (6.4) turns out to be well-defined, this is not yet sufficient for making it useful in practical calculations. The conditions (6.5) and (6.6) appear to translate into boundary conditions of, e. g., the Dyson-Schwinger equations (4.66). Thus, these equations would have multiple solutions, each corresponding to different gauges, and the corresponding boundary conditions will select among them. Conversely this implies that such equations for gauge-invariant quantities would have only a single solution, no matter what gauge-dependent quantities are input. It is at the current time not yet clear whether this is indeed how it works.

However, this also shows how the problem is vanishing in perturbation theory once more. The perturbative series is not a self-consistency equation like the Dyson-Schwinger equation, but a functional, i. e. creates a unique solution, and always has a non-negative Faddeev-Popov operator.

6.2 Non-perturbative BRST symmetry

As section 4.5 has shown BRST symmetry is instrumental in establishing the physical state space. However, the Gribov-Singer ambiguity also breaks perturbative BRST symmetry. This is once more best seen in Landau gauge. Consider the BRST transformations (4.36-4.39). The BRST transformations of the gauge field, (4.36), is an ordinary gauge transformation, which is just a bit strange because the gauge transformation function ξu^b is constructed from the gauge field. However, because the perturbative Landau gauge has a unique solution this transformation cannot be anything but the identity transformation for the gauge field. Thus, any gauge transformation which would change the gauge field would automatically break Landau gauge. Thus, the remaining transformations (4.37-4.39) may alter the ghost fields and the Nakanishi-Lautrup fields, but cannot do anything on the gauge field. This implies that if the ghost and the Nakanishi-Lautrup field would be integrated out, the whole path integral would become trivial under a BRST transformation, as the Faddeev-Popov determinant must also be invariant as it then only involves the gauge boson field.

This already immediately implies that Gribov copies modify the situation, as both the gauge field and the Faddeev-Popov determinant changes when changing from one Gribov copy to another. Especially, as these can be applied before introducing the ghosts and Nakanishi-Lautrup field, they act not differently than a BRST transformation, except that they are not perturbative BRST transformations.

On the other hand, since the BRST transformation should not alter the Landau gauge condition, it follows for the BRST transformed field that

$$0 = \partial^\mu (A_\mu^a + \delta_B A_\mu^a) = \partial^\mu \delta_B A_\mu^a = \partial^\mu D_\mu^{ab}(\lambda u_b)$$

and thus a valid BRST transformation in Landau gauge requires that the Faddeev-Popov operator has a zero mode, and thus breaks the perturbative quantization in section 4.4, just as expected from the Gribov-Singer ambiguity in section 2.6.2, and shows that the BRST transformation in Landau gauge really just transfers between different Gribov copies.

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

$$\phi_i \rightarrow \phi_i + \delta\phi_i \tag{6.7}$$

all correlation functions remain invariant as well. This implies that the correlation functions of operators not invariant under (6.7), e. g. $\langle\phi_i\rangle$, have to vanish identically. 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 symmetry⁵, 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. The trivial reason is just that no direction 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 equally likely. As

⁵Any 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.

a consequence, the correlation functions are no longer unbiased 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,

$$\begin{aligned} C_{MM} &= \langle MM \rangle \\ M &= \sum_i s_i \end{aligned} \tag{6.8}$$

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 Fujikawa-Hirschfeld gauge can be used to recover⁶ a well-defined non-perturbative BRST symmetry, obeying the same algebra (4.41-4.43) as the perturbative version. The BRST transformations (4.36-4.39) are, in fact, just an ordinary gauge transformation for the gauge boson fields. This becomes evident when the ghost fields are integrated out. Leaving matter aside, the integrand of the path integral only involves the gauge boson fields. The BRST symmetry is thus 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 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.

⁶The following is not mathematical rigorous, but there are possibilities, using lattice gauge theory, to make it much more so.

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 all⁷ possible targets of a BRST transformation, i. e. all Gribov copies. But this is just a Fujikawa-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. 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.

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 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. But by analogy it is likely that BRST symmetry then works like in the perturbative Landau gauge, i. e. its acts only on the ghost fields non-trivially.

However, this does not yet makes statements whether BRST symmetry also continues to define physical states. For this also the linearity of the asymptotic states in the elementary fields was necessary in section 4.5.2.4. Beyond perturbation theory, this is no longer

⁷It 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.

true, and asymptotic states can be composite. This requires to classify gauge-invariant states first.

6.3 Gauge-invariance beyond perturbation theory

As the previous discussion shows, it is not possible to just simply extend the construction of section 4.5.2 to identify physical degrees of freedom. Especially, beyond perturbation theory asymptotic states are not necessarily anymore just the elementary states, but can now be composite. Just think of the hydrogen atom in QED. Thus, the argument that there is no target available for the BRST transformation is no longer valid⁸. Especially, the transverse projected part of sA_μ^a is

$$\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) D^{\mu ab}(\lambda c^b) = gf^{abc} u^b \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) A^{\nu c},$$

which is only in the limit $g \rightarrow 0$ zero, and thus non-perturbatively the gauge field belongs to V_2 , rather than V_1 , and is thus unphysical, if BRST would work. Also, the basic assumption of perturbation theory that asymptotic fields can only be linear combinations of the elementary fields breaks down.

Thus, this requires to construct physical states in a different way. The obvious choice is to have only gauge-invariant states as physical states, as they are automatically also BRST invariant, as this is only a particular gauge transformation. The natural question is then, whether there exists a kind of 'gauge-invariant', or at least non-perturbatively BRST-invariant if BRST symmetry exists, version of the gauge bosons or of the matter fields. As it turns out, the answer to this is quite different for Abelian and non-Abelian gauge theories.

6.3.1 Abelian gauge theories and the Dirac phase

Consider first the Abelian gauge field, which transforms according to (2.7), $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu g$. If the gauge field is contracted with a transverse projector, this yields for the gauge-transformed field in momentum space

$$\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) A^{\mu'} = \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) (A^\mu + p^\mu g(p)) = \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) A^\mu.$$

⁸Alternatively, it can be shown that a unitary time evolution cannot yield free fields, and thus asymptotic states need to be interacting. This is once more Haag's theorem at work.

Thus, the transverse projected gauge-field is invariant under a gauge transformation. The longitudinal part behaves like

$$\frac{p_\mu p_\nu}{p^2} A^{\mu'} = \frac{p_\mu p_\nu}{p^2} A^\mu + \frac{p_\nu}{p^2} g(p)$$

and thus carries all the changes, but does also include (longitudinal) information from the gauge field.

To get rid of this part, define a physical photon field A_μ^P as

$$A_\mu^P = \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2} \right) A^\nu,$$

which is manifestly transverse, and thus has only two degrees of freedom, and is invariant under gauge transformations. Seems to be pretty simple, but there is a catch.

This catch appears when trying to invert the Fourier transform, and return to position space. This yields

$$\int d^4 p A_\mu^P(p) e^{ipx} = \int d^4 p \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2} \right) A^\nu(p) e^{ipx}.$$

Now, this is no longer a simple expression, because of the factor $1/p^2$. Rather, what is needed is an integral over all of space of $A_\mu(x)$,

$$\begin{aligned} A_\mu^P(x) &= g_{\mu\nu} \int d^4 y \delta(x-y) A^\nu(y) - \frac{1}{4\pi} \int d^4 y \frac{\partial_\mu \partial_\nu}{|x-y|} A^\nu(y) \\ &= g_{\mu\nu} A^\nu(x) - \frac{1}{4\pi} \int d^4 y \frac{\partial_\mu \partial_\nu}{|x-y|} A^\nu(y). \end{aligned} \quad (6.9)$$

Because

$$\begin{aligned} &\left(g_{\mu\nu} \int d^4 y \delta(x-y) - \frac{1}{4\pi} \int d^4 y \frac{\partial_\mu \partial_\nu}{|x-y|} \right) \partial^\nu g(y) \\ &= \partial_\mu g(x) - \frac{1}{4\pi} \int d^4 y \frac{\partial^2 \partial_\mu g(y)}{|x-y|} = \partial_\mu g - \partial_\mu g = 0, \end{aligned} \quad (6.10)$$

the so defined field is indeed gauge-invariant. However, it is not a local object anymore, as the second integral is not reduced. Thus passing to a physical vector field implies even in QED giving up locality. Note that the physical field is entirely transverse, as any other contraction of it vanishes by construction. Thus, it carries only two degrees of freedom (in four dimensions). In fact, as the non-local term is essentially a solution to the partial differential equations determining the vector potential at fixed electric and magnetic field, it is possible to regain an expression in terms of these fields.

