Quantum Field Theory II

Lecture Notes

ETH Zurich, 2026 FS

Prof. N. Beisert

Contents

0 Overview

After having learned the basic concepts of Quantum Field Theory in QFT I, in particular the theory of Quantum Electrodynamics, we can now go on to complete the foundations of QFT in QFT II. The aim of this lecture course is to be able to formulate the Standard Model of Particle Physics and perform calculations in it. We shall cover the following topics:

- path integral formulation
- non-abelian gauge theory
- renormalisation
- quantum symmetries
- spontaneous symmetry breaking

More concretely, the topics can be explained as follows:

Path Integral Formulation. In QFT I we have applied the canonical quantisation framework to fields. The path integral is an alternative framework for performing equivalent computations. In many situations it is more direct, more efficient or simply more convenient to use. It is however not built upon the common intuition of quantum mechanics.

Non-Abelian Gauge Theory. We have already seen how to formulate the vector field for use in electrodynamics. The vector field for chromodynamics is similar, but it adds the important concept of self-interactions which makes the field have a very different physics. The underlying model is called non-abelian gauge theory or Yang–Mills theory.

Renormalisation. We will take a fresh look at renormalisation, in particular concerning the consistency of gauge theory and the global features of renormalisation transformations.

Quantum Symmetries. We will consider how symmetries work in the path integral framework. This will also give us some awareness of quantum violations of symmetry, so-called anomalies.

Spontaneous Symmetry Breaking. The electroweak interactions are mediated by massive vector particles. Naively, they lead to non-renormalisable models, but by considering spontaneous symmetry breaking one can accommodate them in gauge theories.

0.1 Prerequisites

- Quantum Field Theory I (concepts, start from scratch)
- classical and quantum mechanics
- electrodynamics, mathematical methods in physics

0.2 Contents

Indicated are the approximate number of lecture minutes for each chapter. Altogether, the course consists of around 40 lectures of 45 minutes.

0.3 References

There are many text books and lecture notes on quantum field theory. Here is a selection of well-known ones:

- N. Beisert, "Quantum Field Theory I", lecture notes, see course homepage.
- M. E. Peskin, D. V. Schroeder, "An Introduction to Quantum Field Theory", Westview Press (1995)
- C. Itzykson, J.-B. Zuber, "Quantum Field Theory", McGraw-Hill (1980)
- P. Ramond, "Field Theory: A Modern Primer", Westview Press (1990)
- M. Srendnicki, "Quantum Field Theory", Cambridge University Press (2007)
- M. Kaku, "Quantum Field Theory", Oxford University Press (1993)
- online: M. Gaberdiel, A. Gehrmann-De Ridder, "Quantum Field Theory II", lecture notes, `http://edu.itp.phys.ethz.ch/fs11/11FSQFT2/`
- ITP lecture archive `http://edu.itp.phys.ethz.ch/select.php?title=QFT2`

0.4 Acknowledgements

I am grateful to Elias Furrer for a list of corrections on an earlier version of these lecture notes.

1 Path Integral for Quantum Mechanics

We start the lecture course by introducing the path integral in the simple setting of quantum mechanics in canonical quantisation.

1.1 Motivation

The path integral is a framework to formulate quantum theories. It was developed initially by Dirac (1933) and Feynman (1948). It is particularly useful for relativistic quantum field theory because it allows a manifestly covariant formulation.

Why? In QFT I we have relied on canonical quantisation to formulate a quantum theory of relativistic fields. During the first half of that course, we have encountered and overcome several difficulties in quantising scalar, spinor and vector fields.

- Canonical quantisation is intrinsically not relativistically covariant due to the specialisation of time. Nevertheless, at the end of the day, results turned out covariant as they should. In between, we had to manipulate some intransparent expressions.
- Gauge fixing the massless vector field was not exactly fun.
- Canonical quantisation is based on non-commuting field operators. The operator algebra makes manipulations rather tedious.
- Moreover one has to deal with ordering ambiguities when quantising a classical expression.
- Despite their final simplicity, deriving Feynman rules was a lengthy effort.
- We can treat interacting models perturbatively, but it is hard to formulate what finite or strong coupling means.

The path integral method avoids many of the above problems:

- It does not single out a particular time or relativistic frame in any way. A priori, it is a fully covariant framework.
- It uses methods of functional analysis rather than operator algebra. The fundamental quantities are perfectly commuting objects (or sometime anti-commuting Graßmann numbers).
- It is based directly on the classical action functional. Operator ordering issues do not have to be considered (although there is an equivalent of operator ordering).
- Gauge fixing for massless vector fields has a few complications which are conveniently treated in the path integral framework.

- Feynman rules can be derived directly and conveniently.
- The path integral can be formulated well for finite or strong coupling, and some information can sometimes be extracted (yet the methods of calculation are usually restricted to the perturbative regime).
- It is a different formulation and interpretation of quantum theory.

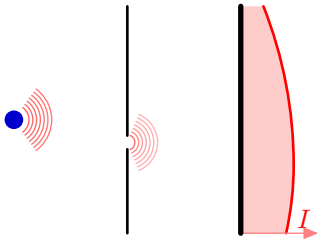
Why Not? In fact, one could directly use the path integral to formulate a quantum theory without first performing canonical quantisation. There are however a few shortcoming of the path integral which are good reasons to understand the canonical framework first:

- The notion of states is not as evident in the path integral.
- Similarly, operators and their algebras are not natural concepts of the path integral.
- Therefore, the central feature of unitarity remains obscure in the path integral.
- Canonical quantisation of fields connects immediately to the conventional treatment of quantum mechanics.

Multiple Slits. But what is the path integral? It is a method to compute the interference of quantum mechanical waves by considering *all* trajectories.

A standard way to illustrate the path integral is to consider multiple-slit interference patterns: Consider a source which emits particles or waves to a screen where they will be detected. Suppose the particles have a well-defined de Broglie wavelength λ. We then insert hard obstacles into the path and observe the interference pattern on the screen.

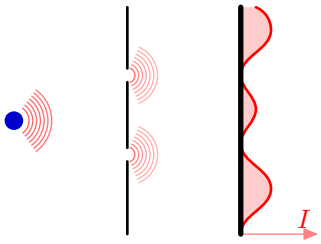
Suppose we first put an obstacle with a single slit.



(1.1)

A sufficiently small slit (compared to λ) would act as a new point-like source, and we would observe no structures on the screen.[1]

Opening a second slit in the obstacle produces a non-trivial interference pattern.

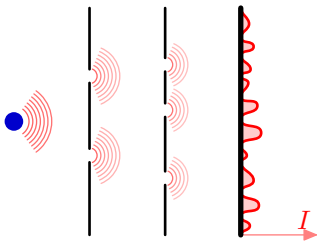


(1.2)

[1] One would indeed observe a non-trivial interference pattern when the size is of the same order as λ. We will discuss this case further below.

The wave can propagate through both slits, but the two waves arrive with a different relative phase at the screen thus producing a pattern of constructive and destructive interference.

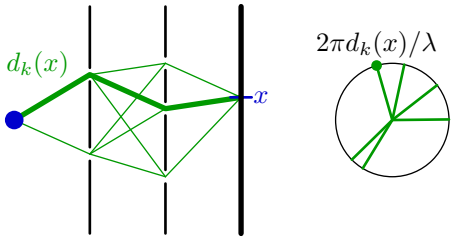
Now we put another obstacle with slits at different locations.



(1.3)

The waves which have passed the first obstacle will now hit the second obstacle, and only a tiny fraction of them will pass this obstacle. Although there may not be a classical straight path connecting the source to the screen, a very weak interference pattern can be observed.

How to compute the interference pattern? Each slit can be viewed to act as a light source for the next layer of obstacles. Importantly, the relative phase at each slit is well-defined. This is what creates the interference pattern.[2] To determine the intensity at a specific point on the screen, we thus collect all paths connecting it to the source via the various slits. The source, slits and screen are connected by straight lines and we measure their overall length $d_k(x)$.



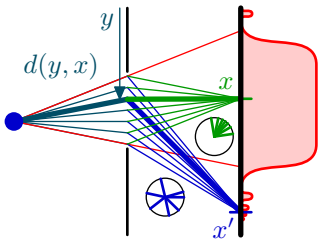
(1.4)

Neglecting the decrease of amplitude for circular waves in this simplistic thought experiment, the intensity is given by

$$I(x) = \big|A(x)\big|^2, \qquad A(x) \sim \sum_k \exp\big(2\pi i d_k(x)/\lambda\big). \tag{1.5}$$

As all the $d_k(x)$ have a distinct dependence on x, the resulting intensity will crucially depend on x.

We can now also consider a single slit of an arbitrarily large aperture compared to λ. In this case, the wave can pass at every point within the slit.



(1.6)

[2]For uncorrelated phases one would not obtain interference.

The distance $d(y,x)$ from source to screen now depends also on the position y within the slit. To obtain the amplitude we should integrate over it

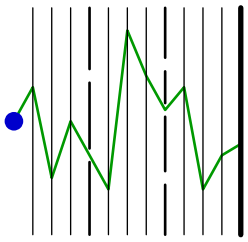
$$A(x) \sim \int_{\min}^{\max} \mathrm{d}y \exp\bigl(2\pi \mathring{\imath} d(y,x)/\lambda\bigr). \tag{1.7}$$

This also leads to a non-trivial interference pattern in $I(x) = |A(x)|^2$.

When the size of the slit is large compared to λ, one expects to find a rather sharp image of the slit on the screen. This is because the wave character is not very relevant to the problem. This fact can be understood as constructive and destructive interference of correlated waves: In a straight line behind the slit, all trajectories passing the slit will have approximately the same length. They will be in phase and there is constructive interference. For points in the classical shadow of the obstacle, the trajectory between source and screen must bend. The various trajectories have lengths which differ strongly on the scale of λ. Hence destructive interference is expected.

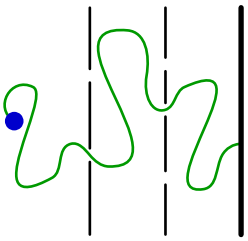
If the above considerations are correct, we can compute the interference pattern for an arbitrary array of obstacles. We could even put an obstacle at some distance which does not block waves at all. does not block waves at all. We would have to integrate over all intermediate positions at this non-obstacle, but the result should still be correct.

We can now be even more extreme, and put non-obstacles at many different locations.



(1.8)

The calculation will be more tedious, but the answer should still remain the same. If we keep adding virtual layers, we eventually have to integrate over all curved paths connecting source to destination, not just the straight ones.[3]

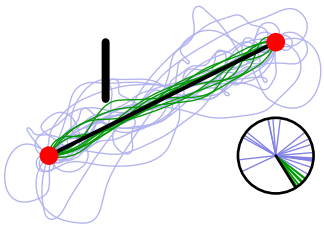


(1.9)

How comes that eventually the curvature has no effect at all? The point is that non-classical paths average out: A straight trajectory has the shortest length. A trajectory which is slightly curved has a length which is just a tiny bit larger. It

[3]Here, one may (or may not) restrict to paths that do not curve back along the horizontal direction.

will also see almost the same obstacles. Hence there is constructive interference. For reasonably large curvature, there are many trajectories which have relatively different lengths.



(1.10)

Thus they interfere destructively, and will effectively not contribute to the interference pattern no matter whether they hit an obstacle or not.

The above describes the path integral method for calculating interference patterns of waves. Let us now apply it more formally to a generic quantum mechanical system.

1.2 Transition Amplitudes

In the remainder of this chapter we shall explicitly use hats to denote an operator $\hat{F}$ corresponding to a classical function $F(p,q)$ of phase space. We shall also make $\hbar$ explicit everywhere.

In the following we shall consider transition amplitudes in a simple quantum mechanical model, and derive a path integral formulation for them. The model is described by a classical Hamiltonian function $H(q,p)$ and its quantum counterpart operator $\hat{H}$. Up to ordering issues of $\hat{q},\hat{p}$ we thus have

$$\hat{H} = H(\hat{q},\hat{p}). \tag{1.11}$$

Our goal is to express the quantum transition amplitude as a weighted sum over all conceivable intermediate field configurations using merely the classical Hamiltonian $H(q,p)$.

Transition Amplitude. We want to compute the transition amplitude $A_{\rm f,i}$ between position $q_{\rm i}$ at time $t_{\rm i}$ and position $q_{\rm f}$ at time $t_{\rm f}$ [4]

$$A_{\rm f,i} = \langle q_{\rm f}, t_{\rm f} | q_{\rm i}, t_{\rm i}\rangle = \langle q_{\rm f}|\hat{U}(t_{\rm f},t_{\rm i})|q_{\rm i}\rangle, \tag{1.12}$$

where $\hat{U}(t_{\rm f},t_{\rm i})$ is the time evolution operator. In the case of a time-independent Hamiltonian it reads

$$\hat{U}(t_{\rm f},t_{\rm i}) = \exp\bigl(-\mathring{\imath}\hbar^{-1}(t_{\rm f}-t_{\rm i})\hat{H}\bigr). \tag{1.13}$$

[4] The symbol $|q\rangle$ denotes a state localised at position q. The symbol $|q,t\rangle := \hat{U}(t_{\rm ref},t)|q\rangle$ denotes a state in the Heisenberg picture which is perfectly localised at position q and time t transported back to the reference time slice at time $t_{\rm ref}$.

Otherwise $\hat{U}(t_\mathrm{f}, t_\mathrm{i})$ is determined by the differential equation

$$\mathrm{d}\hat{U}(t_\mathrm{f}, t_\mathrm{i}) = -\mathring{\imath}\hbar^{-1}\hat{H}(t_\mathrm{f})\hat{U}(t_\mathrm{f}, t_\mathrm{i})\,\mathrm{d}t_\mathrm{f} + \mathring{\imath}\hbar^{-1}\hat{U}(t_\mathrm{f}, t_\mathrm{i})\hat{H}(t_\mathrm{i})\,\mathrm{d}t_\mathrm{i}\,. \tag{1.14}$$

We want to find an expression for $A_{\mathrm{f,i}}$ which merely uses the classical Hamiltonian $H(q,p)$ instead of the operator $\hat{H}$.

Time Slices. First we interrupt the time evolution at some intermediate time $t_\mathrm{f} > t_k > t_\mathrm{i}$ by using the group property of time evolution

$$\hat{U}(t_\mathrm{f}, t_\mathrm{i}) = \hat{U}(t_\mathrm{f}, t_k)\hat{U}(t_k, t_\mathrm{i}). \tag{1.15}$$

Now we insert a complete set of position states at time t_k

$$1 = \int \mathrm{d}q_k\, |q_k\rangle\langle q_k|. \tag{1.16}$$

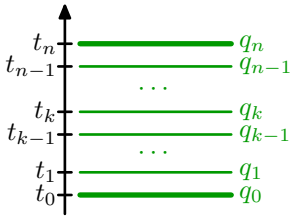
Altogether, we obtain an identity

$$A_{\mathrm{f,i}} = \int \mathrm{d}q_k\, A_{\mathrm{f},k} A_{k,\mathrm{i}}. \tag{1.17}$$

We iterate the subdivision n times to obtain a very fine resolution for the time evolution

$$A_{\mathrm{f,i}} = \int \mathrm{d}q_{n-1} \ldots \mathrm{d}q_1\, A_{\mathrm{f},n-1} \ldots A_{1,\mathrm{i}} = \int \prod_{k=1}^{n-1} \mathrm{d}q_k \prod_{k=1}^{n} A_{k,k-1}, \tag{1.18}$$

where we define $q_0 := q_\mathrm{i}$, $t_0 := t_\mathrm{i}$ and $q_n := q_\mathrm{f}$, $t_n := t_\mathrm{f}$.



(1.19)

Classical Evolution. The intermediate time slices allow us to approximate quantum time evolution by the classical one which works well for sufficiently small time intervals.

Consider the elementary transition amplitude

$$A_{k,k-1} = \langle q_k|\hat{U}(t_k, t_{k-1})|q_{k-1}\rangle. \tag{1.20}$$

For a sufficiently small time interval we can approximate the time evolution by the Hamiltonian

$$\hat{U}(t_k, t_{k-1}) = 1 - \mathring{\imath}\hbar^{-1}(t_k - t_{k-1})\hat{H} + \mathcal{O}\big((t_k - t_{k-1})^2\big). \tag{1.21}$$

We should now evaluate

$$\langle q_k|\hat{H}|q_{k-1}\rangle = \langle q_k|H(\hat{q},\hat{p})|q_{k-1}\rangle. \tag{1.22}$$

This expression certainly depends on the ordering of factors $\hat{q}$ and $\hat{p}$ in $H(\hat{q},\hat{p})$ which is not determined by the classical Hamiltonian function $H(q,p)$. If we order all $\hat{q}$ to be to the left of all $\hat{p}$ we will get

$$\langle q_k|\hat{H}|q_{k-1}\rangle = \langle q_k|H(q_k,\hat{p})|q_{k-1}\rangle. \tag{1.23}$$

If instead we order all $\hat{q}$ to be to the right of all $\hat{p}$ we will get

$$\langle q_k|\hat{H}|q_{k-1}\rangle = \langle q_k|H(q_{k-1},\hat{p})|q_{k-1}\rangle. \tag{1.24}$$

We may choose to order the factors such that $\hat{q}$ evaluates to the average $\bar{q}_k = \frac{1}{2}q_{k-1} + \frac{1}{2}q_k$

$$\langle q_k|\hat{H}|q_{k-1}\rangle = \langle q_k|H(\bar{q}_k,\hat{p})|q_{k-1}\rangle. \tag{1.25}$$

This is called the Weyl ordering of $H(\hat{q},\hat{p})$. In fact, all orderings are equivalent up to simpler terms of order $\hbar$, and the precise definition of $\bar{q}_k$ does not matter be it q_k, q_{k-1}, their average or any other combination that limits to $\bar{q}_k \to q(t_k)$ at $t_{k-1} \to t_k$.

Note that for the transition $\hat{q} \to q$ to work as described, it is crucial that the leading non-trivial order of the expansion consists of only a single factor of $\hat{H}$. For higher powers of $\hat{H}$ the non-trivial operator ordering would prevent us from replacing $\hat{q}$ by some average value $\bar{q}_k$. Hence we need a sufficiently large number of intermediate time slices for a good approximation of time evolution altogether. We arrive at an approximation for $A_{k,k-1}$

$$\langle q_k|q_{k-1}\rangle - \mathring{\imath}\hbar^{-1}(t_k - t_{k-1})\langle q_k|H(\bar{q}_k,\hat{p})|q_{k-1}\rangle + \mathcal{O}\big((t_k - t_{k-1})^2\big). \tag{1.26}$$

In order to evaluate the momentum operator $\hat{p}$, we insert a complete set of momentum eigenstates

$$1 = \int \frac{\mathrm{d}p_k}{2\pi\hbar}\,|p_k\rangle\langle p_k|. \tag{1.27}$$

Insertion into the above approximation yields two terms

$$\begin{aligned}\langle q_k|q_{k-1}\rangle &= \int \frac{\mathrm{d}p_k}{2\pi\hbar}\,\langle q_k|p_k\rangle\langle p_k|q_{k-1}\rangle,\\ \langle q_k|\hat{H}|q_{k-1}\rangle &= \int \frac{\mathrm{d}p_k}{2\pi\hbar}\,\langle q_k|H(\bar{q}_k,\hat{p})|p_k\rangle\langle p_k|q_{k-1}\rangle.\end{aligned} \tag{1.28}$$

We substitute the eigenvalue $H(\bar{q}_k,\hat{p})|p_k\rangle = H(\bar{q}_k,p_k)|p_k\rangle$ and the Fourier exponent $\langle p|q\rangle = \exp(-\mathring{\imath}\hbar^{-1}pq)$ to obtain a classical approximation for $A_{k,k-1}$

$$\int \frac{\mathrm{d}p_k}{2\pi\hbar}\exp\big[-\mathring{\imath}\hbar^{-1}(t_k - t_{k-1})H(\bar{q}_k,p_k) + \mathring{\imath}\hbar^{-1}p_k(q_k - q_{k-1})\big]. \tag{1.29}$$

Here, the Fourier exponents neatly combine with the expansion in small $t_k - t_{k-1}$ to yield one overall exponential function.

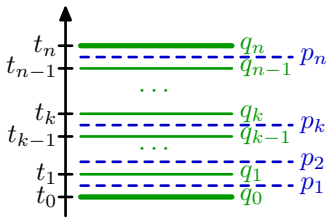
Path Integral in Phase Space. Putting all time slices together we now have an integral expression for the transition amplitude $A_{n,0}$ in terms of the classical Hamiltonian

$$A_{\mathrm{f,i}} \approx A_{n,0} := \int \prod_{k=1}^{n-1} \mathrm{d}q_k \prod_{k=1}^{n} \frac{\mathrm{d}p_k}{2\pi\hbar} \exp\bigl(\mathring{\imath}\hbar^{-1} S_{n,0}[q,p]\bigr). \tag{1.30}$$

Here, we recall the exact additive property $\mathrm{e}^x\,\mathrm{e}^y = \mathrm{e}^{x+y}$ that distinguishes the exponential function e^x from all other functions with the expansion $1 + x + \mathcal{O}(x^2)$. This allows us to combine all integrands into a single exponential of the sum of exponents. The phase of the exponential is then given by the action function

$$S_{n,0}[q,p] := \sum_{k=1}^{n} \bigl(-(t_k - t_{k-1}) H(\bar{q}_k, p_k) + p_k (q_k - q_{k-1})\bigr). \tag{1.31}$$

The transition amplitude $A_{\mathrm{f,i}}$ is approximated by the integral of $\exp(\mathring{\imath}\hbar^{-1} S_{n,0}[q,p])$ over all intermediate positions q_k and all intermediate momenta p_k. The initial and final positions q_0, q_n are held fixed, and there is one momentum integral more to be done than position integrals.



(1.32)

Interestingly, the measure factor for a combined position and momentum integral is just $2\pi\hbar$,

$$\int \frac{\mathrm{d}q\,\mathrm{d}p}{2\pi\hbar}\,. \tag{1.33}$$

This is the volume quantum mechanics associates to an elementary cell in phase space.

We have already convinced ourselves that the quality of the approximation depends on the number of time slices. The exact transition amplitude is obtained by formally taking the limit of infinitely many time slices at infinitesimal intervals. We abbreviate the limit by the so-called path integral

$$A_{\mathrm{f,i}} = \int \mathrm{D}q\,\mathrm{D}p \exp\bigl(\mathring{\imath}\hbar^{-1} S_{\mathrm{f,i}}[q,p]\bigr), \tag{1.34}$$

where the phase is now given as a functional of the path

$$S_{\mathrm{f,i}}[q,p] = \int_{t_\mathrm{i}}^{t_\mathrm{f}} \mathrm{d}t\,\bigl(p(t)\dot{q}(t) - H(q(t),p(t))\bigr) = \int_{\mathrm{i}}^{\mathrm{f}} (p\,\mathrm{d}q - H\,\mathrm{d}t). \tag{1.35}$$

This path integral "integrates" over all paths $(q(t), p(t))$.

Comparing to the above discrete version, the term $\mathrm{d}t\, p\dot{q} \approx p_k(q_k - q_{k-1})$ is responsible for shifting the time slice forward whereas the Hamiltonian governs the evolution of the wave function.

The expression we obtained for the phase factor $S_{\mathrm{f,i}}[q,p]$ is exciting, it is precisely the action in phase space. Note that the associated Euler–Lagrange equations

$$0 = \frac{\delta S}{\delta q(t)} = -\dot{p}(t) - \frac{\partial H}{\partial q}\,(t), \qquad 0 = \frac{\delta S}{\delta p(t)} = +\dot{q}(t) - \frac{\partial H}{\partial p}\,(t), \tag{1.36}$$

are just the Hamiltonian equations of motion.

Here the principle of extremal action for a classical path finds a justification: The action determines a complex phase $S_{\mathrm{f,i}}[q,p]/\hbar$ for each path $(q(t),p(t))$ in phase space. Unless the action is extremal, the phase will vary substantially from one path to a neighbouring one. On average such paths will cancel out from the path integral. Conversely, a path which extremises the action, has a stationary action for all neighbouring paths. These paths will dominate the path integral classically. When quantum corrections are taken into account, the allowable paths can wiggle around the classical trajectory slightly, on the order of $\hbar$.

The path integral for the transition amplitude $A_{\mathrm{f,i}}$, keeps the initial and final positions fixed

$$q(t_{\mathrm{i}}) = q_{\mathrm{i}}, \qquad q(t_{\mathrm{f}}) = q_{\mathrm{f}}, \tag{1.37}$$

whereas the momenta $p(t_{\mathrm{i}})$ and $p(t_{\mathrm{f}})$ are free. The path integral can also compute other quantities where the boundary conditions are specified differently.

Note that the integration measures $\mathrm{D}q$ and $\mathrm{D}p$ typically hide some factors which are hard to express explicitly. Usually such factors can be ignored during a calculation, and are only reproduced in the end by demanding appropriate normalisation.

Finally, we must point out that the path integral may not be well-defined in a mathematical sense, especially because the integrand is highly oscillating. Nevertheless, it is reasonably safe to use the path integral in physics by formally manipulating it by the usual rules and inserting suitable regulators such as $\pm\mathring{\imath}\epsilon$.

Path Integral in Position Space. The above path integral is based on the Hamiltonian formulation where time has a distinguished role. For common physical systems we can transform the path integral back to the original Lagrangian framework.

We merely need to assume that the Hamiltonian is quadratic in the momenta p

$$H(q,p) = \frac{p^2}{2M(q)} + pK(q) + V(q). \tag{1.38}$$

For common physical models this is the case. For example, a classical particle of mass m in a potential $V(q)$ would have $M(q) = m$ and $K(q) = 0$.

Notice that the exponent depends at most quadratically on each momentum. This allows to integrate all momenta out using the Gaußian integral[5]

$$\int_{-\infty}^{+\infty} \mathrm{d}p \exp\bigl(-\tfrac{1}{2}ap^2 + bp + c\bigr) = \sqrt{2\pi/a}\, \exp(b^2/2a + c). \tag{1.39}$$

We obtain an expression for the transition amplitude

$$A_{n,0} = \int \prod_{k=1}^{n-1} \mathrm{d}q_k \prod_{k=1}^{n} \sqrt{\frac{M(\bar{q}_k)}{2\pi\hbar \mathring{i}(t_k - t_{k-1})}}\, \exp\bigl(\mathring{i}\hbar^{-1} S_{n,0}[q]\bigr), \tag{1.40}$$

which depends on

$$\begin{aligned} S_{n,0}[q] = \sum_{k=1}^{n} (t_k - t_{k-1}) \Bigl[& -V(\bar{q}_k) \\ & + \tfrac{1}{2} M(\bar{q}_k) \left(\frac{q_k - q_{k-1}}{t_k - t_{k-1}} - K(\bar{q}_k) \right)^2 \Bigr]. \end{aligned} \tag{1.41}$$

Now the exact path integral for the transition amplitude reads

$$A_{\mathrm{f,i}} = \int \mathrm{D}q \exp\bigl(\mathring{i}\hbar^{-1} S_{\mathrm{f,i}}[q]\bigr), \tag{1.42}$$

where the phase is given by the conventional action for a particle

$$S_{\mathrm{f,i}}[q] = \int_{t_\mathrm{i}}^{t_\mathrm{f}} \mathrm{d}t \, \bigl(\tfrac{1}{2} M(q)(\dot{q} - K(q))^2 - V(q)\bigr) = \int \mathrm{d}t \, L(q, \dot{q}). \tag{1.43}$$

Evidently, this action is classically equivalent to the above action in Hamiltonian form. Here we see that both classical actions yield the same path integral, and are therefore quantum equivalent. This is a special case, and quantum equivalence is not to be expected in general for two equivalent classical actions.

Note that the measure factor hidden in $\mathrm{D}q$ is substantially more complicated compared to the combination $\mathrm{D}q\,\mathrm{D}p$. In fact, it depends on the mass term $M(q)$. For a standard classical particle $M(q) = m$ is independent of q and hence the integration measure amounts to some constant overall factor.

We have finally obtained an expression for the quantum transition amplitude $A_{\mathrm{f,i}}$ which is based just on the classical action in the Lagrangian formulation. There is no need to translate to the Hamiltonian framework at any point of the calculation. In fact, we can generally use the path integral to define transition amplitudes or other quantum mechanical expressions. Note that the precise discretisation of the action $S[q]$, i.e. how to represent each instance of $q(t)$ and $\hat{q}(t)$, can influence the value of the path integral. This effect is equivalent to the choice of operator orderings. We will discuss these ambiguities further below.

[5]The formula is applicable as long as a has a positive real part, however small it may be. In our case, a is purely imaginary, but the usual assumptions of causality allows to attribute a small real positive part to a. For imaginary a, we will use the limiting sign $\sqrt{a} = \mathrm{e}^{\mathring{i}\pi/8} \sqrt{|a|}$.

1.3 Free Particle

Let us discuss a simple example, the free non-relativistic particle. We have derived two expressions for the path integral.

Phase Space. The free particle is defined by the Hamiltonian

$$H(q,p) = \frac{p^2}{2m}\,. \tag{1.44}$$

The discretised action in phase space therefore reads

$$S_{n,0}[q,p] := \sum_{k=1}^{n} \big(-(t_k - t_{k-1})p_k^2/2m + p_k(q_k - q_{k-1})\big). \tag{1.45}$$

We observe that each variable q_k appears only in a product with either p_k or p_{k+1}. The integral over q_k therefore yields a delta-function

$$\int \mathrm{d}q_k \exp\big(\mathring{\imath}\hbar^{-1} q_k(p_k - p_{k+1})\big) = 2\pi\hbar\,\delta(p_k - p_{k+1}), \tag{1.46}$$

which trivialises one of the momentum integrals and completely eliminates one time slice from the sequence without a remainder. Performing this step for all intermediate times, we are left with one Gaußian integral which we readily evaluate

$$\begin{aligned} A_{n,0} &= \int \prod_{k=1}^{n-1} \mathrm{d}q_k \prod_{k=1}^{n} \frac{\mathrm{d}p_k}{2\pi\hbar} \exp(\mathring{\imath}\hbar^{-1} S_{n,0}[q,p]) \\ &= \int \frac{\mathrm{d}p}{2\pi\hbar} \exp\big(-\mathring{\imath}\hbar^{-1}(t_n - t_0)p^2/2m + \mathring{\imath}\hbar^{-1} p(q_n - q_0)\big) \\ &= \sqrt{\frac{m}{2\pi\hbar\mathring{\imath}(t_n - t_0)}} \exp\left(\frac{\mathring{\imath} m(q_n - q_0)^2}{2\hbar(t_n - t_0)}\right). \end{aligned} \tag{1.47}$$

This is the correct transition amplitude $A_{\mathrm{f,i}}$ for a free non-relativistic particle. Actually, the number of intermediate steps n does not matter here due to the simplicity of the problem; usually the limit $n \to \infty$ is required, but here we may conveniently just set $n = 1$.

Position Space. Alternatively, we can start with the classical discretised action

$$S_{n,0}[q] = \tfrac{1}{2} m \sum_{k=1}^{n} \frac{(q_k - q_{k-1})^2}{t_k - t_{k-1}}\,. \tag{1.48}$$

It turns out that every integral over a q_k in

$$A_{n,0} = \int \prod_{k=1}^{n-1} \mathrm{d}q_k \prod_{k=1}^{n} \sqrt{\frac{m}{2\pi\hbar\mathring{\imath}(t_k - t_{k-1})}} \exp\big(\mathring{\imath}\hbar^{-1} S_{n,0}[q]\big), \tag{1.49}$$

simply eliminates the set of variables (q_k, t_k) from their sequences without leaving a gap in the above expression. Eventually, we thus find

$$A_{n,0} = \sqrt{\frac{m}{2\pi\hbar\mathring{\imath}(t_n - t_0)}}\, \exp\!\left(\frac{\mathring{\imath}m(q_n - q_0)^2}{2\hbar(t_n - t_0)}\right). \tag{1.50}$$

Fortunately, the two results agree.

1.4 Operator Insertions

We have found a way to express the transition amplitude $\langle q_\mathrm{f}, t_\mathrm{f}|q_\mathrm{i}, t_\mathrm{i}\rangle$ in terms of a path integral. The next important step is to find a way to insert quantum operators into the transition amplitude.

Single Insertion. Let us therefore insert one quantum operator $\hat{O} = O(\hat{q}, \hat{p})$ at time t_o with $t_\mathrm{f} > t_\mathrm{o} > t_\mathrm{i}$ into the transition amplitude[6]

$$A_{\mathrm{f,o,i}} = \langle q_\mathrm{f}, t_\mathrm{f}|\hat{O}(t_\mathrm{o})|q_\mathrm{i}, t_\mathrm{i}\rangle = \langle q_\mathrm{f}|\hat{U}(t_\mathrm{f}, t_\mathrm{o})\hat{O}\hat{U}(t_\mathrm{o}, t_\mathrm{i})|q_\mathrm{i}\rangle. \tag{1.51}$$

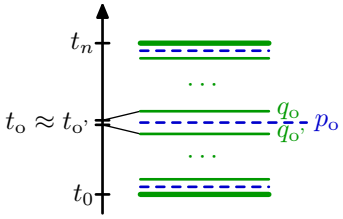
As before, we evaluate this expression by inserting a complete set of position eigenstates before and after the operator

$$A_{\mathrm{f,o,i}} = \int \mathrm{d}q_\mathrm{o}\, \mathrm{d}q'_\mathrm{o}\, A_{\mathrm{f,o}}\langle q_\mathrm{o}|\hat{O}|q'_\mathrm{o}\rangle A_{\mathrm{o',i}}. \tag{1.52}$$

Here we should replace the quantum operator $\hat{O}$ by $O(\bar{q}_\mathrm{o}, \hat{p})$ where $\bar{q}_\mathrm{o}$ is a suitable combination of q_o and q'_o. To evaluate the momentum operator $\hat{p}$ we insert a complete set of momentum eigenstates

$$\begin{aligned} A_{\mathrm{f,o,i}} &= \int \mathrm{d}q_\mathrm{o}\, \mathrm{d}q'_\mathrm{o}\, \frac{\mathrm{d}p_\mathrm{o}}{2\pi\hbar}\, A_{\mathrm{f,o}}\langle q_\mathrm{o}|O(\bar{q}_\mathrm{o}, \hat{p})|p_\mathrm{o}\rangle\langle p_\mathrm{o}|q'_\mathrm{o}\rangle A_{\mathrm{o',i}} \\ &= \int \mathrm{d}q_\mathrm{o}\, \mathrm{d}q'_\mathrm{o}\, \frac{\mathrm{d}p_\mathrm{o}}{2\pi\hbar}\, \exp\!\big(\mathring{\imath}\hbar^{-1}p_\mathrm{o}(q_\mathrm{o} - q'_\mathrm{o})\big) A_{\mathrm{f,o}}O(\bar{q}_\mathrm{o}, p_\mathrm{o})A_{\mathrm{o',i}}. \end{aligned} \tag{1.53}$$

We observe that the alternating pattern of position and momentum integrations is preserved across the operator insertion.



(1.54)

[6] Here we have used the Heisenberg and Schrödinger pictures of the operator related by $\hat{O}(t_\mathrm{o}) = \hat{U}(t_\mathrm{ref}, t_\mathrm{o})\hat{O}\hat{U}(t_\mathrm{o}, t_\mathrm{ref})$ with t_ref describing the reference time slice of the Heisenberg picture.

The complete discretised path integral with operator insertion therefore reads simply[7]

$$A_{\mathrm{f,o,i}} = \int \prod_{k=1}^{n-1} \mathrm{d}q_k \prod_{k=1}^{n} \frac{\mathrm{d}p_k}{2\pi\hbar}\, O(\bar{q}_\mathrm{o}, p_\mathrm{o}) \exp\bigl(\mathring{\imath}\hbar^{-1} S_{n,0}[q,p]\bigr). \tag{1.55}$$

As a continuous path integral it takes the form

$$A_{\mathrm{f,o,i}} = \int \mathrm{D}q\, \mathrm{D}p\, O\bigl(q(t_\mathrm{o}), p(t_\mathrm{o})\bigr) \exp\bigl(\mathring{\imath}\hbar^{-1} S_{\mathrm{f,i}}[q,p]\bigr). \tag{1.56}$$

Time Ordering. It should now be evident how to insert multiple operators into the path integral. Suppose we have n operators $\hat{O}_k$ at times t_k, respectively, with $t_\mathrm{f} > t_n > \ldots > t_2 > t_1 > t_\mathrm{i}$. We insert them into the transition amplitude and obtain a path integral with n operator insertions

$$\begin{aligned} &\langle q_\mathrm{f}, t_\mathrm{f}| \hat{O}_n \ldots \hat{O}_2 \hat{O}_1 |q_\mathrm{i}, t_\mathrm{i}\rangle \\ &= \int \mathrm{D}q\, \mathrm{D}p\, O_1(t_1) O_2(t_2) \ldots O_n(t_n) \exp\bigl(\mathring{\imath}\hbar^{-1} S_{\mathrm{f,i}}[q,p]\bigr). \end{aligned} \tag{1.57}$$

It is crucial that all the operators are in proper time order. Only then all the time evolution operators $U(t_k, t_{k-1})$ will shift time forward as in the above derivation of the path integral.

Conversely, a set of operator insertions in the path integral corresponds to a time-ordered product of quantum operators inserted into the transition amplitude

$$\begin{aligned} &\int \mathrm{D}q\, \mathrm{D}p \prod_{k=1}^{n} O_k(t_k) \exp\bigl(\mathring{\imath}\hbar^{-1} S_{\mathrm{f,i}}[q,p]\bigr) \\ &= \langle q_\mathrm{f}, t_\mathrm{f}|\, \mathrm{T}[\hat{O}_1 \hat{O}_2 \ldots \hat{O}_n] |q_\mathrm{i}, t_\mathrm{i}\rangle. \end{aligned} \tag{1.58}$$

In the form with the time-ordering operator T, the relation holds even for an arbitrary ordering of operator times t_k. Hence the path integral take care of time-ordering automatically.[8]

In some sense, the time ordering enters by the very definition of the path integral: There is precisely one location within the path integral sequence for each time. The trajectory strictly moves forward in time, there cannot be loops in time. This ordering of times is forced upon the plain operator insertions.

[7] Note that we assumed the two consecutive times $t'_\mathrm{o}, t_\mathrm{o}$ before and after the operator insertion as approximately equal. Consequently, the missing Hamiltonian contribution $-(t_\mathrm{o} - t_{\mathrm{o}'})H$ at the operator insertion is small and can be disregarded.

[8] To compute expectation values of quantum operators which are not in proper time order with the path integral is more laborious, but could be handled by integrating over the fields in the intermediate regions of time more than once in order to go back and forth in time.

Equal-Time Commutators. The strict built-in ordering of times makes commutation relations between the operators irrelevant. The only operator algebra we can possibly consider is at equal times. Let us therefore understand how to realise the fundamental commutation relation

$$[\hat{q}, \hat{p}] = \mathring{\imath}\hbar. \tag{1.59}$$

We would like to insert the operator $\hat{O} = [\hat{q}, \hat{p}]$ into the path integral. Unfortunately, there is no classical equivalent $O(q, p)$ to $\hat{O}$ except for the number $\mathring{\imath}\hbar$, but that would amount to postulation rather than derivation.

The trick is to separate the times slightly:[9]

$$O(t) = q(t)p(t - \epsilon) - p(t + \epsilon)q(t). \tag{1.60}$$

The intrinsic time ordering then puts the operators into their desired order. In fact, the discretised path integral knows about the ordering problem and does not even permit ambiguous orderings of positions and momenta: The position variables q_k were defined at times t_k. The associated momenta p_k, however, are not located at t_k, but rather at some time between t_k and t_{k-1} (see earlier figure). This fact is most evident in the expression for the discretised action

$$S_{n,0}[q, p] := \sum_{k=1}^{n} \big(-(t_k - t_{k-1})H(\bar{q}_k, p_k) + p_k(q_k - q_{k-1})\big). \tag{1.61}$$

The above operator is therefore discretised as follows

$$O = q_k p_k - p_{k+1} q_k. \tag{1.62}$$

We insert the operator into the discretised path integral for the transition amplitude $\langle q_\mathrm{f}, t_\mathrm{f}|\hat{O}|q_\mathrm{i}, t_\mathrm{i}\rangle$ of free non-relativistic particle. We can perform most integrations trivially as before

$$\begin{aligned} &\int \frac{\mathrm{d}p_k}{2\pi\hbar}\,\mathrm{d}q_k\,\frac{\mathrm{d}p_{k+1}}{2\pi\hbar}\,q_k(p_k - p_{k+1}) \\ &\quad\cdot \exp\big[-\tfrac{\mathring{\imath}}{2}\hbar^{-1}m^{-1}\big((t_n - t_k)p_{k+1}^2 + (t_k - t_0)p_k^2\big)\big] \\ &\quad\cdot \exp\big(\mathring{\imath}\hbar^{-1}(p_{k+1}q_n - p_k q_0)\big)\exp\big(\mathring{\imath}\hbar^{-1}(p_k - p_{k+1})q_k\big). \end{aligned} \tag{1.63}$$

We then convert the factor q_k to a differential operator acting on the latter exponent, and perform the integral over q_k as the derivative of a delta-function

$$\begin{aligned} &\int \frac{\mathrm{d}p_k}{2\pi\hbar}\,\mathrm{d}p_{k+1}\,(p_k - p_{k+1}) \\ &\quad\cdot \exp\big[-\tfrac{\mathring{\imath}}{2}\hbar^{-1}m^{-1}\big((t_n - t_k)p_{k+1}^2 + (t_k - t_0)p_k^2\big)\big] \\ &\quad\cdot \exp\big(\mathring{\imath}\hbar^{-1}(p_{k+1}q_n - p_k q_0)\big)\left[-\mathring{\imath}\hbar\,\frac{\partial}{\partial p_k}\,\delta(p_{k+1} - p_k)\right]. \end{aligned} \tag{1.64}$$

[9] In fact, the product $q(t)p(t)$ at equal times is ill-defined in the continuous path integral since the insertion $q(t)p(t + \delta t)$ is discontinuous at $\delta t = 0$ (by precisely $\mathring{\imath}\hbar$).

Now the only object that protects the factor $(p_k - p_{k+1})$ from vanishing by means of the delta-function $\delta(p_{k+1} - p_k)$ is the derivative $\partial/\partial p_k$. We therefore integrate by parts and let the derivative act on the remainder of the integrand. Unless it hits the factor $(p_k - p_{k+1})$, the integral must vanish, hence

$$\begin{aligned} &\int \frac{\mathrm{d}p_k}{2\pi\hbar}\,\mathrm{d}p_{k+1}\,\delta(p_{k+1}-p_k)\left[\mathring{\imath}\hbar\,\frac{\partial}{\partial p_k}\,(p_k - p_{k+1})\right] \\ &\qquad \cdot \exp\bigl(-\mathring{\imath}\hbar^{-1}(t_n - t_0)p_k^2/2m + \mathring{\imath}\hbar^{-1}p_k(q_n - q_0)\bigr) \\ &= \mathring{\imath}\hbar \int \frac{\mathrm{d}p}{2\pi\hbar}\exp\bigl(\mathring{\imath}\hbar^{-1}\bigl(-(t_n - t_0)p^2/2m + p(q_n - q_0)\bigr)\bigr). \end{aligned} \tag{1.65}$$

This is precisely $\mathring{\imath}\hbar$ times the transition amplitude $A_{n,0}$. Hence we learn from the path integral that

$$\langle q_\mathrm{f}, t_\mathrm{f}|[\hat{q},\hat{p}]|q_\mathrm{i}, t_\mathrm{i}\rangle = \mathring{\imath}\hbar\langle q_\mathrm{f}, t_\mathrm{f}|q_\mathrm{i}, t_\mathrm{i}\rangle, \tag{1.66}$$

which is fully consistent with canonical quantisation.

This result shows (once again) that quantisation of a classical operator $O(q,p)$ depends crucially on the discretisation $O(\bar{q}_k, p_k)$. The precise choice of $\bar{q}_k$, whether to use q_{k-1}, q_k, their arithmetic mean or something else, has a similar effect as operator ordering in the canonical framework. However, since the canonical commutator $[\hat{q},\hat{p}] = \mathring{\imath}\hbar$ is simple enough, one can always add appropriate terms of $\mathcal{O}(\hbar)$ to $O(q,p)$ to make any given discretisation $O(\bar{q}_k, p_k)$ correspond to the desired quantum operator $\hat{O}$.

2 Path Integral for Fields

In this chapter we generalise the path integral to quantum fields and introduce generating functionals to collect large classes of quantum correlators into uniform expressions.

2.1 Time-Ordered Correlators

We know how to express a quantum mechanical transition amplitude with the path integral. This generalises straight-forwardly to fields (we set $\hbar = 1$ for convenience)

$$\begin{aligned}&\langle \Psi_{\mathrm{f}}, t_{\mathrm{f}}|\, \mathrm{T}(\hat{O}_1 \dots \hat{O}_n)|\Psi_{\mathrm{i}}, t_{\mathrm{i}}\rangle \\ &= \int \mathrm{D}\Psi\, O_1[\Psi] \dots O_n[\Psi] \exp\bigl(\mathring{\imath} S_{\mathrm{f,i}}[\Psi]\bigr).\end{aligned} \tag{2.1}$$

Here $\Psi_{\mathrm{i,f}}$ are the spatial fields at the initial and final time slices, whereas Ψ is a field in spacetime which interpolates between the fields $\Psi_{\mathrm{i}} = \Psi(t_{\mathrm{i}})$ and $\Psi_{\mathrm{f}} = \Psi(t_{\mathrm{f}})$. The action can be written as an integral over the Lagrangian (density)

$$S_{\mathrm{f,i}}[\Psi] := \int_{t_{\mathrm{i}}}^{t_{\mathrm{f}}} \mathrm{d}t\, L\bigl[\Psi(t), \dot{\Psi}(t)\bigr] = \int_{\mathrm{i}}^{\mathrm{f}} \mathrm{d}x^D\, \mathcal{L}\bigl(\Psi(x), \partial_\mu \Psi(x)\bigr). \tag{2.2}$$

This expression is almost covariant, but it still makes reference to two particular time slices. Moreover, we are usually not so much interested in transition amplitudes between particular field configurations, but rather in time-ordered correlators

$$\langle O_1 \dots O_n\rangle := \langle 0|\, \mathrm{T}[\hat{O}_1 \dots \hat{O}_n]|0\rangle. \tag{2.3}$$

To solve these problems, we can apply a trick we have learned in QFT I: A generic state such as $|\Psi_{\mathrm{i}}, t_{\mathrm{i}}\rangle$ can be expected to have some overlap with the ground state $|0\rangle$. Letting the state evolve for some time while adding some friction (let the time have a small imaginary component) makes the state decay to its lowest-energy contribution, i.e. the ground state.

We therefore take the limit $t_{\mathrm{f,i}} \to \pm\infty(1 - \mathring{\imath}\epsilon)$ and obtain a familiar expression for the time-ordered correlator

$$\langle O_1 \dots O_n\rangle = \frac{\displaystyle\int \mathrm{D}\Psi\, O_1[\Psi] \dots O_n[\Psi] \exp\bigl(\mathring{\imath} S[\Psi]\bigr)}{\displaystyle\int \mathrm{D}\Psi \exp\bigl(\mathring{\imath} S[\Psi]\bigr)}. \tag{2.4}$$

Here the path integrals integrate over fields Ψ defined for all of spacetime, and the action is the integral of the Lagrangian density over all of spacetime

$$S[\Psi] = \int \mathrm{d}x^D \, \mathcal{L}\big(\Psi(x), \partial_\mu \Psi(x)\big). \tag{2.5}$$

The term in the denominator accounts for the overlap of the initial and final states with the ground state. It evidently takes care of proper normalisation $\langle 1 \rangle = 1$. Even more, it conveniently eliminates any constant factor in the integration measure $\mathrm{D}\Psi$ allowing us to be somewhat sloppy in defining the latter.

As discussed in QFT I, the slight tilting of the time axis into the complex plane selects Feynman propagators as Green functions. When we keep this in mind, we do not to consider the tilting anymore. The path integral expression for time-ordered[1] correlators is thus perfectly relativistic.

The main application of the path integral in quantum field theory is to compute time-ordered vacuum expectation values. It may also be used to compute different quantities by specifying alternative boundary conditions for the integration over fields Ψ.

2.2 Sources and Generating Functional

In principle, we can now compute correlators such as

$$\big\langle \Psi(x)\, \Psi(y) \big\rangle = \frac{\displaystyle\int \mathrm{D}\Psi \, \Psi(x)\, \Psi(y) \exp\big(\mathring{\imath} S[\Psi]\big)}{\displaystyle\int \mathrm{D}\Psi \exp\big(\mathring{\imath} S[\Psi]\big)} \,. \tag{2.6}$$

In a free theory this expression amounts to performing Gaußian integrals with polynomial prefactors. We will see that the integral effectively evaluates via Wick's theorem where spacetime derivatives of Ψ can be evaluated on the correlation function without derivatives. However, interaction terms in the action $S[\Psi]$ lead to difficult integrals even in the path integral with discretised time.

Sources. Fortunately, there is a standard trick to insert factors in front of the exponential factor using source terms. We define the generating functional $Z[J]$

$$Z[J] := \int \mathrm{D}\Psi \exp\big(\mathring{\imath} S[\Psi] + \mathring{\imath} S_{\mathrm{src}}[\Psi, J]\big) \tag{2.7}$$

as a standard path integral but with an additional source term in the action[2]

$$S_{\mathrm{src}}[\Psi, J] := \int \mathrm{d}x^D \, \Psi(x)\, J(x). \tag{2.8}$$

[1] Here time ordering can be interpreted as causality since fields commute outside the light-cone.

