

Quantum Field Theory II

Lecture Notes

ETH Zurich, FS13

PROF. N. BEISERT

© 2013 Niklas Beisert, ETH Zurich

This document as well as its parts is protected by copyright.
Reproduction of any part in any form without prior written
consent of the author is permissible only for private,
scientific and non-commercial use.

Contents

0	Overview	4
0.1	Prerequisites	5
0.2	Contents	5
0.3	References	5
1	Path Integral for Quantum Mechanics	1.1
1.1	Motivation	1.1
1.2	Path Integral for Transition Amplitude	1.5
1.3	Free Particle	1.10
1.4	Operator Insertions	1.11
2	Path Integral for Fields	2.1
2.1	Time-Ordered Correlators	2.1
2.2	Sources and Generating Functional	2.2
2.3	Fermionic Integrals	2.6
2.4	Interactions	2.9
2.5	Further Generating Functionals	2.14
3	Lie Algebra	3.1
3.1	Lie Groups	3.1
3.2	Lie Algebras	3.2
3.3	Enveloping Algebras	3.5
3.4	Representations	3.6
3.5	Invariants	3.10
3.6	Unitary and Other Algebras	3.13
4	Yang–Mills Theory	4.1
4.1	Classical Gauge Theory	4.1
4.2	Abelian Quantisation Revisited	4.6
4.3	Yang–Mills Quantisation	4.9
4.4	Feynman Rules	4.13
4.5	BRST Symmetry	4.16
5	Renormalisation	5.1
5.1	Dimensional Regularisation	5.1
5.2	Renormalisation	5.5
5.3	Renormalisation Flow	5.13
6	Quantum Symmetries	6.1
6.1	Schwinger–Dyson Equations	6.1
6.2	Slavnov–Taylor Identities	6.3
6.3	Ward–Takahashi Identity	6.5
6.4	Anomalies	6.8
7	Spontaneous Symmetry Breaking	7.1
7.1	Breaking of Global Symmetries	7.1

7.2	Breaking of Gauge Symmetries	7.5
7.3	Electroweak Model	7.9

0 Overview

After having learned the basic concepts of Quantum Field Theory in QFT I, 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
- non-abelian gauge theory
- renormalisation
- symmetries
- spontaneous symmetry breaking

More concretely, the topics can be explained as follows:

Path Integral. 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.

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. Naive 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

1. Path Integral for Quantum Mechanics (3 lectures)
2. Path Integral for Fields (7 lectures)
3. Lie Algebra (5 lectures)
4. Yang–Mills Theory (6 lectures)
5. Renormalisation (6 lectures)
6. Quantum Symmetries (4 lectures)
7. Spontaneous Symmetry Breaking (5 lectures)

Indicated are the approximate number of 45-minute lectures. Altogether, the course consists of 37 lectures which includes one overview lecture.

0.3 References

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

- 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. Srednicki, “Quantum Field Theory”, Cambridge University Press (2007)
- M. Kaku, “Quantum Field Theory”, Oxford University Press (1993)
- online: M. Gaberdiel, A. Gehrman-De Ridder, “Quantum Field Theory II”, lecture notes,
<http://www.itp.phys.ethz.ch/research/qftstrings/archive/11FSQFT2>

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 framework to formulate quantum theories. It was developed mainly by Dirac (1933) and Feynman (1948). It is particularly useful for relativistic quantum field theory.

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 long 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 to the perturbative regime).

- It is a different formulation and interpretation of quantum theories.

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 shortcomings of the path integral which are good reasons to understand the canonical framework:

- 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.



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.



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

¹One 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.

Now we put another obstacle with slits at a different location.



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 obstacle. 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 $d_k(x)$.

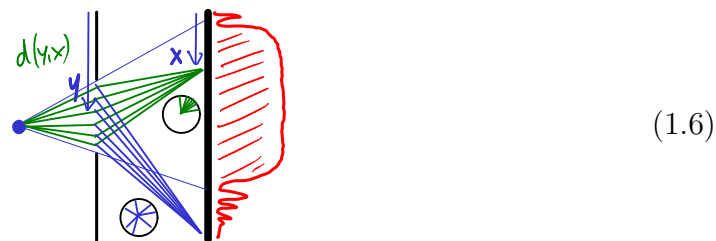


Neglecting the decrease of amplitude for circular waves, the intensity is given by

$$I(x) = |A(x)|^2, \quad A(x) = \sum_k \exp(2\pi i d_k(x)/\lambda). \quad (1.5)$$

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 slit of size comparable to λ . In this case, the wave can pass at every point within the slit.



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

$$A(x) = \int_{\min}^{\max} dy \exp(2\pi i d(y, x)/\lambda). \quad (1.7)$$

²For uncorrelated phases one would not obtain interference.

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 should 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 could compute the interference pattern for an arbitrary array of obstacles. We could for instance put an obstacle at some distance which blocks no waves at all. We would have to integrate over all intermediate positions at this non-obstacle, but the result should still be correct.

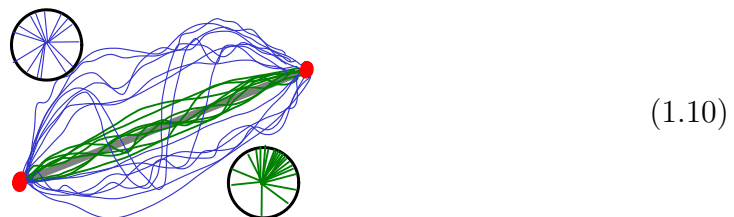
We can now be even more extreme, and put non-obstacles at many different locations.



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, not just the straight ones.



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 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.



Thus they interfere destructively, and will effectively not contribute to the interference pattern no matter if they hit the 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 Path Integral for Transition Amplitude

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.

Start with a classical Hamiltonian function $H(q, p)$. Quantise canonically to get a corresponding Hamiltonian operator \hat{H} . Up to ordering issues of \hat{q}, \hat{p} we thus have

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

Transition Amplitude. We want to compute the transition amplitude $A_{f,i}$ between position q_i at time t_i and position q_f at time t_f ³

$$A_{f,i} = \langle q_f, t_f | q_i, t_i \rangle = \langle q_f | \hat{U}(t_f, t_i) | q_i \rangle \quad (1.12)$$

where $\hat{U}(t_f, t_i)$ is the time evolution operator. In the case of a time-independent Hamiltonian it reads⁴

$$\hat{U}(t_f, t_i) = \exp(-i\hbar^{-1}(t_f - t_i)\hat{H}). \quad (1.13)$$

We want to find an expression for $A_{f,i}$ 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 $t_f > t_k > t_i$ by using the group property of time evolution

$$\hat{U}(t_f, t_i) = \hat{U}(t_f, t_k)\hat{U}(t_k, t_i). \quad (1.14)$$

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

$$1 = \int dq_k |q_k\rangle\langle q_k|. \quad (1.15)$$

Altogether we obtain an identity

$$A_{f,i} = \int dq_k A_{f,k} A_{k,i}. \quad (1.16)$$

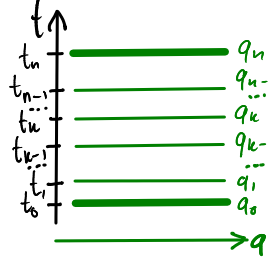
³The symbol $|q\rangle$ denotes a state localised at position q . The symbol $|q, t\rangle := \hat{U}(t_{\text{ref}}, t)|q\rangle$ denotes a state in the Heisenberg picture which is perfectly localised at position q and time t transported back to the reference time slice at time t_{ref} .

⁴Our derivation works perfectly well for a time-dependent Hamiltonian $\hat{H}(t)$ or $H(q, p, t)$, however, we will not make this time-dependence explicit.

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

$$A_{f,i} = \int dq_{n-1} \dots dq_1 A_{f,n-1} \dots A_{1,i} = \int \prod_{k=1}^{n-1} dq_k \prod_{k=1}^n A_{k,k-1}, \quad (1.17)$$

where we define $q_0 := q_i$, $t_0 := t_i$ and $q_n := q_f$, $t_n := t_f$.



$$(1.18)$$

Consider now an elementary transition amplitude

$$A_{k,k-1} = \langle q_k | \hat{U}(t_k, t_{k-1}) | q_{k-1} \rangle. \quad (1.19)$$

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

$$\hat{U}(t_k, t_{k-1}) \simeq \exp(-i\hbar^{-1}(t_k - t_{k-1})\hat{H}) \simeq 1 - i\hbar^{-1}(t_k - t_{k-1})\hat{H}. \quad (1.20)$$

Here it is crucial that the exponent is sufficiently small so that only the first expansion term is relevant at all. In other words we should evaluate

$$\langle q_k | \hat{H} | q_{k-1} \rangle = \langle q_k | H(\hat{q}, \hat{p}) | q_{k-1} \rangle. \quad (1.21)$$

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

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

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

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

Conventionally one orders the factors such that \hat{q} evaluates to the average $\bar{q}_{k-1,k} := \frac{1}{2}q_{k-1} + \frac{1}{2}q_k$

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

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 .

Note that it is crucial that only a single factor of \hat{H} appears. In higher powers of \hat{H} the non-trivial operator ordering would prevent us from replacing \hat{q} by some

average value $\bar{q}_{k-1,k}$. Hence we need a sufficiently large number of intermediate time slices for a good approximation.

In order to evaluate the momentum operator, we insert a complete set of momentum eigenstates

$$1 = \int \frac{dp_k}{2\pi\hbar} |p_k\rangle\langle p_k|. \quad (1.25)$$

This yields

$$\langle q_{k+1}|\hat{H}|q_k\rangle = \int \frac{dp_k}{2\pi\hbar} \langle q_{k+1}|H(\bar{q}_{k-1,k}, \hat{p})|p_k\rangle\langle p_k|q_k\rangle \quad (1.26)$$

We substitute the Fourier exponent $\langle p|q\rangle = \exp(i\hbar^{-1}pq)$ and obtain an approximation for $A_{k,k-1}$

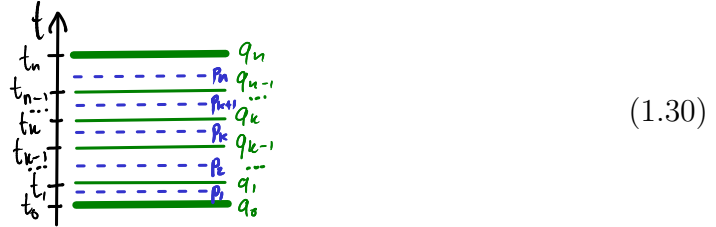
$$\int \frac{dp_k}{2\pi\hbar} \exp(-i\hbar^{-1}(t_k - t_{k-1})H(\bar{q}_{k-1,k}, p_k) + i\hbar^{-1}p_k(q_k - q_{k-1})). \quad (1.27)$$

Putting everything together we now have an integral expression for the transition amplitude $A_{n,0}$ in terms of the classical Hamiltonian

$$A_{f,i} \approx A_{n,0} := \int \prod_{k=1}^{n-1} dq_k \prod_{k=1}^n \frac{dp_k}{2\pi\hbar} \exp(i\hbar^{-1}S_{n,0}[q, p]), \quad (1.28)$$

with the phase of the exponential determined by the function

$$S_{n,0}[q, p] := \sum_{k=1}^n (-(t_k - t_{k-1})H(\bar{q}_{k-1,k}, p_k) + p_k(q_k - q_{k-1})). \quad (1.29)$$



Path Integral in Phase Space. The transition amplitude $A_{f,i}$ is approximated by the integral of $\exp(i\hbar^{-1}S_{n,0}[q, p])$ 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.

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

$$\int \frac{dq dp}{2\pi\hbar}. \quad (1.31)$$

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

$$A_{f,i} = \int Dq Dp \exp(i\hbar^{-1}S_{f,i}[q, p]), \quad (1.32)$$

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

$$S_{f,i}[q, p] = \int_{t_i}^{t_f} dt (p(t)\dot{q}(t) - H(q(t), p(t))) = \int_i^f (pdq - Hdt). \quad (1.33)$$

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

Comparing to the above discrete version, the term $dt p\dot{q} \approx p_k(q_k - q_{k-1})$ 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 $S_{f,i}[q, p]$ is exciting, it is precisely the action in phase space. Note that the associated Euler–Lagrange equations

$$0 = \frac{\delta S}{\delta q(t)} = -\dot{p}(t) - \frac{\partial H}{\partial q}(t), \quad 0 = \frac{\delta S}{\delta p(t)} = +\dot{q}(t) - \frac{\partial H}{\partial p}(t), \quad (1.34)$$

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 $S_{f,i}[q, p]/\hbar$ 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 $A_{f,i}$, keeps the initial and final positions fixed

$$q(t_i) = q_i, \quad q(t_f) = q_f, \quad (1.35)$$

whereas the momenta $p(t_i)$ and $p(t_f)$ are free. The path integral can also compute other quantities where the boundary conditions are specified differently.

Note that the integration measures Dq and Dp 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 $\pm i\epsilon$.

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

$$H(q, p) = \frac{p^2}{2M(q)} + pK(q) + V(q). \quad (1.36)$$

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 integral⁵

$$\int_{-\infty}^{+\infty} dp \exp\left(-\frac{1}{2}ap^2 + bp + c\right) = \sqrt{2\pi/a} \exp(b^2/2a + c). \quad (1.37)$$

We obtain an expression for the transition amplitude

$$A_{n,0} = \int \prod_{k=1}^{n-1} dq_k \prod_{k=1}^n \sqrt{\frac{M(\bar{q}_{k-1,k})}{2\pi\hbar i(t_k - t_{k-1})}} \exp(i\hbar^{-1}S_{n,0}[q]), \quad (1.38)$$

which depends on

$$S_{n,0}[q] = \sum_{k=1}^n (t_k - t_{k-1}) \left[-V(\bar{q}_{k-1,k}) + \frac{1}{2}M(\bar{q}_{k-1,k}) \left(\frac{q_k - q_{k-1}}{t_k - t_{k-1}} - K(\bar{q}_{k-1,k}) \right)^2 \right]. \quad (1.39)$$

Now the exact path integral for the transition amplitude reads

$$A_{f,i} = \int Dq \exp(i\hbar^{-1}S_{f,i}[q]), \quad (1.40)$$

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

$$S_{f,i}[q] = \int_0^n dt \left(\frac{1}{2}M(q)(\dot{q} - K(q))^2 - V(q) \right). \quad (1.41)$$

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 for two equivalent classical actions.

Note that the measure factor hidden in Dq is substantially more complicated compared to the combination $Dq Dp$. 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.

⁵The 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 by the usual assumptions of causality allow to attribute a small real positive part to a .

We have finally obtained an expression for the quantum transition amplitude $A_{f,i}$ 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.

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 particles is defined by the Hamiltonian

$$H(q, p) = \frac{p^2}{2m}. \quad (1.42)$$

The discretised action in phase space therefore reads

$$S_{n,0}[q, p] := \sum_{k=1}^n (-(t_k - t_{k-1})p_k^2/2m + p_k(q_k - q_{k-1})). \quad (1.43)$$

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 the delta-function $\delta(p_{k+1} - p_k)$ which trivialises one of the momentum integrals. Altogether we find

$$\begin{aligned} A_{n,0} &= \int \prod_{k=1}^{n-1} dq_k \prod_{k=1}^n \frac{dp_k}{2\pi\hbar} \exp(i\hbar^{-1}S_{n,0}[q, p]), \\ &= \int \frac{dp}{2\pi\hbar} \exp(i\hbar^{-1}(-(t_n - t_0)p^2/2m + p(q_n - q_0))) \\ &= \sqrt{\frac{m}{2\pi\hbar i(t_n - t_0)}} \exp\left(\frac{im(q_n - q_0)^2}{2\hbar(t_n - t_0)}\right). \end{aligned} \quad (1.44)$$

This is the correct transition amplitude $A_{f,i}$ for a free non-relativistic particle. Actually the number of intermediate steps n does not matter here because of the simplicity of the problem; usually the limit $n \rightarrow \infty$ is required.

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

$$S_{n,0}[q] = \frac{1}{2}m \sum_{k=1}^n \frac{(q_k - q_{k-1})^2}{t_k - t_{k-1}}. \quad (1.45)$$

It turns out that every integral over a q_k in

$$A_{n,0} = \int \prod_{k=1}^{n-1} dq_k \prod_{k=1}^n \sqrt{\frac{m}{2\pi\hbar i(t_k - t_{k-1})}} \exp(i\hbar^{-1}S_{n,0}[q]), \quad (1.46)$$

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

$$A_{n,0} = \sqrt{\frac{m}{2\pi\hbar i(t_n - t_0)}} \exp\left(\frac{im(q_n - q_0)^2}{2\hbar(t_n - t_0)}\right). \quad (1.47)$$

Gladly the two results agree.

1.4 Operator Insertions

We have found a way to express the transition amplitude $\langle q_f, t_f | q_i, t_i \rangle$ in terms of a path integral.

Single Insertion. The next important step is to be able to insert a quantum operator $\hat{O} = O(\hat{q}, \hat{p})$ at time t_o with $t_f > t_o > t_i$ into the transition amplitude⁶

$$A_{f,o,i} = \langle q_f, t_f | \hat{O}(t_o) | q_i, t_i \rangle = \langle q_f | \hat{U}(t_f, t_o) \hat{O} \hat{U}(t_o, t_i) | q_i \rangle. \quad (1.48)$$

As before, we evaluate this expression by inserting a complete set of position eigenstates

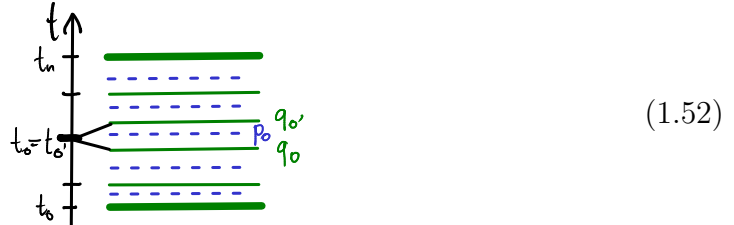
$$A_{f,o,i} = \int dq_o dq'_o A_{f,o} \langle q_o | \hat{O} | q'_o \rangle A_{o',i}. \quad (1.49)$$

Here we should replace the quantum operator \hat{O} by $O(\bar{q}_o, \hat{p})$ where \bar{q}_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

$$\begin{aligned} A_{f,o,i} &= \int dq_o dq'_o \frac{dp_o}{2\pi\hbar} A_{f,o} \langle q_o | O(\bar{q}_o, \hat{p}) | p_o \rangle \langle p_o | q'_o \rangle A_{o',i} \\ &= \int dq_o dq'_o \frac{dp_o}{2\pi\hbar} \exp(i\hbar^{-1} p_o (q_o - q'_o)) A_{f,o} O(\bar{q}_o, p_o) A_{o',i}. \end{aligned} \quad (1.50)$$

We observe that the alternating pattern of position and momentum integrations is preserved across the operator insertion. The complete discretised path integral with operator insertion therefore reads simply⁷

$$A_{f,o,i} \approx \int \prod_{k=1}^{n-1} dq_k \prod_{k=1}^n \frac{dp_k}{2\pi\hbar} O(\bar{q}_o, p_o) \exp(i\hbar^{-1} S_{n,0}[q, p]). \quad (1.51)$$



⁶Here we have used the Heisenberg and Schroedinger pictures of the operator related by $\hat{O}(t_o) = \hat{U}(t_{\text{ref}}, t_o) \hat{O}(t_o, t_{\text{ref}})$.

⁷Note that we assumed the two consecutive times $t_o, t_{o'}$ before and after the operator insertion as identified $t_o = t_{o'}$. This explains the absence of Hamiltonian contribution $-(t_o - t_{o'})H$ at the operator insertion.

As a continuous path integral it takes the form

$$A_{f,o,i} = \int Dq Dp O(q(t_o), p(t_o)) \exp(i\hbar^{-1} S_{f,i}[q, p]). \quad (1.53)$$

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 $t_f > t_n > \dots > t_2 > t_1 > t_i$. We insert them into the transition amplitude and obtain a path integral with n operator insertions

$$\begin{aligned} & \langle q_f, t_f | \hat{O}_n \dots \hat{O}_2 \hat{O}_1 | q_i, t_i \rangle. \\ & = \int Dq Dp O_1(t_1) O_2(t_2) \dots O_n(t_n) \exp(i\hbar^{-1} S_{f,i}[q, p]). \end{aligned} \quad (1.54)$$

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

$$\begin{aligned} & \int Dq Dp \prod_{k=1}^n O_k(t_k) \exp(i\hbar^{-1} S_{f,i}[q, p]). \\ & = \langle q_f, t_f | T(\hat{O}_1 \hat{O}_2 \dots \hat{O}_n) | q_i, t_i \rangle. \end{aligned} \quad (1.55)$$

This relation holds even for an arbitrary ordering of operator times t_k . Hence the path integral performs time-ordering automatically.⁸

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

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

$$[\hat{q}, \hat{p}] = i\hbar. \quad (1.56)$$

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 $i\hbar$, but that would amount to postulation rather than derivation.

The trick is to separate the times slightly:⁹

$$O(t) = q(t)p(t - \epsilon) - p(t + \epsilon)q(t). \quad (1.57)$$

⁸To compute expectation values of quantum operators which are not in proper time order with the path integral is more laborious, but could be handled manually.

⁹In 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 $i\hbar$).

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 between t_k and t_{k-1} . This fact is most evident in the expression for the discretised action

$$S_{n,0}[q, p] := \sum_{k=1}^n \left(-(t_k - t_{k-1}) H(\bar{q}_{k-1,k}, p_k) + p_k (q_k - q_{k-1}) \right). \quad (1.58)$$

The above operator is therefore discretised as follows

$$O = q_k p_k - p_{k+1} q_k. \quad (1.59)$$

We insert the operator into the discretised path integral for the transition amplitude $\langle q_f, t_f | \hat{O} | q_i, t_i \rangle$ of free non-relativistic particle. We can perform most integrations trivially as before

$$\begin{aligned} & \int \frac{dp_k}{2\pi\hbar} dq_k \frac{dp_{k+1}}{2\pi\hbar} q_k (p_k - p_{k+1}) \\ & \cdot \exp\left(-\frac{i}{2}\hbar^{-1}m^{-1}\left((t_n - t_k)p_{k+1}^2 + (t_k - t_0)p_k^2\right)\right) \\ & \cdot \exp\left(i\hbar^{-1}(p_{k+1}q_n - p_kq_0)\right) \exp\left(i\hbar^{-1}(p_k - p_{k+1})q_k\right). \end{aligned} \quad (1.60)$$

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

$$\begin{aligned} & \int \frac{dp_k}{2\pi\hbar} dp_{k+1} (p_k - p_{k+1}) \\ & \cdot \exp\left(-\frac{i}{2}\hbar^{-1}m^{-1}\left((t_n - t_k)p_{k+1}^2 + (t_k - t_0)p_k^2\right)\right) \\ & \cdot \exp\left(i\hbar^{-1}(p_{k+1}q_n - p_kq_0)\right) \left(-i\hbar \frac{\partial}{\partial p_k} \delta(p_{k+1} - p_k) \right). \end{aligned} \quad (1.61)$$

Now the only thing that protects the factor $(p_k - p_{k-1})$ from vanishing by means of the delta function $\delta(p_k - p_{k-1})$ is the derivative $\partial/\partial p_k$. We therefore perform a partial integration 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

$$\begin{aligned} & \int \frac{dp_k}{2\pi\hbar} dp_{k+1} \delta(p_{k+1} - p_k) \left(i\hbar \frac{\partial}{\partial p_k} (p_k - p_{k+1}) \right) \\ & \cdot \exp\left(i\hbar^{-1}\left(-\frac{1}{2}(t_n - t_0)p_k^2/m + p_k(q_n - q_0)\right)\right) \\ & = i\hbar \int \frac{dp}{2\pi\hbar} \exp\left(i\hbar^{-1}\left(-\frac{1}{2}(t_n - t_0)p^2/m + p(q_n - q_0)\right)\right). \end{aligned} \quad (1.62)$$

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

$$\langle q_f, t_f | [\hat{q}, \hat{p}] | q_i, t_i \rangle = i\hbar \langle q_f, t_f | q_i, t_i \rangle, \quad (1.63)$$

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 $O(\bar{q}_{k-1, k}, p_k)$. The precise choice of $\bar{q}_{k-1, k}$, 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 $[\hat{q}, \hat{p}] = i\hbar$ is simple enough, one can always add appropriate terms of $\mathcal{O}(\hbar)$ to $O(q, p)$ to make any given discretisation $O(\bar{q}_{k-1, k}, p_k)$ correspond to the desired quantum operator \hat{O} .

2 Path Integral for Fields

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)

$$\begin{aligned} & \langle \Psi_f, t_f | T(\hat{O}_1 \dots \hat{O}_n) | \Psi_i, t_i \rangle \\ &= \int D\Psi O_1[\Psi] \dots O_n[\Psi] \exp(iS_{f,i}[\Psi]). \end{aligned} \quad (2.1)$$

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

$$S_{f,i}[\Psi] := \int_{t_i}^{t_f} dt L[\Psi(t)] = \int_i^f d^D x \mathcal{L}(\Psi(x), \partial_\mu \Psi(x)). \quad (2.2)$$

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

$$\langle O_1 \dots O_n \rangle := \langle 0 | T(\hat{O}_1 \dots \hat{O}_n) | 0 \rangle. \quad (2.3)$$

To solve these problems, we can apply a trick we have learned in QFT I: A generic state such as $|\Psi_i, t_i\rangle$ 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 $t_{f,i} \rightarrow \pm\infty(1 - i\epsilon)$ and obtain a familiar expression for the time-order correlator

$$\langle O_1 \dots O_n \rangle = \frac{\int D\Psi O_1[\Psi] \dots O_n[\Psi] \exp(iS[\Psi])}{\int D\Psi \exp(iS[\Psi])}. \quad (2.4)$$

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

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

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 $D\Psi$ 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-ordered¹ time-ordered 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

$$\langle \Psi(x)\Psi(y) \rangle = \frac{\int D\Psi \Psi(x)\Psi(y) \exp(iS[\Psi])}{\int D\Psi \exp(iS[\Psi])}. \quad (2.6)$$

In a free theory it amounts to a Gaussian integral with prefactors. The evaluation is complicated by the fact that derivatives of Ψ appear in the action, and it is not immediately clear how to apply the standard methods to put additional factors from of the Gaussian exponent. Therefore, the integral should be discretised which often leads to a involved combinatorics.