Likewise, it is possible to construct an electron field with such a non-local construction. Define the Dirac phase

$$D(x) = \exp \left(-ie \int_x^\infty dy_\mu A_\mu(y) \right)$$

which transforms under a gauge transformation using

$$\exp \left(-ie \int_x^\infty dy_\mu \partial_\mu g(y) \right) = \exp(-ieg(x)) \quad (6.11)$$

just opposite to a matter field. Thus, given a matter field $\psi(x)$, the combination

$$\psi^P(x) = D(x)\psi(x) \quad (6.12)$$

creates a gauge-invariant, physical state, e. g. an observable electron.

This expression is again non-local, as the line integral starts at infinity and connects to some point x . This connection is called a Dirac string. It can be shown that the results of physical processes cannot depend on the detailed form of the Dirac string. However, it punctuates any sphere around x , and thus creates non-contractible paths. Hence, it can also not simply be dismissed. Such a situation, where non-locality is reduced to a single point on any sphere around some local position x is called almost local. Note that the Dirac phase acts like an inverse transporter in the sense of section 2.7. Thus, what the Dirac string actually does is compensating for the arbitrary changes of coordinate systems.

While the field (6.12) is thus the physical electron, and should be used in actual calculations, this can often be reduced to the usual perturbative electron. If the gauge field has a small enough amplitude such that the line integral is of order one, then the Dirac phase can be expanded in the electric coupling constant, yielding $1 + \mathcal{O}(e)$. Thus, any effects from this dressing are of one order higher in the coupling constant than when using the usual elementary field. Hence, in a purely perturbative construction the dressing can be neglected. However, if the line integral of the fields becomes large, of order $1/e$, this is no longer possible. Thus, at strong fields it is necessary to work with (6.12) even in a weakly coupled Abelian gauge theory.

Similar statements hold when considering the physical vector field (2.27). The integrated amplitude is small, if the field is well localized and has also locally a small amplitude. Then the gauge field can be approximated by a δ -function, for which the integral can be solved, yielding just the local field. Hence, the physical field becomes the unphysical field in the small amplitude limit.

6.3.2 Composite states as physical degree of freedoms

When turning to a non-Abelian gauge theories, the first idea would be to do also just (2.27) and (6.12). That is, unfortunately, not possible. The reason is that now the transformation formulas involve the covariant, rather than the ordinary, derivative. Thus, the calculation (2.28) does not work out, and involves an integral over the gauge field. Likewise, (6.11) does not work, as it involves now also an integral over the gauge field.

Of course, it is tempting to add further terms to the Dirac phase or consider a more involved kernel than $\partial_\mu \partial_\nu / x^2$ to counter these effect. At least in the sense of an infinite series, this appears to be possible at first. However, it turns out that this series requires an explicit inversion of the Faddeev-Popov operator. Because of the appearance of zero modes due to Gribov copies, see section 2.6.2, this is not possible. Except in perturbation theory, because there no such zero modes exist. Hence, in perturbation theory such a construction, in the sense of an asymptotic series, is possible. But only then. And in fact, in the same sense as before this reduces the physical states to the elementary ones in a perturbative expansion, and just yields results which are consistent with the BRST construction of chapter 4. Beyond perturbation theory, this is no longer possible.

Hence, beyond perturbation theory it is necessary⁹ to construct gauge-invariant states differently. The key to this are composite operators. Operators like

$$\mathcal{O}(x) = \phi^\dagger(x)\phi(x) \quad (6.13)$$

are manifestly gauge-invariant, even beyond perturbation theory. Likewise any operator build from combinations of the field strength tensor in a similar manner is gauge invariant, e. g. the action itself. Furthermore, so-called Wilson lines

$$W(\mathcal{C}_{x \rightarrow y}) = \exp \left(-ig \int_{\mathcal{C}_{x \rightarrow y}} dx_\mu A^\mu \right),$$

transform like

$$W(x, y) \rightarrow g(x)W(x, y)g(y)^{-1},$$

where $\mathcal{C}_{x \rightarrow y}$ is a path connecting the points x and y . Therefore bilocal objects

$$\mathcal{O}(x, y) = \phi^\dagger(y)W(y, x)\phi(x) \quad (6.14)$$

⁹Note that, despite overwhelming circumstantial evidence, there is no rigorous proof that no alternative exists. Thus, while unlikely, it cannot (yet) be excluded that an analogous construction to the Dirac phase is possible in the non-Abelian case.

are also gauge-invariant. Likewise more complicated objects can be build. Note that the important difference between the Wilson line and the Dirac phase is that even if either endpoint is send to infinity there is still a factor $g(\infty)$ under a gauge transformation. It is this fact that the Wilson line in a non-Abelian gauge theory transforms bilocal rather than local as the Dirac phase does, which differentiates between the Abelian case and the non-Abelian case.

Especially, this shows again how in perturbation theory the elementary fields become physical. To leading order in the coupling constant the Wilson line is again 1. Thus, an expression like (6.14) reduces to $\phi^\dagger(y)\phi(x)$, but the expectation value of this is merely the propagator of the elementary particle. Hence, extended gauge-invariant objects become independent elementary particles in perturbation theory.

This allows also to interpret (6.13) as being two particles at the same place. And this is only a localized bound state. Hence, such operators describe bound states.

Of course, determining the properties of such bound states, or bilocal or more complicated objects, then formally requires non-perturbative methods¹⁰. This will not be detailed here, but is relegated to the corresponding lectures, e. g. lattice field theory or functional methods. And, of course, if the bound state characteristic is only a minor disturbance, this can still be done perturbatively. E. g. at a couple of hundred of GeVs in both QCD and QED the corrections due to the Dirac phase and Wilson lines are negligibly small when it comes to cross sections. This can be intuitively understood¹¹: Shooting particles at each other at such energies they will pass each other so quickly that the bound state formation time is just far too long to have any relevance. Still, any field-theoretical sound approach should at least formulate the starting point of a calculation of any observable quantity first in terms of also non-perturbatively gauge-invariant quantities, and then systematically develop these expression when doing a perturbative calculation. This may lead to potentially different results then if just using the Feynman rules. E. g. the situation at LHC with colliding protons is described by the gauge-invariant initial protons, which are then decomposed into quarks by a technique called factorization, rather than with just three asymptotically non-interacting quarks. This would, e. g., totally miss the gluon content of the proton, which is actually the most relevant contribution at typical LHC energies.

¹⁰In the presence of a Brout-Englert-Higgs effect there is some way to circumvent this. See the lecture on electroweak physics for details

¹¹A word of caution: In other theories this statement may be misleading.

6.3.3 Gauge invariance as redundancy

The previous discussion should have shown that a physical problem needs to be formulated in a manifestly gauge-invariant way. Especially, gauge-dependent quantities cannot be used to characterize physical processes. In QED, it is explicitly possible to transform to such degrees of freedom by (2.27) and (6.12), while in non-Abelian gauge theories composite operators are necessary. Of course, also in QED the non-localities can be foregone by using composite operators only.