[2] The field $J(x)$ is the same source field as discussed in QFT I in connection to propagators.

This source term has the simple property that a functional derivative with respect to the source $J(x)$ produces precisely the field $\Psi(x)$ at the same location

$$\frac{\delta S_{\text{src}}[\Psi, J]}{\delta J(x)} = \int \mathrm{d}y^D \, \Psi(y) \, \delta^D(x-y) = \Psi(x), \tag{2.9}$$

where we have used the elementary functional derivative relation

$$\frac{\delta J(x)}{\delta J(y)} = \delta^D(x-y). \tag{2.10}$$

When the source action is in the exponent, the functional derivative brings down one power of Ψ without altering the exponent

$$\frac{-\mathring{\imath}\delta}{\delta J(x)} \exp\bigl(\mathring{\imath} S_{\text{src}}[\Psi, J]\bigr) = \Psi(x) \exp\bigl(\mathring{\imath} S_{\text{src}}[\Psi, J]\bigr). \tag{2.11}$$

Moreover, the source field does not appear in the original action. Hence, functional derivatives of $Z[J]$ with respect to the source insert factors of $\Psi(x)$ into the path integral

$$\Psi(x) \simeq \frac{-\mathring{\imath}\delta}{\delta J(x)} \,. \tag{2.12}$$

For example, we express two insertions as a double functional derivative of $Z[J]$

$$\frac{-\mathring{\imath}\delta}{\delta J(x)} \frac{-\mathring{\imath}\delta}{\delta J(y)} \, Z[J] = \int \mathrm{D}\Psi \, \Psi(x) \, \Psi(y) \exp\bigl(\mathring{\imath} S[\Psi] + \mathring{\imath} S_{\text{src}}[\Psi, J]\bigr). \tag{2.13}$$

Now we still need to get rid of the source term in the exponent by setting $J = 0$. The time-ordered two-point correlator with proper normalisation term finally takes the form

$$\bigl\langle \Psi(x) \, \Psi(y) \bigr\rangle = Z[J]^{-1} \frac{-\mathring{\imath}\delta}{\delta J(x)} \frac{-\mathring{\imath}\delta}{\delta J(y)} \, Z[J] \bigg|_{J=0} . \tag{2.14}$$

Free Scalar Field. We are now in a position to formally write any time-ordered correlator, but how to compute them in practice? We can only expect to obtain an exact expression for free fields. For simplicity, consider the scalar field; other types of fields will work analogously. The action with source term reads

$$S_0[\phi] + S_{\text{src}}[\phi, j] = \int \mathrm{d}x^D \, \bigl(-\tfrac{1}{2} (\partial\phi)^2 - \tfrac{1}{2} m^2 \phi^2 + j\phi \bigr). \tag{2.15}$$

By partial integration we can make all the derivatives act on a single field

$$S_0[\phi] + S_{\text{src}}[\phi, j] = \int \mathrm{d}x^D \, \bigl(\tfrac{1}{2} \phi (\partial^2 - m^2) \phi + \phi j \bigr). \tag{2.16}$$

This is a Gaußian integral which can be performed by shifting the integration variable ϕ. A complication is that the kernel of the Gaußian function is the derivative operator $(-\partial^2 + m^2)$, and result of the integral depends on its inverse.

We have already determined its inverse in QFT I, it is the propagator $G_{\mathrm{F}}(x-y)$ satisfying[3]

$$(-\partial^2 + m^2) G_{\mathrm{F}}(x-y) = \delta^D(x-y). \tag{2.17}$$

Hence we shift the field

$$\phi(x) = \tilde{\phi}(x) + \int \mathrm{d}y^D \, G_{\mathrm{F}}(x-y)\, j(y), \tag{2.18}$$

and substitute it into the action

$$\begin{aligned} S_0[\phi] + S_{\mathrm{src}}[\phi, j] &= \int \mathrm{d}x^D \, \tfrac{1}{2}\tilde{\phi}(x)\, (\partial^2 - m^2)\tilde{\phi}(x) + W_0[j], \\ W_0[j] &= \int \mathrm{d}x^D \, \mathrm{d}y^D \, \tfrac{1}{2} j(x)\, G_{\mathrm{F}}(x-y)\, j(y). \end{aligned} \tag{2.19}$$

As $\tilde{\phi}$ and j are now well-separated, we can perform the integral over $\tilde{\phi}$. Moreover, we expect that the integration measure does not change, $\mathrm{D}\phi = \mathrm{D}\tilde{\phi}$, when the integration variable is merely shifted, $\phi \to \tilde{\phi}$. Up to the overall constant $Z_0[0]$ we thus get

$$Z_0[j] = Z_0[0] \exp\bigl(\mathring{\imath} W_0[j]\bigr). \tag{2.20}$$

A derivation in momentum space is somewhat simpler because the Gaußian kernel is automatically diagonal. The momentum-space version of $W[j]$ is[4]

$$W_0[j] = \int \frac{\mathrm{d}p^D}{(2\pi)^D} \frac{\frac{1}{2} j(p)\, j(-p)}{p^2 + m^2 - \mathring{\imath}\epsilon} = \int \frac{\mathrm{d}p^D}{(2\pi)^D} \frac{\frac{1}{2}\bigl|j(p)\bigr|^2}{p^2 + m^2 - \mathring{\imath}\epsilon}\,. \tag{2.21}$$

Formally, the prefactor reads[5]

$$Z_0[0] \sim \frac{1}{\sqrt{\mathrm{Det}[-\partial^2 + m^2 - \mathring{\imath}\epsilon]}}\,. \tag{2.22}$$

Wick's Theorem. Let us now compute $\langle \Psi(x)\Psi(y)\rangle$ for the free scalar field. We perform two functional derivatives[6]

$$\begin{aligned} &\frac{-\mathring{\imath}\delta}{\delta J(y)} \frac{-\mathring{\imath}\delta}{\delta J(x)} Z_0[J] \\ &= \frac{-\mathring{\imath}\delta}{\delta J(y)} \int \mathrm{d}z^D \, G_{\mathrm{F}}(x-z)\, J(z)\, Z_0[J] \\ &= -\mathring{\imath} G_{\mathrm{F}}(x-y)\, Z_0[J] \\ &\quad + \int \mathrm{d}z^D \, \mathrm{d}z'^D \, G_{\mathrm{F}}(x-z)\, J(z)\, G_{\mathrm{F}}(y-z')\, J(z')\, Z_0[J]. \end{aligned} \tag{2.23}$$

[3] Due to the tilting of the time axis into the complex plane, we have to choose the Feynman propagator.

[4] The factor of $1/2$ compensates the double counting of $|j(p)| = |j(-p)|$ for a real field $\phi^*(x) = \phi(x)$.

[5] This statement is more or less tautological in QFT since the determinant of an operator is commonly defined via a Gaußian integral. Here, $Z_0[0]$ is a constant (independent of the other fields) and can therefore be ignored.

[6] There are two equivalent factors of J in the exponent of $Z_0[J]$, so the functional derivative acting on them produces a factor of 2 to be cancelled by the prefactor of $1/2$.

Now divide by $Z_0[J]$ and set $J = 0$ to obtain the correlator

$$\big\langle \phi(x)\, \phi(y) \big\rangle = -\mathring{\imath} G_{\mathrm{F}}(x-y). \tag{2.24}$$

This is precisely the expected result.

We can also perform the exercise with more than two fields. The result agrees with Wick's theorem. In fact, the form of $Z_0[J]$ as the exponent of $W_0[J]$ which is a quadratic monomial of J can be viewed as the functional formulation of Wick's theorem:

- The first derivative pulls down a linear term $\delta W_0/\delta J$ from the quadratic exponent.
- Subsequent derivatives can pull further linear terms $\delta W_0/\delta J$ from the exponent. They can also hit the remaining J in some previously pulled $\delta W_0/\delta J$ leaving behind a Feynman propagator $\delta^2 W_0/\delta J^2 \simeq -\mathring{\imath} G_{\mathrm{F}}(x_k - x_l)$.
- Any first derivative term $\delta W_0/\delta J$ that remains after the functional derivatives cause the expression to vanish when $J = 0$.
- As every exponent requires two functional derivatives, all fields $\phi(x_k)$ must be Wick contracted to some other field.
- The product rule of derivatives takes care of the sum of all combinations.
- Setting $J = 0$ in the end corresponds to the time-ordered correlator; it removes all non-trivial normal ordered terms.
- The above arguments are independent of the type of field; other fields work analogously by using their Feynman propagator instead.

Feynman Graphs. Let us introduce a graphical notation for the terms in the generating functional:

- A cross labelled x attached to some object F by a short line indicates an argument x of F

$$\longrightarrow \quad F(x). \tag{2.25}$$

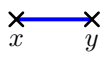
- A solid dot represents a source field $\mathring{\imath} J(x)$. When connected by a short line to some object F, multiply by the corresponding function $F(x)$ and integrate over all positions x

$$\longrightarrow \quad \int \mathrm{d}x^D\, F(x)\, \mathring{\imath} J(x). \tag{2.26}$$

- An empty dot represents a functional derivative $-\mathring{\imath}\delta/\delta J(x)$ which effectively describes a field insertion $\Psi(x)$. When connected to some object F, multiply by $F(x)$ and integrate over all positions x

$$\longrightarrow \quad \int \mathrm{d}x^D\, F(x)\, \frac{-\mathring{\imath}\delta}{\delta J(x)}\,. \tag{2.27}$$

- A long fat (blue) line represents a Feynman propagator $-\mathring{\imath} G_{\mathrm{F}}(x,y)$ between two points x and y



$$\longrightarrow \quad -\mathring{\imath} G_{\mathrm{F}}(x,y). \tag{2.28}$$

- A short thin (black) line represents a direct link to an object, such as a position argument. Alternatively, one may view this line as a delta-function connecting its two end-points

$$\underset{x}{\times}\!\!-\!\!\underset{y}{\times} \quad \longrightarrow \quad \delta^D(x-y). \tag{2.29}$$

Therefore, the exponent $W_0[J]$ in the generating functional is represented by

$$W_0[J] = -\tfrac{\mathring{\imath}}{2}\ \bullet\!\!-\!\!\bullet\ . \tag{2.30}$$

The generating functional $Z_0[J]$ itself is represented by a series of terms with increasing number of propagators

$$Z_0[J] = \sum_{n=0}^{\infty} \frac{Z_0[0]}{n!\,2^n} \Big\} n = Z_0[0]\exp\big(\tfrac{1}{2}\ \bullet\!\!-\!\!\bullet\big). \tag{2.31}$$

The field $\Psi(x)$ is represented by a functional derivative with respect to the source $J(x)$ and thus by the symbol

$$\Psi(x) \simeq -\frac{\mathring{\imath}\delta}{\delta J(x)} = \underset{x}{\times}\!\!-\!\!\circ\ . \tag{2.32}$$

The functional derivative of $W_0[J]$ yields two equivalent terms which compensate the prefactor of $^1\!/_2$. The second derivative is the Feynman propagator

$$\begin{aligned} \underset{x}{\times}\!\!-\!\!\circ\ \tfrac{1}{2}\ \bullet\!\!-\!\!\bullet &= \tfrac{1}{2}\ \underset{x}{\times}\!\!-\!\!\bullet + \tfrac{1}{2}\ \bullet\!\!-\!\!\underset{x}{\times} = \underset{x}{\times}\!\!-\!\!\bullet\ , \\ \underset{y}{\times}\!\!-\!\!\circ\ \underset{x}{\times}\!\!-\!\!\circ\ \tfrac{1}{2}\ \bullet\!\!-\!\!\bullet &= \underset{y}{\times}\!\!-\!\!\circ\ \underset{x}{\times}\!\!-\!\!\bullet = \underset{x}{\times}\!\!-\!\!\underset{y}{\times}\ . \end{aligned} \tag{2.33}$$

Note that the action of a derivative implies that an empty dot grabs any of the solid dots and effectively replaces it by a cross. All pairings between the empty dot and solid dots are summed over. For the first and second functional derivatives of the generating functional $Z[J]$ we can write

$$\begin{aligned} \underset{x}{\times}\!\!-\!\!\circ\ \exp\big(\tfrac{1}{2}\ \bullet\!\!-\!\!\bullet\big) &= \underset{x}{\times}\!\!-\!\!\bullet\ \exp\big(\tfrac{1}{2}\ \bullet\!\!-\!\!\bullet\big), \\ \underset{y}{\times}\!\!-\!\!\circ\ \underset{x}{\times}\!\!-\!\!\circ\ \exp\big(\tfrac{1}{2}\ \bullet\!\!-\!\!\bullet\big) &= \Big(\underset{x}{\times}\!\!-\!\!\bullet\ \underset{y}{\times}\!\!-\!\!\bullet + \underset{x}{\times}\!\!-\!\!\underset{y}{\times}\Big) \\ &\quad \cdot \exp\big(\tfrac{1}{2}\ \bullet\!\!-\!\!\bullet\big). \end{aligned} \tag{2.34}$$

Setting $J = 0$ in the above expressions shows that

$$\big\langle \phi(x)\phi(y) \big\rangle = \underset{x}{\times}\!\!-\!\!\underset{y}{\times}\ . \tag{2.35}$$

2.3 Fermionic Integrals

Before we go on to interactions, let us briefly comment on path integrals for fermionic fields. They work very analogously to bosonic integrals, but there are some important differences.

Graßmann Variables. The distinction between bosonic and fermionic fields in QFT is that the former obey commutation relations while the latter obey anti-commutation relations. For the path integral we work with classical fields: In the classical limit, bosonic fields commute with each other, and they are represented by ordinary numbers. Correspondingly, classical fermionic fields anti-commute with each other. Fermionic fields therefore should take values in anti-commuting numbers, the so-called odd Graßmann numbers, whereas bosonic fields are represented by even Graßmann numbers.

Let us briefly review Graßmann numbers:

- A Graßmann number a can be either even ($\mathbb{Z}_2$-grading $|a| = 0$) or odd ($\mathbb{Z}_2$-grading $|a| = 1$).
- Graßmann numbers can be added as usual (one typically does not mix numbers of different grading).
- The product of two Graßmann numbers is commutative unless both factors are odd in which case the product is anti-commutative.

Consider a set of odd Graßmann variables θ_k. They mutually anti-commute

$$\{\theta_j, \theta_k\} = 0 \qquad \text{or} \qquad \theta_j \theta_k = -\theta_k \theta_j. \tag{2.36}$$

For each variable θ_k we define a corresponding partial derivative $\partial/\partial\theta_k$ by the relation

$$\left[\frac{\partial}{\partial\theta_j}\,\theta_k\right] = \delta_k^j. \tag{2.37}$$

By assuming the derivatives to be operators acting on anything on their right, we may also phrase the defining property as the algebraic anti-commutation relation[7]

$$\frac{\partial}{\partial\theta_j}\,\theta_k + \theta_k\,\frac{\partial}{\partial\theta_j} = \delta_k^j. \tag{2.38}$$

This makes evident that the derivatives themselves are odd Graßmann objects, and therefore they anti-commute

$$\frac{\partial}{\partial\theta_j}\,\frac{\partial}{\partial\theta_k} = -\,\frac{\partial}{\partial\theta_k}\,\frac{\partial}{\partial\theta_j}\,. \tag{2.39}$$

This relation is analogous to the symmetry of partial derivatives for several ordinary (even) numbers x_k.

We will be interested in continuous fermionic fields $\Psi(x)$ rather than some discrete set of fermionic variables θ_k. Here, the transition between arrays and functions is as for bosonic variables: One may approximate a continuous field $\Psi(x)$ on a collection of points x_k as the variables $\Psi_k = \Psi(x_k)$. The field is obtained by going to the limit where the points x_k densely populate the desired domain. A field thus acts as an infinitely large collection of variables. For functional derivatives for fermionic fields we will assume the fundamental relation

$$\frac{\delta\Psi(x)}{\delta\Psi(y)} = \delta^D(x-y). \tag{2.40}$$

[7]This relation holds with any further term multiplied to the right of the expression. Applying the Leibniz rule to the anti-commutator ensures that the derivative $\partial/\partial\theta_j$ only acts on θ_k.

Berezin Integral. To define a fermionic path integral we need to understand how to integrate over an odd Graßmann variable θ

$$\int \mathrm{d}\theta\, f(\theta) = ?\,. \tag{2.41}$$

It makes sense to require that the integral of a total derivative vanishes[8]

$$\int \mathrm{d}\theta\, \frac{\partial}{\partial\theta}\, f(\theta) = 0. \tag{2.42}$$

Now due to their anti-commutativity, odd Graßmann numbers square to zero, $\theta^2 = 0$, and hence a generic function $f(\theta)$ can be expanded as $f(\theta) = f_0 + \theta f_1$ with two coefficients f_0 and f_1. We substitute this expansion into the integral of a total derivative

$$0 = \int \mathrm{d}\theta\, \frac{\partial}{\partial\theta}\, (f_0 + \theta f_1) = \int \mathrm{d}\theta\, f_1. \tag{2.43}$$

It tells us that the integral of a constant must vanish. We can now integrate a generic function

$$\int \mathrm{d}\theta\, (f_0 + \theta f_1) = \left(\int \mathrm{d}\theta\, \theta\right) f_1. \tag{2.44}$$

The integral $\int \mathrm{d}\theta\, \theta$ is some undetermined factor, it makes sense to define it as 1. The curious result is that integration of odd Graßmann variables is equivalent to differentiation[9]

$$\int \mathrm{d}\theta\, f(\theta) = \frac{\partial}{\partial\theta}\, f(\theta). \tag{2.45}$$

The so-called Berezin integral over odd Graßmann variables behaves in many other respects like the standard bosonic integral. For the path integral in quantum field theory, the most important concepts are Fourier integrals, delta-functions and Gaußian integrals. Let us consider these now:

Delta-Functions. We can convince ourselves that the defining property of the delta-function

$$\int \mathrm{d}\theta\, \delta(\theta - \alpha) f(\theta) = f(\alpha) \tag{2.46}$$

is solved trivially by[10]

$$\delta(\theta) = \theta. \tag{2.47}$$

Under variable transformations this delta-function behaves as

$$\delta\big(\phi(\theta)\big) = \frac{\partial\phi}{\partial\theta}\, \delta(\theta - \theta_0). \tag{2.48}$$

This is analogous to the transformation of the bosonic delta-function except that the Jacobian of the transformation multiplies rather than divides the delta-function (and no absolute values are taken).

[8] We disregard conceivable boundary terms; Graßmann numbers turn out to have trivial topology.

[9] This implies that the integration measure $\mathrm{d}\theta$ has the dimension of $\partial/\partial\theta$ or $1/\theta$.

[10] Note that the order of terms in $\delta(\theta - \alpha)$ does matter since $\delta(\alpha - \theta) = -\delta(\theta - \alpha)$.

Fourier Integrals. A plain Fourier integral for the Fourier transformation between the odd variables $\theta \to \alpha$ produces a delta-function as usual

$$\int \mathrm{d}\theta \exp(c\theta\alpha) = \int \mathrm{d}\theta \left(1 + c\theta\alpha\right) = c\alpha = c\,\delta(\alpha). \tag{2.49}$$

Note that the coefficient c of the exponent can be an arbitrary (Graßmann even) number.

Gaußian Integrals. To define a Gaußian integral, we need at least two Graßmann odd variables, otherwise the quadratic exponent would vanish by construction. For the simplest Gaußian integral we obtain

$$\int \mathrm{d}\theta_2\, \mathrm{d}\theta_1 \exp(a\theta_1\theta_2) = a. \tag{2.50}$$

To make this expression more reminiscent of a usual n-dimensional Gaußian integral, let us introduce a 2×2 matrix A

$$A = \begin{pmatrix} 0 & +a \\ -a & 0 \end{pmatrix}. \tag{2.51}$$

The result can be expressed as[11]

$$\int \mathrm{d}\theta^n \exp(\tfrac{1}{2}\theta^\intercal A\theta) \sim \sqrt{\det(A)}. \tag{2.52}$$

This result in fact applies to general fermionic Gaußian integrals defined in terms of an $n \times n$ anti-symmetric matrix A.[12] The expression is very similar to the bosonic n-dimensional Gaußian integral for a symmetric matrix S

$$\int \mathrm{d}x^n \exp(-\tfrac{1}{2}x^\intercal S x) \sim \frac{1}{\sqrt{\det(S)}}\,. \tag{2.53}$$

The crucial difference is that the determinant of the matrix appears with positive rather than negative exponent. Moreover, the matrix A does not need to fulfil any positivity requirements as the odd integral is always well-defined.

For complex fields one usually encounters complex Gaußian integrals. One may decompose them into real Gaußian integrals of twice the dimension. For odd integration variables one finds

$$\int \mathrm{d}\theta^n\, \mathrm{d}\bar{\theta}^n \exp(\bar{\theta} M \theta) \sim \det(M), \tag{2.54}$$

whereas the corresponding integral for even variables reads

$$\int \mathrm{d}x^n\, \mathrm{d}\bar{x}^n \exp(-\bar{x} M x) \sim \frac{1}{\det(M)}\,. \tag{2.55}$$

[11] The determinant of an anti-symmetric matrix A is the square of the so-called Pfaffian $\mathrm{Pf}\, A$. The result of the Gaußian integral is thus the Pfaffian.

[12] A fermionic Gaußian integral requires an even number of integration variables n because the determinant of an odd-dimensional anti-symmetric matrix is zero.

In the latter bosonic case, the matrix M should obey some positivity constraints to make the integral convergent, whereas the fermionic integral is indifferent to the signature of M.

Eventually, we are interested in fields. The functional form for fermionic Gaußian integrals reads

$$\begin{aligned} \int \mathrm{D}\Psi \exp(\tfrac{1}{2}\Psi{\cdot}A{\cdot}\Psi) &\sim \sqrt{\operatorname{Det} A}, \\ \int \mathrm{D}\Psi\, \mathrm{D}\bar{\Psi} \exp(\bar{\Psi}{\cdot}A{\cdot}\Psi) &\sim \operatorname{Det} A. \end{aligned} \tag{2.56}$$

Here, $\Psi{\cdot}A{\cdot}\Psi$ is a quadratic form for the field defined by the operator A. In other words, A is some operator acting on a field, for example the Klein–Gordon or Dirac operator, and the scalar product $\Phi{\cdot}\Psi$ between two fields Φ and Ψ takes the form

$$\Phi{\cdot}\Psi = \int \mathrm{d}x^D\, \Phi(x)\Psi(x). \tag{2.57}$$

Summary. Altogether, when dealing with bosonic and fermionic fields we must pay attention to

- the ordering of fields (and pick up appropriate sign factors for reordering),
- the ordering of derivatives and integration measures,
- the exponents of factors associated to integrals.

Otherwise the procedures are much the same. For example, completion of a square is the essential step to solve Gaußian integrals for free fields.

2.4 Interactions

As soon as interaction terms are added to the Lagrangian, the path integral is not Gaußian anymore, and cannot be performed in general. There is a simple trick to formally implement interactions in the path integral. The obtained expression can be evaluated perturbatively which directly leads to the Feynman rules.

Interaction Terms. We split up the action into a free part and interaction terms

$$S[\Psi] = S_0[\Psi] + S_{\mathrm{int}}[\Psi]. \tag{2.58}$$

The free part S_0 is quadratic in the fields and can therefore be integrated exactly. The interaction part S_{int} contains all the remaining terms.[13] The path integral for the generating functional then reads

$$Z[J] = \int \mathrm{D}\Psi \exp\bigl(\mathring{\imath}S_0[\Psi] + \mathring{\imath}S_{\mathrm{int}}[\Psi] + \mathring{\imath}S_{\mathrm{src}}[\Psi, J]\bigr). \tag{2.59}$$

[13] As discussed in QFT I, the interaction part may as well contain quadratic terms. These terms can be used to counterbalance loop effects such that the free action $S_0 = S_{\mathrm{asymp}}$ describes physical particles accurately in the absence of interactions.

We now split up the exponential into interactions and free terms including sources. We can then replace the field Ψ in the interaction part by a functional derivative with respect to the source J

$$\begin{aligned} Z[J] &= \int \mathrm{D}\Psi \exp\bigl(\mathring{\imath} S_{\mathrm{int}}[\Psi]\bigr) \exp\bigl(\mathring{\imath} S_0[\Psi] + \mathring{\imath} S_{\mathrm{src}}[\Psi, J]\bigr) \\ &= \int \mathrm{D}\Psi \exp\left(\mathring{\imath} S_{\mathrm{int}}\left[\frac{-\mathring{\imath}\delta}{\delta J}\right]\right) \exp\bigl(\mathring{\imath} S_0[\Psi] + \mathring{\imath} S_{\mathrm{src}}[\Psi, J]\bigr). \end{aligned} \tag{2.60}$$

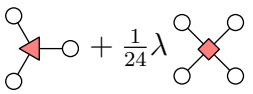
Now the integration variable Ψ appears only in the free and source contributions whereas the interactions are formulated in terms of the source field J. We can thus pull the interactions out of the integral and perform the free integral as described above

$$\begin{aligned} Z[J] &= \exp\left(\mathring{\imath} S_{\mathrm{int}}\left[\frac{-\mathring{\imath}\delta}{\delta J}\right]\right) Z_0[J], \\ Z_0[J] &= \int \mathrm{D}\Psi \exp\bigl(\mathring{\imath} S_0[\Psi] + \mathring{\imath} S_{\mathrm{src}}[\Psi, J]\bigr) = Z_0[0] \exp\bigl(\mathring{\imath} W_0[J]\bigr). \end{aligned} \tag{2.61}$$

Examples and Feynman Graphs. Let us compute the first few terms of $Z[J]$ in a model with cubic and quartic interactions

$$S_{\mathrm{int}}[\Psi] \simeq \int \mathrm{d}x^D \left(-\tfrac{1}{6}\kappa\,\Psi^3 - \tfrac{1}{24}\lambda\,\Psi^4\right). \tag{2.62}$$

In a graphical notation we could represent this as[14]

$\mathring{\imath} S_{\mathrm{int}}[\Psi] \simeq +\frac{1}{6}\kappa$  $+\,\frac{1}{24}\lambda$, (2.63)

where we have made the dependence on the coupling constants κ and λ explicit as an explicit means of power counting. In position space, both vertices represent the integral $-\mathring{\imath}\int \mathrm{d}x$. The generating functional $Z[J]$ can be expanded in powers of S_{int}

$$Z[J] = \left(1 + \mathring{\imath} S_{\mathrm{int}}\left[\frac{-\mathring{\imath}\delta}{\delta J}\right] - \tfrac{1}{2} S_{\mathrm{int}}\left[\frac{-\mathring{\imath}\delta}{\delta J}\right]^2 + \ldots\right) Z_0[J]. \tag{2.64}$$

In the following we will discuss the individual terms arising from this expression. Evidently, the first term is just the free generating functional $Z_0[J]$.

The second term with a cubic interaction has three functional derivatives $-\mathring{\imath}\delta/\delta J$

[14]Note that the graphical notation uses long, fat, blue lines for propagators and short, thin, black lines for direct local connections.

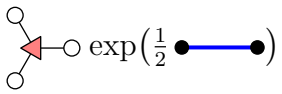
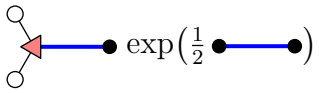
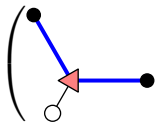
acting on $Z_0[J]$. We let them act one by one:

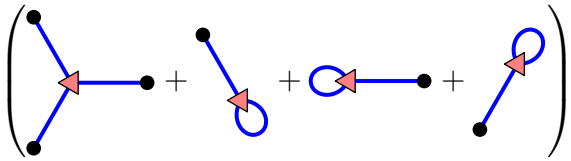
$$\begin{aligned}
&-\tfrac{\mathring{\imath}}{6}\kappa \int \mathrm{d}w \left(\frac{-\mathring{\imath}\delta}{\delta J_w}\right)^3 Z_0[J] \\
&= -\tfrac{\mathring{\imath}}{6}\kappa \int \mathrm{d}w\,\mathrm{d}x \left(\frac{-\mathring{\imath}\delta}{\delta J_w}\right)^2 \left(G_{wx}J_x Z_0[J]\right) \\
&= -\tfrac{\mathring{\imath}}{6}\kappa \int \mathrm{d}w\,\mathrm{d}x\,\mathrm{d}y \left(\frac{-\mathring{\imath}\delta}{\delta J_w}\right) \left(G_{wx}J_x G_{wy}J_y Z_0[J]\right) \\
&\quad -\tfrac{1}{6}\kappa \int \mathrm{d}w \left(\frac{-\mathring{\imath}\delta}{\delta J_w}\right) \left(G_{ww} Z_0[J]\right) \\
&= -\tfrac{\mathring{\imath}}{6}\kappa \int \mathrm{d}w\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\, G_{wx}J_x G_{wy}J_y G_{wz}J_z Z_0[J] \\
&\quad -\tfrac{3}{6}\kappa \int \mathrm{d}w\,\mathrm{d}x\, G_{ww}G_{wx}J_x Z_0[J].
\end{aligned} \tag{2.65}$$

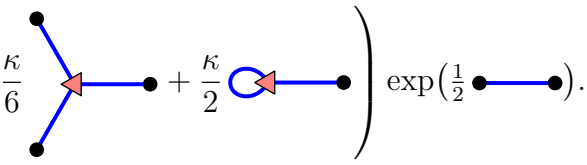
We can also perform the calculation using diagrams. We already understand the free generating functional $Z_0[J]$. It can be viewed as a bag of arbitrarily many propagators

$$Z_0[J] = \exp\big(\tfrac{1}{2}$$ $$\big). \tag{2.66}$$

For the second term, draw a triangle vertex with three source derivatives in front of the bag. Pick some ordering (1,2,3) for the derivatives and let them act on the bag and pull propagators

$$\frac{\kappa}{6}$$  $$= \frac{\kappa}{6}$$  $$= \frac{\kappa}{6}\Big($$  $$\Big) \exp\big(\tfrac{1}{2}$$ $$\big)$$

$$= \frac{\kappa}{6}\Big($$  $$\Big) \cdot \exp\big(\tfrac{1}{2}$$ $$\big)$$

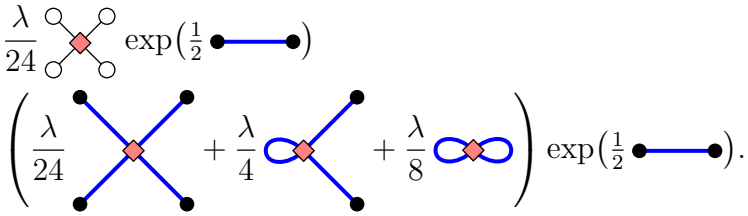
$$= \Big($$  $$\Big) \exp\big(\tfrac{1}{2}$$ $$\big).$$ (2.67)

The construction for a single quartic interaction vertex is similar, there are just

many more (intermediate terms)

$$
\begin{aligned}
&-\tfrac{\mathring{\imath}}{24}\lambda \int \mathrm{d}w \left(\frac{-\mathring{\imath}\delta}{\delta J_w}\right)^4 Z_0[J] \\
&= -\tfrac{\mathring{\imath}}{24}\lambda \int \mathrm{d}w\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}u\, G_{wx}J_xG_{wy}J_yG_{wz}J_zG_{wu}J_uZ_0[J] \\
&\quad -\tfrac{6}{24}\lambda \int \mathrm{d}w\,\mathrm{d}x\,\mathrm{d}y\, G_{ww}G_{wx}J_xG_{wy}J_yZ_0[J] \\
&\quad +\tfrac{3\mathring{\imath}}{24}\lambda \int \mathrm{d}w\, G_{ww}G_{ww}Z_0[J]
\end{aligned}
\tag{2.68}
$$

or graphically



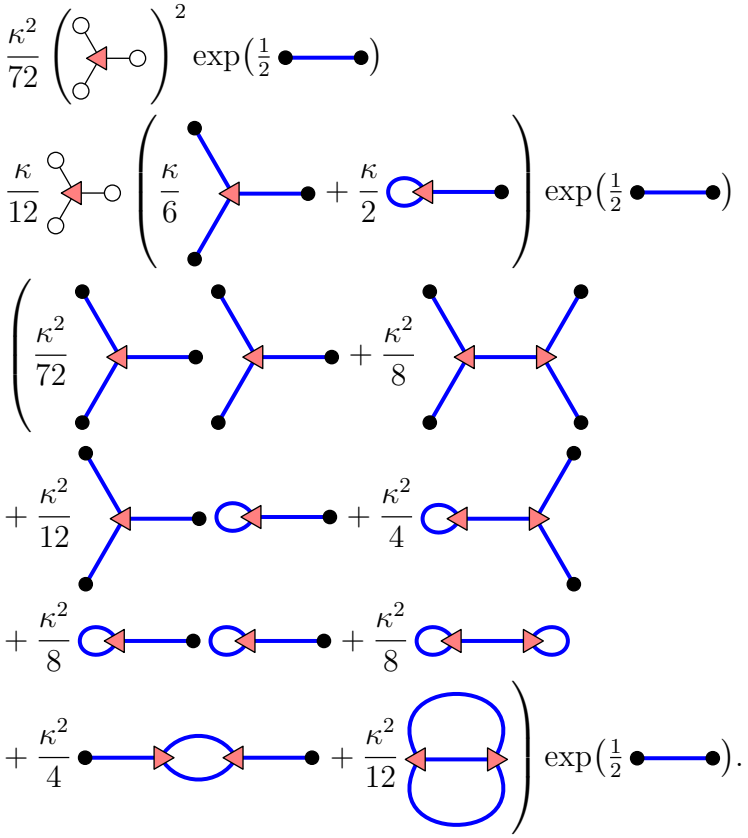
(2.69)

The various instances of equivalent intermediate terms conspire to yield the appropriate symmetry factors for all the graphs that eventually contribute.

Let us finally consider two cubic vertices

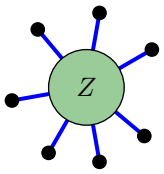
$$
\frac{\mathring{\imath}^2\kappa^2}{2\cdot 6^2}\int \mathrm{d}v\,\mathrm{d}w \left(\frac{-\mathring{\imath}\delta}{\delta J_v}\right)^3\left(\frac{-\mathring{\imath}\delta}{\delta J_w}\right)^3 Z_0[J]. \tag{2.70}
$$

We can start with the single cubic vertex as intermediate result



(2.71)

Altogether the generating functional $Z[J]$ is a collection (sum) of Feynman graphs whose legs are all saturated by source fields J (monomials)

$$Z[J] = \sum$$  $$. \tag{2.72}$$

To obtain a correlation function, all sources J have to be saturated by functional derivatives $-\mathring{\imath}\delta/\delta J$ representing the fields Ψ of the correlator.

Note that symmetry factors are usually different for terms of the generating functional and for the corresponding correlation functions: Source fields J are typically integrated over all of spacetime, thus become indistinguishable and can be freely permuted. This results in large symmetry factors in $Z[J]$. The fields $\Psi(x)$, $\Psi(y)$ in some correlator and corresponding functional derivatives are well distinguishable due to the distinct positions x, y in spacetime. As the external legs cannot be permuted, correlation functions typically have relatively small symmetry factors.

Loop Counting. Next let us understand the role of the quantum parameter $\hbar$ in QFT better. We reinstate $\hbar$ in the generating functional as in the path integral for quantum mechanics discussed above[15]

$$\begin{aligned} Z[J] &= \int \mathrm{D}\Psi \exp\bigl(\mathring{\imath}\hbar^{-1} S[\Psi] + \mathring{\imath}\hbar^{-1} S_{\mathrm{src}}[\Psi, J]\bigr) \\ &= \exp\left(\mathring{\imath}\hbar^{-1} S_{\mathrm{int}}\left[\hbar\,\frac{-\mathring{\imath}\delta}{\delta J}\right]\right) Z_0[J], \\ Z_0[J] &= \exp\left(\tfrac{\mathring{\imath}}{2}\hbar^{-1} \int \mathrm{d}x^D\, \mathrm{d}y^D\, J(x)\, J(y)\, G_{\mathrm{F}}(x,y)\right). \end{aligned} \tag{2.73}$$

Functional derivatives with respect to J receive a factor of $\hbar$. We shall see that the generating functional can be classified by the powers of $\hbar$ as a classical contribution plus a series of quantum corrections.

The above expression for the generating functional shows clearly where powers of $\hbar$ appear within graphs:

- Each interaction vertex contributes a power of $\hbar^{-1}$.
- Each line contributes a power of $\hbar^{-1}$.
- Each junction between a vertex and a line originating from the functional derivative $-\mathring{\imath}\hbar\delta/\delta J$ contributes a power of $\hbar^{+1}$.

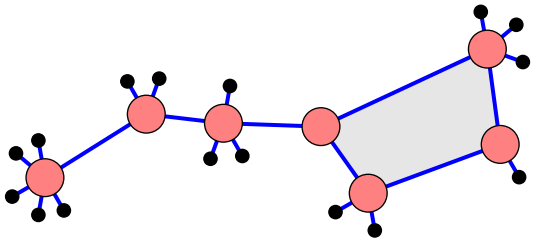
Consider now a graph with

- I internal lines (connecting two vertices),

[15] Of course the action $S[\Psi]$ may also implicitly depend on Planck's constant which would spoil the counting scheme outlined below. Therefore, we set $\hbar = 1$ and only put the (new) constant $\hbar$ in front of the action in the exponent. This $\hbar$ is the formal parameter to count loops.

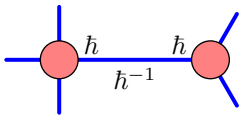
- E external lines (connecting to one vertex only),
- V interaction vertices,
- N connection components,
- L loops.

For example the following graph has the characteristics $I = 7$, $E = 16$, $V = 7$, $N = 1$, $L = 1$:



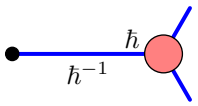
(2.74)

For the counting of powers of $\hbar$ we need to count the vertices, lines and junctions. The latter is not immediately given by the graph parameters, but it can be expressed in terms of the number of internal and external lines: Each internal line contributes $\hbar^{-1}$ for the line itself times $\hbar^{+2}$ from the two junctions with vertices:



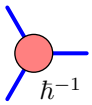
(2.75)

Each external line contributes $\hbar^{-1}$ for the line itself times $\hbar^{+1}$ from a single junction to a vertex:



(2.76)

Each interaction vertex contributes $\hbar^{-1}$:



(2.77)

Altogether we obtain the total power of $\hbar$ associated to the graph

$$\hbar^{I-V}. \tag{2.78}$$

We can reexpress this combination using the number of non-trivial momentum integrals, i.e. the number of loops of the graph. Consider the number of undetermined momenta within a graph:

- one free momentum for each internal and external line (integral),
- one constraint for each vertex (delta-function),
- one constraint for each external line (fixed by momentum inflow),
- relax one constraint for the external lines within each connection component (the constraints due to the vertices already imply overall momentum conservation within this component which determines one of the external momenta in terms of the others).

Performing all trivial momentum integrals which contain delta-functions thus leaves the following number of momentum integrals undetermined

$$L = I + E - V - E + N = I - V + N. \tag{2.79}$$

This combination is the number L of non-trivial momentum integrals, i.e. the number of loops of the graph.

Altogether the powers of $\hbar$ now read

$$\hbar^{I-V} = \hbar^{L-N}. \tag{2.80}$$

This means that each momentum loop is suppressed by one power of $\hbar$. The leading contribution at $L = 0$ is considered to represent classical physics. Moreover the number of connection components N plays a role. We shall soon return to this result.

2.5 Further Generating Functionals

Besides the generating functional $Z[J]$ of all graphs, there are further useful generating functionals which are somewhat simpler to handle and evaluate:

- $W[J]$, the generating functional of connected graphs,
- $T[J]$, the generating functional of connected tree graphs,
- $G[\Psi]$, the generating functional of one-particle irreducible graphs, also known as the (quantum) effective action.

Finally, there is the classical action $S[\Psi]$ which can be viewed as the generating functional of interaction vertices.

Connected Graphs. The connected generating functional $W[J]$ is defined as the logarithm of $Z[J]$ [16]

$$W[J] = -i\hbar \log Z[J], \qquad Z[J] = \exp\bigl(i\hbar^{-1} W[J]\bigr). \tag{2.81}$$

Just like any other generating functional, $W[J]$ can be represented in terms of a sum over Feynman graphs. The difference compared to $Z[J]$ is that $W[J]$ encodes precisely all *connected* Feynman graphs. It is nice to have a formal description of this smaller set of graphs because it allows to easily reproduce all disconnected graphs.

How can this relationship be proved? It is a simple consequence of the symmetry factors of disconnected Feynman graphs. The symmetry factor is the product of

- the symmetry factors of the connected components and
- a factor of $1/n!$ for n equivalent connected components.

[16] This is a pretty general relationship for generating functionals of graphs which also holds in different contexts.

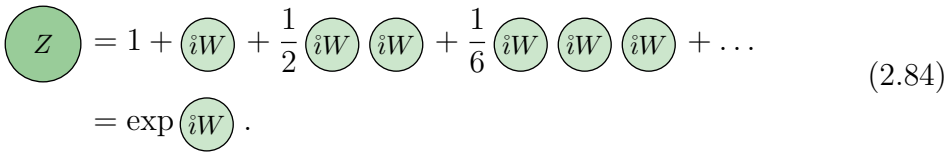
More explicitly, consider a disconnected graph Γ consisting of connected subgraphs Γ_k with multiplicity n_k. The contribution to $Z[J]$ reads[17]

$$\frac{\Gamma[J]}{S(\Gamma)} = \prod_k \frac{1}{n_k!}\left(\frac{\Gamma_k[J]}{S(\Gamma_k)}\right)^{n_k} \in Z[J]. \tag{2.82}$$

Here $S(\Gamma)$ denotes the symmetry factor associated to the graph Γ. The above term actually arises as one term in the multinomial theorem applied to the expansion of the exponential

$$\frac{\Gamma}{S(\Gamma)} \in \frac{1}{n!}\left(\sum\nolimits_k \frac{\Gamma_k}{S(\Gamma_k)}\right)^n \in \exp\left(\sum\nolimits_k \frac{\Gamma_k}{S(\Gamma_k)}\right). \tag{2.83}$$

This means that exponentiating the sum of all connected graphs with appropriate symmetry factors yields the sum of all connected and disconnected graphs with precisely the right symmetry factors. In terms of graphs we can write $Z[J]$ in terms of $W[J]$ as



(2.84)

Tree Graphs. There is a generating functional for the leading classical contributions. This turns out to generate precisely the tree graphs, i.e. those without momentum loops.

The leading classical contributions are the most relevant contributions when $\hbar$ is very small. The quantum constant appears as the inverse $\hbar^{-1}$ in the prefactor to the action in the exponent

$$Z[J] = \int \mathrm{D}\Psi \exp\bigl(\mathring{\imath}\hbar^{-1}S[\Psi] + \mathring{\imath}\hbar^{-1}S_{\mathrm{src}}[\Psi, J]\bigr). \tag{2.85}$$

In the path integral this causes a strongly oscillating integrand. All contributions cancel out almost perfectly unless the exponent is stationary

$$\frac{\delta S[\Psi]}{\delta\Psi(x)} + \frac{\delta S_{\mathrm{src}}[\Psi, J]}{\delta\Psi(x)} = \frac{\delta S[\Psi]}{\delta\Psi(x)} + J(x) = 0. \tag{2.86}$$

Let us assume that there is a single stationary contribution for each source field configuration J which we shall denote by $\Psi = \Psi[J]$.[18] At small $\hbar$ the path integral is dominated by this contribution (up to some irrelevant prefactor)

$$Z[J] \approx \exp\bigl(\mathring{\imath}\hbar^{-1}T[J]\bigr), \tag{2.87}$$

[17] Here, the symbol $X \in Y$ is means "X is a term of the polynomial Y with the same coefficient".

[18] This assumption clearly does not hold because the kinetic terms have a non-trivial kernel (by definition). Each J defines a corresponding equivalence class of solutions Ψ, and clearly not all functions J correspond to actual field configurations. Let us comment on this issue further below.

where we have introduced the functional $T[J]$ for the leading contribution to the exponent

$$T[J] := S\big[\Psi[J]\big] + \int \mathrm{d}x^D \, J(x)\, \Psi[J](x). \tag{2.88}$$

What can we say about $T[J]$?

First of all $T[J]$ is constructed as the leading classical contribution to $W[J]$.

$$T[J] = \lim_{\hbar\to 0} W[J]. \tag{2.89}$$

As such $T[J]$ generates a subclass of the *connected* graphs.

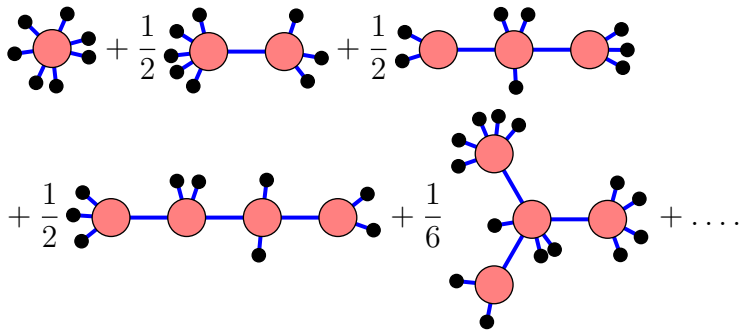
Furthermore, we have learned that the contributions to $Z[J]$ depend on $\hbar$ as $\hbar^{L-N}$. For connected graphs we have $N = 1$, and therefore the graphs in $W[J]$ scale as $\hbar^L$. The limit $\hbar \to 0$ then restricts to graph with $L = 0$, i.e. the graphs in $T[J]$ have no momentum loops. Therefore, $T[J]$ generates precisely the subclass of connected tree graphs within $Z[J]$ or $W[J]$.

Another important observation is that $T[J]$ formally is the Legendre transformation of the action $S[\Psi]$: The source J is constructed as the functional derivative of $S[\Psi]$

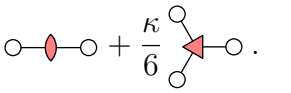
$$J(x) = -\frac{\delta S[\Psi]}{\delta \Psi(x)}\,. \tag{2.90}$$

Moreover, $T[J]$ equals $S[\Psi]$ plus a term $J{\cdot}\Psi = S_{\mathrm{src}}[\Psi, J]$ evaluated at the inverse $\Psi = \Psi[J]$ of the above relation.

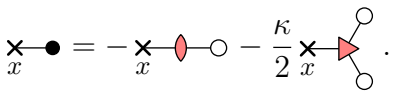
Putting the above insights together, we conclude that the Legendre transformation of some generating functional $S[\Psi]$ is a functional $T[J]$ which generates connected trees from the lines and vertices encoded by $S_0[\Psi]$ and $S_{\mathrm{int}}[\Psi]$, respectively

$T[J] =$  (2.91)

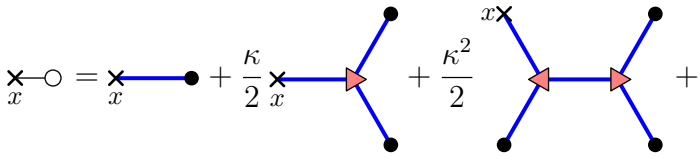
Let us consider this transformation explicitly for the example of a theory with a cubic interaction

$\mathring{\imath}S[\Psi] =$  . (2.92)

The source field J is then defined in terms of the field Ψ

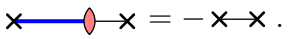
 (2.93)

We need to find the inverse functional $\Psi[J]$ which turns out to be given by the expansion



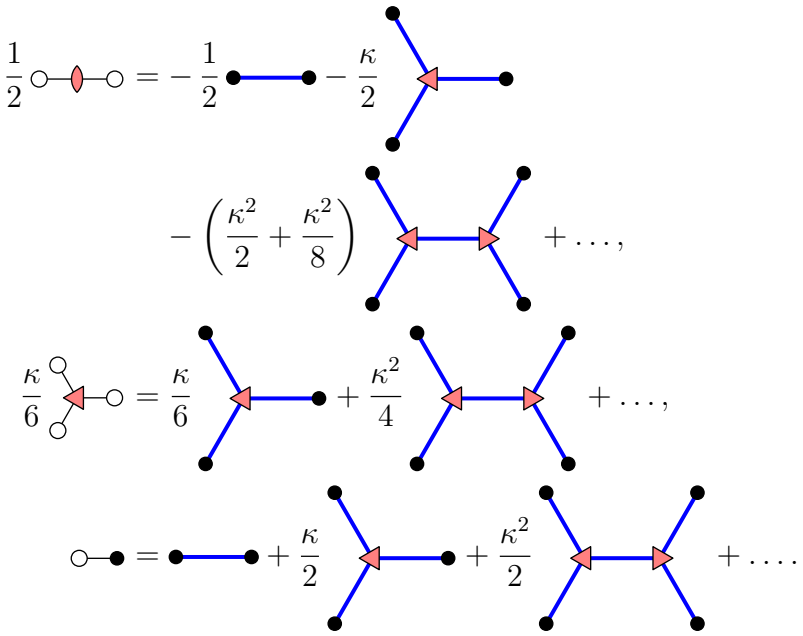
(2.94)

One can confirm by substitution that these two functions are mutually inverse by using the graphical relationship (defining differential equation for Green functions)



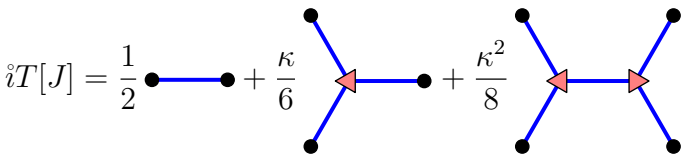
(2.95)

Now substitute $\Psi[J]$ into the two terms of the action $S[\Psi]$ and the Legendre transformation term $S_{\rm src}[\Psi, J]$



(2.96)

The sum of these terms yields the Legendre transform



(2.97)

The various prefactors in the above conspire to yield the canonical symmetry factors of the tree graphs.