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

$$Z[J] := \int D\Psi \exp(iS[\Psi] + iS_{\text{src}}[\Psi, J]) \quad (2.7)$$

as a standard path integral but with an additional source term in the action²

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

This source term has the simple property that a functional derivative w.r.t. the source $J(x)$ produces precisely the field $\Psi(x)$ at the same location

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

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

$$\frac{-i\delta}{\delta J(x)} \exp(iS_{\text{src}}[\Psi, J]) = \Psi(x) \exp(iS_{\text{src}}[\Psi, J]). \quad (2.10)$$

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

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

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

$$\Psi(x) \simeq \frac{-i\delta}{\delta J(x)}. \quad (2.11)$$

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

$$\frac{-i\delta}{\delta J(x)} \frac{-i\delta}{\delta J(y)} Z[J] = \int D\Psi \Psi(x) \Psi(y) \exp(iS[\Psi] + iS_{\text{src}}[\Psi, J]). \quad (2.12)$$

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

$$\langle \Psi(x) \Psi(y) \rangle = Z[J]^{-1} \frac{-i\delta}{\delta J(x)} \frac{-i\delta}{\delta J(y)} Z[J] \Big|_{J=0}. \quad (2.13)$$

Free Scalar Field. Now we can formally write any time-ordered correlators, but how to compute them in practice? We can only expect to obtain an exact expression for free fields. Therefore consider the scalar field. The action with source term reads

$$S[\phi] + S_{\text{src}}[\phi, j] = \int d^D x \left(-\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 + j\phi \right). \quad (2.14)$$

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

$$S[\phi] + S_{\text{src}}[\phi, j] = \int d^D x \left(\frac{1}{2}\phi(\partial^2 - m^2)\phi + \phi j \right). \quad (2.15)$$

This is a Gaussian integral 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 $G_{\text{F}}(x - y)$ satisfying³

$$(-\partial^2 + m^2)G_{\text{F}}(x - y) = \delta^D(x - y). \quad (2.16)$$

Hence we shift the field

$$\phi(x) = \tilde{\phi}(x) + \int d^D y G_{\text{F}}(x - y)j(y). \quad (2.17)$$

and substitute it into the action

$$\begin{aligned} S[\phi] + S_{\text{src}}[\phi, j] &= \int d^D x \frac{1}{2}\tilde{\phi}(x)(\partial^2 - m^2)\tilde{\phi}(x) + W[j]. \\ W[j] &= \int d^D x d^D y \frac{1}{2}j(x)G_{\text{F}}(x - y)j(y). \end{aligned} \quad (2.18)$$

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

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

$$Z[j] = Z[0] \exp(iW[j]). \quad (2.19)$$

A derivation in momentum space is somewhat simpler because the Gaussian kernel is automatically diagonal. The momentum-space version of $W[j]$ is⁴

$$W[j] = \int d^D p \frac{\frac{1}{2}j(p)j(-p)}{p^2 + m^2 - i\epsilon} = \int d^D p \frac{\frac{1}{2}|j(p)|^2}{p^2 + m^2 - i\epsilon}. \quad (2.20)$$

Formally, the prefactor reads⁵

$$Z[0] \sim \frac{1}{\sqrt{\det(-\partial^2 + m^2 - i\epsilon)}}. \quad (2.21)$$

Wick's Theorem. Let us now compute $\langle \Psi(x)\Psi(y) \rangle$. We perform two functional derivatives⁶

$$\begin{aligned} & \frac{-i\delta}{\delta J(y)} \frac{-i\delta}{\delta J(x)} Z[J] \\ &= \frac{-i\delta}{\delta J(y)} \int d^D z G_F(x-z) J(z) Z[J] \\ &= -iG_F(x-y) Z[J] \\ & \quad + \int d^D z d^D z' G_F(x-z) J(z) G_F(y-z') J(z') Z[J]. \end{aligned} \quad (2.22)$$

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

$$\langle \phi(x)\phi(y) \rangle = -iG_F(x-y). \quad (2.23)$$

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 $Z[J]$ as the exponent of $W[J]$ which is a quadratic monomial of J can be viewed as the functional formulation of Wick's theorem:

- The first derivative knocks down a linear term $\delta W/\delta J$ from the quadratic exponent.
- Subsequent derivatives can knock further linear terms $\delta W/\delta J$ from the exponent. They can also hit the remaining J in some $\delta W/\delta J$ leaving behind a Feynman propagator $-iG_F(x_k - x_l)$.

⁴The factor of 1/2 compensates the double-counting of $|j(p)| = |j(-p)|$.

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

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

- Any linear term $\delta W/\delta J$ 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 take 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.

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

$$\times_x \textcircled{F} \longrightarrow F(x). \quad (2.24)$$

- A solid dot connected by a short line to some object F represents a source field $J(x)$ multiplied to the object $F(x)$ integrated over all positions x

$$\bullet \textcircled{F} \longrightarrow \int d^D x F(x) J(x). \quad (2.25)$$

- An empty dot with a label represents a functional derivative $-i\delta/\delta J(x)$ multiplied to the object $F(x)$ integrated over all positions x

$$\circ \textcircled{F} \longrightarrow \int d^D x F(x) \frac{-i\delta}{\delta J(x)}. \quad (2.26)$$

- A long fat (blue) line represents a Feynman propagator $G_F(x, y)$ between two points x and y

$$\times_x \text{---} \times_y \longrightarrow G_F(x, y). \quad (2.27)$$

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

$$W[J] = \frac{1}{2} \bullet \text{---} \bullet \quad (2.28)$$

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

$$Z[J] = \sum_{n=0}^{\infty} \frac{Z[0] i^n}{n! 2^n} \left. \begin{array}{c} \bullet \text{---} \bullet \\ \bullet \text{---} \bullet \\ \bullet \text{---} \bullet \\ \bullet \text{---} \bullet \end{array} \right\} n = \exp \left(\frac{i}{2} \bullet \text{---} \bullet \right) \quad (2.29)$$

The functional derivative of $W[J]$ yields two equivalent terms which compensate the prefactor of $1/2$. The second derivative is the Feynman propagator

$$\begin{aligned} \times_x \circ \frac{1}{2} \bullet \text{---} \bullet &= -\frac{i}{2} \times_x \text{---} \bullet - \frac{i}{2} \bullet \text{---} \times_x = -i \times_x \text{---} \bullet, \\ \times_y \circ \times_x \circ \frac{1}{2} \bullet \text{---} \bullet &= -i \times_y \circ \times_x \text{---} \bullet = -\times_x \text{---} \times_y. \end{aligned} \quad (2.30)$$

For the first and second functional derivatives of the generating functional $Z[j]$ we can write

$$\begin{aligned}
 \times \underset{x}{\circ} \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) &= \times \text{---} \bullet \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right), \\
 \times \underset{y}{\circ} \times \underset{x}{\circ} \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) &= \left(\times \text{---} \bullet \times \text{---} \bullet - i \times \text{---} \times \right) \\
 &\quad \cdot \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right).
 \end{aligned} \tag{2.31}$$

Note that an empty dot grabs any of the solid dots and replaces it by $(-i)$ times a cross. Setting $J = 0$ in the above expressions shows that

$$\langle \phi(x)\phi(y) \rangle = -i \times \text{---} \times. \tag{2.32}$$

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. We have already discussed in QFT I that fermionic fields should take values in odd Grassmann numbers. Let us briefly review Grassmann numbers:

- A Grassmann number a can be either even ($|a| = 0$) or odd ($|a| = 1$).
- Grassmann numbers can be added as usual.
- The products 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

$$\theta_j \theta_k = -\theta_k \theta_j. \tag{2.33}$$

For each variable θ_k we can define a derivative $\partial/\partial\theta_k$. The derivatives themselves are odd Grassmann quantities

$$\frac{\partial}{\partial\theta_j} \frac{\partial}{\partial\theta_k} = -\frac{\partial}{\partial\theta_k} \frac{\partial}{\partial\theta_j}. \tag{2.34}$$

The defining property of derivatives is

$$\frac{\partial}{\partial\theta_j} \theta_k + \theta_k \frac{\partial}{\partial\theta_j} = \delta_k^j. \tag{2.35}$$

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

$$\int d\theta f(\theta) = ?. \tag{2.36}$$

It makes sense to demand that the integral of a total derivative vanishes

$$\int d\theta \frac{\partial}{\partial \theta} f(\theta) = 0. \quad (2.37)$$

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

$$0 = \int d\theta \frac{\partial}{\partial \theta} (f_0 + \theta f_1) = \int d\theta f_1. \quad (2.38)$$

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

$$\int d\theta (f_0 + \theta f_1) = \left(\int d\theta \theta \right) f_1. \quad (2.39)$$

The integral $\int d\theta \theta$ is some undetermined factor, we can define it as 1. The curious result is that integration of odd Grassmann variables is equivalent to differentiation⁷

$$\int d\theta f(\theta) = \frac{\partial}{\partial \theta} f(\theta). \quad (2.40)$$

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 issues 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

$$\int d\theta \delta(\theta - \alpha) f(\theta) = f(\alpha) \quad (2.41)$$

is solved trivially by⁸

$$\delta(\theta) = \theta. \quad (2.42)$$

Under variable transformations this delta function behaves as

$$\delta(\phi(\theta)) = \frac{\partial \phi}{\partial \theta} \delta(\theta - \theta_0). \quad (2.43)$$

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

Fourier Integrals. A plain Fourier integral produces a delta function as usual

$$\int d\theta \exp(c\theta\alpha) = \int d\theta (1 + c\theta\alpha) = c\alpha = c\delta(\alpha). \quad (2.44)$$

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

⁷This implies that the integration measure $d\theta$ has the dimension of $\partial/\partial\theta$ or $1/\theta$.

⁸Note that the order of terms in $\delta(\theta - \alpha)$ does matter since $\delta(\alpha - \theta) = -\delta(\theta - \alpha)$.

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

$$\int d\theta_2 d\theta_1 \exp(a\theta_1\theta_2) = a. \quad (2.45)$$

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

$$A = \begin{pmatrix} 0 & +a \\ -a & 0 \end{pmatrix}. \quad (2.46)$$

The result can be expressed as

$$\int d^n\theta \exp(\frac{1}{2}\theta^T A\theta) \sim \sqrt{\det(A)}. \quad (2.47)$$

This result in fact applies to general fermionic Gaussian integrals defined in terms of an $n \times n$ anti-symmetric matrix A .⁹ This expression is very similar to the bosonic n -dimensional Gaussian integral for a symmetric matrix S

$$\int d^n x \exp(-\frac{1}{2}x^T Sx) \sim \frac{1}{\sqrt{\det(S)}}. \quad (2.48)$$

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 since the odd integral is always well-defined.

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

$$\int d^n\theta d^n\bar{\theta} \exp(\bar{\theta}M\theta) \sim \det(M), \quad (2.49)$$

whereas the corresponding integral for even variables reads

$$\int d^n x d^n \bar{x} \exp(-\bar{x}Mx) \sim \frac{1}{\det(M)}, \quad (2.50)$$

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 .

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.

⁹A fermionic Gaussian integral requires an even number of integration variables n because the determinant of an odd-dimensional anti-symmetric matrix is zero.

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

$$S[\Psi] = S_0[\Psi] + S_{\text{int}}[\Psi]. \quad (2.51)$$

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.¹⁰ The path integral for the generating functional then reads

$$Z[J] = \int D\Psi \exp(iS_0[\Psi] + iS_{\text{int}}[\Psi] + iS_{\text{src}}[\Psi, J]). \quad (2.52)$$

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 w.r.t. the source J

$$\begin{aligned} Z[J] &= \int D\Psi \exp(iS_{\text{int}}[\Psi]) \exp(iS_0[\Psi] + iS_{\text{src}}[\Psi, J]) \\ &= \int D\Psi \exp\left(iS_{\text{int}}\left[-i\frac{\delta}{\delta J}\right]\right) \exp(iS_0[\Psi] + iS_{\text{src}}[\Psi, J]). \end{aligned} \quad (2.53)$$

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

$$\begin{aligned} Z[J] &= \exp\left(iS_{\text{int}}\left[-i\frac{\delta}{\delta J}\right]\right) Z_0[J], \\ Z_0[J] &= \int D\Psi \exp(iS_0[\Psi] + iS_{\text{src}}[\Psi, J]). \end{aligned} \quad (2.54)$$

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

$$S_{\text{int}}[\Psi] \simeq -\frac{1}{6}\kappa\Psi^3 - \frac{1}{24}\lambda\Psi^4 \sim -\frac{1}{6}\kappa \begin{array}{c} \circ \\ | \\ \circ \text{---} \text{---} \circ \\ | \\ \circ \end{array} - \frac{1}{24}\lambda \begin{array}{c} \circ \\ | \\ \circ \text{---} \text{---} \circ \\ | \\ \circ \end{array}. \quad (2.55)$$

The generating functional $Z[J]$ can be expanded in powers of S_{int}

$$Z[J] = \left(1 + iS_{\text{int}}\left[-i\frac{\delta}{\delta J}\right] - \frac{1}{2}S_{\text{int}}\left[-i\frac{\delta}{\delta J}\right]^2 + \dots\right) Z_0[J]. \quad (2.56)$$

¹⁰As 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 describes physical particles accurately in the absence of interactions.

- The first term is just the free generating functional $Z_0[J]$.
- The second term with a cubic interaction has three functional derivatives $-i\delta/\delta J$ acting on $Z_0[J]$. Let each of the derivatives act on

$$\begin{aligned}
& -\frac{i}{6}\kappa \int dw \left(-i\frac{\delta}{\delta J_w}\right)^3 Z_0[J] \\
&= -\frac{i}{6}\kappa \int dw dx \left(-i\frac{\delta}{\delta J_w}\right)^2 (G_{wx}J_x Z_0[J]) \\
&= -\frac{i}{6}\kappa \int dw dx dy \left(-i\frac{\delta}{\delta J_w}\right) (G_{wx}J_x G_{wy}J_y Z_0[J]) \\
&\quad -\frac{1}{6}\kappa \int dw \left(-i\frac{\delta}{\delta J_w}\right) (G_{ww}Z_0[J]) \\
&= -\frac{i}{6}\kappa \int dw dx dy dz G_{wx}J_x G_{wy}J_y G_{wz}J_z Z_0[J] \\
&\quad -\frac{3}{6}\kappa \int dw dx G_{ww}G_{wx}J_x Z_0[J]. \tag{2.57}
\end{aligned}$$

- 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. 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

$$\begin{aligned}
& -\frac{i}{6}\kappa \left(\begin{array}{c} \circ \\ \diagup \\ \circ \\ \diagdown \\ \circ \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) \\
&= -\frac{i}{6}\kappa \left(\begin{array}{c} \circ \\ \diagup \\ \bullet \\ \diagdown \\ \circ \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) \\
&= -\frac{i}{6}\kappa \left(\begin{array}{c} \bullet \\ \diagup \\ \circ \\ \diagdown \\ \bullet \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) - i \left(\begin{array}{c} \bullet \\ \diagup \\ \bullet \\ \diagdown \\ \circ \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) \\
&= -\frac{i}{6}\kappa \left(\begin{array}{c} \bullet \\ \diagup \\ \bullet \\ \diagdown \\ \bullet \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) - i \left(\begin{array}{c} \bullet \\ \diagup \\ \bullet \\ \diagdown \\ \bullet \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) - i \left(\begin{array}{c} \bullet \\ \diagup \\ \bullet \\ \diagdown \\ \bullet \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) - i \left(\begin{array}{c} \bullet \\ \diagup \\ \bullet \\ \diagdown \\ \bullet \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) \\
&= \kappa \left(\begin{array}{c} \bullet \\ \diagup \\ \bullet \\ \diagdown \\ \bullet \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right) - \frac{1}{2} \left(\begin{array}{c} \bullet \\ \diagup \\ \bullet \\ \diagdown \\ \bullet \end{array} \right) \exp\left(\frac{i}{2} \bullet \text{---} \bullet\right). \tag{2.58}
\end{aligned}$$

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

many more (intermediate terms)

$$\begin{aligned}
& -\frac{i}{24}\lambda \int dw \left(-i\frac{\delta}{\delta J_w}\right)^4 Z_0[J] \\
& = -\frac{i}{24}\lambda \int dw dx dy dz du G_{wx}J_x G_{wy}J_y G_{wz}J_z G_{wu}J_u Z_0[J] \\
& \quad -\frac{6}{24}\lambda \int dw dx dy G_{ww}G_{wx}J_x G_{wy}J_y Z_0[J] \\
& \quad +\frac{3i}{24}\lambda \int dw G_{ww}G_{ww}Z_0[J]
\end{aligned} \tag{2.59}$$

or graphically

$$\begin{aligned}
& -\frac{i}{24}\lambda \left(\text{diagram: vertex with 4 external lines} \right) \exp\left(\frac{i}{2} \text{diagram: line with two dots}\right) \\
& = \lambda \left(-\frac{i}{24} \left(\text{diagram: vertex with 4 external lines} \right) - \frac{1}{4} \left(\text{diagram: vertex with 2 external lines and a loop} \right) + \frac{i}{8} \left(\text{diagram: vertex with 2 external lines and a loop} \right) \right) \exp\left(\frac{i}{2} \text{diagram: line with two dots}\right).
\end{aligned} \tag{2.60}$$

- Let us finally consider two cubic vertices

$$\frac{i^2\kappa^2}{2 \cdot 6^2} \int dv dw \left(-i\frac{\delta}{\delta J_v}\right)^3 \left(-i\frac{\delta}{\delta J_w}\right)^3 Z_0[J]. \tag{2.61}$$

We can start with the single cubic vertex as intermediate result

$$\begin{aligned}
& -\frac{1}{72}\kappa^2 \left(\text{diagram: vertex with 3 external lines} \right)^2 \exp\left(\frac{i}{2} \text{diagram: line with two dots}\right) \\
& = -\frac{1}{12}\kappa^2 \left(\text{diagram: vertex with 3 external lines} \right) \left(\frac{1}{6} \left(\text{diagram: vertex with 2 external lines and a line} \right) - \frac{i}{2} \left(\text{diagram: vertex with 2 external lines and a loop} \right) \right) \exp\left(\frac{i}{2} \text{diagram: line with two dots}\right) \\
& = \kappa^2 \left[-\frac{1}{72} \left(\text{diagram: vertex with 3 external lines} \right) \left(\text{diagram: vertex with 2 external lines and a line} \right) + \frac{i}{8} \left(\text{diagram: vertex with 3 external lines and a line} \right) \right. \\
& \quad + \frac{i}{12} \left(\text{diagram: vertex with 3 external lines and a loop} \right) + \frac{1}{4} \left(\text{diagram: vertex with 2 external lines and a loop} \right) \\
& \quad + \frac{1}{8} \left(\text{diagram: vertex with 2 external lines and two loops} \right) - \frac{i}{8} \left(\text{diagram: vertex with 2 external lines and two loops} \right) \\
& \quad \left. + \frac{1}{4} \left(\text{diagram: vertex with 2 external lines and two loops} \right) - \frac{i}{12} \left(\text{diagram: vertex with 2 external lines and two loops} \right) \right] \exp\left(\frac{i}{2} \text{diagram: line with two dots}\right).
\end{aligned}$$

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

$$Z[J] = \sum \text{[Diagram: A central green circle labeled 'Z' with six blue lines radiating outwards to black dots]} . \quad (2.62)$$

To obtain a correlation function, all sources J have to be saturated by functional derivatives $-i\delta/\delta J$ 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:

- All source fields J are indistinguishable and can be freely permuted. This results in large symmetry factors in $Z[J]$.
- The field $\Psi(x), \Psi(y)$ in some correlator and corresponding functional derivatives are well distinguishable due to the distinct positions x, y in spacetime. Hence correlation functions often have relatively small symmetry factors since the external legs cannot be permuted.

Loop Counting. First 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 above¹¹

$$\begin{aligned} Z[J] &= \int D\Psi \exp(i\hbar^{-1}S[\Psi] + i\hbar^{-1}S_{\text{src}}[\Psi, J]), \\ &= \exp\left(i\hbar^{-1}S_{\text{int}}\left[-i\hbar\frac{\delta}{\delta J}\right]\right) Z_0[J], \\ Z_0[J] &= \exp\left(\frac{i}{2}\hbar^{-1} \int d^Dx d^Dy J(x)J(y)G_{\text{F}}(x, y)\right). \end{aligned} \quad (2.63)$$

Functional derivatives w.r.t. 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 $-i\hbar\delta/\delta J$ contributes a power of \hbar^+1 .

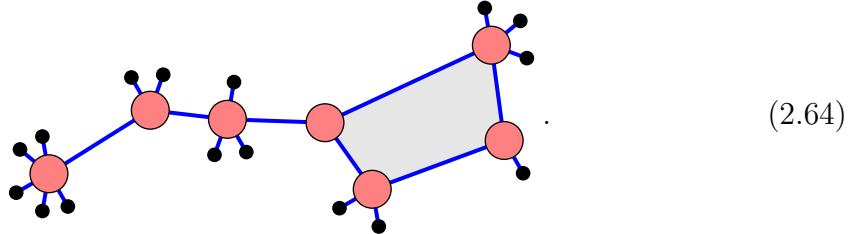
Consider now a graph with

- I internal lines (connecting two vertices),

¹¹Of 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 $I = 7$, $E = 16$, $V = 7$, $N = 1$, $L = 1$



For the counting of powers of \hbar we need to count the vertices, lines and junctions. Since the number of junctions per vertex is flexible, we shall attribute junctions to lines:

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



- Each internal line contributes \hbar^{-1} for the line times \hbar^{+2} from the two junctions with vertices.



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



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

$$\hbar^{I-V}. \quad (2.68)$$

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 line (integral),
- no free momentum for each external line (fixed by momentum inflow),
- one constraint for each vertex (delta function),
- one overall momentum conservation for each connected component merely constrains the external momenta, but none of the internal momenta.

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

$$I - V + N = L. \quad (2.69)$$

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

$$\hbar^{I-V} = \hbar^{L-N}. \quad (2.70)$$

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 $Z[J]$ there are further useful generating functionals which are somewhat simpler to handle and evaluate. These summarise

- $W[J]$ connected graphs,
- $T[J]$ connected tree graphs,
- $G[\Psi]$ one-particle irreducible graphs.

Connected Graphs. The connected generating functional $W[J]$ is defined as the logarithm of $Z[J]$ ¹²

$$W[J] = -i\hbar \log Z[J], \quad Z[J] = \exp(i\hbar^{-1}W[J]). \quad (2.71)$$

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

More explicitly, consider a disconnected graph Γ consisting of connected subgraphs Γ_k with multiplicity n_k . The contribution to $Z[J]$ reads¹³

$$\frac{\Gamma[J]}{S[\Gamma]} = \prod_k \frac{1}{n_k!} \left(\frac{\Gamma_k[J]}{S[\Gamma_k]} \right)^{n_k} \in Z[J]. \quad (2.72)$$

Here $S[\Gamma]$ denotes the symmetry factor associated to the graph Γ . The above term actually arises as one term in the multinomial and exponential

$$\frac{\Gamma}{S[\Gamma]} \in \frac{1}{n!} \left(\sum_k \frac{\Gamma_k}{S[\Gamma_k]} \right)^n \in \exp \left(\sum_k \frac{\Gamma_k}{S[\Gamma_k]} \right). \quad (2.73)$$

¹²This is a pretty general relationship for generating functionals of graphs which also holds in different contexts.

¹³Here, the symbol $X \in Y$ is means “ X is a term of the polynomial Y ”.

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

$$\begin{aligned} \textcircled{Z} &= 1 + i \textcircled{W} - \frac{1}{2} \textcircled{W} \textcircled{W} - \frac{i}{6} \textcircled{W} \textcircled{W} \textcircled{W} + \dots \\ &= \exp \left(i \textcircled{W} \right). \end{aligned} \quad (2.74)$$

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 power \hbar^{-1} as a prefactor to the action in the exponent

$$Z[J] = \int D\Psi \exp(i\hbar^{-1}S[\Psi] + i\hbar^{-1}S_{\text{src}}[\Psi, J]). \quad (2.75)$$

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

$$\frac{\delta S[\Psi]}{\delta \Psi(x)} + \frac{\delta S_{\text{src}}[\Psi, J]}{\delta \Psi(x)} = \frac{\delta S[\Psi]}{\delta \Psi(x)} + J(x) = 0. \quad (2.76)$$

Let us assume that there is a single stationary contribution for each source field configuration J ¹⁴ which we shall denote by $\Psi = \Psi[J]$. At small \hbar the path integral is dominated by this contribution (up to some irrelevant prefactor)

$$Z[J] \approx \exp(i\hbar^{-1}T[J]), \quad (2.77)$$

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

$$T[J] := S[\Psi[J]] + \int d^D x J(x) \Psi[J](x). \quad (2.78)$$

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

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

$$T[J] = \lim_{\hbar \rightarrow 0} W[J]. \quad (2.79)$$

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 $N = 1$, and therefore the graphs in $W[J]$ scale as \hbar^L . The limit $\hbar \rightarrow 0$ then restricts to graph with $L = 0$, i.e. the graphs in $T[J]$ have no

¹⁴At least formally and perturbatively we can make this assumption.

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 defined as the functional derivative of $S[\Psi]$

$$J(x) = -\frac{\delta S[\Psi]}{\delta \Psi(x)}. \quad (2.80)$$

Moreover, $T[J]$ equals $S[\Psi]$ plus a term $J \cdot \Psi$ 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 $S_{\text{int}}[\Psi]$, respectively

$$T[J] = \begin{array}{c} \text{Diagram 1} \\ + \frac{1}{2} \text{Diagram 2} \\ + \frac{1}{2} \text{Diagram 3} \\ + \frac{1}{2} \text{Diagram 4} \\ + \frac{1}{6} \text{Diagram 5} \\ + \dots \end{array} \quad (2.81)$$

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

$$S[\Psi] = -\frac{1}{2} \text{Diagram 1} - \frac{1}{6} \kappa \text{Diagram 2}. \quad (2.82)$$

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

$$\text{Diagram 1} = \text{Diagram 2} + \frac{1}{2} \kappa \text{Diagram 3}. \quad (2.83)$$

We need to find the inverse functional $\Psi[J]$

$$\text{Diagram 1} = \text{Diagram 2} - \frac{1}{2} \kappa \text{Diagram 3} + \frac{1}{2} \kappa^2 \text{Diagram 4} + \dots \quad (2.84)$$

One can confirm by substitution that these two functions are mutually inverse.