It is thus possible to formulate the problem in terms of gauge-invariant quantities. In QED, it is actually possible to transform the path integral into the variables (2.27) and (6.12), i. e. physical degrees of freedom. Of course, because of the Dirac string, this will no longer lead to a local Lagrangian. On the other hand, the Dirac string explains how a non-local effect like the Aharonov-Bohm effect can emerge, or how Bell's inequalities can be understood. Still, the resulting formalism is cumbersome at least, and technically practically impossible to use at worst. The situation is worse for the non-Abelian case. It appears possible to rewrite the path integral in terms of a sum over all possible closed Wilson loops, i. e. Wilson lines in which the path is closed. However, there is an infinity of such Wilson loops, and thus there would be an infinite number of integration variables in the path integral. Thus, while conceptually interesting, this is practically essentially irrelevant.

Still, both aspects together shows that gauge degrees of freedom are, indeed, just auxiliary fields, which can be removed. This is done at the expense of a non-local formulation of quantum theory, and thus essentially insurmountable technical complications at the current time. But it shows that gauge degrees of freedom are unphysical, and a problem needs to be formulated manifestly gauge-invariantly before gauge degrees of freedom can then be used to solve it in a technically convenient way.

6.3.4 Asymptotic state space

To complete the discussion, this leads to the question of observable states. In principle now all physically observable states need to be described by gauge-invariant operators. At the same time, gauge-dependent operators can still describe asymptotic states, but they will belong to an unphysical (gauge-dependent) state space. Thus, just like in section 4.5, the total space is split into two. The total space has an indefinite metric, and contains all gauge-dependent states. The gauge-invariant ones are in the physical subspace, which is a Hilbert space. In principle, this Hilbert space would be sufficient to describe all of physics, but at the expense of being governed by non-local dynamics. Gauge-dependent states

cannot evolve into gauge-invariant, and vice-versa, as the S-matrix itself is gauge-invariant, and can thus not change gauge-dependency. Thus, all observable states (and thus any particles) need to be gauge-invariant, especially in the sense of asymptotic states. In non-Abelian gauge theories, these are composite, and thus need in general a non-perturbative description, though in special cases this may not be quantitatively relevant.

6.4 Topology at the quantum level

6.4.0.1 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

$$\langle j|e^{-H\tau}|i\rangle = \sum_{N_+=0}^{\infty} \sum_{N_-=0}^{\infty} \frac{\delta_{N_+-N_--j+i}}{N_+!N_-!} (K\tau e^{-S_0})^{N_++N_-}$$

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

$$\langle j|e^{-H\tau}|i\rangle = \frac{1}{2\pi} \int_0^{2\pi} d\theta \sum_{N_+=0}^{\infty} \sum_{N_-=0}^{\infty} \frac{e^{i\theta(N_+-N_--j+i)}}{N_+!N_-!} (K\tau e^{-S_0})^{N_++N_-}$$

This can be rearranged as

$$\begin{aligned} \langle j|e^{-H\tau}|i\rangle &= \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{i\theta(i-j)} \sum_{N=0}^{\infty} \frac{2^N \cos^N \theta}{N!} (K\tau e^{-S_0})^N \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{i\theta(i-j)} e^{2K\tau e^{-S_0} \cos \theta} \end{aligned}$$

From this can be read of, that the vacuum state wave functions are $e^{i\theta n} |n\rangle$, and the energy is $E(\theta) = -2K e^{-S_0} \cos \theta$. The full vacuum state, the θ vacuum, is

$$|\theta\rangle = \sum_n e^{i\theta n} |n\rangle \tag{6.15}$$

The lowest energy is negative, as expected, since tunneling lowers the ground state energy. There are nevertheless no excitations 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

$$L = \frac{i\theta}{32\pi^2} G_{\mu\nu}^a \Gamma_{\mu\nu}^a.$$

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.

6.4.0.2 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 so-called Pauli-Villars scheme, which introduces a scale M by adding fictional particles of large mass and unphysical signs in propagators and loops. Some more calculations then yields

$$\frac{d^5 n_I}{d\rho d^4 z} \sim \frac{1}{\rho^5} \exp \left(-\frac{8\pi^2}{g^2} + \frac{11}{3} N_c \log(M\rho) \right)$$

where the prefactor cancels. The complete result is

$$\frac{dn_I}{d\rho d^4 z} = \frac{0.466 e^{-1.679 N_c}}{(N_c - 1)! (N_c - 2)!} \left(\frac{8\pi^2}{g^2} \right)^{2N_c} \frac{1}{\rho^5} e^{-\frac{8\pi^2}{g^2} + \frac{11}{3} N_c \log(M\rho)} \quad (6.16)$$

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.

6.4.1 Light fermions

The presence of different vacua can be connected to the axial charge, i. e. the charge associated with axial symmetry, transformation of spinor fields by $\exp(i\alpha\gamma_5)$.

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

$$\partial_\mu j_\mu^5 = \frac{N_f}{16\pi^2} G_{\mu\nu}^a \Gamma_{\mu\nu}^a \quad (6.17)$$

where the gauge fields are arbitrary background fields. This is essentially an anomaly, which will be discussed in more detail in section 7. The surface term on the right hand side is relevant, since the vacua 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. E. g. in QCD the consequence is that the ninth Goldstone boson η' acquires even in the chiral limit a mass. Another consequence is that in the presence of light fermions, single instantons cannot exist (although there is some exception possible), but correlated instantons are necessary, even in the limit of a dilute gas of instantons. Secondly, fermions 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, spontaneous chiral symmetry breaking can be viewed as a consequence of the presence of instantons.

Chapter 7

Anomalies

7.1 Introduction

There is one particular important property of the standard model, which is very much restricting its structure, and which is recurring in extensions of the standard model. That is the absence of anomalies. 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. Constructing a theory which is at the same time anomaly-free and consistent with the standard model is actually already quite restricting, and therefore anomalies are an important tool to check the consistency of new proposals for physics beyond the standard model. This will be therefore discussed here in some detail.

7.2 Global anomalies

Anomalies fall into two classes, global and local anomalies. Global anomalies refer to the breaking of global symmetries by quantum effects. The most important one of these global anomalies is the breaking of dilatation symmetry. This symmetry corresponds to rescaling all dimensionful quantities, e. g., $x \rightarrow \lambda x$. Maxwell theory, massless QED, Yang-Mills theory, and massless QCD are all invariant under such a rescaling, at the classical level, though not the Higgs sector of the standard model. This is no longer the case at the quantum level. By the dimensional transmutation, surfacing in the renormalization process of section 5.2, 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 photon is massless in massless QED. Of course, it is also massless in massive QED, but there additional explicit breaking of the dilatation symmetry is added due to the lepton mass.

Another example is the so-called axial, or chiral, anomaly, which occurs due to the breaking of the global axial symmetry of baryons. A consequence of it is the anomalously large η' mass. While the dilatation anomaly is quite obvious, the axial anomaly is much more subtle, and therefore deserves some more discussion. In addition, it will be very helpful when generalizing to the local anomalies.

7.2.1 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 group G with generators T and gauge fields in the adjoint representation. The fermionic part of the Lagrangian is then given by

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu(\partial_\mu - igT^a A_\mu^a) - m)\psi = \bar{\psi}(i\gamma_\mu D^\mu - m)\psi$$

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

$$j_\mu^a = \bar{\psi}\gamma_\mu T^a \psi.$$

Due to the axial symmetry, there is also a corresponding axial current

$$j_\mu^{5a} = \bar{\psi}\gamma_5\gamma_\mu T^a \psi.$$

In addition, there are also the singlet currents

$$\begin{aligned} j_\mu &= \bar{\psi}\gamma_\mu \psi \\ j_\mu^5 &= \bar{\psi}\gamma_5\gamma_\mu \psi, \end{aligned}$$

which corresponds to the fermion current and the axial current.