Finally, let us discuss whether the Legendre transformation between $S[\Psi]$ and $T[J]$ is well defined. The convexity criterion which is needed to ensure that there is a one-to-one map between Ψ and J clearly does not hold because the kinetic term has a kernel and is not invertible, and not all source configurations J have a corresponding field configuration Ψ. There are at least two useful resolutions to this issue: One may work at a functional level or in terms of complex analysis

where zeros and poles (of the kinetic term and propagator) are essential features of functions which do not preclude taking their inverses in a global, functional sense. For instance, in momentum space the propagator is obtained as the inverse of the kinetic term even though the kinetic term has a zero at the mass shell. This construction works as far as the propagator is not evaluated precisely at the mass shell, and even there the residue has a proper meaning. Alternatively, one can deal with the many-to-one relationship between Ψ and J using pseudo-inverses and restricting J such that it evaluates to zero at the location of the propagator poles (recalling that sources always attach to propagators within diagrams). In terms of physics, this restriction of the source avoids driving the corresponding field to an infinite amplitude.

Effective Action. We have seen that the tree functional $T[J]$ is the classical limit of the connected functional $W[J]$. Furthermore, $T[J]$ is the Legendre transform of the action $S[\Psi]$. It therefore makes sense to consider also the Legendre transform of $W[J]$. This functional is called the effective action $G[\Psi]$.[19]

Define the effective field functional $\Psi[J]$ as the functional derivative of W

$$\Psi[J] = \frac{\delta W}{\delta J}\,[J], \tag{2.98}$$

and denote its inverse by $J[\Psi]$. The field functional $\Psi[J]$ is related to the quantum field Ψ: It is the vacuum expectation value of a single field Ψ in the presence of a source J [20]

$$\Psi[J](x) = \big\langle \Psi(x) \big\rangle_J := Z[J]^{-1}\,\frac{-\mathring{\imath}\hbar\delta}{\delta J(x)}\,Z[J] = \frac{\delta W}{\delta J(x)}\,[J]. \tag{2.99}$$

The effective action is then defined as

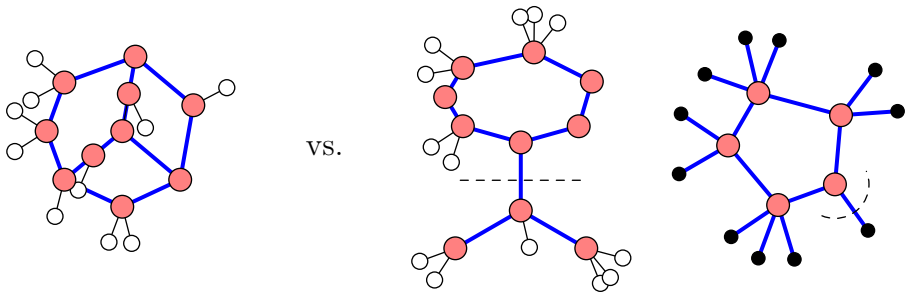
$$G[\Psi] := W\big[J[\Psi]\big] - \int \mathrm{d}x^D\, J[\Psi](x)\,\Psi(x). \tag{2.100}$$

It turns out to generate amputated one-particle irreducible (1PI) Feynman graphs. These are precisely the graphs where cutting a single line will never split the graph into two components. At least two cuts are required to split any graph. Moreover, the external fields Ψ are directly connected to the vertices of the graph without

[19] It is also denoted by Γ. Its argument sometimes takes a different symbol to clearly distinguish it from the quantum field Ψ. However, there is no danger of confusion, and we shall use the same symbol.

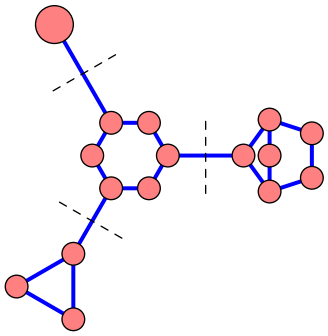
[20] In the presence of a source means promoting the source field J from an auxiliary quantity to an honest constituent of the classical action. A time-ordered vacuum expectation value in the presence of J thus takes the form $\langle O\rangle_J := Z[J]^{-1} O[-\mathring{\imath}\hbar\delta/\delta J] Z[J]$.

Feynman propagators.



(2.101)

How can this be seen? We know that the Legendre transformation switches between vertices ($S[\Psi]$) and connected tree graphs made from these vertices ($T[J]$). Here the starting point is the generating functional of connected graphs $W[J]$ with loops. Let us view the latter as trees made from some effective vertices. How to describe these effective vertices? A defining property of trees is that cutting any line splits a tree into two trees. Therefore, we iteratively split a graph in $W[J]$ into subgraphs by cutting lines such that each cut splits the graph into two components. At the end of the procedure the graph is decomposed into *effective vertices*.[21] These effective vertices have the property that cutting any single line does not split the vertex. Hence $G[\Psi]$ encodes precisely the 1PI graphs.[22]



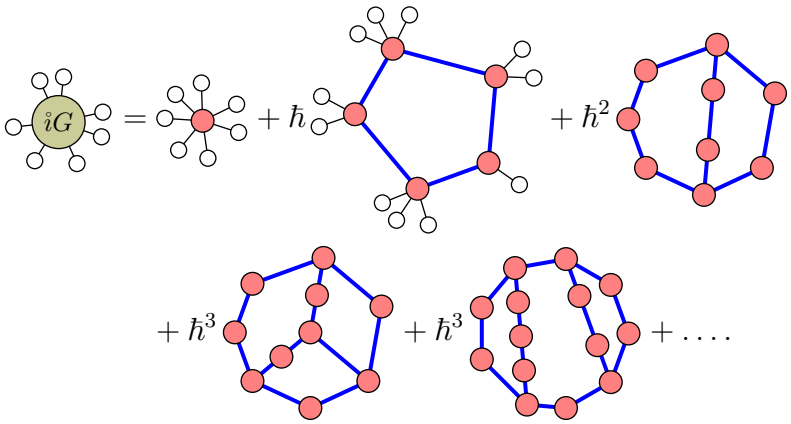
(2.102)

The effective vertices in $G[\Psi]$ certainly contain the elementary vertices in the original action $S[\Psi]$. This is evident since $T[J]$ is the classical limit of $W[J]$, consequently $S[\Psi]$ must be the classical limit of $G[\Psi]$. The higher contributions are

[21] Clearly, this procedure is finite and leads to a unique decomposition of the graph.

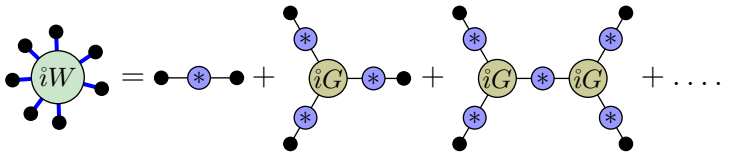
[22] This is is qualitative argument. For a quantitative argument one would need to show that the correct prefactors of the graphs are produced. This turns out to work nicely if the prefactors are given by the natural symmetry factors.

amputated loop graphs



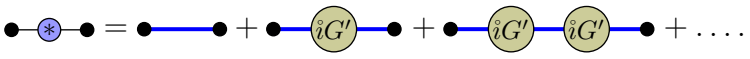
(2.103)

This generating functional comprises a small subset of all graphs. Any connected graph can easily be reassembled as a tree graph consisting of the vertices encoded by the effective action



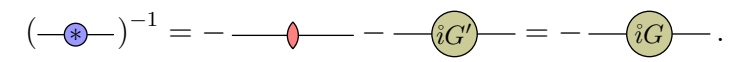
(2.104)

We assume that the effective vertices labelled $i\mathring{G}$ have at least three legs. The *effective propagator* marked by $*$ represents a linear sequence of effective two-point vertices G_2' of arbitrary length



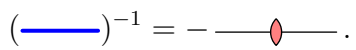
(2.105)

Here, $G_2' = G_2 - S_0$ are those effective two-point vertices which are not already in the kinetic term of the free action S_0.[23] Summing up the geometric series shows that the effective propagator marked by $*$ is minus the inverse of the two-point effective vertex G_2



(2.106)

Note that the above relationship uses the fact that the inverse of the Feynman propagator is minus the free action S_0



(2.107)

The higher order corrections to G_2 are 1PI loop contributions with two external fields. We have already encountered the above geometric series in QFT I as the two-point function W_2 (in fact, it equals the effective propagator marked by $*$). There we had to argue why it is useful to consider the inverse function of W_2 instead of W_2. Here the two-point vertex G_2 as a contribution to the effective action $G[\Psi]$ arises very naturally as the inverse of W_2.

[23] The free action defines the Feynman propagator and therefore it never appears as an interaction vertex.

The reason for calling $G[\Psi]$ the (quantum) effective action is as follows: Quantisation of a classical action $S[\Psi]$ yields the functional $Z[J] = \exp(\mathring{\imath}W[J])$. We may thus say (define) that quantum physics of the classical model described by $S[\Psi]$ is given by $Z[J]$. The corresponding classical functional is the exponential $\exp(\mathring{\imath}T[J])$ of the tree functional $T[J]$. We note that the latter is the Legendre transform of $S[\Psi]$. Therefore, we can say that classical physics is given by exponentiation of the Legendre transform of $S[\Psi]$. Noting further that the connected functional $W[J]$ is the Legendre transform of $G[\Psi]$, we can thus say that quantum physics $Z[J]$ is equivalently described by:

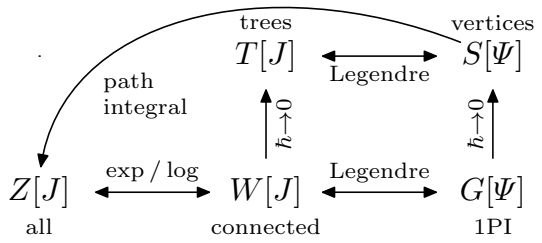
- properly quantising a classical action $S[J]$ or
- taking classical physics of a formally classical model described by the effective action $G[\Psi]$.

The effective action therefore encodes all quantum effects of the model and lets us formally treat it as a classical model based on the quantum effective action $G[\Psi]$ as its action.[24]

Importantly, the effective action encodes all loop effects, and no loops have to be closed to recover the full generating functional $Z[J]$ from $G[\Psi]$. The latter property is particularly useful for renormalisation purposes because finiteness of $G[\Psi]$ is equivalent to finiteness of $Z[J]$.

We can also mention that zeros of the function G_2 in momentum space (or equivalently the poles of the effective propagator) in fact define asymptotic particles for scattering processes. It is tempting to relate the scattering matrix to the effective action $G[\Psi]$, in particular because both objects have amputated legs. There is, however, no immediate relationship between these two objects. The LSZ reduction rather relates S-matrix elements to residues within the functional $Z[J]$. This implies that the S-matrix has disconnected graphs as well as one-particle reducible graphs.

Summary. There are at least five noteworthy generating functionals related as depicted in the following figure:



(2.108)

- The action $S[\Psi]$ is a local functional typically consisting of finitely many terms.
- The connected tree functional $T[J]$ is the Legendre transform of the action $S[\Psi]$. It is non-local (but merely rational) and has infinitely many terms.

[24]Note that the quantum effective action is intrinsically non-local and complex due to its loop diagrams, while the classical action is local and real by construction. Therefore, a hypothetical action $G[\Psi]$ would violate some of the fundamental requirements for a reasonable physical model. However, the latter does not need to stop us from performing calculations.

- The partition functional $Z[J]$ encodes all Feynman graphs. It is the path integral of the action $S[\Psi]$ with source terms.
- The connected functional $W[J]$ is the logarithm of $Z[J]$. Its classical limit is the connected tree functional $T[J]$.
- The effective action $G[\Psi]$ is the Legendre transform of $W[J]$. It encodes precisely the amputated one-particle irreducible (1PI) graphs. Its classical limit is the classical action $S[\Psi]$.

3 Yang–Mills Theory

In QFT I we have discussed quantum electrodynamics as a model of photons interacting with electrons and positrons. The interactions take place only between the particle species, each particle species is free on its own. Here we will introduce Yang–Mills theory as a more elaborate gauge theory, where the vector particles interact among themselves. For example, this is essential to describe the electrically charged vector bosons $W^{\pm}$ which have to interact with photons. Yang–Mills theory is also the foundation for quantum chromodynamics (QCD).

3.1 Classical Gauge Theory

We start with the construction of the classical model.

Electrodynamics. Let us briefly review electrodynamics coupled to spinor matter. Its fundamental fields are the electromagnetic potential A_μ, the Dirac spinor field ψ and its conjugate $\bar\psi$. The Lagrangian reads

$$\mathcal{L}_{\text{QED}} = \bar\psi(\gamma^\mu D_\mu - m)\psi - \tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{3.1}$$

The electromagnetic field strength $F_{\mu\nu}$ and the covariant derivative D_μ associated to the gauge potential A_μ acting on a field of charge q are defined as

$$F_{\mu\nu} = (\partial_\mu A_\nu) - (\partial_\nu A_\mu), \qquad D_\mu = \partial_\mu - \mathring{\imath} q A_\mu. \tag{3.2}$$

The above Lagrangian turns out to have a large amount of symmetry or redundancy. Define a gauge transformation as the following transformation with a transformation parameter field $\alpha(x)$

$$\begin{aligned} \psi'(x) &= \exp\bigl(+\mathring{\imath} q\alpha(x)\bigr)\psi(x), \\ \bar\psi'(x) &= \exp\bigl(-\mathring{\imath} q\alpha(x)\bigr)\bar\psi(x), \\ A'_\mu(x) &= A_\mu(x) + \bigl(\partial_\mu\alpha(x)\bigr). \end{aligned} \tag{3.3}$$

It is not immediately evident that the local symmetry transformation is compatible with the various derivatives appearing in the Lagrangian. However, the consistent use of the gauge covariant derivative D_μ guarantees gauge invariance by means of the transformation law[1]

$$D'_\mu = \exp\bigl(+\mathring{\imath} q\alpha(x)\bigr) D_\mu \exp\bigl(-\mathring{\imath} q\alpha(x)\bigr). \tag{3.4}$$

[1]In this chapter we assume (partial and covariant) derivatives to be operators which act on everything on their right unless restricted by brackets. E.g., in this notation a derivative can be written in two equivalent ways as $[\partial_\mu, X] = (\partial_\mu X)$. Note that this relationship remains true when supplemented by any terms on the right.

The covariant derivative simply pushes the transformation factor from the right to the left across the derivative. Moreover the field strength $F_{\mu\nu}$ can be defined as a commutator of two covariant derivatives which immediately shows that it is gauge invariant

$$F_{\mu\nu} = \mathring{\imath} q^{-1}[D_\mu, D_\nu], \qquad F'_{\mu\nu} = F_{\mu\nu}. \tag{3.5}$$

The local symmetries in quantum electrodynamics are described by the abelian Lie group U(1). This is the group of complex phase rotations $\psi \mapsto \exp(\mathring{\imath} q\alpha)\psi$.

Non-Abelian Gauge Group. Based on the above discussion is it straight-forward to enlarge the local symmetries by adding additional fields. Suppose we have N types of spinor fields ψ^j, $j = 1, \ldots, N$, with equal masses m. Consider the mass term

$$-m\bar{\psi}_j\psi^j. \tag{3.6}$$

It is not only invariant under N copies of U(1) transformations $\psi^j \to \exp(\mathring{\imath} q\alpha_j)\psi^j$, but it rather has the larger symmetry

$$\psi'^j = U^j{}_k\psi^k, \qquad \bar{\psi}'_j = (U^*)_j{}^k\bar{\psi}_k = \bar{\psi}_k(U^\dagger)^k{}_j, \tag{3.7}$$

where U is an $N \times N$ matrix which must be unitary $U^\dagger = U^{-1}$ such that (using index-free vector and matrix notation)

$$\bar{\psi}'\psi' = \bar{\psi}U^\dagger U\psi = \bar{\psi}U^{-1}U\psi = \bar{\psi}\psi. \tag{3.8}$$

The unitary matrices form the Lie group called U(N).

In the above discussion we can now replace the transformation factor $\exp(\mathring{\imath} q\alpha(x))$ by $U(x)$ everywhere. We merely have to pay attention that, unlike the phase factor $\exp(\mathring{\imath} q\alpha)$, matrices do not commute in general.

$$\psi'(x) = U(x)\,\psi(x), \qquad \bar{\psi}'(x) = \bar{\psi}(x)\,U(x)^{-1}. \tag{3.9}$$

To make the kinetic terms invariant, we must demand that the gauge covariant derivative transforms as

$$D'_\mu = UD_\mu U^{-1}. \tag{3.10}$$

Suppose D_μ takes the same expression as above in terms of the gauge potential A_μ (we drop the charge factor q for the time being)

$$D_\mu = \partial_\mu - \mathring{\imath} A_\mu. \tag{3.11}$$

The gauge potential A_μ must therefore transform according to

$$A'_\mu = UA_\mu U^{-1} + \mathring{\imath} U(\partial_\mu U^{-1}). \tag{3.12}$$

This shows that the gauge potential itself must be an $N \times N$ matrix. However, we can restrict it further by imposing a reality condition. The reality condition must be compatible with the gauge transformation rule. We make use of the unitarity

property of U and write the transformation of the hermitian conjugate of the potential A^μ

$$\begin{aligned}(A'_\mu)^\dagger &= UA_\mu^\dagger U^{-1} - \mathring{\imath}(\partial_\mu U)U^{-1} \\ &= UA_\mu^\dagger U^{-1} + \mathring{\imath}U(\partial_\mu U^{-1}).\end{aligned} \tag{3.13}$$

The latter step follows from the identity

$$0 = (\partial_\mu 1) = \big(\partial_\mu(UU^{-1})\big) = (\partial_\mu U)U^{-1} + U(\partial_\mu U^{-1}). \tag{3.14}$$

Altogether it shows that $A_\mu^\dagger$ transforms exactly as A_μ itself, and therefore the gauge transformation is consistent with the assumption of a hermitian gauge potential

$$A_\mu^\dagger = A_\mu. \tag{3.15}$$

Consequently, it makes sense to impose this property on the gauge field in order to reduce its independent components to a minimum.

Finally we can define the field strength as before

$$F_{\mu\nu} = \mathring{\imath}[D_\mu, D_\nu] = (\partial_\mu A_\nu) - (\partial_\nu A_\mu) - \mathring{\imath}[A_\mu, A_\nu]. \tag{3.16}$$

This field is hermitian $(F_{\mu\nu})^\dagger = F_{\mu\nu}$ and it transforms covariantly under gauge transformations

$$F'_{\mu\nu} = UF_{\mu\nu}U^{-1}. \tag{3.17}$$

Gauge Transformations. Finite gauge transformations can be written using the exponential map of a matrix of transformation parameters α

$$U = \exp(\mathring{\imath}\alpha). \tag{3.18}$$

Unitarity of U translates to hermiticity of α

$$\alpha^\dagger = \alpha. \tag{3.19}$$

The transformation rules can be summarised as follows

$$\begin{aligned}\psi' &= U\psi, \\ \bar\psi' &= \bar\psi U^{-1}, \\ A'_\mu &= UA_\mu U^{-1} + \mathring{\imath}U(\partial_\mu U^{-1}), \\ D'_\mu &= UD_\mu U^{-1}, \\ F'_{\mu\nu} &= UF_{\mu\nu}U^{-1}.\end{aligned} \tag{3.20}$$

When handling long expressions, the full gauge transformation rules are sometimes hard to handle. In physics one often considers only the infinitesimal transformations. When they apply everywhere, they are as good as finite transformation for practical purposes.

Consider now the infinitesimal transformation where only the linear orders in $\delta\alpha$ are relevant

$$U = 1 + \mathring{\imath}\,\delta\alpha + \dots. \tag{3.21}$$

The infinitesimal transformation parameter $\delta\alpha$ must be hermitian, $\delta\alpha^\dagger = \delta\alpha$. The infinitesimal transformation rules then read

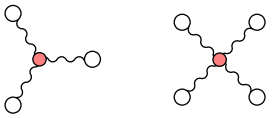
$$\begin{aligned}
\delta\psi' &= \mathring{\imath}\,\delta\alpha\,\psi, \\
\delta\bar{\psi}' &= -\mathring{\imath}\bar{\psi}\,\delta\alpha, \\
\delta A_\mu &= (\partial_\mu \delta\alpha) + \mathring{\imath}[\delta\alpha, A_\mu] = [D_\mu, \delta\alpha], \\
\delta D_\mu &= \mathring{\imath}[\delta\alpha, D_\mu], \\
\delta F_{\mu\nu} &= \mathring{\imath}[\delta\alpha, F_{\mu\nu}].
\end{aligned} \tag{3.22}$$

Importantly, a finite transformation by matrix conjugation with U translates to commutators with $\delta\alpha$.

Non-Abelian Gauge Theory. We are now in the position to write a Lagrangian which is manifestly invariant under gauge transformations. First we need to find a suitable kinetic term for the gauge field. The usual term $F^{\mu\nu}F_{\mu\nu}$ taken as a matrix product is merely gauge covariant, $F'^2 = UF^2U^{-1}$; putting it inside a trace makes it properly invariant

$$\mathcal{L}_{\text{YM}} = -\frac{1}{2g^2}\,\text{tr}\,F^{\mu\nu}F_{\mu\nu}. \tag{3.23}$$

Interestingly, this model is interacting due to the extra non-linear terms in A_μ within $F_{\mu\nu}$. This Lagrangian extends electromagnetism and the model is called Yang–Mills theory. The prefactor $g = g_{\text{YM}}$ in fact serves as a coupling constant as we shall see later. The graphical representation for the interactions in pure Yang–Mills theory is a cubic and a quartic vertex between gauge fields.[2]

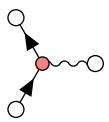


(3.24)

The spinor matter sector for a non-abelian gauge theory (Yang–Mills theory coupled to matter) takes precisely the same form as before

$$\mathcal{L}_{\text{spinor}} = \bar{\psi}(\gamma^\mu D_\mu - m)\psi. \tag{3.25}$$

Note that here ψ and $\bar{\psi}$ are vectors and D_μ is a matrix in an N-dimensional space. The resulting vertex is analogous to the vertex in QED.



(3.26)

The model with gauge group SU(3) and one copy of the 3-vector of spinor fields ψ for each quark flavour (up, down, strange, ...) is called quantum chromodynamics (QCD).

[2]The non-abelian gauge fields are often depicted by curly lines. Here we recycle the wiggly lines otherwise used for abelian photons.

Equations of Motion. Let us briefly summarise the equation of motion for non-abelian gauge theory. They directly reflect the equations of motion for electrodynamics, but due to the non-abelian nature of the gauge group, the partial derivatives need to be replaced by covariant derivatives.

The homogeneous Maxwell equation, also known as a Bianchi identity, derives from the Jacobi identity of three covariant derivatives. For Yang–Mills theory it can be written as

$$(D^{\mathrm{ad}}_\rho F_{\mu\nu}) + (D^{\mathrm{ad}}_\nu F_{\rho\mu}) + (D^{\mathrm{ad}}_\mu F_{\nu\rho}) = 0 \qquad \text{or} \qquad \varepsilon^{\mu\nu\rho\sigma}(D^{\mathrm{ad}}_\rho F_{\mu\nu}) = 0. \tag{3.27}$$

Here the covariant derivative D^{ad}_μ of a gauge covariant field X such as the field strength $F_{\mu\nu}$ can be written as a commutator[3]

$$(D^{\mathrm{ad}}_\mu X) := [D_\mu, X] = (\partial_\mu X) - \mathring{\imath}[A_\mu, X]. \tag{3.28}$$

The extension of the inhomogeneous Maxwell equation reads

$$\frac{2}{g^2}\,(D^{\mathrm{ad}}_\mu F^{\mu\nu}) = J^\nu. \tag{3.29}$$

The current J^ν is a hermitian $N \times N$ matrix of fermion bilinears

$$(J^\nu)^j{}_k = \mathring{\imath}\bar{\psi}_k \gamma^\nu \psi^j. \tag{3.30}$$

Finally, the matter equations of motion take the usual Dirac form with a covariant derivative to replace the ordinary one

$$(D^\mu \gamma_\mu - m)^j{}_k\, \psi^k = 0. \tag{3.31}$$

Note that the equations of motion imply that the current is covariantly conserved[4]

$$(D^{\mathrm{ad}}_\mu J^\mu) = 0. \tag{3.32}$$

3.2 Abelian Quantisation Revisited

Before we can quantise the Yang–Mills theory, we should first discuss the path integral formulation of electrodynamics. In QFT I, we have noticed that gauge invariance makes the naive vector field propagator ill-defined. The solution was to break gauge invariance by adding a gauge-fixing term. We will now consider how to achieve this in the path integral formulation.

[3] In physics, the definition of the covariant derivative $D_\mu X$ is often assumed to depend implicitly on the transformation rules of the field X on which it acts. We might thus abbreviate $D^{\mathrm{ad}}_\mu X$ by $D_\mu X$ for a field X which transforms by conjugation with U.

[4] The additional terms in the covariant conservation rule involving the gauge potential A_μ render the applicability of Noether's theorem more subtle.

Naive Path Integral. Consider the action for the electromagnetic field with sources[5]

$$S_{\mathrm{EM}}[A,J] = \int \mathrm{d}x^4 \left(-\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} - J^\mu A_\mu\right). \tag{3.33}$$

The action is invariant under gauge transformations

$$S_{\mathrm{EM}}[A',J] = S_{\mathrm{EM}}[A,J] = S_{\mathrm{EM}}\big[[A],J\big] \tag{3.34}$$

with

$$A'_\mu = A_\mu + (\partial_\mu \alpha), \tag{3.35}$$

provided that the source field is conserved,

$$\partial_\mu J^\mu = 0. \tag{3.36}$$

Naively, the path integral reads

$$Z[J] = \int \mathrm{D}A \exp\big(\mathring{\imath} S_{\mathrm{EM}}[A,J]\big). \tag{3.37}$$

Gauge invariance of the action implies that the integrand of the path integral is constant along all gauge-equivalent field configurations A. We may thus formally split up the path integral into integrals over the gauge equivalence classes and over the gauge orbits:

$$Z[J] = \left(\int \mathrm{D}\alpha\right) \cdot \int \mathrm{D}[A] \exp\big(\mathring{\imath} S_{\mathrm{EM}}\big[[A],J\big]\big). \tag{3.38}$$

This shows that the path integral over the gauge potential A over-counts the admissible field configurations.[6] In fact, the relevant issue is that the kinetic term for A is non-invertible. This is because the gauge transformations act trivially on it thus implying zero modes.

We would like to divide out the divergent integral $\int \mathrm{D}\alpha$ over gauge orbits, and keep just the integral $\int \mathrm{D}[A]$ over gauge equivalence classes. However, it is not evident how to properly define the corresponding measure $\mathrm{D}[A]$ ensuring that it is gauge invariant and that it obeys the properties of common path integral measures for matter fields.[7] The gauge-fixing procedure that we will construct in the following essentially provides a proper notion for $\mathrm{D}[A]$.

[5] The current J can be an external field or it may represent couplings to matter fields such as $J^\mu = -\mathring{\imath}\bar{\psi}\gamma^\mu\psi$. The latter is conserved provided that the field ψ obeys the equation of motion. Strictly speaking, coupling to charged scalar fields is not of this form due to the terms quadratic in gauge fields. Nevertheless, scalar fields can be added to the action without complications as long as the latter remains gauge invariant.

[6] On the one hand, the gauge orbits are infinite-dimensional spaces and therefore the path integral will formally be infinite. On the other hand, the path integral diverges anyway, so this by itself is not a criterion.

[7] As we have not defined path integrals and their measure factors in detail, we might abstractly define the measure $\mathrm{D}[A]$ to do its intended job. However, the resulting measure will likely not obey valuable identities for partial integration or for variable transformations that are used in deriving Feynman diagrams. We assume that the full measure $\mathrm{D}A$ for the gauge potential obeys the relevant identities as for matter fields. Any adjustments are to be carried by explicit action terms $\exp(\mathring{\imath} S_{\mathrm{measure}}[A])$.

Gauge Fixing. Ideally, we should count each physical configuration only once. To this end we introduce a gauge-fixing condition $G[A](x) = 0$ where G is some functional that is assumed to vanish for all x for exactly one configuration A in each gauge orbit. For example, we could use the functional map $G[A](x) = \partial^\mu A_\mu(x)$ which enforces the Lorenz gauge.[8] In fact, we will use a slightly more general gauge condition[9]

$$G[A, \Omega](x) = \partial^\mu A_\mu(x) - \Omega(x), \tag{3.39}$$

where $\Omega(x)$ is some function that defines the gauge. This generalisation is useful since we have already seen in QFT I that $\partial^\mu A_\mu$ does not identically vanish in the quantum theory.

We would like to restrict the path integral by inserting a delta-functional Δ

$$\int \mathrm{D}A\, \Delta\big[G[A, \Omega]\big] \exp\big(\mathring{\imath} S_{\mathrm{EM}}[A, J]\big). \tag{3.40}$$

However, this insertion has the undesirable feature that it is not evidently related to the original formulation of $Z[J]$. Furthermore, the result clearly depends on the functional variation of G at $G = 0$ due to[10]

$$\Delta\big[G[A, \Omega]\big] = \frac{\Delta\big[A - A_0[\Omega]\big]}{\mathrm{Det}\big[\delta G[A, \Omega]/\delta A\big]}\,. \tag{3.41}$$

In other words, the result depends on our choice of gauge-fixing function.

Instead, we use the assumption that the functional $G[A', \Omega]$ vanishes precisely at one point of the gauge orbit, where A' is the gauge potential A transformed by the field α

$$A'_\mu = A_\mu + (\partial_\mu \alpha). \tag{3.42}$$

This implies the delta-functional identity[11]

$$1 = \int \mathrm{D}\alpha\, \Delta\big[G[A', \Omega]\big]\, \mathrm{Det}_{x,y}\left[\frac{\delta G[A', \Omega](x)}{\delta \alpha(y)}\right], \tag{3.43}$$

where the 1 on the left-hand side represents the number of solutions of $G[A', \Omega] = 0$ over a gauge orbit.[12]

[8] This condition only fixes the gauge up to some residual gauge transformations, namely those with $\partial^\mu \delta A_\mu = \partial^2 \alpha = 0$. This particular deviation from our assumptions turns out to be permissible.

[9] Other gauges are acceptable as well, but details in the following discussion may have to be adapted.

[10] This is the analogue of the reparametrisation rule for delta-functions $\delta(f(x)) = \delta(x - x_0)/|f'(x_0)|$ where $f(x_0) = 0$.

[11] The functional determinant Det is approximated by an ordinary determinant when the coordinates x and y are discretised.

[12] For our choice of the gauge-fixing functional G, the left-hand side is not actually 1 (but rather infinite). Importantly, the result appears to not depend significantly on A or Ω so that its subsequent insertion into the path integral has a minor effect.

A convenient feature of the above gauge-fixing functional map G is that its functional derivative in the determinant is independent of A, Ω and α. It simply reads

$$\frac{\delta G[A',\Omega](x)}{\delta\alpha(y)} = \partial^2\delta^4(x-y). \tag{3.44}$$

Here, the delta-function acts as a functional equivalent of a unit matrix, hence we can write the identity as

$$1 = \mathrm{Det}[\partial^2]\int \mathrm{D}\alpha\, \Delta\big[G[A',\Omega]\big]. \tag{3.45}$$

The trick used by Faddeev and Popov is to insert the identity into the original path integral for the electromagnetic field

$$Z[J] = \mathrm{Det}[\partial^2]\int \mathrm{D}\alpha\,\mathrm{D}A\,\Delta\big[G[A',\Omega]\big]\exp\big(\mathring{\imath}S_{\mathrm{EM}}[A,J]\big). \tag{3.46}$$

We now perform a reparametrisation of the integration variable A to transform A' to A. This reparametrisation is both a shift and a gauge transformation, hence $\mathrm{D}A$ and $S_{\mathrm{EM}}[A,J]$ remain unchanged

$$Z[J] = \mathrm{Det}[\partial^2]\bigg(\int \mathrm{D}\alpha\bigg)\int \mathrm{D}A\,\Delta\big[G[A,\Omega]\big]\exp\big(\mathring{\imath}S_{\mathrm{EM}}[A,J]\big). \tag{3.47}$$

We have now separated the path integral over the gauge-fixed potential A from a divergent prefactor which does not depend on A.

Averaging. Now we must somehow implement the gauge-fixing delta-functional in the path integral. One option would be to solve for one component of A_μ, e.g. A_0. Doing so would violate Lorentz invariance. Preserving Lorentz symmetry is one of the aims of the path integral formalism.

An alternative is to employ another trick, and integrate out Ω with a Gaußian potential

$$S_\Omega[\Omega] = -\frac{1}{2\xi}\int \mathrm{d}x^4\,\Omega^2, \qquad \int \mathrm{D}\Omega\exp\big(\mathring{\imath}S_\Omega[\Omega]\big) = \mathrm{const}\,. \tag{3.48}$$

Since $Z[J]$ only formally depends on Ω this merely introduces another constant factor into the original path integral

$$Z[J] \sim \int \mathrm{D}\Omega\,\mathrm{D}A\,\Delta\big[G[A,\Omega]\big]\exp\big(\mathring{\imath}S_{\mathrm{EM}}[A,J] + \mathring{\imath}S_\Omega[\Omega]\big). \tag{3.49}$$

We now use the delta-functional (whose argument is linear in Ω) to perform the path integral over Ω by substituting $\Omega = \partial\cdot A$ and obtain

$$Z[J] \sim \int \mathrm{D}A\exp\big(\mathring{\imath}S_{\mathrm{EM}}[A,J] + \mathring{\imath}S_{\mathrm{gf}}[A]\big) \tag{3.50}$$

with an additional contribution to the action

$$S_{\mathrm{gf}}[A] = -\frac{1}{2\xi}\int \mathrm{d}x^4\,(\partial\cdot A)^2. \tag{3.51}$$

Discussion. Finally, we have recovered the Lagrangian of the electromagnetic field with the standard Lorenz gauge-fixing term. Up to some infinite but constant prefactors the new path integral is equal to the original path integral

$$\begin{aligned} Z[J] &= \int \mathrm{D}A \exp\bigl(\mathring{\imath} S_{\mathrm{EM}}[A,J]\bigr) \\ &= \mathrm{const} \cdot \int \mathrm{D}A \exp\bigl(\mathring{\imath} S_{\mathrm{EM}}[A,J] + \mathring{\imath} S_{\mathrm{gf}}[A]\bigr). \end{aligned} \tag{3.52}$$

Since all constant prefactors drop out of correlation functions, we can use the action with gauge-fixing term as the starting point for quantising electrodynamics.

Note, however, that the path integrals are only equivalent when the source field J is conserved. Therefore, only correlation functions of gauge invariant combinations of the fields are meaningful in the gauge-fixed form.[13]

We have used a modified Lorenz gauge condition $\partial{\cdot}A = \Omega$ where Ω is an auxiliary dynamical field. This field is governed by a free Lagrangian without derivative terms which essentially force it (quantum mechanically close) to zero. This resulting gauge fixing of the electrodynamics Lagrangian is also known as R_ξ-gauge. The parameter ξ can be fixed at will, and no physical, gauge-invariant quantities may depend on it. A convenient choice is $\xi = 1$ for Feynman gauge or the limit $\xi \to 0$ (after quantisation) for Landau gauge.

The resulting Feynman propagator for the vector field in momentum space is

$$(G_{\mathrm{F}})_{\mu\nu}(p) = \frac{\eta_{\mu\nu} - (1-\xi) p_\mu p_\nu / p^2}{p^2 - \mathring{\imath}\epsilon}\,. \tag{3.53}$$

Note that this propagator does not respect the strict Lorenz gauge as can be seen by contracting the propagator with p^μ

$$p^\mu (G_{\mathrm{F}})_{\mu\nu}(p) = \frac{\xi p_\nu}{p^2 - \mathring{\imath}\epsilon}\,. \tag{3.54}$$

This term would have to be zero in a strict Lorenz gauge, nevertheless it is acceptable because the result is proportional to p_ν. The above relationship can be understood better when formulated in position space

$$\mathring{\imath}\bigl\langle \partial{\cdot}A(x)\, A_\nu(y)\bigr\rangle = \xi \partial_{y,\nu} G_{\mathrm{F}}(x-y) = \partial_\nu \alpha_x(y). \tag{3.55}$$

The correlator of $\partial{\cdot}A$ with another gauge field yields a result which is proportional to a gauge transformation of the other gauge field by $\alpha_x(y) = \xi G_{\mathrm{F}}(x-y)$ which is effectively trivial. In practice, the non-vanishing term does not matter since the other side of the propagator should couple to a conserved current such as J^μ.

[13] Of course we may also compute more general correlation functions, but these are potentially unrelated to the original path integral and thus meaningless.

3.3 Yang–Mills Quantisation

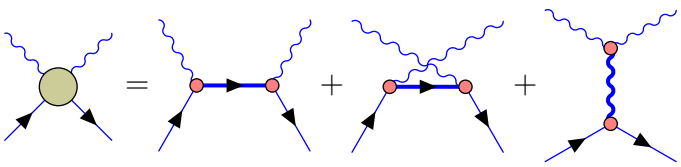
We are now prepared to apply to the path integral quantisation to the vector fields of Yang–Mills theory. Here some further complications arise due to the non-abelian nature of the gauge group. These lead to the introduction of further fields.

Some Puzzles. If we were over-confident we could now postulate the Feynman rules for Yang–Mills theory. When reading them off from a straight-forward non-abelian generalisation of the above gauge-fixed action,

$$Z[J] = \text{const} \cdot \int \mathrm{D}A \exp\bigl(\mathring{i}S_{\mathrm{YM}}[A,J] + \mathring{i}S_{\mathrm{gf}}[A]\bigr) \tag{3.56}$$

we run into some problems. Let us briefly sketch without derivations some undesirable effects related to improper gauge fixing:

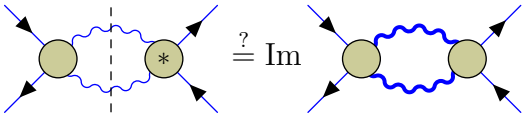
We can consider the process of two matter particles annihilating into two vector particles



(3.57)

In the abelian process the resulting photons automatically have only physical polarisation vectors. This turns out not to hold in the non-abelian generalisation, some of the produced vector particles have unphysical polarisations. They arise from from the third diagram which is not present for abelian gauge theory. Such unphysical behaviour in a process is a reliable sign that something is wrong.

To some extent we may choose to ignore such contributions by projecting the external states to physical polarisations. This makes sense because the above construction applies to gauge-invariant quantities and configurations only. Nevertheless, this interpretation causes problems concerning unitarity: The optical theorem relates the square of the above tree process to the imaginary part of a loop process

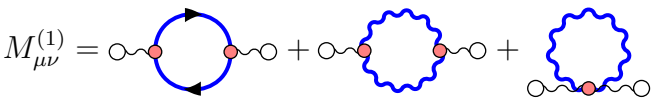


(3.58)

- All the fields of the gauge-fixed formulation of the model propagate in loops. This includes the unphysical polarisation of the vector field.
- The truncated tree process discards the unphysical polarisations.

Consequently, unitarity fails or the loop contributions predicted by the path integral do not accurately describe quantum effects in Yang–Mills theory.

A further clue is provided by the loop correction to the effective vertex of two gluons



(3.59)

In the abelian theory, the result turns out to be properly transverse and thus gauge-invariant

$$M^{(1)}_{\mu\nu}(p) = M^{(1)}(p)(p^2\eta_{\mu\nu} - p_\mu p_\nu), \qquad p^\nu M^{(1)}_{\mu\nu}(p) = 0. \tag{3.60}$$

The non-abelian result, however, is not transverse

$$M^{(1)}_{\mu\nu}(p) = M^{(1)}_1(p)\, p^2\, \eta_{\mu\nu} + M^{(1)}_2(p)\, p_\mu p_\nu, \qquad p^\nu M^{(1)}_{\mu\nu}(p) \neq 0. \tag{3.61}$$

The violation of transversality can actually be traced to the second diagram involving only gluons.

It turns out that indeed loop effects are not correctly captured by a naive postulation of Feynman rules. We have to reinvestigate the quantisation and properly generalise it to non-abelian gauge theory.

Non-Abelian Gauge Fixing. Let us now go through the above quantisation procedure but with non-abelian fields. The action with source terms reads

$$S_{\mathrm{YM}}[A, J] = -\frac{1}{2g^2}\int \mathrm{d}x^4\, \mathrm{tr}\, F^{\mu\nu}F_{\mu\nu} - \int \mathrm{d}x^4\, \mathrm{tr}\, J^\mu A_\mu. \tag{3.62}$$

We can choose the same gauge-fixing function as before

$$G[A, \Omega](x) = \partial^\mu A_\mu(x) - \Omega(x), \tag{3.63}$$

and fix the gauge by means of the identity

$$1 = \int \mathrm{D}U\, \Delta\big[G[A', \Omega]\big]\, \mathrm{Det}_{(xa,yb)}\left[\frac{\delta G[A', \Omega]^a(x)}{-\mathring{\imath}(\delta U\, U^{-1})^b(y)}\right]. \tag{3.64}$$

Here, $U(x)$ is a group element, and the integration measure DU is assumed to be invariant under group multiplication. In other words, for a fixed V, $\mathrm{D}(UV) = \mathrm{D}U = \mathrm{D}(VU)$. Moreover, the variation δU of a group element U is expressed through the hermitian combination

$$-\mathring{\imath}\, \delta U\, U^{-1}. \tag{3.65}$$

To compute the operator in the determinant, consider the transformation of the gauge potential

$$A'_\mu = UA_\mu U^{-1} + \mathring{\imath}U(\partial_\mu U^{-1}). \tag{3.66}$$

Its variation can be expressed neatly as a covariant derivative of the combination $\delta U\, U^{-1}$ [14]

$$\begin{aligned} \delta A'_\mu &= -\mathring{\imath}\left(\partial_\mu(\delta U\, U^{-1})\right) - [A'_\mu, \delta U\, U^{-1}] \\ &= -\mathring{\imath}[D'_\mu, \delta U\, U^{-1}] = -\mathring{\imath}\big(D'^{\mathrm{ad}}_\mu(\delta U\, U^{-1})\big). \end{aligned} \tag{3.67}$$

[14] The desired combination can always be achieved by writing $\delta U = (\delta U\, U^{-1})U$.

Hence we can write the identity as

$$1 = \int \mathrm{D}U\, \Delta\big[G[A', \Omega]\big]\, \mathrm{Det}\big[\partial^\mu D'^{\mathrm{ad}}_\mu\big]. \tag{3.68}$$

Note that the determinant now depends on A' and therefore it cannot be pulled out of the integral.

We insert the identity into the path integral

$$\begin{aligned} Z[J] = \int \mathrm{D}U\, \mathrm{D}A\, \Delta\big[G[A', \Omega]\big]\, \mathrm{Det}\big[\partial^\mu D'^{\mathrm{ad}}_\mu\big] \\ \cdot \exp\big(\mathring{i} S_{\mathrm{YM}}[A, J]\big). \end{aligned} \tag{3.69}$$

Next we perform a gauge transformation by U^{-1} to bring all gauge potentials to the original form.[15]

$$\begin{aligned} Z[J] = \left(\int \mathrm{D}U\right) \int \mathrm{D}A\, \Delta\big[G[A, \Omega]\big]\, \mathrm{Det}\big[\partial^\mu D^{\mathrm{ad}}_\mu\big] \\ \cdot \exp\big(\mathring{i} S_{\mathrm{YM}}[A, J]\big). \end{aligned} \tag{3.70}$$

Now the integral over the gauge orbit is isolated and can be dropped.

Faddeev–Popov Ghosts. The crucial difference to the abelian case is that the determinant still depends on the gauge potential and cannot be pulled in front of the integral. The final trick used by Faddeev and Popov is to reconstruct the determinant as a fermionic path integral. Consider a pair of hermitian fermionic fields $C, \tilde{C}$ [16] belonging to the same space as each component A_μ of the gauge potential. The field C is called the ghost field, $\tilde{C}$ is called the anti-ghost. Define its action as a quadratic functional of the new fields[17]

$$\begin{aligned} S_{\mathrm{gh}}[C, \tilde{C}, A] &= \frac{2\mathring{i}}{g^2} \int \mathrm{d}x^4\, \mathrm{tr}\, \tilde{C}(\partial^\mu D^{\mathrm{ad}}_\mu C) \\ &= -\frac{2\mathring{i}}{g^2} \int \mathrm{d}x^4\, \mathrm{tr}(\partial^\mu \tilde{C})[D_\mu, C]. \end{aligned} \tag{3.71}$$

A quadratic action leads to a Gaußian path integral which can be evaluated in closed form as a determinant

$$\int \mathrm{D}C\, \mathrm{D}\tilde{C} \exp\big(\mathring{i} S_{\mathrm{gh}}[C, \tilde{C}, A]\big) = \mathrm{Det}\big[\partial^\mu D^{\mathrm{ad}}_\mu\big]. \tag{3.72}$$

[15] Strictly speaking the source term is not invariant under gauge transformations. However, it is effectively invariant as long as we extract only gauge-invariant correlation functions from the generating functional $Z[J]$. When the source J is viewed as representing gauge-invariant couplings to matter fields, the gauge transformation of the matter fields induces the desired transformation for J.

[16] Both fields are hermitian in the sense that $C^\dagger = C$ and $\tilde{C}^\dagger = \tilde{C}$ where hermitian conjugation also flips the ordering of (fermionic) fields without introducing signs for their permutation.

[17] The normalisation of a quadratic term is inessential because it only affects a constant prefactor.

The argument of the determinant is precisely the operator acting on C. We use this identity to convert the determinant to some extra path integrals

$$\begin{aligned} Z[J] = \int \mathrm{D}C\,\mathrm{D}\tilde{C}\,\mathrm{D}A\,\Delta\big[G[A,\Omega]\big] \\ \cdot \exp\big(\mathring{\imath}S_{\mathrm{YM}}[A,J] + \mathring{\imath}S_{\mathrm{gh}}[C,\tilde{C},A]\big). \end{aligned} \tag{3.73}$$

Finally, we can treat the delta-functional as before in the abelian case by averaging over the field Ω. We then obtain the gauge-fixed Yang–Mills path integral

$$\begin{aligned} Z[J] = \int \mathrm{D}C\,\mathrm{D}\tilde{C}\,\mathrm{D}A \\ \cdot \exp\big(\mathring{\imath}S_{\mathrm{YM}}[A,J] + \mathring{\imath}S_{\mathrm{gh}}[C,\tilde{C},A] + \mathring{\imath}S_{\mathrm{gf}}[A]\big). \end{aligned} \tag{3.74}$$

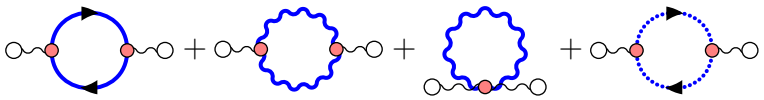
with gauge-fixing term

$$S_{\mathrm{gf}}[A] = -\frac{1}{\xi g^2}\int \mathrm{d}x^4\,\mathrm{tr}(\partial\cdot A)^2. \tag{3.75}$$

Discussion. We observe that the proper quantisation of the Yang–Mills action requires the introduction of additional scalar fermionic fields, so-called Faddeev–Popov ghosts. These fields are called ghosts because they obey the wrong statistics for scalar fields.

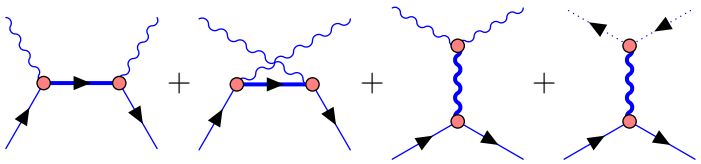
Usually such fields would violate basic properties of QFT. In this situation it actually solves a problem since the gauge potential itself has some unphysical modes. The unphysical behaviour of these two fields conspires to cancel each other out precisely.

For example, the one-loop contribution to the two-point effective vertex receives another contribution from a ghost loop (dotted line)

$M^{(1)}_{\mu\nu} =$  . (3.76)

This ghost has precisely the same unphysical behaviour as the pure gauge loop. Since it is a fermion loop, it receives an extra minus sign, so that the two unphysical contributions cancel exactly. The final result for the one-loop two-point effective vertex therefore has a proper polarisation structure.

Note, however, that unphysical polarisations do remain for asymptotic vector fields as do the ghost fields. In the above example of two matter fields annihilating we thus have an extra contribution producing a pair of ghosts

 . (3.77)

As we shall see later, the unphysical particles can now be projected out safely without violating unitarity.

Note that the Faddeev–Popov ghosts are not required for all gauges. For example, the light cone gauge is typically formulated in terms of the gauge field only. The point is that the vector field in the light cone gauge has no unphysical modes, hence no ghosts are needed to cancel their effects.