Now substitute $\Psi[J]$ into the two terms of the action $S[\Psi]$ and the Legendre

transformation term $S_{\text{src}}[\Psi, J]$

$$\begin{aligned}
 -\frac{1}{2} \circ \text{---} \circ &= -\frac{1}{2} \bullet \text{---} \bullet + \frac{1}{2} \kappa \begin{array}{c} \bullet \\ \diagdown \\ \text{---} \\ \diagup \\ \bullet \end{array} \\
 & - \left(\frac{1}{2} + \frac{1}{8}\right) \kappa^2 \begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \text{---} \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + \dots,
 \end{aligned} \tag{2.85}$$

$$\begin{aligned}
 -\frac{1}{6} \kappa \begin{array}{c} \circ \\ \diagdown \\ \text{---} \\ \diagup \\ \circ \end{array} &= -\frac{1}{6} \kappa \begin{array}{c} \bullet \\ \diagdown \\ \text{---} \\ \diagup \\ \bullet \end{array} + \frac{1}{4} \kappa^2 \begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \text{---} \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + \dots,
 \end{aligned} \tag{2.86}$$

$$\begin{aligned}
 \circ \text{---} \bullet &= \bullet \text{---} \bullet - \frac{1}{2} \kappa \begin{array}{c} \bullet \\ \diagdown \\ \text{---} \\ \diagup \\ \bullet \end{array} + \frac{1}{2} \kappa^2 \begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \text{---} \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + \dots
 \end{aligned} \tag{2.87}$$

The sum of these terms yields the Legendre transform

$$T[J] = \frac{1}{2} \bullet \text{---} \bullet - \frac{1}{6} \kappa \begin{array}{c} \bullet \\ \diagdown \\ \text{---} \\ \diagup \\ \bullet \end{array} + \frac{1}{8} \kappa^2 \begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \text{---} \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + \dots \tag{2.88}$$

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

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 $G[\Psi]$.¹⁵

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

$$\Psi[J] = \frac{\delta W}{\delta J} [J], \tag{2.89}$$

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 .¹⁶

$$\Psi[J](x) = \langle \Psi(x) \rangle_J := Z[J]^{-1} \frac{-i\hbar\delta}{\delta J(x)} Z[J] = \frac{\delta W}{\delta J(x)} [J]. \tag{2.90}$$

¹⁵It 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.

¹⁶In 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 $\langle O \rangle_J := Z[J]^{-1} O[-i\hbar\delta/\delta J] Z[J]$.

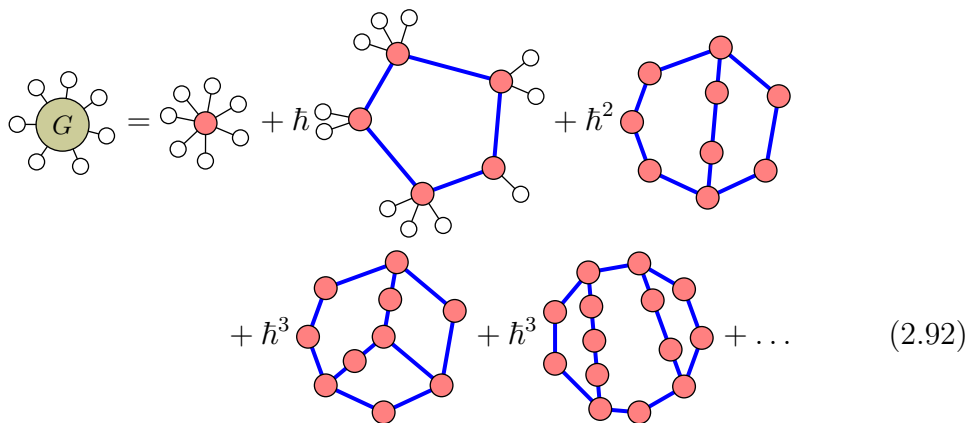
The effective action is then defined as

$$G[\Psi] := W[J[\Psi]] - \int d^D x J[\Psi](x)\Psi(x). \quad (2.91)$$

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 without Feynman propagators.

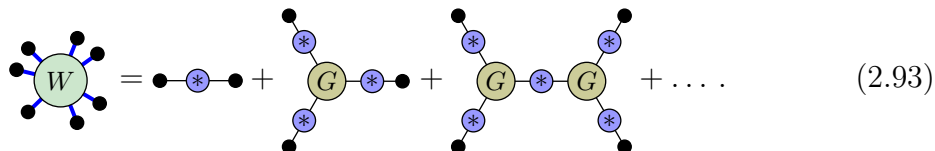
How can this be seen? We know that the Legendre transformation switches between vertices ($S[\Psi]$) and connected tree graphs made from these vertices ($T[J]$). Here the starting point is the generating functional of connected graphs $W[J]$ with loops. Let us view the latter as trees made from some 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.¹⁷ These effective vertices have the property that cutting any single line does not split the vertex. Hence $G[\Psi]$ encodes precisely the 1PI graphs.

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 amputated loop graphs



$$G = \text{tree} + \hbar \text{loop}_4 + \hbar^2 \text{loop}_6 + \hbar^3 \text{loop}_8 + \hbar^3 \text{loop}_{10} + \dots \quad (2.92)$$

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



$$W = \text{tree} + G \text{ tree} + G \text{ tree} G + \dots \quad (2.93)$$

¹⁷Clearly, this procedure is finite and leads to a unique decomposition of the graph.

Here we assume that the effective vertices labelled 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

$$\bullet - (\ast) - \bullet = \bullet - \bullet - \bullet + \bullet - \textcircled{G'} - \bullet + \bullet - \textcircled{G'} - \textcircled{G'} - \bullet + \dots \quad (2.94)$$

Here, $G'_2 = G_2 - S_0$ are those effective two-point vertices which are not already in the free action S_0 .¹⁸ Summing up the geometric series shows that the effective propagator marked by $*$ is minus the inverse of the two-point effective vertex G_2

$$(-\ast-)^{-1} = \text{---} \text{---} \text{---} - \text{---} \textcircled{G'} \text{---} = -\text{---} \textcircled{G} \text{---} . \quad (2.95)$$

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

$$(\text{---})^{-1} = \text{---} \text{---} \text{---} . \quad (2.96)$$

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 .

The reason for calling $G[\Psi]$ the (quantum) effective action is as follows: Quantising a classical theory described by $S[\Psi]$ yields the functional $Z[J]$ and its logarithm $W[J]$. The same functional $W[J]$ is obtained by constructing tree graphs of the effective action $G[\Psi]$. Put differently: The functional $W[J]$ describes

- quantum physics for a model described by the classical (local) action $S[\Psi]$ as well as
- classical physics for the model described by the effective (non-local) action $G[\Psi]$.

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, since 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 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 in fact relates S-matrix elements to residues of the functional $Z[J]$. This implies that the S-matrix has disconnected graphs as well as one-particle reducible graphs.

¹⁸The free action defines the Feynman propagator and therefore it never appears as an interaction vertex.

Summary. There are at least five noteworthy generating functionals.

$$\begin{array}{ccccc}
 S[\Psi] & \xleftrightarrow{\text{Lgdr.}} & T[J] & & S[\Psi] \\
 \uparrow & & \uparrow & & \downarrow \\
 \hbar \rightarrow 0 & & \hbar \rightarrow 0 & & \text{P.I.} \\
 \downarrow & & \downarrow & & \downarrow \\
 G[\Psi] & \xleftrightarrow{\text{Lgdr.}} & W[J] & \xleftrightarrow{\text{log/exp}} & Z[J]
 \end{array} \tag{2.97}$$

- 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.
- 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 Lie Algebra

Symmetries are ubiquitous in physics. Mathematically they are described by the concept of groups. They can be discrete (such as lattice symmetries in condensed matter physics) or continuous (such as the spacetime symmetries in quantum field theory). The latter are known as Lie groups. Lie groups play an important role in quantum field theory, where they serve as global spacetime and flavour symmetries. Furthermore they prominently appear as local gauge symmetries in Yang–Mills theory which is a generalisation of electrodynamics.¹

3.1 Lie Groups

Yang–Mills theory is often explained in terms of $N \times N$ unitarity matrices. The latter form a Lie group. Let us therefore discuss Lie groups, Lie algebras, enveloping algebras and their relationship.

Definition. A *Lie group* G is a group that is also a smooth manifold. The group multiplication $G \times G \rightarrow G$ must be a smooth map.

Example. The set of unitary matrices evidently defines a (compact) smooth manifold, and it is a group with a smooth composition rule.

Lie groups can be distinguished by several useful properties:

- They can be *simple*, *semi-simple* or composite.
- 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).

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 does not have non-trivial connected normal sub-groups.² Since only *normal* subgroups H of G can be used to define quotient spaces G/H with a *group* structure, simplicity essentially means that the group cannot be reduced to a smaller group.
- A semi-simple Lie group is a direct product of simple Lie groups.
- Composite Lie groups are direct or non-direct products of simple or abelian Lie groups.

¹As a motivation, it makes sense to get familiar with classical Yang–Mills theory at the beginning of the next chapter before considering the more abstract topics of Lie groups in this chapter.

²A one-dimensional abelian Lie group is sometimes considered to be simple.

Example. Consider a group of matrices which takes one of the following two block forms.

$$\begin{array}{cc} \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} & \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \\ \text{semi-simple} & \text{composite} \end{array} \quad (3.1)$$

Suppose $*$ denotes matrices which form simple or semi-simple Lie groups and \star denotes general matrices. The former group is a direct product of simple groups, hence it is semi-simple. It is not simple because it has normal subgroups. The latter group is not a direct product, hence it is not semi-simple. Furthermore it contains normal subgroups: These consist of matrices with a non-trivial off-diagonal matrix \star and some of the diagonal elements $*$ replaced by the unit matrix 1.

In Yang–Mills theory we will encounter mostly direct products of compact, finite-dimensional, simple or abelian, real Lie groups. Furthermore, most of the results presented below will apply only to particular types of Lie groups, in most cases at least to finite-dimensional, simple Lie groups.

3.2 Lie Algebras

Lie groups are curved manifolds which makes them somewhat hard to investigate and apply. 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.

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 small) neighbourhood of 1 must again be in a (slightly larger) neighbourhood of 1 since multiplication is 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 T_1G of the Lie group G at the unit element 1. This tangent space is called the *Lie algebra* \mathfrak{g} associated to G

$$\mathfrak{g} := T_1G. \quad (3.2)$$

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 ³ such that⁴

$$\exp(0) = 1, \quad d\exp(0) = \text{id}, \quad \exp(na) = \exp(a)^n. \quad (3.3)$$

This map is called the *exponential map*. We can construct the exponential map as the limit⁵

$$\exp(a) = \lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n. \quad (3.4)$$

Here $1 + a/n$ is understood as 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 $m : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ (more precisely on some neighbourhoods of 0 in \mathfrak{g}) such that

$$\exp(a)\exp(b) = \exp(m(a, b)). \quad (3.5)$$

Now it is clear that $m(a, 0) = m(0, a) = x$ hence

$$m(a, b) = a + b + \mathcal{O}(a)\mathcal{O}(b). \quad (3.6)$$

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. This composition law is therefore 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

$$m(a, b) = a + b + m_2(a, b) + m_{\geq 3}(a, b). \quad (3.7)$$

From the above discussion 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

$$0 = m_2(a + b, a + b) - m_2(a, a) - m_2(b, b) = m_2(a, b) + m_2(b, a). \quad (3.8)$$

³It 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.

⁴For a map f between two manifolds $f : A \rightarrow B$, its derivative df at a point a is a linear map between the corresponding tangent spaces $df(a) : T_a A \rightarrow T_{f(a)} B$. In this case the derivative $d\exp$ is defined as the identity map id on $T_0 \mathfrak{g} = \mathfrak{g} = T_1 G$.

⁵This is commonly achieved by transport via a particular vector field.

Lie Brackets. The leading deviation $2m_2(a, b)$ from linearity is called the *Lie bracket* $[x, y]_{\text{Lie}}$.

$$\approx \exp(\epsilon^2 [a, b]_{\text{Lie}}) \quad (3.9)$$

We can thus compute it as the limit of a commutator of Lie group elements

$$[a, b]_{\text{Lie}} = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon^2} \exp^{-1}(\exp(\epsilon a) \exp(\epsilon b) \exp(-\epsilon a) \exp(-\epsilon b)). \quad (3.10)$$

The Lie bracket has the following properties:

- It is bilinear.
- It is anti-symmetric

$$[a, b]_{\text{Lie}} = -[b, a]_{\text{Lie}}. \quad (3.11)$$

- It satisfies the Jacobi identity⁶

$$[[a, b]_{\text{Lie}}, c]_{\text{Lie}} + [[b, c]_{\text{Lie}}, a]_{\text{Lie}} + [[c, a]_{\text{Lie}}, b]_{\text{Lie}} = 0. \quad (3.12)$$

The latter property follows from the Hall–Witt identity for three group elements x, y, z and their inverses $\bar{x}, \bar{y}, \bar{z}$

$$\begin{aligned} 1 &= \bar{y}((y\bar{x}\bar{y}x)\bar{z}(\bar{x}yxy)z)y \\ &\quad \bar{z}((z\bar{y}\bar{z}y)\bar{x}(\bar{y}zy\bar{z})x)z \\ &\quad \bar{x}((x\bar{z}\bar{x}z)\bar{y}(\bar{z}xz\bar{x})y)x. \end{aligned} \quad (3.13)$$

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.

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.

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.

⁶In 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.

3.3 Enveloping Algebras

In quantum physics, we are not only interested in Lie algebra elements and their Lie brackets, but we would also like to be able to take associative products of them.

Definition. There is a useful concept to embed a Lie algebra into a (much) bigger associative algebra with unit element. This is the (universal) *enveloping algebra* $U(\mathfrak{g})$ of a Lie algebra \mathfrak{g} . Its elements are ordered polynomials of the elements of the Lie algebra \mathfrak{g} . For example, if $a, b, c \in \mathfrak{g}$ then the monomials $abc, acb, baac, \dots$, as well as their linear combinations, are elements of $U(\mathfrak{g})$. The product of two monomials X, Y is defined by their concatenation,

$$X \cdot Y = XY \quad (3.14)$$

Moreover, the Lie brackets are encoded in the enveloping algebra by identifying them with commutators, i.e. for all $a, b \in \mathfrak{g}$

$$[a, b]_{\text{Lie}} \equiv [a, b] = ab - ba. \quad (3.15)$$

To make this identification compatible with products we must also demand the following: If $[a, b]_{\text{Lie}} = c$ then for any $X, Y \in U(\mathfrak{g})$

$$XabY - XbaY \equiv XcY. \quad (3.16)$$

Therefore the enveloping algebra can be defined as the space

$$U(\mathfrak{g}) = \bigoplus_{n=0}^{\infty} \mathfrak{g}^{\otimes n} / ([\cdot, \cdot] \equiv [\cdot, \cdot]_{\text{Lie}}). \quad (3.17)$$

Since the identification of commutators effectively removes all asymmetries within monomials (by shifting them to shorter monomials), we can also write the vector space of the enveloping algebra as a symmetric tensor product \odot

$$U(\mathfrak{g}) \simeq \bigoplus_{n=0}^{\infty} \mathfrak{g}^{\odot n}. \quad (3.18)$$

Properties. The enveloping algebra $U(\mathfrak{g})$ clearly includes the Lie algebra \mathfrak{g} as the linear polynomials. Furthermore, it (more precisely its completion) also includes the associated Lie group G since the exponential map acts within $U(\mathfrak{g})$

$$\exp(X) = \lim_{n \rightarrow \infty} \left(1 + \frac{X}{n}\right)^n = \sum_{n=0}^{\infty} \frac{1}{n!} X^n. \quad (3.19)$$

In that sense the enveloping algebra $U(\mathfrak{g})$ contains a Lie algebra \mathfrak{g} , the associated Lie group G as well as many other elements which belong to neither of the two.⁷

⁷In physics these three terms are sometimes confused. In those cases it is safe to assume the enveloping algebra.

3.4 Representations

Let us now introduce some basic elements of Lie theory, in particular those which will be relevant to Yang–Mills theory. Fields 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 endomorphisms of some vector space \mathbb{V} (operators, linear maps or matrices acting on the space)

$$R : G \rightarrow \text{End}(\mathbb{V}). \quad (3.20)$$

The representation must reflect the composition law of G by operator composition (matrix multiplication), i.e. for all $h, k, l \in G$

$$hk = l \implies R(h)R(k) = R(l). \quad (3.21)$$

The operator $R(h)$ is called a representation of the element h .

The definition of a representation R of a Lie algebra \mathfrak{g} is analogous. It is a linear map from the algebra to endomorphisms of some space \mathbb{V}

$$R : \mathfrak{g} \rightarrow \text{End}(\mathbb{V}). \quad (3.22)$$

The representation must reflect the Lie bracket of \mathfrak{g} as a commutator $[A, B] := AB - BA$ of operators, i.e. for all $a, b, c \in \mathfrak{g}$

$$[a, b]_{\text{Lie}} = c \implies [R(a), R(b)] = R(c). \quad (3.23)$$

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.

Finally, the enveloping algebra $U(\mathfrak{g})$ has representations as well

$$R : U(\mathfrak{g}) \rightarrow \text{End}(\mathbb{V}). \quad (3.24)$$

They must obey the multiplication rule for all $X, Y, Z \in U(\mathfrak{g})$

$$XY = Z \implies R(X)R(Y) = R(Z). \quad (3.25)$$

This rule is the same as for Lie groups, hence we will not specify this case explicitly in what follows.

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 .⁸ For example, only a discrete subset of the representation of $\mathfrak{g} = \mathfrak{u}(1) = \mathbb{R}$ lift to representations of $G = U(1) = S^1$ because the latter must respect the fact that rotations by 2π are trivial. Finally, representations of \mathfrak{g} trivially lift to $U(\mathfrak{g})$.

⁸It should hold when G is simply connected.

The space \mathbb{V} is called a representation space or a module of the group, algebra or enveloping algebra (in physics it is sometimes also sloppily called a representation). The dimension D^R of the representation R is defined as the dimension of \mathbb{V}

$$D^R = \dim R := \dim \mathbb{V}. \quad (3.26)$$

Adjoint and Trivial Representations. A distinguished representation of any Lie algebra \mathfrak{g} is the adjoint representation on the Lie algebra itself

$$\text{ad} : \mathfrak{g} \rightarrow \text{End}(\mathfrak{g}), \quad \text{ad}(a)b = [a, b]_{\text{Lie}}. \quad (3.27)$$

The Jacobi identity ensures that ad is a proper representation since it guarantees that for any $a, b \in \mathfrak{g}$

$$\text{ad}([a, b]_{\text{Lie}}) = [\text{ad}(a), \text{ad}(b)]. \quad (3.28)$$

This becomes clear once applied to some $c \in \mathfrak{g}$ and using anti-symmetry of the Lie brackets

$$\begin{aligned} \text{ad}([a, b]_{\text{Lie}})c &= [[a, b]_{\text{Lie}}, c]_{\text{Lie}}, \\ -\text{ad}(a)\text{ad}(b)c &= -[a, [b, c]_{\text{Lie}}]_{\text{Lie}} = [[b, c]_{\text{Lie}}, a]_{\text{Lie}}, \\ +\text{ad}(b)\text{ad}(a)c &= +[b, [a, c]_{\text{Lie}}]_{\text{Lie}} = [[c, a]_{\text{Lie}}, b]_{\text{Lie}}. \end{aligned} \quad (3.29)$$

The vector fields always transform in the adjoint representation of the gauge group.

The adjoint representation for the Lie group (and equivalently for the enveloping algebra) reads⁹

$$\text{Ad} : G \rightarrow \text{End}(\mathfrak{g}), \quad \text{Ad}(h)b = h b h^{-1}. \quad (3.30)$$

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 $h \in G$

$$\text{triv}(a) = 0, \quad \text{triv}(h) = \text{id}. \quad (3.31)$$

Finally, there always exists a representation on the enveloping algebra since it naturally contains the Lie algebra and group

$$\text{env} : \mathfrak{g}, G \rightarrow \text{End}(U(\mathfrak{g})), \quad \text{env}(h)X = hX, \quad \text{env}(a)X = aX, \quad (3.32)$$

This representation is useful for formal reasons, since in some sense it is maximal and contains all other representations.

Sums and Tensor Products. Representations can be combined to form bigger representations. There are two main operations: direct sums and tensor products. In Yang–Mills theory these operations correspond to collections and products of fields.

⁹Here the product $b' = h b h^{-1}$ 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, Lie algebra or an enveloping 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 $(\mathbb{V}_1, \mathbb{V}_2)$

$$R_{1\oplus 2} = \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix}. \quad (3.33)$$

Tensor product representations $R_{1\otimes 2}$ of an element of the group (or enveloping algebra) on the space $\mathbb{V}_{1\otimes 2} = \mathbb{V}_1 \otimes \mathbb{V}_2$ are defined as

$$R_{1\otimes 2}(h) = R_1(h) \otimes R_2(h), \quad (3.34)$$

The definition for a Lie algebra element is slightly different

$$R_{1\otimes 2}(a) = R_1(a) \otimes \text{id} + \text{id} \otimes R_2(a). \quad (3.35)$$

Evidently, all three definitions of R satisfy the axioms of representations provided that R_1, R_2 are proper representations.

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 $P \in \text{End}(\mathbb{V} \otimes \mathbb{V})$

$$P(v_1 \otimes v_2) := v_2 \otimes v_1. \quad (3.36)$$

It can be used to construct two projectors P^\pm onto the symmetric and anti-symmetric subspaces \mathbb{V}^\pm of $\mathbb{V} \otimes \mathbb{V}$

$$P^\pm = \frac{1}{2}(\text{id} \pm P), \quad (3.37)$$

where id is the identity acting on $\mathbb{V} \otimes \mathbb{V}$. Now the definitions of the tensor product representation $R \otimes R$ commute with the permutation operator P , and hence also with the projectors P^\pm . Therefore it will act as a direct sum of representation R^\pm on the subspaces \mathbb{V}^\pm

$$R \otimes R = R^+ \oplus R^-, \quad R^\pm = P^\pm(R \otimes R)P^\pm. \quad (3.38)$$

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.¹⁰

Properties. The structure of a representation can have various properties:

- It can be *reducible* or *irreducible* (simple). An irreducible representation has no proper sub-representation. It is also called *irrep* for short.

¹⁰A useful tool in this respect are Young tableaux/diagrams, see below.

- It can be *decomposable* or *indecomposable*. A decomposable representation can be written as a direct sum of sub-representations.
- Two representations R_1, R_2 are *equivalent* if they are related by the same similarity transformation M for all $a \in \mathfrak{g}$

$$MR_1(a)M^{-1} = R_2(a). \quad (3.39)$$

- A representation can be finite or infinite-dimensional. Note that finite-dimensional Lie algebras can have infinite-dimensional representations and vice versa.

There is only a slight difference between reducibility and decomposability. We illustrate the difference as follows: A representation can be brought to one of the following three forms

$$\begin{array}{ccc} \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} & \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} & \begin{pmatrix} * & * \\ * & * \end{pmatrix} \\ \text{decomposable} & \text{indecomposable} & \text{indecomposable} \\ \text{reducible} & \text{reducible} & \text{irreducible} \end{array} \quad (3.40)$$

Each entry represents a sub-matrix which is either trivial (0) or non-trivial (*). The first case has two proper sub-representations. Due to the block off-diagonal elements, only one of the diagonal blocks of the second case is a sub-representation.¹¹ The third case cannot be brought to a form where an off-diagonal block is zero.

Example. Tensor products of representations are typically reducible. In the case of identical representations we have seen above how to decompose them into sub-representations.

An irreducible representation can have several other characterising properties:

- It can be real, complex or quaternionic depending on the type of the matrix entries.¹² Note that the reality type of representations is not directly related to the reality type of the algebra.
- Furthermore a real, complex or quaternionic representation of a Lie group (algebra) is said to be unitary (antihermitian), orthogonal (antisymmetric) or symplectic (antisymmetric) when all its matrices have this property.

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.

Note that the properties of the adjoint representation reflects the properties of the corresponding Lie group or algebra. For instance, a Lie algebra is simple if and only if its adjoint representation is non-trivial and irreducible. The adjoint representation has the same reality properties as the Lie algebra.

¹¹This case can be avoided when restricting to particular classes of representations such as the ones appearing in Yang–Mills theory.

¹²Evidently, one can write complex numbers as matrices over the real numbers and so on. Therefore the largest applicable class of numbers is meant.

3.5 Invariants

Lie algebras have several invariant objects which serve various important purposes. For instance the quadratic form provides a natural metric for the Lie algebra manifold. Consequently, it can be used to translate between vectors and covectors. In the Yang–Mills theory, they are not only important for defining Feynman propagators, but also relevant for consistency issues such as 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 $K \in (\mathfrak{g}^*)^{\odot 2}$, known as the Killing form. Symmetry and invariance means that for all $a, b, c \in \mathfrak{g}$

$$K(a, b) = K(b, a), \quad K([c, a]_{\text{Lie}}, b) + K(a, [c, b]_{\text{Lie}}) = 0. \quad (3.41)$$

This form can be constructed easily using some representation R , for example using the adjoint representations $R = \text{ad}$

$$\text{Tr } R(a)R(b) \sim K(a, b). \quad (3.42)$$

The Killing form of a semi-simple Lie algebra is invertible. Therefore there is a corresponding invariant element $C_2 \in \mathfrak{g}^{\odot 2}$ called the Casimir invariant. Thought of as a matrix, it is the inverse of the Killing form K . It obeys the invariance property for all $a \in \mathfrak{g}$

$$(\text{ad}(a) \otimes \text{id} + \text{id} \otimes \text{ad}(a))C_2 = 0. \quad (3.43)$$

The quadratic invariant can also be viewed as a quadratic element of the enveloping algebra $U(\mathfrak{g})$. The invariance property implies that it is in the centre of $U(\mathfrak{g})$, i.e. it commutes with all $X \in U(\mathfrak{g})$

$$C_2 X = X C_2. \quad (3.44)$$

Therefore an irreducible representation of C_2 is necessarily proportional to the identity operator

$$R(C_2) \sim \text{id}^R. \quad (3.45)$$

For example, for $\mathfrak{g} = \mathfrak{su}(2)$ the Casimir invariant C_2 is the total spin operator.

For most Lie algebras, there are analogous invariant polynomials $C_n \in \mathfrak{g}^{\odot n}$ of higher degree n . Which ones of these higher invariant exists depends on the Lie algebra. The quadratic invariant exists for every simple Lie algebra.

There is a curious relationship between the invariants of \mathfrak{g} and the topology of the associated (connected, universal covering) Lie group G : In terms of cohomology, the group manifold is equivalent to a product of odd-dimensional spheres

$$G \simeq \prod_k S^{2n_k-1}. \quad (3.46)$$

For every independent invariant of degree n_k there is a sphere of dimension $2n_k - 1$.¹³

¹³This well-known result of Lie groups is again not immediately relevant for our course.