Naively, the divergences of these equations can be calculated using the Dirac equation.

$$\begin{aligned} \partial^\mu j_\mu^a &= -i\bar{\psi}(g\tau^b\gamma_\mu A_b^\mu - m)\tau^a\psi - i\bar{\psi}\tau^a(-g\tau^b\gamma_\mu A_b^\mu + m)\psi \\ &= ig\bar{\psi}[\tau^a, \tau^b]\gamma_\mu A_b^\mu\psi = -gf^{abc}A_b^\mu\bar{\psi}\gamma_\mu\tau_c\psi = -gf_c^{ab}A_b^\mu j_\mu^c. \end{aligned}$$

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

$$D_\mu^{ab} j_b^\mu = 0. \quad (7.1)$$

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

$$D_\mu^{ab} j_b^\mu = 2im\bar{\psi}\gamma_5\tau^a\psi = 2imp^a, \quad (7.2)$$

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

$$\begin{aligned} \partial_\mu j^\mu &= 0 \\ \partial^\mu j_\mu^5 &= 2im\bar{\psi}\gamma_5\psi = 2imp^0. \end{aligned}$$

Hence, the number of fermions is, as expected, a conserved current. The axial current is only conserved for massless fermions. This is the result that axial, and chiral, symmetry gets explicitly broken, already classically, by a mass term.

In a theory like the standard model, where parity is broken, left-handed and right-handed fermions

$$\begin{aligned} \psi_L &= \frac{1 - \gamma_5}{2}\psi \\ \psi_R &= \frac{1 + \gamma_5}{2}\psi \end{aligned}$$

do not couple in the same way to the gauge-fields

$$\mathcal{L} = \bar{\psi}_L i\gamma_\mu D_L^\mu \psi_L + \bar{\psi}_R i\gamma_\mu D_R^\mu \psi_R,$$

with $D_L \neq D_R$, and no mass term is permitted due to gauge invariance. Thus, the gauge currents are recombined into covariantly conserved left-handed and right-handed currents as

$$\begin{aligned} j_\mu^{aL} &= \frac{1}{2}(j_\mu^a - j_\mu^{5a}) \\ j_\mu^{aR} &= \frac{1}{2}(j_\mu^a + j_\mu^{5a}) \\ D_L j_\mu^L &= 0 \\ D_R j_\mu^R &= 0, \end{aligned}$$

and a similar recombination for the singlet currents.

7.2.2 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 the Ward-Takahashi identities of section 4.7. In particular, take Ward identities for correlation functions of the form

$$T_{\mu\nu\rho}^{ijk} = \langle T j_{\mu}^i j_{\nu}^j j_{\rho}^k \rangle,$$

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 Lorentz index is dropped in the last case. Calculating the corresponding identities for a local¹ 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

$$\partial_x^\mu T_{\mu\nu\rho}^{VVA}(x, y, z) = \partial_y^\nu T_{\mu\nu\rho}^{VVA}(x, y, z) = 0 \quad (7.3)$$

$$\partial_z^\rho T_{\mu\nu\rho}^{VVA}(x, y, z) = 2m T_{\mu\nu}^{VVP}(x, y, z), \quad (7.4)$$

directly implementing the relations (7.1) and (7.2). 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 group generators 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

$$\begin{aligned}T_{\mu\nu\rho}^{VVA}(p_1, p_2, p_3 = -p_1 - p_2) = & \quad (7.5) \\ -i^3 \int \frac{d^4k}{(2\pi)^4} & \left(\text{tr} \gamma_\mu (\gamma_\alpha k^\alpha - m)^{-1} \gamma_\nu (\gamma^\beta k_\beta - \gamma_\beta p_2^\beta - m)^{-1} \gamma_\rho \gamma_5 (\gamma_\gamma k^\gamma + \gamma_\gamma p_1^\gamma - m)^{-1} \right. \\ & \left. + \text{tr} \gamma_\nu (\gamma_\alpha k^\alpha - m)^{-1} \gamma_\mu (\gamma^\beta k_\beta - \gamma_\beta p_1^\beta - m)^{-1} \gamma_\rho \gamma_5 (\gamma_\gamma k^\gamma + \gamma_\gamma p_2^\gamma - m)^{-1} \right).\end{aligned}$$

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.

¹This is a technical trick, and in the end $\beta(x)$ will be taken to be x -independent.

Here, it permits to use a Pauli-Villar regulator with a mass M , which is technically more simple than other possibilities. A Pauli-Villar regulator is based essentially on subtracting the same diagram with opposite sign, but all masses replaced by very large masses M , which eventually will be send to infinity in the renormalization process. Dimensional regularization would be more subtle, as it depends on the way the matrix γ_5 is analytically continued. This problem does not appear for Pauli-Villar regularization, as the calculation remains in four dimensions.

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

$$\begin{aligned} p_1^\mu T_{\mu\nu\rho}^{VVA}(p_1, p_2, p_3 = -p_1 - p_2) = & \quad (7.6) \\ -i^3 \int \frac{d^4 k}{(2\pi)^4} & \left(\text{tr} - (\gamma_\alpha k^\alpha - m)^{-1} \gamma_\nu (\gamma^\beta k_\beta - \gamma_\beta p_2^\beta - m)^{-1} \gamma_\rho \gamma_5 \right. \\ & \text{tr}(\gamma_\gamma k^\gamma + \gamma_\gamma p_1^\gamma - m)^{-1} \gamma_\nu (\gamma^\beta k_\beta - \gamma_\beta p_2^\beta - m)^{-1} \gamma_\rho \gamma_5 \\ & + \text{tr}(\gamma_\gamma k^\gamma + \gamma_\gamma p_2^\gamma - m)^{-1} \gamma_\nu (\gamma_\alpha k^\alpha - m)^{-1} \gamma_\rho \gamma_5 \\ & \left. + \text{tr} - (\gamma_\gamma k^\gamma + \gamma_\gamma p_2^\gamma - m)^{-1} \gamma_\nu (\gamma^\beta k_\beta - \gamma_\beta p_1^\beta - m)^{-1} \gamma_\rho \gamma_5 + (m \rightarrow M) \right). \end{aligned}$$

This rather lengthy expression is now a finite integral. It is therefore permissible to reshuffle the momenta like $k \rightarrow k + p_2$ in the first term and $k \rightarrow k + p_2 - p_1$ 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 (7.3) works in the same way.

The situation changes drastically for the axial Ward identity (7.4). The expression (7.5) is still divergent, so before doing anything, it will again be regulated using a Pauli-Villar regulator, to make it well-defined. To evaluate (7.4) requires multiplication with $p_3 = -p_1 - p_2$, which can be 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

$$\begin{aligned}
p_3^\rho T_{\mu\nu\rho}^{VV A}(p_1, p_2, p_3 = -p_1 - p_2) &= 2i \int \frac{d^4 k}{(2\pi)^4} \\
&\left(m \text{tr} \left(\gamma_\mu (k_\alpha \gamma^\alpha - m)^{-1} \gamma_\nu (\gamma_\beta k^\beta - \gamma_\beta p_2^\beta - m)^{-1} \gamma_5 (\gamma_\gamma k^\gamma + p_1^\gamma \gamma_\gamma - m)^{-1} \right) \right. \\
& m \text{tr} \left(\gamma_\mu (k_\alpha \gamma^\alpha - m)^{-1} \gamma_\nu (\gamma_\beta k^\beta - \gamma_\beta p_1^\beta - m)^{-1} \gamma_5 (\gamma_\gamma k^\gamma + p_2^\gamma \gamma_\gamma - m)^{-1} \right) \\
& M \text{tr} \left(\gamma_\mu (k_\alpha \gamma^\alpha - M)^{-1} \gamma_\nu (\gamma_\beta k^\beta - \gamma_\beta p_2^\beta - M)^{-1} \gamma_5 (\gamma_\gamma k^\gamma + p_1^\gamma \gamma_\gamma - M)^{-1} \right) \\
& \left. M \text{tr} \left(\gamma_\mu (k_\alpha \gamma^\alpha - M)^{-1} \gamma_\nu (\gamma_\beta k^\beta - \gamma_\beta p_1^\beta - M)^{-1} \gamma_5 (\gamma_\gamma k^\gamma + p_2^\gamma \gamma_\gamma - M)^{-1} \right) \right)
\end{aligned}$$

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

$$\text{tr} \gamma_\mu \gamma_\nu \gamma_\rho \gamma_\sigma \gamma_5 = -4i \epsilon_{\mu\nu\rho\sigma}. \quad (7.7)$$

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 (7.6) 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 (7.5).