Another example where no ghosts are needed is the abelian theory. Here the gauge field never forms closed loops on its own. Consequently, the ghost field does not couple to the gauge field.

3.4 BRST Symmetry

In quantising the Yang–Mills fields, we have traded in full gauge invariance for a well-defined gauge field propagator. We can observe that the one-loop correction to the gauge field propagator leaves the polarisation structure of the gauge field propagator unchanged. This comes at the price of introducing ghost fields with unphysical properties such as violating the spin-statistics theorem. It remains to be understood how to properly restrict to processes of the physical particles and why the structure of gauge fixing is maintained by renormalisation to higher perturbative orders. There is still the coordinate-independent part of gauge symmetry left over in the Lorenz gauge. The global gauge symmetry ensures that only globally covariant structures can appear in renormalisation, but this is clearly insufficient for either of the above causes.

Becchi, Rouet, Stora and, independently, Tyutin found an extra symmetry of the gauge-fixed Lagrangian with unphysical and ghost degrees of freedom, the so-called BRST transformation. This transformation is an essential tool for understanding physical states and how to filter out their interaction processes from arbitrary correlators that one may formulate in the gauge-fixed model.

Auxiliary Field. To see the symmetry more clearly, we first rewrite the Lagrangian. Consider the following Lagrangian of the gauge potential A_μ, the scalar fermionic ghost fields $C, \tilde{C}$ and an auxiliary scalar bosonic field B

$$\mathcal{L}_{\mathrm{BRST}} = \frac{2}{g^2}\,\mathrm{tr}\Big(\mathring{\imath}\tilde{C}(\partial^\mu D^{\mathrm{ad}}_\mu C) + B(\partial^\mu A_\mu) + \tfrac{1}{2}\xi B^2\Big). \tag{3.78}$$

The field B is free and non-propagating, and it can be integrated out exactly by replacing it by its saddle point solution $B = -\xi^{-1}\partial^\mu A_\mu$. This yields precisely the conventional gauge-fixing and ghost terms[18]

$$\mathcal{L}_{\mathrm{BRST}} \simeq \mathcal{L}_{\mathrm{ghost}} + \mathcal{L}_{\mathrm{gf}}. \tag{3.79}$$

[18] One may choose to work without the auxiliary field B, but then some properties (closure of the BRST algebra, existence of K_{BRST}) are obscured.

BRST Transformation. The BRST transformation is defined as the following variation of the fields[19]

$$\begin{aligned} \delta A_\mu &= \delta\epsilon(D^{\rm ad}_\mu C), & \delta\tilde{C} &= \mathring{\imath}\,\delta\epsilon\, B, \\ \delta C &= \mathring{\imath}\,\delta\epsilon\, CC = \tfrac{\mathring{\imath}}{2}\delta\epsilon\{C,C\}, & \delta B &= 0. \end{aligned} \tag{3.80}$$

Note that the transformation parameter $\delta\epsilon$ must be an anti-commuting Graßmann odd number in order to match the statistics on both sides of the equation. It is straight-forward to confirm that the Lagrangian is invariant under the BRST transformation

$$\delta\mathcal{L}_{\rm BRST} = 0. \tag{3.81}$$

Likewise, the Yang–Mills Lagrangian $\mathcal{L}_{\rm YM}$ is invariant, $\mathcal{L}_{\rm YM} = 0$, since the BRST transformation on the ordinary fields is a plain gauge transformation with transformation parameter $\delta\epsilon\, C$, i.e. a field-dependent gauge transformation.

Let us write the BRST transformation as a fermionic operator Q such that

$$\delta = \mathring{\imath}\,\delta\epsilon\, Q. \tag{3.82}$$

It acts on the quantum fields by the following commutation relations

$$\begin{aligned} [Q, A_\mu] &= -\mathring{\imath}(D^{\rm ad}_\mu C), & \{Q,\tilde{C}\} &= B, \\ \{Q, C\} &= CC, & [Q, B] &= 0. \end{aligned} \tag{3.83}$$

This BRST operator has two important properties: it is hermitian and it squares to zero

$$Q^\dagger = Q, \qquad Q^2 = 0. \tag{3.84}$$

These properties can be verified easily by hermitian conjugation of its action and by acting with Q twice. In fact, we can write the BRST Lagrangian as a BRST variation of some other combination $K_{\rm BRST}$

$$\mathcal{L}_{\rm BRST} = QK_{\rm BRST}, \qquad K_{\rm BRST} = \frac{2}{g^2}\,\mathrm{tr}\big(\tilde{C}(\partial^\mu A_\mu) + \tfrac{1}{2}\xi\tilde{C}B\big). \tag{3.85}$$

It is then evident that $Q\mathcal{L}_{\rm BRST} = Q^2K_{\rm BRST} = 0$. The Yang–Mills Lagrangian, on the other hand, cannot be written as a BRST variation of something, since the gauge field strength $F_{\mu\nu}$ is never produced by the action of Q

$$\mathcal{L}_{\rm YM} \neq QK_{\rm YM}. \tag{3.86}$$

Another new symmetry is the ghost number $N_{\rm gh}$. It counts the number of ghosts C minus the number of anti-ghosts $\tilde{C}$. The gauge and auxiliary fields carry no ghost number.

[19] In view of the above, supposing that $B = -\xi^{-1}\partial^\mu A_\mu$, the relation $\delta B = 0$ holds only provided that C satisfies its equation of motion. This is not surprising given that $B = -\xi^{-1}\partial^\mu A_\mu$ is the equation of motion for the auxiliary field, i.e. BRST symmetry relates these two equations of motion.

BRST Cohomology. The BRST operator Q squares to zero, $Q^2 = 0$, in close analogy to the exterior derivative $\mathrm{d} := \mathrm{d}x^\mu \, \partial_\mu$. Such operators define a cohomology which often encodes some important information in quantum physics. For example, the de Rham cohomology of the exterior derivative probes the global topology of differentiable manifolds. The BRST cohomology turns out to describe physical states and physical operators.

Suppose we go back to canonical quantisation with a Hilbert space $\mathbb{V}$. The BRST symmetry leads to an associated operator Q acting on $\mathbb{V}$. There are two types of distinguished states:

- *Closed* states are annihilated by Q

$$Q|\Psi\rangle = 0 \quad \leftrightarrow \quad |\Psi\rangle \in \mathbb{V}_{\mathrm{cl}}. \tag{3.87}$$

- *Exact* states are in the image of Q

$$|\Psi\rangle = Q|\Omega\rangle \text{ for some } |\Omega\rangle \in \mathbb{V} \quad \leftrightarrow \quad |\Psi\rangle \in \mathbb{V}_{\mathrm{ex}}. \tag{3.88}$$

By construction exact states are also closed $\mathbb{V}_{\mathrm{ex}} \subset \mathbb{V}_{\mathrm{cl}}$, but not necessarily vice versa. The BRST cohomology $\mathbb{H}_{\mathrm{phys}}$ is defined as the quotient space

$$\mathbb{H}_{\mathrm{phys}} := \mathbb{V}_{\mathrm{cl}}/\mathbb{V}_{\mathrm{ex}}. \tag{3.89}$$

This space (restricted to ghost number zero) turns out to define the physical states. In other words, physical states are defined as equivalence classes $|\Psi\rangle + \mathbb{V}_{\mathrm{ex}}$ of states in the original Hilbert space

$$|\Psi\rangle \simeq |\Psi\rangle + Q|\Omega\rangle \qquad \text{with} \qquad Q|\Psi\rangle = 0. \tag{3.90}$$

We have postulated that the BRST cohomology (of ghost number zero) defines physical states. This makes sense only if physical answers depend on equivalence classes but not on their representatives. The discussion is similar to the corresponding discussion of equivalence classes of the vector field in QFT I.

Let us see that quantum correlators depend on the equivalence classes only. The key insight is that the BRST operator is hermitian. This implies that exact states $|\Psi\rangle = Q|\Omega\rangle$ have zero overlap with closed states $|\Phi\rangle$

$$\langle\Phi|\Psi\rangle = \langle\Phi|Q|\Omega\rangle = \langle\Phi|Q^\dagger|\Omega\rangle = 0. \tag{3.91}$$

Therefore the scalar product is independent of the representative $|\Psi'\rangle = |\Psi\rangle + Q|\Omega\rangle$ of an equivalence class

$$\langle\Phi|\Psi'\rangle = \langle\Phi|\Psi\rangle + \langle\Phi|Q|\Omega\rangle = \langle\Phi|\Psi\rangle. \tag{3.92}$$

One-Particle States. Let us discuss the physical modes of the gauge and ghost fields in terms of BRST cohomology. For the discussion of single-particle states it suffices to restrict to the linear terms in the BRST transformation and substitute

the equation of motion for $B = -\xi^{-1}\partial^\mu A_\mu$. Expressed in momentum space, the resulting transformations read

$$\begin{array}{ll} QA_\mu \sim p_\mu C, & Q\tilde{C} \sim p^\mu A_\mu, \\ QC \sim 0, & p^\mu Q A_\mu \sim 0. \end{array} \tag{3.93}$$

We can then make the following statements:

- The BRST operator maps $\tilde{C}$ to $p^\mu A_\mu$. Hence the anti-ghost field $\tilde{C}$ is not closed.[20]
- The vector field A_μ is mapped to $p_\mu C$. Thus one of the four modes of A_μ is not closed.
- The image of $\tilde{C}$ is $p^\mu A_\mu$. Therefore one of the four modes of A_μ is exact. It is different from the mode which is not closed since exact states are closed by construction.
- The image of A_μ is C. Therefore the ghost field C is exact.
- The two remaining polarisation modes of the field A_μ are closed but not exact.

We observe that the anti-ghost $\tilde{C}$ and one of the modes of A_μ are not closed and therefore not physical. Furthermore the ghost C and another mode of A_μ are exact and therefore equivalent to the trivial state in $\mathbb{V}_{\text{phys}}$. The two remaining modes of A_μ are physical because they are closed but not exact. These are precisely the two helicity modes of the massless vector field.

We conclude that the cohomology for one-particle particles describes correctly the set of physical particles. The discussion for higher number of particles is not as straight-forward due to non-linear effects.

S-Matrix Consistency. A related issue is unitarity of the physical scattering matrix. The BRST operator is a symmetry; therefore it commutes with the Hamiltonian and with the S-matrix[21]

$$[Q, S] = 0. \tag{3.94}$$

Consider some physical initial state $|\Psi\rangle$. It is scattered to a final state $S|\Psi\rangle$ which is closed

$$QS|\Psi\rangle = SQ|\Psi\rangle = 0. \tag{3.95}$$

As expected, changing the representative of the initial state changes the representative of the final state

$$S|\Psi'\rangle = S|\Psi\rangle + SQ|\Omega\rangle = S|\Psi\rangle + QS|\Omega\rangle. \tag{3.96}$$

We can now consider unitarity of the S-matrix $S^\dagger S = 1$. Towards deducing the optical theorem we should insert the decomposition of unity between the S-matrices

$$1 = \sum_\Omega |\Omega\rangle\langle\Omega^*|, \tag{3.97}$$

[20] And it evidently cannot be made closed in any linear combination of single-particle states.

[21] Also the ghost number operators N_{gh} is a symmetry.

where $\{|\Omega\rangle\}$ is some basis of the full Hilbert space $\mathbb{V}$ and $\{\langle\Omega^*|\}$ is the corresponding dual basis of $\mathbb{V}^*$ with $\langle\Omega^*|\Omega'\rangle = \delta_{\Omega,\Omega'}$.[22] For our purposes it makes sense to pick a basis of exact, physical and the remaining states

$$\mathbb{V} = \mathbb{H}_{\text{op}} \oplus \mathbb{H}_{\text{phys}} \oplus \mathbb{V}_{\text{ex}}, \qquad \mathbb{H}_{\text{op}} := \mathbb{V}/\mathbb{V}_{\text{cl}}. \tag{3.98}$$

Let us denote the remaining states in $\mathbb{H}_{\text{op}}$ as "open" because they are in the complement of the closed states. They represent the preimage of the exact states under Q, and dividing out the closed states, essentially turns the map $Q : \mathbb{H}_{\text{op}} \to \mathbb{V}_{\text{ex}}$ into a bijection. This decomposition has some counter-intuitive features because the space $\mathbb{V}_{\text{ex}}$ is null in the scalar product of states. Let us discuss the implications briefly:[23] We multiply the decomposition of unity by Q from the left to find

$$Q1 = \sum_{\Omega \text{ open}} Q|\Omega\rangle\langle\Omega^*|. \tag{3.99}$$

Now we multiply the decomposition of unity by Q from the right noting that the image of this operator is $\mathbb{V}_{\text{ex}}$

$$1Q = \sum_{\Omega \text{ exact}} |\Omega\rangle\langle\Omega^*|Q. \tag{3.100}$$

By comparing the two expressions, we can deduce that Ω being exact implies that Ω^* is open and vice versa. Therefore, the decomposition of unity takes the form

$$1 = \sum_{\Omega \text{ physical}} |\Omega\rangle\langle\Omega^*| + \sum_{\Omega \text{ exact}} |\Omega\rangle\langle\Omega^*| + \sum_{\Omega^* \text{ exact}} |\Omega\rangle\langle\Omega^*|. \tag{3.101}$$

Let us now turn to the formulation of the optical theorem by considering matrix elements of the unitarity statement for two physical states

$$\langle\Phi|\Psi\rangle = \langle\Phi|S^\dagger S|\Psi\rangle = \sum_{\Omega}\langle\Phi|S^\dagger|\Omega\rangle\langle\Omega^*|S|\Psi\rangle. \tag{3.102}$$

The above decomposition of unity is very convenient because the exact states $|\Omega\rangle$ and $\langle\Omega^*|$ are cancelled in the combination $\langle\Phi|S^\dagger|\Omega\rangle$ and $\langle\Omega^*|S|\Psi\rangle$, respectively. The optical theorem thus reduces to

$$\langle\Phi|\Psi\rangle = \sum_{\Omega \text{ physical}} \langle\Phi|S^\dagger|\Omega\rangle\langle\Omega|S|\Psi\rangle. \tag{3.103}$$

In the optical theorem one is therefore free to restrict the sum over intermediate states to physical intermediate states as desired (as long as the external states are physical). However, one can also perform the sum over all intermediate states including unphysical ones. The result was shown to be the same, therefore the

[22] We neglect subtleties regarding dualisation in infinite-dimensional spaces.

[23] A simple analogue is Minkowski space $\mathbb{V} = \mathbb{R}^{D-1,1}$ with a fixed null vector n^μ and the operator $Qv = (n{\cdot}v)n$ which evidently squares to zero. The following discussion can be worked out very explicitly in this case.

contributions from unphysical gauge modes must cancel exactly against the contributions from ghost modes. The applicability of the full intermediate Hilbert space is crucial because unphysical gauge and ghost fields do propagate within loops.

In conclusion, it is consistent to restrict the S-matrix to physical external states, even though loop amplitudes are computed using all the fields.

Independence of Gauge Fixing. Finally, let us consider the relevance of the gauge-fixing terms of the Lagrangian. We have already seen that $\mathcal{L}_{\text{BRST}}$ is exact, $\mathcal{L}_{\text{BRST}} = QK_{\text{BRST}}$, while $\mathcal{L}_{\text{YM}}$ is physical. Both terms contribute to the Hamiltonian and therefore they govern the time-evolution and scattering of states.

Consider a physical state $|\Psi\rangle$ and act with $\mathcal{L}_{\text{BRST}}$. It can be written as

$$\mathcal{L}_{\text{BRST}}|\Psi\rangle = (QK_{\text{BRST}})|\Psi\rangle = Q\big(K_{\text{BRST}}|\Psi\rangle\big). \tag{3.104}$$

It shows that whenever the gauge-fixing term is applied to a physical state, the result must be an exact state which is equivalent to the zero state.

This sketch of an argument demonstrates that the precise formulation of the gauge fixing does not influence the correlators of physical states. In particular all physical correlators are independent of the gauge-fixing parameter ξ.

3.5 Feynman Rules

We can now derive the Feynman rules for gauge-fixed Yang–Mills theory.

Gauge Sector. In the above considerations we have assumed that the fields A_μ, $F_{\mu\nu}$, C and $\tilde{C}$ are $N \times N$ hermitian matrices. We further note that these matrices frequently appear within commutators, from which any contributions proportional to the unit matrix drop out, both in the input as well as in the output. Consequently, it makes sense to remove degrees of freedom proportional to the unit matrix from these fields by restricting the matrices to be traceless. Effectively, this reduces the gauge group from $\mathrm{U}(N)$ to its simple subgroup $\mathrm{SU}(N)$.[24]

For the purposes of Feynman diagrams, we have to enumerate the various matrix components. For the gauge group $\mathrm{U}(N)$ or $\mathrm{SU}(N)$ this can be done conveniently by using row and column indices; the resulting approach will be outlined further below. The more common and more universal approach is to introduce some basis T_a of $N \times N$ hermitian traceless matrices enumerated by an index

[24]The formulation of Yang–Mills models with gauge group $\mathrm{U}(N)$ vs. $\mathrm{SU}(N)$ makes no big difference. Note, however, that a $\mathrm{U}(N)$ gauge group enables a non-trivial kinetic term $(\operatorname{tr} F^{\mu\nu})(\operatorname{tr} F_{\mu\nu})$ in addition to the standard term $\operatorname{tr}(F^{\mu\nu}F_{\mu\nu})$. These two terms become relevant in the quantum regime because they renormalise differently. It therefore makes good sense to consider the $\mathrm{U}(1)$ and $\mathrm{SU}(N)$ subgroups separately.

$a = 1, \ldots, N^2 - 1$. The matrix-valued fields X are then expanded in terms of components X^a as

$$\begin{aligned} A_\mu &= g T_a A^a_\mu, \qquad & C &= g T_a C^a, \\ F_{\mu\nu} &= g T_a F^a_{\mu\nu}, \qquad & \tilde{C} &= g T_a \tilde{C}^a. \end{aligned} \tag{3.105}$$

Here g is a normalisation factor which we identify with the Yang–Mills coupling $g = g_{\mathrm{YM}}$.

Let us first express the field strength in terms of components

$$\begin{aligned} F_{\mu\nu} &= (\partial_\mu A_\nu) - (\partial_\nu A_\mu) - \mathring{\imath}[A_\mu, A_\nu] \\ &= g T_a (\partial_\mu A^a_\nu) - g T_a (\partial_\nu A^a_\mu) - \mathring{\imath} g^2 [T_b, T_c] A^b_\mu A^c_\nu \\ &= g T_a \big((\partial_\mu A^a_\nu) - (\partial_\nu A^a_\mu) + g f_{bc}{}^a A^b_\mu A^c_\nu \big). \end{aligned} \tag{3.106}$$

Here we have expanded the commutator $[T_b, T_c]$ in the basis T_a of hermitian traceless matrices as

$$[T_b, T_c] = \mathring{\imath} f_{bc}{}^a T_a, \tag{3.107}$$

where we make use of the fact that the trace of a commutator must vanish. The constants $f_{bc}{}^a$ are called structure constants, and they depend on the basis T^a used for their construction. Hence we find for the field strength

$$F^a_{\mu\nu} = (\partial_\mu A^a_\nu) - (\partial_\nu A^a_\mu) + g f_{bc}{}^a A^b_\mu A^c_\nu. \tag{3.108}$$

We further define the trace of a product of two matrices T_a as the matrix k_{ab}

$$k_{ab} := 2 \operatorname{tr}(T_a T_b). \tag{3.109}$$

It would make sense to normalise the basis T^a canonically such that $k_{ab} = \delta_{ab}$. Here we keep k_{ab} serving as an abstract metric and define k^{ab} to be its inverse matrix. The Yang–Mills Lagrangian then reads

$$\mathcal{L}_{\mathrm{YM}} = -\frac{1}{2g^2} \operatorname{tr} F^{\mu\nu} F_{\mu\nu} = -\tfrac{1}{4} k_{ab} F^{a,\mu\nu} F^b_{\mu\nu}. \tag{3.110}$$

This is similar to one copy of the photon Lagrangian for each component A^a_μ of the matrix-valued gauge potential. Note, however, that the field strength $F^a_{\mu\nu}$ now depends non-linearly on the gauge potential A^a_μ.

Next consider the ghost and gauge-fixing terms. In components they read

$$\begin{aligned} \mathcal{L}_{\mathrm{gh}} &= \mathring{\imath} k_{ab} \tilde{C}^a \big(\partial^\mu (D^{\mathrm{ad}}_\mu C)^b \big), \\ \mathcal{L}_{\mathrm{gf}} &= -\tfrac{1}{2} \xi^{-1} \, k_{ab} (\partial^\mu A^a_\mu)(\partial^\nu A^b_\nu). \end{aligned} \tag{3.111}$$

Here the covariant derivative of the ghost field C reduces to

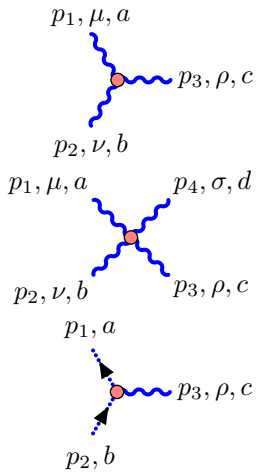
$$(D^{\mathrm{ad}}_\mu C)^a = (\partial_\mu C^a) + g f_{bc}{}^a A^b_\mu C^c. \tag{3.112}$$

Altogether, the pure gauge sector consists of the gauge and ghost fields. In the Feynman graphs the gauge fields are commonly denoted by curly or wiggly lines,

and the ghosts by straight dotted or dashed lines. Their momentum space Feynman propagators read

$$\begin{aligned} \mu, a \overset{p}{\sim} \nu, b &\to \frac{-\mathring{\imath} k^{ab}\big(\eta_{\mu\nu} - (1-\xi) p_\mu p_\nu / p^2\big)}{p^2 - \mathring{\imath}\epsilon}\,, \\ a \overset{p}{\cdots\!\blacktriangleright\!\cdots} b &\to \frac{-\mathring{\imath} k^{ab}}{p^2 - \mathring{\imath}\epsilon}\,. \end{aligned} \tag{3.113}$$

There are three types of vertices arising from the above Lagrangian: cubic and quartic interactions of the gauge field as well as an interaction of one gauge field with a ghost line[25]



$$\begin{aligned} &-g f_{abc}(p_1 - p_2)^\rho \eta^{\mu\nu} \\ \to\ &-g f_{abc}(p_2 - p_3)^\mu \eta^{\nu\rho} \\ &-g f_{abc}(p_3 - p_1)^\nu \eta^{\rho\mu}, \end{aligned}$$

$$\begin{aligned} &-\mathring{\imath} g^2 f_{abe} f_{cdf} k^{ef}\big(\eta^{\mu\rho}\eta^{\nu\sigma} - \eta^{\nu\rho}\eta^{\mu\sigma}\big) \\ \to\ &-\mathring{\imath} g^2 f_{ace} f_{bdf} k^{ef}\big(\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\nu\rho}\eta^{\mu\sigma}\big) \\ &-\mathring{\imath} g^2 f_{ade} f_{bcf} k^{ef}\big(\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\nu\sigma}\eta^{\mu\rho}\big), \end{aligned}$$

$$-g f_{abc} p_1^\mu. \tag{3.114}$$

For these terms, the upper index of the structure constants has been lowered with the abstract metric k_{ab}

$$f_{abc} := k_{ad} f_{bc}{}^d. \tag{3.115}$$

We note that all four polarisations of the gauge field as well as both ghost and anti-ghost fields are admissible for creating asymptotic (massless) particle states. However, the resulting amplitudes will only describe physically applicable processes if and only if all external legs are physical in the sense of (linearised) BRST symmetry. This removes all ghost fields and two gauge field polarisations from the external legs. It is even sufficient to restrict to physical polarisations for the incoming particles of a scattering process: Here, scattering from a perfectly physical incoming state can produce unphysical excitations in the outgoing state. However, the particular structure of BRST symmetry explains why such outgoing states may be discarded without causing violations of unitarity.

Matter Couplings. Let us now consider spinor matter fields forming an N-dimensional vector. To this end, we introduce a basis v_k, $k = 1, \ldots, N$ of $\mathbb{C}^N$

[25] Note that the gauge structure of the quartic vertex is equivalent to two cubic vertices connected by a propagator. In fact, one can eliminate the quartic interaction by means of a non-propagating auxiliary field interacting via a cubic term. This would make all vertices cubic and proportional to the structure constants of the gauge group.

and a dual basis $\bar{v}^k$ of the dual vector space with $\bar{v}^k v_j = \delta^k_j$. The components $(T_a)^j{}_k$ of the basis T^a of matrices are thus obtained as

$$(T_a)^k{}_j := \bar{v}^k T_a v_j \qquad \text{or} \qquad T_a v_k \stackrel{!}{=} (T_a)^j{}_k v_j. \tag{3.116}$$

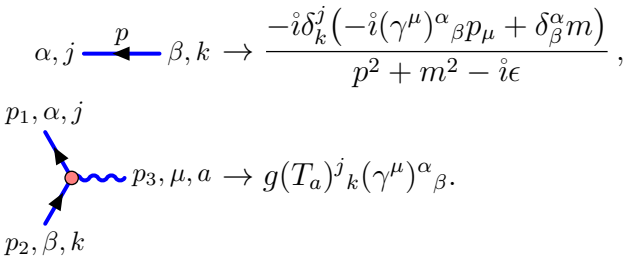
The kinetic term for spinor matter in Yang–Mills theory reads

$$\begin{aligned} \mathcal{L}_{\text{spinor}} &= \bar{\psi}(\gamma^\mu D_\mu - m)\psi \\ &= \bar{\psi}_k\big(\gamma^\mu (D_\mu \psi)^k - m\psi^k\big) \end{aligned} \tag{3.117}$$

with the matter covariant derivative

$$(D_\mu X)^k = (\partial_\mu X^k) - \mathring{\imath} g (T_a)^k{}_j A^a_\mu X^j. \tag{3.118}$$

This results in the following Feynman rules involving the spinor field



(3.119)

3.6 Planar Limit

Finally, we will consider a gauge theory whose gauge group has a very large rank N.[26] Naively one would think that computations in such a theory become rather involved. However, 't Hooft realised that the converse is true if one restricts attention to the leading behaviour at large rank. In particular, the limit allows to classify graphs by their two-dimensional genus, and the graph of lowest genus contribute most dominantly. The so-called *planar* or *large-N limit* has since become a useful tool to obtain approximations[27] and new formal insights.[28]

Ribbon Graphs. Let us therefore investigate the behaviour of Feynman diagrams at large rank. More concretely, we will consider a pure U(N) Yang–Mills theory without matter fields, and determine the N-dependency of a Feynman diagram which is invariant under (global) gauge transformations. To that end we expand the gauge potential in the basis of matrix elements $(A_\mu)^i{}_k$. For a U(N)

[26] The following discussion is not essential for the understanding of the following chapters, it merely represents an exciting advanced topic within QFT.

[27] Even though the gauge groups in the standard model admittedly have a very low rank, curiously, the large-N approximation are applicable to some extent there, partially because the rank typically enters as a square.

[28] In particular, the large-N limit introduces two-dimensional structures to higher-dimensional gauge theory which are reminiscent of string theory aspects.

gauge group all of these component fields are independent. We sketch the form of the Yang–Mills Lagrangian as follows[29]

$$\mathcal{L} \sim \frac{N}{\lambda}\left[(\partial A)^i{}_k(\partial A)^k{}_i + (\partial A)^i{}_j A^j{}_k A^k{}_i + \mathcal{O}(A^4)\right]. \tag{3.120}$$

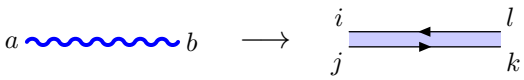
Here we paid no attention to the precise spacetime structure, but merely highlighted the matrix structure and the dependence on the parameters of the model. In particular, we traded in the Yang–Mills coupling g_{YM} for the *'t Hooft coupling*

$$\lambda := g_{\mathrm{YM}}^2 N, \tag{3.121}$$

which is meant to be held fixed in the limit $N \to \infty$.[30] The gauge field propagator then possesses the following gauge group structure

$$\langle A^i{}_j A^k{}_l \rangle \sim \frac{\delta^i_l \delta^k_j}{N}\,. \tag{3.122}$$

This structure has a useful graphical notation if we depict δ^i_j by a directed line from j to i. Then the gauge field propagator becomes a pair of parallel lines directed in opposite directions:

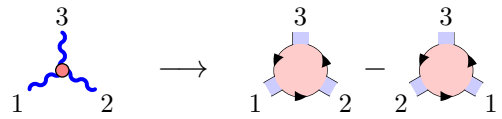


(3.123)

This is the *double-line notation* for gauge fields. The cubic vertex is proportional to the structure constants. As we have seen above, it can be expanded as a trace

$$f_{abc} \sim \mathrm{tr}\big([T_a, T_b]T_c\big). \tag{3.124}$$

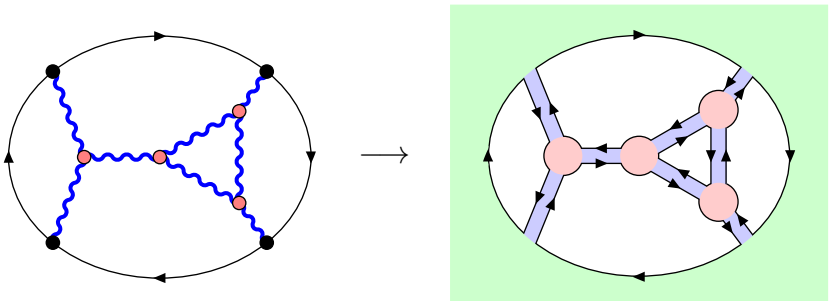
Correspondingly, the cubic vertex connects the indices of three gauge fields in a cyclic fashion. There are two inequivalent cyclic orderings for three fields and the cubic vertex is their difference:



(3.125)

Likewise, the quartic vertex can be written as some combination of traces.

We can now assemble the gluon propagators and vertices into a so-called *ribbon graph* or *fat graph*, e.g.:



(3.126)

[29] Ghost fields have precisely the same matrix structure as the gauge fields, and there is no need to discuss them separately.

[30] We will drop the constant λ from here on. Just like g^2_{YM} or $\hbar$ it will essentially count loops.

In order to complete the graph and make it properly gauge-invariant, we saturate all the indices of the external gluons with additional traces.[31]

Large-N Scaling. In the above representation it is easy to compute the N-dependence of the graph. By construction there are no open indices, and thus the directed lines form closed cycles which are known as *index loops*. Each index loop provides a factor of

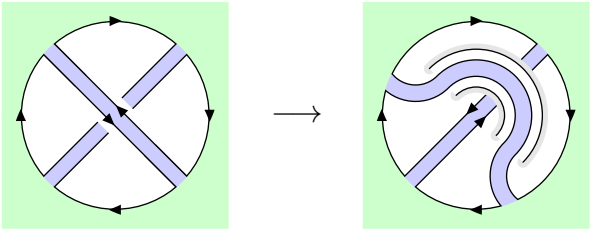
$$\delta^i_j \delta^j_k \delta^k_{\cdots} \cdots \delta^{\cdots}_i = \delta^i_i = \operatorname{tr} 1 = N. \tag{3.127}$$

Furthermore there is an explicit factor of N for each vertex and an explicit factor of N^{-1} for each propagator. The overall N-dependence of the graph therefore reads

$$N^{F+V-E}, \tag{3.128}$$

where F denotes the number of closed index loops (faces), V the number of vertices and E the number of propagators (edges). In the above example, these numbers are $F = 5$, $V = 4$, $E = 8$ yielding the overall dependence N^1.

As such, the exponent of N depends very much on the details of the graph, but we can use Euler's formula for planar graphs to rewrite it in a useful way. In order to apply the formula, we have to draw the graph on a two-dimensional surface in such a way that no two edges cross. It is easy to see that such a surface always exists: if the graph cannot be drawn on a given surface without crossings, one can add handles to the surface via which some edges can be routed past others without having them cross on the surface:



(3.129)

Adding such handles to the surface increases its topological genus. The genus of a given graph is defined as the minimum genus of a surface onto which the graph can be drawn without crossings. Moreover, the traces to contract indices of the external gluons should be interpreted as boundaries of the surface (no part of the graph should reside within the trace). *Euler's formula* then takes the form

$$V - E + F = 2C - 2G - T. \tag{3.130}$$

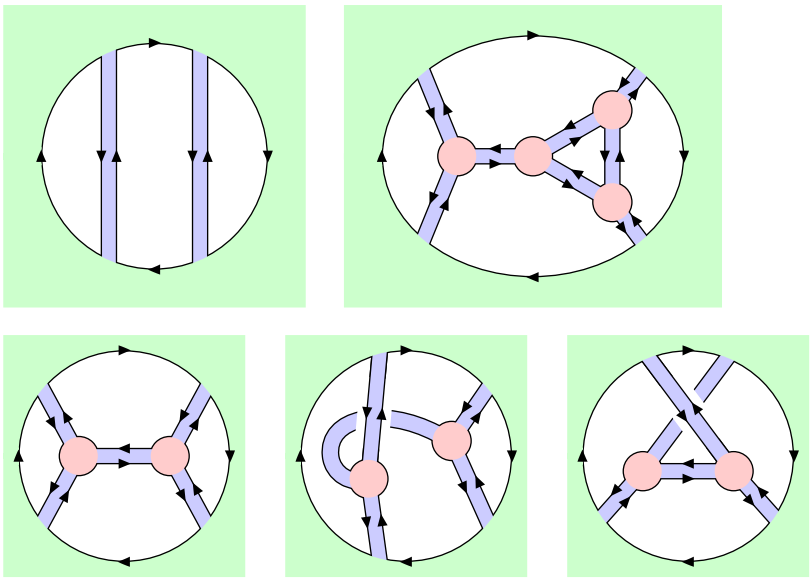
Here, C denotes the number of connection components, G the total genus (number of handles) of all component surfaces and T the number of traces alias boundaries. In total the N-dependence of the graph is given by

$$N^{2C-2G-T} = N^\chi. \tag{3.131}$$

[31] This happens naturally when considering correlation functions of local composite operators (e.g. $\operatorname{tr} F_{\mu\nu} F_{\rho\sigma}$), Wilson loops ($\operatorname{tr} \mathrm{P} \exp \oint A$) as well as colour-ordered scattering amplitudes.

The quantity χ is called the *Euler characteristic* of the surface which is a number to characterise its topology. In particular, if we restrict to connected graphs with a fixed number of external traces, the exponent is bounded from above. Graphs may then differ by their genus, and the graphs with low genus will contribute dominantly if N is large. In the strict large-N limit only the graphs of lowest admissible genus (typically genus $G = 0$ corresponding to a sphere) contribute to a given process. This implies a substantial combinatorial simplification because the number of graphs at fixed genus grows much slower with an increasing number of vertices than the number of all graphs. In other words, only a tiny fraction of all graphs has the lowest genus.

Example. Let us compare some ribbon graphs contributing to the correlator of a trace of four gluons $\langle \operatorname{tr} A^4 \rangle$:



(3.132)

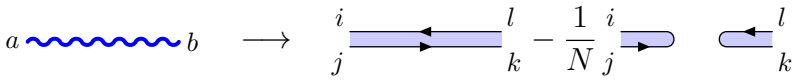
By counting closed index loops, propagators and vertices one find the following large-N behaviours N^{F-E+V} for these graphs

$$N^{3-2=1}, \qquad N^{5-8+4=1}, \qquad N^{4-5+2=1}, \qquad N^{2-5+2=-1}, \qquad N^{2-5+2=-1}. \tag{3.133}$$

Indeed, the surface (white area with graph removed) of the former three graphs is a disc ($C = 1$, $G = 0$, $T = 1$) with Euler characteristic $\chi = 1$. For the latter two it is a disc with one handle ($C = 1$, $G = 1$, $T = 1$) with Euler characteristic $\chi = -1$. We thus confirm that the large-N behaviour depends on the graph topology rather than the detailed structure of its vertices and connections.

Beyond Pure Yang–Mills Theory. The above discussion can be extended beyond the case of pure U(N) Yang–Mills theory in several ways.

It is fairly easy to reduce the gauge group from U(N) to the simple subgroup SU(N). In this case, the gauge field must be traceless which implies the following form of propagator:



(3.134)

It is easy to see that the second part of the propagator (corresponding to the abelian subalgebra) effectively does not couple to gluon vertices. Therefore this restriction of the gauge group introduces changes only for propagators whose ends both couple to external traces (or matter fields).

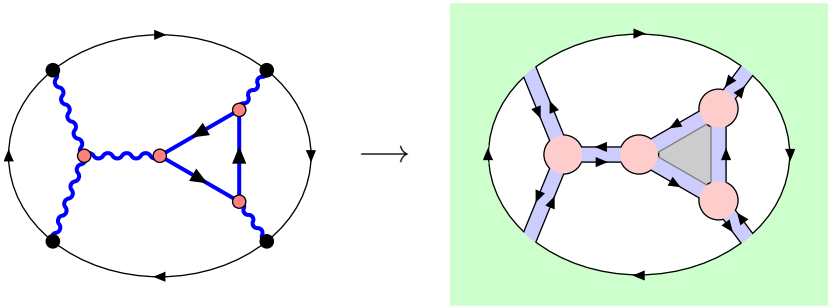
The above discussion can also be generalised to the orthogonal and symplectic gauge groups $\mathrm{SO}(N)$ and $\mathrm{Sp}(N)$. In that case, also graphs on unoriented surfaces come into play leading to additional topologies and associated Euler characteristics.

Finally, we can add matter fields. Typically, matter fields transform in the defining representation, and consequently their propagators are represented by single lines rather than double lines.[32] These single lines should be drawn on additional boundary components of the surface. The N-dependence is then given by[33]

$$N^{2C-2G-T-B} = N^{\chi}, \tag{3.135}$$

where B denotes the number of boundaries for fields in the defining representation.

Example. We replace the gluon loop in the above Feynman diagram by a loop of matter fields in the defining representation:



(3.136)

This graph has $F = 4$ index loops, $E = 5$ gluon propagators and $V = 1$ gluon vertices. Its N-dependence reads $N^{4-1+5} = N^0$ in agreement with the Euler characteristic $\chi = 0$ of an annulus (disc with a hole).

Duality to String Theory. The above considerations led 't Hooft to conjecture the emergence of a string field theory in the large-N limit of gauge theory: In string field theory, an amplitude on a two-dimensional string worldsheet with Euler characteristic χ is proportional to $g_{\mathrm{s}}^{-\chi}$, where g_{s} is the string coupling constant. This is in agreement with the N-dependence in gauge theory upon identifying

$$g_{\mathrm{s}} \sim \frac{1}{N}\,. \tag{3.137}$$

This analogy between gauge and string theory structures was made more concrete in the so-called AdS/CFT-correspondence. The latter conjectures the exact duality between some gauge and string theories, and the large-N limit is an essential tool to understand it.

[32] Matter fields in the adjoint representation are completely analogous to gluons in terms of their N-dependence. The large-N behaviour of matter fields in higher representations is not captured well by graph theory.

[33] Note that only the gluon propagators and pure gluon vertices contribute explicit factors of N^{-1} and N, respectively. Matter propagators and vertices do not contribute to E and V which is consistent with them residing on the boundary.

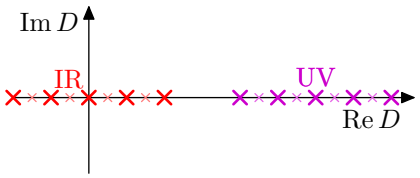
4 Renormalisation

We have touched upon loop effects, divergences and regularisation and renormalisation at the very end of QFT I. Here we will present a broadly applicable regulator, apply the renormalisation to gauge theory and discuss the effects of changing the renormalisation scale.

4.1 Dimensional Regularisation

In QFT I we have used a plain momentum cutoff to regularise particular divergent loop integrals. It served its purpose, but it is hard to define consistently for all conceivable processes, it breaks Poincaré symmetry and it violates gauge symmetry.

A modern regulator without these shortcomings is dimensional regularisation. It is based on the fact that the degree of divergence of loop contributions crucially depends on the dimension D of spacetime. For a sufficiently large D, every graph develops UV-divergences (from integrating over large momenta) whereas for small D some graphs can have IR divergences (from integrating over small or light-like momenta). The idea of dimensional regularisation is to assume the number of spacetime dimensions D to be a continuous (real or complex) parameter. Here, the divergences appear as poles in D at integer values.



(4.1)

Basic Integral. The basic integral to consider in dimensional regularisation is

$$I_n(X) = \int \frac{-\mathring{\imath}\,\mathrm{d}\ell^D}{(2\pi)^D} \frac{1}{(\ell^2 + X - \mathring{\imath}\epsilon)^n}\,. \tag{4.2}$$

For convenience, we first Wick rotate the integral

$$I_n(X) = \int \frac{\mathrm{d}\ell_{\mathrm{E}}^D}{(2\pi)^D} \frac{1}{(\ell_{\mathrm{E}}^2 + X - \mathring{\imath}\epsilon)^n}\,. \tag{4.3}$$

Now the integrand is spherically symmetric, therefore we transform to spherical coordinates

$$I_n(X) = \mathrm{Vol}(S^{D-1}) \int_0^\infty \frac{\mathrm{d}\ell_{\mathrm{E}}}{(2\pi)^D} \frac{\ell_{\mathrm{E}}^{D-1}}{(\ell_{\mathrm{E}}^2 + X - \mathring{\imath}\epsilon)^n}\,. \tag{4.4}$$

The volume of a $(D-1)$-sphere equals

$$\mathrm{Vol}(S^{D-1}) = 2\pi^{D/2}/\,\Gamma(D/2). \tag{4.5}$$

The remaining radial integral after a change of variables $\ell_{\mathrm{E}}^2 = z$ is a Beta-function in D

$$\frac{1}{2}\int_0^\infty \frac{\mathrm{d}z}{(2\pi)^D}\,\frac{z^{D/2-1}}{(z+X-\mathring{\imath}\epsilon)^n} = \frac{\Gamma(D/2)\,\Gamma(n-D/2)}{2(2\pi)^D\,\Gamma(n)\,(X-\mathring{\imath}\epsilon)^{n-D/2}}\,. \tag{4.6}$$

Altogether the integral reads

$$I_n(X) = \frac{\Gamma(n-D/2)}{(4\pi)^{D/2}\,\Gamma(n)}\,\frac{1}{(X-\mathring{\imath}\epsilon)^{n-D/2}}\,. \tag{4.7}$$

The resulting expression has the nice feature that it is defined not only for integer D, but also for continuous values of the dimensionality of spacetime. We may thus use the above formula as the definition of the integral for continuous D.[1] We have seen in QFT I that all integrals arising from Feynman diagrams can be brought into this form. Hence the above formula defines an extrapolation of arbitrary Feynman diagrams to non-integer dimensions.

We observe an important feature: The Gamma-function in the numerator has poles for even integers $D \geq 2n$. Indeed, the integral has a UV-divergence when $D \geq 2n$.[2] However, for non-integer values of D, the function is perfectly finite. So indeed we can use a continuous number of spacetime dimensions D as a regulator for the integral. For four-dimensional theories, a convenient choice is to take D slightly below 4 [3]

$$D = 4 - 2\epsilon. \tag{4.8}$$

Dimensional regularisation provides a finite result for arbitrary Feynman integrals at arbitrary number of loops (as long as D is not rational). Moreover it does not need to introduce a preferred point or direction which would violate Poincaré symmetry.

Vectors. For models with spin, we will have to deal with vectors in the numerator of the integral, for example[4]

$$\begin{aligned} I_n^\mu(X) &= \int \frac{-\mathring{\imath}\,\mathrm{d}\ell^D}{(2\pi)^D}\,\frac{\ell^\mu}{(\ell^2+X-\mathring{\imath}\epsilon)^n} = 0, \\ I_n^{\mu\nu}(X) &= \int \frac{-\mathring{\imath}\,\mathrm{d}\ell^D}{(2\pi)^D}\,\frac{\ell^\mu\ell^\nu}{(\ell^2+X-\mathring{\imath}\epsilon)^n}\,. \end{aligned} \tag{4.9}$$

[1] This does not imply that the formula is morally correct or unique. It is merely a convenient and smooth interpolation between integer dimensions.

[2] In fact, the integral diverges also for odd D, so a pole in D is a sufficient condition for a divergent integral, but not a necessary one. For instance, in odd dimensions, poles typically occur only at even loop orders.

[3] A factor of 2 for ϵ conveniently cancels a factor of $^1/_2$ which is often associated to D.

[4] Note that ϵ as the deviation from $D=4$ and ϵ of the $\mathring{\imath}\epsilon$ prescription for propagators are clearly distinct and independent parameters. Nevertheless we will use the same symbol for both objects.

Evidently, the first integral vanishes because the integrand is an anti-symmetric function. An alternative derivation uses Lorentz symmetry: The result must transform as a vector. However, there is no vector at our disposal after the integral is done. Hence the result must vanish. The second integral can be performed by a similar trick: The only Lorentz-invariant bi-vector is the metric tensor $\eta_{\mu\nu}$, so the integral must be proportional to it. To find the factor of proportionality, take the trace with $\eta_{\mu\nu}$. This yields

$$\begin{aligned} I_n^{\mu\nu}(X) &= \frac{\eta^{\mu\nu}}{D}\,\eta_{\rho\sigma} I_n^{\rho\sigma}(X) = \frac{\eta^{\mu\nu}}{D}\,\big(I_{n-1}(X) - X I_n(X)\big) \\ &= \frac{\Gamma(n-1-D/2)}{2(4\pi)^{D/2}\,\Gamma(n)}\,\frac{\eta^{\mu\nu}}{(X-\mathring{\imath}\epsilon)^{n-1-D/2}}\,. \end{aligned} \tag{4.10}$$

Note that one has to pay attention to the dimensionality of vectors and matrices. For consistency all vectors should be defined in $D = 4 - 2\epsilon$ dimensions. Likewise the Minkowski metric must be defined in $D = 4 - 2\epsilon$ dimensions, which implies that its trace equals D,[5]

$$\delta^\mu_\mu = \eta_{\mu\nu}\eta^{\mu\nu} = D = 4 - 2\epsilon. \tag{4.11}$$

Although, the difference between D and 4 is small, it can be important in the case of divergences which contribute factors of $1/(D-4)$.

Spinors. Furthermore, spinor models are formulated in terms of gamma-matrices. A set of Dirac gamma-matrices satisfying the corresponding Clifford algebra exists in arbitrary spacetime dimension D.[6] The dimension of the spinor representation grows exponentially with D, so for evaluating a trivial trace we must set

$$\operatorname{tr} 1 = 2^{D/2}. \tag{4.12}$$

Everything else must be derived strictly from the Clifford algebra relation

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}, \tag{4.13}$$

whose Minkowski metric belongs to D dimensions. For instance, some of the bracketing identities generalise as

$$\begin{aligned} \gamma^\mu\gamma_\mu &= \tfrac{1}{2}\{\gamma^\mu, \gamma_\mu\} = D, \\ \gamma^\mu\gamma^\nu\gamma_\mu &= \{\gamma^\mu, \gamma^\nu\}\gamma_\mu - \gamma^\nu\gamma^\mu\gamma_\mu = -(D-2)\gamma^\nu, \\ \gamma^\mu\gamma^{[\rho\sigma]}\gamma_\mu &= \ldots = (D-4)\gamma^{[\rho\sigma]}. \end{aligned} \tag{4.14}$$

Particularly, the last one has a non-trivial right-hand side away from $D = 4$ which can become relevant in connection to $1/\epsilon$ divergences.

[5] At the end of the day, the external momentum vectors will be defined in exactly four dimensions, which may lead to some confusion.

[6] The existence of Weyl and Majorana spinors depends discretely on the particular choice of D modulo 8, therefore these representations are not suitable for dimensional regularisation.

A complication of dimensional regularisation is that there is no universal generalisation of the totally antisymmetric tensor $\varepsilon_{\mu\nu\rho\sigma}$ in continuous dimensions D. For spinors it means that the additional gamma-matrix γ^5 does not exist immediately. This shortcoming is related to important physical effects such as anomalies of quantum symmetries. A cure to this problem is to formally introduce a rank-4 antisymmetric tensor $\epsilon_{\mu\nu\rho\sigma}$ with[7]

$$\varepsilon_{0123} = +1. \tag{4.15}$$

The resulting matrix γ^5 is then similar to the usual γ^5 in four dimensions, but it has slightly different algebraic properties which become relevant in connection to divergences.

Introduction of a Scale. Let us consider a simple QFT model in dimensional regularisation. For example, the Lagrangian of a scalar field with a quartic interaction reads

$$\mathcal{L} = -\tfrac{1}{2}(\partial\phi)^2 - \tfrac{1}{2}m^2\phi^2 - \tfrac{1}{24}\lambda\phi^4. \tag{4.16}$$

Consider now the mass dimension. For $D = 4$, ϕ, m, ∂ all have unit mass dimension and λ is a dimensionless quantity. All terms in the Lagrangian have mass dimension 4 so that the action is properly dimensionless.