Structure Constants. For non-abelian gauge theories it is convenient to choose some basis T_a for a 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

$$a = ia^a T_a. \quad (3.47)$$

On 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 i . For instance, a group element h is typically parametrised via the exponential map as

$$h = \exp(ia^a T_a). \quad (3.48)$$

Now that we have a basis, we can expand the Lie brackets in terms of the *structure constants* f_{ab}^c

$$[T_a, T_b]_{\text{Lie}} = if_{ab}^c T_c. \quad (3.49)$$

For a real algebra, the structure constants f_{ab}^c must be real.

Furthermore we expand the invariant bilinear form in this basis as

$$K(T_a, T_b) = k_{ab}. \quad (3.50)$$

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

$$k_{ab} = \delta_{ab}. \quad (3.51)$$

It is convenient to make this assumption, but for the time being we shall keep the notation k_{ab} and introduce its inverse k^{ab} .

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

$$f_{abc} = f_{ab}^d k_{dc}. \quad (3.52)$$

These structure constants are totally anti-symmetric in all three indices

$$f_{abc} = -f_{bac} = -f_{acb}. \quad (3.53)$$

Antisymmetry of the former two indices follows by definition. Antisymmetry of the latter indices follows from invariance of K

$$0 = K([T_c, T_a]_{\text{Lie}}, T_b) + K(T_a, [T_c, T_b]_{\text{Lie}}) = if_{cab} + if_{cba} \quad (3.54)$$

Now we can also consider a representation T_a^R of the basis T_a , namely

$$T_a^R = R(T_a). \quad (3.55)$$

Evidently the matrices T_a^R obey the commutator identity

$$[T_a^R, T_b^R] = if_{ab}^c T_c^R. \quad (3.56)$$

A special case is the adjoint representation

$$(T_a^{\text{ad}})T_b = [T_a, T_b]_{\text{Lie}} = if_{ab}^c T_c. \quad (3.57)$$

Therefore, the adjoint representation T_a^{ad} is a matrix whose elements are the structure constants

$$(T_a^{\text{ad}})_b^c = if_{ab}^c. \quad (3.58)$$

Constants for Representations. We have seen above that the invariant elements obey identities for representations which hold up certain factors of proportionality. These factors can be used to characterise representations. They also appear as characteristic quantities in computations of Feynman diagrams. Let us therefore discuss them in more detail.

For an irreducible representation R we know that the trace of two representation matrices yields the Killing form

$$\mathrm{Tr} T_a^R T_b^R = B^R k_{ab}. \quad (3.59)$$

The factor of proportionality B^R depends on the particular representation R . It also depends on the normalisation as we shall discuss later. For the adjoint representation we can write this relation in terms of structure constants

$$-f_{ac}{}^d f_{bd}{}^c = B^{\mathrm{ad}} k_{ab}. \quad (3.60)$$

An irreducible representation of the quadratic Casimir invariant $C_2 := k^{ab} T_a T_b \in \mathrm{U}(\mathfrak{g})$ now takes the form

$$R(C_2) = k^{ab} T_a^R T_b^R = C_2^R \mathrm{id}^R. \quad (3.61)$$

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 $j(j+1)$.

By taking the trace of the above identity we can establish a relationship between the constants B^R and C_2^R

$$C_2^R \mathrm{Tr} \mathrm{id}^R = k^{ab} \mathrm{Tr} T_a^R T_b^R = B^R k^{ab} k_{ab}. \quad (3.62)$$

Now $\mathrm{Tr} \mathrm{id}^R = D^R$ is the dimension of the representation R . Furthermore $k^{ab} k_{ab} = \mathrm{Tr} \mathrm{id}^{\mathrm{ad}} = D^{\mathrm{ad}} = \dim \mathfrak{g}$ is the dimension of the algebra. We thus have the relation

$$D^{\mathrm{ad}} B^R = D^R C_2^R. \quad (3.63)$$

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

$$B^{\mathrm{ad}} = C_2^{\mathrm{ad}}. \quad (3.64)$$

This can also be seen directly

$$C_2^{\mathrm{ad}} \delta_c^e = (k^{ab} T_a^{\mathrm{ad}} T_b^{\mathrm{ad}})_c{}^f = -k^{ab} f_{ac}{}^d f_{bd}{}^e = B^{\mathrm{ad}} k_{cf} k^{fe}. \quad (3.65)$$

Also the higher invariants C_n have characterising eigenvalues C_n^R on irreducible representations R

$$R(C_n) = C_n^R \mathrm{id}. \quad (3.66)$$

Note that there is some arbitrariness in the definition of the above constants B^R and C_n^R . We can simultaneously rescale all generators T_a and the structure constants $f_{ab}{}^c$ by a common factor of α

$$(T')_a = \alpha T_a, \quad (f')_{ab}{}^c = \alpha f_{ab}{}^c. \quad (3.67)$$

This transformation does not change any of the above relations. However, the constants related to the invariants must rescale with a factor of α^2

$$(B')^R = \alpha^2 B^R, \quad (C'_2)^R = \alpha^2 C_2^R. \quad (3.68)$$

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.

Moreover, we are in principle free to rescale the definition of the structure constants $[T_a, T_b]_{\text{Lie}} \sim f_{ab}^c T_c$ and of the invariants k_{ab} , B and C_n . This would add further arbitrariness to their definitions.

3.6 Unitary and Other Algebras

Finite-dimensional simple Lie algebras have been classified. Let us review the classification and then discuss the unitary algebras in more detail.

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)	(3.69)
\mathfrak{c}_n	$\mathfrak{sp}(2n)$	symplectic	
\mathfrak{d}_n	$\mathfrak{so}(2n)$	orthogonal (even)	

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 $\mathfrak{sl}(N)$

$$A \in \mathfrak{sl}(N) : \quad \text{Tr } A = 0. \quad (3.70)$$

- Anti-symmetric $N \times N$ matrices w.r.t. an invertible symmetric metric M (e.g. $M = 1$) close under taking commutators. Therefore they form the orthogonal Lie algebra $\mathfrak{so}(N)$

$$A \in \mathfrak{so}(N) : \quad A^T = -MAM^{-1}, \quad M^T = M. \quad (3.71)$$

- 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 $\mathfrak{sp}(N)$

$$A \in \mathfrak{sp}(N) : \quad A^T = -EAE^{-1}, \quad E^T = -E. \quad (3.72)$$

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.¹⁴ 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

$$\begin{array}{ll}
 \mathfrak{a}_n: & \text{---} \circ \text{---} \circ \text{---} \cdots \cdots \cdots \circ \text{---} \circ \text{---} \circ \\
 \mathfrak{b}_n: & \text{---} \circ \text{---} \circ \text{---} \cdots \cdots \cdots \circ \text{---} \circ \text{---} \Rightarrow \circ \\
 \mathfrak{c}_n: & \text{---} \circ \text{---} \circ \text{---} \cdots \cdots \cdots \circ \text{---} \circ \text{---} \Leftarrow \circ \\
 \mathfrak{d}_n: & \text{---} \circ \text{---} \circ \text{---} \cdots \cdots \cdots \circ \text{---} \circ \text{---} \begin{array}{l} \circ \\ \circ \end{array} \\
 \mathfrak{e}_6: & \begin{array}{c} \circ \\ \text{---} \circ \text{---} \circ \text{---} \circ \text{---} \circ \text{---} \circ \end{array} \\
 \mathfrak{e}_7: & \begin{array}{c} \circ \\ \text{---} \circ \text{---} \circ \text{---} \circ \text{---} \circ \text{---} \circ \end{array} \\
 \mathfrak{e}_8: & \begin{array}{c} \circ \\ \text{---} \circ \text{---} \circ \text{---} \circ \text{---} \circ \text{---} \circ \text{---} \circ \end{array} \\
 \mathfrak{f}_4: & \text{---} \circ \text{---} \circ \text{---} \Rightarrow \circ \text{---} \circ \\
 \mathfrak{g}_2: & \text{---} \circ \text{---} \Rightarrow \Rightarrow \circ
 \end{array} \tag{3.73}$$

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, the unitary algebra is a real form of the complex linear group.

Unitary Algebras. Among the simple Lie groups, the special unitary groups $SU(N)$ have the simplest structure. Moreover, the groups with $N = 2$ and $N = 3$ along with the abelian group $U(1)$ are the gauge groups of the standard model. Let us therefore discuss some features of the special unitary group and its algebra.

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 $U(N)$ is N^2 . Triviality of the determinant removes another real degree of freedom so that

¹⁴This is merely a teaser. We will not enter Lie theory as deeply in this course.

the dimension of $SU(N)$ is $N^2 - 1$.

$$D^{\text{ad}} = N^2 - 1. \quad (3.74)$$

The corresponding Lie algebra $\mathfrak{su}(N)$ is the commutator algebra of traceless anti-hermitian matrices A . Correspondingly, they have $N^2 - 1$ real degrees of freedom.

Since the Lie group and algebra are both formulated in terms of matrices, we know that

$$G, \mathfrak{g} \subset \text{End}(\mathbb{C}^N). \quad (3.75)$$

Therefore it is straight-forward to construct an N -dimensional representation, the so-called defining or fundamental representation (with $v \in \mathbb{C}^N$)

$$\text{def}(A)v = Av, \quad \text{def}(U)v = Uv, \quad D^{\text{def}} = N. \quad (3.76)$$

There is in fact another N -dimensional representation, the conjugate defining representation

$$\overline{\text{def}}(A)v = A^*v = -A^\top v, \quad \overline{\text{def}}(U)v = U^*v, \quad D^{\overline{\text{def}}} = N. \quad (3.77)$$

The adjoint representation is a representation on the $(N^2 - 1)$ -dimensional space \mathfrak{g} ¹⁵

$$\text{ad}(A)B = [A, B], \quad \text{Ad}(U)B = UBU^{-1}. \quad (3.78)$$

So we know that the unitary algebras have at least four (irreducible) representations: trivial, adjoint, defining and conjugate defining. Tensor products of these lead to infinitely many more irreducible representations.

Tensor Product Representations. 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

$$\text{def} \otimes \text{def} = \text{sym} \oplus \text{alt}. \quad (3.79)$$

Here sym and alt denote the symmetric and anti-symmetric or alternating components of the full tensor product. For the unitary algebras these representations turn out to be irreducible.

The same applies to the tensor product of two conjugate defining representations,

$$\overline{\text{def}} \otimes \overline{\text{def}} = \overline{\text{sym}} \oplus \overline{\text{alt}}. \quad (3.80)$$

¹⁵It is sometime 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 a $(N^2 - 1)$ -dimensional vector space rather than a subspace of matrices.

Here $\overline{\text{sym}}$ and $\overline{\text{alt}}$ denote the conjugate representations of sym and alt, respectively. The tensor product of mixed defining representations turns out to decompose into irreps as well

$$\text{def} \otimes \overline{\text{def}} = \text{ad} \oplus \text{triv}. \quad (3.81)$$

This decomposition does not follow from general arguments such as the tensor products of identical representations. It is a characteristic feature of the unitary algebras $\mathfrak{su}(N)$.

Finite-dimensional irreps of the unitary algebras $\mathfrak{su}(N)$ have been classified and their tensor products are well understood. A useful tool in this regard are Young diagrams (or tableaux) and Dynkin labels.¹⁶

Young diagrams are heaps of boxes. They describe certain symmetrisation classes or equivalently representations of the symmetric group. 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 written as a dot. For example, for $N = 5$

$$\text{def} = \square, \quad \overline{\text{def}} = \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array}, \quad \text{ad} = \begin{array}{|c|} \hline \square \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array}, \quad \text{triv} = \bullet. \quad (3.82)$$

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

$$\begin{aligned} \square \otimes \square &= \square \square \oplus \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array}, \\ \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \otimes \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} &= \begin{array}{|c|} \hline \square \square \\ \hline \square \square \\ \hline \square \square \\ \hline \square \square \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array}, \\ \square \otimes \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} &= \begin{array}{|c|} \hline \square \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \oplus \bullet. \end{aligned} \quad (3.83)$$

Young tableaux 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 rewriting of the shape of Young tableaux in terms of a set of integer numbers, e.g. for $N = 5$

$$\text{def} = [1000], \quad \overline{\text{def}} = [0001], \quad \text{ad} = [1001]. \quad (3.84)$$

Each of these numbers is associated to one of the dots of the Dynkin diagram for the Lie algebra.

$$[abcd] \equiv \overset{a}{\circ} - \overset{b}{\circ} - \overset{c}{\circ} - \overset{d}{\circ}. \quad (3.85)$$

Therefore the classification in terms of Dynkin labels applies to all simple Lie algebras, and not just to $\mathfrak{su}(N)$.

¹⁶This 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 or fundamental representation of $\mathfrak{su}(N)$ maps to $N \times N$ anti-hermitian traceless matrices. The representation T_a^{def} of the basis T_a therefore consists of hermitian traceless matrices

$$(T_a^{\text{def}})^\dagger = T_a^{\text{def}}, \quad \text{Tr } T_a^{\text{def}} = 0. \quad (3.86)$$

This is one of the standard representations in which unitary gauge theories are formulated. It is also the representation in which the quark fields of quantum chromodynamics transform.

There is also the conjugate defining representation

$$\overline{T_a^{\text{def}}} = -(T_a^{\text{def}})^* = -(T_a^{\text{def}})^\top. \quad (3.87)$$

It has very much the same properties as the original defining representation, but it is not equivalent to it.

The defining representation has a completeness relation. It can be formulated in terms of two trace identities which hold for any $N \times N$ matrix X

$$\begin{aligned} k^{ab} \text{Tr}(T_a^{\text{def}} X) T_b^{\text{def}} &= B^{\text{def}} (X - N^{-1} \text{Tr } X), \\ k^{ab} T_a^{\text{def}} X T_b^{\text{def}} &= B^{\text{def}} (\text{Tr } X - N^{-1} X). \end{aligned} \quad (3.88)$$

These identities are useful in simplifying various expressions that appear in gauge theory.

Example. We can express the structure constants as a trace

$$\text{Tr}([T_a^{\text{def}}, T_b^{\text{def}}] T_c^{\text{def}}) = i B^{\text{def}} f_{abc}. \quad (3.89)$$

For example, the contractions of two structure constants then reads

$$\begin{aligned} & - (B^{\text{def}})^2 k^{cf} f_{abc} f_{def} \\ &= k^{cf} \text{Tr}([T_a^{\text{def}}, T_b^{\text{def}}] T_c^{\text{def}}) \text{Tr}([T_d^{\text{def}}, T_e^{\text{def}}] T_f^{\text{def}}) \\ &= B^{\text{def}} \text{Tr}([T_a^{\text{def}}, T_b^{\text{def}}][T_d^{\text{def}}, T_e^{\text{def}}]). \end{aligned} \quad (3.90)$$

The above completeness relation also has a useful graphical representation reminiscent of Feynman graphs and rules. Both identities can be written as

$$\text{Diagram} = B^{\text{def}} \text{Diagram} - \frac{B^{\text{def}}}{N} \text{Diagram} \quad (3.91)$$

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 useful consequence of the completeness relations for the defining representation is that the anti-commutator of two hermitian matrices is again hermitian. After removing the trace one can write it in terms of the defining representation of $\mathfrak{su}(N)$

$$\{T_a^{\text{def}}, T_b^{\text{def}}\} = d_{ab}{}^c T_c^{\text{def}} + 2N^{-1} B^{\text{def}} k_{ab}. \quad (3.92)$$

This is interesting because it allows to define a set of totally symmetric cubic structure constants d_{abc}

$$\text{Tr}(\{T_a^{\text{def}}, T_b^{\text{def}}\} T_c^{\text{def}}) = B^{\text{def}} d_{abc}. \quad (3.93)$$

These exist only for the unitary algebras with $N > 2$. For the orthogonal and symplectic algebras this invariant vanishes identically.¹⁷

Invariants and Constants. 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 central element of the enveloping algebra

$$C_3 = d^{abc} T_a T_b T_c \in U(\mathfrak{su}(N)). \quad (3.94)$$

The following table summarises the constants B^R , C_2^R and C_3^R relative to B^{def} for a few simple representations

R	D^R	$B^R N / B^{\text{def}}$	$C_2^R N / B^{\text{def}}$	$C_3^R N^2 / (B^{\text{def}})^2$
def	N	N	$(N^2 - 1)$	$+(N^2 - 4)(N^2 - 1)$
$\overline{\text{def}}$	N	N	$(N^2 - 1)$	$-(N^2 - 4)(N^2 - 1)$
ad	$N^2 - 1$	$2N^2$	$2N^2$	0
sym	$\frac{1}{2}N(N+1)$	$N(N+2)$	$2(N-1)(N+2)$	$2(N^2-4)(N-1)(N+4)$
alt	$\frac{1}{2}N(N-1)$	$N(N-2)$	$2(N+1)(N-2)$	$2(N^2-4)(N+1)(N-4)$

(3.95)

The representations listed are the defining (def), the conjugate defining ($\overline{\text{def}}$), the adjoint (ad) as well as the symmetric (sym) and antisymmetric (alt) products of two defining def, see below.

As explained above, we have the freedom to set B^{def} to any desirable value. It is often declared to equal $1/2$. Another good choice is

$$B^{\text{def}} = \frac{1}{N}. \quad (3.96)$$

Some other useful choices are 1 or $1/2N$.

The 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.

¹⁷The only exception is $\mathfrak{so}(6)$ which is isomorphic to $\mathfrak{su}(4)$.

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

$$T_a^{\text{def}} = \frac{1}{2}\sigma_a. \quad (3.97)$$

A special feature of $\mathfrak{su}(2)$ is that the defining representation is equivalent to the conjugate defining representation, namely

$$T_a^{\overline{\text{def}}} = \sigma_2^{-1} T_a^{\text{def}} \sigma_2. \quad (3.98)$$

More concretely we have the following relations¹⁸

$$\text{def} \simeq \overline{\text{def}}, \quad \text{sym} \simeq \overline{\text{sym}} \simeq \text{ad}, \quad \text{alt} \simeq \overline{\text{alt}} \simeq \text{triv}. \quad (3.99)$$

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 $a = 1, \dots, 8 = D^{\text{ad}}$. For the purposes of Yang–Mills theory it is almost always inconvenient to use these matrices directly rather than their complement relations.

For $N \geq 3$ the two defining representations are inequivalent. However, some relations remain. For $N = 3$ they include

$$\text{alt} \simeq \overline{\text{def}}, \quad \overline{\text{alt}} \simeq \text{def}. \quad (3.100)$$

For $N = 4$ there is yet a different set of relations, in particular $\text{alt} \simeq \overline{\text{alt}}$.

Orthogonal and Symplectic Algebras. Superficially, the orthogonal and symplectic algebras are close relatives of the unitary algebras. There are, however, important differences that distinguish them. Let us just mention a few of these.

The orthogonal and symplectic algebras merely have a single defining representation rather than the pair observed for unitary algebras.

The tensor products of two (identical) defining representations now splits into three irreps: For the orthogonal algebras the symmetric representation further splits into a trivial and a traceless component. Furthermore the anti-symmetric component is equivalent to the adjoint representation

$$\text{def} \otimes \text{def} = (\text{sym}' \oplus \text{triv}) \oplus \text{ad}. \quad (3.101)$$

The appearance of the trivial sub-representation is related to the existence of the invariant metric tensor that defines orthogonality.

The situation is opposite for the symplectic algebras: Here the anti-symmetric representation further splits into a trivial and a traceless component. Furthermore the symmetric component is equivalent to the adjoint representation

$$\text{def} \otimes \text{def} = \text{ad} \oplus (\text{alt}' \oplus \text{triv}). \quad (3.102)$$

¹⁸Note that these relations are consistent with the above table of characteristic constants for representation when $N = 2$.

The appearance of the trivial sub-representation is related to the existence of the invariant tensor that defines the symplectic structure.

Most of the properties, relations and tools of unitary algebras generalise in some way to the orthogonal and symplectic algebras. Usually they will be somewhat more elaborate than their basic unitary counterparts.

4 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 charged vector bosons W^\pm . Yang–Mills theory is also the foundation for quantum chromodynamics.

4.1 Classical Gauge Theory

Spinor 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

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \quad (4.1)$$

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

$$F_{\mu\nu} = (\partial_\mu A_\nu) - (\partial_\nu A_\mu), \quad D_\mu = \partial_\mu - iqA_\mu. \quad (4.2)$$

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 $\alpha(x)$

$$\begin{aligned} \psi'(x) &= \exp(+iq\alpha(x))\psi(x), \\ \bar{\psi}'(x) &= \exp(-iq\alpha(x))\bar{\psi}(x), \\ A'_\mu(x) &= A_\mu(x) + (\partial_\mu\alpha(x)). \end{aligned} \quad (4.3)$$

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

$$D'_\mu = \exp(+iq\alpha(x))D_\mu \exp(-iq\alpha(x)). \quad (4.4)$$

The covariant derivative simply pushes the 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 invariant¹

$$F_{\mu\nu} = iq^{-1} [D_\mu, D_\nu], \quad F'_{\mu\nu} = F_{\mu\nu}. \quad (4.5)$$

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

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

$$- m \bar{\psi}_\alpha \psi^\alpha. \quad (4.6)$$

It is not only invariant under N copies of U(1) transformations $\psi^\alpha \rightarrow \exp(iq\alpha_\alpha)\psi^\alpha$, but it rather has the larger symmetry

$$\psi'^\alpha = U^\alpha_\beta \psi^\beta, \quad \bar{\psi}'_\alpha = (U^*)_\alpha^\beta \bar{\psi}_\beta = \bar{\psi}_\beta (U^\dagger)^\beta_\alpha, \quad (4.7)$$

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)

$$\bar{\psi}' \psi' = \bar{\psi} U^\dagger U \psi = \bar{\psi} U^{-1} U \psi = \bar{\psi} \psi. \quad (4.8)$$

The unitary matrices form a group since the identity matrix is evidently unitary, inverses exist ($U^{-1} = U^\dagger$) and since

$$(U_1 U_2)^\dagger = U_2^\dagger U_1^\dagger = U_2^{-1} U_1^{-1} = (U_1 U_2)^{-1}. \quad (4.9)$$

This group is the Lie group called U(N).

Non-Abelian Gauge Potential. In the above discussion we can now replace the transformation factor $\exp(iq\alpha(x))$ by $U(x)$ everywhere. We merely have to pay attention that, unlike the phase factor $\exp(iq\alpha)$, matrices do not commute in general.

$$\psi'(x) = U(x)\psi(x), \quad \bar{\psi}'(x) = \bar{\psi}(x)U(x)^{-1}. \quad (4.10)$$

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

$$D'_\mu = U D_\mu U^{-1}. \quad (4.11)$$

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)

$$D_\mu = \partial_\mu - iA_\mu. \quad (4.12)$$

¹Here, the covariant derivatives are defined as operators which act on everything on their right unless restricted by brackets. Alternatively, one could supplement the first equation by some function or field to the right of $F_{\mu\nu}$ and $[D_\mu, D_\nu]$ on which the derivatives act. In that case the equation should hold for any such function.

The gauge potential A_μ must therefore transform according to

$$A'_\mu = UA_\mu U^{-1} + iU(\partial_\mu U^{-1}). \quad (4.13)$$

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

$$(A'_\mu)^\dagger = UA_\mu^\dagger U^{-1} - i(\partial_\mu U)U^{-1}. \quad (4.14)$$

We note that $0 = (\partial_\mu(UU^{-1})) = (\partial_\mu U)U^{-1} + U(\partial_\mu U^{-1})$. Altogether it shows that A_μ^\dagger transforms exactly as A_μ itself, and therefore the gauge transformation is consistent with the assumption of a hermitian gauge potential

$$A_\mu^\dagger = A_\mu. \quad (4.15)$$

It therefore makes sense to impose this property on the gauge field in order to reduce its independent components to a minimum.

Non-Abelian Gauge Theory. Finally we can define the field strength as before

$$F_{\mu\nu} = i[D_\mu, D_\nu] = (\partial_\mu A_\nu) - (\partial_\nu A_\mu) - i[A_\mu, A_\nu]. \quad (4.16)$$

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

$$F'_{\mu\nu} = UF_{\mu\nu}U^{-1}. \quad (4.17)$$

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 $F^{\mu\nu}F_{\mu\nu}$ is only gauge covariant; putting it inside a trace makes it properly invariant

$$\mathcal{L}_{\text{YM}} = -\frac{1}{2g_{\text{YM}}^2} \text{Tr} F^{\mu\nu} F_{\mu\nu}. \quad (4.18)$$

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_{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



$$(4.19)$$

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

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

$$\mathcal{L}_{\text{spinor}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi. \quad (4.20)$$

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.



Scalar matter can just as well be coupled to the gauge field. For a vector of complex scalars we would write

$$\mathcal{L}_{\text{scalar}} = -(\bar{D}^\mu \bar{\phi})(D_\mu \phi) - m^2 \bar{\phi}\phi. \quad (4.22)$$

The kinetic term for the scalar field (dashed line) necessarily comes along with a cubic and a quartic vertex.



Infinitesimal Gauge Transformations. 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.

Finite transformations can be written using the exponential map of a matrix of transformation parameters α

$$U = \exp(i\alpha) \quad (4.24)$$

Unitarity of U translates to hermiticity of α

$$\alpha^\dagger = \alpha. \quad (4.25)$$

The transformation rules can be summarised as follows

$$\begin{aligned} \psi' &= U\psi, \\ \bar{\psi}' &= \bar{\psi}U^{-1}, \\ A'_\mu &= UA_\mu U^{-1} + iU(\partial_\mu U^{-1}), \\ D'_\mu &= UD_\mu U^{-1}, \\ F'_{\mu\nu} &= UF_{\mu\nu}U^{-1}. \end{aligned} \quad (4.26)$$

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

$$U = 1 + i\delta\alpha. \quad (4.27)$$

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

$$\begin{aligned}
\delta\psi' &= i\delta\alpha\psi, \\
\delta\bar{\psi}' &= -i\bar{\psi}\delta\alpha, \\
\delta A_\mu &= (\partial_\mu\delta\alpha) + i[\delta\alpha, A_\mu] = [D_\mu, \delta\alpha], \\
\delta D_\mu &= i[\delta\alpha, D_\mu], \\
\delta F_{\mu\nu} &= i[\delta\alpha, F_{\mu\nu}].
\end{aligned} \tag{4.28}$$

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

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 derives from the Jacobi identity of three covariant derivatives. For Yang–Mills theory it can be written alternatively as

$$[D_\rho, F_{\mu\nu}] + [D_\nu, F_{\rho\mu}] + [D_\mu, F_{\nu\rho}] = 0 \quad \text{or} \quad \varepsilon^{\mu\nu\rho\sigma}[D_\rho, F_{\mu\nu}] = 0. \tag{4.29}$$

Here the covariant derivative of a gauge covariant field X such as the field strength $F_{\mu\nu}$ can be written as a commutator³

$$[D_\mu, X] = \partial_\mu X - i[A_\mu, X]. \tag{4.30}$$

The extension of the inhomogeneous Maxwell equation reads

$$[D_\mu, F^{\mu\nu}] = J^\nu. \tag{4.31}$$

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

$$(J^\nu)^\alpha_\beta = -\bar{\psi}_\beta \gamma^\nu \psi^\alpha. \tag{4.32}$$

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

$$(D^\mu \gamma_\mu - m)^\alpha_\beta \psi^\beta = 0. \tag{4.33}$$

Note that the equations of motion imply that the current is covariantly conserved

$$[D_\mu, J^\mu] = 0. \tag{4.34}$$

³In physics, the definition of the covariant derivative $D_\mu X$ is often assumed to depend implicitly on the transformation rules of the field X . In this notation we may also write $(D_\mu X)$ instead of $[D_\mu, X]$ for a field X that transforms by conjugation with U .