The term involving the regulator can then be easily 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

$$\begin{aligned}
ip_3^\rho T_{\mu\nu\rho}^{VV A}(p_1, p_2) &= 2mi T_{\mu\nu}^{VVP}(p_1, p_2) + \lim_{M \rightarrow \infty} 8iM^2 \epsilon_{\mu\nu\rho\sigma} p_\rho^1 p_\sigma^2 \times \frac{i}{16\pi^2} \frac{-1}{2M^2} \\
&= 2mi T_{\mu\nu}^{VVP}(p_1, p_2) + \frac{1}{2\pi^2} \epsilon^{\mu\nu\rho\sigma} p_\rho^1 p_\sigma^2
\end{aligned} \quad (7.8)$$

Thus, the Ward identity (7.4) is violated. The the value of 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 same holds true for the charged version, and the only thing changing is the appearance of a prefactor d^{abc} of the coupling matrices T^a in charge space, which turn out to be

$$d^{abc} = \frac{1}{2} \text{tr} (\{T^a, T^b\} T^c), \quad (7.9)$$

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 (7.6). 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

$$\partial^\mu j_\mu^a = \frac{f_\pi}{\sqrt{2}} M_\pi^2 \pi^a, \quad (7.10)$$

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.

7.3 Local anomalies

In contrast to the global anomalies, the local anomalies are a more severe problem. A local anomaly occurs, when a quantum effect breaks a local gauge symmetry. The consequence of this would be that observable quantities depend on the gauge, and therefore the theory makes no sense. Thus, such anomalies may not occur. There are two possibilities how such anomalies can be avoided. One is that no such anomalies occurs, i. e., the path integral measure must be invariant under the symmetry. The second is by anomaly cancellation, i. e., some parts of the measure are not invariant under the symmetry, but the sum of all such anomalous terms cancel. It is the latter mechanism which makes the standard model anomaly-free. However, the price to pay for this is that the matter content of the standard model has to follow certain rules. It is thus rather important to understand how this comes about. Furthermore, any chiral gauge theory beyond the standard model faces similar, or even more severe, problems.

Already the classical result (7.2) is indicating that the current is only covariantly conserved. The latter equation implies that only for massless fermions there will be no gauge anomaly. However, this is not a problem, as only zero-mass fermions are admitted

to the standard model anyway, and all apparent fermion masses are generated by the Brout-Englert-Higgs effect and the strong interactions dynamically. But for the standard model this is still modified. Due to the parity violation, it is necessary to consider a current for left-handed and right-handed fermions separately, where the corresponding left-handed and right-handed covariant derivatives for the left-handed and right-handed currents appear.

In principle, it is possible to do the same one-loop calculation in a gauge theory, and the final result is quite similar. However, it may still be questioned whether this is an artifact of perturbation theory. It is not, and to show this it is useful to derive the local anomaly for gauge theories using a different approach. In a path integral approach, this becomes particularly clear, as it can be shown that the anomaly stems from the fact that the path-integral measure for fermions, $\mathcal{D}\psi\mathcal{D}\bar{\psi}$ is not invariant under chiral gauge transformations, and therefore the anomaly arises. It is, of course, invariant under vectorial gauge transformations, and thus theories like QCD need not to be considered, as will be confirmed below. This also shows that the anomaly is a pure quantum phenomenon, as the measure is part of the quantization process. In this course, it will also be shown that the global anomaly remains beyond perturbation theory in much the same way, and that the one-loop result is actually already the full answer.

7.3.1 Anomalies as a quantum effect

To see that this is a relevant effect, it is important to remember how Ward-Takahashi identities are obtained in general. As discussed in section 4.7, any well-defined symmetry transformation should leave the partition function unchanged, i. e.

$$0 = \delta Z = \delta \int \mathcal{D}\phi e^{iS+i\int d^4x j\phi}, \quad (7.11)$$

where ϕ is for simplicity a non-Grassmann field, which changes under the transformation as $\phi \rightarrow \phi + \epsilon f(\phi, x)$, with f some arbitrary function and ϵ infinitesimal. Performing the variation yields

$$0 = \int \mathcal{D}\phi e^{iS+i\int d^4x j\phi} \int d^4x \left(i \left(\frac{\delta S}{\delta \phi} + j \right) f + \frac{\delta f}{\delta \phi} \right), \quad (7.12)$$

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-Takahashi identities. The third term is new in a quantum theory, and gives the contribution of the Jacobian,

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

This is a genuine quantum contribution. This was dropped so far, but here it will become the source of the anomaly. Here it also becomes evident that the term anomaly is actually a misnomer. There is nothing anomalous about them. They are just a quantum effect. 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.

7.3.2 Full expression for the anomaly

To check this, rotate first to Euclidean time, by replacing $t \rightarrow it$ 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. The following uses now many of the methods already encountered in chapter.

Start by expanding the fermion fields in orthonormal eigenfunctions ψ_n of the Dirac operator,

$$\begin{aligned}\psi(x) &= \sum_n a_n \psi_n(x) \\ \bar{\psi}(x) &= \sum_n \psi_n^\dagger(x) \bar{b}_n,\end{aligned}$$

which satisfy

$$i\gamma_\mu D^\mu \psi_n = \lambda_n \psi_n \quad (7.13)$$

$$-i\gamma_\mu D^\mu \psi_n^\dagger = \lambda_n \psi_n^\dagger. \quad (7.14)$$

This permits to rewrite the path integral as an infinite product of integrations over the coefficients,

$$\mathcal{D}\psi \mathcal{D}\bar{\psi} = \Pi_m da_m d\bar{b}_m, \quad (7.15)$$

keeping in mind that these differentials are Grassmannian.

Now, a local chiral transformation $\beta(x)$

$$\psi \rightarrow e^{i\beta(x)\gamma_5} \psi,$$

then corresponds to a linear transformation of the coefficients

$$a_m \rightarrow C_{mn} a_n = a'_m,$$

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

$$\frac{1}{(\det C)^2} = e^{-2\text{tr} \ln C} = e^{-2\text{tr} \delta C}, \quad (7.16)$$

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

$$a'_m = \int d^4x \psi_m^\dagger (1 + i\beta\gamma_5) \psi_n a_n = (1 + \delta c_{mn}) a_n. \quad (7.17)$$

Inserting this result into (7.16) yields for the Jacobian of the infinitesimal transformation

$$J = \exp \left(-2i \int d^4x \beta \psi_m^\dagger \gamma_5 \psi_m \right), \quad (7.18)$$

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

$$\psi_m^\dagger \gamma_5 \psi_m \rightarrow \lim_{\tau \rightarrow 0} \psi_m^\dagger \gamma_5 e^{-\lambda_m^2 \tau} \psi_m, \quad (7.19)$$

where the limit has to be performed at the end of the calculation only. Expanding the Gaussian and using the relations (7.13-7.14), this expression can be rewritten as

$$\lim_{\tau \rightarrow 0} \psi_m^\dagger \gamma_5 e^{-\lambda_m^2 \tau} \psi_m = \lim_{\tau \rightarrow 0} \text{tr} \left(\gamma_5 e^{-\tau(\gamma_\mu D^\mu)^\dagger \gamma_\nu D^\nu} \right). \quad (7.20)$$

The exponential can be rewritten as

$$(\gamma_\mu D^\mu)^\dagger \gamma_\nu D^\nu = -D_\mu D^\mu + \frac{i}{4} [\gamma^\mu, \gamma^\nu] F_{\mu\nu}^a \tau_a. \quad (7.21)$$

The limit is still ill-defined. It is necessary to regularize the expression in a suitable way. This is achieved by the so-called 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

$$(\partial_\tau + \Delta_x)G(x, y, \tau) = 0 \quad (7.22)$$

$$G(x, y, 0) = \delta(x - y). \quad (7.23)$$

Which is solved by the formal expression

$$G(x, y, \tau) = e^{-\Delta_x \tau} = \sum_m e^{-\tau \lambda_m} \psi_m^\dagger(y) \psi_m(x).$$