In dimensional regularisation there are two choices to obtain a dimensionless action $S = \int \mathrm{d}x^D\, \mathcal{L}$:

- We can keep the dimension of all involved quantities fixed. Then we have to multiply the Lagrangian by some overall dimensionful factor

  $$\mathcal{L} = \mu^{-2\epsilon}\big(-\tfrac{1}{2}(\partial\phi)^2 - \tfrac{1}{2}m^2\phi^2 - \tfrac{1}{24}\lambda\phi^4\big). \tag{4.17}$$

  Here μ is some regularisation constant with unit mass dimension.
- The alternative is to adjust the mass dimension of the field ϕ to $1-\epsilon$. Then the kinetic terms have the desired mass dimension $4-2\epsilon$. The interaction term, however, has mass dimension $4-4\epsilon$. We can compensate it either by attributing a mass dimension 2ϵ to λ, or by factoring out this mass dimension and keeping λ dimensionless

  $$\mathcal{L} = -\tfrac{1}{2}(\partial\phi)^2 - \tfrac{1}{2}m^2\phi^2 - \tfrac{1}{24}\mu^{+2\epsilon}\lambda\phi^4. \tag{4.18}$$

Note that both choices are equivalent up to a rescaling of the field ϕ by μ^ϵ.

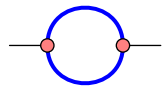
In the first choice, $\mu^{-2\epsilon}$ is an overall prefactor of the action. It acts like Planck's constant $1/\hbar$ which we have already observed to count loops in Feynman graphs. A contribution at L loops therefore comes along with a factor of

$$\lambda^L\, \mu^{2\epsilon L}. \tag{4.19}$$

This is in fact the only place where the new scale can appear. The dependence on μ therefore is always very mild since its exponent ϵ is always very small. It is logarithmic rather than polynomial.

[7] Strictly speaking this requires $D \geq 4$, but we may assume $\epsilon \leq 0$.

Expansion. Consider the standard massive scalar bubble integral

$$I = \quad \text{  }$$
$$= \frac{\mu^{2\epsilon}}{2} \int \frac{-\mathring{\imath}\,\mathrm{d}\ell^D}{(2\pi)^D} \frac{1}{\ell^2 + m^2 - \mathring{\imath}\epsilon} \frac{1}{(p-\ell)^2 + m^2 - \mathring{\imath}\epsilon} \tag{4.20}$$
$$= \frac{\mu^{2\epsilon}}{2} \int_0^1 \mathrm{d}z \int \frac{-\mathring{\imath}\,\mathrm{d}\ell^D}{(2\pi)^D} \frac{1}{\left(\ell^2 + z\bar{z}p^2 + m^2 - \mathring{\imath}\epsilon\right)^2}\,.$$

In the second step we have combined the denominators and shifted the integration variable to centre the integrand. The integral has the desired form and we use the dimensionally regularised expression

$$I = \frac{\Gamma(\epsilon)}{32\pi^2} \int_0^1 \mathrm{d}z \left(\frac{z\bar{z}p^2 + m^2 - \mathring{\imath}\epsilon}{4\pi\mu^2}\right)^{-\epsilon}. \tag{4.21}$$

Next we expand the integrand in ϵ [8]

$$I = \frac{1}{32\pi^2} \int_0^1 \mathrm{d}z \left(\frac{1}{\epsilon} - \gamma_\mathrm{E} - \log\frac{z\bar{z}p^2 + m^2 - \mathring{\imath}\epsilon}{4\pi\mu^2} + \mathcal{O}(\epsilon)\right). \tag{4.22}$$

We have performed the relevant parts of this integral in QFT I, here we obtain

$$\begin{aligned} I = &-\frac{1}{16\pi^2} \sqrt{\frac{p^2 + 4m^2 - \mathring{\imath}\epsilon}{-p^2}} \arctan \sqrt{\frac{-p^2}{p^2 + 4m^2 - \mathring{\imath}\epsilon}} \\ &+ \frac{1}{32\pi^2} \left(\frac{1}{\epsilon} - \gamma_\mathrm{E} + \log(4\pi) + 2 - \log\frac{m^2}{\mu^2}\right) + \mathcal{O}(\epsilon). \end{aligned} \tag{4.23}$$

The dimensional regularisation agrees with the cut-off scheme up to terms of order ϵ and up to some constant shifts. The constant shifts can actually be absorbed by an appropriate change of the renormalisation scale μ. Alternatively, they can be cancelled directly by counterterms. One of these shift terms is a pole at $\epsilon = 0$ which indicates the divergence in $D = 4$. Such poles are typical for dimensional regularisation.

We have terminated the expansion at ϵ^0. This makes sense since we are mainly interested in the physics in $D = 4$. What about the terms at positive orders of ϵ? Are they relevant? Yes and no!

- Some of the sub-leading terms in ϵ are required even at $D = 4$. This is the case whenever there are divergences in the Feynman diagram that can undo the suppression of $\mathcal{O}(\epsilon)$ terms.
- The higher-order terms in ϵ typically have a more complicated structure which is reminiscent of higher-loop contributions. In fact they compete with higher-loop terms and all terms are necessary to construct a consistent final result.

[8] Integration and expansion do not necessarily always commute, but here it is not a problem.

- After all divergences have been removed, we will eventually set $\epsilon = 0$ and thus drop these terms.
- Note that setting $\epsilon = 0$ too early may discard some essential terms from the final answer and make it inconsistent. For example, the external states in a scattering process are usually defined in $D = 4$. However, this may lead to violations of unitarity since it is conceivable that the corresponding integrals over the phase space of intermediate particles diverge. In those cases, the difference between a simplified phase space with $D = 4$ and the actual phase space with $D = 4 - 2\epsilon$ may be important.

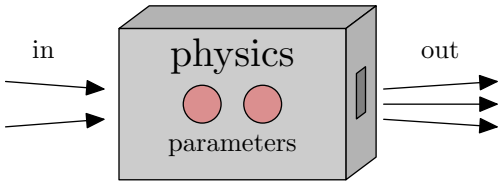
4.2 Renormalisation of Couplings

Dimensional regularisation makes Feynman integrals finite at the cost of introducing a regularisation parameter ϵ and a regularisation scale μ. The physically relevant model is at $\epsilon = 0$, but we cannot send ϵ to zero because typical results have some singularity at $\epsilon = 0$. We first need to remove these singularities by redefining the parameters and fields of the model. This allows to remove the regulator by sending $\epsilon \to 0$ and recover physically meaningful results. The crucial question is whether this so-called renormalisation procedure can render all observables finite at the same time.

Philosophy. Before we discuss the technical implementation of renormalisation, let us revisit two generic questions concerning its philosophy:

- Is it permissible to redefine the coupling constants?
- If so, is it permissible to redefine them by an infinite amount?

Towards answering the first question, it is important to realise that neither the bare parameters of the Lagrangian nor the renormalised parameters are *physical*. Neither of their values can actually be measured directly.[9] Effectively, it is only relevant that the model makes accurate predictions of the physics. The predictions certainly depend on the values of the (bare or renormalised) parameters, and therefore the latter need to be tuned to make the predictions as precise as possible. As there is no deeper meaning to the values of the parameters, they can be redefined.



(4.24)

Analogy. A practical analogy in this context is the tuning of a radio to receive a specific station (which in fact is similar to a physics experiment in many regards):

[9] To this end it makes sense to distinguish the physical mass of a particle representing the location of its propagator pole from the bare mass parameter in the Lagrangian or some renormalised mass parameter. Even though these three concepts are often directly related in the classical formulation, quantum effects introduce corrections.

arguably, the optimal strategy in terms of the result is to tune the radio such that the received signal displays the highest signal to noise ratio. After tuning, the setting of the dial will be close to the design frequency of the station. However, small deviations are to be expected due to the imprecise translation between the writing on the dial and the receiver frequency which can be influenced by a multitude of factors (e.g., writing misplaced, elements malfunction electronics not gauged, quark oscillations dependent on temperature, voltage not at design value).

Towards answering the second question, it is important to realise that the regularised model is merely an approximation to physics.[10] As discussed above, the values of the regularised model's bare parameters have no significance, except that they should be tuned such that the model's predictions are in optimal agreement with physics. In this case all parameters are properly finite, and the above considerations apply. When we remove the regulator, the values of the parameters may change as we keep tuning them to maintain (and improve) agreement with physics. Whether or not the bare parameters will be infinite in the unregularised limit is of no concern as their values are not immediately measurable. The renormalised parameters merely serve as finite substitutes to allow tuning to physics in a situation where the bare parameters are infinite. However, just as their bare counterparts, their values are not significant.

Example. An example of renormalisation is the energy of the electromagnetic field of a point charge in classical electrodynamics. Due to the divergent field near the point source, the energy of the associated electromagnetic field configuration receives an infinite contribution from the neighbourhood of the point source. To improve this result we regularise the point charge by giving it a finite radius. Then the field energy is perfectly finite but diverges as the radius approaches zero. Next, we renormalise the vacuum energy to exactly counterbalance the (large but finite) contribution from the charge. We can do this, because there is no physical significance to the vacuum energy. Finally, we can send the radius to zero making sure to tune the vacuum energy such that the overall energy remains finite in the limit. To make better sense of this method we conclude with two remarks: A point charge is an idealisation, in reality all charges will inevitably have some finite radius and all results will be finite. Even if this argument would not apply to the elementary particles such as the electrons and quarks, there is still the fact that the vacuum energy is not measurable and it should not worry us to find divergences in unphysical constants or results. In this case, only the difference of energies, where the vacuum energy cancels out, can be relevant.

Counterterms. Consider the dimensionally regulated massive scalar bubble integral

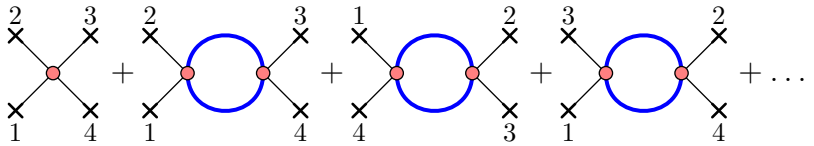
$$I = \ldots + \frac{1}{32\pi^2}\left(\frac{1}{\epsilon} - \gamma_{\mathrm{E}} + \log(4\pi) + 2 - \log\frac{m^2}{\mu^2}\right) + \mathcal{O}(\epsilon). \tag{4.25}$$

In a scalar field theory with quartic interactions it can arise as the one-loop correction to the quartic interaction. When cubic interactions are included it also

[10] Quite likely the original model is not even meant to be an exact description of physics, but should only serve as a good approximation at sufficiently low energy scales.

serves as a one-loop correction to the cubic interaction and to the kinetic term, specifically to the mass.

These three correction terms have a simple pole at $\epsilon = 0$. We can remove the divergences by adding counterterms for the quartic interaction, the cubic interaction and the mass. Specifically, let us consider the quartic coupling and its divergent one-loop corrections in the effective action[11]



(4.26)

which yields

$$-\mathring{\imath}\bar{\lambda} + \mathring{\imath}\bar{\lambda}^2 I(s) + \mathring{\imath}\bar{\lambda}^2 I(t) + \mathring{\imath}\bar{\lambda}^2 I(u) + \dots . \tag{4.27}$$

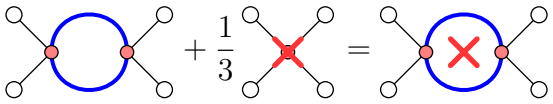
A suitable relationship between the renormalised quartic coupling constant λ and the bare coupling $\bar{\lambda} = \bar{\lambda}_{\text{MS}}$ used in the Lagrangian is called the minimal subtraction (MS)

$$\bar{\lambda}_{\text{MS}} = \lambda + \frac{3\lambda^2}{32\pi^2\epsilon} + \mathcal{O}(\lambda^3). \tag{4.28}$$

The relationship $\bar{\lambda} = \bar{\lambda}(\lambda)$ defines a new coupling constant $\lambda = \lambda(\bar{\lambda})$ which agrees with $\bar{\lambda}$ at leading order. The order-λ^2 contribution defines a counterterm that removes precisely the $1/\epsilon$ term in the above integral. Effectively, it changes the regularised integral to the subtracted integral $I_{\text{MS}} = I - 1/32\pi^2\epsilon$

$$I_{\text{MS}} = \dots + \frac{1}{32\pi^2}\left(-\gamma_{\text{E}} + \log(4\pi) + 2 - \log\frac{m^2}{\mu^2}\right) + \mathcal{O}(\epsilon). \tag{4.29}$$

In terms of diagrams the divergent loop has a counterterm that makes it perfectly finite



(4.30)

This result can be generalised to a method of constructing minimal counterterms proportional to pole term $1/\epsilon$. In the dimensionally regularised integrals this removes all corresponding poles $1/\epsilon$ and makes them finite. At one loop the minimal subtraction scheme can be conveniently implemented at the level of regularised integrals by simply discarding the $1/\epsilon$ terms.

Inspection of many cases makes obvious that each pole $1/\epsilon$ is accompanied by an Euler constant $-\gamma_{\text{E}}$ and a logarithm $\log(4\pi)$. Importantly they all arise in the universal combination $1/\epsilon - \gamma_{\text{E}} + \log(4\pi)$. It is therefore convenient to modify the scheme to that counterterms are proportional to this combination.

$$\bar{\lambda}_{\overline{\text{MS}}} = \lambda + 3\lambda^2 \frac{1/\epsilon - \gamma_{\text{E}} + \log(4\pi)}{32\pi^2} + \mathcal{O}(\lambda^3). \tag{4.31}$$

[11] All these terms are independent of the cubic coupling.

This scheme is called MS-bar, and it is the scheme commonly applied in conjunction with dimensional regularisation. The subtraction of the above integral reads

$$I_{\overline{\text{MS}}} = \ldots + \frac{1}{32\pi^2}\left(2 - \log\frac{m^2}{\mu^2}\right) + \mathcal{O}(\epsilon). \tag{4.32}$$

Finally, one could also choose to remove more terms of the integral such as the integer 2 or the term $\log(m^2/\mu^2)$. However, these terms are not universal in the sense that they may appear with different coefficients in different loop integrals. Note that the redefinition of the coupling constant must be independent of the momenta. In other words, merely shifts of the integrals are permissible, but the kinematical dependence (hidden in the above expressions) remains unchanged.

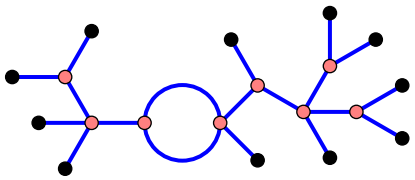
Another important observation is that the above choices of counterterms are related by a change of the renormalisation scale. For instance, using the renormalisation scale $\mu_{\overline{\text{MS}}}$ in the MS scheme with

$$\mu_{\overline{\text{MS}}} = \frac{\exp(\frac{1}{2}\gamma_{\text{E}})}{\sqrt{4\pi}}\,\mu_{\text{MS}} \tag{4.33}$$

is equivalent to the MS-bar scheme with renormalisation scale μ. Hence, there is no physical significance to the constant μ. Nevertheless, the operation of changing the renormalisation scale μ has important physical consequences as we shall discuss later.

Divergent Subgraphs. We have successfully renormalised a few divergent one-loop graphs. There are, however, more divergent graphs at this level, for which we have no more counterterms at our disposal. So the question is whether the earlier counterterms automatically cancel the divergences in a all other graphs.

The divergent graphs are in fact loops where the external legs connect to tree graphs.

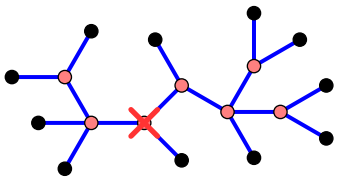


(4.34)

The loop itself is the same as above, hence we already understand its counterterm. One important difference is that the divergence is no longer local since the local divergence of the loop is connected to (non-local) trees.[12] This implies that no local counterterm could possibly compensate the divergence of the above graph. Nevertheless, this is not a problem since the counterterm can also act as a vertex

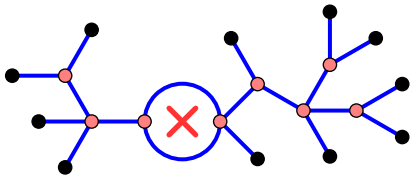
[12]Noting that $\int \mathrm{d}p\, \mathrm{e}^{ipx}\, p^n = 2\pi(-i)^n \partial^n \delta(x)$, locality in position space corresponds to polynomial dependence on the momenta in momentum space. Conversely, the Fourier transform of a non-polynomial function is necessarily non-zero away from the origin.

within a tree graph.



(4.35)

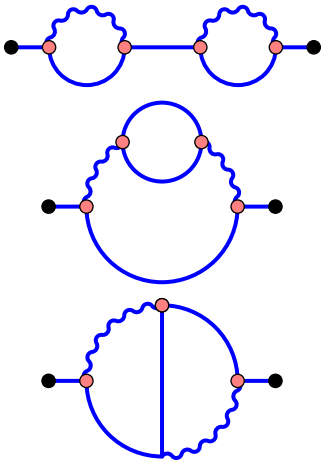
It cancels precisely the divergence of the loop embedded into a tree because the dressing by trees in fact does not make a difference.



(4.36)

One can convince oneself that the above argument works in general at one loop. A simpler way to show it is to consider the effective action G: It contains only the above three divergent plain loop graphs. Dressing a loop by trees certainly makes the graph one-particle reducible. Therefore the available three counterterms are sufficient to make G finite at one loop. Converting to the connected functional W does not introduce additional divergences, hence the one-loop renormalisation is successful.

Higher Loops. How about renormalisation at higher loops? The worry is that now each loop integral can be divergent on its own and create non-local divergences in connection with the other loops, e.g.



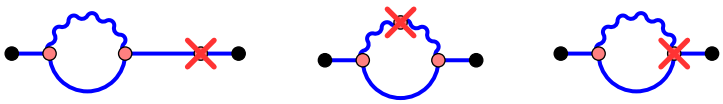
(4.37)

In particular we can distinguish three different situations:

- Disconnected loops: This is the easiest case, since the individual loops have a lower perturbative order. Even though the combined divergence is non-local, the corresponding counterterms at lower orders are sufficient to remove the divergence completely. This follows from finiteness of the effective action at lower orders.

- Nested loops: A divergent loop resides within another loop. The outer loop integral may appear convergent at first sight. However, when the inner loop is performed, the outer loop also becomes divergent.
- Overlapping loops: Two divergent loops share some propagators, but they are not nested. Since the two integrals cannot be disentangled, it is not immediately clear whether their divergences influence each other and whether they produce a non-local combined divergence.

All of the above loop integrals produce non-local divergences. Quite amazingly, none of these are severe: At higher loops the counterterms from lower orders now also appear inside loops:



(4.38)

For the same reasons as before, these cancel the divergences due to the individual loops. When all lower-order counterterms have been taken into account there is only a simple divergence left to be removed. Despite the fact that all subdivergences of the higher-loop graph are genuinely non-local, the residual divergence turns out to be perfectly local. It can be cancelled by a new counterterm.

Are these cancellations an outright miracle?

- An intuitive argument (with potential flaws) is that (UV)-divergences are local in position space. They arise from singular configurations in the integration domain where loops are contracted to a point.[13] Such a divergence can be compensated by a local counterterm at this point which has precisely the same non-local structure as the divergent part of the loop graph. The argument works for nested and overlapping loops as well because divergences still originate from contracted loops only.
- A formal argument for full consistency of QFT renormalisation by local counterterms was found by Bogoliubov, Parasiuk, Hepp and Zimmermann (BPHZ). This is too lengthy to present here. It is based on removing the divergences order-by-order as described above.

4.3 Renormalisation Flow

Physical results should not depend on the arbitrary choice of regularisation and renormalisation scheme. However, we have already seen that the parameters of quantum field theory models must be adjusted to the choice of renormalisation scheme in order to obtain the desired result.

Here we shall discuss the dependence of the parameters and the results on the renormalisation scheme. In particular, we have seen that the various schemes

[13] This argument works best in euclidean space where null distances do not exist and will not cause potential additional divergences.

introduce a renormalisation scale which consequently appears as a new parameter for all results. We shall therefore discuss the dependence on it, called the renormalisation (group) flow, along with its implications for the coupling constants.

Callan–Symanzik Equation. Let us consider a simple quantum field theory for a single massless field ϕ with a single dimensionless coupling constant λ and a regularisation scale μ.[14] For example, the latter can be a momentum cutoff scale or the scale parameter of dimensional regularisation. We shall be interested in n-point correlation functions G_n in momentum space which depend on the momenta p_k, as well as λ and μ

$$G_n(p_k, \lambda, \mu). \tag{4.39}$$

All of the above quantities are considered after regularisation, renormalisation and subsequent removal of the regulator; in particular, they are finite. The regularisation scale μ is a remainder from the regularisation procedure, and we would like to understand how quantities depend on it.

Suppose we change the renormalisation scale by some amount. Then we should be able to compensate the change of any physical variable by a change of the coupling constant and the normalisation of the n external fields[15]

$$\frac{\partial G_n}{\partial \mu} + \frac{\partial \lambda}{\partial \mu}\frac{\partial G_n}{\partial \lambda} - \frac{n}{\phi}\frac{\partial \phi}{\partial \mu} G_n = 0. \tag{4.40}$$

It is called the Callan–Symanzik equation. It completely determines the dependence on the renormalisation scale μ and thus makes it an irrelevant parameter since every physical quantity can be computed at arbitrary μ.

The Callan–Symanzik equation has the general form

$$\left(\mu \frac{\partial}{\partial \mu} + \beta(\lambda)\frac{\partial}{\partial \lambda} + n\gamma(\lambda)\right) G_n(p_k, \lambda, \mu) = 0, \tag{4.41}$$

where we have introduced the conventional functions $\beta(\lambda)$ and $\gamma(\lambda)$

$$\beta(\lambda) = \mu \frac{\partial \lambda}{\partial \mu}, \qquad \gamma(\lambda) = -\frac{\mu}{\phi}\frac{\partial \phi}{\partial \mu}. \tag{4.42}$$

These dimensionless functions are universally defined for the QFT model at hand independently of the particular observable one may be interested in. Therefore they cannot depend on the external momenta, and since they are dimensionless they must depend only on λ which is the only universal dimensionless parameter of the model.

[14] The following discussion applies just as well for several fields with several coupling constants. Masses and dimensionful coupling constants are also welcome, but require a generalised discussion in terms of conformal properties.

[15] This follows, for instance, from performing the rescaling in the path integral formulation of the generating functional.

A useful alternative form of the Callan–Symanzik equation follows by a common rescaling of all the momenta, fields and the renormalisation scale. Such a rescaling is trivial as there are no other scales involved in the calculation

$$\Big(\sum\nolimits_k p_k \frac{\partial}{\partial p_k} - nd + \mu \frac{\partial}{\partial \mu}\Big) G_n = 0. \tag{4.43}$$

Here d represents the canonical mass dimension of the field $\phi(p)$. Combining this relationship with the Callan–Symanzik equation yields the equation

$$\Big(\sum\nolimits_k p_k \frac{\partial}{\partial p_k} - \beta(\lambda) \frac{\partial}{\partial \lambda} - n\big(d + \gamma(\lambda)\big)\Big) G_n = 0. \tag{4.44}$$

This form is useful since it tells that changing the scale of all external momenta is equivalent to

- a change of the coupling strength λ together with a
- rescaling the external fields according to a deformed mass dimension $d + \gamma(\lambda)$.

This has some crucial physical consequences:

- The strength of the coupling constant depends on the scale of the external momenta. This effect is called a *running coupling*, and it is governed by the beta-function $\beta(\lambda)$.
- The mass dimension of the field ϕ effectively departs from its canonical value d to $d + \gamma(\lambda)$ when interactions are turned on. The coefficient $\gamma(\lambda)$ is called the anomalous dimension of the field ϕ. It is governed by renormalisation of the kinetic term, also known as wave function renormalisation.
- The running of the coupling constant can also be interpreted as an anomalous dimension of the coupling constant. The coupling constant develops an effective mass dimension once interactions are turned on. This effect is called *dimensional transmutation*. It follows from the Callan–Symanzik equation which essentially relates the coupling constant to the renormalisation scale, and therefore allows us to remove the coupling constant λ in favour of the scale μ.
- The anomalous dimensions of the fields and the coupling constants are directly related to the UV-divergences in QFT: They would never arise in a naive treatment of QFT without regularisation and renormalisation. One could argue that, since anomalous dimensions are indeed physical phenomena, a naive treatment of QFT ends up producing infinities in response to this conflict.

Running Coupling. The Callan–Symanzik equation is particularly strong and easiest to interpret for a two-point function G_2 which depends on a *single* momentum scale p only. Equivalently, we can consider a higher-point function G for a particular momentum configuration which depends dominantly on a single overall momentum scale p. The Callan–Symanzik equation thus reads

$$\Big(p \frac{\partial}{\partial p} - \beta(\lambda) \frac{\partial}{\partial \lambda} - nd - n\gamma(\lambda)\Big) G(p, \lambda, \mu) = 0. \tag{4.45}$$

Let us introduce the running coupling function $\tilde{\lambda}(z,\lambda)$ with $z = p/\mu$ which is defined by a differential equation and an initial condition

$$\left(p\,\frac{\partial}{\partial p} - \beta(\lambda)\,\frac{\partial}{\partial \lambda}\right)\tilde{\lambda}(p/\mu,\lambda) = 0, \qquad \tilde{\lambda}(1,\lambda) = \lambda. \tag{4.46}$$

Furthermore, we make a dimensionless ansatz for the function G based on the dimensionality of its parameters and its result

$$G(p,\lambda,\mu) = p^{nd}\tilde{G}\big(p/\mu, \tilde{\lambda}(p/\mu,\lambda)\big). \tag{4.47}$$

By construction, the function $\tilde{G}$ is a dimensionless function of dimensionless parameters. It simplifies the Callan–Symanzik equation as follows

$$\left(z\,\frac{\partial}{\partial z} - n\gamma\big(\lambda(z,\tilde{\lambda})\big)\right)\tilde{G}(z,\tilde{\lambda}) = 0, \tag{4.48}$$

where $\lambda(z,\tilde{\lambda})$ is the inverse function of $\tilde{\lambda}(z,\lambda)$. This equation is now easy enough to solve as

$$\tilde{G}(z,\tilde{\lambda}) = \exp\left(\int_1^z \frac{\mathrm{d}z'}{z'}\, n\gamma\big(\lambda(z',\tilde{\lambda})\big)\right)\tilde{F}(\tilde{\lambda}) \tag{4.49}$$

with a function $\tilde{F}(\tilde{\lambda})$ to be determined by the dynamics of the QFT.

The running coupling constant $\tilde{\lambda}$ depends on the (relative) momentum scale p/μ. However, since several particles are involved in each interaction, which momentum does p refer to? The momentum dependence can be interpreted in several ways:

- The elementary interaction of QED involves a single photon line. One could argue that p is an (approximate) measure of the virtuality of this photon. For example, consider $W^{\pm}$ pair production at a particle collider with intermediate photons. Its virtuality (centre-of-mass energy) in the electroweak regime (100 GeV) determines the effective coupling strength of the electromagnetic field. Here the effective fine structure constant is significantly different than for everyday electromagnetic processes (including production of electron-positron pairs).
  In non-abelian gauge theories, the association of energy to a specific particle is not as straight-forward since the cubic gauge vertex makes no distinction between the three contributing gluons.
- More accurately, the above equations hold when all the momenta involved in the process are rescaled by exactly the same amount. Here the equations predict the shift of the coupling due to the scaling factor.
- To understand the precise dependence on all involved momenta, one must calculate the actual loop integrals.

Dimensional Regularisation. Let us now show how regularisation and renormalisation of divergences leads to non-trivial coefficient functions $\beta(\lambda)$ and $\gamma(\lambda)$ in the Callan–Symanzik equation.

We start with some correlation function which we compute in the dimensional regularisation scheme. The bare result $\bar{G}$ in terms of the bare coupling constant $\bar{\lambda}$ is expressed as $\bar{G}(p_k, \bar{\lambda}, \mu, \epsilon)$.

We know that in dimensional regularisation the renormalisation scale μ appears in a very predictable fashion: Up to some potential rescalings it only appears in the combination $\bar{\lambda}\mu^{2\epsilon}$. This combination has a non-trivial mass dimension which we should compensate by some appropriate powers of the momenta p_k. Moreover the overall mass dimension of $\bar{G}$ is nd. We can thus write the most general result in the following form[16]

$$\bar{G}(p, \bar{\lambda}, \mu, \epsilon) = p^{nd} \bar{F}\big((\mu/p)^{2\epsilon}\bar{\lambda}, \epsilon\big). \tag{4.50}$$

It is straight-forward to confirm that this function obeys the Callan–Symanzik equation with the bare parameter functions

$$\bar{\beta}(\bar{\lambda}) = -2\epsilon\,\bar{\lambda}, \qquad \bar{\gamma}(\bar{\lambda}) = 0. \tag{4.51}$$

The non-zero beta-function tells us that the coupling constant has an effective mass dimension which tends to zero when the regulator ϵ is removed.

The crucial point is that we cannot yet remove the regulator since the bare result $\bar{G}$ typically has a pole at $\epsilon = 0$. This calls for renormalisation which takes the generic form

$$\bar{G}_n(p_k, \bar{\lambda}, \mu, \epsilon) = N(\bar{\lambda})^n G_n(p_k, \lambda(\bar{\lambda}), \mu, \epsilon). \tag{4.52}$$

The renormalised result G_n is now finite with divergences hidden in the field renormalisation $N(\bar{\lambda})$ and coupling constant renormalisation $\lambda(\bar{\lambda})$.

Assuming that the Callan–Symanzik equation holds in terms of the bare variables, it also holds for the renormalised variables with the renormalised functions given by the transformation

$$\begin{aligned} \beta\big(\lambda(\bar{\lambda})\big) &= \bar{\beta}(\bar{\lambda})\,\frac{\partial\lambda}{\partial\bar{\lambda}}\,(\bar{\lambda}), \\ \gamma\big(\lambda(\bar{\lambda})\big) &= \bar{\gamma}(\bar{\lambda}) + \frac{\bar{\beta}(\bar{\lambda})}{N(\bar{\lambda})}\,\frac{\partial N}{\partial\bar{\lambda}}\,(\bar{\lambda}). \end{aligned} \tag{4.53}$$

Renormalisation therefore preserves the Callan–Symanzik equation with some set of deformed parameters.

Suppose in dimensional regularisation the renormalisation functions take the following generic form with poles at $\epsilon = 0$

$$\begin{aligned} \lambda &= \bar{\lambda} + \bar{\lambda}^2(b_1/\epsilon + b_2 + \ldots) + \ldots, \\ \bar{\lambda} &= \lambda - \lambda^2(b_1/\epsilon + b_2 + \ldots) + \ldots, \\ N(\bar{\lambda}) &= 1 + \bar{\lambda}(c_1/\epsilon + c_2 + \ldots) + \ldots. \end{aligned} \tag{4.54}$$

[16]For simplicity we shall not distinguish between the individual momenta p_k and simply refer to them by p. Additional individual momenta can be interpreted as additional scalar parameters which play no role for our considerations.

Substituting these expansions into the above expressions we obtain perfectly regular expressions for β and γ at $\epsilon \to 0$

$$\begin{aligned} \beta(\lambda) &= -2\epsilon\lambda - 2\lambda^2(b_1 + b_2\epsilon + \ldots) + \ldots, \\ \gamma(\lambda) &= -2\lambda(c_1 + c_2\epsilon + \ldots) + \ldots. \end{aligned} \tag{4.55}$$

Taking the limit $\epsilon \to 0$, we find the renormalised expressions

$$\beta(\lambda) \to -2b_1\lambda^2 + \ldots, \qquad \gamma(\lambda) \to -2c_1\lambda + \ldots. \tag{4.56}$$

Curiously, they are non-vanishing, so the renormalisation procedure has in fact achieved something non-trivial. This was only possible because of the divergences in the bare correlators. So the divergences are not just an inconvenient feature of QFT, they actually are related to non-trivial behaviour of the quantum model such as running coupling constants.

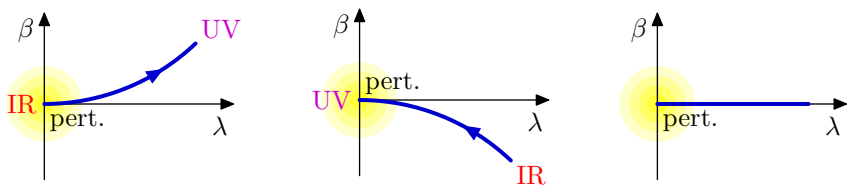
We should mention that the above functions β and γ are not uniquely defined: For example, we are free to perform a further finite perturbative redefinition of the coupling $\lambda \to \lambda' = \lambda + \mathcal{O}(\lambda^2)$. This leaves the classical model unchanged, but it modifies the higher-loop contributions to β and γ.

Fixed Points of the Flow. Here we discuss the qualitative behaviour of the QFT model depending on its beta-function $\beta(\lambda)$. As in our above examples, we assume that $\beta(0) = 0$ and we can compute the first few correction terms in λ by means of Feynman diagrams.

Changing the scale μ changes the bare parameters of the QFT model. This effect is called renormalisation flow. It is relevant since dimensional analysis teaches us that a shift of μ is equivalent to an opposite shift of momenta. The renormalisation flow therefore tells us how the QFT model behaves when we go to higher or lower energy scales.

A point λ_* with $\beta(\lambda_*) = 0$ is called a fixed point of the renormalisation flow. Changing the scale μ does not change the strength of the effective interactions. The best known example is the free theory at the trivial fixed point $\lambda_* = 0$. Since there is no coupling constant in the free theory, there is no flow. The free theory (without masses) is typically conformal.

As we go to a small but finite coupling λ there are three cases to be distinguished.



(4.57)

- $\beta(\lambda) > 0$: The coupling constant decreases as we go to small momenta or large distances. In the infrared region, the theory is weakly coupled and therefore our perturbative analysis is well justified here. Conversely in the ultraviolet region of small distances and large momenta, the effective coupling constant grows.

Eventually, the coupling may become too large, so that we cannot trust our perturbative analysis anymore. This situation is called an *IR-stable fixed point*. An example for such a model is QED[17]

$$\beta_{\mathrm{QED}}(q) = \frac{q^3}{12\pi^2} + \mathcal{O}(q^5). \tag{4.58}$$

The screening of electric charges makes the coupling weaker for long distances but much stronger for short distances.

- $\beta(\lambda) < 0$: This corresponds to a UV-stable fixed point which is the opposite of the above case. At very short distances or large momenta the coupling becomes weak, and this is where perturbation theory can be trusted. Conversely, at long distances, perturbation theory does not provide a good description of the model.
  The key example of a *UV-stable fixed point* is QCD, or more generally Yang–Mills theory. For gauge group $\mathrm{SU}(N)$ and n_{f} flavours of quarks, its perturbative beta-function takes the form

$$\beta_{\mathrm{YM}}(g) = -\frac{g^3}{16\pi^2}\left(\frac{11}{3}\,N - \frac{4}{3}\,n_{\mathrm{f}}\right) + \mathcal{O}(g^5). \tag{4.59}$$

  The model is called *asymptotically free* (in the UV). Constituent quarks within hadrons can be treated well in perturbation theory when their scattering with other particles is hard. On the other hand, perturbative QCD breaks down at low energies. For instance, at the scale of hadron masses the strong interactions become very strong, and quarks are eternally bound into mesons and hadrons. This effect is closely related to confinement.
  Asymptotically free theories are somewhat restricted: Among the standard interacting QFT's in four dimensions, only Yang–Mills theories can have negative beta-functions. Furthermore, all matter contributes with a positive sign to the beta-function, so only a certain amount of matter fields are admissible in asymptotically free theories.
- A very rare third option is $\beta(\lambda) = 0$ for a range of λ's. In this case a certain class of divergent contributions is absent, and the theory is called *finite*.[18] This implies that the coupling constant does not depend on the scale, even though there are non-trivial interactions. We have a one-parameter family of conformal field theories with non-trivial interactions.
  The best-understood example is $\mathcal{N} = 4$ maximally supersymmetric Yang–Mills theory which has a host of exiting features next to its conformality.

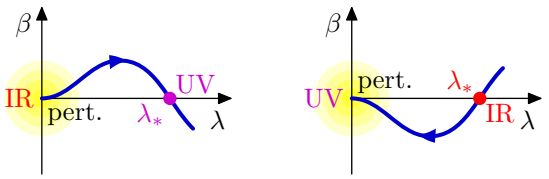
Now we would like to follow the renormalisation flow towards larger values of the coupling. We can do this only abstractly, since perturbation theory does not apply,

[17] The term on the right-hand side reflects the contribution from the electron field at very high energies. Other types of charged elementary particles equally contribute to the running of the electromagnetic coupling (unit charge). Interestingly, their contribution is activated only if the photon has sufficiently large virtuality to create a pair of these particles. Therefore, at energies around several MeV, the running of the electromagnetic coupling q is governed by the electron field alone. Below $2m_e$, there is no running at all.

[18] Note that there can still be infinite renormalisation of the fields.

but we can suppose we know $\beta(\lambda)$ exactly. We assume that $\beta(\lambda)$ returns to zero at some point λ_*.[19] This is a fixed point of the renormalisation flow. This point describes a conformal field theory with non-trivial interactions in general. Superficially this model may look rather different from the original perturbative model. For instance, the degrees of freedom may have effectively changed into something else. Also note that the value of λ_* may not be meaningful since it could be changed by a finite reparametrisation $\lambda \to \lambda'(\lambda)$.

Again we can distinguish two cases.



(4.60)

- $\beta(\lambda) > 0$ for $\lambda < \lambda_*$: The fixed point λ_* is called UV-stable. As the scale of momenta increases, the flow of the coupling constant eventually halts, and the model approaches a conformal model.
- $\beta(\lambda) < 0$ for $\lambda < \lambda_*$: This fixed point λ_* is called IR stable. As we go to larger distances, we approach a conformal model.

In these cases, the renormalisation flow interpolates between two conformal theories, one in the IR, the other in the UV. For correlation function this implies scaling laws for short and for long distances. The scaling laws are different since the models are distinct CFT's. One of them is a free field theory, the other one is typically a highly non-local field theory.

[19] Alternatively, $\beta(\lambda)$ could approach infinity which is not very interesting. A finite or vanishing asymptotic value of β could be viewed as an unfortunate choice of the coordinate λ.

5 Lie Algebra

We have introduced Yang–Mills theory in terms of $N \times N$ unitarity matrices for the gauge group, $N \times N$ hermitian matrices for the gauge potential and N-dimensional vectors for the spinor matter. The set of $N \times N$ unitary matrices embodies the Lie group $\mathrm{U}(N)$. A Lie group is a particular type of group whose elements form a smooth manifold (continuous set) rather than a discrete set. Lie groups play an important role in quantum field theory, where they also serve as spacetime and flavour symmetries. In this chapter we present some basic elements of Lie theory which allow us to formulate very general classes of Yang–Mills models by specifying their gauge symmetry and representations thereof.

5.1 Lie Groups

We start by discussing Lie groups, Lie algebras and their relationship.

Definition. A *Lie group* G is a group that is also a smooth manifold. The group multiplication $\mathrm{G} \times \mathrm{G} \to \mathrm{G}$ must be a smooth map.

Lie groups can be distinguished by several useful properties:

- they can be real or complex (as a manifold);
- they can be compact or non-compact (as a manifold);
- they can be simply connected, connected or disconnected (as a manifold);
- they can be finite- or infinite-dimensional (as a manifold);
- they can be abelian or non-abelian (as a group);
- they can be simple, semi-simple or composite (see below).

Example. The set of $N \times N$ unitary matrices defines a non-simple, real, compact, connected, finite-dimensional smooth manifold, and it is a group with a smooth composition rule. It is called $\mathrm{U}(N)$, which we used for introducing the gauge group in Yang–Mills theory.

Composition. Simple Lie groups serve as fundamental building blocks for more general Lie groups:

A *simple* Lie group is a connected non-abelian Lie group which has no non-trivial connected normal sub-groups. A subgroup H of G is *normal* if it is invariant under conjugation, i.e. $ghg^{-1} \in \mathrm{H}$ for all $h \in \mathrm{H}, g \in \mathrm{G}$. For normal subgroups the coset space G/H has a group structure, called the *quotient group*. Therefore, simplicity essentially means that the group cannot be reduced to a smaller group.

Composite Lie groups can be seen as direct or non-direct products of simple or abelian Lie groups. In particular, a *semi-simple* Lie group is a direct product of simple Lie groups, and many important results of Lie theory apply to this class.

Example. Consider four groups of matrices which take one of the following block forms:

$$\begin{array}{cccc} \begin{pmatrix} * \end{pmatrix} & \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} & \begin{pmatrix} * & \star \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & \star \\ 0 & 1 \end{pmatrix} \\ \text{simple} & \text{semi-simple} & \text{composite} & \text{abelian} \\ \mathrm{SU}(4) & \mathrm{SU}(2) \times \mathrm{SU}(2) & \text{Poincaré} & \text{translations} \end{array} \tag{5.1}$$

Suppose $*$ denotes (sub)matrices which from a simple Lie group and $\star$ denotes general matrices. The first group, e.g. SU(4), is simple by definition. The second, e.g. $\mathrm{SU}(2) \times \mathrm{SU}(2)$, is a direct product of simple groups, hence it is semi-simple. It is not simple because it has normal subgroups. The third group is not a direct product, hence it is not semi-simple. Furthermore it is not simple because it contains normal subgroups: One of them is composed from matrices where the upper-left submatrix $*$ is replaced by the unit matrix (fourth group). An example for groups of this form is the Poincaré group. It can be written as 5×5 matrices whose upper-left 4×4 submatrix $*$ belongs to the Lorentz group and the upper-right 4×1 submatrix $\star$ is a translation vector. The subgroup of translations is normal.

Example. The group $\mathrm{U}(N)$ is non-simple: A unitary matrix can be decomposed into an overall phase factor and a unimodular matrix (with unit determinant). The former is described by the (abelian) Lie group U(1) which is abelian, non-simply connected and (formally) non-simple. The latter matrices form the Lie group $\mathrm{SU}(N)$ which is non-abelian, simply connected and simple.[1]

In Yang–Mills theory we typically encounter direct products of compact, finite-dimensional, simple or abelian, real Lie groups. We will focus our attention to such groups; results presented below will not necessarily apply beyond.

5.2 Lie Algebras

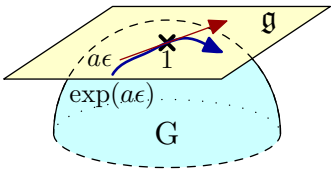
Lie groups are curved manifolds complicates their investigation and applications. Many aspects of Lie groups can be discussed in a linearised fashion in terms of vector spaces. The tangent space at the identity is equipped with a natural product; it is called a Lie algebra. In particular, the gauge potential $A_\mu(x)$ in Yang–Mills theory takes values in the Lie algebra associated to the gauge group.

Tangent Space at Identity. The unit element 1 of a Lie group G is a special point of the manifold. For instance, the product of two elements in a (sufficiently

[1] There is a small overlap between these two subgroups consisting of root-of-unity phase factors $\exp(2\pi i \mathbb{Z} N)$ which form the finite group $\mathbb{Z}_N$ describing integer addition modulo N. The group $\mathrm{SU}(N)/\mathbb{Z}_N$ is called $\mathrm{PSU}(N)$ and it is non-simply connected.

small) neighbourhood of 1 must again be in a (slightly larger) neighbourhood of 1 because multiplication in smooth. The neighbourhood of 1 can thus be understood as a region of G which is (partially) closed under multiplication. The multiplication of such elements can be understood in terms of the tangent space $\mathrm{T}_1\mathrm{G}$ of the Lie group G at the unit element 1. This tangent space is called the *Lie algebra* $\mathfrak{g}$ associated to G

$$\mathfrak{g} := \mathrm{T}_1\mathrm{G}. \tag{5.2}$$



How to define a suitable multiplication for the algebra $\mathfrak{g}$? We want it to reflect the multiplication in G, therefore we need a map between $\mathfrak{g}$ and G, at least in the neighbourhood of 1.

Exponential Map. Define a smooth map exp from a neighbourhood of 0 in $\mathfrak{g}$ to a neighbourhood of 1 in G [2] [3] such that[4]

$$\exp(0) = 1, \qquad \mathrm{d}\exp(0) = \mathrm{id}, \qquad \exp(na) = \exp(a)^n. \tag{5.3}$$

This map is called the *exponential map*. We can construct the exponential map as the limit[5]

$$\exp(a) = \exp(a/n)^n = \lim_{n\to\infty}\Big(1 + \frac{a}{n}\Big)^n. \tag{5.4}$$

Here the tangent space element $1 + a/n$ [6] approximates an element of the Lie group in the neighbourhood of 1, and raising it to some power is achieved by group multiplication. Evidently, the definition makes sense only for sufficiently large n, hence the limit.

Multiplication. The exponential map allows to pull back the group multiplication to the Lie algebra. Define a smooth map $m : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ (more precisely on some neighbourhoods of 0 in $\mathfrak{g}$) such that

$$\exp(a)\exp(b) =: \exp\big(m(a,b)\big). \tag{5.5}$$

Now it is clear that $m(a,0) = m(0,a) = a$ hence

$$m(a,b) = a + b + \mathcal{O}(a)\mathcal{O}(b). \tag{5.6}$$

[2] In order to understand some of the following abstract relations, it can make sense to consider G and $\mathfrak{g}$ to be sets of matrices which can be compared directly.

[3] It can be extended to the whole Lie algebra $\mathfrak{g}$ and the connected component G_0 of the Lie group G which includes the identity element 1.

[4] For a map f between two manifolds $f : A \to B$, its derivative $\mathrm{d}f$ at a point a is a linear map between the corresponding tangent spaces $\mathrm{d}f(a) : \mathrm{T}_a A \to \mathrm{T}_{f(a)}B$. In this case the derivative $\mathrm{d}\exp$ is defined as the identity map id on $\mathrm{T}_0\mathfrak{g} = \mathfrak{g} = \mathrm{T}_1\mathrm{G}$.

[5] This is commonly achieved by transport via a particular vector field.

[6] In this notation 1 serves as the base point in G and a/n describes a vector in the corresponding tangent space $\mathrm{T}_1\mathrm{G}$.

We see that the pull back is approximated by vector addition. Addition is a natural operation for vector spaces, and the result is evident from smoothness properties. Consequently, this composition law is not very interesting since it tells nothing about multiplication in the underlying Lie group.

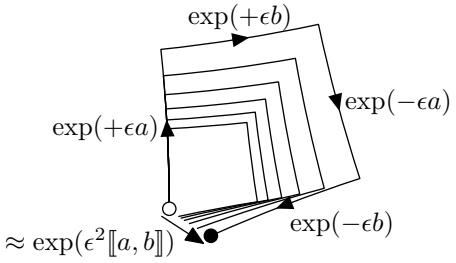
Let us therefore understand the deviation from linearity. We write explicitly the terms quadratic and of higher orders in a, b as

$$m(a,b) = a + b + m_2(a,b) + m_{\geq 3}(a,b). \tag{5.7}$$

From the above discussion and smoothness of the map $m(a,b)$ we know that m_2 must be bilinear. We furthermore know that $m(a,a) = 2a$. This implies that $m_2(a,a) = 0$. Together with bilinearity we conclude that m_2 is anti-symmetric

$$0 = m_2(a+b, a+b) - m_2(a,a) - m_2(b,b) = m_2(a,b) + m_2(b,a). \tag{5.8}$$

Lie Brackets. The leading deviation $2m_2(a,b)$ from linearity is called the *Lie bracket* $[\![a,b]\!]$.



(5.9)

We can thus compute it as the limit of a commutator of Lie group elements

$$[\![a,b]\!] = \lim_{\epsilon\to 0} \frac{1}{\epsilon^2} \exp^{-1}\big[\exp(\epsilon a)\exp(\epsilon b)\exp(-\epsilon a)\exp(-\epsilon b)\big]. \tag{5.10}$$

The Lie bracket has the following properties:

- It is bilinear.
- It is anti-symmetric
$$[\![a,b]\!] = -[\![b,a]\!]. \tag{5.11}$$
- It satisfies the Jacobi identity[7]
$$\big[\![[\![a,b]\!],c]\!\big] + \big[\![[\![b,c]\!],a]\!\big] + \big[\![[\![c,a]\!],b]\!\big] = 0. \tag{5.12}$$

The latter property follows from the Hall–Witt identity for three group elements x, y, z and their inverses $\bar{x}, \bar{y}, \bar{z}$

$$\begin{aligned} 1 = {} & \bar{y}\big((y\bar{x}\bar{y}x)\bar{z}(\bar{x}yx\bar{y})z\big)y \\ & \cdot \bar{z}\big((z\bar{y}\bar{z}y)\bar{x}(\bar{y}zy\bar{z})x\big)z \\ & \cdot \bar{x}\big((x\bar{z}\bar{x}z)\bar{y}(\bar{z}xz\bar{x})y\big)x. \end{aligned} \tag{5.13}$$

The leading order (cubic) contribution of each line equals one of the three terms of the Jacobi identity. Note that the conjugation of the double commutator does not affect the leading order terms, but it is necessary to make the identity hold for generic groups upon dropping all brackets.

[7] In fact the Jacobi identity is an axiom that turns an anti-symmetric bilinear map into a proper Lie bracket. For an associative algebra the commutator $[A,B] := AB - BA$ defines Lie brackets for which the Jacobi identity is in fact an identity.

Lie Algebras. We can also turn the above derivation of Lie algebras around and use it to construct an associated Lie group: A vector space $\mathfrak{g}$ equipped with a Lie bracket satisfying the above properties is called a Lie algebra. The exponential map defines a (simply connected) Lie group G associated to the Lie algebra. To this end, we note that all higher orders of the multiplication map m on $\mathfrak{g}$ are uniquely determined in terms of the Lie bracket by means of the Baker–Campbell–Hausdorff identity.