4.2 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.

Naive Path Integral. Consider the action for the electromagnetic field with sources ⁴

$$S_{\text{EM}}[A, J] = \int d^4x \left(-\frac{1}{4} F^{\mu\nu} F_{\mu\nu} - J^\mu A_\mu \right). \quad (4.35)$$

The action is invariant under gauge transformations

$$S_{\text{EM}}[A', J] = S_{\text{EM}}[A, J], \quad (4.36)$$

provided that the source field is conserved,

$$\partial_\mu J^\mu = 0. \quad (4.37)$$

Naively, the path integral reads

$$Z[J] = \int DA \exp(iS_{\text{EM}}[A, J]). \quad (4.38)$$

Gauge invariance of the action implies that the integrand of the path integral is constant along all gauge-equivalent field configurations A

$$\begin{aligned} Z[J] &= \int DA \exp(iS_{\text{EM}}[A, J]) \\ &= \left(\int D\alpha \right) \cdot \int D[A] \exp(iS_{\text{EM}}) = \infty. \end{aligned} \quad (4.39)$$

Since these gauge orbits are infinite-dimensional spaces, the path integral will formally not converge.⁵ In other words, the path integral over the gauge potential A is badly over-counting the admissible field configurations. We would like to divide out the divergent integral $\int D\alpha$ over gauge orbits, and keep just the integral $\int D[A]$ over gauge equivalence classes, but a priori we do not understand how to do it in practice.

⁴The sources can be an external field or represent couplings to matter fields such as $J^\mu = \bar{\psi} \gamma^\mu \psi$. 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.

⁵One may of course argue that the path integral diverges anyway, but this type of divergence can be considered worse.

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 vanishes for one point in the gauge orbit. For example, we could use the functional map $G[A](x) = \partial^\mu A_\mu(x)$ which enforces the Lorenz gauge.⁶ In fact, we will use a slightly more general gauge⁷

$$G[A, \Omega](x) = \partial^\mu A_\mu(x) - \Omega(x). \quad (4.40)$$

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

$$\int DA \Delta[G[A, \Omega]] \exp(iS_{\text{EM}}[A, J]). \quad (4.41)$$

However, this insertion has the undesirable feature that it is not evidently related to the original formulation of $Z[J]$. Furthermore, the result depends on the functional derivative of G at 0.⁸

Instead, let us consider the identity⁹

$$1 = \int D\alpha \Delta[G[A', \Omega]] \text{Det}_{x,y} \left(\frac{\delta G[A', \Omega](x)}{\delta \alpha(y)} \right), \quad (4.42)$$

where A' is the gauge potential A transformed by the field α

$$A'_\mu = A_\mu + (\partial_\mu \alpha). \quad (4.43)$$

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

$$\frac{\delta G[A', \Omega](x)}{\delta \alpha(y)} = \partial^2 \delta(x - y). \quad (4.44)$$

Hence we can write the identity as

$$1 = \text{Det}(\partial^2) \int D\alpha DA \Delta[G[A', \Omega]]. \quad (4.45)$$

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

$$Z[J] = \text{Det}(\partial^2) \int D\alpha DA \Delta[G[A', \Omega]] \exp(iS_{\text{EM}}[A, J]). \quad (4.46)$$

⁶This condition almost fixes the gauge up to some residual gauge transformations which are acceptable.

⁷Other gauges are acceptable as well, but the following discussion may have to be adapted.

⁸This is the analog of the reparametrisation rule for delta functions $\delta(f(x)) = \delta(x - x_0)/|f'(x_0)|$ where $f(x_0) = 0$.

⁹The functional determinant is approximated by an ordinary determinant when the coordinates x and y are discretised.

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 DA and $S_{\text{EM}}[A, J]$ remain unchanged

$$Z[J] = \text{Det}(\partial^2) \left(\int D\alpha \right) \int DA \Delta[G[A, \Omega]] \exp(iS_{\text{EM}}[A, J]). \quad (4.47)$$

We have now separated the path integral over 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

$$S_\Omega[\Omega] = -\frac{1}{2\xi} \int d^4x \Omega^2, \quad \int D\Omega \exp(iS_\Omega[\Omega]) = \text{const.} \quad (4.48)$$

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

$$Z[J] \sim \int D\Omega DA \Delta[G[A, \Omega]] \exp(iS_{\text{EM}}[A, J] + iS_\Omega[\Omega]). \quad (4.49)$$

We now use the delta functional to perform the path integral over Ω and obtain

$$Z[J] \sim \int DA \exp(iS_{\text{EM}}[A, J] + iS_{\text{gf}}[A]) \quad (4.50)$$

with an additional contribution to the action

$$S_{\text{gf}}[A] = -\frac{1}{2\xi} \int d^4x (\partial \cdot A)^2. \quad (4.51)$$

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

$$\begin{aligned} Z[J] &= \int DA \exp(iS_{\text{EM}}[A, J]) \\ &= \text{const.} \cdot \int DA \exp(iS_{\text{EM}}[A, J] + iS_{\text{gf}}[A]). \end{aligned} \quad (4.52)$$

Since all constant prefactor drop out of correlation function, 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.¹⁰

We have used a modified Lorentz 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. There is no need to impose the gauge condition.

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

$$(G_{\text{F}})_{\mu\nu}(p) = \frac{\eta_{\mu\nu} - (1 - \xi)p_{\mu}p_{\nu}/p^2}{p^2 - i\epsilon}. \quad (4.53)$$

Note that this propagator does not respect the Lorenz gauge exactly since $\partial \cdot A$ in correlators yields the propagator

$$p^{\nu}(G_{\text{F}})_{\mu\nu}(p) = \frac{\xi p_{\mu}}{p^2 - i\epsilon} \quad (4.54)$$

which does not vanish identically. This is an effect of the generalised Lorenz gauge. In practice, the non-vanishing does not matter since the other side of the propagator usually couples to the source J^{μ} which is conserved, $p_{\mu}J^{\mu}(p) = 0$.

4.3 Yang–Mills Quantisation

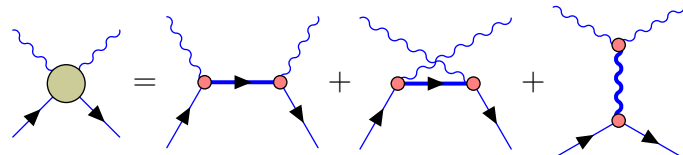
We are now prepared to apply to the path integral quantisation to the vector fields of Yang–Mills theory. Here some further complications arise due to the non-abelian nature of the gauge group. These lead to the introduction of further fields.

Some Puzzles. If we are 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,

$$Z[J] = \text{const.} \cdot \int DA \exp(iS_{\text{YM}}[A, J] + iS_{\text{gf}}[A]). \quad (4.55)$$

we run into some problems. These can be seen in several ways:

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



$$= \text{[Three diagrams with wavy lines]} \quad (4.56)$$

In the abelian process the resulting photons automatically have only physical polarisation vectors. This turns out not to hold in the non-abelian generalisation,

¹⁰Of course we may also compute more general correlation functions, but these are potentially unrelated to the original path integral and thus meaningless.

some of the produced vector particles have unphysical polarisations. They arise 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 applied to gauge-invariant quantities and configurations only. Nevertheless, this interpretation causes problems w.r.t. unitarity: The optical theorem relates the square of the above tree process to the imaginary part of a loop process.

$$\text{Tree process with cut} \stackrel{?}{=} \text{Im} \text{ Loop process} \quad (4.57)$$

- 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

$$M_{\mu\nu}^{(1)} = \text{Tree with loop} + \text{Tree with gluon loop} + \text{Tree with ghost loop} \quad (4.58)$$

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

$$M_{\mu\nu}^{(1)}(p) = M^{(1)}(p)(p^2\eta_{\mu\nu} - p_\mu p_\nu), \quad p^\nu M_{\mu\nu}^{(1)}(p) = 0. \quad (4.59)$$

The non-abelian result, however, is not transverse

$$M_{\mu\nu}^{(1)}(p) = M_1^{(1)}(p)p^2\eta_{\mu\nu} + M_2^{(1)}(p)p_\mu p_\nu, \quad p^\nu M_{\mu\nu}^{(1)}(p) \neq 0. \quad (4.60)$$

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 reads with source terms reads

$$S_{\text{YM}}[A, J] = -\frac{1}{2g_{\text{YM}}^2} \int d^4x \text{Tr} F_{\mu\nu} F^{\mu\nu} - \int d^4x \text{Tr} J^\mu A_\mu. \quad (4.61)$$

We can choose the same gauge fixing function as before

$$G[A, \Omega](x) = \partial^\mu A_\mu(x) - \Omega(x), \quad (4.62)$$

and fix the gauge by means of the identity

$$1 = \int DU \Delta[G[A', \Omega]] \text{Det}_{(xa,yb)} \left(\frac{\delta G[A', \Omega]^a(x)}{\delta \alpha^b(y)} \right). \quad (4.63)$$

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 , $D(UV) = DU = D(VU)$. Moreover, the variation δU of a group element U is expressed through the algebra element $\delta \alpha$ as

$$\delta U = i\delta \alpha U. \quad (4.64)$$

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

$$A'_\mu = U A_\mu U^{-1} + iU(\partial_\mu U^{-1}). \quad (4.65)$$

Its variation can be expressed as a covariant derivative of $\delta \alpha$

$$\delta A'_\mu = (\partial_\mu \delta \alpha) - i[A'_\mu, \delta \alpha] = [D'_\mu, \delta \alpha] = ((D'_\mu)^{\text{ad}} \delta \alpha). \quad (4.66)$$

Hence we can write the identity as

$$1 = \int DU \Delta[G[A', \Omega]] \text{Det}(\partial^\mu (D'_\mu)^{\text{ad}}). \quad (4.67)$$

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

We insert the identity into the path integral

$$\begin{aligned} Z[J] &= \int DU DA \Delta[G[A', \Omega]] \text{Det}(\partial^\mu (D'_\mu)^{\text{ad}}) \\ &\quad \cdot \exp(iS_{\text{YM}}[A, J]). \end{aligned} \quad (4.68)$$

Next we perform a gauge transformation by U^{-1} to bring all gauge potentials to the original form.¹¹

$$\begin{aligned} Z[J] &= \left(\int DU \right) \int DA \Delta[G[A, \Omega]] \text{Det}(\partial^\mu D_\mu^{\text{ad}}) \\ &\quad \cdot \exp(iS_{\text{YM}}[A, J]). \end{aligned} \quad (4.69)$$

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 fermionic fields \bar{C}, C

¹¹Strictly speaking the source term is not invariant under gauge transformations. We can view the source J as representing gauge-invariant couplings to other fields or other gauge-invariant operators. In that sense the source term can be considered perfectly invariant.

belonging to the same space as each component A_μ of the gauge potential. The field C is called the ghost field, \bar{C} is called the anti-ghost. Define its action as a quadratic functional of the new fields¹²

$$\begin{aligned} S_{\text{gh}}[C, \bar{C}, A] &= \frac{2}{g_{\text{YM}}^2} \int d^4x \text{Tr} \bar{C} (\partial^\mu D_\mu^{\text{ad}} C) \\ &= \frac{2}{g_{\text{YM}}^2} \int d^4x \text{Tr} \bar{C} (\partial^\mu [D_\mu, C]). \end{aligned} \quad (4.70)$$

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

$$\int DC D\bar{C} \exp(iS_{\text{gh}}[C, \bar{C}, A]) = \text{Det}(\partial^\mu (D'_\mu)^{\text{ad}}). \quad (4.71)$$

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

$$\begin{aligned} Z[J] &= \int DC D\bar{C} DA \Delta[G[A, \Omega]] \\ &\quad \cdot \exp(iS_{\text{YM}}[A, J] + iS_{\text{gh}}[C, \bar{C}, A]). \end{aligned} \quad (4.72)$$

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

$$\begin{aligned} Z[J] &= \int DC D\bar{C} DA \\ &\quad \cdot \exp(iS_{\text{YM}}[A, J] + iS_{\text{gh}}[C, \bar{C}, A] + iS_{\text{gf}}[A]). \end{aligned} \quad (4.73)$$

with gauge fixing term

$$S_{\text{gf}}[A] = -\frac{1}{\xi g_{\text{YM}}^2} \int d^4x \text{Tr}(\partial \cdot A)^2. \quad (4.74)$$

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.

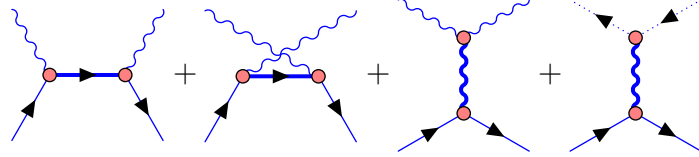
For example, the one-loop contribution to the two-point effective vertex receives another contribution from a ghost loop (dotted line)

$$M_{\mu\nu}^{(1)} = \text{diagram 1} + \text{diagram 2} + \text{diagram 3} + \text{diagram 4}. \quad (4.75)$$

¹²The normalisation of a quadratic term is inessential because it only affects a constant prefactor.

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


(4.76)

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.

4.4 Feynman Rules

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

Gauge Sector. First we write the fields in components of the Lie algebra. In the above considerations we have assumed that the fields A_μ , $F_{\mu\nu}$, C and \bar{C} are $N \times N$ hermitian matrices. In other words we have implicitly used the defining representation of $U(N)$. We furthermore would like to restrict to traceless matrices, and the corresponding group $SU(N)$.

We would like to obtain the Feynman rules for the most general case. Therefore we write the fields A_μ^a in a basis T_a of the Lie algebra \mathfrak{g} . The following relations express the matrix-values fields in terms of the defining representation of $SU(N)$.

$$\begin{aligned} A_\mu &= gT_a^{\text{def}} A_\mu^a, & C &= gT_a^{\text{def}} C^a, \\ F_{\mu\nu} &= gT_a^{\text{def}} F_{\mu\nu}^a, & \bar{C} &= gT_a^{\text{def}} \bar{C}^a. \end{aligned} \quad (4.77)$$

Here g is a normalisation factor that we will later adjust to be the coupling constant of the gauge field.

Let us first express the field strength in terms of components

$$\begin{aligned} F_{\mu\nu} &= (\partial_\mu A_\nu) - (\partial_\nu A_\mu) - i[A_\mu, A_\nu] \\ &= gT_a^{\text{def}} (\partial_\mu A_\nu^a) - gT_a^{\text{def}} (\partial_\nu A_\mu^a) - ig^2 [T_b^{\text{def}}, T_c^{\text{def}}] A_\mu^b A_\nu^c \\ &= gT_a^{\text{def}} ((\partial_\mu A_\nu^a) - (\partial_\nu A_\mu^a) + gf_{bc}^a A_\mu^b A_\nu^c). \end{aligned} \quad (4.78)$$

Hence we find

$$F_{\mu\nu}^a = (\partial_\mu A_\nu^a) - (\partial_\nu A_\mu^a) + gf_{bc}{}^a A_\mu^b A_\nu^c. \quad (4.79)$$

The Yang–Mills Lagrangian now reads

$$\begin{aligned} \mathcal{L}_{\text{YM}} &= -\frac{1}{2g_{\text{YM}}^2} \text{Tr} F^{\mu\nu} F_{\mu\nu} \\ &= -\frac{g^2}{2g_{\text{YM}}^2} \text{Tr}(T_a^{\text{def}} T_b^{\text{def}}) F^{a,\mu\nu} F_{\mu\nu}^b \\ &= -\frac{g^2 B^{\text{def}}}{2g_{\text{YM}}^2} k_{ab} F^{a,\mu\nu} F_{\mu\nu}^b. \end{aligned} \quad (4.80)$$

In order to obtain a canonical prefactor we shall set¹³ $g = g_{\text{YM}}/\sqrt{2B^{\text{def}}}$. Altogether we have

$$\mathcal{L}_{\text{YM}} = -\frac{1}{4} k_{ab} F^{a,\mu\nu} F_{\mu\nu}^b. \quad (4.81)$$

This is similar to one copy of the photon Lagrangian for each of the components of the Lie algebra. Note, however, that the field strength $F_{\mu\nu}^a$ now depends non-linearly on the gauge potential A_μ^a .

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

$$\begin{aligned} \mathcal{L}_{\text{gh}} &= k_{ab} \bar{C}^a (\partial^\mu (D_\mu^{\text{ad}} C)^b), \\ \mathcal{L}_{\text{gf}} &= -\frac{1}{2} \xi^{-1} k_{ab} (\partial^\mu A_\mu^a) (\partial^\nu A_\nu^b). \end{aligned} \quad (4.82)$$

Here the covariant derivative of the ghost field C is defined as

$$(D_\mu^{\text{ad}} C)^a = (\partial_\mu C^a) + gf_{bc}{}^a A_\mu^b C^c. \quad (4.83)$$

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 dashed lines. Their momentum space Feynman propagators read

$$\begin{aligned} \mu, a \begin{array}{c} p \\ \text{wiggly line} \\ \nu, b \end{array} &\rightarrow \frac{-ik^{ab}(\eta_{\mu\nu} - (1 - \xi)p_\mu p_\nu/p^2)}{p^2 - i\epsilon}, \\ a \begin{array}{c} p \\ \text{dashed line} \\ b \end{array} &\rightarrow \frac{-ik^{ab}}{p^2 - i\epsilon}. \end{aligned} \quad (4.84)$$

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

¹³For the choice $B^{\text{def}} = \frac{1}{2}$ this would simply be $g = g_{\text{YM}}$. However, we shall not fix B^{def} and continue to work with g instead of g_{YM} .

with a ghost line

$$\begin{aligned}
& \begin{array}{l} p_1, \mu, a \\ \text{---} \\ p_2, \nu, b \\ \text{---} \\ p_3, \rho, c \end{array} \rightarrow \begin{array}{l} -gf_{abc}(p_1 - p_2)^\rho \eta^{\mu\nu} \\ -gf_{abc}(p_2 - p_3)^\mu \eta^{\nu\rho} \\ -gf_{abc}(p_3 - p_1)^\nu \eta^{\rho\mu}, \end{array} \\
& \begin{array}{l} p_1, \mu, a \\ \text{---} \\ p_2, \nu, b \\ \text{---} \\ p_3, \rho, c \\ \text{---} \\ p_4, \sigma, d \end{array} \rightarrow \begin{array}{l} -ig^2 f_{abe} f_{cdf} k^{ef} (\eta^{\mu\rho} \eta^{\nu\sigma} - \eta^{\nu\rho} \eta^{\mu\sigma}) \\ -ig^2 f_{ace} f_{bdf} k^{ef} (\eta^{\mu\nu} \eta^{\rho\sigma} - \eta^{\nu\rho} \eta^{\mu\sigma}) \\ -ig^2 f_{ade} f_{bcf} k^{ef} (\eta^{\mu\nu} \eta^{\rho\sigma} - \eta^{\nu\sigma} \eta^{\mu\rho}), \end{array} \\
& \begin{array}{l} p_1, a \\ \text{---} \\ p_2, b \\ \text{---} \\ p_3, \rho, c \end{array} \rightarrow -gf_{abc} p_1^\mu. \tag{4.85}
\end{aligned}$$

Matter Couplings. In our previous discussion of matter fields we assumed that the gauge group is $SU(N)$ and that there are N matter fields. In other words, they transformed in the defining representation of $SU(N)$.

Let us now consider spinor matter fields transforming in some representation R of the gauge group. The kinetic term for spinor matter in Yang–Mills theory reads

$$\mathcal{L}_{\text{spinor}} = \bar{\psi}(i\gamma^\mu D_\mu^R - m)\psi. \tag{4.86}$$

We now expand the matter field $\psi = v_k \psi^k$ in a basis v_k of \mathbb{V} . The conjugate field $\bar{\psi} = \bar{\psi}_k v^k$ is expanded in the dual basis v^k of the dual space \mathbb{V}^* . The component Lagrangian now reads

$$\mathcal{L}_{\text{spinor}} = \bar{\psi}_k (i\gamma^\mu (D_\mu^R \psi)^k - m\psi^k) \tag{4.87}$$

with the matter covariant derivative

$$(D_\mu^R X)^k = (\partial_\mu X^k) - ig(T_a^R)_j{}^k A_\mu^a X^j. \tag{4.88}$$

This results in the following Feynman rules involving the spinor field

$$\begin{aligned}
& \begin{array}{l} \alpha, j \\ \text{---} \\ p \\ \text{---} \\ \beta, k \end{array} \rightarrow \frac{-i\delta_j^k ((\gamma^\mu)^\alpha{}_\beta p_\mu + \delta_\beta^\alpha m)}{p^2 + m^2 - i\epsilon}, \\
& \begin{array}{l} p_1, \alpha, j \\ \text{---} \\ p_2, \beta, k \\ \text{---} \\ p_3, \mu, a \end{array} \rightarrow ig(T_a^R)_j{}^k (\gamma^\mu)^\alpha{}_\beta. \tag{4.89}
\end{aligned}$$

The Feynman rules for a interactions with scalar fields are slightly different, so let us display this case briefly¹⁴

$$\begin{aligned}
\mathcal{L}_{\text{scalar}} &= -\frac{1}{2}(D^{R,\mu}\phi) \cdot (D_\mu^R \phi) - m^2 \phi \cdot \phi \\
&= -\frac{1}{2}(D^{R,\mu}\phi)^k (D_\mu^R \phi)^k - \frac{1}{2}m^2 \phi^k \phi^k. \tag{4.90}
\end{aligned}$$

¹⁴Here we assume a real representation R as opposed to the complex representation for spinors. We suppose an orthonormal basis $v_j \cdot v_k = \delta_{jk}$.

The resulting Feynman rules now include an extra quartic interaction $\phi^2 A^2$

$$\begin{aligned}
 j \text{ --- } p \text{ --- } k &\rightarrow \frac{-i\delta^{jk}}{p^2 + m^2 - i\epsilon}, \\
 \begin{array}{c} p_1, j \\ \diagdown \\ \text{---} \\ \diagup \\ p_2, k \end{array} &\rightarrow ig(T_a^R)_{jk}(p_1^\mu - p_2^\mu), \\
 \begin{array}{c} p_1, j \\ \diagdown \\ \text{---} \\ \diagup \\ p_2, k \end{array} &\rightarrow ig^2(T_a^R)_{jl}(T_b^R)_{kl}\eta^{\mu\nu}.
 \end{aligned} \tag{4.91}$$

Finally, we can discuss interaction between the fields. In renormalisable field theories, there are three types of interactions: cubic and quartic interactions of the scalars as well as Yukawa interactions between two spinors and a scalar

$$\begin{aligned}
 \mathcal{L}_{\text{cubic}} &= \frac{1}{6}c_{jkl}\phi^j\phi^k\phi^l, \\
 \mathcal{L}_{\text{quartic}} &= \frac{1}{24}c_{jklm}\phi^j\phi^k\phi^l\phi^m, \\
 \mathcal{L}_{\text{Yukawa}} &= c^k_{lj}\bar{\psi}_j\psi^k\phi^j.
 \end{aligned} \tag{4.92}$$

Gauge symmetry requires the terms to be invariant under gauge transformations. This implies that the constants c_{jkl} , c_{jklm} and c^k_{lj} are invariant tensors of the groups. The existence of such tensors crucially depends on the specific group and representations of the fields.¹⁵ For example, a symmetric cubic tensor c_{jkl} is rather rare whereas a symmetric quartic tensor c_{jklm} exists much more frequently.

4.5 BRST Symmetry

In quantising the Yang–Mills fields, we have traded in gauge invariance for a well-defined propagator. There is still a global gauge symmetry left over, but it is not sufficient to guarantee many of the nice properties of the gauge fixed theory. Becchi, Rouet, Stora and, independently, Tyutin found an extra symmetry of the gauge-fixed Lagrangian, the so-called BRST transformation. Despite its apparent simplicity, it is a very powerful tool for understanding physical states and unitarity.

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, \bar{C} and an auxiliary scalar bosonic field B

$$\mathcal{L}_{\text{BRST}} = \frac{2}{g_{\text{YM}}^2} \text{Tr} \left(\bar{C}(\partial^\mu[D_\mu, C]) + B(\partial^\mu A_\mu) + \frac{1}{2}\xi B^2 \right). \tag{4.93}$$

¹⁵The tensor product of the involved representations must contain a trivial representation so that c does not transform under gauge transformations.

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

$$\mathcal{L}_{\text{BRST}} \simeq \frac{2}{g_{\text{YM}}^2} \text{Tr} \left(\bar{C}(\partial^\mu [D_\mu, C]) - \frac{1}{2}\xi^{-1}(\partial^\mu A_\mu)^2 \right). \quad (4.94)$$

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

$$\begin{aligned} \delta A_\mu &= \delta\epsilon [D_\mu, C], \\ \delta C &= i\delta\epsilon CC = -\frac{i}{2}\delta\epsilon \{C, C\}, \\ \delta \bar{C} &= -\delta\epsilon B, \\ \delta B &= 0. \end{aligned} \quad (4.95)$$

Note that the transformation parameter $\delta\epsilon$ must be a fermionic 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

$$\delta\mathcal{L}_{\text{BRST}} = 0. \quad (4.96)$$

Likewise, the Yang–Mills Lagrangian \mathcal{L}_{YM} is invariant since the BRST transformation on the ordinary fields is a plain gauge transformation with transformation parameter ϵC .