This is already the expression (7.20). Without proof, it can now be shown that this heat kernel can be expanded for small τ as

$$G(x, y, \tau) \rightarrow_{\tau \rightarrow 0} \frac{1}{(4\pi\tau)^2} \exp^{-\frac{(x-y)^2}{4\tau}} \sum_{j=0}^{\infty} a_j(x, y) \tau^j.$$

Inserting this expansion into (7.18) yields

$$\ln J = -2i \lim_{\tau \rightarrow 0} \frac{1}{(4\pi\tau)^2} \int d^4 x \beta \sum_j \tau^j \text{tr} \gamma_5 a_j.$$

For $\tau \rightarrow 0$, the first term does not contribute, as a_0 has to be equal to one because of the condition (7.23). 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 satisfies (7.22) a so-called descent equation

$$-\Delta a_{j-1} = j a_j.$$

Since $a_0 = 1$, a_1 can be obtained algebraically from (7.21). 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 the anomaly

$$J = \exp \left(-\frac{i}{32\pi^2} \int d^4 x \beta \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^a \right). \quad (7.24)$$

Hence, the Jacobian is non-trivial, and will contribute in the Ward identities (4.73). 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

$$\int d^4 x \text{tr} \epsilon^{\mu\nu\rho\sigma} \partial_\mu \left(i A_\nu^a \partial_\rho A_\sigma^a + \frac{2}{3} f^{abc} A_\nu^a A_\rho^b A_\sigma^c \right). \quad (7.25)$$

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 anomaly 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 then just an explicit calculation, inserting the Jacobian (7.24) into the Ward-Takahashi identity (4.73). This will yield (7.8) with (7.9) inserted.

7.3.3 Anomaly cancellation

For the standard model it is most interesting to consider the case that left-handed and right-handed fermions are coupled with different gauge fields. Due to the different sign of γ_5 in the corresponding projector, this will reemerge as a different sign of the anomaly, yielding

$$k^\rho T_{\mu\nu\rho}^{V^a V^b A^c}(p, q, k) = 2m T_{\mu\nu}^{V^a V^b P^c}(p, q, k) + \frac{\text{tr}\{\tau_L^a, \tau_L^b\} \tau_L^c - \text{tr}\{\tau_R^a, \tau_R^b\} \tau_R^c}{2} \frac{1}{3\pi^2} \epsilon_{\mu\nu\rho\sigma} p^\rho q^\sigma,$$

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 from this expression: the left-hand side should vanish, if there is no massless pseudo-scalar particle in the theory, which is true for the standard model. On the right-hand side, the first term will indeed do so, if the fermion mass is zero. This is already required due to parity violation in the standard model. But for the second term this is not obvious.

There are now two possibilities how to obtain an anomaly-free theory. Either, the theory is anomaly-free, if each of the remaining terms is individually zero, or they cancel. Indeed, the expression $\text{tr}\{\tau^a, \tau^b\} \tau^c$, the so-called symmetric structure constant, is zero for all (semi-)simple Lie groups, except for $SU(N \geq 3)$ and $U(1)$. Unfortunately, these are precisely those appearing in the standard model, except for the $SU(2)$ of weak isospin. For the group $SU(3)$ of QCD, this is actually not a problem, since QCD is vectorial, and thus² $\tau_L = \tau_R$, and the terms cancel for each flavor individually. Thus remains only the part induced by the hypercharge.

In this case, each generation represents an identical contribution to the total result, as the generations are just identical copies concerning the generators. It is thus sufficient to consider one generation. The right-handed contributions are all singlets under the weak isospin, and thus they only couple vectorially to electromagnetism, and therefore yield

²Actually, unitarily equivalent is sufficient.

zero. The contributions from the left-handed doublet contain then the generators of the weak isospin, τ^a , and the electric charge $Q = \tau^3 + \mathbf{1}y/2$. The possible combinations contributing are

$$\text{tr}t^a\{\tau^b, \tau^c\} \quad (7.26)$$

$$\text{tr}Q\{\tau^a, \tau^b\} \quad (7.27)$$

$$\text{tr}\tau^a Q^2 \quad (7.28)$$

$$\text{tr}Q^3. \quad (7.29)$$

The contribution (7.26) vanishes, as this is a pure SU(2) expression. The term (7.29) is not making a difference between left and right, and is therefore also vanishing. It turns out that (7.27) and (7.28) lead to the same result, so it is sufficient to investigate (7.28). Since the isospin group is SU(2), the anti-commutator of two Pauli matrices just gives a Kronecker- δ times a constant, yielding in total

$$\text{tr}Q\{\tau^a, \tau^b\} = \frac{1}{2}\delta^{ab} \sum_f Q_f,$$

where Q_f is the electric charge of the member f of the generation in units of the electric charge. It has to vanish to prevent any gauge anomaly in the standard model, which is fulfilled:

$$\sum_f Q_f = (0 - 1) + N_c \left(\frac{2}{3} - \frac{1}{3} \right) = -1 + \frac{N_c}{3} = 0.$$

Therefore, there is no gauge anomaly in the standard model. However, this is only possible, because the electric charges have certain ratios, and the number of colors N_c is three. This implies that the different sectors of the standard model, the weak isospin, the strong interactions, and electromagnetism, very carefully balance each other, to provide a well-defined theory. Such a perfect combination is one of the reasons to believe that the standard model is part of a larger theory, which imposes this structure.

7.4 Relation to topology

There is an interesting twist for the quantity making up the Jacobian

$$\frac{1}{64\pi^2} \int d^4x \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^a = -\frac{i}{512\pi^4} \int d^4x \text{tr} \epsilon^{\mu\nu\rho\sigma} \partial_\mu \left(i A_\nu^a \partial_\rho A_\sigma^a + \frac{2}{3} f^{abc} A_\nu^a A_\rho^b A_\sigma^c \right)$$

Evidently, is again the topological charge (3.10) of the gauge field configuration. Since this quantity was obtained in the context of anomalies 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. The 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,

$$n^+ - n^- = \int d^4x \sum_{m, \lambda_m=0} \psi_m^\dagger \gamma_5 \psi_m.$$

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 7.3.2. 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 (7.19), and thus this will lead to the same result as in section 7.3.2. Thus, the final answer is

$$n^+ - n^- = k = \frac{1}{64\pi^2} \int d^4x \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^a$$

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 chiral 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 gauge formulation.

7.5 Witten anomalies

There is actually a further possible anomaly for fermions, the so-called Witten anomaly, which is also connected to the parity violation. It is also a gauge anomaly, and has therefore to be canceled as well. This occurs if the number of Weyl fermion states is even, as is the case, e. g., in the standard model. This would not be the case, if, e. g., there would be a single triplet of fermions charged under the weak isospin. In technicolor theories, or other theories beyond the standard model, this is a constraint, as in such theories multiplets with an odd number of fermions may appear, e. g. when the chirally coupled fermions are additionally charged under different gauge groups or representations, leading to an odd number of fermions. This has then to be canceled by additional fermions. This is a problem exclusively applying to the $\text{Sp}(N)$ gauge groups, and to $\text{SU}(2)$ of the weak interactions because $\text{SU}(2) \approx \text{Sp}(1)$, as well as $\text{O}(N < 6)$ groups, except for $\text{SO}(2)$.