A Lie algebra is anti-commutative and non-associative. In comparison, a Lie group is typically non-commutative and associative. The relationship between these properties of Lie groups and algebras is as follows:

- The Lie bracket of the Lie algebra is a measure for non-commutativity of the associated Lie group. Abelian Lie groups therefore have trivial associated Lie brackets.
- Associativity of Lie groups translates to the Jacobi identity of Lie algebras.

Classification. Finite-dimensional, complex, simple Lie groups have been classified: There are four infinite families $\mathfrak{a}_n$, $\mathfrak{b}_n$, $\mathfrak{c}_n$ and $\mathfrak{d}_n$. They are related to the three main classes of matrix algebras

type	matrix algebra symbol	matrix group name
$\mathfrak{a}_n$	$\mathfrak{sl}(n+1)$	special linear
$\mathfrak{b}_n$	$\mathfrak{so}(2n+1)$	orthogonal (odd)
$\mathfrak{c}_n$	$\mathfrak{sp}(2n)$	symplectic
$\mathfrak{d}_n$	$\mathfrak{so}(2n)$	orthogonal (even)

(5.14)

Here, the index of the algebras denotes their rank, a characteristic number for simple Lie algebras.

The matrix algebras have particular properties which are stable under taking commutators

- The commutator of two $N \times N$ matrices is traceless. Hence, traceless matrices form the special linear Lie algebra $\mathfrak{sl}(N)$

$$A \in \mathfrak{sl}(N): \qquad \operatorname{tr} A = 0. \tag{5.15}$$

- Anti-symmetric $N \times N$ matrices with respect to an invertible symmetric metric M (e.g. $M = 1$) close under taking commutators. Therefore they form the orthogonal Lie algebra $\mathfrak{so}(N)$

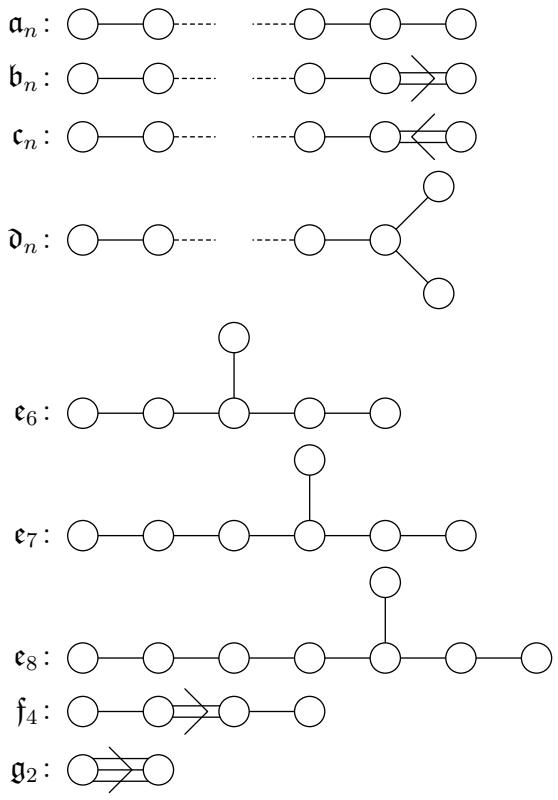
$$A \in \mathfrak{so}(N): \qquad A^{\mathsf{T}} = -MAM^{-1}, \qquad M^{\mathsf{T}} = M. \tag{5.16}$$

- Anti-symmetric $N \times N$ matrices with respect to an invertible anti-symmetric metric E close under taking commutators. They form the symplectic Lie algebra $\mathfrak{sp}(N)$

$$A \in \mathfrak{sp}(N): \qquad A^{\mathsf{T}} = -EAE^{-1}, \qquad E^{\mathsf{T}} = -E. \tag{5.17}$$

Furthermore, there are five exceptional simple Lie algebra $\mathfrak{g}_2$, $\mathfrak{f}_4$, $\mathfrak{e}_6$, $\mathfrak{e}_7$ and $\mathfrak{e}_8$ which are related to octonionic numbers.

Many properties of these simple Lie algebras can be encoded into so-called *Dynkin diagrams*.[8] A Dynkin diagram consists of a collection of dots connected by lines of various styles. The number of dots equals the rank of the algebra. The Dynkin diagrams of the simple finite-dimension Lie groups are as follows:



(5.18)

In analogy to simple Lie algebras, also the irreducible representations can be classified. Note that for a given Lie algebra there exists a particular set of irreducible representations which a characteristic spectrum of admissible dimensions.

A complex Lie algebra may have several corresponding real forms. For example, $\mathfrak{sl}(N,\mathbb{R})$ is a real form of the complex algebra $\mathfrak{sl}(N,\mathbb{C}) = \mathfrak{sl}(N) = \mathfrak{a}_{N-1}$ obtained by restricting all matrix elements to be real. The special unitary algebra $\mathfrak{su}(N)$ is another real form of it. It is spanned by anti-hermitian traceless $N \times N$ matrices. In our introduction to Yang–Mills theory, the gauge potential was a matrix of this form.[9] More generally, the gauge potential A_μ should take values in the Lie algebra $\mathfrak{g}$ associated to the gauge group G.

[8] This is merely a teaser. We will not discuss deeper structures of Lie theory in this course.

[9] In physics, one often introduces a conventional factor of i to convert to hermitian matrices (which have a real spectrum).

5.3 Representations

Matter fields in Yang–Mills theory have particular transformation properties under gauge transformations. Mathematically, their transformation rules are described by representations.

Definition. A representation R of a Lie group G is a map from the group to *automorphisms* of some vector space $\mathbb{V}$ (invertible operators, linear maps or matrices acting on the space)

$$R : \mathrm{G} \to \mathrm{Aut}(\mathbb{V}). \tag{5.19}$$

The representation must reflect the composition law of G by operator composition (matrix multiplication), i.e. for all $h, k \in \mathrm{G}$

$$R(hk) = R(h)R(k). \tag{5.20}$$

Consequently, it satisfies the group properties related to the unit element and inverse, namely

$$R(1) = \mathrm{id}, \qquad R(h^{-1}) = R(h)^{-1}. \tag{5.21}$$

The definition of a representation R of a Lie algebra $\mathfrak{g}$ is analogous. It is a *linear map* from the algebra to *endomorphisms* (linear operators) of some space $\mathbb{V}$

$$R : \mathfrak{g} \to \mathrm{End}(\mathbb{V}). \tag{5.22}$$

The representation must reflect the Lie bracket of $\mathfrak{g}$ as a commutator $[A, B] := AB - BA$ of operators, i.e. for all $a, b \in \mathfrak{g}$

$$R\big(\llbracket a, b \rrbracket\big) = [R(a), R(b)]. \tag{5.23}$$

The representation of Lie brackets in terms of commutators is compatible with their properties: The commutator is evidently bilinear, anti-symmetric, and it obeys the Jacobi identity.

By focusing on the unit element, it is clear that every representation of G has an associated representation of $\mathfrak{g}$. The converse does not exactly hold due to potential global issues of the group G. For example, consider the group $\mathrm{G} = \mathrm{U}(1) = \mathrm{S}^1$ of rotations of the complex plane. Its one-dimensional representations are given by the n-fold rotations $R_n(\varphi) := (\mathrm{e}^{\mathrm{i}n\varphi})$ which are labelled by the integer $n \in \mathbb{Z}$. The corresponding Lie algebra is the one-dimensional abelian algebra $\mathfrak{g} = \mathfrak{u}(1) = \mathbb{R}$, and its one-dimensional representations are given by $R_x(\mathrm{d}\varphi) := (x\,\mathrm{d}\varphi)$ with a real parameter $x \in \mathbb{R}$. Then for every R_n of G there is a R_x of $\mathfrak{g}$ with $x = n \in \mathbb{Z}$, but there is no corresponding representation of G for a R_x with real $x \neq \mathbb{Z}$. The reason for the discrepancy is that the group multiplication must respect the fact that rotations by 2π are trivial, while such topological issues are not observed in terms of a tangent space.

The operators $R(h) \in \mathrm{Aut}(\mathbb{V})$ and $R(a) \in \mathrm{End}(\mathbb{V})$ are called *representations of the elements* $h \in \mathrm{G}$ and $a \in \mathfrak{g}$, respectively.

The space $\mathbb{V}$ is called a *representation space* or a *module* of the group or algebra.[10] The dimension $\dim R$ of the representation R is defined as the dimension of $\mathbb{V}$

$$\dim R := \dim \mathbb{V}. \tag{5.24}$$

In Yang–Mills theory, fields belong to a representation space $\mathbb{V}$, while the representation R of the Lie group G describes large gauge transformations. The representation R of the Lie algebra not only describes infinitesimal gauge transformations, but also the coupling of fields to the gauge potentials.

Adjoint and Trivial Representations. A distinguished representation of any Lie algebra $\mathfrak{g}$ is the adjoint representation on the Lie algebra itself

$$\mathrm{ad} : \mathfrak{g} \to \mathrm{End}(\mathfrak{g}), \qquad \mathrm{ad}(a)b := [\![a, b]\!]. \tag{5.25}$$

The Jacobi identity ensures that ad is a proper representation since it guarantees that for any $a, b \in \mathfrak{g}$

$$\mathrm{ad}\big([\![a, b]\!]\big) = \big[\mathrm{ad}(a), \mathrm{ad}(b)\big]. \tag{5.26}$$

This becomes clear once applied to some $c \in \mathfrak{g}$ and using anti-symmetry of the Lie brackets

$$\begin{aligned} \mathrm{ad}\big([\![a, b]\!]\big)c &= \big[\![[\![a, b]\!], c]\!\big], \\ -\mathrm{ad}(a)\,\mathrm{ad}(b)c &= -\big[\![a, [\![b, c]\!]]\!\big] = \big[\![[\![b, c]\!], a]\!\big], \\ +\mathrm{ad}(b)\,\mathrm{ad}(a)c &= +\big[\![b, [\![a, c]\!]]\!\big] = \big[\![[\![c, a]\!], b]\!\big]. \end{aligned} \tag{5.27}$$

The vector gauge potentials in Yang–Mills theory always belong to the Lie algebra of the gauge group, and they transform in the adjoint representation.

The adjoint representation for the Lie group reads[11]

$$\mathrm{Ad} : \mathrm{G} \to \mathrm{Aut}(\mathfrak{g}), \qquad \mathrm{Ad}(h)b := hbh^{-1}. \tag{5.28}$$

Another representation which exists in any Lie algebra is the trivial one (typically defined on a one-dimensional space $\mathbb{V}$). For $a \in \mathfrak{g}$ and $h \in \mathrm{G}$

$$\mathrm{triv}(a) := 0, \qquad \mathrm{triv}(h) := \mathrm{id}\,. \tag{5.29}$$

In Yang–Mills theory, fields transforming in the trivial representation would not interact (directly) with the gauge fields.

Sums and Decompositions. Representations can be combined to form bigger representations. Conversely, representations can be decomposed to their building blocks.

[10] In physics the representation space is sometimes also sloppily referred to as a representation.

[11] Here the product $b' = hbh^{-1}$ of group and algebra elements can be understood in terms of the group product $h(1 + \epsilon b)h^{-1} = (1 + \epsilon b')$ with the middle factor in the neighbourhood of 1.

Suppose R_1, R_2 are two representations of a Lie group or algebra on the spaces $\mathbb{V}_1, \mathbb{V}_2$. The *direct sum* $R_{1\oplus 2}$ of two representations R_1, R_2 on the space $\mathbb{V}_{1\oplus 2} = \mathbb{V}_1 \oplus \mathbb{V}_2$ is defined as the block-diagonal matrix acting on the space $(\mathbb{V}_1, \mathbb{V}_2)$

$$R_{1\oplus 2} = \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix}. \tag{5.30}$$

Evidently, the dimension of the direct sum representation equals the sum of dimensions of the constituent representations.

$$\dim R_{1\oplus 2} = \dim R_1 + \dim R_2. \tag{5.31}$$

The direct sum operation can be reversed to decompose representations to their building blocks. There are three cases to be distinguished:

- A representation R is *decomposable* if it can be written as a direct sum of proper sub-representations $R = R_1 \oplus R_2$.
- A representation R is *irreducible* if it has no proper sub-representation, i.e. $\mathbb{V}$ has no proper subspace $\mathbb{V}_1$ on which the representation closes ($R\mathbb{V}_1 \subset \mathbb{V}_1$). Such a representation is also called *irrep* for short, and it can be considered an elementary building block for representations. An irrep is indecomposable by definition.
- There is a third class of representations which are *reducible* but *indecomposable*. These have proper sub-presentations, but cannot be written as a direct sum of such. Representations of this type can be written as upper-triangular block matrices.

Fortunately, the third case is not encountered for unitary representations which are most relevant to quantum mechanics, and in particular to Yang–Mills theory. For our purposes we may assume representations to be either irreducible or fully decomposable (to irreps).

One practical complication in decompositions is that one first has to identify appropriate subspaces $\mathbb{V}_1, \mathbb{V}_2$ of $\mathbb{V}$ such that the representation matrices take on a block-diagonal form. Two relevant concepts are conjugation and equivalence: A representation R_1 can be *conjugated* to a representation R_2 by applying one and the same similarity transformation $M : \mathbb{V}_1 \to \mathbb{V}_2$ for all $g \in \mathrm{G}$ or all $a \in \mathfrak{g}$

$$R_2 = M\, R_1\, M^{-1}. \tag{5.32}$$

Any two such representations are called *equivalent*. Classification of representations, particularly irreps, is normally considered modulo equivalence.

Tensor Products. Another main operation to combine representations is the tensor product. The *tensor product representation* $R_{1\otimes 2}$ of an element $h \in \mathrm{G}$ of the group on the space $\mathbb{V}_{1\otimes 2} = \mathbb{V}_1 \otimes \mathbb{V}_2$ is defined as

$$R_{1\otimes 2}(h) = R_1(h) \otimes R_2(h), \tag{5.33}$$

The definition for a Lie algebra element $a \in \mathfrak{g}$ is slightly different

$$R_{1\otimes 2}(a) = R_1(a) \otimes \mathrm{id} + \mathrm{id} \otimes R_2(a). \tag{5.34}$$

Evidently, both definitions of $R_{1\otimes 2}$ satisfy the axioms of representations provided that R_1, R_2 are proper representations. The dimension is given as the product of dimensions of the factors

$$\dim R_{1\otimes 2} = (\dim R_1) \cdot (\dim R_2). \tag{5.35}$$

Example. Tensor products of two or more identical representations R are special since their tensor product naturally decomposes into certain symmetric components. For example, for the tensor product $\mathbb{V} \otimes \mathbb{V}$ of two identical spaces $\mathbb{V}$ we can define a permutation operator $\mathrm{ex} \in \mathrm{End}(\mathbb{V} \otimes \mathbb{V})$

$$\mathrm{ex}(v_1 \otimes v_2) := v_2 \otimes v_1. \tag{5.36}$$

It can be used to construct two projectors $\mathit{\Pi}^{\pm}$ onto the symmetric and anti-symmetric subspaces $\mathbb{V}^{\pm}$ of $\mathbb{V} \otimes \mathbb{V}$

$$\mathit{\Pi}^{\pm} = \tfrac{1}{2}(\mathrm{id} \pm \mathrm{ex}), \tag{5.37}$$

where 'id' denotes the identity acting on $\mathbb{V} \otimes \mathbb{V}$. Now the definitions of the tensor product representation $R^{\otimes 2}$ commute with the permutation operator ex, and hence also with the projectors $\mathit{\Pi}^{\pm}$. Therefore it will act as a direct sum of representation $R^{\pm}$ on the subspaces $\mathbb{V}^{\pm}$

$$R^{\otimes 2} = R^{+} \oplus R^{-}, \qquad R^{\pm} = \mathit{\Pi}^{\pm} R^{\otimes 2}. \tag{5.38}$$

For more than two identical representations, tensor products will split into several components with particular permutation symmetries determined by representations of the discrete permutation group.[12]

Real Groups and Algebras. For complex Lie groups and algebras all representations are complex matrices. However, for real groups and algebras, there are further properties to characterise (irreducible) representations:

- A representation can be real, complex or quaternionic depending on the type of the matrix entries.[13] [14]
- Furthermore, a representation of a real Lie group (algebra) is said to be unitary (anti-hermitian), orthogonal (anti-symmetric) or symplectic (anti-symmetric) when all its matrices have this property.

[12] A useful tool in this respect are Young tableaux or diagrams, see below.

[13] Evidently, one can rewrite complex numbers as matrices over the real numbers, but also interpret real numbers as complex numbers without trivial imaginary part, and correspondingly for quaternions. Typically one would choose the largest number field which makes the matrices as small as possible and then the smallest number field which captures all matrix elements.

[14] The $\mathfrak{su}(2)$ representations with non-integer spin j are in fact quaternionic and of quaternionic dimension $j + {}^1\!/_2$, while in physics they are often considered to be complex and of complex dimension $2j + 1$.

Unitary representations are evidently relevant to quantum physics. Unitary representations of finite-dimensional compact Lie groups are finite-dimensional. Unitary representations of finite-dimensional non-compact Lie groups must be infinite-dimensional.

5.4 Invariants

Lie algebras have several invariant objects which serve various important purposes. For instance the quadratic form provides a natural metric for the Lie group manifold. Consequently, it can be used to translate between vectors and covectors. In Yang–Mills theory, they are not only important for defining Feynman propagators as well as couplings between fields, but also relevant for consistency issues involving quantum anomalies of gauge symmetry.

Invariant Elements. Consider a simple (finite-dimensional) Lie algebra $\mathfrak{g}$. Up to rescaling it has a unique invariant symmetric bilinear form $K : \mathfrak{g} \times \mathfrak{g} \to \mathbb{K}$ known as the *Killing form* ($\mathbb{K} = \mathbb{R}, \mathbb{C}$ denotes the number field over which the Lie algebra $\mathfrak{g}$ is defined). Symmetry and invariance means that for all $a, b, c \in \mathfrak{g}$

$$K(a,b) = K(b,a), \qquad K\big([\![c,a]\!],b\big) + K\big(a,[\![c,b]\!]\big) = 0. \tag{5.39}$$

A form with this property can be constructed easily using some representation R

$$\operatorname{tr} R(a)R(b) = B^R K(a,b). \tag{5.40}$$

Symmetry and invariance follow from cyclicity of the trace. Due to uniqueness of K in a simple Lie algebra, all these forms must be equivalent up to a factor of proportionality B^R which depends on the particular representation R.[15] Note that it also depends on the normalisation of K as we shall discuss later.

The Killing form of a semi-simple Lie algebra is invertible. Therefore there is a corresponding invariant element $C_2 \in \mathfrak{g} \otimes \mathfrak{g}$ called the *Casimir invariant*. Thought of as a matrix, C_2 is the inverse of the Killing form $K \in \mathfrak{g}^* \otimes \mathfrak{g}^*$. It obeys the invariance property for all $a \in \mathfrak{g}$

$$\big(\operatorname{ad}(a) \otimes \mathrm{id} + \mathrm{id} \otimes \operatorname{ad}(a)\big)C_2 = 0. \tag{5.41}$$

For example, for $\mathfrak{g} = \mathfrak{su}(2)$ the Casimir invariant C_2 serves as the squared angular momentum operator.

The invariance property is particularly relevant for representations of $C_2 \in \mathfrak{g} \otimes \mathfrak{g}$ which one typically defines as[16]

$$R(C_2) := (R \mathbin{\dot\otimes} R)C_2. \tag{5.42}$$

[15] Typically, it is defined in terms of the adjoint representation $R = \operatorname{ad}$ with the constant $B^{\mathrm{ad}} = 1$.

[16] The symbol $\dot\otimes$ here means that the a representations R is applied to each tensor factor of C_2 and the resulting two matrices are joined by matrix multiplication. The meaning of this abstract definition may become clearer later when it is expanded in a basis of $\mathfrak{g}$.

The above invariance condition then reads for all $a \in \mathfrak{g}$

$$\big[R(a), R(C_2)\big] = 0. \tag{5.43}$$

By a version of Schur's lemma for Lie algebras, this implies that an irreducible representation of C_2 is necessarily proportional to the identity operator

$$R(C_2) = C_2^R \,\mathrm{id}^R\,. \tag{5.44}$$

The eigenvalue C_2^R is a characteristic quantity of the irreducible representation R. For example, the constant C_2^j for a spin-j representation of $\mathfrak{su}(2)$ evidently is proportional to $j(j+1)$.

The factors $D^R := \dim R$, B^R and C_2^R (among others) can be used to characterise irreducible representations. They also appear as characteristic quantities in computations of Feynman diagrams. Let us therefore discuss them in more detail.

By taking the trace of the identity $R(C_2) = C_2^R \,\mathrm{id}^R$ on both sides we can establish a relationship between the constants B^R and C_2^R

$$C_2^R \,\mathrm{tr}\,\mathrm{id}^R = \mathrm{tr}\, R(C_2) = \mathrm{tr}(R \dot\otimes R)C_2 = B^R K(C_2). \tag{5.45}$$

The latter identity uses the above construction of the Killing form K. Then $K(C_2) = \mathrm{tr}\,\mathrm{id}^{\mathrm{ad}} = D^{\mathrm{ad}} = \dim \mathfrak{g}$ is the dimension of the algebra,[17] and for the left-hand side $\mathrm{tr}\,\mathrm{id}^R = D^R$ is the dimension of the representation R. We thus find the relation

$$D^{\mathrm{ad}} B^R = D^R C_2^R. \tag{5.46}$$

Therefore all the constants B^R can be determined in terms of the constants C_2^R. In particular, for the adjoint representation it implies the equality of the two constants

$$B^{\mathrm{ad}} = C_2^{\mathrm{ad}}. \tag{5.47}$$

Note that there is some arbitrariness in the definition of the Killing form and the above constants B^R and C_n^R: The common rescaling by factors of β

$$K(a,b) \to \beta K(a,b), \qquad B^R \to \beta^{-1} B^R, \qquad C_2^R \to \beta^{-1} C_2^R, \tag{5.48}$$

leaves all above relationships intact. This result shows that there is no universal definition for the B^R and C_2^R. However, fixing one of them, fixes all the others. A common normalisation of the Killing form is $K(a,b) := \mathrm{tr}\,\mathrm{ad}(a)\,\mathrm{ad}(b)$, the constant for the adjoint then equals $B^{\mathrm{ad}} = C_2^{\mathrm{ad}} = 1$. However, this may neither be a convenient normalisation nor the typical one used in physics in the case of $\mathfrak{su}(N)$. We therefore will not fix the normalisation, allowing to easily adjust to one's favourite normalisation.

For most Lie algebras, there are analogous totally symmetric invariant polynomials $C_n \in \mathfrak{g}^{\otimes n}$ of higher degree n. Which ones of these higher invariants exist depends on the Lie algebra. The quadratic invariant C_2 exists for every simple Lie algebra.

[17] Note that C_2 is the inverse of K and their contraction yields the dimension of the space $\mathfrak{g}$.

For example, $\mathfrak{su}(2)$ has only the quadratic invariant C_2, whereas $\mathfrak{su}(3)$ also has a cubic invariant C_3. Also the higher invariants C_n have eigenvalues C_n^R on irreducible representations R

$$R(C_n) = C_n^R \,\mathrm{id}^R, \tag{5.49}$$

thus offering a way to characterise irreps R in terms of the values C_n^R.

Finally, there is a curious relationship between the invariants of $\mathfrak{g}$ and the topology of the associated (compact, connected, universal covering) Lie group G: In terms of cohomology, the group manifold is equivalent to a product of odd-dimensional spheres

$$\mathrm{G} \simeq \prod_k \mathrm{S}^{2n_k-1}. \tag{5.50}$$

For every independent invariant of degree n_k there is a sphere of dimension $2n_k - 1$.[18] For example we have

$$\mathrm{SU}(2) \simeq \mathrm{S}^3, \qquad \mathrm{SU}(3) \simeq \mathrm{S}^3 \times \mathrm{S}^5 \tag{5.51}$$

in agreement with their spectra of Casimir invariants.

Structure Constants. For non-abelian gauge theories it is convenient to choose some basis T_a for the Lie algebra $\mathfrak{g}$. The elements T_a are called generators of $\mathfrak{g}$. In quantum physics, the basis is typically chosen to be imaginary

$$a = \mathring{\imath} a^c T_c \in \mathfrak{g}. \tag{5.52}$$

One the one hand, this choice usually makes the representation of T_a hermitian which is convenient because it guarantees real eigenvalues. On the other hand, it introduces several unnecessary factors of $\mathring{\imath}$. For instance, a group element h is typically parametrised via the exponential map as

$$h = \exp(\mathring{\imath} a^c T_c) \in \mathrm{G}. \tag{5.53}$$

Now that we have a basis, we can expand the Lie brackets in terms of the *structure constants* $f_{ab}{}^c$

$$[\![T_a, T_b]\!] = \mathring{\imath} f_{ab}{}^c T_c. \tag{5.54}$$

For a real algebra, the structure constants $f_{ab}{}^c$ must be real.

Furthermore we expand the invariant bilinear form in this basis as

$$K(T_a, T_b) = k_{ab}. \tag{5.55}$$

For a real, compact, simple Lie algebra, the matrix k_{ab} is positive-definite. By a suitable choice of (orthonormal) basis T_a we can adjust k_{ab} to be the unit matrix

[18] This well-known relationship for Lie groups has some relevance for chiral anomalies within quantised Yang–Mills theory.

$k_{ab} = \delta_{ab}$. It may be convenient to make this assumption, but we shall keep the notation k_{ab} general and introduce k^{ab} as its inverse.[19]

For semi-simple Lie algebras the matrix k_{ab} is non-degenerate. It can be used to raise and lower indices. For instance we can uniformise the indices of the structure constants

$$f_{abc} := f_{ab}{}^{d} k_{dc}. \tag{5.56}$$

These structure constants are totally anti-symmetric in all three indices

$$f_{abc} = -f_{bac} = -f_{acb}. \tag{5.57}$$

Anti-symmetry of the former two indices follows by definition. Anti-symmetry of the latter indices follows from invariance of K

$$0 = K\big([\![T_c, T_a]\!], T_b\big) + K\big(T_a, [\![T_c, T_b]\!]\big) = \mathring{\imath} f_{cab} + \mathring{\imath} f_{cba}. \tag{5.58}$$

Now we can also consider a representation T_a^R of the basis T_a, namely

$$T_a^R := R(T_a). \tag{5.59}$$

Evidently the matrices T_a^R obey the commutator identity

$$[T_a^R, T_b^R] = \mathring{\imath} f_{ab}{}^{c} T_c^R. \tag{5.60}$$

A special case is the adjoint representation

$$T_a^{\mathrm{ad}} T_b = \mathrm{ad}(T_a) T_b = [\![T_a, T_b]\!] = \mathring{\imath} f_{ab}{}^{c} T_c. \tag{5.61}$$

Therefore, the adjoint representation T_a^{ad} is a matrix whose elements are the structure constants

$$(T_a^{\mathrm{ad}})_b{}^{c} = \mathring{\imath} f_{ab}{}^{c}. \tag{5.62}$$

Let us briefly discuss the expansion of the Killing form and Casimir invariant in terms of the basis T_a of generators. For an irreducible representation R we know that the trace of two representation matrices yields the Killing form

$$\mathrm{tr}\big(T_a^R T_b^R\big) = B^R k_{ab}. \tag{5.63}$$

A representation of the quadratic Casimir invariant C_2 now takes the form

$$R(C_2) = k^{ab}\, T_a^R T_b^R. \tag{5.64}$$

For the adjoint representation we can write the above relation in terms of structure constants

$$-f_{ac}{}^{d} f_{bd}{}^{c} = B^{\mathrm{ad}} k_{ab} = C_2^{\mathrm{ad}} k_{ab}. \tag{5.65}$$

This relationship allows to compute the matrix k_{ab} from the structure constants.

[19] Note that the introduction of generators T_a expands the issues related to normalisation somewhat.

5.5 Yang–Mills Theory

We are now well-prepared to formulate Yang–Mills models for general gauge groups and matter field representations. We will assume a simple gauge group G; composite gauge groups can be achieved by combining Yang–Mills models for the components.

Gauge Field Sector. The fields A_μ, $F_{\mu\nu}$, C and $\tilde{C}$ of gauge-fixed Yang–Mills theory all have a similar form and transformation properties. In particular, they take values in the Lie algebra $\mathfrak{g}$ associated to the gauge group G. We expand them in an (imaginary) basis T^a for $\mathfrak{g}$:

$$\begin{aligned} A_\mu &= gT_a A^a_\mu, & C &= gT_a C^a, \\ F_{\mu\nu} &= gT_a F^a_{\mu\nu}, & \tilde{C} &= gT_a \tilde{C}^a. \end{aligned} \tag{5.66}$$

The field strength $F^a_{\mu\nu}$ is related to the gauge potential A^a_μ via the structure constants $f_{ab}{}^c$ for the basis T^a of $\mathfrak{g}$:

$$F^a_{\mu\nu} = (\partial_\mu A^a_\nu) - (\partial_\nu A^a_\mu) + g f_{bc}{}^a A^b_\mu A^c_\nu. \tag{5.67}$$

The Yang–Mills Lagrangian with a canonically normalised kinetic term is expressed using the inverse Killing form k_{ab}

$$\mathcal{L}_{\text{YM}} = -\tfrac{1}{4} k_{ab} F^{a,\mu\nu} F^b_{\mu\nu}. \tag{5.68}$$

This form of the Yang–Mills Lagrangian is universal. The choice of gauge algebra enters only via the structure constants, and the coupling constant g serves as their prefactor.

Matter Couplings. Matter fields must transform into themselves under gauge transformations. Earlier, we assumed the transformation of an N-dimensional vector of fields to be governed by the defining representation of $\mathfrak{su}(N)$. More generally, the fields must take values in some space $\mathbb{V}$ for a representation R of $\mathfrak{g}$.

The kinetic terms for matter fields rely on partial derivatives which must be accompanied by gauge potentials for properly covariant behaviour under gauge transformation. The covariant derivative D^R_μ in some representation R is given by

$$D^R_\mu := \partial_\mu - \mathring{\imath} R(A_\mu) = \partial_\mu - \mathring{\imath} g A^a_\mu T^R_a. \tag{5.69}$$

This covariant derivative acts on some field $X \in \mathbb{V}$ which we expand as $X = X^k v_k$ in some basis v_k of $\mathbb{V}$ according to

$$(D^R_\mu X)^k = (\partial_\mu X^k) - \mathring{\imath} g (T^R_a)^k{}_j A^a_\mu X^j. \tag{5.70}$$

We can now apply the covariant derivative to the construction of the kinetic term for a Dirac spinor field ψ in some representation R

$$\begin{aligned} \mathcal{L}_{\text{spinor}} &= \bar{\psi}\big(\gamma^\mu D^R_\mu - m\big)\psi \\ &= \bar{\psi}_k\big(\gamma^\mu (D^R_\mu \psi)^k - m\psi^k\big), \end{aligned} \tag{5.71}$$

Here, we have expanded the conjugate field $\bar{\psi} = \bar{\psi}_k v^k$ in the dual basis v^k of the dual space $\mathbb{V}^*$. Importantly, the representation must be unitary so that the derivative terms are properly real and normalised with a positive signature. Note that the representation R may or may not be reducible. For a reducible representation, unitarity ensures that the representation decomposes into a direct sum of irreducible representations. This case can thus be understood, equivalently, as a collection of spinor fields of (potentially different) irreducible representations of the gauge group G. In any case, this Lagrangian results in the following Feynman rules for the spinor field(s)

$$\begin{aligned}
&\alpha, j \xleftarrow{\ p\ } \beta, k \ \to \ \frac{-\mathring{\imath}\delta^j_k\big(-\mathring{\imath}(\gamma^\mu)^\alpha{}_\beta p_\mu + \delta^\alpha_\beta m\big)}{p^2 + m^2 - \mathring{\imath}\epsilon}\,, \\
&\begin{array}{l} p_1, \alpha, j \\ \qquad\quad\ p_3, \mu, a \\ p_2, \beta, k \end{array} \ \to\ g(T^R_a)^j{}_k(\gamma^\mu)^\alpha{}_\beta .
\end{aligned} \tag{5.72}$$

We may also couple a set of scalar fields $\phi \in \mathbb{V}$ to non-abelian gauge fields using the same covariant derivative in the corresponding representation R. For scalar fields, it makes sense to choose a real representation R and associated vector space $\mathbb{V}$ [20] [21]

$$\begin{aligned}
\mathcal{L}_{\text{scalar}} &= -\tfrac{1}{2}(D^{R,\mu}\phi)^\intercal (D^R_\mu \phi) - \tfrac{1}{2}m^2 \phi^\intercal \phi \\
&= -\tfrac{1}{2}(D^{R,\mu}\phi)^k (D^R_\mu \phi)^k - \tfrac{1}{2}m^2 \phi^k \phi^k .
\end{aligned} \tag{5.73}$$

Here we have assumed an orthonormal basis v_k of $\mathbb{V}$ with $v_j^\intercal v_k = \delta_{jk}$. The resulting Feynman rules now include an extra quartic interaction $\phi^2 A^2$

$$\begin{aligned}
&j \xrightarrow{\ p\ } k \ \to\ \frac{-\mathring{\imath}\delta_{jk}}{p^2 + m^2 - \mathring{\imath}\epsilon}\,, \\
&\begin{array}{l} p_1, j \\ \qquad\quad\ p_3, \mu, a \\ p_2, k \end{array} \ \to\ \mathring{\imath} g (T^R_a)_{jk}(p_1^\mu - p_2^\mu), \\
&\begin{array}{ll} p_1, j & \quad p_4, \nu, b \\ \qquad\quad & \\ p_2, k & \quad p_3, \mu, a \end{array} \ \to\ -\mathring{\imath} g^2 (T^R_a)_{jl}(T^R_b)_{lk}\eta^{\mu\nu} .
\end{aligned} \tag{5.74}$$

We point out that the coupling of (further) vector fields is severely constrained by renormalisability: On the one hand, massive vector fields generally lead to non-renormalisable models and it makes little sense to consider them in the QFT context. On the other hand, massless vector fields in some representation R effectively extend the gauge algebra $\mathfrak{g}$ to some non-simple Lie algebra $\mathfrak{g} \oplus \mathbb{V}$. Fields with higher spin or helicity have similar issues, and we can safely ignore these here.

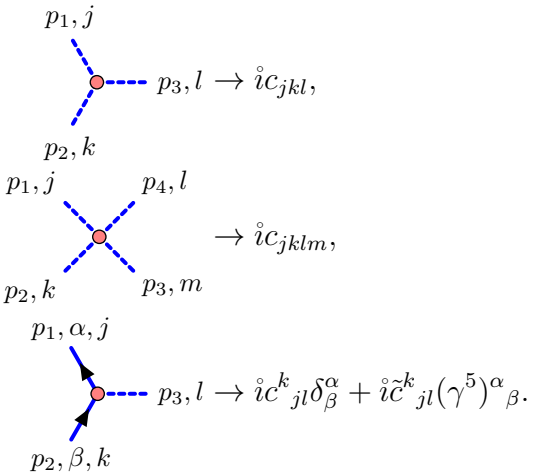
[20] A complex representation R can be realised as the direct sum of R and its conjugate $\bar{R}$.

[21] The Clifford algebra structure of spinor fields is complex and thus complex gauge representations can interact with it non-trivially.

Finally, we can discuss interactions between the fields. In renormalisable field theories in four dimensions, there are three types of interactions: cubic and quartic interactions of the scalars as well as Yukawa interactions between two spinors and a scalar[22] [23]

$$\begin{aligned}\mathcal{L}_{\text{cubic}} &= \tfrac{1}{6} c_{jkl}\, \phi^j \phi^k \phi^l, \\ \mathcal{L}_{\text{quartic}} &= \tfrac{1}{24} c_{jklm}\, \phi^j \phi^k \phi^l \phi^m, \\ \mathcal{L}_{\text{Yukawa}} &= c^k{}_{lj}\, \bar\psi_k \psi^l \phi^j + \tilde c^k{}_{lj}\, \bar\psi_k \gamma^5 \psi^l \phi^j .\end{aligned} \tag{5.75}$$

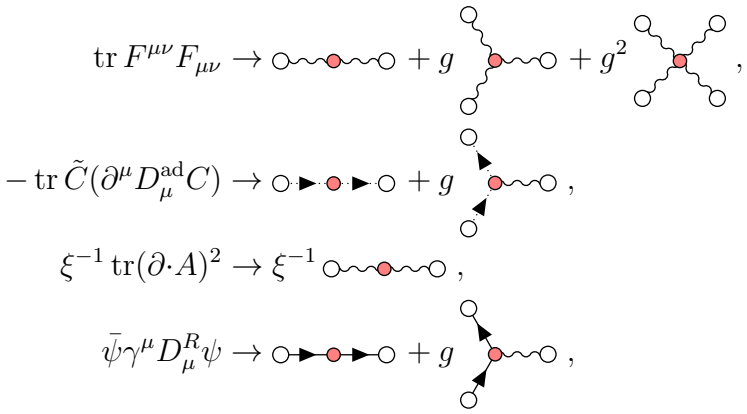
Gauge symmetry requires the terms to be invariant under gauge transformations. This implies that the constants c_{jkl}, c_{jklm} and $c^k{}_{lj}$, $\tilde c^k{}_{lj}$ are invariant tensors of the gauge group. The existence of such tensors crucially depends on the specific group and representations of the fields.[24]



(5.76)

Renormalisation. Let us finally discuss the renormalisation of Yang–Mills theory coupled to matter, in particular its consistency.

In non-abelian gauge theory we have various couplings between the gauge, ghost and matter fields

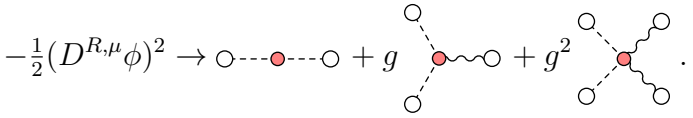


(5.77)

[22] Interaction terms which are not manifestly real must be accompanied by their complex conjugate.

[23] Other renormalisable interaction terms potentially involving covariant derivatives and field strengths can exist in spacetime dimensions other than $D = 4$.

[24] The tensor product of the corresponding representations must contain a trivial representation so that c does not transform under gauge transformations.



Furthermore, there are some pure matter couplings such as mass terms, cubic and quartic couplings of the scalars as well as Yukawa couplings which we will not consider here.

For gauge invariance it is crucial that the various interactions have a very specific form: For instance, every cubic vertex should contain a factor of g and some gauge invariant tensor relative to the kinetic term with one gauge leg removed. Likewise the quartic vertices have these factors squared.

All the above couplings have dimension 4 and therefore are expected to require renormalisation. On general grounds we should expect that all couplings renormalise independently. It is conceivable that each graph may have a counterterm with an individual coefficient. This would be troublesome for two reasons: At leading order all the gauge couplings were determined by a universal constant g and the discrete choice of representation. Renormalisation would introduce an individually tunable constant for each type of field, and thus reduce the predictive power of the model. Gauge invariance would break down which in turn would break renormalisability of the vector field. Does gauge invariance survive renormalisation?

Fortunately, it turns out that the above vertices renormalise coherently due to gauge symmetry. Although we had to break gauge symmetry to quantise the theory, there are two parts left: On the one hand, global transformations under the gauge group remain. This symmetry ensures that all vertices are governed by invariant tensors of the gauge group. Moreover, the relative coefficients are fixed by the Ward–Takahashi identity which is the quantum analogue of the conservation of Noether currents. Therefore, the coupling g acts as a universal renormalised gauge coupling constant. On the other hand, BRST symmetry emerges. This symmetry ensures that the ghosts and the unphysical polarisations of the vector field are equivalent up to signs which make their overall contribution cancel exactly.

The above 5 classes of terms lead to 6 independent renormalisation coefficients:

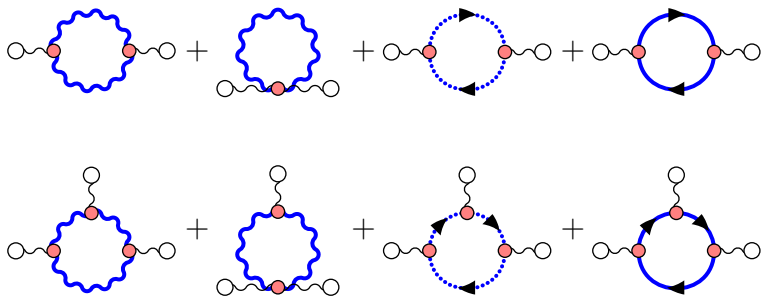
- one renormalisation coefficient for each class of terms and
- one overall renormalisation coefficient for the relative contributions to each term.

They are interpreted as follows:

- one coefficient to renormalise the kinetic term of each type of field, namely the gauge, ghost, spinor and scalar fields;
- one coefficient to (potentially) renormalise the constant ξ;
- one coefficient to renormalise the gauge coupling g.

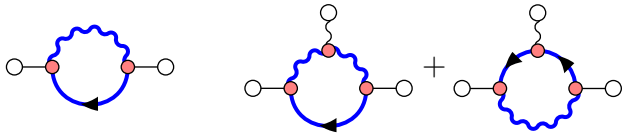
Note that the field renormalisation coefficients have no physical significance on their own. Nevertheless, they are important coefficients which contribute to each graph via the external legs.

One-Loop Renormalisation in Yang–Mills Theory. Let us discuss the renormalisation of the Yang–Mills coupling g. In order to extract it, we need to consider the kinetic term and the cubic coupling of gauge fields. For simplicity, we shall discard the contributions of scalar fields.



(5.78)

It takes some patience to evaluate these graphs since there are many contributions, and since the coupling of gauge fields has many terms. To simplify the calculation one can instead consider the kinetic term of some other field, e.g. the spinor or even better the ghost, along with its coupling to the gauge field. Note that the kinetic term for the gauge field is still relevant since it fixes the normalisation of the gauge field in the interaction vertex. The graphs with two external spinors (or equivalently ghosts) take the form:



(5.79)

We shall not evaluate the graphs in full detail, let us merely investigate their structure in terms of the gauge algebra.

When some matter fields in representation R form a loop with coupling to two external gluons, the resulting gauge algebra structure reads

$$\operatorname{tr} T_a^R T_b^R = B^R k_{ab}. \tag{5.80}$$

For three external gluons the structure is[25]

$$\operatorname{tr}[T_a^R, T_b^R] T_c^R = \mathring{\imath} f_{ab}{}^d \operatorname{tr} T_d^R T_c^R = \mathring{\imath} B^R f_{abc}. \tag{5.81}$$

We observe that both gauge algebra structures match the structures of the quadratic and cubic terms in the Yang–Mills action. Moreover, both coefficients are proportional to the constant B^R related to the representation of the matter fields.

An analogous result holds for loops made from gluons and ghosts. For two external gluons we find

$$f_{ac}{}^d f_{bd}{}^c = -B^{\mathrm{ad}} k_{ab}. \tag{5.82}$$

[25]The appearance of a commutator is not immediately obvious. It follows from a parity symmetry which reverses the order of external gluons. The absence of this parity symmetry for chiral spinors can lead to an undesirable effect called gauge anomaly which spoils gauge symmetry at loop level.

For three gluons the result reduces to an analogous combination[26]

$$\begin{aligned} f_{ad}{}^{e} f_{be}{}^{f} f_{cf}{}^{d} &= \tfrac{1}{2} f_{ad}{}^{e} f_{be}{}^{f} f_{cf}{}^{d} - \tfrac{1}{2} f_{db}{}^{e} f_{ae}{}^{f} f_{cf}{}^{d} - \tfrac{1}{2} f_{ba}{}^{e} f_{de}{}^{f} f_{cf}{}^{d} \\ &= -\tfrac{1}{2} f_{ab}{}^{e} f_{ed}{}^{f} f_{cf}{}^{d} = \tfrac{1}{2} B^{\mathrm{ad}} f_{abc}. \end{aligned} \tag{5.83}$$

We learn that the renormalisation of the gluon normalisation and of the gauge coupling can both be expressed as some linear combination of B^{ad} and B^R.

For completeness, let us discuss the corresponding processes for two external spinors in representation R of the gauge algebra (and equivalently for two ghosts in representation $R = \mathrm{ad}$). The correction to the kinetic term directly yields the quadratic Casimir eigenvalue in representation R

$$k^{ab} T_a^R T_b^R = C_2^R. \tag{5.84}$$

We can now add an external gluon in two inequivalent places. When it is attached to the matter line we obtain

$$\begin{aligned} k^{bc} T_b^R T_a^R T_c^R &= k^{bc} [T_b^R, T_a^R] T_c^R + k^{bc} T_a^R T_b^R T_c^R \\ &= \mathring{\imath} f_a{}^{dc} T_d^R T_c^R + k^{bc} T_a^R T_b^R T_c^R \\ &= -\tfrac{1}{2} f_a{}^{dc} f_{dc}{}^{e} T_e^R + C_2^R T_a^R \\ &= -\tfrac{1}{2} B^{\mathrm{ad}} T_a^R + C_2^R T_a^R \\ &= (C_2^R - \tfrac{1}{2} B^{\mathrm{ad}}) T_a^R. \end{aligned} \tag{5.85}$$

When it is attached to the gluon line instead we obtain the structure

$$f_a{}^{bc} T_b^R T_c^R = \tfrac{1}{2} f_a{}^{bc} [T_b^R, T_c^R] = \tfrac{\mathring{\imath}}{2} f_a{}^{bc} f_{bc}^{d} T_d^R = \tfrac{\mathring{\imath}}{2} B^{\mathrm{ad}} T_a^R. \tag{5.86}$$

Again the gauge theory structures come out as expected, but we obtain a different set of constants, namely C_2^R and B^{ad}.

How can the latter possibly yield the same renormalisation of the gauge coupling which we determined to be a linear combination of B^R and B^{ad}?[27] It turns out that C_2^R merely contributes to the spinor field normalisation. The dependence on B^R enters through the gluon normalisation.

Altogether both approaches to compute the gauge coupling renormalisation yield the same final result after the respective loop integrals are evaluated. It is usually expressed in terms of the so-called beta-function[28]

$$\beta(g) = -\frac{g^3}{16\pi^2} \left(\frac{11}{3} B^{\mathrm{ad}} - \frac{4}{3} \sum_{\substack{\text{Dirac} \\ \text{spinors}}} B^R - \frac{1}{6} \sum_{\substack{\text{real} \\ \text{scalars}}} B^R \right) + \mathcal{O}(g^5). \tag{5.87}$$

[26] First divide the term into two equal ones, and apply a Jacobi identity to one of them. After some shuffling of indice, two terms cancel and the third produces the final result.

[27] Note that C_2^R and B^R differ by constants depending on the gauge algebra and the representation size.

[28] We discussed above that B^R is subject to normalisation issues. Changing its normalisation also affects the structure constants f_{abc}. The latter always appear in conjunction with the coupling constant g in such a way that the result for β is consistent.

Here the first sum is over the representations R of Dirac spinor fields and the second sum over the representations R of real scalar fields.[29] Note that the different prefactors of the B's are characteristic numbers for type of particles.[30] In the special case of N_f flavours of Dirac fermions in a $\mathrm{SU}(N_\mathrm{c})$ gauge theory we recover the famous combination $\frac{11}{3}N_\mathrm{c} - \frac{4}{3}N_\mathrm{f}$.

5.6 Unitary Algebras

Among the simple Lie groups, the special unitary groups $\mathrm{SU}(N)$ have the simplest structure. Moreover, the groups with $N = 2$ and $N = 3$ along with the abelian group $\mathrm{U}(1)$ serve as the gauge groups of the standard model. Let us therefore discuss some features of the special unitary groups and their associated algebras.

Definition. A complex $N \times N$ matrix U has N^2 complex degrees of freedom or $2N^2$ real ones. Unitarity, $U^\dagger = U^{-1}$, constrains half of them, so that the real dimension of $\mathrm{U}(N)$ is N^2. Triviality of the determinant, $\det U = 1$, removes another real degree of freedom so that the dimension of $\mathrm{SU}(N)$ is $N^2 - 1$

$$\dim \mathrm{SU}(N) = N^2 - 1. \tag{5.88}$$

The corresponding Lie algebra $\mathfrak{su}(N)$ is the commutator algebra of traceless anti-hermitian matrices, $A = -A^\dagger$, $\mathrm{tr}\, A = 0$. Correspondingly, it has $N^2 - 1$ real degrees of freedom.

Defining Representations. Since the Lie group and algebra are both formulated in terms of matrices, we know that

$$\mathfrak{su}(N), \mathrm{SU}(N) \subset \mathrm{End}(\mathbb{C}^N). \tag{5.89}$$

Therefore it is straight-forward to construct an N-dimensional representation, the so-called *defining* or *fundamental representation*[31] (with $v \in \mathbb{C}^N$)[32]

$$\mathrm{def}(A)v = Av, \qquad \mathrm{def}(U)v = Uv, \qquad D^{\mathrm{def}} = N. \tag{5.90}$$

There is in fact another N-dimensional representation, the *conjugate defining representation*[33]

$$\overline{\mathrm{def}}(A)v = A^*v = -A^\intercal v, \qquad \overline{\mathrm{def}}(U)v = U^*v, \qquad D^{\overline{\mathrm{def}}} = N. \tag{5.91}$$

[29] Weyl spinors have half as many degrees of freedom and therefore are counted with half the prefactor while complex scalars evidently contribute with the double prefactor.

[30] For instance, it makes physical sense that a particle with spin has a stronger influence on $\beta(g)$ than a scalar particle.

[31] The notion of fundamental representation may include more than just the defining representation(s).

[32] The defining representation is complex and its complex dimension is $D^{\mathrm{def}} = N$.