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

$$\delta = \delta\epsilon Q. \quad (4.97)$$

This BRST operator has the important property that it is nilpotent

$$Q^2 = 0, \quad (4.98)$$

as can be verified easily by acting with Q^2 on all fields. In fact we can write the BRST Lagrangian as a BRST variation of some other function K_{BRST}

$$\mathcal{L}_{\text{BRST}} = QK_{\text{BRST}}, \quad K_{\text{BRST}} = \frac{2}{g_{\text{YM}}^2} \text{Tr} \left(-\bar{C}(\partial^\mu A_\mu) - \frac{1}{2}\xi\bar{C}B \right). \quad (4.99)$$

It is then evident that $Q\mathcal{L}_{\text{BRST}} = Q^2K_{\text{BRST}} = 0$. 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

$$\mathcal{L}_{\text{YM}} \neq QK_{\text{YM}}. \quad (4.100)$$

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

BRST Cohomology. The BRST operator Q squares to zero, $Q^2 = 0$, in close analogy to the exterior derivative d . Such operators define a cohomology which often encodes some important information. 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 operators.

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

- *Closed* states are annihilated by Q

$$Q|\Psi\rangle = 0 \quad \leftrightarrow \quad |\Psi\rangle \in \mathbb{H}_{\text{cl}}. \quad (4.101)$$

- *Exact* states are in the image of Q

$$|\Psi\rangle = Q|\Omega\rangle \text{ for some } |\Omega\rangle \in \mathbb{H} \quad \leftrightarrow \quad |\Psi\rangle \in \mathbb{H}_{\text{ex}}. \quad (4.102)$$

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

$$\mathbb{H}_{\text{phys}} = \mathbb{H}_{\text{cl}}/\mathbb{H}_{\text{ex}}. \quad (4.103)$$

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

$$|\Psi\rangle \simeq |\Psi\rangle + Q|\Omega\rangle \quad \text{with} \quad Q|\Psi\rangle = 0. \quad (4.104)$$

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.

The key insight is that the BRST operator is hermitian

$$Q^\dagger = Q. \quad (4.105)$$

This implies that exact states $|\Psi\rangle = Q|\Omega\rangle$ have zero overlap with closed states $|\Phi\rangle$

$$\langle\Phi|\Psi\rangle = \langle\Phi|Q|\Omega\rangle = \langle\Phi|Q^\dagger|\Omega\rangle = 0. \quad (4.106)$$

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

$$\langle\Phi|\Psi'\rangle = \langle\Phi|\Psi\rangle + \langle\Phi|Q|\Omega\rangle = \langle\Phi|\Psi\rangle. \quad (4.107)$$

One-Particle States. Evidently, we should choose a vacuum state $|0\rangle$ in the BRST cohomology. As we will not define the vacuum state explicitly, we cannot discuss its BRST properties in detail.

Next, let us confirm that single-particle states in the BRST cohomology are indeed the physical excitations of a vector field:

- The BRST operator maps \bar{C} to B . Hence the anti-ghost field \bar{C} is not closed.¹⁶
- In momentum space $A_\mu(p)$ is mapped to $p_\mu C(p)$.¹⁷ Thus one of the four modes of A_μ is not closed.
- The image of \bar{C} is the field B . Hence B is exact. Moreover, the equations of motion identify the auxiliary field $B(p)$ with $p^\mu A_\mu(p)$. Therefore one of the four modes of A_μ is exact on shell. It is different from the mode which is not closed since exact states are closed by construction.
- The image of A_μ is the ghost field. Therefore the ghost field C is exact.
- The two remaining polarisation modes of the field A_μ are closed and but not exact.

We observe that the anti-ghost \bar{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{H}_{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-matrix¹⁸

$$[Q, S] = 0. \quad (4.108)$$

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

$$QS|\Psi\rangle = SQ|\Psi\rangle = 0. \quad (4.109)$$

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

$$S|\Psi'\rangle = S|\Psi\rangle + SQ|\Omega\rangle = S|\Psi\rangle + QS|\Omega\rangle. \quad (4.110)$$

We can now consider unitarity of the S-matrix

$$S^\dagger S = 1. \quad (4.111)$$

The optical theorem is the formulation of this statement in terms of the matrix elements of S . We enclose the unitarity statement in physical states

$$\langle\Phi|S^\dagger S|\Psi\rangle = \langle\Phi|\Psi\rangle. \quad (4.112)$$

At this point we should insert the decomposition of unity between the S-matrices. A priori we should insert the unit operator on the full Hilbert space \mathbb{H} . Using our

¹⁶And it evidently cannot be made closed in any linear combination of single-particle states.

¹⁷For the discussion of single-particle states we ignore the terms non-linear in the fields.

¹⁸Also the ghost number operators N_{gh} is a symmetry.

scalar product this is non-trivial since we know that the space \mathbb{H}_{ex} is null. Its dual within \mathbb{H} is the quotient space $\mathbb{H}/\mathbb{H}_{\text{cl}}$.¹⁹ The decomposition thus has be written using some map $*$ which interchanges \mathbb{H}_{ex} with $\mathbb{H}/\mathbb{H}_{\text{cl}}$ and leaves the residual physical space $\mathbb{H}_{\text{cl}}/\mathbb{H}_{\text{ex}} = \mathbb{H}_{\text{phys}}$ invariant

$$\begin{aligned} 1 &= \sum_{\Omega} |\Omega^*\rangle \langle \Omega| \\ &= \sum_{\Omega \text{ physical}} |\Omega\rangle \langle \Omega| + \sum_{\Omega \text{ exact}} |\Omega^*\rangle \langle \Omega| + \sum_{\Omega \text{ exact}} |\Omega\rangle \langle \Omega^*|. \end{aligned} \quad (4.113)$$

This somewhat unusual decomposition of unity is very convenient since all the exact states are cancelled by either $\langle \Phi|S^\dagger$ or $S|\Psi\rangle$.²⁰ The optical theorem can thus be expressed as

$$\langle \Phi|S^\dagger S|\Psi\rangle = \sum_{\Omega \text{ physical}} \langle \Phi|S^\dagger|\Omega\rangle \langle \Omega|S|\Psi\rangle = \langle \Phi|\Psi\rangle. \quad (4.114)$$

In the optical theorem one is therefore free restrict the sum to physical intermediate states. However, one can also perform the sum over all intermediate states including unphysical ones. The result was shown to be the same, therefore the contributions from unphysical modes of the gauge field must cancel exactly against the contributions from ghost fields. The applicability of the full intermediate Hilbert space is crucial since unphysical gauge and ghost fields 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, $\mathcal{L}_{\text{BRST}} = K_{\text{BRST}}$, while \mathcal{L}_{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

$$\mathcal{L}_{\text{BRST}}|\Psi\rangle = (QK_{\text{BRST}})|\Psi\rangle = Q(K_{\text{BRST}}|\Psi\rangle). \quad (4.115)$$

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 parameter ξ of the Lorenz gauge.

¹⁹This is because there is exactly one exact state for every non-closed equivalence class.

²⁰Alternatively one can argue that every contribution of an exact state from one S-matrix is cancelled by the other.

5 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.

5.1 Dimensional Regularisation

In QFT I we have used a plain momentum cutoff to regularise some 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 strongly 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.

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

$$I_n(X) = \int \frac{-id^D \ell}{(2\pi)^D} \frac{1}{(\ell^2 + X - i\epsilon)^n}. \quad (5.1)$$

For convenience, we first Wick rotate the integral

$$I_n(X) = \int \frac{d^D \ell_E}{(2\pi)^D} \frac{1}{(\ell_E^2 + X - i\epsilon)^n}. \quad (5.2)$$

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

$$I_n(X) = \text{Vol}(S^{D-1}) \int_0^\infty \frac{d\ell_E}{(2\pi)^D} \frac{\ell_E^{D-1}}{(\ell_E^2 + X - i\epsilon)^n}. \quad (5.3)$$

The volume of a $(D - 1)$ sphere equals $\text{Vol}(S^{D-1}) = 2\pi^{D/2} / \Gamma(D/2)$. The remaining radial integral after a change of variables $\ell_E^2 = z$ is a Beta function

$$\frac{1}{2} \int_0^\infty \frac{dz}{(2\pi)^D} \frac{z^{D/2-1}}{(z + X - i\epsilon)^n} = \frac{\Gamma(D/2) \Gamma(n - D/2)}{2(2\pi)^D \Gamma(n) (X - i\epsilon)^{n-D/2}}. \quad (5.4)$$

Altogether the integral reads

$$I_n(X) = \frac{\Gamma(n - D/2)}{(4\pi)^{D/2} \Gamma(n)} \frac{1}{(X - i\epsilon)^{n-D/2}}. \quad (5.5)$$

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 4³

$$D = 4 - 2\epsilon. \quad (5.6)$$

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 and Spinors. For models with spin, we will have to deal with vectors in the numerator of the integral, for example⁴

$$\begin{aligned} I_n^\mu(X) &= \int \frac{-id^D \ell}{(2\pi)^D} \frac{\ell^\mu}{(\ell^2 + X - i\epsilon)^n} = 0, \\ I_n^{\mu\nu}(X) &= \int \frac{-id^D \ell}{(2\pi)^D} \frac{\ell^\mu \ell^\nu}{(\ell^2 + X - i\epsilon)^n}. \end{aligned} \quad (5.7)$$

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

¹This does not imply that the formula is morally correct or unique. It is merely a convenient and smooth interpolation between integer dimensions.

²In 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.

³A factor of 2 for ϵ conveniently cancels a factor of 1/2 which is often associated to D .

⁴Note that ϵ as the deviation from $D = 4$ and in the $i\epsilon$ prescription for propagators is clearly distinct. Nevertheless we will use the same symbol for both objects.

trace with $\eta_{\mu\nu}$. This yields

$$\begin{aligned} I_n^{\mu\nu}(X) &= \frac{\eta^{\mu\nu}}{D} \eta_{\rho\sigma} I_n^{\rho\sigma}(X) = \frac{\eta^{\mu\nu}}{D} (I_{n-1}(X) - X I_n(X)) \\ &= \frac{\Gamma(n-1-D/2)}{2(4\pi)^{D/2} \Gamma(n)} \frac{\eta^{\mu\nu}}{(X-i\epsilon)^{n-1-D/2}}. \end{aligned} \quad (5.8)$$

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

$$\delta_\mu^\mu = \eta_{\mu\nu} \eta^{\mu\nu} = D = 4 - 2\epsilon. \quad (5.9)$$

Although, the difference between D and 4 is small, it can be important in the case of divergences which contribute factors of $1/(D-4)$.

Furthermore, spinor models are formulated in terms of Dirac gamma matrices. Their Clifford algebra depends on the number of spacetime dimensions, and for consistency, this should also equal D . The spinor representation grows exponentially with D , so for evaluating a trivial trace we must set

$$\text{Tr } 1 = 2^{D/2}. \quad (5.10)$$

Everything else must be derived strictly from the Clifford algebra relation

$$\{\gamma^\mu, \gamma^\nu\} = -2\eta^{\mu\nu}. \quad (5.11)$$

For instance, some of the bracketing identities generalise as

$$\begin{aligned} \gamma^\mu \gamma_\mu &= \frac{1}{2} \{\gamma^\mu, \gamma_\mu\} = -D, \\ \gamma^\mu \gamma^\nu \gamma_\mu &= \{\gamma^\mu, \gamma^\nu\} \gamma_\mu - \gamma^\nu \gamma^\mu \gamma_\mu = (D-2) \gamma^\nu. \end{aligned} \quad (5.12)$$

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 an antisymmetric tensor $\epsilon_{\mu\nu\rho\sigma}$ with⁶

$$\varepsilon_{0123} = +1. \quad (5.13)$$

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.

⁵At the end of the day, the external momentum vectors will be defined in exactly four dimensions, which may lead to some confusion.

⁶Strictly speaking this requires $D \geq 4$, but we may assume $\epsilon < 0$.

Introduction of a Scale. Let us consider a simple QFT model in dimensional regularisation. For example, the Lagrangian of a scalar with a quartic interaction reads

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{24}\lambda\phi^4. \quad (5.14)$$

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 $S = \int d^D x \mathcal{L}$:

- We can keep the dimension of all involved quantities fixed. Then we have to multiply the Lagrangian by some overall dimensionful factor

$$\mathcal{L} = \mu^{-2\epsilon} \left(-\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{24}\lambda\phi^4 \right). \quad (5.15)$$

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

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{24}\mu^{2\epsilon}\lambda\phi^4. \quad (5.16)$$

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

$$\lambda^L \mu^{2\epsilon L}. \quad (5.17)$$

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.

Expansion. Consider the standard massive scalar bubble integral

$$\begin{aligned} I &= \text{---} \circ \text{---} \\ &= \frac{\mu^{2\epsilon}}{2} \int \frac{-id^D \ell}{(2\pi)^D} \frac{1}{\ell^2 + m^2 - i\epsilon} \frac{1}{(p - \ell)^2 + m^2 - i\epsilon} \\ &= \frac{\mu^{2\epsilon}}{2} \int_0^1 dz \int \frac{-id^D \ell}{(2\pi)^D} \frac{1}{(\ell^2 + z\bar{z}p^2 + m^2 - i\epsilon)^2}. \end{aligned} \quad (5.18)$$

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 regulated expression

$$I = \frac{\Gamma(\epsilon)}{32\pi^2} \int_0^1 dz \left(\frac{z\bar{z}p^2 + m^2 - i\epsilon}{4\pi\mu^2} \right)^{-\epsilon}. \quad (5.19)$$

Next we expand the integrand in ϵ ⁷

$$I = \frac{1}{32\pi^2} \int_0^1 dz \left(\frac{1}{\epsilon} - \gamma_E - \log \frac{z\bar{z}p^2 + m^2 - i\epsilon}{4\pi\mu^2} + \mathcal{O}(\epsilon) \right). \quad (5.20)$$

We have performed the relevant parts of this integral in QFT I, here we obtain

$$I = -\frac{1}{16\pi^2} \sqrt{\frac{p^2 + 4m^2 - i\epsilon}{-p^2}} \arctan \sqrt{\frac{-p^2}{p^2 + 4m^2 - i\epsilon}} + \frac{1}{32\pi^2} \left(\frac{1}{\epsilon} - \gamma_E + \log(4\pi) + 2 - \log \frac{m^2}{\mu^2} \right) + \mathcal{O}(\epsilon). \quad (5.21)$$

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.
- 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.

5.2 Renormalisation

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 as 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

⁷Integration and expansion do not necessarily always commute, but here it is not a problem.

regulator by sending $\epsilon \rightarrow 0$ and recover physically meaningful results. The crucial question is whether this so-called renormalisation procedure can render all observables finite at the same time.

Counterterms. Consider the dimensionally regulated massive scalar bubble integral

$$I = \dots + \frac{1}{32\pi^2} \left(\frac{1}{\epsilon} - \gamma_E + \log(4\pi) + 2 - \log \frac{m^2}{\mu^2} \right) + \mathcal{O}(\epsilon). \quad (5.22)$$

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 serves as a one-loop correction to the cubic interaction and to the kinetic term, specifically to the mass.

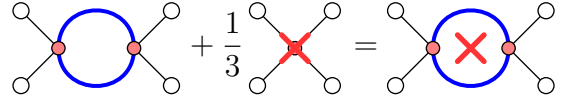
These three correction terms have a simple pole at $\epsilon = 0$. We can remove the divergences by adding counterterms for the quartic interaction, the cubic interaction and the mass. Specifically, a suitable redefinition of the quartic coupling constant λ is⁸

$$\lambda'_{\text{MS}} = \lambda + 3\lambda^2 \frac{1}{32\pi^2\epsilon} + \mathcal{O}(\lambda^3). \quad (5.23)$$

This counterterm removes precisely the $1/\epsilon$ term in the above integral. Effectively, it changes the regularised integral to the subtracted integral $I_{\text{MS}} = I - 1/32\pi^2\epsilon$

$$I_{\text{MS}} = \dots + \frac{1}{32\pi^2} \left(-\gamma_E + \log(4\pi) + 2 - \log \frac{m^2}{\mu^2} \right) + \mathcal{O}(\epsilon). \quad (5.24)$$

In terms of diagrams the divergent loop has a counterterm that makes it perfectly finite



$$\text{Diagram 1} + \frac{1}{3} \text{Diagram 2} = \text{Diagram 3}. \quad (5.25)$$

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. The method is called the minimal subtraction (MS) scheme. At one loop it 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 $-\gamma_E$ and a logarithm $\log(4\pi)$. Importantly they all arise in the universal combination $1/\epsilon - \gamma_E + \log(4\pi)$. It is therefore convenient to modify the scheme to that counterterms are proportional to this combination.

$$\lambda'_{\text{MS}} = \lambda + 3\lambda^2 \frac{1/\epsilon - \gamma_E + \log(4\pi)}{32\pi^2} + \mathcal{O}(\lambda^3). \quad (5.26)$$

⁸ λ'_{MS} is the bare coupling constant appearing in the Lagrangian and λ is the renormalised coupling.

This scheme is called $\overline{\text{MS}}$, and it is the scheme commonly applied in conjunction with dimensional regularisation. The subtraction of the above integral reads

$$I_{\overline{\text{MS}}} = \dots + \frac{1}{32\pi^2} \left(2 - \log \frac{m^2}{\mu^2} \right) + \mathcal{O}(\epsilon). \quad (5.27)$$

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.

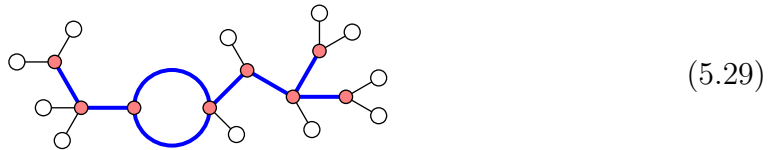
Another important observation is that the above choices of counterterms are related by a change of the renormalisation scale. For instance, using the renormalisation scale $\mu_{\overline{\text{MS}}}$ in the MS scheme with

$$\mu_{\overline{\text{MS}}} = \frac{\exp(\frac{1}{2}\gamma_E)}{\sqrt{4\pi}} \mu \quad (5.28)$$

is equivalent to the $\overline{\text{MS}}$ 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 all other graphs.

The divergent graphs are in fact loops where the external legs connect to tree graphs

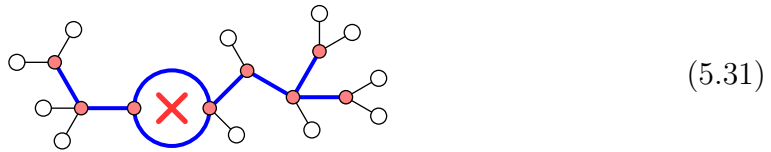


The loop itself is the same as above, hence we already understand its counterterm. One important difference is that the divergence is no longer local since the local divergence of the loop is connected to (non-local) trees. This implies that no local counterterm could possibly compensate the divergence of the above graph.

Nevertheless, this is not a problem since the counterterm can also act as a vertex within a tree graph.

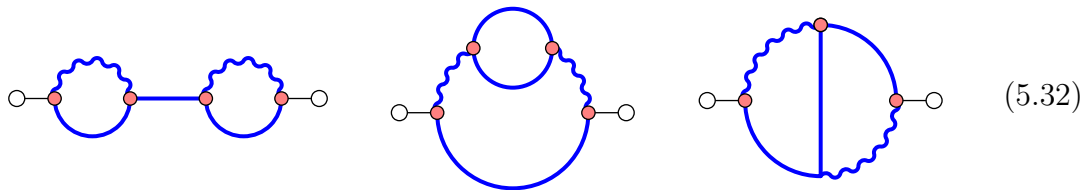


It cancels precisely the divergence of the loop embedded into a tree because the dressing by trees in fact does not make a difference



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 more severe divergences, e.g.

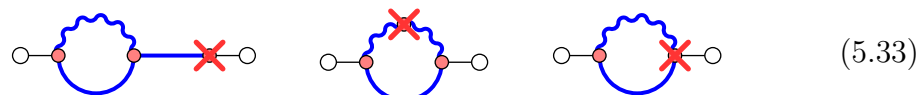


In particular we can distinguish three different situations:

- Disconnected loops: This is the easiest case, since each loop has a lower perturbative order. The corresponding counterterms at lower orders are sufficient to remove the divergence.
- 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 more severe combined divergence.

Quite amazingly, none of these problems is severe at all:

- At higher loops the counterterms from lower orders can also appear inside loops.



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 only a simple divergence left to be removed.

- Despite the fact that all subdivergences of the graph are heavily 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.⁹ Such a divergence can be compensated by a local counterterm. 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.

Gauge Theory. 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

$$\begin{aligned}
\text{Tr } F^{\mu\nu} F_{\mu\nu} &\rightarrow \text{diagram 1} + g \text{diagram 2} + g^2 \text{diagram 3}, \\
-\text{Tr } \bar{C}(\partial^\mu D_\mu^{\text{ad}} C) &\rightarrow \text{diagram 4} + g \text{diagram 5}, \\
\xi^{-1} \text{Tr}(\partial \cdot A)^2 &\rightarrow \xi^{-1} \text{diagram 6}, \\
i\bar{\psi}\gamma^\mu D_\mu^R \psi &\rightarrow \text{diagram 7} + g \text{diagram 8}, \\
-\frac{1}{2}(D^{R,\mu}\phi)^2 &\rightarrow \text{diagram 9} + g \text{diagram 10} + g^2 \text{diagram 11}.
\end{aligned} \tag{5.34}$$

Furthermore, there are some pure matter couplings such as mass terms, cubic and quartic couplings of the scalars as well as Yukawa couplings which are of no relevance 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

⁹This argument works best in euclidean space where null distances do not exist.

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?

Gladly, it turns out that the above vertices renormalise coherently. The reason is symmetry. Although we had to break gauge symmetry to quantise the theory, there are two parts left.

- Manifest global symmetry under the gauge group: 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 analog of the conservation of Noether currents. Therefore the coupling g acts as a universal renormalised gauge coupling constant.
- BRST symmetry: 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 therefore lead to 6 independent renormalisation coefficients:

- One renormalisation coefficient for each line.
- In total one relative renormalisation coefficient for the elements in each line.

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. These are all independent and they have no physical relevance on their own. Nevertheless, they are important coefficients which contribute to each graph via the external legs.
- One coefficient to (potentially) renormalise the constant ξ .
- One coefficient to renormalise the gauge coupling g .

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.

$$(5.35)$$

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 ghost, and 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 ghosts or spinors take the form

$$(5.36)$$

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

$$\text{Tr } T_a^R T_b^R = B^R k_{ab}. \quad (5.37)$$

For three external gluons the structure is¹⁰

$$\text{Tr}[T_a^R, T_b^R] T_c^R = i f_{ab}^d \text{Tr } T_d^R T_c^R = i B^R f_{abc}. \quad (5.38)$$

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

$$f_{ac}^d f_{bd}^c = -B^{\text{ad}} k_{ab}. \quad (5.39)$$

For three gluons the result reduces to an analogous combination¹¹

$$\begin{aligned} f_{ad}^e f_{be}^f f_{cf}^d &= \frac{1}{2} f_{ad}^e f_{be}^f f_{cf}^d - \frac{1}{2} f_{db}^e f_{ae}^f f_{cf}^d - \frac{1}{2} f_{ba}^e f_{de}^f f_{cf}^d \\ &= -\frac{1}{2} f_{ab}^e f_{ed}^f f_{cf}^d = \frac{1}{2} B^{\text{ad}} f_{abc}. \end{aligned} \quad (5.40)$$

¹⁰The 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.

¹¹First divide the term into two equal ones, and apply a Jacobi identity to one of them. After some shuffling of indices, two terms cancel and the third produces the final result.

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 $R = \text{ad}$). The correction to the kinetic term directly yields the quadratic Casimir eigenvalue in representation R

$$k^{ab}T_a^R T_b^R = C_2^R. \quad (5.41)$$

We can now add an external gluon in two inequivalent places. When it is attached to the matter line we obtain

$$\begin{aligned} k^{bc}T_b^R T_a^R T_c^R &= k^{bc}[T_b^R, T_a^R]T_c^R + k^{bc}T_a^R T_b^R T_c^R \\ &= -\frac{1}{2}f_a^{dc}f_{dc}^e T_e^R + C_2^R T_a^R \\ &= -\frac{1}{2}B^{\text{ad}}T_a^R + C_2^R T_a^R \\ &= (C_2^R - \frac{1}{2}B^{\text{ad}})T_a^R. \end{aligned} \quad (5.42)$$

When it is attached to the gluon line instead we obtain the structure

$$f_a^{bc}T_b^R T_c^R = \frac{1}{2}f_a^{bc}[T_b^R, T_c^R] = \frac{i}{2}f_a^{bc}f_{bc}^d T_d^R = \frac{i}{2}B^{\text{ad}}T_a^R. \quad (5.43)$$

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} ?¹² 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 function¹³

$$\beta(g) = \frac{g^3}{16\pi^2} \left(\frac{11}{3} B^{\text{ad}} - \frac{4}{3} \sum_{\text{Dirac spinors}} B^R - \frac{1}{6} \sum_{\text{real scalars}} B^R \right). \quad (5.44)$$

Here the first sum is over the representations R of Dirac spinor fields¹⁴ and the second sum over the representations R of real scalar fields. Note that the different prefactors of the B 's are characteristic numbers for type of particles.¹⁵ In the special case of N_f flavours of Dirac fermions in a $\text{SU}(N_c)$ gauge theory we recover the famous combination $\frac{11}{3}N_c - \frac{4}{3}N_f$.

¹²Note that C_2^R and B^R differ by constants depending on the gauge algebra and the representation size.

¹³We 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.

¹⁴Weyl spinors have half as many degrees of freedom and therefore are counted with half the prefactor.

¹⁵For instance, it makes physical sense that a particle with spin has a stronger influence on $\beta(g)$ than a scalar particle.

5.3 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 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 .¹⁶ 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

$$G_n(p_k, \lambda, \mu). \quad (5.45)$$

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 renormalisation of the n external fields¹⁷

$$\frac{\partial G_n}{\partial \mu} + \frac{\partial \lambda}{\partial \mu} \frac{\partial G_n}{\partial \lambda} - \frac{n}{\phi} \frac{\partial \phi}{\partial \mu} G_n = 0. \quad (5.46)$$

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

$$\left(\mu \frac{\partial}{\partial \mu} + \beta(\lambda) \frac{\partial}{\partial \lambda} + n\gamma(\lambda) \right) G_n(p_k, \lambda, \mu) = 0. \quad (5.47)$$

where we have introduced the conventional functions $\beta(\lambda)$ and $\gamma(\lambda)$

$$\beta(\lambda) = \mu \frac{\delta \lambda}{\delta \mu}, \quad \gamma(\lambda) = -\frac{\mu}{\phi} \frac{\delta \phi}{\delta \mu}. \quad (5.48)$$

These dimensionless functions are universally defined for the QFT model at hand independently of the particular observable one may be interested in. Therefore

¹⁶The 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.

¹⁷This follows, for instance, from performing the rescaling in the path integral formulation of the generating functional.

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.