The reason can be most easily illustrated by considering the path-integral with the fermions integrated out. For n Weyl fermions, the expression is

$$Z = \int \mathcal{D}A_\mu (\det i\gamma_\mu D^\mu)^{\frac{n}{2}} e^{iS}, \quad (7.30)$$

with S the usual gauge-field action. The problem arises, as it can be proven that for each gauge-field configuration of a gauge theory with an affected gauge group there exists a gauge-transformed one such that

$$(\det i\gamma_\mu D^\mu)^{\frac{1}{2}} = -(\det i\gamma_\mu D^{\mu'})^{\frac{1}{2}},$$

where $'$ denotes gauge-transformed. The proof is somewhat involved, but essentially boils down to the fact that the determinant has to be defined in terms of a product of eigenvalues. For $\text{Sp}(N)$ gauge theories as well as the groups $\text{O}(N < 6)$ it can then be shown that there exist gauge-transformations, which are topologically non-trivial, such that one of the non-zero eigenvalues changes sign. Mathematically, the reason is that the fourth homotopy group of these groups is non-trivial and actually is \mathbb{Z}_2 or \mathbb{Z}_2^2 . Hence, the integrand of the path integral (7.30) exists twice on each gauge orbit, but with opposite signs. Thus, the partition function vanishes, and all expectation values become ill-defined $0/0$ constructs. Thus, such a theory is ill-defined, as there is no continuous deformation of the gauge group possible to introduce a suitable definition, similar to L'Hospital's rule.

In the standard model, the problem does not arise, because the number of Weyl flavors of the fermions is even since only Dirac fermions appear. One could also hope that, since the gauge group of the standard model is actually $\text{S}(\text{U}(3) \times \text{U}(2)) \approx \text{SU}(3)/\mathbb{Z}_3 \times \text{SU}(2)/\mathbb{Z}_2 \times \text{U}(1)$, this problem would not arise. The reason for this division is that only for this particular

gauge group the matter field representation becomes single-valued, as is necessary for them to be meaningful. However, because $SU(2)/Z_2 \approx SO(3)$ instead of $Sp(2)$, this does not help, as the fourth homotopy group of $SO(3)$ is also non-trivial, and the problem persists.

Thus, adding further sectors to the standard model, or embedding it in a grand-unified theory, must respect this fact, to avoid triggering the Witten anomaly.

7.6 Anomalies in the BRST framework

In the context of the antifield construction of section 4.6, the anomalies play again a role as possible obstructions to the existence of a solution W to the quantum master equation (4.68).

Analyzing the obstructions to the existence of a solution to the quantum master equation can be done most easily from a direct \hbar expansion. To order \hbar^0 the quantum master equation yields

$$\frac{1}{2}(S, S) = 0, \quad (7.31)$$

which is the classical master equation. This equation is certainly fulfilled, since there is no obstruction to the existence of S .

To next order \hbar , the quantum master equation yields

$$sM_1 = (S, M_1) = i\Delta S. \quad (7.32)$$

Given the \hbar^0 -term S , this equation has a solution for M_1 if $s\Delta S = (S, \Delta S) = 0$. This condition is necessary but in general not sufficient. To prove that the condition $(S, \Delta S) = 0$ holds, note that

$$\Delta(\alpha, \beta) = (\Delta\alpha, \beta) - (-1)^{\epsilon_\alpha}(\alpha, \Delta\beta), \quad (7.33)$$

which is stated here without proof, but follows immediately from the properties of Δ . This property uses $\Delta^2 = 0$ and the generalized Leibniz rule

$$\Delta(\alpha\beta) = (\Delta\alpha)\beta + (-1)^{\epsilon_\alpha}\alpha\Delta\beta + (-1)^{\epsilon_\alpha}(\alpha, \beta). \quad (7.34)$$

For $\alpha = \beta = S$, the left-hand side of (7.33) vanishes by $(S, S) = 0$. In view of the gradings of S , the right-hand side yields $2(\Delta S, S)$. Thus, $(\Delta S, S) = 0$, that is, ΔS is closed, as requested.

This does not imply that ΔS is exact, however, unless the cohomological group $H^1(s)$ vanishes at ghost number one, using that ΔS has ghost number one. But (7.32) requires ΔS to be exact. If ΔS is not exact, there is no M_1 and therefore, no way to define a BRST invariant measure such that the quantum averages do not depend on the gauge

fixing fermion. This presumably signals a serious pathology of the theory. If ΔS is exact, M_1 exists and the problem of existence of the next term M_2 can be investigated. It is straightforward to show that it is again $H^1(s)$ that measures the potential obstructions to the existence of this next term M_2 , as well as the existence of the subsequent terms M_3 etc..

In particular, if it can be shown that $H^1(s)$ vanishes, it is guaranteed that a solution of the quantum master equation exists. If $H^1(s) \neq 0$, further work is required since it is necessary to check that no obstruction emerges. Note that W should be perturbatively a local functional (with possibly infinite coupling constants), so that the relevant space in which to compute the cohomology is that of local functionals. A detailed discussion of this is, however, beyond the scope of this lecture.

Index

- 't Hooft symbol, 43
- 1PI correlation function, 107

- Abelian tensor field, 32
- Action, 50
- Aharonov-Bohm effect, 21
- Algebra, 9
 - Representation, 9
- Almost local, 152
- Amputated correlation function, 106
- Anomalous dimension, 117
- Anomaly, 159, 169
 - Axial, 160
 - Cancellation, 165, 171
 - Chiral, 160
 - Global, 159
 - Local, 165
 - Topological charge, 171
 - Value, 164
- Anti-field, 77
- Anti-field formalism, 71
 - Anomaly, 174
- Anti-instanton, 38
- Anti-self-duality, 46
- Antibracket, 77
- Asymptotic freedom, 136
- Asymptotic safety, 137
- Atiyah-Singer index theorem, 172
- Axial charge, 158

- Axial symmetry, 158

- Baker-Campbell-Hausdorff formula, 51
- Banks-Zak fixed point, 137
- Bare mass, 111
- β_0 , 136
- β function, 117, 135
 - Expansion, 119
- Bloch-Nordsieck theorem, 138
- Bound state, 154
- BPHZ
 - BPHZ subtraction scheme, 120
- BPHZ theorem, 119
- BRST, 68
 - Algebra, 73
 - Charge, 72
 - Cohomology, 74
 - Exactness, 70
 - Matter, 71
 - Nilpotency, 70
 - Subspaces, 73
 - Symmetry, 68
 - Transformation, 68, 69, 78
 - Gauge-fixed, 85
 - Quantum, 88
- BRST symmetry, 146
 - Breaking, 146
 - Non-perturbative, 148

- Callan-Symanzik equation, 117
- Cartan subalgebra, 11
- Casimir, 12
- Center, 10
- Channel, 96
- Charge, 5
 - Electric, 25
- Chern class, 42
- Chern-Simmons characteristic, 40
- Chiral symmetry
 - Breaking, 158
- Chiral transformation
 - Local, 162, 167
- Classical contribution, 107
- Closed, 74
- Coboundary, 74
- Cocycle, 74
- Collective coordinates, 46
- Collective excitation, 36
- Color wave, 18
- Composite operator, 126, 153
- Composite state, 126
- Conformal, 137
- Connected correlation function, 106
- Connected diagram, 105
- Connection, 27
- Conservation law, 4
- Correlation function, 53, 96
 - Full, 105
- Counter-term
 - Coupling, 113
 - Mass, 111
 - Wave-function, 112
- Coupling constant, 16
- Coupling universality, 28
- Covariant derivative, 25, 26
 - Adjoint representation, 16
 - Arbitrary representation, 28
- Cross-section, 95
- Current
 - Axial, 160
 - Axial color, 161
 - Left-handed, 161
 - Right-handed, 161
 - Singlet, 161
- Custodial symmetry, 29
- Cutoff, 110
- Decay, 95
- Defect, 47
- Descent equation, 169
- Determinant of an operator, 55
- Dilatation symmetry, 159
- Dimension
 - Anomalous, 119
 - Canonical, 118
 - Engineering, 118
 - Naive, 118
- Dimensional transmutation, 115, 122, 136
- Dirac phase, 152
- Dirac string, 152
- Disconnected diagram, 102
- Distribution theory, 54
- Double-well potential, 37
- Dressing function, 5
- Dual field-strength tensor, 6, 43
- Dynkin index, 12
- Dyson-Schwinger equations, 86, 92, 93
- Elastic scattering, 99
- Electromagnetism, 5
- Electron
 - Observable, 152