[33] Note that there are some complications related to different number fields of representations in real algebras: For example, the direct sum $\mathrm{def} \oplus \overline{\mathrm{def}}$ of a complex representation def and its complex conjugate $\overline{\mathrm{def}}$ can be given a real structure; it thus has real dimension $D^{\mathrm{def}\oplus\overline{\mathrm{def}}} = 2N$. Put differently, the real representation $\mathrm{def} \oplus \overline{\mathrm{def}}$ is decomposable, but only over the complex numbers.

The adjoint representation is a representation on the (N^2-1)-dimensional space $\mathfrak{g}$ [34] [35]

$$\mathrm{ad}(A)B = [A,B], \qquad \mathrm{Ad}(U)B = UBU^{-1}, \qquad D^{\mathrm{ad}} = N^2 - 1. \tag{5.92}$$

So we know that the special unitary algebras have at least four (irreducible) representations: trivial, adjoint, defining and conjugate defining.

These representations suffice to formulate chromodynamics and the standard model: In Yang–Mills theories, the gauge fields transform in the adjoint representation and thus belong to the gauge algebra. The matter fields such as the quarks typically transform in the defining representations, and their conjugates in the conjugate defining representations. They may also transform in trivial representations in which case they do not couple to the corresponding gauge fields.

Tensor Product Representations. The unitary algebras have infinitely many finite-dimensional representations. They can be constructed from the above elementary representations by iteratively taking tensor products and decomposing them into irreps. Let us consider tensor products of the defining representations.

We have already seen that tensor products of identical representations are decomposable into various symmetric components. The tensor product of two original defining representations can thus be written as

$$\mathrm{def} \otimes \mathrm{def} = \mathrm{sym} \oplus \mathrm{alt}. \tag{5.93}$$

Here sym and alt denote the symmetric and anti-symmetric or alternating components of the full tensor product. They have dimensions $N(N+1)/2$ and $N(N-1)/2$ which sum up to the product $N \cdot N$. For the unitary algebras these representations turn out to be irreducible.

The same applies to the tensor product of two conjugate defining representations,

$$\overline{\mathrm{def}} \otimes \overline{\mathrm{def}} = \overline{\mathrm{sym}} \oplus \overline{\mathrm{alt}}. \tag{5.94}$$

Here $\overline{\mathrm{sym}}$ and $\overline{\mathrm{alt}}$ denote the conjugate representations of sym and alt, respectively.

Finally, let us consider the tensor product of mixed defining representations. Based on general considerations, the tensor product of two conjugate representations contains the adjoint and the trivial representation

$$\mathrm{def} \otimes \overline{\mathrm{def}} = \mathrm{ad} \oplus \mathrm{triv}. \tag{5.95}$$

In this case there are no further components due to matching of dimensions: $N \cdot N = (N^2-1) + 1$.

[34] It is sometimes tempting to confuse the defining with the adjoint representation since both can be formulated simply in terms of the $N \times N$ matrix A. The distinction becomes clearer when viewing the representation space $\mathfrak{g}$ as an (N^2-1)-dimensional vector space rather than a subspace of matrices.

[35] The adjoint representation is real and its real dimension is $D^{\mathrm{ad}} = N^2 - 1$.

Young Diagrams and Dynkin Labels. Finite-dimensional irreps of the unitary algebras $\mathfrak{su}(N)$ have been classified and their tensor products are well understood. Useful tools in this regard are Young diagrams[36] and Dynkin labels.[37]

Young diagrams are stackings of boxes. They describe certain symmetrisation classes or equivalently representations of the symmetric group S_K. A single box corresponds to the defining representation. The conjugate fundamental representation corresponds to a vertical stack of $N-1$ boxes. The adjoint representation adds a single box to the right of the topmost box. The trivial representation has no box which we denote by a dot. For example, for $N=5$

$$\text{def} = \square, \qquad \overline{\text{def}} = \begin{array}{|c|}\hline\\\hline\\\hline\\\hline\end{array}, \qquad \text{ad} = \begin{array}{|c|c|}\hline\\\hline\\\hline\\\hline\end{array}, \qquad \text{triv} = \bullet. \tag{5.96}$$

Horizontal stacks of boxes correspond to symmetrisation; vertical stacks correspond to anti-symmetrisation. For example, the above decomposition of the tensor products of two defining representations can be expressed as follows

$$\begin{aligned} \square \otimes \square &= \square\square \oplus \begin{array}{|c|}\hline\\\hline\end{array}, , \\ \begin{array}{|c|}\hline\\\hline\\\hline\end{array} \otimes \begin{array}{|c|}\hline\\\hline\\\hline\end{array} &= \begin{array}{|c|c|}\hline\\\hline\\\hline\end{array} \oplus \begin{array}{|c|}\hline\\\hline\end{array}, , \\ \square \otimes \begin{array}{|c|}\hline\\\hline\\\hline\end{array} &= \begin{array}{|c|c|}\hline\\\hline\\\hline\end{array} \oplus \bullet. \end{aligned} \tag{5.97}$$

Young diagrams are useful tools to work with representations of $\mathfrak{su}(N)$. For instance, there are rules for computing the dimension of irreps and rules for the decomposition of tensor products.

Dynkin labels are a convenient rewriting of the shape of Young diagram in terms of a set of non-negative integer numbers l_k. Suppose n_i $(i=1,\ldots,N)$ denotes the number of boxes in the i-th row of the Young diagram. Then the corresponding Dynkin labels are defined as

$$[l_1,\ldots,l_{N-1}] \qquad \text{with} \qquad l_k := n_k - n_{k+1}. \tag{5.98}$$

For example, the elementary representations for $\mathfrak{su}(N)$ are expressed in terms of $N-1$ Dynkin labels as

$$\begin{aligned} \text{triv} &= [00\ldots00], , \qquad \text{def} = [10\ldots00], , \\ \text{ad} &= [10\ldots01], , \qquad \overline{\text{def}} = [00\ldots01]. \end{aligned} \tag{5.99}$$

The Dynkin labels l_k are associated to the nodes k of the Dynkin diagram of the underlying algebra $\mathfrak{a}_{N-1}$

$$[l_1,l_2,\ldots,l_{N-1}] \equiv \overset{l_1}{\bigcirc}\!\!-\!\!\overset{l_2}{\bigcirc}\!\!-\cdots-\!\!\overset{l_{N-1}}{\bigcirc}\ . \tag{5.100}$$

In that sense, the Dynkin labels are a natural means of describing representations in more general algebras.

[36] Young diagrams are called Young tableaux when filled with particular numbers

[37] This topic is far beyond the scope of this course, but a very nice topic of group theory and combinatorics, so let us give a flavour.

Completeness Relations. The defining representation of $\mathfrak{su}(N)$ maps elements of $\mathfrak{g}$ to $N \times N$ traceless anti-hermitian matrices. The representation T_a^{def} of the basis T_a therefore consists of traceless *hermitian* matrices[38]

$$(T_a^{\text{def}})^\dagger = T_a^{\text{def}}, \qquad \operatorname{tr} T_a^{\text{def}} = 0. \tag{5.101}$$

The matrices T_a^{def} are used to translate between the two standard formulations of unitary gauge theories: One formulation assumes gauge fields to be traceless hermitian matrices A; the other one uses real component fields A^a in a basis of the Lie algebra. Consequently, the two are related by $A \sim T_a^{\text{def}} A^a$.

An important property of the representation matrices T_a^{def} is that they also form a basis of $N \times N$ matrices (over the complex numbers) when supplemented with the unit matrix id. The completeness relation reads

$$k^{ab} (T_a^{\text{def}})^i{}_j \, (T_b^{\text{def}})^k{}_l = B^{\text{def}} \left(\delta^i_l \delta^k_j - \frac{1}{N} \, \delta^i_j \delta^k_l \right). \tag{5.102}$$

As explained above, we have the freedom to set B^{def} to any desirable value and also fix the normalisation of k_{ab}. In an orthonormal basis with $k_{ab} = \delta_{ab}$, it is often declared to equal $B^{\text{def}} = {}^1\!/_2$. Some other useful and popular choices are $1/N$, 1 and $1/2N$.

The completeness relation can be rephrased in terms of two trace identities which hold for any $N \times N$ matrix X

$$\begin{aligned} k^{ab} \operatorname{tr}(T_a^{\text{def}} X) \, T_b^{\text{def}} &= B^{\text{def}} \big(X - N^{-1} \operatorname{tr} X \big), , \\ k^{ab} T_a^{\text{def}} X T_b^{\text{def}} &= B^{\text{def}} \big(\operatorname{tr} X - N^{-1} X \big). \end{aligned} \tag{5.103}$$

These identities are useful in simplifying various expressions that appear in gauge theory.

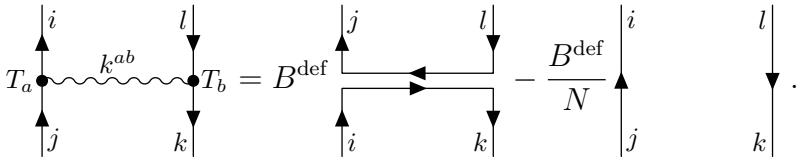
Example. We can express the structure constants as a trace

$$\operatorname{tr}\big([T_a^{\text{def}}, T_b^{\text{def}}] T_c^{\text{def}} \big) = \mathring{\imath} B^{\text{def}} f_{abc}. \tag{5.104}$$

For example, the contraction of two structure constants then reads

$$\begin{aligned} & -(B^{\text{def}})^2 k^{cf} f_{abc} f_{def} \\ &= k^{cf} \operatorname{tr}\big([T_a^{\text{def}}, T_b^{\text{def}}] T_c^{\text{def}} \big) \operatorname{tr}\big([T_d^{\text{def}}, T_e^{\text{def}}] T_f^{\text{def}} \big) \\ &= B^{\text{def}} \operatorname{tr}\big([T_a^{\text{def}}, T_b^{\text{def}}] [T_d^{\text{def}}, T_e^{\text{def}}] \big). \end{aligned} \tag{5.105}$$

The above completeness relation also has a useful graphical representation reminiscent of Feynman graphs and rules. Both identities can be written as



. (5.106)

[38] When $\mathfrak{su}(N)$ is viewed as a matrix algebra, it makes sense to identify T_a^{def} with T_a and drop the label def from all matrices T_a^{def} below for convenience.

Here a curly line depicts an algebra index and the directed straight line depicts the flow of the fundamental representation. Inserting a matrix X in two different places results in the above completeness relations which are thus equivalent.

Symmetric Structure Constants. A fact with useful consequences for the defining representation is that the anti-commutator of two hermitian matrices is again hermitian. Using the above completeness relations and after removing traces, this fact can be recast as a statement for the defining representation of $\mathfrak{su}(N)$

$$\{T_a^{\text{def}}, T_b^{\text{def}}\} = d_{ab}{}^c T_c^{\text{def}} + 2N^{-1} B^{\text{def}} k_{ab}. \tag{5.107}$$

This is interesting because it allows to define a set of totally symmetric cubic structure constants d_{abc}

$$\operatorname{tr}\big(\{T_a^{\text{def}}, T_b^{\text{def}}\} T_c^{\text{def}}\big) = B^{\text{def}} d_{abc}. \tag{5.108}$$

These exist only for the unitary algebras with $N > 2$. For the orthogonal and symplectic algebras this cubic combinations of representation matrices vanishes identically.[39]

The symmetric structure constants obey certain Jacobi-like identities among themselves and with the ordinary structure constants f_{abc}. They ensure that d_{abc} is an invariant tensor and can be used to define a cubic invariant

$$C_3 = d^{abc}\, T_a \otimes T_b \otimes T_c \in \mathfrak{su}(N)^{\otimes 3}. \tag{5.109}$$

Cases $\mathfrak{su}(2)$ and $\mathfrak{su}(3)$. The cases of $N = 2$ and $N = 3$ have some rather evident special features, and they are the Lie groups most relevant to the standard model.

For $\mathfrak{su}(2)$, the three hermitian traceless 2×2 Pauli matrices σ_a furnish a basis for the defining representation. In some normalisation we have

$$T_a^{\text{def}} = \tfrac{1}{2}\sigma_a. \tag{5.110}$$

The completeness relations for Pauli matrices take the form

$$\operatorname{tr}(\sigma_a X)\sigma_a = X - \tfrac{1}{2}\operatorname{tr}(X), \qquad \sigma_a X \sigma_a = -X + 2\operatorname{tr}(X). \tag{5.111}$$

A special feature of $\mathfrak{su}(2)$ is that the defining representation is equivalent to the conjugate defining representation, namely

$$T_a^{\overline{\text{def}}} = \sigma_2^{-1} T_a^{\text{def}} \sigma_2. \tag{5.112}$$

[39] The cubic invariant d_{abc} does not exist for orthogonal algebras, except for the case $\mathfrak{so}(6)$ which is isomorphic to $\mathfrak{su}(4)$. There, it needs to be defined via the spinor rather than the vector representation.

More concretely we have the following relations among the representations which we have already encountered[40]

$$\mathrm{def} \simeq \overline{\mathrm{def}}, \qquad \mathrm{sym} \simeq \overline{\mathrm{sym}} \simeq \mathrm{ad}, \qquad \mathrm{alt} \simeq \overline{\mathrm{alt}} \simeq \mathrm{triv}. \tag{5.113}$$

For the defining representations of $\mathfrak{su}(3)$ there is an obvious generalisation of the Pauli matrices to 3×3 hermitian traceless matrices. These are called the Gell-Mann matrices λ_a with $a = 1, \ldots 8 = D^{\mathrm{ad}}$. For the purposes of Yang–Mills theory it is almost always inconvenient to use these matrices directly rather than their algebraic relations.

For $N \geq 3$ the two defining representations are inequivalent. However, some relations remain. For $N = 3$ they include

$$\mathrm{alt} \simeq \overline{\mathrm{def}}, \qquad \overline{\mathrm{alt}} \simeq \mathrm{def}. \tag{5.114}$$

For $N = 4$ there is yet a different set of relations, in particular $\mathrm{alt} \simeq \overline{\mathrm{alt}}$ which becomes the vector representation of the equivalent orthogonal algebra $\mathfrak{so}(6)$.

[40]Note that these relations are consistent with the above table of characteristic constants for representation when $N = 2$.

6 Quantum Symmetries

Symmetries are crucial for quantum field theories as they provide stability in the renormalisation process. Furthermore, symmetries reduce the gauge field to its two physical polarisation modes.

Here we discuss the notion of symmetries at quantum level in the path integral framework. We derive a couple of identities which are analogues of Noether's theorem for QFT.

6.1 Schwinger–Dyson Equations

Before we consider quantum symmetries, let us discuss the equation of motion at quantum level which are called the Schwinger–Dyson equations.

Equations of Motion. Quantum fields obey the equations of motion. In the path integral they originate from an integrand which is a total derivative

$$\begin{aligned} 0 &= \int \mathrm{D}\Psi \, \frac{\delta}{\delta\Psi(x)} \exp\bigl(\mathring{\imath}S[\Psi] + \mathring{\imath}S_{\text{src}}[\Psi, J]\bigr) \\ &= \mathring{\imath} \int \mathrm{D}\Psi \left(\frac{\delta S}{\delta\Psi(x)} + J(x) \right) \exp\bigl(\mathring{\imath}S[\Psi] + \mathring{\imath}S_{\text{src}}[\Psi, J]\bigr). \end{aligned} \tag{6.1}$$

The first term in the brackets generates the Euler–Lagrange equations, the second one is a source term which is usually removed by setting $J = 0$ in the end. In terms of the generating functional this identity is known as the Schwinger–Dyson equation

$$\frac{\delta S}{\delta\Psi(x)} \left[-\mathring{\imath} \, \frac{\delta}{\delta J} \right] Z[J] = -J(x) Z[J]. \tag{6.2}$$

For example, it implies the expectation value of the classical equation of motion to hold

$$\bigl\langle \delta S[\Psi]/\delta\Psi(x) \bigr\rangle = 0. \tag{6.3}$$

In the presence of some other operator(s) $O[\Psi]$ in the correlator some more care is needed because the functional derivatives in the operator can remove the source J

before it is set to zero

$$\begin{aligned}\big\langle O[\Psi]\,\delta S/\delta\Psi(x)\big\rangle &= Z[J]^{-1}O\bigg[-\mathring{\imath}\,\frac{\delta}{\delta J}\bigg]\,\frac{\delta S}{\delta\Psi(x)}\,\bigg[-\mathring{\imath}\,\frac{\delta}{\delta J}\bigg]Z[J]\bigg|_{J=0}\\ &= -Z[J]^{-1}O\bigg[-\mathring{\imath}\,\frac{\delta}{\delta J}\bigg]J(x)Z[J]\bigg|_{J=0}\\ &= \mathring{\imath}Z[J]^{-1}\,\frac{\delta O}{\delta\Psi(x)}\,\bigg[-\mathring{\imath}\,\frac{\delta}{\delta J}\bigg]Z[J]\bigg|_{J=0}\\ &= \mathring{\imath}\big\langle\delta O/\delta\Psi(x)\big\rangle.\end{aligned} \tag{6.4}$$

Here, one of the fields $\Psi \simeq -\mathring{\imath}\delta/\delta J$ must act on the free $J(x)$, the other fields in O amount to $\delta O/\delta\Psi(x)$ and act on $Z[J]$. This result shows that the equations of motion are satisfied in time-ordered correlators unless there are other fields present at the same point of spacetime.[1]

Comments. Let us add a few assorted comments on the Schwinger–Dyson equation.

The Schwinger–Dyson equation is the quantum counterpart of the classical Euler–Lagrange equations. The latter play an important role for symmetries, for example the conservation of currents and charges depends on them. For quantum currents the above considerations imply that additional operators may alter their conservation laws in expectation values.

A convenient form of the Schwinger–Dyson equation uses correlators in the presence of sources J

$$\big\langle\delta S[\Psi]/\delta\Psi(x) + J(x)\big\rangle_J = 0. \tag{6.5}$$

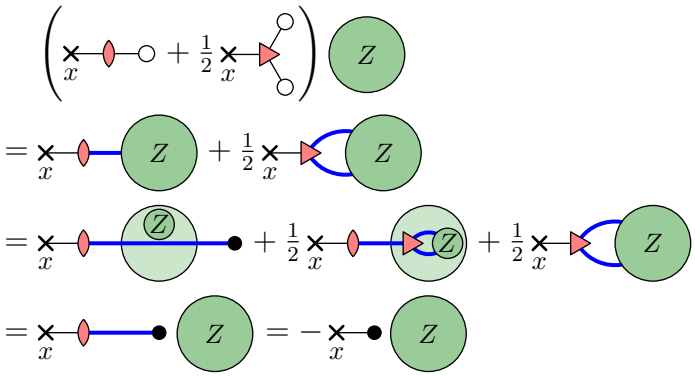
This is precisely the expectation value of the Euler–Lagrange equations in the presence of (arbitrary) sources. Expanding this equation in powers of J is equivalent to the above Schwinger–Dyson equations with arbitrary operator insertions. The benefit of this form is that it is fairly simple to manipulate. However, some care has to be taken because the field $\Psi(x)$ within the expectation value should be interpreted as $-\mathring{\imath}\delta/\delta J(x)$, and therefore it does not commute with J.

Commonly, the generating functional $Z[J]$ is defined as a path integral. Its integrand is a complex phase given by the action $\hbar^{-1}S[\Psi]$ plus a source term. Alternatively the generating functional can be viewed as the solution to the Schwinger–Dyson differential equation which is formulated in terms of the action S.

Feynman Graphs. It is instructive to see how the Schwinger–Dyson equation acts on Feynman diagrams. To that end we split up the variation of the action into free and interacting parts $\delta S_0/\delta\Psi(x)$ and $\delta S_{\text{int}}/\delta\Psi(x)$. The free action grabs one

[1] Operators at coincident spacetime points are typically ill-defined and require regularisation. The Schwinger–Dyson equation can be used to define the result of particular combinations of coincident operators.

end of a Feynman propagator, and converts it to a delta-function. The other end of the propagator may be attached to a source or to a vertex. The contribution from the first case corresponds to the term $J(x)Z[J]$. The contribution from the second case is cancelled precisely by the interacting part of the action.



(6.6)

Other Functionals. The Schwinger–Dyson equation can also be formulated for the generating functional of connected graphs $W[J]$ and for the effective action $G[\Psi]$.

We start with the Schwinger–Dyson equation with the constant $\hbar$ made explicit

$$\frac{\delta S}{\delta \Psi(x)}\left[-\mathring{\imath}\hbar\,\frac{\delta}{\delta J}\right]Z[J] = -J(x)Z[J]. \tag{6.7}$$

For the connected graphs one finds

$$\frac{\delta S}{\delta \Psi(x)}\left[-\mathring{\imath}\hbar\,\frac{\delta}{\delta J} + \frac{\delta W}{\delta J}\right] = -J(x). \tag{6.8}$$

Curiously, the generating functional now appears in the argument of the variation of the action.[2] Note that the functional derivative $\delta/\delta J$ in the argument of $\delta S/\delta\Psi$ can act only on the other argument $\delta W/\delta J$. Hence, the statement arises upon expansion of S which is typically a polynomial in the fields.

The Schwinger–Dyson equation for the effective action takes a rather implicit form

$$\frac{\delta S}{\delta \Psi(x)}\left[\Psi + \mathring{\imath}\hbar\left(\frac{\delta^2 G}{\delta \Psi^2}\right)^{-1}\frac{\delta}{\delta \Psi}\right] = \frac{\delta G}{\delta \Psi(x)}\,. \tag{6.9}$$

We observe that setting $\hbar = 0$ implies that the leading order of the effective action G coincides with the classical action S (up to a constant shift).

6.2 Slavnov–Taylor Identities

In classical field theory Noether's theorem guarantees one conserved current and associated conserved charge for every continuous global symmetry. A similar statement should also hold for quantum field theories.

[2] This is due to $\exp(-b(x))a(\partial/\partial x)\exp(b(x)) = a(\partial/\partial x + b'(x))$.

Noether's Theorem. Suppose the action $S[\Psi]$ is invariant under a global symmetry

$$0 = \frac{\delta S}{\delta \alpha} = \int \mathrm{d}x^D \, \frac{\delta \Psi(x)}{\delta \alpha} \, \frac{\delta S}{\delta \Psi(x)} \, . \tag{6.10}$$

Vanishing of the integral on the right-hand side implies that the integrand is a total derivative.[3] We define the Noether current $N^\mu(x)$ implicitly by generating the total derivative

$$\frac{\delta \Psi(x)}{\delta \alpha} \, \frac{\delta S}{\delta \Psi(x)} =: \partial_\mu N^\mu(x). \tag{6.11}$$

Note that the left-hand side of this equation vanishes on solutions of the Euler–Lagrange equations, so the current is conserved subject to the equations of motion

$$\partial_\mu N^\mu = \frac{\delta \Psi}{\delta \alpha} \, \frac{\delta S}{\delta \Psi} = 0. \tag{6.12}$$

Generating Functional. We have already derived the quantum analog of the Euler–Lagrange equation, namely the Schwinger–Dyson equation[4]

$$\frac{\delta S}{\delta \Psi(x)} \left[-\mathring{\imath} \, \frac{\delta}{\delta J} \right] Z[J] = -J(x) Z[J]. \tag{6.13}$$

Multiplying this identity by the field variation we find

$$\left(\frac{\delta \Psi(x)}{\delta \alpha} \, \frac{\delta S}{\delta \Psi(x)} \right) \left[-\mathring{\imath} \, \frac{\delta}{\delta J} \right] Z[J] = - \, \frac{\delta \Psi(x)}{\delta \alpha} \left[-\mathring{\imath} \, \frac{\delta}{\delta J} \right] J(x) Z[J]. \tag{6.14}$$

We substitute the Noether current to obtain the Slavnov–Taylor equation

$$\partial_\mu N^\mu \left[-\mathring{\imath} \, \frac{\delta}{\delta J} \right] (x) \, Z[J] = - \, \frac{\delta \Psi(x)}{\delta \alpha} \left[-\mathring{\imath} \, \frac{\delta}{\delta J} \right] \big(J(x) Z[J] \big). \tag{6.15}$$

We may integrate this equation over x to see that in the integral form the Noether current drops out

$$\int \mathrm{d}x^D \, \frac{\delta \Psi(x)}{\delta \alpha} \left[-\mathring{\imath} \, \frac{\delta}{\delta J} \right] \big(J(x) Z[J] \big) = 0. \tag{6.16}$$

This tells us that Z is invariant up to issues related to the source J.

The content of the Slavnov–Taylor identity is clearer when expanded in powers of J. Its expansion coefficients are identities between quantum correlation functions. For instance, the leading order is the classical conservation law of the expectation value

$$\big\langle \partial \cdot N(x) \big\rangle = 0. \tag{6.17}$$

[3] We ignore any potential boundary terms.

[4] Recall that a functional $F[-\mathring{\imath}\delta/\delta J]$ acting on $Z[J]$ should be interpreted as the functional $F[\Psi]$ within a quantum correlator $\langle \ldots \rangle_J$.

The term at first order reads

$$\big\langle \Psi(y)\,\partial_\mu N^\mu(x)\big\rangle = \mathring{\imath}\delta^D(x-y)\big\langle \delta\Psi(x)/\delta\alpha\big\rangle. \tag{6.18}$$

Here, the current is conserved unless there is another field at coincident location. Note that the extra term equals the symmetry variation of the other field. Therefore, one can view the current divergence $\partial\cdot N(x)$ to be a symmetry generator localised at x.[5]

After integration over x we find the global field variation

$$\big\langle \delta\Psi(y)/\delta\alpha\big\rangle = 0. \tag{6.19}$$

In words it tells us that the expectation value of the change of Ψ vanishes. This is obvious since the transformed field Ψ' is related to Ψ by symmetry.

The same holds at second order

$$\big\langle (\delta\Psi(y)/\delta\alpha)\Psi(z)\big\rangle + \big\langle \Psi(y)(\delta\Psi(z)/\delta\alpha)\big\rangle = 0 \tag{6.20}$$

telling that the expectation value of the change of the product $\Psi(y)\Psi(z)$ vanishes.

Altogether these relationships are no big surprises. In fact it makes sense to write the Slavnov–Taylor equation in terms of a correlator in the presence of sources[6]

$$\big\langle \partial_\mu N^\mu(x) + (\delta\Psi(x)/\delta\alpha)J(x)\big\rangle_J = 0. \tag{6.21}$$

The argument of the correlator is just the classical conservation of the Noether current when sources are included. The pleasant feature of this equation is that additional operators can be introduced into the correlator by means of the sources.

Example. Consider the interacting complex scalar field with Lagrangian

$$\mathcal{L} = -\partial^\mu\phi^*\,\partial_\mu\phi - m^2\phi^*\phi - \tfrac{1}{4}\lambda(\phi^*\phi)^2. \tag{6.22}$$

Rotation of the fields by a phase $\delta\phi = \mathring{\imath}\delta\alpha\,\phi$, $\delta\phi^* = -\mathring{\imath}\delta\alpha\,\phi^*$ is a continuous global symmetry. We define the associated Noether current via

$$\mathring{\imath}\phi\,\partial^2\phi^* - \mathring{\imath}\phi^*\partial^2\phi = \partial_\mu N^\mu \quad\Longrightarrow\quad N^\mu = \mathring{\imath}\phi\,\partial^\mu\phi^* - \mathring{\imath}\phi^*\partial^\mu\phi. \tag{6.23}$$

We can now write the above Slavnov–Taylor equation in terms of a correlator with sources

$$\big\langle \partial_\mu N^\mu(x)\big\rangle_{j,j^*} = -\mathring{\imath}\big\langle \phi(x)\,j(x) - \phi^*(x)\,j^*(x)\big\rangle_{j,j^*}. \tag{6.24}$$

Disabling the sources leads to the classical conservation law for the expectation value

$$\big\langle \partial_\mu N^\mu(x)\big\rangle = 0. \tag{6.25}$$

[5] This is analogous to the Noether charge acting as a symmetry generator in the canonical framework.

[6] The second term in this correlator contains both fields Ψ and conjugate sources J. For consistency one has to assume the non-trivial commutation relation $[J(x),\Psi(y)] = \mathring{\imath}\delta(x-y)$ and thus the ordering does matter in this case.

However, the conservation law is deformed when additional fields are inserted into the correlator via $\phi = -\mathring{\imath}\delta/\delta j$

$$\big\langle \partial_\mu N^\mu(x)\, \phi(y) \big\rangle = -\delta^D(x-y) \big\langle \phi(y) \big\rangle. \tag{6.26}$$

Integration over x yields the statement that the expectation value of a charged quantity vanishes

$$\big\langle \phi(y) \big\rangle = 0. \tag{6.27}$$

For the insertion of two fields we find

$$\big\langle \partial_\mu N^\mu(x)\, \phi(y)\, \phi^*(z) \big\rangle = \big(\delta^D(x-z) - \delta^D(x-y)\big) \big\langle \phi(y)\, \phi^*(z) \big\rangle. \tag{6.28}$$

Note that this identity holds exactly even in the presence of interactions. Here the integration over x yields a trivial statement. In other words, the symmetry permits a non-trivial expectation value of the uncharged correlator $\langle \phi(y)\phi^*(z) \rangle$.

6.3 Ward–Takahashi Identity

QED is a gauge theory with one abelian gauge symmetry. In the quantisation process we have broken gauge invariance, but a global U(1) invariance remains. The Slavnov–Taylor identity for the latter is called Ward–Takahashi identity. It guarantees consistent renormalisation of the coupling constant (charges).

Slavnov–Taylor Identities. The gauge-fixed QED Lagrangian with sources has the following terms

$$\begin{aligned} \mathcal{L}_{\text{QED}} &= \bar\psi(\gamma^\mu D_\mu - m)\psi - \tfrac{1}{4} F^{\mu\nu} F_{\mu\nu}, \\ \mathcal{L}_{\text{gf}} &= -\tfrac{1}{2}\xi^{-1}(\partial\cdot A)^2, \\ \mathcal{L}_{\text{src}} &= \bar\psi\rho + \bar\rho\psi - J^\mu A_\mu. \end{aligned} \tag{6.29}$$

Here the covariant derivatives are defined as

$$D_\mu\psi = (\partial_\mu - \mathring{\imath} q A_\mu)\psi, \qquad D_\mu\bar\psi = (\partial_\mu + \mathring{\imath} q A_\mu)\bar\psi. \tag{6.30}$$

We note that the global part of the gauge transformation acts as

$$\delta\psi = \mathring{\imath} q\, \delta\alpha\, \psi, \qquad \delta\bar\psi = -\mathring{\imath} q\, \delta\alpha\, \bar\psi. \tag{6.31}$$

We construct the Noether current N^μ from the action via its divergence $\partial\cdot N$ which has to equal $(\delta\Psi/\delta\alpha)(\delta S/\delta\Psi)$

$$\partial_\mu N^\mu = -\mathring{\imath} q \bar\psi \gamma^\mu (D_\mu\psi) - \mathring{\imath} q (D_\mu\bar\psi)\gamma^\mu\psi. \tag{6.32}$$

This equation has the electromagnetic current of the spinor field as its solution

$$N^\mu = -\mathring{\imath} q \bar\psi \gamma^\mu \psi. \tag{6.33}$$

The associated Slavnov–Taylor identity reads[7]

$$\left\langle \partial_\mu N^\mu - \mathring{\imath} q \bar{\psi}\rho + \mathring{\imath} q \bar{\rho}\psi \right\rangle_{\rho,\bar{\rho},J} = 0. \tag{6.34}$$

This equation is one version of the Ward–Takahashi identity which involves the electromagnetic current.

The identity is usually expressed in terms of a gauge field which couples to the electromagnetic current. The current N^μ also appears in the equations of motion for the electromagnetic field, let us therefore consider its Schwinger–Dyson equation

$$\left\langle N^\mu - \partial_\nu F^{\mu\nu} + \xi^{-1}\partial^\mu(\partial\cdot A) - J^\mu \right\rangle_{\rho,\bar{\rho},J} = 0. \tag{6.35}$$

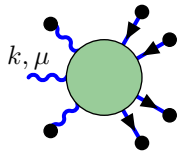
Assembling the two identities yields another version of the Ward–Takahashi identity which is encountered frequently

$$\left\langle -\xi^{-1}\partial^2(\partial\cdot A) + \partial\cdot J - \mathring{\imath} q \bar{\psi}\rho + \mathring{\imath} q \bar{\rho}\psi \right\rangle_{\rho,\bar{\rho},J} = 0. \tag{6.36}$$

Momentum Space. The identity is conventionally expressed in momentum space as follows:

$$\begin{aligned} &\left\langle \xi^{-1}k^2 k\cdot A(k) + k\cdot J(k) \right\rangle_{\rho,\bar{\rho},J} \\ &= q\int \frac{\mathrm{d}p^4}{(2\pi)^4} \left\langle \bar{\psi}(-p)\rho(k+p) - \bar{\rho}(k+p)\psi(-p) \right\rangle_{\rho,\bar{\rho},J}. \end{aligned} \tag{6.37}$$

The first term picks a Feynman graph with an external photon line whose momentum is k and whose polarisation index is μ.



(6.38)

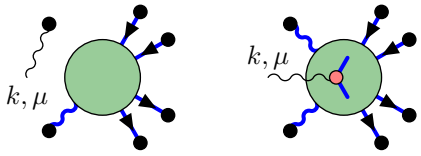
It multiplies the graph by $\xi^{-1}k^2 k_\mu$. The external photon line connects to the remainder of the graph by the propagator

$$G_{\mathrm{F}}^{\mu\nu}(k) = \frac{\eta^{\mu\nu} - (1-\xi)k^\mu k^\nu/k^2}{k^2 - \mathring{\imath}\epsilon}\,. \tag{6.39}$$

Multiplied by the prefactor we obtain

$$\xi^{-1}k^2 k_\mu G_{\mathrm{F}}^{\mu\nu}(k) = k^\nu. \tag{6.40}$$

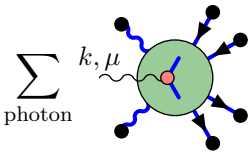
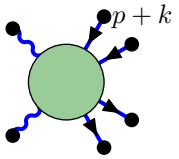
This means we actually consider a graph with one amputated photon line whose polarisation vector is contracted with k. Now there are two options:



(6.41)

[7]Here the ordering of fields and sources merely follows the spinor contractions; we disregard the proper quantum ordering.

- The other side of the photon propagator is connected to a source J, i.e. we consider a plain photon propagator (disconnected from the remainder of the graph). In this case we obtain $\partial{\cdot}J$ which cancels precisely the corresponding term in the Ward–Takahashi identity. This part of the identity is somewhat trivial as it concerns only a plain photon propagator.
- Otherwise, the photon line connects to a non-trivial (sub)graph. The remaining two terms involving spinors and sources must cancel this graph. They pick a spinor line of a graph with one fewer photon line. They shift the spinor momentum by k, and they multiply the graph by the charge $\pm q$ of the particle. The sum over all external spinor lines should then precisely cancel the sum over all photon insertions as explained above

$$k_\mu \sum_{\text{photon}}$$  $$= \sum_{\text{matter}} (\pm q)$$  . (6.42)

Two-Point Function, Vertex and Scattering. Let us consider two examples, a photon two-point function and a spinor-photon vertex:

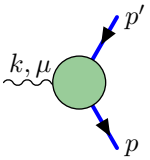
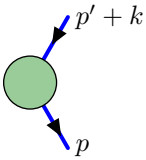
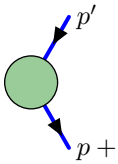
We have already seen that the plain photon propagator $G_{\mathrm{F}}^{\mu\nu}(k)$ satisfies the identity, and it absorbs the term $\partial{\cdot}J$ completely. How about the loop correction $M^{\mu\nu}(k)$? Consider the graph $\xi^{-1}k^2k_\mu M^{\mu\nu}(k)$ with one external photon line amputated and contracted with its own momentum. As there are no external spinor lines, this graph must vanish exactly. We learn that the loop corrections to the propagator must be exactly transversal

$$M^{\mu\nu}(k) \sim \eta^{\mu\nu} - k^\mu k^\nu/k^2. \tag{6.43}$$

The only violation of transversality originates from the gauge fixing term in the plain propagator. This identity substantiates our earlier claim in the context of introduction of ghost fields.

Next, consider a quantum vertex $M_\mu(k,p,p')$ of two spinor fields and a photon where the photon line of momentum k was already amputated. Denote by $M_0(p,p')$ the quantum propagator of the spinor field. The Ward–Takahashi identity implies[8]

$$-\mathring{\imath}k{\cdot}M(k,p,p') = \mathring{\imath}qM_0(p,p'+k) - \mathring{\imath}qM_0(p+k,p'). \tag{6.44}$$

$$-\mathring{\imath}k_\mu \cdot$$  $$= \mathring{\imath}q \cdot$$  $$- \mathring{\imath}q \cdot$$  . (6.45)

There is also a version of the Ward–Takahashi identity for elements of the scattering matrix where all external lines are on-shell. This identity is called the

[8] This identity holds exactly at tree level and at loop level. Note that in this form all the constants and fields are bare. However, there is an analogous renormalised expression.

Ward identity. Often the Ward identity is considered for photons in the limit of zero momentum where the contributing diagrams are somewhat simpler to evaluate. This is a convenient limit for the comparison between the various counterterm coefficients since the latter do not depend on the momentum.

The Ward–Takahashi identities can also be formulated for non-abelian gauge theories. They contain quite a few more terms due to non-linear effects in the gauge fields and due to the ghosts. Effectively, they yield a similar set of constraints.

6.4 Anomalies

Above we have discussed equations that hold due to symmetries in the QFT. An important aspect of symmetries in QFT is that they may break due to quantum effects. Such symmetries are called anomalous.

Scaling Anomaly. A massless scalar field with quartic interactions in four dimensions is classically invariant under scale transformations since the model has no dimensionful parameter. Moreover, the classical field theory is invariant under conformal transformations. These are global symmetry transformations which can be treated with the above framework.

However, we have seen that the theory has a non-trivial beta-function which effectively associates an anomalous dimension to the coupling constant λ. The associated running of the coupling is a physically observable effect. This implies that quantum corrections actually destroy scale and conformal invariance. The QFT model is conformal only at the non-interacting point $\lambda = 0$.

The mechanism of this conformal anomaly is as follows:

- The naive QFT treatment leads to divergences.
- Regularisation manifestly breaks conformal symmetry due to the introduction of a scale, e.g. the dimensional regularisation scale μ or the cutoff scale Λ.
- Renormalisation removes the divergences.
- Although renormalisation leaves some freedom in defining the effective coupling constant, there is no way to recover scaling or conformal symmetry.
- Therefore scaling symmetry is anomalous.

When a symmetry is anomalous, it does not mean that we lose the associated Slavnov–Taylor identities entirely. The anomaly merely deforms them, but we can still make use of them.

Axial Anomaly in Massless QED. A prominent type of anomaly in QFT is the axial anomaly. We have learned in QFT I that a massless Dirac spinor field has a chiral symmetry

$$\psi \to \exp(-\mathring{\imath}\gamma^5\beta)\psi \tag{6.46}$$

in addition to its usual electric charge symmetry $\psi \to \exp(\mathring{\imath}\alpha)\psi$. The associated Noether current is the axial current

$$N_{\mathrm{A}}^{\mu} = -\mathring{\imath}\bar{\psi}\gamma^5\gamma^{\mu}\psi. \tag{6.47}$$

It is classically conserved, even in the presence of the electromagnetic field

$$\partial{\cdot}N_{\mathrm{A}} = -\mathring{\imath}\bar{\psi}\gamma^5\gamma^{\mu}(D_{\mu}\psi) - \mathring{\imath}(D_{\mu}\bar{\psi})\gamma^5\gamma^{\mu}\psi = 0. \tag{6.48}$$

The naive Slavnov–Taylor identity for the axial current thus reads

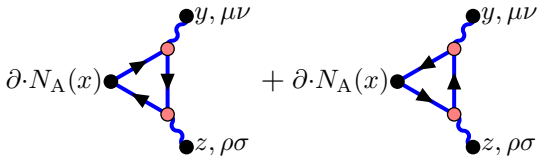
$$\left\langle \partial_{\mu}N_{\mathrm{A}}^{\mu} - \mathring{\imath}\bar{\rho}\gamma^5\psi - \mathring{\imath}\bar{\psi}\gamma^5\rho \right\rangle_{\rho,\bar{\rho},J} = 0. \tag{6.49}$$

Consider now the correlator of the axial current with two photons

$$\left\langle \partial{\cdot}N_{\mathrm{A}}(x)\, F_{\mu\nu}(y)\, F_{\rho\sigma}(z) \right\rangle \stackrel{?}{=} 0. \tag{6.50}$$

By the above identity the correlator should be zero because there are no further fermions to which the extra terms in the identity could couple.

Let us discuss the two contributions from one-loop graphs



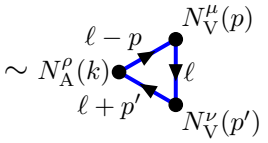
. (6.51)

We define the underlying loop integral in momentum space as

$$L^{\mu\nu\rho}(p,p') := I^{\mu\nu\rho}(p,p') + I^{\nu\mu\rho}(p',p) \tag{6.52}$$

with the vector-vector-axial triangle integral

$$I^{\mu\nu\rho}(p,p') := \int \frac{-\mathring{\imath}\,\mathrm{d}\ell^4}{(2\pi)^4}\,\mathrm{tr}\left[\gamma^5\gamma^{\rho}\,\frac{\gamma{\cdot}(\ell-p)}{(\ell-p)^2}\,\gamma^{\mu}\,\frac{\gamma{\cdot}\ell}{\ell^2}\,\gamma^{\nu}\,\frac{\gamma{\cdot}(\ell+p')}{(\ell+p')^2}\right] \tag{6.53}$$

$\sim$ 

.

The above correlator consists of the contraction $k_{\rho}L^{\mu\nu\rho}(p,p')$. Now write $k = -p - p' = (\ell - p) - (\ell + p')$, split the integrand into two terms and permute one past γ^5. Each term then cancels one of the adjacent spinor propagators. Permuting another term past γ^5 yields

$$k_{\rho}I^{\mu\nu\rho} = \int \frac{-\mathring{\imath}\,\mathrm{d}\ell^4}{(2\pi)^4}\,\mathrm{tr}\left[\gamma^5\gamma^{\mu}\,\frac{\gamma{\cdot}\ell}{\ell^2}\,\gamma^{\nu}\left(\frac{\gamma{\cdot}(\ell+p')}{(\ell+p')^2} - \frac{\gamma{\cdot}(\ell-p)}{(\ell-p)^2}\right)\right]. \tag{6.54}$$

It is tempting to argue that this integral vanishes:

- The trace of gamma-matrices produces the anti-symmetric tensor ε in four dimensions, $\mathrm{tr}\,\gamma^5\gamma^{\mu}\gamma^{\nu}\gamma^{\rho}\gamma^{\sigma} = 4\mathring{\imath}\varepsilon^{\mu\nu\rho\sigma}$.

- When writing the expression as a difference of two integrals, each one should depend on either p or p' but not on both.
- Altogether there are only three vectors or indices at our disposal: μ, ρ and either p or p'. There is nothing to act as the fourth index of ε.

This conclusion is however too naive: One must pay attention to the fact that the two above integrals are divergent. Subtracting two divergent integrals is an ambiguous operation. The result may in principle be adjusted to any desired value, be it zero, finite or infinite.

In fact, the original integral $I^{\mu\nu\rho}$ diverges linearly which implies that its value depends on the choice of integration boundary. In other words, a shift of the loop momentum ℓ by $\delta\ell$ is not trivial, but it changes the integral $I^{\mu\nu\rho}$ by a finite amount

$$\delta I^{\mu\nu\rho} = -\frac{\mathring{i}}{8\pi^2}\,\varepsilon^{\mu\nu\rho\sigma}\delta\ell_\sigma. \tag{6.55}$$

Importantly, such a shift can modify the conservation of currents and therefore the integration boundary has to be defined carefully.

Even though there is some residual arbitrariness in the definition of (the divergent) integral $I^{\mu\nu\rho}$, it does obey some unambiguous relations, which can be seen as obstructions to symmetries. Concretely, the definition of the loop integral $I^{\mu\nu\rho}(p,p')$ is cyclically symmetric in its three legs, but only up to a shift of the integration variable. Non-triviality of the latter implies a non-trivial cyclic symmetry relationship

$$\begin{aligned} I^{\mu\nu\rho}(p,p') &= I^{\nu\rho\mu}(p',k) - \frac{\mathring{i}}{8\pi^2}\,\varepsilon^{\mu\nu\rho\sigma}p'_\sigma \\ &= I^{\rho\mu\nu}(k,p) + \frac{\mathring{i}}{8\pi^2}\,\varepsilon^{\mu\nu\rho\sigma}p_\sigma. \end{aligned} \tag{6.56}$$

For the relevant symmetric combination $L^{\mu\nu\rho}$ this implies a relationship between the conservation of the axial and vector currents

$$k_\rho L^{\mu\nu\rho}(p,p') = k_\rho L^{\rho\mu\nu}(k,p) + \frac{\mathring{i}}{4\pi^2}\,\varepsilon^{\mu\nu\rho\sigma}p_\rho p'_\sigma. \tag{6.57}$$

By choosing the integration such that the vector current is exactly conserved, $k_\rho L^{\rho\mu\nu}(k,p) = 0$,[9] which is crucial for a consistent gauge symmetry in the quantum theory, we find that the axial current is not conserved

$$k_\rho L^{\mu\nu\rho}(p,p') = \frac{\mathring{i}}{4\pi^2}\,\varepsilon^{\mu\nu\rho\sigma}p_\rho p'_\sigma. \tag{6.58}$$

The pitfall in the above considerations depends on the regularisation scheme. We have argued using a momentum cut-off. However the same conclusion can be reached in other schemes.

[9]For instance, this is achieved by an appropriate counterterm which arises in the same places as the loop integral.

In dimensional regularisation the above derivation makes the mistake to assume that γ^5 anti-commutes with all γ^μ. First of all γ^5 is a strictly 4-dimensional concept. One can still define an anti-symmetric product of 4 gamma-matrices which imitates the behaviour of γ^5. However, there are discrepancies at $\mathcal{O}(\epsilon)$. For instance, it anti-commutes with 4 gamma-matrices but it commutes with the remaining (-2ϵ) ones.

$$\gamma^5\gamma^\mu \neq -\gamma^\mu\gamma^5. \tag{6.59}$$

The resulting factor of ϵ is then cancelled by a factor of $1/\epsilon$ from a divergent loop integral.

It is also instructive to investigate the integral $L^{\mu\nu\rho}$ in position space. The anomaly originates from a distributional contribution when taking a derivative at a singularity

$$\begin{aligned}&\frac{\partial}{\partial z^\rho}\,\mathrm{tr}\left[\gamma^5\gamma^\rho\,\frac{\gamma\cdot(z-x)}{(z-x)^4}\,\gamma^\mu\,\frac{\gamma\cdot(x-y)}{(x-y)^4}\,\gamma^\nu\,\frac{\gamma\cdot(y-z)}{(y-z)^4}\right]\\ &\sim \varepsilon^{\mu\nu\rho\sigma}\,\partial_\rho\delta^4(z-x)\,\partial_\sigma\delta^4(z-y),\end{aligned} \tag{6.60}$$

which is analogous to the identity $\partial^2 x^{-2} \sim \delta^4(x)$ required for the scalar propagator.

Curiously, a careful evaluation in either scheme yields the same result. It can be summarised in the anomalous Slavnov–Taylor identity known as the Adler–Bell–Jackiw identity

$$\left\langle \partial_\mu N^\mu_{\mathrm{A}} - \mathring{\imath}\bar\rho\gamma^5\psi - \mathring{\imath}\bar\psi\gamma^5\rho - q^2(4\pi)^{-2}\,\varepsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}\right\rangle_{\rho,\bar\rho,J} = 0. \tag{6.61}$$

Adler and Bardeen showed that this identity is exact, there are no further corrections from higher loop orders.

Let us discuss the additional term in the Adler–Bell–Jackiw identity. We note that it has a special form: it can be written as a total derivative of a *local* quantity. This allows to redefine the axial Noether current by a compensating local term involving the gauge fields

$$N'^\mu_{\mathrm{A}} = N^\mu_{\mathrm{A}} - \frac{q^2}{8\pi^2}\,\varepsilon^{\mu\nu\rho\sigma}A_\nu F_{\rho\sigma}. \tag{6.62}$$

which obeys the non-anomalous Slavnov–Taylor identity. However, in contradistinction to N_{A}, the redefined axial current is no longer gauge invariant due to the presence of a bare gauge potential A. This again points at a clash between axial and gauge symmetry in the quantum theory.

A related feature of the anomaly term $F \wedge F = \mathrm{d}(A \wedge F)$ is that its integral is topological; it does not depend on the local degrees of freedom of the gauge field. As the Slavnov–Taylor identity was derived in integral form while discarding all potential boundary contributions, it is perhaps not too surprising to observe that the anomaly term has such a structure.

Comments. Anomalies are an exciting topic of quantum field theory which we cannot treat in detail in this course. Let us nevertheless make a few more remarks:

Anomalies are often related to topological issues of spacetime and of the gauge groups. The corresponding anomalous Slavnov–Taylor identity is typically one-loop exact.[10]

Gauge fields can also be coupled to the axial current. In fact, the electroweak sector of the standard model relies on such couplings. The presence of an axial anomaly would be disastrous since it would violate gauge invariance. Interestingly, all such anomalies cancel in the standard model due to a particular arrangement of spinor fields. This is called anomaly cancellation and it represents one of the consistency requirements of QFT on the particle spectrum.