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 since there are no other scales involved in the calculation

$$\left(\sum_k p_k \frac{\partial}{\partial p_k} - nd + \mu \frac{\partial}{\partial \mu} \right) G_n = 0. \quad (5.49)$$

Here d represents the canonical mass dimension of the field $\phi(p)$. Combine this with the Callan–Symanzik equation yields the equation

$$\left(\sum_k p_k \frac{\partial}{\partial p_k} - \beta(\lambda) \frac{\partial}{\partial \lambda} - n(d + \gamma(\lambda)) \right) G_n = 0 \quad (5.50)$$

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. This 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 is governed by a single overall momentum scale p . The Callan–Symanzik equation thus reads

$$\left(p \frac{\partial}{\partial p} - \beta(\lambda) \frac{\partial}{\partial \lambda} - nd - n\gamma(\lambda) \right) G(p, \lambda, \mu) = 0. \quad (5.51)$$

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

$$\left(p \frac{\partial}{\partial p} - \beta(\lambda) \frac{\partial}{\partial \lambda} \right) \tilde{\lambda}(p/\mu, \lambda) = 0, \quad \tilde{\lambda}(1, \lambda) = \lambda. \quad (5.52)$$

Furthermore we make an ansatz depending on the running coupling and with appropriate overall scaling

$$G(p, \lambda, \mu) = p^{nd} \tilde{G}(p/\mu, \tilde{\lambda}(p/\mu, \lambda)). \quad (5.53)$$

It simplifies the Callan–Symanzik equation as follows

$$\left(z \frac{\partial}{\partial z} - n\gamma(\lambda(z, \tilde{\lambda})) \right) \tilde{G}(z, \tilde{\lambda}) = 0, \quad (5.54)$$

where $\lambda(z, \tilde{\lambda})$ is the inverse function of $\tilde{\lambda}(z, \lambda)$. This equation is now easy enough to solve as

$$\tilde{G}(z, \tilde{\lambda}) = \exp \left(\int_1^z \frac{dz'}{z'} n\gamma(\lambda(z', \tilde{\lambda})) \right) \tilde{F}(\tilde{\lambda}) \quad (5.55)$$

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 Z^\pm 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.
- 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. Suppose we compute some correlation function in the dimensional regularisation scheme. The bare result \bar{G} in terms of the bare coupling constant $\bar{\lambda}$ is expressed as $\bar{G}(p_k, \bar{\lambda}, \mu, \epsilon)$.

Now in dimensional regularisation the renormalisation scale μ appears in a very predictable fashion: Up to some potential rescalings it only appears in the combination $\lambda\mu^{2\epsilon}$. 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 form¹⁸

$$\bar{G}(p, \bar{\lambda}, \mu, \epsilon) = p^{nd} \bar{F}(\bar{\lambda}(\mu/p)^{2\epsilon}, \epsilon). \quad (5.56)$$

This function obeys the Callan–Symanzik equation with the parameters

$$\bar{\beta}(\bar{\lambda}) = -2\epsilon \bar{\lambda}, \quad \bar{\gamma}(\bar{\lambda}) = 0. \quad (5.57)$$

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

$$\bar{G}_n(p_k, \bar{\lambda}, \mu, \epsilon) = N(\bar{\lambda})^n G_n(p_k, \lambda(\bar{\lambda}), \mu, \epsilon) \quad (5.58)$$

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 original variables, the Callan–Symanzik equation also holds for the renormalised variables with the renormalised functions

$$\begin{aligned} \beta(\lambda(\bar{\lambda})) &= \bar{\beta}(\bar{\lambda}) \frac{\partial \lambda}{\partial \bar{\lambda}}(\bar{\lambda}), \\ \gamma(\lambda(\bar{\lambda})) &= \bar{\gamma}(\bar{\lambda}) + \frac{\bar{\beta}(\bar{\lambda})}{N(\bar{\lambda})} \frac{\partial N}{\partial \bar{\lambda}}(\bar{\lambda}). \end{aligned} \quad (5.59)$$

Renormalisation therefore preserves the Callan–Symanzik equation with deformed parameters.

Suppose in dimensional regularisation the renormalisation functions take the following generic form with poles at $\epsilon = 0$

$$\begin{aligned} \lambda &= \bar{\lambda} + \bar{\lambda}^2(b_1/\epsilon + b_2 + \dots) + \dots, \\ \bar{\lambda} &= \lambda - \lambda^2(b_1/\epsilon + b_2 + \dots) + \dots, \\ N(\bar{\lambda}) &= 1 + \bar{\lambda}(c_1/\epsilon + c_2 + \dots) + \dots \end{aligned} \quad (5.60)$$

Substituting these into the above expressions we obtain perfectly regular expressions for β and γ at $\epsilon \rightarrow 0$

$$\begin{aligned} \beta(\lambda) &= -2\epsilon\lambda - 2\lambda^2(b_1 + b_2\epsilon + \dots) + \dots, \\ \gamma(\lambda) &= -2\lambda(c_1 + c_2\epsilon + \dots) + \dots \end{aligned} \quad (5.61)$$

Taking the limit $\epsilon \rightarrow 0$, we find the renormalised expressions

$$\beta(\lambda) \rightarrow -2b_1\lambda^2 + \dots, \quad \gamma(\lambda) \rightarrow -2c_1\lambda + \dots \quad (5.62)$$

¹⁸For 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.

Curiously, they are non-vanishing, so the renormalisation procedure has actually achieved something non-trivial. This was only possible because of the divergences in the bare correlators. So the divergences are actually not just an inconvenient feature of QFT, they actually are related to non-trivial behaviour of the quantum model such as running coupling constants.

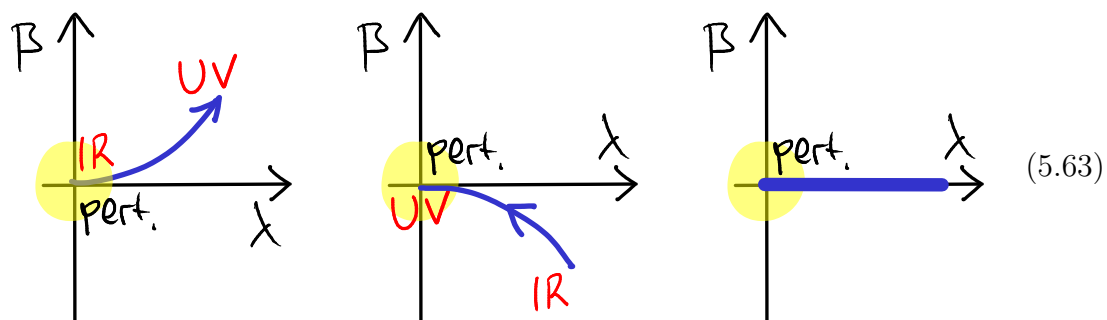
We should mention that the above functions β and γ are not uniquely defined: For example, we are free to perform a further finite perturbative redefinition of the coupling $\lambda \rightarrow \lambda' = \lambda + \mathcal{O}(\lambda^2)$. This leaves the classical model unchanged, but it modifies the 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 w.r.t. the renormalisation scale. 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.



- $\beta(\lambda) > 0$: 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 a *IR stable fixed point*. An example for such a model is QED. The screening of electric charges makes the coupling weak for long distances but very strong 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. It 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.

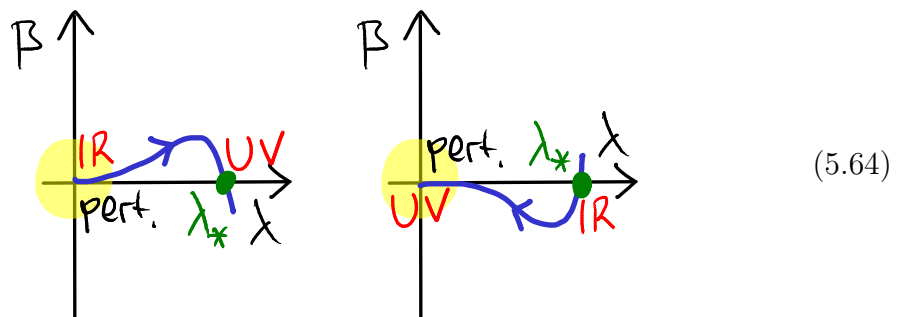
- A very rare third option is $\beta(\lambda) = 0$ for a range of λ 's. In this case a certain class of divergent contributions is absent, and the theory is called *finite*.¹⁹ This implies that the coupling constant does not depend on the scale, even though there are non-trivial interactions. We have 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, but we can suppose we know $\beta(\lambda)$ exactly. We assume that $\beta(\lambda)$ returns to zero at some point λ_* .²⁰ 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 $\lambda \rightarrow \lambda'(\lambda)$.

Again we can distinguish two cases.



- $\beta(\lambda) > 0$ 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.

¹⁹Note that there can still be infinite renormalisation of the fields.

²⁰Alternatively, $\beta(\lambda)$ 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 .

- $\beta(\lambda) < 0$ 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.

6 Quantum Symmetries

Symmetries are crucial for quantum field theories since they provide stability in the renormalisation process. In particular, symmetries reduce the gauge field to two 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.

Equations of Motion. Quantum fields obey the equations of motion. In the path integral they originate from an integrand which is a total derivative

$$\begin{aligned} 0 &= \int D\Psi \frac{\delta}{\delta\Psi(x)} \exp(iS[\Psi] + iS_{\text{src}}[\Psi, J]) \\ &= i \int D\Psi \left(\frac{\delta S}{\delta\Psi(x)} + J(x) \right) \exp(iS[\Psi] + iS_{\text{src}}[\Psi, J]). \end{aligned} \quad (6.1)$$

The first term in the brackets generates the Euler–Lagrange equations, the second one 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

$$\frac{\delta S}{\delta\Psi(x)} \left[-i \frac{\delta}{\delta J} \right] Z[J] = -J(x)Z[J]. \quad (6.2)$$

For example this implies the expectation value of the classical equation of motion to hold

$$\langle \delta S[\Psi] / \delta\Psi(x) \rangle = 0. \quad (6.3)$$

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

$$\begin{aligned} \langle O[\Psi] \delta S / \delta\Psi(x) \rangle &= Z[J]^{-1} O \left[-i \frac{\delta}{\delta J} \right] \frac{\delta S}{\delta\Psi(x)} \left[-i \frac{\delta}{\delta J} \right] Z[J] \Big|_{J=0} \\ &= - Z[J]^{-1} O \left[-i \frac{\delta}{\delta J} \right] J(x) Z[J] \Big|_{J=0} \\ &= i Z[J]^{-1} \frac{\delta O}{\delta\Psi(x)} \left[-i \frac{\delta}{\delta J} \right] Z[J] \Big|_{J=0} \\ &= i \langle \delta O / \delta\Psi(x) \rangle. \end{aligned} \quad (6.4)$$

Here one of the fields $\Psi \simeq -i\delta/\delta J$ 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

$$\langle \delta S[\Psi]/\delta\Psi(x) + J(x) \rangle_J = 0. \quad (6.5)$$

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 $-i\delta/\delta J(x)$, 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 end of a Feynman propagator, and converts it to a delta function. The other end of the propagator may be attached to a vertex or to an operator in the correlator. The contribution from the first case is cancelled precisely the interacting part of the action. The contribution from the second case corresponds to varying the operator by the field.

$$\left(-\underset{x}{\bullet} \text{---} \circ - \frac{1}{2} \underset{x}{\bullet} \text{---} \triangle \right) Z$$

$$= -\underset{x}{\bullet} \text{---} Z - \frac{1}{2} \underset{x}{\bullet} \text{---} Z$$

¹Operators 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.

$$\begin{aligned}
&= - \begin{array}{c} \times \\ x \end{array} \begin{array}{c} \color{red}{\circ} \\ \color{blue}{\text{---}} \\ \bullet \end{array} \begin{array}{c} \textcircled{Z} \\ \color{blue}{\text{---}} \\ \bullet \end{array} - \frac{1}{2} \begin{array}{c} \times \\ x \end{array} \begin{array}{c} \color{red}{\circ} \\ \color{blue}{\text{---}} \\ \color{red}{\rightarrow} \end{array} \begin{array}{c} \textcircled{Z} \\ \color{blue}{\text{---}} \\ \bullet \end{array} - \frac{1}{2} \begin{array}{c} \times \\ x \end{array} \begin{array}{c} \color{red}{\circ} \\ \color{blue}{\text{---}} \\ \color{red}{\rightarrow} \end{array} \begin{array}{c} \color{blue}{\text{---}} \\ \bullet \end{array} \begin{array}{c} \textcircled{Z} \\ \color{blue}{\text{---}} \\ \bullet \end{array} \\
&= - \begin{array}{c} \times \\ x \end{array} \begin{array}{c} \color{red}{\circ} \\ \color{blue}{\text{---}} \\ \bullet \end{array} \begin{array}{c} \textcircled{Z} \\ \color{blue}{\text{---}} \\ \bullet \end{array} = - \begin{array}{c} \times \\ x \end{array} \begin{array}{c} \bullet \\ \color{blue}{\text{---}} \\ \bullet \end{array} \begin{array}{c} \textcircled{Z} \\ \color{blue}{\text{---}} \\ \bullet \end{array}
\end{aligned} \tag{6.6}$$

Other Functionals. The Schwinger–Dyson equation can also be formulated for the generating functional of connected graphs $W[J]$ and the effective action $G[\Psi]$.

We start with the Schwinger–Dyson equation with the constant \hbar made explicit

$$\frac{\delta S}{\delta \Psi(x)} \left[-i\hbar \frac{\delta}{\delta J} \right] Z[J] = -J(x)Z[J]. \tag{6.7}$$

For the connected graphs one finds

$$\frac{\delta S}{\delta \Psi(x)} \left[-i\hbar \frac{\delta}{\delta J} + \frac{\delta W}{\delta J} \right] = -J(x). \tag{6.8}$$

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 $\delta W/\delta J$.

The Schwinger–Dyson equation for the effective action takes a rather implicit form

$$\frac{\delta S}{\delta \Psi(x)} \left[\Psi + i\hbar \left(\frac{\delta^2 G}{\delta \Psi^2} \right)^{-1} \frac{\delta}{\delta \Psi} \right] = \frac{\delta G}{\delta \Psi(x)}. \tag{6.9}$$

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 charge for every continuous global symmetry. A similar statement should also hold for quantum field theories.

Noether’s Theorem. Suppose the action $S[\Psi]$ is invariant under a global symmetry

$$\frac{\delta S}{\delta \alpha} = \int d^D x \frac{\delta \Psi(x)}{\delta \alpha} \frac{\delta S}{\delta \Psi(x)} = 0. \tag{6.10}$$

Vanishing of the integral on the r.h.s. implies the existence of the Noether current $N^\mu(x)$ which is (implicitly) defined by

$$\frac{\delta \Psi(x)}{\delta \alpha} \frac{\delta S}{\delta \Psi(x)} = \partial_\mu N^\mu(x). \tag{6.11}$$

This current is conserved on solutions of the Euler–Lagrange equations because the l.h.s. of this equation vanishes.

²This is due to $\exp(-b(x))a(\partial/\partial x)\exp(b(x)) = a(\partial/\partial x + b'(x))$.

Generating Functional. We have already derived the quantum analog of the Euler–Lagrange equation, namely the Schwinger–Dyson equation³

$$\frac{\delta S}{\delta \Psi(x)} \left[-i \frac{\delta}{\delta J} \right] Z[J] = -J(x)Z[J]. \quad (6.12)$$

Multiplying this identity by the field variation we find

$$\left(\frac{\delta \Psi(x)}{\delta \alpha} \frac{\delta S}{\delta \Psi(x)} \right) \left[-i \frac{\delta}{\delta J} \right] Z[J] = -\frac{\delta \Psi(x)}{\delta \alpha} \left[-i \frac{\delta}{\delta J} \right] J(x)Z[J]. \quad (6.13)$$

We substitute the Noether current to obtain the Slavnov–Taylor equation

$$\partial_\mu N^\mu \left[-i \frac{\delta}{\delta J} \right] (x)Z[J] = -\frac{\delta \Psi(x)}{\delta \alpha} \left[-i \frac{\delta}{\delta J} \right] (J(x)Z[J]). \quad (6.14)$$

In the integral form the Noether current drops out

$$\int d^D x \frac{\delta \Psi(x)}{\delta \alpha} \left[-i \frac{\delta}{\delta J} \right] (J(x)Z[J]) = 0. \quad (6.15)$$

The content of the above statements is clearer when expanded in powers of J . Their expansion coefficients are identities between quantum correlation functions. For instance, the leading order of the differential form is the classical conservation law of the expectation value

$$\langle \partial \cdot N(x) \rangle = 0. \quad (6.16)$$

The term at first order yields in differential form reads

$$\langle \Psi(y) \partial_\mu N^\mu(x) \rangle = i\delta^D(x-y) \langle \delta \Psi(x) / \delta \alpha \rangle. \quad (6.17)$$

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 .⁴

After integration over x we find the global field variation

$$\langle \delta \Psi(y) / \delta \alpha \rangle = 0. \quad (6.18)$$

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

$$\langle (\delta \Psi(y) / \delta \alpha) \Psi(z) \rangle + \langle \Psi(y) (\delta \Psi(z) / \delta \alpha) \rangle = 0 \quad (6.19)$$

telling that the expectation value of the change of the product $\Psi(y)\Psi(z)$ vanishes.

³Recall that a functional $F[-i\delta/\delta J]$ acting on $Z[J]$ should be interpreted as the functional $F[\Psi]$ within a quantum correlator $\langle \dots \rangle_J$.

⁴This is somewhat analogous to the Noether charge acting as a symmetry generator in the canonical framework.

Correlators with Sources. 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

$$\langle \partial_\mu N^\mu(x) + J(x) (\delta\Psi(x)/\delta\alpha) \rangle_J = 0. \quad (6.20)$$

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 fields can be introduced into the correlator by means of the sources.

Example. Consider the interacting complex scalar field with Lagrangian

$$\mathcal{L} = -\partial^\mu \phi^* \partial_\mu \phi - m^2 \phi^* \phi - \frac{1}{4} \lambda (\phi^* \phi)^2. \quad (6.21)$$

Rotation of the fields by a phase $\delta\phi = i\delta\alpha\phi$, $\delta\phi^* = -i\delta\alpha\phi^*$ is a continuous global symmetry. We define the associated Noether current via

$$i\phi \partial^2 \phi^* - i\phi^* \partial^2 \phi = \partial_\mu N^\mu \quad \Rightarrow \quad N^\mu = i\phi \partial^\mu \phi^* - i\phi^* \partial^\mu \phi. \quad (6.22)$$

We can now write the above Slavnov–Taylor equation in terms of a correlator with sources

$$\langle \partial_\mu N^\mu(x) \rangle_{j,j^*} = -i \langle \phi(x) j(x) - \phi^*(x) j^*(x) \rangle_{j,j^*}. \quad (6.23)$$

Disabling the sources leads to the classical conservation law for the expectation value

$$\langle \partial_\mu N^\mu(x) \rangle = 0. \quad (6.24)$$

However, the conservation law is deformed when additional fields are inserted into the correlator via $\phi = -i\delta/\delta j$

$$\langle \partial_\mu N^\mu(x) \phi(y) \rangle = -\delta^D(x-y) \langle \phi(y) \rangle. \quad (6.25)$$

Integration over x yields the statement that the expectation value of a charged quantity vanishes

$$\langle \phi(y) \rangle = 0. \quad (6.26)$$

For insertion of two fields we find

$$\langle \partial_\mu N^\mu(x) \phi(y) \phi^*(z) \rangle = -(\delta^D(x-y) - \delta^D(x-z)) \langle \phi(y) \phi^*(z) \rangle. \quad (6.27)$$

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

$$\begin{aligned}\mathcal{L}_{\text{QED}} &= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \\ \mathcal{L}_{\text{gf}} &= -\frac{1}{2}\xi^{-1}(\partial\cdot A)^2, \\ \mathcal{L}_{\text{src}} &= \bar{\psi}\rho + \bar{\rho}\psi - J^\mu A_\mu.\end{aligned}\tag{6.28}$$

Here the covariant derivatives are defined as

$$D_\mu\psi = (\partial_\mu - iqA_\mu)\psi, \quad D_\mu\bar{\psi} = (\partial_\mu + iqA_\mu)\bar{\psi}.\tag{6.29}$$

We note that the global part of the gauge transformation acts as

$$\delta\psi = iq\delta\alpha\psi, \quad \delta\bar{\psi} = -iq\delta\alpha\bar{\psi}.\tag{6.30}$$

We construct the Noether current N^μ from the action via its divergence $\partial\cdot N$

$$\partial_\mu N^\mu = q\bar{\psi}\gamma^\mu(D_\mu\psi) + q(D_\mu\bar{\psi})\gamma^\mu\psi.\tag{6.31}$$

This equation has the electromagnetic current of the spinor field as its solution

$$N^\mu = q\bar{\psi}\gamma^\mu\psi.\tag{6.32}$$

The associated Slavnov–Taylor identity reads

$$\langle\partial_\mu N^\mu - iq\bar{\psi}\rho + iq\bar{\rho}\psi\rangle_{\rho,\bar{\rho},J} = 0.\tag{6.33}$$

This equation is one version of the Ward–Takahashi identity which involves the electromagnetic current.

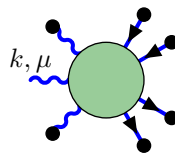
The identity is usually expressed in terms of a gauge field which couples to the electromagnetic current. The current N^μ also appears in the equations of motion for the electromagnetic field, let us therefore consider its Schwinger–Dyson equation

$$\langle q\bar{\psi}\gamma^\mu\psi - \partial_\nu F^{\mu\nu} + \xi^{-1}\partial^\mu(\partial\cdot A) - J^\mu\rangle_{\rho,\bar{\rho},J} = 0.\tag{6.34}$$

Assembling the two identities yields another version of the Ward–Takahashi identity

$$\langle -\xi^{-1}\partial^2(\partial\cdot A) + \partial\cdot J - iq\bar{\psi}\rho + iq\bar{\rho}\psi\rangle_{\rho,\bar{\rho},J} = 0.\tag{6.35}$$

Momentum Space. The identity is conventionally expressed in momentum space as follows: The first term picks a Feynman graph with an external photon line whose momentum is k and whose polarisation index is μ .



$$\tag{6.36}$$

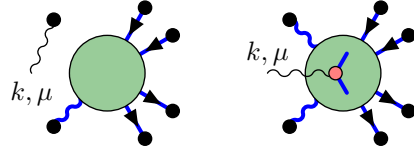
It multiplies the graph by $\xi^{-1}k^2k_\mu$. The external photon line connects to the remainder of the graph by the propagator

$$G_F^{\mu\nu}(k) = \frac{\eta^{\mu\nu} - (1 - \xi)k^\mu k^\nu / k^2}{k^2 - i\epsilon}. \quad (6.37)$$

Multiplied by the prefactor we obtain

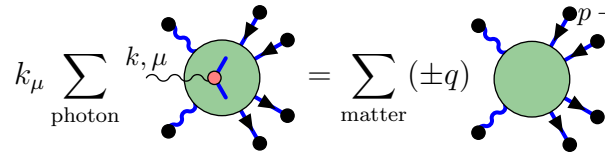
$$\xi^{-1}k^2k_\mu G_F^{\mu\nu}(k) = k^\nu. \quad (6.38)$$

This means we actually consider a graph with one amputated photon line whose polarisation vector is contracted with k . Now there are two options:



$$(6.39)$$

- 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.



$$k_\mu \sum_{\text{photon}} k_\mu \sum_{\text{matter}} (\pm q) \quad (6.40)$$

Propagator and Vertex. Let us consider two examples, a photon propagator and a spinor-photon vertex:

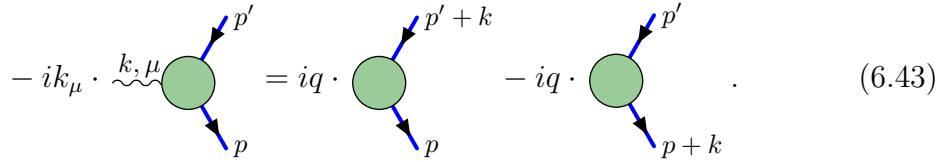
We have already seen that the plain photon propagator $G_F^{\mu\nu}(k)$ satisfies the identity, and it absorbs the term $\partial \cdot J$ completely. How about the loop correction $M^{\mu\nu}(k)$? Consider the graph $\xi^{-1}k^2k_\mu M^{\mu\nu}(k)$ with one external photon line amputated and contracted 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

$$M^{\mu\nu}(k) \sim \eta^{\mu\nu} - k^\mu k^\nu / k^2. \quad (6.41)$$

The only violation of transversality originates from the gauge fixing term in the plain propagator. This identity 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 implies⁵

$$-ik \cdot M(k, p, p') = iqM_0(p, p' + k) - iqM_0(p + k, p'). \quad (6.42)$$



$$-ik_\mu \cdot \text{---} \overset{k, \mu}{\sim} \text{---} \text{---} \text{---} \overset{p'}{\nearrow} \text{---} \overset{p}{\searrow} \text{---} = iq \cdot \text{---} \text{---} \overset{p'+k}{\nearrow} \text{---} \overset{p}{\searrow} \text{---} - iq \cdot \text{---} \text{---} \overset{p'}{\nearrow} \text{---} \overset{p+k}{\searrow} \text{---} . \quad (6.43)$$

Further Comments. 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 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 special 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 .

⁵This 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.

- 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. We have learned in QFT I that a massless Dirac spinor field has a chiral symmetry

$$\psi \rightarrow \exp(-i\gamma^5\beta)\psi \quad (6.44)$$

in addition to its usual electric charge symmetry $\psi \rightarrow \exp(i\alpha)\psi$. The associated Noether current is the axial current

$$N_A^\mu = \bar{\psi}\gamma^5\gamma^\mu\psi. \quad (6.45)$$

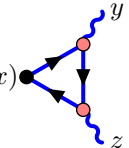
It is classically conserved, even in the presence of the electromagnetic field

$$\partial \cdot N_A = \bar{\psi}\gamma^5\gamma^\mu(D_\mu\psi) + (D_\mu\bar{\psi})\gamma^5\gamma^\mu\psi = 0. \quad (6.46)$$

The naive Slavnov–Taylor identity for the axial current thus reads

$$\langle \partial_\mu N_A^\mu - i\bar{\rho}\gamma^5\psi - i\bar{\psi}\gamma^5\rho \rangle_{\rho,\bar{\rho},J} = 0. \quad (6.47)$$

Consider now the correlator of the axial current with two photons

$$\langle \partial \cdot N_A(x) F_{\mu\nu}(y) F_{\rho\sigma}(z) \rangle \rightarrow \partial \cdot N_A(x) \cdot \text{diagram} \quad (6.48)$$


By the above identity it should be zero. The resulting integral in momentum space has the form

$$\int \frac{d^4\ell}{(2\pi)^4} \text{Tr} \left[\gamma^5(k \cdot \gamma) \frac{\gamma \cdot (\ell - p)}{(\ell - p)^2} \gamma^\mu \frac{\gamma \cdot \ell}{\ell^2} \gamma^\rho \frac{\gamma \cdot (\ell + p')}{(\ell + p')^2} \right]. \quad (6.49)$$

Now write $k = (\ell + p') - (\ell - p)$, 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 the form

$$\int \frac{d^4\ell}{(2\pi)^4} \text{Tr} \left[\gamma^5\gamma^\mu \frac{\gamma \cdot \ell}{\ell^2} \gamma^\rho \left(\frac{\gamma \cdot (\ell + p')}{(\ell + p')^2} - \frac{\gamma \cdot (\ell - p)}{(\ell - p)^2} \right) \right]. \quad (6.50)$$

There are some formal reasons why this integral should vanish

- The trace of gamma matrices produces the antisymmetric tensor ε in four dimensions, $\text{Tr} \gamma^5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma \sim \varepsilon^{\mu\nu\rho\sigma}$.