- Elitzur's theorem, 147
- η' , 158
- Euler constant, 123
- Exact, 74
- Faddeev-Popov
 - Determinant, 60
 - Ghost, 66
 - Operator, 60, 65
- Feynman gauge, 83
- Feynman parametrization, 130
- Feynman propagator, 98
- Feynman rules, 103, 107
 - Matter, 105
 - Yang-Mills theory, 105
- Field configuration, 52
- Field-strength tensor
 - Yang-Mills theory, 16
- Flat direction, 57
- Flavor symmetry, 29
- Flavors, 62
- Form factor, 5
- Freedman-Townsend model, 33
- Fujikawa-Hirschfeld gauge, 141
- Functional
 - Functional δ -function, 59
 - Derivative, 54
 - Gaussian integral, 55
 - Integral, 54
 - Power series, 54
 - Substitution, 56
 - Translational invariance, 56
- Functional analysis, 54
- Fundamental modular domain, 144
- Gauge
 - Axial, 7, 62
 - Condition, 7, 59, 65, 83
 - Copy, 59
 - Coulomb, 7, 62
 - Covariant, 61
 - Feynman, 62
 - Invariance, 6
 - Landau, 7, 59
 - Lorentz, 7
 - Non-strict, 83
 - Parameter, 62, 66, 83
 - Radiation, 7
 - R_ξ gauge, 61
 - Strict, 83
 - Temporal, 40
 - Transformation
 - Gauge field, 15
 - Large, 24
- Gauge boson
 - Scattering, 127
- Gauge field
 - Geometry, 18
 - Yang-Mills, 15
- Gauge orbit
 - Residual, 143
- Gauge symmetry, 6
 - Off-shell closed, 31
 - On-shell closed, 31
 - On-shell reducible, 33
 - Reducible, 32
- Gauge theory, 6
 - Abelian, 6
- Gauge transformation
 - Harmonic, 7
 - Matter, 24
 - Maxwell theory, 6
 - Yang-Mills, 15

- Gauge-fixing fermion, 82
- Gaussian integral, 55
- Gell-Mann matrices, 11
- Generalized Action, 78
- Generalized action
 - Antisymmetric Abelian Tensor, 79
 - Non-minimal part, 82
 - Yang-Mills theory, 79
- Generating functional, 57
- Generator, 9
- Ghost, 66
 - Propagator, 133
 - Self-energy, 133
- Ghost number, 72
- Ghosts of ghosts, 77
- Grading, 77
- Grassmann
 - Parity, 69
- Green's function, 96
- Gribov
 - Copy, 23
 - Horizon, 142
 - Problem, 22
 - Region, 142
 - First, 141
- Gribov-Singer ambiguity, 22
- Group, 8
 - Abelian, 8
 - Continuous, 9
 - Identity element, 8
 - Invariants, 13
 - Inverse element, 8
 - Non-Abelian, 8
 - Orbit, 13
 - Rank, 11
- Haag's theorem, 139
- Haar measure, 12
- Heat kernel, 168
- Hodge dual, 6
- Infraparticle, 138
- Infrared divergencies
 - Lattice regularization, 125
 - Lower dimensions, 125
- Infrared singularities, 137
- Instanton, 36, 38
 - Yang-Mills theory, 45
- Internal space, 4
- Invariant tensor, 13
- Jacobi identity, 6, 9
- Jacobian, 166
- Jacobian determinant, 56
- Kinoshita-Lee-Nauenberg theorem, 138
- Koszul-Tate differential, 80
- Kugo-Ojima quart mechanism, 75
- Λ_{QCD} , 136
- Landau gauge, 62, 65, 67, 83
 - Absolute, 144
 - Minimal, 144
- Landau pole, 136
- Lattice Monte Carlo simulation, 58
- Leibnitz rule, 54
- Levi-Civita tensor, 11
- Lie algebra, 9
- Lie group, 8
- Linear- σ model, 4, 91
- Local symmetry, 6
- Longitudinal differential, 80
- Master equation
 - Classical, 78
 - Quantum, 87

- Matter, 24
- Maxwell equations, 6
- Maxwell theory, 5
- Minimal coupling, 25
- Moduli, 46
 - Space, 46
- MOM, 120
- Momentum subtraction scheme, 120
- Monopole, 47
- Nakanishi-Lautrup field, 67
- Negative-norm states, 71
- Noether current, 5
- Noether's theorem, 4
- One-particle irreducibility, 107
- Operator
 - Eigenvalue, 55
- Orbit, 13
 - Trivial, 13
- Parallel transporter, 26
- Partition function, 52
- Path integral, 50
 - Quantum mechanics, 50
- Pauli matrices, 10
- Perturbation theory, 95
 - ϕ^4 model, 91
- Physical operators, 74
- Physical photon field, 19, 151
- Physical state space, 71
- Physical subspace, 74
- Positronium, 126
- Primitively divergent, 116
- Proper correlation function, 106
- Pure gauge configuration, 40
- QED, 62
- Quantum action, 87
- Quantum master equation, 174
- Quartet mechanism, 75
- Radiative corrections, 107
- Reconstruction theorem, 57
- Regularization, 108
 - Cutoff, 109
 - Dimensional, 121
 - Heat kernel, 168
 - Lattice, 50
 - Pauli-Villars, 157, 163
 - Symmetry breaking, 120
- Regulator
 - Infrared, 125
 - Lattice, 124
- Renormaliability
 - Power counting, 114
- Renormalization, 108
 - Composite operator, 126
 - Coupling, 113
 - Mass, 111
 - Multiplicative, 113
 - Wave-function, 112
- Renormalization scale, 115, 116
- Renormalization scheme, 108, 111, **115**
 - MS, 123
 - $\overline{\text{MS}}$, 123
 - Pole scheme, 114
- Renormalization-group improvement, 119
- Representation, 8
 - (Pseudo-)real, 29
 - Adjoint, 11
 - Faithful, 8
 - Fundamental, 11
 - Linear, 14
 - Non-linear, 14

- Trivial, 8
- Resummation, 110, 133
- Rigid
 - Identity, 89, 91
 - Symmetry, 91
- Running, 108, 116
 - Coupling, 135
- S-matrix, 96
- Scale, 136
- Scale invariance, 115
- Self-duality, 43
- Slavnov-Taylor identity, 89, 92
- Source, 56
 - Composite operator, 126
- Sphaleron, 46
- Standard model
 - Anomaly, 170
 - Anomaly cancellation, 171, 174
- STI, 92
- Stochastic quantization, 143
- Strong CP problem, 157
- Structure constant, 9
- $su(2)$ algebra, 10
- $su(3)$ algebra, 11
- Subtraction, 120
- Sudakov logarithm, 138
- Symmetric configuration, 119
- Symmetry
 - Global, 3
- Symmetry breaking, 147
 - Spontaneous, **147**
- Symmetry current, 5
- Symmetry factor, 104
- Tadpole diagram, 109
- Tadpole integral, 130
- Tensor integral, 129
- θ vacuum, 156
- Thomson limit, 25, 114
- Time evolution operator, 96
- Time ordering, 96
- Topological
 - Charge, 41, 42
 - Excitation, 36
 - Term, 157
- Topologically trivial path, 39
- Total ghost number, 77
- Transfer matrix, 52
- Transition amplitude, 50
- Transition matrix element, 95, 96
- Tree-level, 107
- Triangle graph, 162
- Triviality, 137
- Tunneling, 37
- Tunneling amplitude, 157
- $U(1)$, 10
- Vacuum contribution, 102
- Vacuum-to-vacuum transition amplitude, 96
- Valence particle, 36
- Vertex function, 106
- Virtual particle, 102
- Vortex, 47
- Ward-Takahashi identity, 89, 91
- Wick rotation, 37
- Wigner-Eckart theorem, 5
- Wilson line, 153
- Wilson loop, 155
- Wilson-Fisher fix point, 137
- Winding number, 40
- Witten anomaly, 173

WKB approximation, 37

WTI, 91

Yang-Mills theory, 14

 Gauge symmetry, 31

 Lagrangian, 17

Yukawa model, 109

Zwanziger Lagrangian, 142