Note that the coupling of a chiral spinor to a gauge field effectively is a combination of vector and axial vector currents. Hence, chiral spinor fields are the origin of the so-called chiral anomaly. The chiral anomaly for quarks can be observed in meson decay.

Anomalies are universal statements of QFT. They state that certain symmetries are incompatible with the quantisation procedure. An anomaly can only occur for non-manifest symmetries which are not respected by the regulator. When the regulator is removed after renormalisation, some of these non-manifest symmetries may turn out intact if the renormalisation parameters are chosen appropriately. If there is no way to recover the symmetry using local counterterms, the symmetry is anomalous.

The amount by which the symmetry is broken is a universal statement independent of the regularisation scheme. The anomaly is typically attributed to non-invariance of the integration measure $\mathrm{D}\Psi$ of the path integral.

Note that even though anomalies are universal, some details may depend on the implementation. Some anomalies refer to pairs of symmetries which cannot be realised simultaneously. The renormalisation process can be adjusted to preserve one symmetry at the cost of the other. Alternatively the other symmetry can be preserved with suitable counterterms. A third option is to break both symmetries. For example, the axial current can be redefined to be exactly conserved, but only at the cost of an anomaly for the vector current. The latter is highly undesirable since it would definitely break gauge invariance. Therefore, one usually makes sure that the vector current is non-anomalous.

[10] Dependence on a continuous parameter can disturb the topological properties of anomaly terms, and only the one-loop contributions to the effective action are independent of $\hbar$.

7 Spontaneous Symmetry Breaking

In nature we observe three kinds of vector particles $W^\pm$ and Z^0 which are massive. In general, models with massive vector fields are non-renormalisable since the massive vector propagator screws up the power counting schemes. The only renormalisable implementation of vector fields is by means of gauge theory. Gauge symmetry, however, forces vector particles to be massless. Breaking gauge symmetry by brute force renders the model non-renormalisable. The way out of this paradoxical situation is spontaneous symmetry breaking.

7.1 Breaking of Global Symmetries

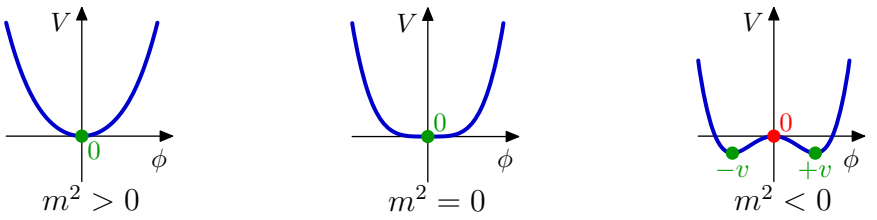
We start by introducing the concept of spontaneous symmetry breaking in the simple context of global symmetries. Spontaneous symmetry breaking merely obscures symmetries; the QFT models still possess such symmetries, but they are not realised in an obvious fashion. The existence of the symmetry has an interesting consequence for the particle spectrum which can be observed directly.

Breaking of a Discrete Symmetry. We have discussed various kinds of interacting scalar field theories. The Lagrangian had the basic form

$$\mathcal{L} = -\tfrac{1}{2}(\partial\phi)^2 - \tfrac{1}{2}m^2\phi^2 - \tfrac{1}{24}\lambda\phi^4. \tag{7.1}$$

We have dropped a potential cubic term in order to have a minimal amount of symmetry, namely the global discrete $\mathbb{Z}_2$ symmetry $\phi \to -\phi$.

So far we have always assumed the mass squared m^2 to be either zero or positive in order to have a physical interpretation for the field excitations. However, it is perfectly permissible to assume a negative value for m^2. The relevant point is that the potential is dominated by the quartic interaction ϕ^4 for large ϕ. Consequently, it is bounded from below (for positive λ) irrespectively of the magnitude or sign of the mass term. The potential has one of the following shapes



(7.2)

Let us therefore consider the case of negative m^2 expressed as a positive $\mu^2 = -m^2$

$$\mathcal{L} = -\tfrac{1}{2}(\partial\phi)^2 + \tfrac{1}{2}\mu^2\phi^2 - \tfrac{1}{24}\lambda\phi^4. \tag{7.3}$$

Evidently, the expansion of the field around a classical value $\phi = 0$ leads to undesirable tachyonic excitations. In other words, $\phi = 0$ is a stationary point of the potential; it is however not a local minimum, but rather a local maximum.

Since the potential is bounded from below, there must be a minimum somewhere around which we can quantise the model. In fact, there are two local minima at

$$\phi = \pm v, \qquad v = \sqrt{6\mu^2/\lambda}\,. \tag{7.4}$$

The constant field $\phi(x) = \pm v$ is a solution to the equation of motion. We can therefore expand the field ϕ around this solution

$$\phi(x) = \pm v + \eta(x), \tag{7.5}$$

where η serves as the quantum field. The expectation value of the quantum field in the vacuum is defined to be zero, hence v is the vacuum expectation value of $\phi(x)$

$$\big\langle \phi(x) \big\rangle = \pm v. \tag{7.6}$$

We obtain a new Lagrangian in the new field η

$$\mathcal{L} = -\tfrac{1}{2}(\partial\eta)^2 + 3\mu^4/2\lambda - \mu^2\eta^2 \mp \sqrt{\lambda\mu^2/6}\,\eta^3 - \tfrac{1}{24}\lambda\eta^4. \tag{7.7}$$

This is again a Lagrangian in the same class of scalar field theories. Let us discuss the terms that arise:

- The derivative term and the quartic interactions have not changed.
- The constant term does not actually play a role, we can disregard it.
- The term linear in η receives two contributions which cancel exactly. This is the condition of the minimum.
- We obtain a regular mass term with mass $m = \sqrt{2}\,\mu$.
- There is a new cubic term with coefficient fixed by the quartic coupling and mass term.

The cubic term evidently violates the $\mathbb{Z}_2$ symmetry $\eta \to -\eta$, but altogether the original symmetry is still present in the somewhat obscure form

$$\eta \to \mp 2v - \eta. \tag{7.8}$$

This symmetry survives quantisation, but it is not compatible with the vacuum expectation value of the fields. The symmetry is said to be *spontaneously broken*.

The model has two possible ground state configurations $\phi = \pm v$, both are global minima of the potential. Which one should we choose as the ground state? There is no criterion to prefer one state over the other. This is the generic situation for spontaneous symmetry breaking: Since symmetry does not leave the ground state invariant, it must map it to some other state. This state has equivalent properties and therefore is also a ground state. Hence, spontaneous symmetry breaking implies degeneracy of the ground state. Which one of the two we choose makes no difference since all physics is the same for both states.[1] We can thus set $\phi(x) = +v + \eta(x)$ without prejudice.[2]

[1] One might also consider linear combinations of ground states. In our situation, however, any transition between ground states would require an infinite amount of energy due to the infinite extent and resolution of spacetime. Each individual ground state will thus lead a life of its own, all of them would be completely equivalent, and mixing of such states is possible but irrelevant.

[2] Fixing one of the ground states may appear paradoxical in the path integral framework: It

Breaking of a Continuous Symmetry. Next, consider a generalisation of the above model to N scalar fields Φ^a with a global $\mathrm{O}(N)$ symmetry

$$\mathcal{L} = -\tfrac{1}{2}(\partial\Phi)^2 + \tfrac{1}{2}\mu^2\Phi^2 - \tfrac{1}{8}\lambda(\Phi^2)^2. \tag{7.9}$$

The potential has a local maximum at $\Phi = 0$, and a family of global minima at the spherical surface defined by

$$\Phi^2 = v^2, \qquad v = \sqrt{2\mu^2/\lambda}\,. \tag{7.10}$$

As discussed above, all of these points are related by $\mathrm{O}(N)$ symmetry and are therefore equivalent. Without prejudice we can choose any representative, e.g.

$$\Phi_0 = (0, \ldots, 0, v). \tag{7.11}$$

We thus introduce the quantum fields η and ϕ^a with $a = 1, \ldots, N-1$

$$\Phi^a(x) = \begin{cases} \phi^a(x) & \text{for } a = 1, \ldots, N-1, \\ v + \eta(x) & \text{for } a = N. \end{cases} \tag{7.12}$$

Substitution in the Lagrangian yields the terms

$$\begin{aligned} \mathcal{L} = &-\tfrac{1}{2}(\partial\phi)^2 - \tfrac{1}{8}\lambda(\phi^2)^2 \\ &- \tfrac{1}{2}(\partial\eta)^2 - \mu^2\eta^2 - \sqrt{\lambda\mu^2/2}\,\eta^3 - \tfrac{1}{8}\lambda\eta^4 \\ &+ \mu^4/2\lambda - \sqrt{\lambda\mu^2/2}\,\phi^2\eta - \tfrac{1}{4}\lambda\phi^2\eta^2. \end{aligned} \tag{7.13}$$

It describes a scalar field with mass $m = \sqrt{2}\,\mu$ and $N-1$ massless scalar fields which interact in a particular pattern. It has a manifest $\mathrm{O}(N-1)$ symmetry of the scalars ϕ^a which enlarges to the original $\mathrm{O}(N)$ symmetry in a non-evident fashion.

The fact that there are $N-1$ massless fields is remarkable. It is a consequence of Goldstone's theorem.

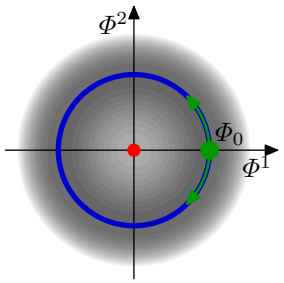
Goldstone's Theorem. Goldstone's theorem asserts that for every spontaneously broken continuous global symmetry there is one massless particle with the same quantum numbers. For ordinary symmetries the particle is a scalar, called the Goldstone boson.

One can understand the statement at the level of the Lagrangian of a model of scalar fields Φ^a with a potential $V(\Phi)$:

- The ground state is specified by a (global) minimum Φ_0 of $V(\Phi)$.
- The mass matrix is given by the curvature (second-order derivatives) of the potential at this point.

sums over all field configurations which appears to include all ground states, so how can we specialise to one? In fact, we should specify some boundary conditions for the fields. The vacuum expectation value is thus implemented as the field value in asymptotic regions of spacetime.

- Suppose some of the symmetries map Φ_0 to a different point, and are therefore spontaneously broken. For each of these continuous symmetries there is a corresponding direction in the space of Φ's where the potential $V(\Phi)$ is constant or flat.



(7.14)

  Consequently, the eigenvalue of the mass matrix along these directions is zero.
- The remaining symmetries do not change Φ_0, hence they are unbroken.
- The potential along the other directions is unrestricted by symmetry. There can be additional massive and massless modes, but their number or properties are not (directly) determined by symmetry.

In the above example, the vacuum breaks the global $\mathrm{O}(N)$ symmetry to the group $\mathrm{O}(N-1)$ which preserves the vector Φ_0. There are $N-1$ broken symmetry generators which form a vector of the subgroup. These are associated to the massless fields ϕ^a via Goldstone's theorem. The remaining field η is massive.

Goldstone's theorem holds for any type of fundamental field, and it applies to the full quantum theory. This is quite clear in the canonical framework in connection to Noether's theorem which guarantees the existence of a conserved current N^μ with associated conserved charge Q:

- In the canonical framework a symmetry is generated by the Noether charge Q. Suppose a Noether charge Q does not annihilate the ground state $|0\rangle$. In this case the associated symmetry is said to be broken.
- The charge Q is the spatial integral over $N^0(\vec{x})$. In momentum space this translates to the statement $Q \sim N^0(\vec{p}=0)$.
- Since Q is conserved, N^0 at $\vec{p}=0$ has only the mode of energy $E=0$. This mode generates some non-trivial excitation of the ground state.
- Assuming that the above statement can be viewed as the limiting case $\vec{p}\to 0$, $N^\mu(p)$ should contain an excitation of the vacuum which has a non-trivial limit only for $E\to 0$. This excitation therefore is massless.
- Note that this guarantees that the quantum numbers of the particle match those of the Noether charge N^0 and the broken symmetry generator.

In our example, the Noether current reads $(a,b=1,\dots N)$

$$N_\mu^{ab} = \Phi^a \partial_\mu \Phi^b - \Phi^b \partial_\mu \Phi^a. \tag{7.15}$$

Expressed in terms of the new fields we find $(a,b=1,\dots N-1)$

$$N_\mu^{ab} = \phi^a \partial_\mu \phi^b - \phi^b \partial_\mu \phi^a, \qquad N_\mu^{Na} = v\partial_\mu \phi^a + \eta \partial_\mu \phi^a - \phi^a \partial_\mu \eta. \tag{7.16}$$

We note that for the latter $N-1$ currents, there is a contribution linear in the fields. These currents N_μ^{Na} indeed create (and annihilate) the massless particles ϕ^a

from the ground state $|0\rangle$. Conversely, the unbroken currents N_μ^{ab} are quadratic in the fields and always consist of one creation and one annihilation operator, so they preserve the particle number.

There are also situations where Noether's theorem is less obvious. For example, in QCD it can be used to argue for the lightest pseudoscalar mesons. QCD is a theory of spinor and gauge fields. It is strongly coupled at the nuclear scale which is substantially higher than the fundamental quark masses. At nuclear scales, it makes sense to approximate the quark masses by zero. Then the quark fields have an (approximate) axial symmetry. Suppose the axial symmetry is spontaneously broken (by a scalar expectation value of quark bilinears). This would lead to the existence of massless pseudoscalar mesons. However, since axial symmetry is not exact in the presence of fermion masses, one would expect the pseudoscalar mesons to have a reasonably small mass as well. This is the case for the pions which one could consider as the pseudo-Goldstone bosons of a spontaneously *and* explicitly broken axial SU(2) flavour symmetry relating the up and down quarks. Additionally, this axial SU(2) may be subject to anomalies.

7.2 Breaking of Gauge Symmetries

The picture changes again when the spontaneously broken symmetry belongs to a local symmetry. The coupling of currents to gauge fields leads to massive vector fields instead of Goldstone bosons.

Higgs Mechanism. The above model for $N = 2$ is equivalent to a complex scalar field ϕ, ϕ^*. We couple it to the electromagnetic field to obtain scalar QED, but we assume that the mass term has the opposite sign

$$\mathcal{L}_{\text{SQED}} = -(D^\mu\phi)^*(D_\mu\phi) + \mu^2\phi^*\phi - \tfrac{1}{4}\lambda(\phi^*\phi)^2 - \tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{7.17}$$

The covariant derivative is defined as $D_\mu\phi = \partial_\mu\phi - \mathring{\imath}qA_\mu\phi$.

Here the ground state solution is specified by a point on a circle in the complex plane

$$\phi^*\phi = \tfrac{1}{2}v^2, \qquad v = \sqrt{4\mu^2/\lambda}, \tag{7.18}$$

We could go ahead as before and introduce a shifted quantum field

$$\phi(x) = \frac{1}{\sqrt{2}}\big(v + \eta(x) + \mathring{\imath}\rho(x)\big). \tag{7.19}$$

This would lead to the same form of Lagrangian as above with the exception of the kinetic term which yields

$$\begin{aligned} -(D^\mu\phi)^*(D_\mu\phi) = &-\tfrac{1}{2}(\partial\eta)^2 - \tfrac{1}{2}(\partial\rho)^2 - \tfrac{1}{2}q^2v^2A^2 + qvA_\mu\partial^\mu\rho \\ &+ \text{interactions}. \end{aligned} \tag{7.20}$$

As it stands this is not a very pleasant form due to the quadratic coupling between the gauge field A_μ and the scalar field ρ. We must first diagonalise by shifting the gauge field appropriately

$$A_\mu(x) = V_\mu(x) + (qv)^{-1}\partial_\mu\rho(x). \tag{7.21}$$

This shift changes the kinetic terms in the Lagrangian substantially

$$-(D^\mu\phi)^*(D_\mu\phi) = -\tfrac{1}{2}(\partial\eta)^2 - \tfrac{1}{2}q^2v^2V^2 + \text{interactions}. \tag{7.22}$$

Note that the kinetic term for the field ρ has cancelled out exactly. What remains is one scalar field η and a mass term for the vector field with mass $m = qv$.

This is the Higgs effect: The coupling of the gauge field to a conserved current of a spontaneously broken symmetry lends the gauge field a mass. Moreover, one of the scalar fields is absorbed completely by this process.

There is an alternative parametrisation of the fields which makes the above statement more obvious. The crucial observation is that the above reparametrisation of the gauge field is a gauge transformation. We thus split the complex scalar field into absolute value and complex phase

$$\phi(x) = \frac{1}{\sqrt{2}}\big(v + \eta(x)\big)\exp\big(\mathring{\imath}\alpha(x)\big). \tag{7.23}$$

By means of a gauge transformation we can remove the phase from the field ϕ without changing any of the terms in the Lagrangian

$$\phi(x) = \frac{1}{\sqrt{2}}\big(v + \eta(x)\big). \tag{7.24}$$

It appears as though one degree of freedom is lost. However, it is compensated by the loss of gauge symmetry since the field ϕ is now real and must not be rotated in the complex plane. Gauge symmetry effectively removes one degree of freedom from the gauge field. In that sense, we have merely transferred the complex phase of ϕ into the field A_μ which henceforth should be considered a vector field V_μ rather than a gauge field. The resulting Lagrangian reads

$$\begin{aligned}\mathcal{L}_{\text{SQED}} = &-\tfrac{1}{2}(\partial\eta)^2 - \tfrac{1}{4}v^2\lambda\eta^2 \\ &- \tfrac{1}{2}(\partial^\mu V^\nu)(\partial_\mu V_\nu) + \tfrac{1}{2}(\partial\cdot V)^2 - \tfrac{1}{2}q^2v^2V^2 \\ &+ \tfrac{1}{16}v^4\lambda - \tfrac{1}{4}\lambda v\eta^3 - \tfrac{1}{16}\lambda\eta^4 - q^2vV^2\eta - \tfrac{1}{2}q^2V^2\eta^2.\end{aligned} \tag{7.25}$$

We thus obtain a scalar field η with mass $\sqrt{v^2\lambda/2}$ and a vector field V_μ with mass qv. The scalar field interacts with itself and with the square of the vector field.

The Lagrangian has a very special form: Altogether it has 2 mass terms and 5 interaction terms (the coefficient of $\eta V_\mu\partial^\mu\eta$ is zero). Using the $\mathbb{Z}_2$ symmetry $\eta \to -2v - \eta$ these numbers reduce to 4 coefficients. However, there are only three parameters v, q and λ, so there is one constraint among the set of conceivable parameters. This feature is owed to the gauge symmetry in the original formulation of the Lagrangian. It will be crucial for a successful renormalisation.

Comments. Above we have presented an example of the Higgs mechanism for an abelian gauge symmetry. The Higgs mechanism works analogously for a non-abelian gauge symmetry: For each broken generator of the gauge group, the associated gauge field acquires a mass. Goldstone bosons never appear for broken gauge symmetries.

The configuration of massive scalar fields very much depends on the particular model. In principle, spontaneous symmetry breaking can be achieved without additional massive scalar fields. For example, sigma models on group manifolds and coset spaces thereof are particular implementations. However, these models typically have non-polynomial interactions and therefore are non-renormalisable. Spontaneous symmetry breaking in models with polynomial interactions requires the presence of at least one massive scalar field.[3] In that sense, the Higgs mechanism implies a massive scalar field, the so-called Higgs particle.

Note that the number of independent particle modes is preserved by the Higgs mechanism: We start with some scalar fields and some massless vector fields. For each broken generator of the gauge algebra one of the scalar fields disappears in favour of a massive vector field. A massless vector field has two on-shell degrees of freedom whereas a massive one has three. The additional degree of freedom originates from the scalar field; the gauge field eats a scalar field to become massive.[4]

Renormalisation. We could now proceed and quantise the above Lagrangian. The model includes a massive vector field whose propagator is

$$G^{\mu\nu}(p) = \frac{\eta^{\mu\nu} + p^\mu p^\nu / m^2}{p^2 + m^2} \,. \tag{7.26}$$

The trouble with this massive vector field propagator is that for large momenta p it approaches a finite value rather than decaying quadratically. This makes the UV-behaviour far worse than for an ordinary gauge field. Moreover, the appearance of the inverse of the mass spoils power counting. This generically renders the model non-renormalisable.

Now the above Lagrangian is very special, and the UV-divergences in fact cancel such that the model remains renormalisable. This nice behaviour evidently is due to gauge symmetry in the original model: One could argue that the model with spontaneous symmetry breaking is essentially just the same as the model where gauge symmetry is intact; the choice of the ground state should make no difference concerning renormalisability.

More practically one can argue that renormalisability is a question of the UV. When momenta are reasonably large, the mass terms are insignificant, it does not matter whether $\mu^2 \ll |p^2|$ is positive or negative.

[3] In the above examples, this is the "magnitude" direction of the scalar potential.

[4] In fact, the scalar origin of the additional mode for the massive vector can be observed at very high energies: There the masses of particles are largely irrelevant and the massive vector decomposes into a massless vector and a scalar.

Gauge Fixing. A rigorous approach to show renormalisability explicitly is given by the so-called R_ξ-gauges. It is analogous to gauge fixing by the Faddeev–Popov method. Let us therefore discuss this method.

We return to the first parametrisation of the complex scalar field ϕ as

$$\phi(x) = \frac{1}{\sqrt{2}}\big(v + \eta(x) + \mathring{\imath}\rho(x)\big). \tag{7.27}$$

The momentum terms therefore read

$$\begin{aligned} -(D^\mu\phi)^*(D_\mu\phi) = &-\tfrac{1}{2}(\partial\eta)^2 - \tfrac{1}{2}(\partial\rho)^2 - \tfrac{1}{2}q^2v^2A^2 + qvA_\mu\partial^\mu\rho \\ &+ \text{interactions.} \end{aligned} \tag{7.28}$$

Now we need to take care of the off-diagonal quadratic term between A and ρ. Above we have used a gauge transformation to remove it. Here this is achieved by adding an extra term to the gauge fixing functional $G[A, \Omega]$

$$G = \partial^\mu A_\mu - \Omega - \xi q v \rho. \tag{7.29}$$

Next we insert a delta-functional together with a compensating functional determinant into the path integral

$$\int \mathrm{D}\alpha\, \mathrm{Det}[\delta G'/\delta\alpha]\Delta[G']. \tag{7.30}$$

We then eliminate the delta-functional by integrating over the auxiliary field Ω with Gaussian weight $S_\Omega = -\frac{1}{2}\xi^{-1}\int \mathrm{d}x^4\, \Omega^2$. This leads to a gauge fixing term

$$\mathcal{L}_{\mathrm{gf}} = -\tfrac{1}{2}\xi^{-1}(\partial\cdot A)^2 - \tfrac{1}{2}\xi q^2v^2\rho^2 + qv\partial^\mu A_\mu\rho. \tag{7.31}$$

The argument of the functional determinant is given by the gauge variation of G

$$\delta G = \partial^2\delta\alpha - \xi q^2 v(v+\eta)\delta\alpha. \tag{7.32}$$

As opposed to the case of plain QED, this does depend on the fields, and therefore cannot be disregarded. As for the non-abelian gauge field, we replace the functional determinant by the introduction of the fermionic ghost fields $C, \tilde{C}$. Their Lagrangian reads

$$\mathcal{L}_{\mathrm{gh}} = \mathring{\imath}\tilde{C}(\partial^2 - \xi q^2v^2 - \xi q^2 v\eta)C. \tag{7.33}$$

Collecting the various Lagrangians, we find the kinetic terms

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{\mathrm{SQED}} + \mathcal{L}_{\mathrm{gf}} + \mathcal{L}_{\mathrm{gh}} \\ &= -\tfrac{1}{2}(\partial^\mu A^\nu)(\partial_\mu A_\nu) + \tfrac{1}{2}(1-\xi^{-1})(\partial\cdot A)^2 - \tfrac{1}{2}q^2v^2A^2 \\ &\quad - \tfrac{1}{2}(\partial\eta)^2 - \tfrac{1}{4}\lambda v^2\eta^2 \\ &\quad - \tfrac{1}{2}(\partial\rho)^2 - \tfrac{1}{2}\xi q^2v^2\rho^2 - \mathring{\imath}\partial^\mu\tilde{C}\partial_\mu C - \mathring{\imath}\xi q^2v^2\tilde{C}C \\ &\quad + \text{interactions.} \end{aligned} \tag{7.34}$$

The gauge field needs further scrutiny, let us write its propagator

$$G^{\mu\nu}(p) = \frac{1}{p^2 + q^2v^2}\left(\eta^{\mu\nu} - (1-\xi)\,\frac{p^\mu p^\nu}{p^2 + \xi q^2 v^2}\right). \tag{7.35}$$

This propagator has several interesting features:

- For large momenta it decays quadratically.
- No mass scale appears in the denominator on its own.
- For $\xi \to \infty$ it approaches the propagator of a massive vector field.
- The 3 transverse modes have mass qv.
- The longitudinal mode has mass $\sqrt{\xi}qv$ as can be seen from the expression

$$p_\mu G^{\mu\nu}(p) = \frac{\xi p^\nu}{p^2 + \xi q^2 v^2}\,. \tag{7.36}$$

This means we have the following particles:

- a massive vector particle A_μ of mass qv.
- a scalar Higgs particle η with mass $\sqrt{\lambda/2}\,v$,
- the longitudinal mode of A_μ of mass $\sqrt{\xi}qv$.
- a scalar particle ρ of mass $\sqrt{\xi}qv$,
- a pair of ghost and anti-ghost particles $C, \tilde{C}$ of mass $\sqrt{\xi}qv$,

The former two are the physical particles of the model. The latter four are unphysical excitations, two bosons and two fermions, all with the same mass. Just as in the case of Faddeev–Popov gauge fixing, their contributions conspire to cancel out from all physical processes.

Moreover it can be shown that all physical processes are independent of the gauge fixing parameter ξ. Several particular values of ξ are relevant:

- In the limit of $\xi \to \infty$ we recover the second parametrisation of the Lagrangian discussed above. In that case all the additional unphysical excitations become infinitely massive and decouple from physics. In the limit $\xi \to \infty$ the model is perfectly unitary. However, the standard massive vector field renders the theory superficially non-renormalisable.
- For finite values of ξ the model is perfectly renormalisable. However, it contains 4 additional excitation modes which are partially non-unitary. Unitarity is recovered only in the sector of physical states. The discussion is analogous to the case of unbroken gauge symmetry.
- The value $\xi = 1$ is convenient for computations since the unphysical particles have the same mass as the vector field, and since the vector field propagator has a plain constant numerator.

7.3 Electroweak Model

At the very end of this course, we discuss the Glashow–Weinberg–Salam model of the electroweak interactions along with its symmetry breaking mechanism.

Electroweak Interactions. A basic process of the electroweak interactions is beta-decay in which a neutron decays into a proton, an electron and an anti-neutrino

$$n \to p^+ + e^- + \bar{\nu}. \tag{7.37}$$

At the level of constituent quarks the process can be understood as the transition of a down into an up quark

$$d^{-1/3} \to u^{+2/3} + e^- + \bar{\nu}. \tag{7.38}$$

These are four fundamental fields, so the interaction should somehow be modelled by the electroweak theory. However, an interaction of 4 spinor fields has mass dimension at least 6 and consequently leads to a non-renormalisable model.

The resolution to this problem is to couple two pairs of spinors by an intermediate field W^-

$$d^{-1/3} \to u^{+2/3} + W^- \to u^{+2/3} + e^- + \bar{\nu}. \tag{7.39}$$

Some characteristics of this interaction are as follows:

- For low energies, the decay is largely independent of the involved particle momenta. This implies that the intermediate particle is very massive, so that its intermediate propagator is practically constant.
- The interactions between spinors and the field $W^\pm$ are via the vector and the axial couplings. The coupling for both types of interactions turns out to be equal leading to maximal violation of parity in these processes.
- A renormalisable coupling to a vector field requires a gauge theory. Since the intermediate particles are massive, gauge symmetry must be spontaneously broken.
- The intermediate particle W^- is electrically charged, therefore the electroweak gauge group must be non-abelian. In fact, it is $\mathrm{SU}(2) \times \mathrm{U}(1)$ where the $\mathrm{U}(1)$ factor is required to obtain electromagnetism after symmetry breaking.[5]

Let us first discuss the symmetry breaking mechanism, and then turn to the quarks and leptons as spinor matter particles.

Spontaneous Symmetry Breaking. The electroweak theory is based on the gauge group $\mathrm{SU}(2) \times \mathrm{U}(1)$. We denote the associated gauge potentials by the 2×2 traceless hermitian matrix W_μ and the real field B_μ (which may act as a 2×2 matrix proportional to unity where needed). The coupling constants for these groups are denoted by g and g', respectively. We define the corresponding field strengths as

$$\begin{aligned} W_{\mu\nu} &= \partial_\mu W_\nu - \partial_\nu W_\mu - \mathring{\imath} g [W_\mu, W_\nu], \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu. \end{aligned} \tag{7.40}$$

The contribution to the Lagrangian reads

$$\mathcal{L}_{\text{gauge}} = -\tfrac{1}{2} \operatorname{tr} W^{\mu\nu} W_{\mu\nu} - \tfrac{1}{4} B^{\mu\nu} B_{\mu\nu}. \tag{7.41}$$

[5]The U(1) of electromagnetism is not identical to the U(1) factor of the electroweak theory since the latter does not couple to the charged vector fields.

Spontaneous symmetry breaking is achieved by a doublet of complex scalar fields denoted by H. It transforms in the (iso)spin-$^1/_2$ representation of SU(2) and has a (hyper)charge of $^1/_2$ of U(1). Hence, the covariant derivative of H is defined as

$$D_\mu H = \partial_\mu H - \mathring{i} g W_\mu H - \tfrac{\mathring{i}}{2} g' B_\mu H. \tag{7.42}$$

We design the scalar potential to have a global minimum at $|H|^2 = v^2/2$. The scalar contribution to the Lagrangian therefore reads

$$\mathcal{L}_{\text{scalar}} = -(D^\mu H)^\dagger (D_\mu H) - \tfrac{1}{2}\lambda\big(H^\dagger H - \tfrac{1}{2}v^2\big)^2. \tag{7.43}$$

We discuss spontaneous symmetry breaking, and we will use the unitary gauge because it describes the physics most immediately. Therefore we use the SU(2) gauge symmetry to rotate the scalar field to the configuration

$$H(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + \eta(x) \end{pmatrix}, \tag{7.44}$$

where $\eta(x)$ is a real field called the Higgs field. Let us furthermore parametrise the SU(2) gauge field as

$$W = \frac{1}{2}\begin{pmatrix} +W^0 & \sqrt{2}W^+ \\ \sqrt{2}W^- & -W^0 \end{pmatrix}. \tag{7.45}$$

We can now work out the masses for the various fields by extracting the quadratic contributions to the Lagrangian

$$\begin{aligned} \mathcal{L} = &-\tfrac{1}{2}W^{-,\mu\nu}W^+_{\mu\nu} - \tfrac{1}{4}W^{0,\mu\nu}W^0_{\mu\nu} - \tfrac{1}{4}B^{\mu\nu}B_{\mu\nu} \\ &- \tfrac{1}{4}g^2v^2W^{-,\mu}W^+_\mu - \tfrac{1}{8}v^2(gW^{0,\mu} - g'B^\mu)(gW^0_\mu - g'B_\mu) \\ &- \tfrac{1}{2}(\partial\eta)^2 - \tfrac{1}{2}\lambda v^2\eta^2 + \text{interactions}. \end{aligned} \tag{7.46}$$

Now we observe mixing between the fields W^0 and B. This can be resolved by a rotation of the basis of gauge fields[6]

$$\begin{pmatrix} A \\ Z^0 \end{pmatrix} = \frac{1}{\sqrt{g^2 + g'^2}}\begin{pmatrix} g & g' \\ -g' & g \end{pmatrix}\begin{pmatrix} B \\ W^0 \end{pmatrix}. \tag{7.47}$$

The coupling constants are commonly parametrised using the Weinberg angle θ_{W}

$$g = g_{\text{EW}}\cos\theta_{\text{W}}, \qquad g' = g_{\text{EW}}\sin\theta_{\text{W}}. \tag{7.48}$$

After the rotation we find

$$\begin{aligned} \mathcal{L} = &-\tfrac{1}{2}W^{-,\mu\nu}W^+_{\mu\nu} - \tfrac{1}{4}Z^{0,\mu\nu}Z^0_{\mu\nu} - \tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} \\ &- \tfrac{1}{4}g^2v^2W^{-,\mu}W^+_\mu - \tfrac{1}{8}v^2g^2_{\text{EW}}Z^{0,\mu}Z^0_\mu \\ &- \tfrac{1}{2}(\partial\eta)^2 - \tfrac{1}{2}\lambda v^2\eta^2 + \text{interactions}. \end{aligned} \tag{7.49}$$

[6]The orthogonality of the transformation preserves the structure of the derivative terms.

Now the mass terms are easily read off

$$m_W = \tfrac{1}{2} g_{\mathrm{EW}} \cos(\theta_{\mathrm{W}})\, v, \qquad m_Z = \tfrac{1}{2} g_{\mathrm{EW}}\, v, \qquad m_\eta = \sqrt{\lambda}\, v. \tag{7.50}$$

The field A turns out to be massless.

The masslessness of A is due to the fact that a U(1) subgroup of the gauge symmetry is unbroken. This subgroup is a mixture of the original U(1) and one component of the SU(2). It represents the electromagnetic field. It makes sense to inspect the covariant derivative of the Higgs field

$$\begin{aligned} gW + \tfrac{1}{2} g' B = \frac{g_{\mathrm{EW}}}{2} \begin{pmatrix} \cos(2\theta_{\mathrm{W}}) Z^0 & \sqrt{2}\cos(\theta_{\mathrm{W}}) W^+ \\ \sqrt{2}\cos(\theta_{\mathrm{W}}) W^- & Z^0 \end{pmatrix} \\ + \frac{g_{\mathrm{EW}}}{2} \sin(2\theta_{\mathrm{W}}) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned} \tag{7.51}$$

This tells us two things:

- The (lower) component of the scalar field representing the Higgs particle is uncharged. The (upper) component which has been gauged away has charge $+1$.[7]

$$H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix}. \tag{7.52}$$

- In general, the electromagnetic charge is the sum of one component of the SU(2) isospin (in our choice of vacuum it is measured by $\frac{1}{2}\sigma_3$ for a isospin $1/2$ representation) and of the hypercharge.
- The coupling strength of the electromagnetic field is given by

$$q = \tfrac{1}{2} g_{\mathrm{EW}} \sin(2\theta_{\mathrm{W}}). \tag{7.53}$$

Fermions and Masses. A peculiar feature of the electroweak interactions for the fermionic matter fields is that they are chiral; they are not invariant under spacetime parity transformations. Chiral couplings of spinors to a vector field are modelled by the axial current $\bar{\psi}\gamma^5\gamma^\mu\psi$ instead of the more familiar vector current $\bar{\psi}\gamma^\mu\psi$. The matrix γ^5 receives an extra sign under parity transformations because it is based on the totally antisymmetric tensor $\varepsilon^{\mu\nu\rho\sigma}$.

Superficially one could combine vector and axial couplings. In nature, however, it was found that the vector and axial couplings have exactly the same strength. This finding is in line with theory: Couplings to vector fields should arise only by means of some covariant derivative. In other words, the form of the Dirac equation typically predicts vector couplings. However, we can split up the Dirac spinor into two chiral spinors,

$$\psi = \psi_{\mathrm{L}} + \psi_{\mathrm{R}}, \qquad \psi_{\mathrm{L/R}} = \tfrac{1}{2}(1 \pm \gamma^5)\psi. \tag{7.54}$$

[7] These statements are independent of the choice of ground state for the field H. The latter defines which components of the gauge group correspond to the electromagnetic field. The Higgs field must be uncharged, since otherwise its vacuum expectation value would attribute an electromagnetic charge to the ground state thus breaking the symmetry.

Then the derivative terms for the two components decouple entirely[8] [9]

$$\bar{\psi}\gamma^\mu\partial_\mu\psi = \bar{\psi}_{\mathrm{L}}\gamma^\mu\partial_\mu\psi_{\mathrm{L}} + \bar{\psi}_{\mathrm{R}}\gamma^\mu\partial_\mu\psi_{\mathrm{R}}. \tag{7.55}$$

If we let the left and right components transform in different representations of the gauge group, we naturally obtain chiral couplings

$$\begin{aligned}&\bar{\psi}_{\mathrm{L}}\gamma^\mu D^{\mathrm{L}}_\mu\psi_{\mathrm{L}} + \bar{\psi}_{\mathrm{R}}\gamma^\mu D^{\mathrm{R}}_\mu\psi_{\mathrm{R}}\\ &= \bar{\psi}\gamma^\mu\big(\partial_\mu + \tfrac{1}{2}gA^a_\mu(T^{\mathrm{L}}_a + T^{\mathrm{R}}_a)\big)\psi - \tfrac{1}{2}gA^a_\mu\bar{\psi}\gamma^5\gamma^\mu(T^{\mathrm{R}}_a - T^{\mathrm{L}}_a)\psi.\end{aligned} \tag{7.56}$$

Since the choice of representation in compact groups is typically discrete, there are only discrete choices for the vector and axial couplings. For example, when one of the representations is trivial, the vector and axial couplings have equal magnitude.

In the electroweak model, the representations for the spinor fields follow a particular pattern: The left-handed components transform in isospin-$^1/_2$ representations of SU(2) which unite two types of fields: electrons and neutrinos as well as up and down quarks. Conversely, the right-handed components form (two) singlets. The U(1) hypercharges are arranged such that the electromagnetic charges fit the observed values. The following tables list the representations of the particles in terms of third component of isospin I_3, hypercharge Y as well as multiplicity in terms of colour within QCD

field	I_3	Y	C
$L_{\mathrm{L}} = \begin{pmatrix}\nu_{\mathrm{L}}\\ e_{\mathrm{L}}\end{pmatrix}$	$\begin{pmatrix}+1/2\\ -1/2\end{pmatrix}$	$-1/2$	1
$Q_{\mathrm{L}} = \begin{pmatrix}u_{\mathrm{L}}\\ d_{\mathrm{L}}\end{pmatrix}$	$\begin{pmatrix}+1/2\\ -1/2\end{pmatrix}$	$+1/6$	3

field	I_3	Y	C
ν_{R}	0	0	1
e_{R}	0	-1	1
u_{R}	0	$+2/3$	3
d_{R}	0	$-1/3$	3

(7.57)

We shall consider only the first of the three generations of spinor fields. The other two families have analogous properties in all aspects that we will discuss.

An interesting result of symmetry breaking is that while the coupling to all vector fields of SU(2) × U(1) violates parity strongly, the coupling to the resulting electromagnetic field is via the usual vector current $\bar{\psi}\gamma^\mu\psi$. This may appear surprising, but it is a consequence of current conservation: After symmetry breaking, the spinor fields are massive (see below) which violates conservation of the axial currents, but a remaining gauge symmetry can only couple to conserved currents.

[8] Note that the complex conjugate of a right-handed spinor is left-handed. Therefore, every model can be represented by left-handed spinors only. The maximal symmetry acting on N Dirac spinors therefore is U($2N$). However, in many models this group splits into $\mathrm{U}(N)_{\mathrm{L}} \times \mathrm{U}(N)_{\mathrm{R}}$ (left/right or vector/axial) by mass terms or other interactions. For example in the standard model, the complex conjugate of a right-handed quark transforming as a 3 of SU(3) is a left-handed quark transforming in the conjugate representation $\bar{3}$ of SU(3). The latter is different from what is commonly called the left-handed quarks which transform as a 3 of SU(3).

[9] In dimensional regularisation it is difficult to represent γ^5, and it should be avoided in the kinetic terms. A single chiral spinor should be supplemented by a spinor of opposite chirality which does not interact with any other field. In that way all matrices γ^5 can be shifted to the interactions.

A consequence of unequal representations for the chiral components is that the usual Dirac mass term is not allowed

$$\bar{\psi}\psi = \bar{\psi}_{\rm R}\psi_{\rm L} + \bar{\psi}_{\rm L}\psi_{\rm R}. \tag{7.58}$$

For instance, it would break conservation of axial currents, and thus violate the gauge symmetry. There is another type of mass term for Dirac fields, called a Majorana mass, but it is allowed only for real representations. All representations in the standard model, however, are complex, except for a right-handed neutrino which is a singlet.[10]

Just like the vector particles, the fermions can only acquire masses by means of the Higgs field which has the following charges:

field	I_3	Y	C
$H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix}$	$\begin{pmatrix} +1/2 \\ -1/2 \end{pmatrix}$	$+1/2$	1

(7.59)

The Higgs field can couple via Yukawa interactions to the spinor fields. There are 4 gauge-invariant couplings

$$\bar{\nu}_{\rm R} H^{\intercal} \varepsilon L_{\rm L}, \qquad \bar{e}_{\rm R} H^{\dagger} L_{\rm L}, \qquad \bar{u}_{\rm R} H^{\intercal} \varepsilon Q_{\rm L}, \qquad \bar{d}_{\rm R} H^{\dagger} Q_{\rm L}, \tag{7.60}$$

where ε is the 2×2 antisymmetric matrix. The vacuum expectation value of the Higgs field then generates a mass term for each of the particle species: neutrinos, electrons, up and down quarks.[11]

Anomalies. The electroweak theory involves chiral couplings of a gauge field to fermions. This interaction is potentially dangerous since it can lead to gauge anomalies. However, the charges of the spinors are arranged in a particular pattern that avoids the anomalies.

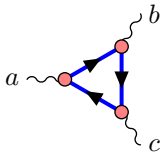
field	I_3	Y	C
$L_{\rm L} = \begin{pmatrix} \nu_{\rm L} \\ e_{\rm L} \end{pmatrix}$	$\begin{pmatrix} +1/2 \\ -1/2 \end{pmatrix}$	$-1/2$	1
$Q_{\rm L} = \begin{pmatrix} u_{\rm L} \\ d_{\rm L} \end{pmatrix}$	$\begin{pmatrix} +1/2 \\ -1/2 \end{pmatrix}$	$+1/6$	3

field	I_3	Y	C
$\nu_{\rm R}$	0	0	1
$e_{\rm R}$	0	-1	1
$u_{\rm R}$	0	$+2/3$	3
$d_{\rm R}$	0	$-1/3$	3

(7.61)

Let us briefly discuss these consistency conditions.

Chiral anomalies in four dimensions can arise only for chiral couplings to spinors, i.e. for the electroweak groups SU(2) and U(1). The anomaly originates from a triangle diagram



$$\longrightarrow \quad d^{abc} \tag{7.62}$$

[10] In fact such a mass term is appealing for the see-saw mechanism which can explain the smallness of neutrino masses.

[11] Since the Yukawa couplings determine the masses of spinor fields the coupling of matter to the Higgs field is proportional to the mass. Therefore the Higgs field strongly couples to the top quark which is by far the most massive spinor field.

which is proportional to some symmetric structure d^{abc} of the three involved gauge fields. However, there is no anomaly for SU(2) alone because there are no symmetric structure constants d^{abc} in this group. The symmetric structure d^{abc} trivially exists only for the abelian group U(1). Moreover, there are anomalies for combined groups. Here, one of the indices of d^{abc} belongs to U(1), the other two to the symmetric invariant form k^{bc} of SU(2). These mixed anomalies can arise for $\mathrm{U}(1) \times \mathrm{SU}(2)$ and for $\mathrm{U}(1) \times \mathrm{SU}(3)$. Finally, there a mixed anomaly between U(1) and gravity. If any of these anomalies is realised, one of the involved gauge symmetries would be badly broken in the quantum theory.

Consider first the pure U(1) anomaly. We have to sum over all spinors propagating in the loop. At each coupling a factor of the hypercharge is inserted. Hence the overall contribution to the anomaly is proportional to

$$A_{\mathrm{U}(1)} \sim \mathrm{tr}_{\mathrm{L}}\, Y^3 - \mathrm{tr}_{\mathrm{R}}\, Y^3. \tag{7.63}$$

This combination surprisingly equals zero

$$2 \cdot 1(-1/2)^3 + 2 \cdot 3(+1/6)^3 - (0)^3 - (-1)^3 - 3(+2/3)^3 - 3(-1/3)^3 = 0. \tag{7.64}$$

The next anomaly mixes U(1) and SU(2). Only the left-handed lepton and quark doublets couple to SU(2), and they couple with the same strength. The third coupling is the hypercharge, hence the anomaly is proportional to

$$A_{\mathrm{U}(1)\times\mathrm{SU}(2)} \sim \mathrm{tr}_{\mathrm{L}}\, Y. \tag{7.65}$$

Here we find zero

$$(-1/2) + 3(+1/6) = 0. \tag{7.66}$$

There is a mixed anomaly of U(1) and the SU(3) of quantum chromodynamics. This one involves only the quarks. Since the quarks all transform in the same fundamental representation, we only need to sum over the hypercharges

$$A_{\mathrm{U}(1)\times\mathrm{SU}(3)} \sim \mathrm{tr}_{\mathrm{Q,L}}\, Y - \mathrm{tr}_{\mathrm{Q,R}}\, Y. \tag{7.67}$$

Again, this combination equals zero

$$2(+1/6) - (+2/3) - (-1/3) = 0. \tag{7.68}$$

Finally, there is a potential mixed gravitational anomaly. Gravity is not part of the electroweak theory or the standard model, so we need not pay attention to this anomaly. Nevertheless, let us investigate it. Gravity couples to all particles with equal strength, so here we find

$$A_{\mathrm{U}(1)\times\mathrm{gravity}} \sim \mathrm{tr}_{\mathrm{L}}\, Y - \mathrm{tr}_{\mathrm{R}}\, Y. \tag{7.69}$$

This equals

$$2(-1/2) + 2 \cdot 3(-1/2) - (0) - (-1) - 3(+2/3) - 3(-1/3) = 0. \tag{7.70}$$

It is interesting to observe that there is no gravitational anomaly. Hence, the electroweak theory can be consistently coupled to non-trivial gravitational backgrounds or even to quantum gravity (as far as that makes sense on its own).

The anomaly cancellations are rather curious. They could be interpreted in several ways:

- The structure of the standard model is actually constrained much more than it may seem at first sight. In particular, the fact that there are three colours for the strong nuclear interactions does play a significant role.[12]
- The standard model may be embedded into some theory with a larger symmetry group at higher energies, a so-called grand unified theory, where the anomaly cancellations are more evident.

[12]In fact, the required gauge invariance of Yukawa couplings implies several anomaly cancellations. Therefore, the cancellation of mixed anomalies is a consequence of the quarks and leptons being massive.

Schedule of Lectures

The following table lists the locations in the text at the beginning of each lecture:

17.02. 1 [20]:	0.	Overview
17.02. 2 [54]:	1.1.	... § Multiple Slits.
20.02. 1 [42]:	1.2.	... § Classical Evolution.
20.02. 2 [38]:	1.2.	... § Path Integral in Position Space.
24.02. 1 [44]:	1.4.	... § Time Ordering.
24.02. 2 [50]:	2.2.	Sources and Generating Functional
03.03. 1 [42]:	2.3.	Fermionic Integrals
03.03. 2 [51]:	2.4.	Interactions
06.03. 1 [24]:	2.4.	... § Loop Counting.
06.03. 2 [62]:	2.5.	Further Generating Functionals
10.03. 1 [46]:	2.5.	... § Effective Action.
10.03. 2 [47]:	3.	Yang–Mills Theory
17.03. 1 [43]:	3.2.	Abelian Quantisation Revisited
17.03. 2 [49]:	3.3.	Yang–Mills Quantisation
20.03. 1 [44]:	3.4.	BRST Symmetry
20.03. 2 [44]:	3.4.	... § S-Matrix Consistency.
24.03. 1 [45]:	3.5.	... § Gauge Sector.
24.03. 2 [48]:	3.6.	... § Large-N Scaling.
27.03. 1 [46]:	4.	Renormalisation
27.03. 2 [46]:	4.1.	... § Introduction of a Scale.
31.03. 1 [50]:	4.2.	... § Counterterms.
14.04. 1 [44]:	4.3.	Renormalisation Flow
14.04. 2 [47]:	4.3.	... § Dimensional Regularisation.
17.04. 1 [41]:	5.	Lie Algebra
17.04. 2 [47]:	5.2.	... § Multiplication.
21.04. 1 [40]:	5.3.	Representations
21.04. 2 [46]:	5.3.	... § Tensor Products.
24.04. 1 [47]:	5.4.	... § Structure Constants.
24.04. 2 [38]:	5.5.	... § Matter Couplings.
28.04. 1 [57]:	5.5.	... § One-Loop Renormalisation in Yang–Mills Theory.
28.04. 2 [44]:	5.6.	... § Tensor Product Representations.
05.05. 1 [41]:	6.	Quantum Symmetries
05.05. 2 [46]:	6.2.	Slavnov–Taylor Identities
08.05. 1 [42]:	6.3.	... § Momentum Space.
08.05. 2 [50]:	6.4.	... § Axial Anomaly in Massless QED.
12.05. 1 [47]:	7.	Spontaneous Symmetry Breaking
12.05. 2 [42]:	7.1.	... § Goldstone's Theorem.
19.05. 1 [43]:	7.2.	Breaking of Gauge Symmetries

19.05. 2 [46]: 7.2. ... § Gauge Fixing.
26.05. 1 [40]: 7.3. ... § Spontaneous Symmetry Breaking.
26.05. 2 [46]: 7.3. ... § Fermions and Masses.