- The two resulting integrals 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 p or p' . There is nothing to act as the fourth index of ε .
- Furthermore, there is another copy of the same integral with the photon lines exchanged. This appears to cancel against the above integral due to statistics.

This conclusion is however too naive: One must pay attention to the fact that the above integral is 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. Nevertheless, the result for the overall integral is finite and it yields a well-defined, non-zero quantity proportional to $\varepsilon^{\mu\nu\rho\sigma}p_\rho p'_\sigma$.

The pitfall in the above considerations depends on the regularisation scheme:

- In a cutoff regularisation scheme, one has to pay attention to the boundary of the integration region. In particular, the evaluation of integrals involves a shift. Since the original integral is linearly divergent, such a shift may well result in a finite difference.
- In dimensional regularisation the mistake was 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 antisymmetric product of 4 gamma matrices which imitates the behaviour of γ^5 . However, there are discrepancies at $\mathcal{O}(\epsilon)$. For instance, it anticommutes with 4 gamma matrices but it commutes with the remaining (-2ϵ) ones.

$$\gamma^5\gamma^\mu \neq -\gamma^\mu\gamma^5. \quad (6.51)$$

The resulting factor of ϵ is then cancelled by a factor of $1/\epsilon$ from a divergent loop integral.

Curiously, a careful evaluation in either of the two schemes yields the same result. It can be summarised in the anomalous Slavnov–Taylor identity known as the Adler–Bell–Jackiw identity

$$\langle \partial_\mu A^\mu - i\bar{\rho}\gamma^5\psi - i\bar{\psi}\gamma^5\rho - q^2(4\pi)^{-2}\varepsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \rangle_{\rho,\bar{\rho},J} = 0. \quad (6.52)$$

Note that the additional term has the expected structure $\varepsilon^{\mu\nu\rho\sigma}p_\rho p'_\sigma$ which is perfectly local. Adler and Bardeen showed that this identity is exact, there are no further corrections from higher loop orders.

Comments. Anomalies are an exciting topic of quantum field theory related to topological issues of spacetime and gauge groups. It is treated in more detail in the course on advanced field theory. Let us nevertheless make a few more remarks:

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. Gladly, 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. An anomalous Slavnov–Taylor identity is often one-loop exact.

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 made 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.

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 to 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

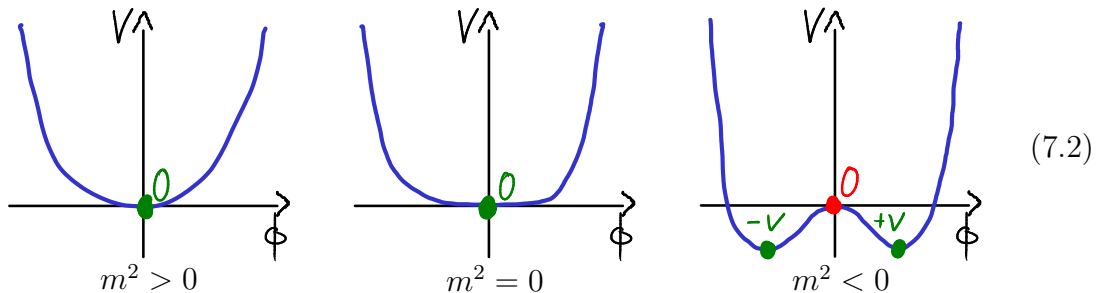
We start by introducing the concept of spontaneous symmetry breaking in the simple context of global symmetries. Spontaneous symmetry breaking merely obscures symmetries; the QFT models still possess such symmetries, but they are not realised in an obvious fashion. The existence of the symmetry has an 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

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{24}\lambda\phi^4. \quad (7.1)$$

We have dropped a potential cubic term in order to have a minimal amount of symmetry, namely the global discrete \mathbb{Z}_2 symmetry $\phi \rightarrow -\phi$.

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 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



Let us therefore consider the case of negative m^2 expressed as a positive $\mu^2 = -m^2$

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 + \frac{1}{2}\mu^2\phi^2 - \frac{1}{24}\lambda\phi^4. \quad (7.3)$$

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.¹

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

$$\phi = \pm v, \quad v = \sqrt{6\mu^2/\lambda}. \quad (7.4)$$

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

$$\phi(x) = \pm v + \eta(x), \quad (7.5)$$

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 $\phi(x)$

$$\langle \phi(x) \rangle = \pm v. \quad (7.6)$$

We obtain a new Lagrangian in the new field η

$$\mathcal{L} = -\frac{1}{2}(\partial\eta)^2 + 3\mu^4/2\lambda - \mu^2\eta^2 \mp \sqrt{\lambda\mu^2/6}\eta^3 - \frac{1}{24}\lambda\eta^4. \quad (7.7)$$

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 $\eta \rightarrow -\eta$, but altogether the original symmetry is still present in the somewhat obscure form

$$\eta \rightarrow \mp 2v - \eta. \quad (7.8)$$

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

¹In fact, a local maximum is troublesome only due to perturbation theory. When we turn off the interactions, the potential all of a sudden becomes unbounded from below. In other words, perturbation theory tries to expand around a tachyonic model. If somehow the interacting theory could be treated non-perturbatively, the expansion point $\phi = 0$ could be suitable as well.

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.² We can thus set $\phi(x) = +v + \eta(x)$ without prejudice.³

Breaking of a Continuous Symmetry. Next, consider a generalisation of the above model to N scalar fields Φ^a with a global $O(N)$ symmetry

$$\mathcal{L} = -\frac{1}{2}(\partial\Phi)^2 + \frac{1}{2}\mu^2\Phi^2 - \frac{1}{8}\lambda(\Phi^2)^2. \quad (7.9)$$

The potential has a local maximum at $\Phi = 0$, and a family of global minima at the spherical surface defined by

$$\Phi^2 = v^2, \quad v = \sqrt{2\mu^2/\lambda}. \quad (7.10)$$

As discussed above, all of these points are related by $O(N)$ symmetry and are therefore equivalent. Without prejudice we can choose any representative, e.g.

$$\Phi_0 = (0, \dots, 0, v). \quad (7.11)$$

We thus introduce the quantum fields η and ϕ^a with $a = 1, \dots, N - 1$

$$\Phi^a(x) = \begin{cases} \phi^a(x) & \text{for } a = 1, \dots, N - 1, \\ v + \eta(x) & \text{for } a = N. \end{cases} \quad (7.12)$$

Substitution in the Lagrangian yields the terms

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2}(\partial\phi)^2 - \frac{1}{8}\lambda(\phi^2)^2 \\ & -\frac{1}{2}(\partial\eta)^2 - \mu^2\eta^2 - \sqrt{\lambda\mu^2/2}\eta^3 - \frac{1}{8}\lambda\eta^4 \\ & + \mu^4/2\lambda - \sqrt{\lambda\mu^2/2}\phi^2\eta - \frac{1}{4}\lambda\phi^2\eta^2. \end{aligned} \quad (7.13)$$

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 $O(N - 1)$ symmetry of the scalars ϕ^a which enlarges to the original $O(N)$ 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.

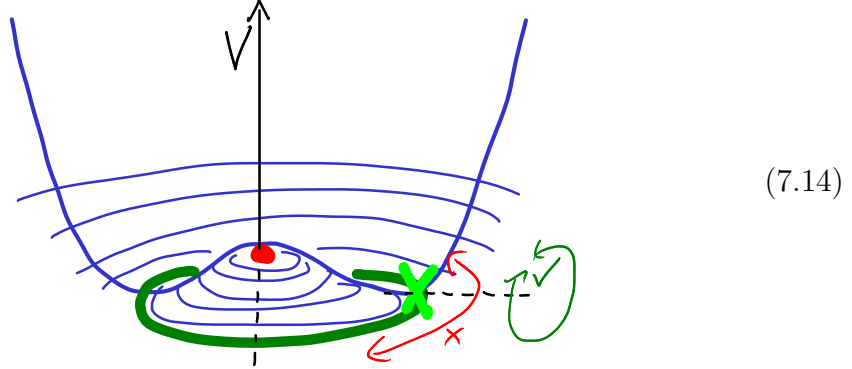
Goldstone's Theorem. Goldstone's theorem asserts that for every spontaneously broken continuous global symmetry there is one massless particle with the same quantum numbers. For ordinary symmetries the particle is a scalar, called the Goldstone boson.

One can understand the statement at the level of the Lagrangian of a model of scalar fields Φ^a :

²One could also consider linear combinations of ground states. In our situation, however, the mixing is strongly discouraged because transitions between the ground states have infinite energy.

³Fixing one of the ground states may appear paradoxical in the path integral framework: It 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.

- The ground state is specified by a (global) minimum Φ_0 of the scalar potential $V(\Phi)$.
- The mass matrix is given by the curvature (second-order derivatives) of the potential at this point.
- Suppose some of the symmetries map Φ_0 to a different point, and are therefore spontaneously broken. For each of these symmetries there is a corresponding direction in the space of Φ 's where the potential $V(\Phi)$ is constant or flat.



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 $O(N)$ symmetry to the group $O(N - 1)$ 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 fields, 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 $\vec{p} \rightarrow 0$, $N^\mu(p)$ should contain an excitation of the vacuum which has a non-trivial limit only for $E \rightarrow 0$. 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 ($a, b = 1, \dots, N$)

$$N_\mu^{ab} = \Phi^a \partial_\mu \Phi^b - \Phi^b \partial_\mu \Phi^a. \quad (7.15)$$

Expressed in terms of the new fields we find ($a, b = 1, \dots, N - 1$)

$$N_\mu^{ab} = \phi^a \partial_\mu \phi^b - \phi^b \partial_\mu \phi^a, \quad N_\mu^{Na} = v \partial_\mu \phi^a + \eta \partial_\mu \phi^a - \phi^a \partial_\mu \eta. \quad (7.16)$$

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.

7.2 Breaking of Gauge Symmetries

The picture changes again when the spontaneously broken symmetry belong 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

$$\mathcal{L}_{\text{SQED}} = -(D^\mu \phi)^* (D_\mu \phi) + \mu^2 \phi^* \phi - \frac{1}{4} \lambda (\phi^* \phi)^2 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu}. \quad (7.17)$$

The covariant derivative is defined as $D_\mu \phi = \partial_\mu \phi - iq A_\mu \phi$.

Here the ground state solution is specified by a point on a circle in the complex plane

$$\phi^* \phi = \frac{1}{2} v^2, \quad v = \sqrt{4\mu^2/\lambda}, \quad (7.18)$$

We could go ahead as before and introduce a shifted quantum field

$$\phi(x) = \frac{1}{\sqrt{2}} (v + \eta(x) + i\rho(x)). \quad (7.19)$$

This would lead to the same form of Lagrangian as above with the exception of the kinetic term which yields

$$-(D^\mu \phi)^* (D_\mu \phi) = -\frac{1}{2} (\partial\eta)^2 - \frac{1}{2} (\partial\rho)^2 - \frac{1}{2} q^2 v^2 A^2 + qv A_\mu \partial^\mu \rho + \text{interactions}. \quad (7.20)$$

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

$$A_\mu(x) = V_\mu(x) + (qv)^{-1}\partial_\mu\rho(x). \quad (7.21)$$

This shift changes the kinetic terms in the Lagrangian substantially

$$-(D^\mu\phi)^*(D_\mu\phi) = -\frac{1}{2}(\partial\eta)^2 - \frac{1}{2}q^2v^2V^2 + \text{interactions}. \quad (7.22)$$

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:⁴ 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

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x)) \exp(i\alpha(x)). \quad (7.23)$$

By means of a gauge transformation we can remove the phase from the field ϕ without changing any of the terms in the Lagrangian

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x)). \quad (7.24)$$

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

$$\begin{aligned} \mathcal{L}_{\text{SQED}} = & -\frac{1}{2}(\partial\eta)^2 - \frac{1}{4}v^2\lambda\eta^2 \\ & -\frac{1}{2}(\partial^\mu V^\nu)(\partial_\mu V_\nu) + \frac{1}{2}(\partial\cdot V)^2 - \frac{1}{2}q^2v^2V^2 \\ & + \frac{1}{16}v^4\lambda - \frac{1}{4}\lambda v\eta^3 - \frac{1}{16}\lambda\eta^4 - q^2vV^2\eta - \frac{1}{2}q^2V^2\eta^2. \end{aligned} \quad (7.25)$$

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 $\eta \rightarrow -2v - \eta$ 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.

⁴In October 2013 we will learn which names of the set Anderson, Brout, Englert, Guralnik, Hagen, Higgs, Kibble and 't Hooft we should “officially” associate to the effect.

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.⁵ 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.⁶

Renormalisation. We could now proceed and quantise the above Lagrangian. The model includes a massive vector field whose propagator is

$$G^{\mu\nu}(p) = \frac{\eta^{\mu\nu} + p^\mu p^\nu / m^2}{p^2 + m^2}. \quad (7.26)$$

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 really 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 μ^2 is positive or negative.

⁵In the above examples, this is the “magnitude” direction of the scalar potential.

⁶In 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.

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

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\rho(x)). \quad (7.27)$$

The momentum terms therefore read

$$-(D^\mu \phi)^*(D_\mu \phi) = -\frac{1}{2}(\partial\eta)^2 - \frac{1}{2}(\partial\rho)^2 - \frac{1}{2}q^2v^2A^2 + qvA_\mu\partial^\mu\rho + \text{interactions.} \quad (7.28)$$

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 $G[A, \Omega]$

$$G = \partial^\mu A_\mu - \Omega - \xi qv\rho. \quad (7.29)$$

Next we insert a delta functional together with a compensating functional determinant into the path integral

$$\int D\alpha \det(\delta G'/\delta\alpha) \Delta[G']. \quad (7.30)$$

We then eliminate the delta functional by integrating over the auxiliary field Ω with Gaussian weight $S_\Omega = -\frac{1}{2}\xi^{-1} \int d^4x \Omega^2$. This leads to a gauge fixing term

$$\mathcal{L}_{\text{gf}} = -\frac{1}{2}\xi^{-1}(\partial \cdot A)^2 - \frac{1}{2}\xi q^2v^2\rho^2 + qv\partial^\mu A_\mu\rho. \quad (7.31)$$

The argument of the functional determinant is given by the gauge variation of G

$$\delta G = \partial^2\delta\alpha - \xi q^2v(v + \eta)\delta\alpha. \quad (7.32)$$

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, \bar{C} . Their Lagrangian reads

$$\mathcal{L}_{\text{gh}} = \bar{C}(\partial^2 - \xi q^2v^2 - \xi q^2v\eta)C. \quad (7.33)$$

Collecting the various Lagrangians, we find the kinetic terms

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{\text{SQED}} + \mathcal{L}_{\text{gf}} + \mathcal{L}_{\text{gh}} \\ &= -\frac{1}{2}(\partial^\mu A^\nu)(\partial_\mu A_\nu) + \frac{1}{2}(1 - \xi^{-1})(\partial \cdot A)^2 - \frac{1}{2}q^2v^2A^2 \\ &\quad - \frac{1}{2}(\partial\eta)^2 - \frac{1}{4}\lambda v^2\eta^2 \\ &\quad - \frac{1}{2}(\partial\rho)^2 - \frac{1}{2}\xi q^2v^2\rho^2 - \partial^\mu \bar{C}\partial_\mu C - \xi q^2v^2\bar{C}C \\ &\quad + \text{interactions.} \end{aligned} \quad (7.34)$$

The gauge field needs further scrutiny, let us write its propagator

$$G^{\mu\nu}(p) = \frac{1}{p^2 + q^2v^2} \left(\eta^{\mu\nu} - (1 - \xi) \frac{p^\mu p^\nu}{p^2 + \xi q^2v^2} \right). \quad (7.35)$$

This propagator has several interesting features:

- For large momenta it decays quadratically.
- No mass scale appears in the denominator on its own.
- For $\xi \rightarrow \infty$ 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

$$p_\mu G^{\mu\nu}(p) = -\frac{\xi p^\nu}{p^2 + \xi q^2 v^2}. \quad (7.36)$$

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, \bar{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 $\xi \rightarrow \infty$ 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 $\xi \rightarrow \infty$ 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

$$n \rightarrow p^+ + e^- + \bar{\nu}. \quad (7.37)$$

At the level of constituent quarks the process can be understood as the transition of a down into an up quark

$$d^{-1/3} \rightarrow u^{+2/3} + e^- + \bar{\nu}. \quad (7.38)$$

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^-

$$d^{-1/3} \rightarrow u^{+2/3} + W^- \rightarrow u^{+2/3} + e^- + \nu. \quad (7.39)$$

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 $SU(2) \times U(1)$ where the $U(1)$ factor is required to obtain electromagnetism after symmetry breaking.⁷

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 $SU(2) \times U(1)$. We denote the associated gauge potentials by the 2×2 traceless matrix W_μ and the field B_μ acting as a 2×2 matrix proportional to unity. The coupling constants for these groups are denoted by g and g' , respectively. We define the corresponding field strengths as

$$W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - ig[W_\mu, W_\nu], \quad (7.40)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (7.41)$$

The contribution to the Lagrangian reads

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{2} \text{Tr} W^{\mu\nu} W_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu}. \quad (7.42)$$

Spontaneous symmetry breaking is achieved by a doublet of complex scalar fields denoted by H . It transforms in the (iso)spin- $\frac{1}{2}$ representation of $SU(2)$ and has a (hyper)charge of $\frac{1}{2}$ of $U(1)$. Hence, the covariant derivative of H is defined as

$$D_\mu H = \partial_\mu H - igW_\mu H - \frac{i}{2}g'B_\mu H. \quad (7.43)$$

⁷The $U(1)$ 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.

We design the scalar potential to have a global minimum at $|H|^2 = \frac{1}{2}v^2$. The scalar contribution to the Lagrangian therefore reads

$$\mathcal{L}_{\text{scalar}} = -(D^\mu H)^\dagger (D_\mu H) - \frac{1}{2}\lambda(H^\dagger H - \frac{1}{2}v^2)^2. \quad (7.44)$$

We discuss spontaneous symmetry breaking, 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

$$H(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \eta(x) \end{pmatrix}, \quad (7.45)$$

where $\eta(x)$ is a real field called the Higgs field. Let us furthermore parametrise the SU(2) gauge field as

$$W = \frac{1}{2} \begin{pmatrix} +W^0 & \sqrt{2}W^+ \\ \sqrt{2}W^- & -W^0 \end{pmatrix}. \quad (7.46)$$

We can now extract the masses for the various fields by extracting the quadratic contributions to the Lagrangian

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2}W^{-,\mu\nu}W_{\mu\nu}^+ - \frac{1}{4}W^{0,\mu\nu}W_{\mu\nu}^0 - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \\ & - \frac{1}{4}g^2v^2W^- \cdot W^+ - \frac{1}{8}v^2(gW^0 - g'B)^2 \\ & - \frac{1}{2}(\partial\eta)^2 - \frac{1}{2}\lambda v^2\eta^2 + \text{interactions}. \end{aligned} \quad (7.47)$$

Now we observe mixing between the fields W^0 and B . This can be resolved by a rotation of the basis of gauge fields⁸

$$\begin{pmatrix} A \\ Z^0 \end{pmatrix} = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ -g' & g \end{pmatrix} \begin{pmatrix} B \\ W^0 \end{pmatrix}. \quad (7.48)$$

The coupling constants are commonly parametrised using the Weinberg angle θ_W

$$g = g_{\text{EW}} \cos \theta_W, \quad g' = g_{\text{EW}} \sin \theta_W. \quad (7.49)$$

After the rotation we find

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2}W^{-,\mu\nu}W_{\mu\nu}^+ - \frac{1}{4}Z^{0,\mu\nu}Z_{\mu\nu}^0 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \\ & - \frac{1}{4}g^2v^2W^- \cdot W^+ - \frac{1}{8}v^2g_{\text{EW}}^2(Z^0)^2 \\ & - \frac{1}{2}(\partial\eta)^2 - \frac{1}{2}\lambda v^2\eta^2 + \text{interactions}. \end{aligned} \quad (7.50)$$

Now the mass terms are easily read off

$$m_W = \frac{1}{2}g_{\text{EW}} \cos(\theta_W) v, \quad m_Z = \frac{1}{2}g_{\text{EW}} v, \quad m_\eta = \sqrt{\lambda} v. \quad (7.51)$$

The field A remains 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

⁸The orthogonality of the transformation preserves the structure of the derivative terms.

component of the SU(2). It represents the electromagnetic field. It makes sense to inspect the covariant derivative of the Higgs field

$$gW + \frac{1}{2}g'B = \frac{g_{EW}}{2} \begin{pmatrix} \cos(2\theta_W)Z^0 & \sqrt{2}\cos(\theta_W)W^+ \\ \sqrt{2}\cos(\theta_W)W^- & Z^0 \end{pmatrix} + \frac{g_{EW}}{2} \sin(2\theta_W) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} A. \quad (7.52)$$

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 +1.⁹

$$H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix}. \quad (7.53)$$

- 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 $\frac{1}{2}$ representation) and of the hypercharge.
- The coupling strength of the electromagnetic field is given by

$$q = \frac{1}{2}g_{EW} \sin(2\theta_W). \quad (7.54)$$

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,

$$\psi = \psi_L + \psi_R, \quad \psi_{L/R} = \frac{1}{2}(1 \pm \gamma^5)\psi. \quad (7.55)$$

⁹These 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^{10 11}

$$i\bar{\psi}\gamma^\mu\partial_\mu\psi = i\bar{\psi}_L\gamma^\mu\partial_\mu\psi_L + i\bar{\psi}_R\gamma^\mu\partial_\mu\psi_R. \quad (7.56)$$

If we let the left and right components transform in different representations of the gauge group, we naturally obtain chiral couplings

$$\begin{aligned} & i\bar{\psi}_L\gamma^\mu D_\mu^L\psi_L + i\bar{\psi}_R\gamma^\mu D_\mu^R\psi_R \\ &= \bar{\psi}\gamma^\mu\left(i\partial_\mu + \frac{1}{2}gA_\mu^a(T_a^L + T_a^R)\right)\psi + \frac{1}{2}gA_\mu^a\bar{\psi}\gamma^5\gamma^\mu(T_a^R - T_a^L)\psi. \end{aligned} \quad (7.57)$$

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- $\frac{1}{2}$ 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	field	I_3	Y	C
$L_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$-\frac{1}{2}$	1	ν_R	0	0	1
$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$+\frac{1}{6}$	3	e_R	0	-1	1
				u_R	0	$+\frac{2}{3}$	3
				d_R	0	$-\frac{1}{3}$	3

(7.58)

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) \times 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.

¹⁰Note 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 U(N)_L \times U(N)_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).

¹¹In 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

$$\bar{\psi}\psi = \bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R. \quad (7.59)$$

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.¹²

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	(7.60)
$H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix}$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$+\frac{1}{2}$	1	

The Higgs field can couple via Yukawa interactions to the spinor fields. There are 4 gauge-invariant couplings

$$\bar{\nu}_R H^\top \varepsilon L_L, \quad \bar{e}_R H^\dagger L_L, \quad \bar{u}_R H^\top \varepsilon Q_L, \quad \bar{d}_R H^\dagger Q_L, \quad (7.61)$$

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.¹³

Anomalies. The electroweak theory involves chiral couplings of a gauge field to fermions. This interaction is a 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	field	I_3	Y	C	(7.62)
$L_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$-\frac{1}{2}$	1	ν_R	0	0	1	
$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$+\frac{1}{6}$	3	e_R	0	-1	1	
				u_R	0	$+\frac{2}{3}$	3	
				d_R	0	$-\frac{1}{3}$	3	

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


(7.63)

¹²In fact such a mass term is appealing for the see-saw mechanism which can explain the smallness of neutrino masses.

¹³Since 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 is 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 indices of d^{abc} belongs to U(1), the other two to the symmetric invariant form k^{bc} . These mixed anomalies can arise for U(1) \times SU(2) and for U(1) \times 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

$$A_{U(1)} \sim \text{Tr}_L Y^3 - \text{Tr}_R Y^3. \quad (7.64)$$

This combination surprisingly equals zero

$$2 \cdot 1 \left(-\frac{1}{2}\right)^3 + 2 \cdot 3 \left(+\frac{1}{6}\right)^3 - (0)^3 - (-1)^3 - 3 \left(+\frac{2}{3}\right)^3 - 3 \left(-\frac{1}{3}\right)^3 = 0. \quad (7.65)$$

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

$$A_{U(1) \times SU(2)} \sim \text{Tr}_L Y \quad (7.66)$$

Here we find zero

$$\left(-\frac{1}{2}\right) + 3 \left(-\frac{1}{2}\right) = 0. \quad (7.67)$$

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

$$A_{U(1) \times SU(3)} \sim \text{Tr}_{Q,L} Y - \text{Tr}_{Q,R} Y. \quad (7.68)$$

Again, this combination equals zero

$$2 \left(+\frac{1}{6}\right) - \left(+\frac{2}{3}\right) - \left(-\frac{1}{3}\right) = 0. \quad (7.69)$$

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

$$A_{U(1) \times \text{gravity}} \sim \text{Tr}_L Y - \text{Tr}_R Y. \quad (7.70)$$

This equals

$$2 \left(-\frac{1}{2}\right) + 2 \cdot 3 \left(-\frac{1}{2}\right) - (0) - (-1) - 3 \left(+\frac{2}{3}\right) - 3 \left(-\frac{1}{3}\right) = 0. \quad (7.71)$$

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.¹⁴
- 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.

¹⁴In fact, the required invariance of Yukawa couplings implies several anomaly cancellations. Therefore, the cancellation of mixed anomalies are a consequence of the quarks and leptons being massive.